EFFECT OF PROSODY ON WORD LIST RECALL

EFFECT OF WORD STRESS PATTERNS ON THE SERIAL RECALL OF WORD LISTS

BY LAN HU, B.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2023)

Hamilton, Ontario (Department of Linguistics and Languages)

TITLE: Effect of word stress patterns on the serial recall of word lists

AUTHOR: Lan Hu, B.A. (McMaster University)

SUPERVISOR: Dr. Elisabet Service

NUMBER OF PAGES: x, 60

Lay Abstract

Many daily experiences require us to remember a list of items, such as drafting a grocery list and reciting phone numbers. How well we can remember a list can be influenced by various characteristics of the list. This thesis looks at a seldom studied characteristic: prosody. Prosody in the English language can be realized via word stress, which is the amount of emphasis we place on certain parts of a word. This thesis explores the impact of word stress on our ability to remember a list of common English words. It reports data from an online survey and a laboratory experiment. Results show that mixed word stress patterns in a list lead to better memory for said list. Overall, this thesis offers new suggestions on the role of prosody in memory. As well, it offers a novel set of data that supports current theories in memory.

Abstract

This thesis examines the effects of prosody on serial recall. Serial recall is an experimental task commonly used to evaluate the capacity of short-term memory. The Working Memory model by Baddeley and Hitch is a theoretical framework that describes the inner operation of short-term memory. Its hierarchies are supported by empirical evidence, but details of the core mechanisms remain unclear. In an attempt to refine the framework, this thesis investigated prosody as a factor in serial recall accuracy. Two behavioural experiments were conducted on native speakers of Canadian English. In the first experiment, the explicit awareness of word stress was examined. Results showed a main effect of word stress type, where iambic words received higher stress identification scores compared to trochaic words. In the second experiment, an immediate serial recall task was used to examine serial recall of word lists. The lists consisted of disyllabic words from Canadian English sources. The lists had mixed or uniform stress patterns. A main effect of list stress patterns was found, where mixed lists elicited better recall of the order of list items compared to uniform lists. Overall, the present thesis offers a new interpretation on how word stress is represented in the short-term memory. It adds support to the proposed interaction between short-term and long-term memory.

Acknowledgements

I would like to express my gratitude to the many individuals who supported me through my thesis journey. My supervisor Dr. Elisabet Service and committee member Dr. Jean Saint-Aubin, whose expertise and guidance made this thesis possible. My colleagues Chelsea Whitwell and Heather Wild, who set great examples of academic excellence and perseverance. A special thanks to Dr. Daniel Pape for chairing my defence, as well as my research centre manager Dr. Maxime Dagenais for always being supportive and resourceful. And last but not least, my deepest and warmest gratitude for my best friend MD, my family WZ, and E.

Table of Contents

List of Tables

List of Figures

Declaration of Academic Achievement

The conceptualization of this thesis was conducted by myself, with collaboration with Dr. Elisabet Service. Analyses and interpretations on data were conducted by myself, with guidance from Dr. Elisabet Service. This thesis was written and edited by myself, with feedback from Dr. Elisabet Service, Dr. Jean Saint-Aubin, and Dr. Daniel Pape.

1 Introduction

Memory plays a vital role in the cognitive functions of human beings. The ability to process, retain, and retrieve information is core to many tasks and decisions; be it an infant recognizing their mom's voice, a child studying the contents taught in class, or an adult memorizing their banking passwords.

Given the pertinence of memory in human cognition, vast efforts have been made to better understand its operation. In the field of cognitive psychology, it is commonly accepted that memory functions differently in the long-term than in the short-term (Cowan, 2008 and Camina & Güell, 2017). Based on current understanding, long-term memory is responsible for storing and retrieving information in the long run, whereas short-term memory is responsible for processing and retrieving information in the moment. Short-term memory is frequently studied in conjunction with other fields for real-life applications. In clinical settings, deficits in short-term memory often co-occur with developmental disorders, such as dyslexia and autism (Preis et al., 1997; Swank, 1999; Jarrold et al., 2000; Clegg et al., 2005; Roesch & Chondrogianni, 2021; Ahufinger et al., 2022; Peeters et al. 2023). In educational settings, it has been reported that the capacity of verbal short-term memory can correlate to second language learning abilities and outcomes (Verhagen & Leseman, 2016; Révész et al., 2017; Michel et al., 2019). A clearer understanding of short-term memory would facilitate interventions in the above fields. In turn, this thesis examines representations of temporal characteristics of speech in verbal short-term memory. The thesis manipulates a lesser studied dimension of speech, called prosody. Section 1 provides key concepts and current issues on short-term memory and prosody. Sections 2 and 3 describe and interpret the experimental work conducted for this thesis. Section 4 summarizes contributions made by this thesis.

1.1 Models and Theories in Short-Term Memory

1.1.1 Working Memory

The Working Memory model by Baddeley and Hitch (1974) has become the most widely accepted framework in memory[1]. Working Memory was originally proposed as a short-term memory system containing three subsystems: the *central executive*, the *visuospatial sketchpad*, and the *phonological loop* (Figure 1a). The subsystems were proposed to interact with each other to process information, each performing distinct functions. In a review of the model, Baddeley (2010) suggested that the central executive governs the other subsystems. It directs attention to the information in need of processing, and delegates the information to the appropriate subsystem. The visuospatial sketchpad processes visual-spatial information. It can send feedback to the central executive to influence the direction of attention. In contrast, the phonological loop processes verbal and auditory information, but does not influence the other subsystems. Both the visuospatial sketchpad and the phonological loop are believed to have limited capacity, and hence can only store a finite amount of information for a short duration of time (Baddeley, 1992).

Figure 1a. The original Working Memory model proposed by Baddeley and Hitch (1974). Figure taken from Baddeley (2000).

[1] In this thesis, "short-term memory" is defined as the cognitive process itself. "Working Memory" is defined as the theoretical framework that describes (mostly) short-term memory.

More recently, the Working Memory model has been updated to include another component called the *episodic buffer* (Baddeley, 2000; Figure 1b). This buffer is proposed to be capable of combining information from multiple sensory modalities, as well as facilitating the transfer of information between short-term and long-term memory. In addition, based on clinical and experimental evidence, the phonological loop is now proposed to act as a gateway to long-term memory, particularly in the language domain (Baddeley et al. 1988; Baddeley et al., 1998).

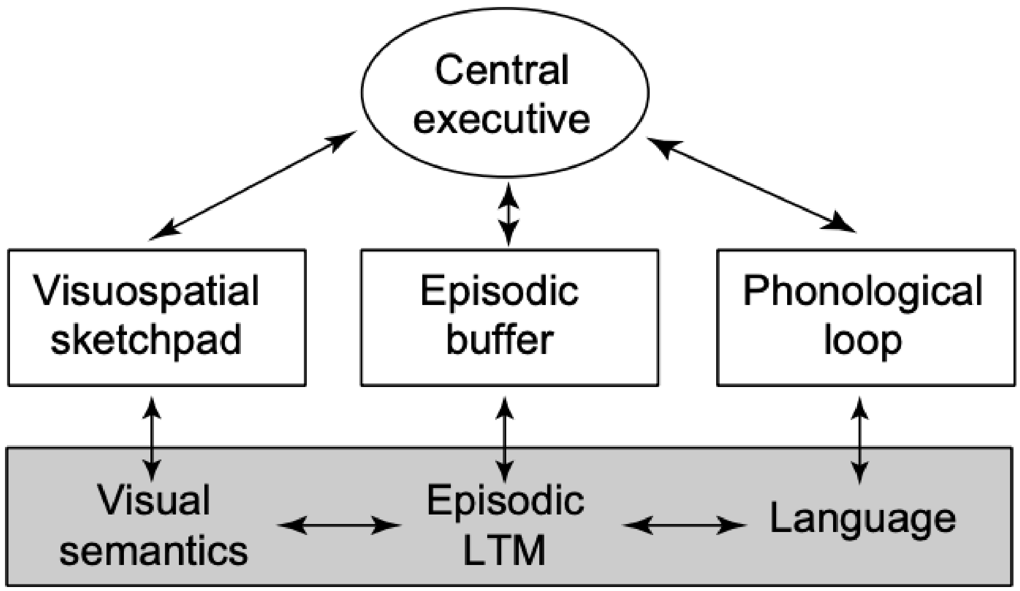

Figure 1b. The present Working Memory model, taken from Baddeley (2000). The shaded areas represent long-term memory and relevant cognitive domains. The unshaded areas represent short-term memory.

1.1.2 The Phonological Loop

In order to understand the relationship between language and the phonological loop, one must first recognize the operation of the phonological loop. It is proposed to consist of a *phonological short-term store* and an *articulatory rehearsal process* (Baddeley et al., 1984). Under this account, the short-term store attends to speech-based information and represents it at the phoneme level. A phoneme is the smallest unit of sound that contains a meaningful difference in speech ("Encyclopedia of Neuroscience," 2009). For example, the "h-" and "b-" sounds in the words "hat" and "bat" are phonemes, because a change in the sounds lead to a change in the meaning of the words.

The phonological representations decay rapidly from the short-term store, unless they are actively refreshed via the articulatory rehearsal process (Baddeley et al., 1984). The rehearsal is proposed to preserve representations in the short-term store as a person silently repeats those representations to oneself, akin to using inner speech. While rehearsal is not conceptualized as having a limited capacity, the short-term store is. Thus, only a limited amount of information can be retained in short-term memory at a time (Buchsbaum, 2013; Cowan, 2001).

The existence of the phonological loop is well-supported by neuropsychology. Clinical studies continue to identify brain areas that correspond to components inside the phonological loop (Aboitiz et al., 2010; Papagno et al., 2017). The present-day model of the phonological loop accounts for both visual and auditory information (Figure 2; Vallar & Papagno, 2002). It is grounded in the assumption that both orthographic and auditory information are converted into phonological representations before processing. This provides a theoretical interpretation for the clinically observed connection between language and the phonological loop.

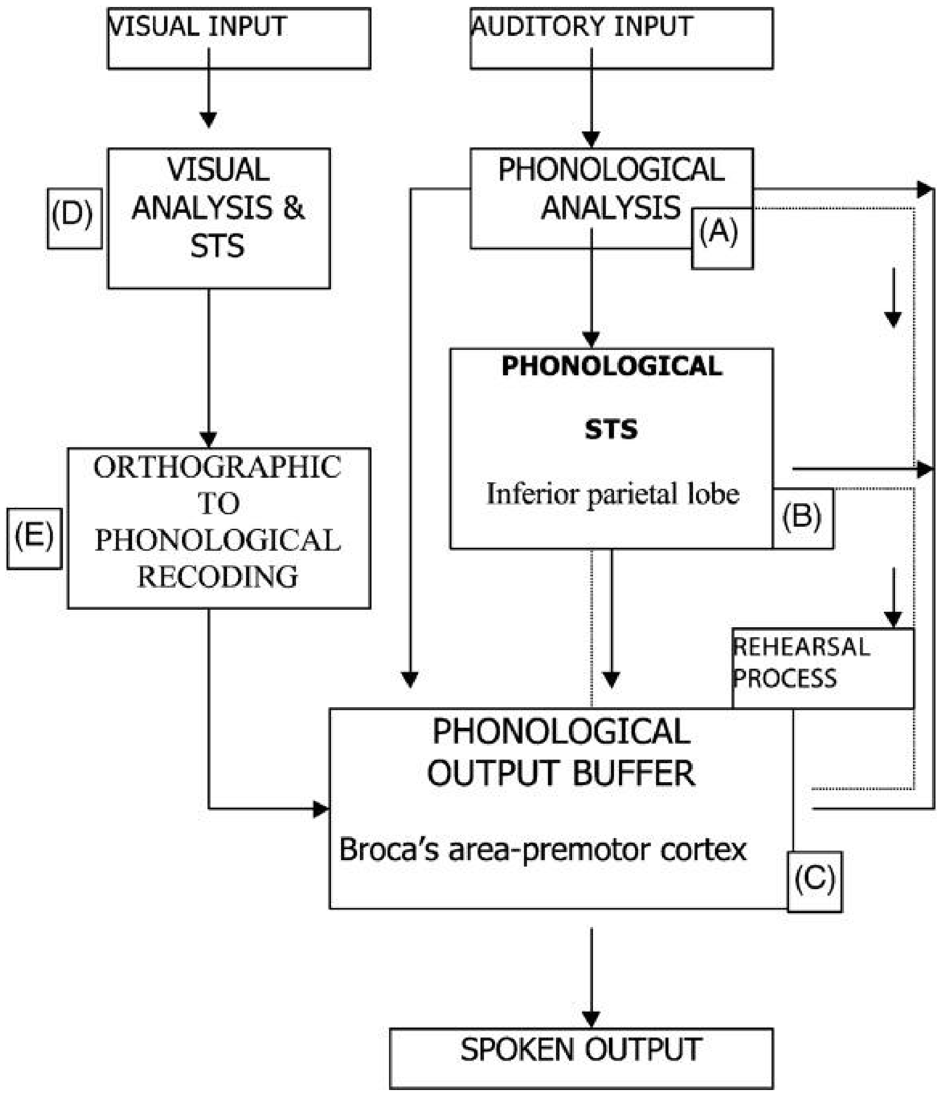

Figure 2. The proposed flow of operations inside the phonological loop*, taken from Vallar and Papagno (2002).*

1.1.3 The Word Length Effect

Aside from clinical evidence, behavioural evidence also supports the existence of the phonological loop. The first iteration of the Working Memory model documents a benchmark phenomenon called the Word Length Effect (Baddeley et al., 1975; Baddeley & Hitch, 1974). The effect describes the poorer recall accuracy for a list of long words compared to a list of short words. The cause of the effect was attributed to the longer amount of time needed to rehearse a list of long words. Since representations decay quickly from the phonological short-term store, not all long words in a list can be rehearsed in time before the decay starts. The unrehearsed words decay from the phonological short-term store, resulting in poor recall accuracy for lists of long words.

Since the initial reports of the Word Length Effect, however, several studies have documented its disappearance under certain experimental manipulations. One of the manipulations is known as the *articulatory suppression*. It is achieved by asking the participant to rehearse out loud irrelevant speech sequences (e.g., "hi-ya, hi-ya, hi-ya, …") during the memorization of lists of words (Richardson & Baddeley, 1975). Under the Working Memory model, the articulatory suppression is thought to have disrupted the articulatory rehearsal process, resulting in a dysfunctional phonological loop. Some authors, however, argue that the Word Length Effect is erased because the phonological loop is disabled in the presence of articulatory suppression; instead, lexical representations are summoned to recall words (Romani et al., 2005). These conflicting views pose the need to clarify the underlying mechanism of the phonological loop.

The Word Length Effect is also affected when factors other than articulatory suppression are manipulated for the stimulus words. These linguistic factors include phonological complexity (Service, 1998), phonemic word length (Lovatt et al., 2000), and certain lexical properties: notably neighbourhood size (Jalbert et al., 2011; Guitard et al., 2018) and semantic relatedness (Kowialiewski et al., 2022). These findings introduce alternative explanations to how the phonological loop functions internally.

Overall, the disappearance of the Word Length Effect raises questions to the processes and representations involved in verbal short-term memory. At the very least, it has become clear that word length is not the sole determinant to recall accuracy. Other factors, often correlated with phonemic word length, should be examined for their roles in recall. Findings will, in turn, reflect and refine the inner structure and functions of the phonological loop.

1.2 Prosody as a Factor Affecting Short-Term Memory Capacity

The capacity of verbal short-term memory is commonly measured through serial recall. Serial recall refers to a type of experimental task, where the participant is asked to recall a list of items in the original presentation order (Haberlandt, 2011). More recent studies show that the serial recall accuracy for word lists is affected by the arousal value of the words (Landry et al., 2022); but not the valence of the words (Bireta et al., 2012), biological sex of the participants (Voyer et al., 2021), or rehearsal strategies during memorization (Oberauer, 2019). One factor that has been seldom studied is the prosody of language, which is defined below.

1.2.1 Defining Prosody

In linguistics, the term prosody covers intonation, stress, and rhythm ("The Oxford Handbook," 2021). The current thesis primarily investigates stress at the word level. Word stress refers to the amount of emphasis a speaker puts on a syllable. In English, words may contain stressed and unstressed syllables. Word stress is usually achieved by manipulating acoustic qualities, so that a specific part of the word becomes more contrastive over other parts. The acoustic qualities available for manipulation include pitch, duration, loudness, vowel quality, as well as a combination of some or all of the above (Fudge, 2016; Ghosh & Levis, 2021). Longer words may contain a *primary stress* and a *secondary stress*. A primary stress is the most contrastive part of the word in regard to the above-mentioned acoustic qualities. In comparison, a secondary stress may have a noticeable prominence, but not enough to exceed the primary stress. As an example, the word "newspaper" contains three syllables, with "news-" carrying the primary stress, and "pa-" carrying the secondary stress. The remaining syllable, "-per", is unstressed.

Given that all multisyllabic English words contain stressed and unstressed syllables, the sequence of these syllables in a word creates different word stress types, called *trochaic* and *iambic*. Words with a trochaic stress type begin with a stressed syllable (e.g., "e-le-phant", in which the first syllable "e-" is stressed). In comparison, words with an iambic stress type begin with an unstressed syllable (e.g., "fan-tas-tic", in which the first syllable "fan-" is unstressed) (Kim et al., 2018). When words combine into lists or sentences, the trochaic and iambic stress types can further create groupings of stress, which contributes to prosody in speech[2]. Based on this logic, speech that contains repeating groups of prosodic patterns has *prosodic regularity*, whereas speech that contains varying prosodic patterns has *prosodic irregularity*.

1.2.2 Prosody as a Language Facilitator

There is a large literature demonstrating the key role of prosody in language perception and language learning in normally developing populations. In infants, prosody is found to be a very reliable cue for speech segmentation when listening to speech (Saffran et al., 1996b), even more so than other linguistic features such as phonology (Mattys et al., 1999). Similarly, prosody is found to aid the identification of word boundaries more than pitch does in the early years of first language acquisition (Suppanen et al., 2019). Infants born in English-speaking environments learn to segment trochaic words by 7.5 months of age, and iambic words by 10.5 months of age

[2] The author recognizes that the prosodic profile of speech comes from more factors than just word stress. As such, findings will be primarily interpreted in the context of word stress, rather than higher-level rhythmicity.

(Jusczyk et al., 1999). This suggests that the innate differentiation between trochaic and iambic word stress is developed at a very young age in native speakers of English. Throughout first language acquisition, prosody is used as a statistical cue for word learning, as well as an attentional cue for sentence-level comprehension (Erickson & Thiessen, 2015). Higher sensitivity to prosody in early childhood has been reported to predict the level of spelling and reading skills later in life (Holliman et al., 2014; Critten et al., 2021). In adults, the effects of prosodic processing can carry over to other linguistic domains such as reading comprehension (Colombo & Zevin, 2009), lexical acquisition (Christophe & Dupoux, 1996), and second language learning (Ordin & Nespor, 2016; Saksida et al., 2021).

When the awareness of prosody is underdeveloped, problems occur with language development and linguistic skills. In children with developmental language disorders, a connection has been found between difficulty in prosody perception and difficulty in phonological processing and reading comprehension (Goswami et al., 2011; Cumming et al., 2015). Furthermore, in populations with autism spectrum disorder, weakness in prosody perception and production is common among different forms of autism (Loveall et al., 2021; Diehl et al., 2019).

All in all, it is evident that prosody as a linguistic feature facilitates language development, and that sensitivity to prosody can serve as a predictor of language skills.

1.2.3 Prosody in Memory

The relationship between prosody and memory is less understood compared to that between prosody and language skills. There are only a handful of immediate serial recall experiments that manipulate prosody as a variable. In Boucher (2006), when recalling auditorily presented lists of French phrases, participants showed a preference during recall to insert word-level stress in phrases that exceeded four syllables in length. This supports the idea that functions of the verbal short-term memory involve the grouping and organization of prosody. In Archibald et al. (2009), when recalling auditorily presented lists of English pseudowords, participants showed higher recall accuracy for lists that contained word-level stress, compared to lists that contained neutral prosody. In another English pseudoword study by Morgan et al. (2014), participants recalled auditorily presented lists that contained trochaic pseudowords better than lists that contained iambic or monotone pseudowords. Together, the two pseudoword studies suggest that prosody, in particular word stress, may facilitate short-term memory performance. However, this conclusion cannot be extended to real English words. This is because among native speakers of English, the level of phonological awareness differs for real words versus pseudowords (Martin et al., 2020).

In addition to the scarcity of real word studies, little evidence is available to address which kind of prosody is beneficial for list recall; prosodic regularity or irregularity. On one hand, prosodic regularity can be argued to enhance memory due to its high predictability. Many behavioural studies show that predictable rhythms in speech leads to faster reactions (Shields et al., 1974; Pitt & Samuel, 1990; Quené & Port, 2005; Rassili & Ordin, 2022). The authors of these studies argue that the faster reaction times are resulted from strategic attention, where the listener extracts rhythmic regularity from the speech input, and focuses attention on those specific points in time. Such strategic attention may aid to memorize stimuli. On the other hand, prosodic irregularity can be argued to enhance memory due to its salience. Many neuropsychology studies demonstrate that unique auditory input attracts attention

(Large & Jones, 1999; McAuley & Jones, 2003; Ranganath & Rainer, 2003). The increased attention may lead to better encoding of the stimulus, and thus better recall performance.

The above contradictory possibilities were evaluated in Kimball et al. (2020). The experiment used a serial recall task and auditory word lists. The lists contained congruent or incongruent word stress. In the incongruent lists, all words were trochaic with the exception of one being iambic, or vice versa. The authors reported best recall accuracy for words that differed in their stress pattern from their context. A caveat to this finding, however, is that it can be explained by the von Restorff effect, also known as the isolation effect. The effect describes a memory advantage for stimuli that differ from the rest (Parker et al., 1998). While the effect is most commonly found for visual and orthographic stimuli, it has also been documented for acoustic stimuli in serial recall tasks (Holmes & Arbogast, 1979). Another limitation to Kimball and colleagues' work is that it reflects recall accuracy at a single word level. The conclusions cannot be extended to memory capacity at the list level, nor can they be used to further explain the underlying mechanism of verbal short-term memory. Hence, there is a critical need for a study, which 1) uses real English words, 2) examines the effects of prosodic regularity and irregularity on recall accuracy, 3) examines recall at the word list level, and 4) is able to expand its implications to the Working Memory model.

1.3 The Present Study

This thesis addresses the empirical gap identified in the Introduction. It uses English real words as stimuli and tests on native speakers of Canadian English. Unlike many existing studies that construct stimuli from American or British English, all stimuli in this thesis are developed from Canadian English sources. Additionally, this thesis is among the first to investigate the role of word stress patterns on word list recall. To achieve this, the thesis is organized into two separate experiments. *Experiment I* surveys the awareness of word stress among native Canadian English speakers. *Experiment II* tests recall accuracy for word lists with varying word stress patterns.

The research questions are as follows: 1) Does prosodic property, such as the list-internal consistency of word stress in a language input, lead to differences in recall accuracy? 2) How may the findings refine the current version of the Working Memory model, specifically the capacity-limiting factors affecting the phonological loop?

2 Experiment I: Word Stress Awareness

In order to understand how often native speakers of English attune to word stress, a survey is conducted to gauge their level of awareness when asked explicitly. Results from this survey informs stimulus development in the second experiment.

2.1 Methods

2.1.1 Participants

One hundred and forty-seven university students (25 males[3]; mean age=19.2 years, *sd*=2.2) were recruited through the McMaster PNB Research Participation System and the McMaster Linguistics Research Participation System. All participants provided written consent and received course credit as compensation for their participation. The study received ethics clearance from the McMaster University Research Ethics Board (MREB #6006).

2.1.2 Stimuli

Initially, 447 words were sourced from the MRC Psycholinguistic Database (hereby MRC). Parameters included number of syllables (*n*=2-4), familiarity (≥500), concreteness (≥500), and imageability (≥500). The words were cross-referenced with the Gage Canadian English Dictionary (hereby Gage). Duplicate entries, as well as words whose values in MRC contradicted values in Gage, were removed. Words that can function as proper nouns (e.g., "china") were removed. In addition, the words were cross-referenced with the Strathy Corpus of Canadian English (hereby Strathy). Words

[3] Such sex imbalance was typical of the user pool from the recruitment systems at the time. At the time of this study, around 21% of the users self-identified as male, 78% as female, and 1% as other or declined to answer. More is discussed in section 2.3.2.

of less than 99 Strathy frequency were removed. The remaining words were ranked by Strathy raw frequency. After prioritizing all available iambic words to be included, the top most frequent trochaic words were selected to make up a stimulus pool of 120 words (25 iambic; see Table 1 for a detailed count).

Table 1. Experiment I, counts of stimuli, split by conditions.

	iambic	**trochaic**	**Total**
2-syllable	8	80	88
3-syllable	16	14	30
4-syllable	1	1	2
Total	25	95	120

Six words from the stimulus pool were randomly chosen as the control stimuli. They were presented twice during the survey as an indicator of rater reliability of the participants. All other words were presented only once to each participant. Presentation order was individually randomized in the experimental trials, so that no participant saw the words in the same order as another participant.

2.1.3 Procedure

The survey was hosted online through LimeSurvey licensed under McMaster University. Participants were redirected to the survey after signing up for the study via one of the participation systems. Then, participants completed the survey at their preferred pace, with a recommended timeframe of 40 minutes. This timeframe was determined through pilot sessions and literature on the ideal length of online surveys, which was believed to be 20 minutes (Revilla & Ochoa, 2017). Since it was not possible to reduce the current study to such a short span while maintaining the desired data quantity, the experimental trials were placed at the start of the survey in order to maximize data quality (Neuert, 2021). In addition, multiple break opportunities were provided at fixed intervals throughout the survey to modularize its length.

The survey began with a consent letter, followed by a stress identification task, demographic questionnaire, and finished with a debrief and re-consent letter. In the stress identification task, participants were presented one word at a time in a multiple-choice question format. In each trial, participants were asked to identify where the main stress is in the presented word. Participants could either choose from one of the provided answer options, or they could enter a custom response. The answer options were the individual syllables of the given word, as indicated in Gage. Figure 3 below showcases an example of a trial. Responses were scored into quantitative data before analysis. In each trial, the response was compared to word stress listed in Gage. The response was assigned a score of 1 if it was a syllable with primary stress, 0.5 if it was a syllable with secondary stress, and 0 if it was an unstressed syllable. Table 2 shows how each response received a score.

Occasionally, participants had entered custom responses that differed in spelling, but were in fact identifying the same syllable as one of the provided answer options (e.g., for the stimulus *hockey*, entering a custom response of *hoc* when the provided answer options were *hock* and *ey*). In this case, the custom response was treated as the same as the corresponding answer option and scored accordingly (e.g., a response of *hoc* received the same score as *hock* would).

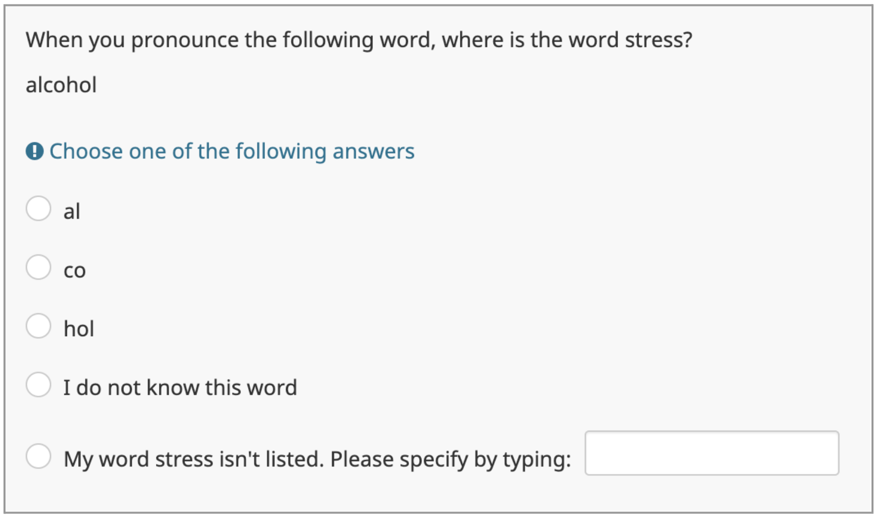

Figure 3. Experiment I, an example trial in the stress identification task.

Table 2. Experiment I, scoring criteria and a scoring example in the stress identification task.

Criteria: The response is...... stress option listed in Gage Canadian English dictionary.	**Score**
the primary	1.0
the secondary	0.5
neither the primary or secondary	0.0

Example scoring for the stimulus *newspaper*: Gage indicates *news-* as the primary stress, *-pa-* as the secondary stress, and *-per* as the unstressed syllable.

If the participant had responded......	**Score**
news	1.0
pa	0.5
per	0.0

In the stress identification task, all participants completed the practice trials (n=3) before proceeding to the experimental trials (n=120). During the practice trials, word stress was defined to the participants as "the part of the word that is pronounced with the most amount of strength, loudness, or emphasis". Participants were encouraged to complete the task with the way they personally pronounce the word. Three short break opportunities were provided to the participants, one after every 40 experimental trials.

In the demographic questionnaire, information about the participant's age, gender, education, language background, and musical experience was collected. A copy of the demographic questionnaire is attached in Appendix A.

2.1.4 Variables

A wide selection of word frequency indices was included as independent control variables. Notably, Strathy frequency covers time periods up to 2010, making it a more recent measure than Brown, Kucera-Francis, and Thorndike-Lorge frequencies. Strathy is therefore included to account for language variation over time, as well as to respect that the participant pool for this study were generally born in the 2000s.

Table 3. Experiment I, variables and their definitions.

Variable group	Variable	Definition	Value range
Dependent			
	Score	The quantified accuracy of a participant's response to judging which syllable in the word contains the word stress.	0.0, 0.5, or 1.0 [categorical]
	meanScore	For a given stimulus, the averaged accuracy of all Scores from all participants.	0.0-1.0 [continuous]
Independent			
	strathyFRQ	Strathy frequency. The occurrence frequency of a word in written and spoken form combined. Based on Strathy Corpus of Canadian English.	149–71545 [discrete]
	brownFRQ	Brown verbal frequency. The occurrence frequency of a word in spoken form. Based on London-Lund Corpus of English Conversation by Brown (1984).	1–533 [discrete]
	kf_FRQ	Kucera-Francis written frequency. The occurrence frequency of a word in written form. Based on the norms of Kucera and Francis (1967).	1–847 [discrete]
	tl_FRQ	Thorndike-Lorge frequency. Based on the L count of Thorndike and Lorge (1942).	7–3993 [discrete]
	nSYL	The number of syllables in a word. Based on Gage Canadian English Dictionary.	2, 3, or 4 [categorical]
	stressType	The order of stress and unstressed syllables in a word. Based on Gage Canadian English Dictionary.	Trochaic or Iambic [categorical]

2.1.5 Statistical Considerations

Statistical power. Due to the low prevalence of four-syllable words in the stimulus pool, statistical power was expected to be poor for their corresponding level in the *nSYL* variable. Therefore, four-syllable words were excluded from analysis (*n*=2).

Native language. Participants who self-reported their native language[4] to not be English were excluded from analysis (*n*=5).

Rater reliability. A rater reliability check was performed for every remaining participant based on six control stimuli. These stimuli were presented twice in the survey, and the two responses were compared against each other for consistency. Each consistent pair of responses earned the participant one rater reliability point, totalling a maximum of six points. Participants who received three or less reliability points were excluded from analysis (*n*=32).

Completion time. A descriptive analysis was done on the amount of time that the remaining participants took to complete the survey (*M*=2525 seconds or about 42 minutes; *sd*=2284 seconds or about 38 minutes). Participants whose survey completion time was two standard deviations above the mean were excluded from analysis (*n*=7). In addition, since the survey was recommended to be completed in 40 minutes, participants whose survey completion time was less than half of the recommended time were also excluded from analysis (*n*=18). While the latter exclusion criterion had rarely been evaluated in literature, it was done in precaution of poor data quality from participants speeding through the survey.

[4] Native language was defined to the participants as "the language [they] first pick up from [their] environments since birth". In cases of bilingualism – which was defined as "simultaneously picking up English and another language since birth" – participants were instructed to report English as their native language, and the other language as their second language.

2.2 Results

After implementing all the exclusion criteria, 85 participants (15 males; mean age =19.4 years, *sd*=2.6) were included in the analysis. Responses in the practice trials were not analyzed.

Descriptive statistics. The stress type of a word led to a difference in the aggregated judgements of word stress. On average, iambic words received a slightly higher score compared to trochaic words (M_{iambic}=.76, sd_{iambic}=.11 compared to $M_{trochaic}$=.71, $sd_{trochaic}$=.11). Within each stress type, the number of syllables in a word led to further differences. Disyllabic words received higher scores compared to trisyllabic words (Figure 4).

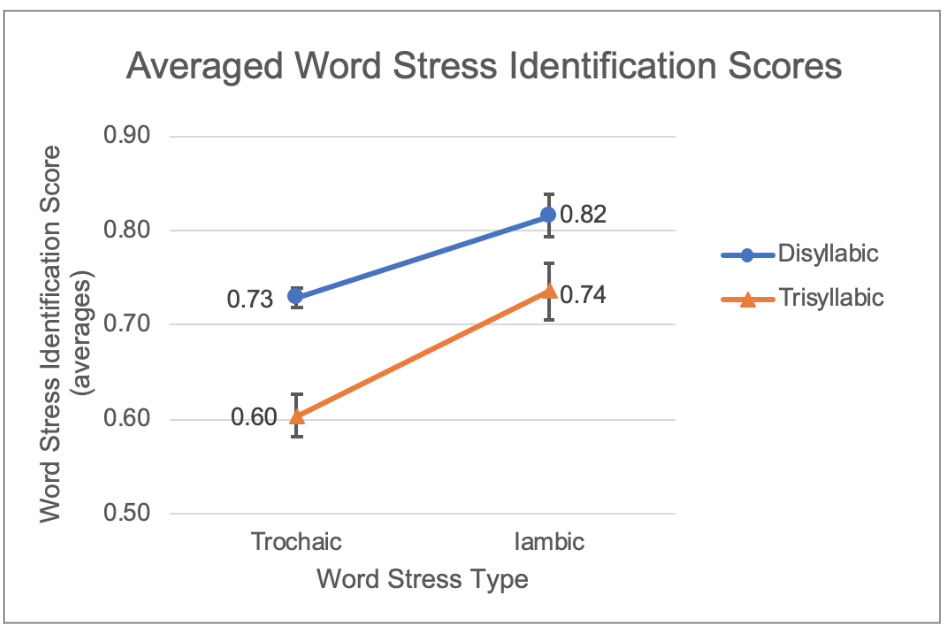

Figure 4. Experiment I, averaged word stress identification scores in each condition group. Error bars represent standard error.

ANOVA. An ANOVA was conducted between the dependent variable *meanScore* and the fixed factors *nSYL* and *stressType*. Both fixed factors were found to have a significant effect on *meanScore* (p<.001 for both). The interaction between *stressType* and *nSYL* did not reach significance (F<1).

Table 4. ANOVA for the dependent variable meanScore *and the fixed factors,* nSYL *and* stressType.

	Sum of Squares	df	Mean Square	F	p
nSYL	0.15334	1	0.15334	16.213	<.001
stressType	0.17762	1	0.17762	18.780	<.001
nSYL * stressType	0.00768	1	0.00768	0.812	0.369
Residuals	1.07821	114	0.00946	-	-

Correlation. A Spearman correlation was computed between the dependent variable *meanScore* and the number of syllables and the frequency indices. Spearman's rho was used over Pearson's r, as many of the frequency measures were ordinal rather than interval. Results revealed a significant, weak, negative correlation between *meanScore* and *nSYL* (r_s=-.222, p=.016). None of the frequency-measuring independent variables had significant correlation with *meanScore*.

Table 5. Correlation between the dependent variable meanScore *and all the independent variables*. Significance is shaded in grey.

		strathyFRQ	brownFRQ	kf_FRQ	tl_FRQ	nSYL
meanScore	Spearman's rho	-0.153	0.117	-0.167	-0.102	**-0.222**
	p-value	0.097	0.293	0.073	0.280	**0.016**

2.3 Discussion

Experiment I asked native speakers of Canadian English to identify word stress in commonly used English words via an online survey. There were three main findings. First, the stress type of a word had a significant effect on the accuracy of word stress identification. Second, the number of syllables within a word had a significant effect on the accuracy of word stress identification. Third, the frequency indices of words did not have any significant correlation with the accuracy of word stress identification. Interpretations follow in the order of the findings.

2.3.1 Interpretation of the Results

Main effect of stress type. Iambic words elicited higher identification scores than trochaic words. In English, iambic words naturally occur less often compared to trochaic words, at about a 1:4 ratio (Graf Estes & Bowen, 2013; Mateu & Sundara, 2022). This means the average native English speaker encounters fewer iambic words in their lifetime. According to learning theories such as statistical learning and exposure-based perceptual learning (Saffran et al., 1996a; Watanabe et al., 2001), linguistic features infrequent in the input are difficult to learn. Therefore, a speaker who has less exposure to the iambic stress type would be less aware of that stress feature, leading to lower identification scores. A contrary argument can be made, however, that the relative rarity of iambic words draws attention upon itself. Evidence from visual perception studies suggests that novel stimuli can enhance transient attention, and consequently, perception (Schomaker & Meeter, 2012; van Kesteren, 2012; Duszkiewicz et al., 2019; Reichardt et al., 2023). Nevertheless, these studies used visual stimuli such as geometric shapes and images, which are fundamentally different from the text stimuli used in this experiment. Based on the Working Memory model, visual inputs are processed in a different subsystem compared to language inputs. At the time of this thesis, few studies had examined the effect of novelty in word stress perception and identification. Still, the author suspects that the relative novelty of iambic words leads to heightened attention, if not the strategic use of rote memorization. In turn, speakers may gain more awareness of iambic word stress type, leading to higher identification scores for this stress feature.

Main effect of number of syllables. Disyllabic words elicited higher stress identification scores than trisyllabic words. This finding is expected, because disyllabic words had fewer available answer options than trisyllabic words. Even if participants were responding by chance, the chosen option was statistically more likely to be correct for disyllabic words (50% chance) compared to trisyllabic words (33% chance).

No interaction effect between stress type and number of syllables. The absence of significance for the interaction suggests that the main effect of stress type applies to all stimuli, regardless of their number of syllables. This is also demonstrated in Figure 4 by the similar slopes of the lines for disyllabic and trisyllabic words. Combined with everything discussed so far, it becomes clear that in a survey like Experiment I, stress type is a more meaningful factor than number of syllables to word stress awareness when scored at the word level rather than syllable level. The effect of the number of syllables is very likely a result of answer availability. This implies that when studying word stress, the number of syllables of the words should be controlled to specific value(s) rather than a range. Put simply, disyllabic words should be examined separately from trisyllabic words, quadrisyllabic words, and so on. Hence, Experiment II of this thesis only uses disyllabic words in the word lists.

No correlation between the frequency indices and word stress identification scores. Interestingly, this result was found across all four referenced frequency scales (Strathy, Brown, Kucera-Francis, and Thorndike-Lorge). This suggests that within a group of words that all have high frequency values, the frequency value of the individual words do not have a major influence on word stress awareness. Future studies can be done to examine whether the same finding persists for a group of words with a wider frequency range. For immediate implications, studies that wish to examine word stress should restrict their stimulus pool to high-frequency words, or statistically control for frequency effects via mixed effects modelling.

Comparison to previous findings is difficult, as existing literature on word stress perception largely centers around performance by second language learners of English. Therefore, findings of Experiment I were primarily used to inform the stimulus design of Experiment II, rather than answering the research questions of this thesis. Caveats to the interpretation of Experiment I are discussed below.

2.3.2 Limitations

Presentation mode. The words in Experiment were presented as text stimuli, which may not elicit the same stream of cognitive processes as auditory stimuli. In hindsight, an auditory perception task would have been a more direct tool to measure participants' awareness of word stress at the perceptual level. Nonetheless, a downside to an auditory perception task would have been the need to screen for hearing deficiencies in participants. In addition, some studies suggest a deficiency in word stress perception in special populations, such as those with autism, although this field of research is pioneering and the findings are often conflicting (Paul et al., 2005; Peppé et al., 2007; Kargas et al., 2016; reviews by McCann & Peppé, 2003; O'Connor, 2007). This adds the complication to potentially screen and exclude participants with autism spectrum disorder. A follow-up study can be done with the same set of stimulus words from Experiment I, but designed for an auditory perception task. Comparison of results will inform the validity of measuring word stress awareness via a survey design.

Sex imbalance in participants. In research methods literature, evidence exists for a disproportionately high participation rate for female participants compared to their male counterparts, notably in epidemiological studies (Markanday et al., 2013; Ryan et al., 2019) and survey studies (Smith, 2008; Mohajer & Jan, 2019; Becker, 2022). In the latter type of studies, sex imbalance persists even when the potential participants hold equal occupational status and social power. Given this precedence, the sampling rate of male and female participants in Experiment I should be considered typical of its participant pool.

3 Experiment II: Word List Recall

With a better understanding of word stress awareness among native speakers of Canadian English, a second experiment was conducted to study the impact of word stress patterns in a list. This experiment was designed as an immediate serial recall task. Given that disyllabic and trisyllabic words were found to evoke different levels of word stress awareness in Experiment I, this experiment only used disyllabic words in the lists, so that the number of syllables was eliminated as a confound.

3.1 Methods

3.1.1 Participants

Forty-seven university students (11 males[5]; mean age=19.0 years, *sd*=1.4) were recruited through the McMaster Psychology, Neuroscience and Behaviour Research Participation System. None had participated in Experiment I. All participants provided written consent and received either cash or course credit as compensation for their participation. The study received ethics clearance from the McMaster University Research Ethics Board (MREB #6006).

3.1.2 Stimuli

Word lists were created by using words that were sourced from MRC in a similar fashion to that in Experiment I. A total of 979 words with the following parameters: number of syllables (n=2), familiarity (350-700), concreteness (350-700), and imageability (350-700) were selected. After cross-referencing Gage and Strathy, the following types of words were removed: duplicates, homophones (e.g., “naval” and

[5] Similar to Experiment I, the sex imbalance here was typical of the pool of potential participants.

"navel"), words that have discrepant syllable counts in the three corpora (e.g., "tire" was labelled as disyllabic in MRC but monosyllabic in Gage), words that can have either trochaic or iambic stress type in the same word category (e.g., "trombone" as a noun can be pronounced as either "TROM-bone" or "trom-BONE"), and words that can function as proper names (e.g., "lily"). The remaining words were ranked by their Strathy word frequency per million, and the lower 25 percent of words were removed. This resulted in a word pool of 668 words (52 iambic).

From the word pool, words were randomly selected to form word lists. Each list contained six words. No word was ever used twice, either in the same list or in different lists. In total, 48 lists (24 with uniform stress patterns) were generated. Uniform lists contained trochaic words only. Mixed lists contained two iambic and four trochaic words. In mixed lists, words at serial position 1 and 6 were always trochaic, and words from position 2 to 5 shuffled between iambic and trochaic. Manipulation was intentionally done from position 2 to 5 and not at 1 or 6, to avoid muddling any findings with primacy and recency effects. Table 6 shows an example from each condition of the lists. To allow for re-examination and/or replication, a copy of all lists is attached in Appendix B.

A female native speaker of Canadian English recorded the lists into acoustic format. Recording took place in a soundproof booth in McMaster's Centre for Advanced Research in Experimental and Applied Linguistics (hereafter ARiEAL). The acoustic equipment included a RODE brand microphone, audio interface, and the computer software Audacity. Given that the recording material was human speech, the chosen sampling rate and bit rate was 44.1 kHz and 24 bit, respectively. The speaker was aware of the topic of the study and was instructed to articulate the lists as naturally they could.

During the recording session, the natural speech rate of the speaker was approximately one word per second. For words that can be pronounced with either trochaic or iambic stress type (e.g., "produce"), the speaker took care to pronounce them in the specific stress type indicated by the researcher. After the recording session, the audio files were loudness-normalized as necessary in Audacity to ensure they had consistent playback volume.

Table 6. Experiment II, word list design and examples.

Stress pattern	Sub stress pattern	Example
Mixed	112121	segment, weapon, **produce**, woodland, **suspect**, dairy [list_ID: WLR_3_VA]
	121211	candy, **exchange**, ceiling, **canoe**, honey, hour [list_ID: WLR_10_VA]
	121121	swimming, **cement**, onion, congress, **recruit**, banker [list_ID: WLR_17_VA]
Uniform	111111	carpet, lesson, outbreak, column, hammer, liar [list_ID: WLR_28_VA]

The initial iteration of the lists was labelled as version A. To counterbalance any ordering effects inside a list, a version B was created from version A by reordering the words in each list while maintaining the word selections and word stress patterns inside the list. The same reordering pattern was applied to all lists when transforming them from version A to B. For example, the first word in a list in version A always became the fourth word in a list in version B, the second word in version A always became the sixth word in version B, and so on. When recording the lists into audio format, version A and B were recorded under the same protocols. During data collection, participants were randomly assigned to either versions A or B. In the experimental trials, the presentation order of the lists was individually randomized, so that no participant listened to the lists in the same order as another participant.

3.1.3 Procedure

Testing took place in a quiet research lab in ARiEAL and lasted approximately 40 minutes per participant. The experiment started with a presentation of a consent form, followed by a word list recall task, the same demographic questionnaire as in Experiment I, and finally a debriefing and presentation of a re-consent form. The word list recall task was programmed using PsychoPy. The demographic questionnaire was hosted digitally using LimeSurvey licensed under McMaster University.

In the word list recall task, participants sat in front of a computer and listened to auditorily presented word lists through headphones. Participants were free to choose a comfortable viewing distance and audio volume for themselves. In each trial, participants saw a fixation cross at the centre of the computer screen for 500 milliseconds, then heard the word list for 7200 milliseconds. Immediately after, participants saw three exclamation marks at the centre of the computer screen. They had been instructed that this was a prompt to start saying out loud the list they had just heard. Participants were allowed as much time as they wanted to orally recall the list. After they had finished recalling, the participants pressed the spacebar on the computer keyboard to continue to the next trial.

Instructions for the word list recall task were verbally explained to the participants by a researcher. To ensure good understanding, participants were asked to paraphrase the instructions back to the researcher. Participants were instructed to maintain the order of the words to the best of their abilities. If they did not remember a particular word, they were instructed to say “blank” in place of that word.

All participants were presented with practice trials (*n*=3) and a clarification opportunity before accessing the experimental trials (*n*=47). A short break was provided to the participants after every eight experimental trials, and the computer screen displayed the calculated progress of the task in percentage (e.g., “You have completed 33% of the trials”).

Responses in the word list recall task were scored for quantitative data before analysis. Under the free scoring criteria, participants received a score of 1 for every recalled word, regardless of whether or not the word had been recalled in its serial position originally presented in the list. Under the serial scoring criteria, participants received a score of 1 for every recalled word, only if the word was recalled in its original serial position. All other words received 0. In both types of scoring, participants were required to enunciate all syllables in the word in order to receive a score of 1. No half scores were given. Table 9 shows how the same response could receive different scores under free versus serial scoring.

Table 7. Experiment II, the same response is assigned different scores under free and serial scoring.

	Free scoring	**Serial scoring**
Word list: carpet, lesson, outbreak, column, hammer, liar	1, 1, 0, 1, 0, 1	0, 0, 0, 0, 1, 0
Participant response: “lesson, carpet, liar, *blank*, hammer.”	[list_freeScore=4]	[list_serialScore=1]

3.1.4 Variables

The main independent variable was the stress pattern of the word lists (i.e., *list_stressPattern*), with two levels: uniform and mixed.

For recall accuracy variables, two types of scoring criteria were used (i.e., free and serial). Serial scoring conforms to criteria that are typically found in serial recall experiments: only words recalled in the serial position of presentation are considered correct. In contrast, free scoring is less strict and ignores the serial position of a response when judging the correctness.

Table 8. Experiment II, variables and their definitions.

Variable group	Variable	Definition	Value range
Dependent	list_freeScore & list_serialScore	For a given participant, the average number of words recalled in all lists. Calculated using free scoring or serial scoring.	0.00-6.00 [continuous]
	proportionRecalled_freeScore & proportionRecalled_serialScore	For a given serial position, the average of counts of recalled words divided by the counts of presented words. Calculated using free scoring or serial scoring.	0.00-1.00 [continuous]
Independent	list_stressPattern	The inclusion and exclusion of iambic words within a list. Uniform lists contain trochaic words only. Mixed lists contain trochaic and iambic words in certain orders.	Uniform or Mixed [categorical]
	list_subStressPattern	A direct representation of the ordering of trochaic and iambic words within a list. A numeral 1 represents a trochaic word, and a numeral 2 represents an iambic word.	111111, 112121, 121211, or 121121 [categorical]
	word_serialPosition	The original position of a word in the list when the list was presented to the participants.	1, 2, 3, 4, 5, or 6 [categorical]

3.1.5 Statistical Considerations

Native language. Participants who self-reported their native language[6] to not be English were excluded from analysis (*n*=2).

Procedural failure. Participants who frequently exhibited distracted behaviours (e.g., severe fidgeting, staring at the ceiling of the room), who declined to wear headphones, or who experienced major background noise were excluded from analysis (*n*=4). Due to a programming error, one stimulus (list_ID: WLR_1) from the word list recall task was consistently omitted for all participants in all versions of the experiment. It was, therefore, not possible to include this particular stimulus in the analysis.

3.2 Results

After implementing all exclusion criteria, 41 participants (8 males; mean age =18.9 years, *sd*=1.3) were included in the analysis. Responses in the practice trials were not analyzed.

Descriptive statistics. The stress pattern of the lists led to a difference in the scores. On average, mixed lists elicited higher scores compared to uniform lists. This was observed in both free scoring and serial scoring (Table 9 and Figure 5).

Table 9. Experiment II, average scores split by the stress pattern of lists.

	list_stressPattern	N	Mean	Median	Standard deviation
list_freeScore	uniform	41	4.03	4.00	0.57
	mixed	41	4.12	4.00	0.64
list_serialScore	uniform	41	2.82	2.79	0.97
	mixed	41	2.95	3.00	1.09

[6] Here, native language and bilingualism were defined to the participants in the same way as in Experiment I.

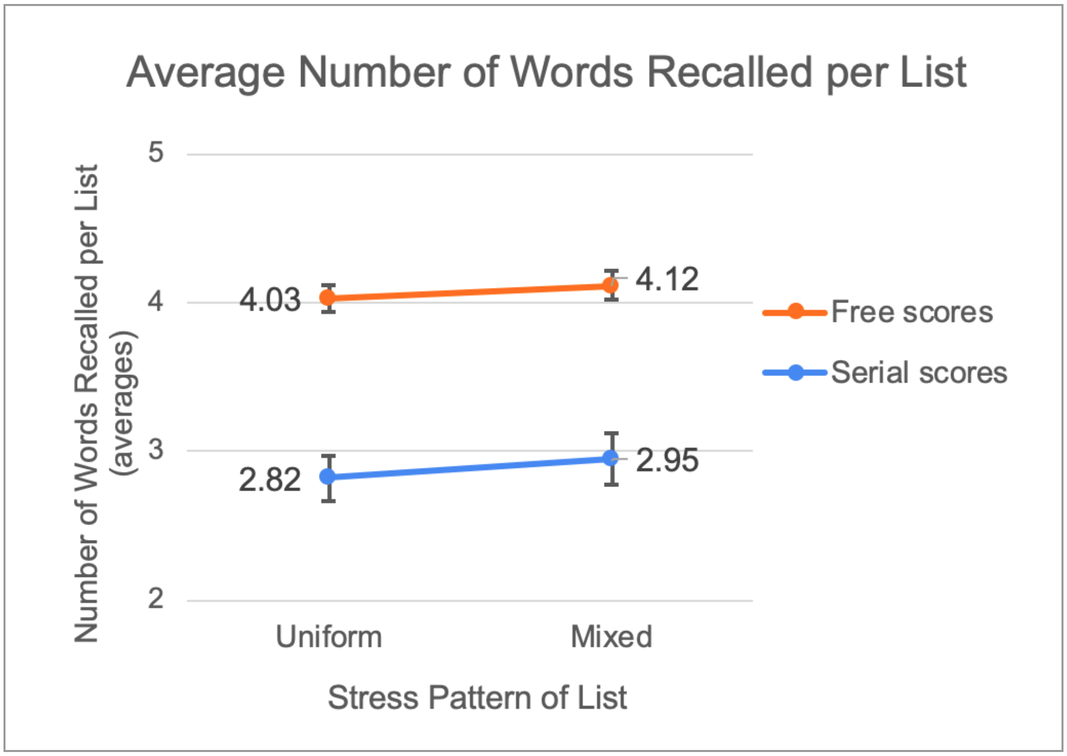

Figure 5. Experiment II, average number of words recalled per list for uniform and mixed lists. Error bars represent standard error.

The serial position of the word inside a list led to differences in the scores. Proportions of successful recalls were calculated for each serial position, using both free and serial scoring. Trends reflected primacy and recency effects that are typically observed in serial recall experiments. In Table 10a, the proportion declined through position 1 to 5, then slightly improved at position 6. Highest scores occurred at position 1, and lowest scores occurred at position 5. The same trend is captured in Figure 6, where all lines have a negative slope from position 1 to 5, then show slightly higher values at position 6.

When the serial position of words is further split by the stress pattern of lists, a difference in recall proportion can be seen at serial position 3. Specifically, lists with a mixed stress pattern led to higher proportions recalled here compared to lists with a uniform stress pattern. This difference was observed in both free and serial scoring (Table 10b). In Figure 6, the only set of non-contiguous error bars occurs at position 3. From serial position 2 to 5, mixed lists (dotted line) are shown to have either higher or similar proportions recalled compared to uniform lists (solid line).

Table 10a. Experiment II, descriptives for word recall proportion at each serial position.

	Free scoring		**Serial scoring**	
word_serialPosition	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**
1	0.88	0.33	0.84	0.37
2	0.74	0.44	0.64	0.48
3	0.67	0.47	0.45	0.50
4	0.61	0.49	0.34	0.47
5	0.53	0.50	0.24	0.43
6	0.65	0.48	0.38	0.48

Table 10b. Experiment II, descriptives for word recall proportion at each serial position, split by the stress pattern of lists. Statistics of interest are bolded.

		Free scoring		**Serial scoring**	
word_serial Position	**list_stress Pattern**	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**
1	uniform	0.89	0.32	0.85	0.35
	mixed	0.87	0.34	0.83	0.38
2	uniform	0.74	0.44	0.62	0.48
	mixed	0.75	0.44	0.65	0.48
3	**uniform**	**0.64**	**0.47**	**0.42**	**0.50**
	mixed	**0.70**	**0.47**	**0.49**	**0.50**
4	uniform	0.61	0.48	0.33	0.48
	mixed	0.61	0.49	0.35	0.47
5	uniform	0.51	0.50	0.23	0.44
	mixed	0.55	0.50	0.24	0.41
6	uniform	0.65	0.48	0.37	0.49
	mixed	0.65	0.47	0.38	0.47

Figure 6. Experiment II, word recall proportion at each serial position, split by the stress pattern of lists (uniform vs. mixed). Region of interest occurs at serial position 3. Error bars represent standard error.

ANOVA. Repeated measures ANOVAs were conducted for the dependent variables *list_freeScore* and *list_serialScore*. The repeated measures factor was *list_stressPattern*, with two levels (uniform and mixed). No between-subject factors or covariates were computed. Type 3 sum of squares were used. No sphericity corrections were needed, as the repeated measures factor only has two levels. Results revealed no significance under free scoring ($F(1,40)=3.68$, $p=.062$), but reached significance under serial scoring ($F(1,40)=4.39$, $p=.043$).

Table 11a. Experiment II, repeated measures ANOVA under free scoring.

Within Subjects Effects

	Sum of Squares	**df**	**Mean Square**	**F**	**p**	η^2_p
list_stressPattern	0.166	1	0.1660	3.68	0.062	0.084
Residual	1.807	40	0.0452	-	-	-

Between Subjects Effects

	Sum of Squares	**df**	**Mean Square**	**F**	**p**	η^2_p
Residual	27.5	40	0.688	-	-	-

Table 11b. Experiment II, repeated measures ANOVA under serial scoring.

Within Subjects Effects

	Sum of Squares	**df**	**Mean Square**	**F**	**p**	η^2_p
list_stressPattern	0.318	1	0.3184	4.39	0.043	0.099
Residual	2.901	40	0.0725	-	-	-

Between Subjects Effects

	Sum of Squares	**df**	**Mean Square**	**F**	**p**	η^2_p
Residual	81.8	40	2.05	-	-	-

3.3 Discussion

Experiment II asked native speakers of Canadian English to recall lists of common English words presented auditorily for a serial recall task. The main finding was that the stress pattern of word lists had a significant effect on order recall accuracy, with lists that included a mix of iambic and trochaic words being recalled better than lists with uniformly trochaic words.

3.3.1 Interpretation of the Results

Free versus serial scores. Before discussing the other results, it is important to recognize that the scoring criteria used in this experiment represent different aspects of serial recall. Free scoring reflects *item* recall accuracy, because it only concerns the contents of the items being remembered. In this experiment, the items were words inside a list. On the other hand, serial scoring reflects *item* and *order* recall accuracy. It accounts for not only the words themselves, but also the sequencing of the words inside the list. Taken together, the differences between free and serial scores represent outcomes of processing *order* information in addition to item information. Descriptively, serial scores were consistently lower than free scores, regardless of whether they were calculated at the list level (Table 9) or word level (Table 10a). This reflects a lower capacity for maintaining and/or recalling order information, relative to item information. Inferentially, the ANOVAs produced significance with serial scores but not free scores. This implies the list stress patterns made an impact on order recall accuracy, but not reliably on item recall accuracy. Overall, differences were observed between the two aspects of serial recall.

The question remains why the effect of prosody was seen more easily when serial scoring was used. A simple explanation is the temporal order nature of word stress. As defined in section 1.2.1, word stress is realized through syllables that contrast each other in a multitude of acoustic qualities. The temporal sequence of these syllables creates different word-level stress types (trochaic and iambic), and the temporal sequence of trochaic and iambic words creates various list-level stress patterns (uniform and mixed). Fundamentally, a list's stress pattern is *order* information unfolded throughout time. Because serial scoring captures memory for order information, but item scoring does not, serial scoring produces a more easily observed effect of list stress patterns.

Increasing evidence suggests that for language inputs, item information and order information are processed separately. Majerus (2019) provides an extensive review containing behavioural, clinical, and neuroimaging studies. He concludes that item information in the short-term memory directly interacts with linguistic knowledge in the long-term memory. In contrast, order information is arbitrary and non-linguistic, and therefore processed via non-linguistic mechanisms. These conclusions explain the consistently lower serial scores observed in the current experiment. If item information can draw extra help from linguistic knowledge, but order information cannot, then the latter is disadvantaged in maintaining itself in short-term memory. The same explanation can be applied to the ANOVA results. If order information is less solidly encoded, then it would be more susceptible to additional manipulations such as prosody. Therefore, the list stress patterns created enough impact on serial scores (i.e., significance for *item* and *order* recall accuracy), but the impact was not strong enough to reliably influence free scores (i.e., no significance for *item* recall accuracy).

Majerus is not the first author to suggest a direct link between short-term memory and long-term memory. In studies of serial recall tasks, Saint-Aubin and LeBlanc (2005) as well as Poirier et al. (2015) have shown a modulating effect of long-term representations on short-term memory. The long-term representations include word frequency and semantics, which are linguistic knowledge stored in the long-term memory. For the effect of word frequency in particular, it has been found that low-frequency words are better recalled when they are embedded in a list of high-frequency words, compared to a list of equally low-frequency words (Saint-Aubin & LeBlanc, 2005). This suggests the distinctiveness of the word relative to its surrounding context may boost recall accuracy. Similarly, one can speculate that a rarer word stress pattern contributes to the distinctiveness of a list, and therefore influences list-level recall accuracy. This interpretation is dissected below.

Word stress as a probabilistic feature of words. In a Japanese pseudoword recall study, Tanida et al. (2015) reported effects of prosody that were mediated by phonotactic frequency. Phonotactic frequency describes how commonly a sound, or a combination of sounds, is observed in a language (Hay, 2003). Again, Tanida and colleagues' results suggest that long-term knowledge, such as knowing the probability of hearing a particular acoustic feature of a word in a certain context, influences recall accuracy. The same logic applied to other acoustic features (e.g., word stress) can explain the current results. If word stress is processed in short-term memory based on probabilistic cues, then iambic words are recognized as a rarer cue than trochaic words in English. As a result, lists that contain a mix of iambic and trochaic words (i.e., mixed lists) obtain higher distinctiveness compared to lists that contain purely trochaic words (i.e., uniform lists). In turn, the distinctiveness of mixed lists leads to the overall higher recall accuracy at the list level (seen in Table 9).

The probabilistic interpretation seems viable. It is less susceptible to the von Restorff effect discussed in section 1.2.3, because the distinctiveness of the mixed lists does not equate isolated uniqueness. Throughout the experiment, participants heard near-equal amounts of mixed lists and uniform lists (n=23 and 24, respectively). Thus, the probabilistic interpretation is more likely to reflect the nature of representation of word stress in the short-term memory. The interpretation also fits with developmental evidence. As introduced in section 1.2.2, English-speaking infants detect and distinguish iambic and trochaic word stress quite early in development. Children and adult listeners continue to extract prosodic features from speech inputs and rely on them as probabilistic cues for language comprehension (Redford & Oh, 2016; Tyler & Cutler, 2009). In addition, the interpretation does not conflict with the neuropsychological evidence introduced in section 1.2.3, which describes increased attention to unique auditory input. While the distinctiveness of mixed lists may have evoked increased attention by the central executive, in theory, such attention is beneficial. It facilitates processing at the phonological loop, which upholds the interaction between short-term and long-term memory. Overall, the probabilistic interpretation corroborates Baddeley's revised Working Memory model (2000).

One more question remains, however, on why a descriptive difference between uniform and mixed lists was observed specifically at serial position 3 (Figure 6). At the time of this thesis, the author could not find any literature that reports the same serial position effect. A purely speculative answer is that in mixed lists, position 3 is the critical region where expectations of a regular prosodic pattern are broken. A major flaw to this speculation, however, is that it only accounts for one subtype of the mixed lists in Experiment II. As visualized in Table 12, only the subtype *112121* "surprises" the participants at position 3 by breaking the prosodic regularity with an iambic word. The other subtypes, *121211* and *121121*, actually contain an iambic word earlier in the list, at position 2. And arguably, the other subtypes break prosodic regularity not at the first occurrence of an iambic word, but rather later in the list, when the alternation between trochaic and iambic words is interrupted. The data in the present thesis was not coded for serial position analysis. To allow for such analysis, the raw data table needs to be expanded to six times its initial size. In future work, the author will analyze each subtype of the mixed lists separately and examine potential effects of serial position. Further, with word-level data coding, the effects of individual words (e.g., semantics and neighbourhood size of each word in a list) can also be taken into account using mixed-effects modelling.

Table 12. Experiment II, individual word type at serial position 3 (shaded in grey) in each subtype of mixed lists.

	word type at each serial position [#]					
list_subStressPattern	[1]	[2]	[3]	[4]	[5]	[6]
11**2**1**2**1	trochaic	trochaic	***iambic***	trochaic	***iambic***	trochaic
1**2**1**2**11	trochaic	***iambic***	trochaic	***iambic***	trochaic	trochaic
1**2**11**2**1	trochaic	***iambic***	trochaic	trochaic	***iambic***	trochaic

Finally, to directly answer the research questions of this thesis: 1) Yes, prosodic properties of a language input, such as list-internal word stress patterns, can lead to differences in recall accuracy. In particular, lists of mixed stress patterns elicited better serial recall accuracy. 2) The present findings offer more support to the notion that short-term memory and long-term memory are not standalone systems. While the connection between the two has been formalized in Baddeley's revised Working Memory model (2000), this thesis offers one explanation to how the two systems can interact. Experiment II offers preliminary evidence that short-term memory relies on long-term, probabilistic knowledge when processing prosodic information in language materials. The results, however, fail to explain the capacity-limiting factors affecting the phonological loop. It remains a complicated issue that invites more work.

3.3.2 Limitations

A few shortcomings remain to be mentioned for Experiment II. For one, the need for further analysis to explore the mystery of serial position 3 has been detailed above. Similar to Experiment I, the sex imbalance in participants may have created female-biased results. This is not a flaw of the study design but rather a limitation of the available participant pool, for the same reasons discussed in section 2.3.2. The other weaknesses are inspected below.

Lexical properties in stimulus words. While the author has exerted utmost effort to control for word frequency in the lists, the upper and lower boundaries of the frequency range were determined arbitrarily. Thus, the effects found in Experiment II are susceptible to other lexical properties of the words, such as their semantics. A post-hoc correction would be to collect native speakers' ratings on the semantic relatedness of words inside each list. Then, list-internal semantic relatedness can be used as a control variable in a re-analysis of Experiment II. Another lexical property that could have confounded the results is the neighbourhood size of the words. The author had intended to conduct an additional logistic mixed-effects analysis on recall of each word, which would be able to account for potential effects from the lexical properties. Due to timeline limitations, the analysis is left for future research.

Individual differences in participants. A tentative mixed-effects analysis explored individual differences in recall accuracy. Language and music aptitudes were quantified from qualitative responses to a demographic questionnaire. Musicianship, but not multilingualism, significantly influenced recall accuracy. These post-hoc exploratory analyses prompt the consideration that participants' sensitivity to prosody may vary greatly due to past experiences. An individual differences study may be conducted in the future to establish aptitude profiles for various types of prosodic tasks. The tasks can involve both linguistic and non-linguistic stimuli to better understand if prosodic aptitude is domain-specific or domain-general.

Presentation mode. The majority of the existing literature on serial recall reports results from orthographic stimuli that were presented visually. Notably, the probabilistic interpretation in the present discussion is based on works that used orthographic stimuli. In contrast, this experiment used auditory stimuli that offered the most direct representation of word stress. In cases of words that can be read with multiple stress types (e.g., "produce" as a noun versus verb), the auditory stimuli automatically enforce one option over the other. This eliminates the extra cognitive load that would have been required of the participant if they were to read a textual presentation of the word lists. While using auditory stimuli has a clear advantage, it also limits comparison to previous works. A follow-up study can be done with the same set of word lists, presented as text. Results will advise whether the prosodic effects found here survive a different modality of presentation.

4 Conclusions

This thesis examines the effect of prosody, in particular word stress, on the recall accuracy of word lists. It provides novel data in two behavioural experiments. Results from an online survey show that among native English speakers, iambic words are associated with higher levels of word stress awareness compared to trochaic words. Results from an immediate serial recall task show that the stress patterns of word lists can influence recall accuracy. Mixed lists are associated with better item and order recall accuracy compared to uniform lists. The latter finding is attributed to the higher distinctiveness of mixed stress patterns, relative to uniform stress patterns in English. It presents a new interpretation of word stress as a probabilistic cue in short-term memory tasks. Overall, the findings support the concept of direct interaction between short-term and long-term memory in human cognition.

References

Aboitiz, F., Aboitiz, S., & García, R. R. (2010). The Phonological Loop: A Key Innovation in Human Evolution. *Current Anthropology*, *51*(1), 55–65. https://doi.org/10.1086/650525

Archibald, L. M. D., Gathercole, S. E., & Joanisse, M. F. (2009). Multisyllabic nonwords: More than a string of syllables. *The Journal of the Acoustical Society of America, 125*(3), 1712–1722. https://doi.org/10.1121/1.3076200

Baddeley, A. (1992). Working Memory: The Interface between Memory and Cognition. *Journal of Cognitive Neuroscience, 4*(3), 281–288. https://doi.org/10.1162/jocn.1992.4.3.281

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review, 105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Hower (Eds.), *The psychology of learning and motivation* (pp. 47–89). Academic Press, Inc. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology, 36A*(2), 233–252. https://doi.org/10.1080/14640748408402157

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*(5), 586–595. https://doi.org/10.1016/0749-596X(88)90028-9

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575–589. https://doi.org/10.1016/S0022-5371(75)80045-4

Becker, R. (2022). Gender and Survey Participation. An event-history analysis of the gender effects of survey participation in a probability-based multi-wave panel study with a sequential mixed-mode design. *Methods, data, analyses, 16*(1), 3–32. https://doi.org/10.12758/mda.2021.08

Binder, M. D., Hirokawa, N., & Windhorst, U. (Eds.). (2009). *Encyclopedia of Neuroscience*. Springer, Berlin, Heidelberg. https://doi-org.libaccess.lib.mcmaster.ca/10.1007/978-3-540-29678-2_4553

Bireta, T. J., Guitard, D., Neath, I., & Surprenant, A. M. (2021). Valence does not affect serial recall. *Canadian Journal of Experimental Psychology, 75*(1), 35–47. https://doi.org/10.1037/cep0000239

Boucher, V. J. (2006). On the Function of Stress Rhythms in Speech: Evidence of a Link with Grouping Effects on Serial Memory. *Language and Speech, 49*(4), 495–519. https://doi.org/10.1177/00238309060490040301

Buchsbaum, B. R. (2013). The role of consciousness in the phonological loop: hidden in plain sight. *Frontiers in Psychology, 4*, 496–496. https://doi.org/10.3389/fpsyg.2013.00496

Camina, E., & Güell, F. (2017). The Neuroanatomical, Neurophysiological and Psychological Basis of Memory: Current Models and Their Origins. *Frontiers in Pharmacology, 8*, 438–438. https://doi.org/10.3389/fphar.2017.00438

Chen, A., & Gussenhoven, C. (Eds.). (2021). *The Oxford handbook of language prosody*. Oxford University Press.

Christophe, A., & Dupoux, E. (1996). Bootstrapping lexical acquisition: The role of prosodic structure. *Linguistic Review, 13*(3-4), 383–412. https://doi.org/10.1515/tlir.1996.13.3-4.383

Critten, S., Holliman, A. J., Hughes, D. J., Wood, C., Cunnane, H., Pillinger, C., & Deacon, S. H. (2021). A longitudinal investigation of prosodic sensitivity and emergent literacy. *Reading & Writing, 34*(2), 371–389. https://doi.org/10.1007/s11145-020-10077-7

Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders - a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry, 46*(2), 128–149. https://doi.org/10.1111/j.1469-7610.2004.00342.x

Colombo, L., & Zevin, J. (2009). Stress priming in reading and the selective modulation of lexical and sub-lexical pathways. *PloS One, 4*(9), e7219–e7219. https://doi.org/10.1371/journal.pone.0007219

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. The Behavioral and Brain Sciences, 24(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*(323–338). https://doi.org/10.1016/S0079-6123(07)00020-9

Cumming, R., Wilson, A., & Goswami, U. (2015). Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: a new look at a perceptual hypothesis. *Frontiers in Psychology, 6*, 972–972. https://doi.org/10.3389/fpsyg.2015.00972

Diehl, J. J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics, 30*(3), 385–404. https://doi.org/10.1017/S0142716409090201

Duszkiewicz, A. J., McNamara, C. G., Takeuchi, T., & Genzel, L. (2019). Novelty and Dopaminergic Modulation of Memory Persistence: A Tale of Two Systems. *Trends in Neurosciences, 42*(2), 102–114. https://doi.org/10.1016/j.tins.2018.10.002

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*(Sep), 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Fudge, E. (2015). *English Word-Stress*. Taylor and Francis. https://doi.org/10.4324/9781315674810

Ghosh, M., & Levis, J. M. (2021). Vowel Quality and Direction of Stress Shift in a Predictive Model Explaining the Varying Impact of Misplaced Word Stress: Evidence From English. *Frontiers in Communication, 6*. https://doi.org/10.3389/fcomm.2021.628780

Goswami, U., Wang, H.-L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal Sensory Deficits in Developmental Dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience, 23*(2), 325–337. https://doi.org/10.1162/jocn.2010.21453

Graf Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology, 114*(3), 405–417. https://doi.org/10.1016/j.jecp.2012.10.002

Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M., & Neath, I. (2018). Word Length, Set Size, and Lexical Factors: Re-Examining What Causes the Word Length Effect. Journal of Experimental Psychology. *Learning, Memory, and Cognition, 44*(11), 1824–1844. https://doi.org/10.1037/xlm0000551

Haberlandt, A.K. (2011). Serial Recall. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (Eds.), *Encyclopedia of Clinical Neuropsychology*. Springer, New York, NY. https://doi.org/10.1007/978-0-387-79948-3_1157

Hay, J. (2003). *Causes and consequences of word structure*. Routledge.

Holliman, A., Critten, S., Lawrence, T., Harrison, E., Wood, C., & Hughes, D. (2014). Modeling the Relationship Between Prosodic Sensitivity and Early Literacy. *Reading Research Quarterly, 49*(4), 469–482. https://doi.org/10.1002/rrq.82

Holmes, C. B., & Arbogast, R. (1979). An Auditory Von Restorff Effect. *The Journal of General Psychology, 101*(2), 199–204. https://doi.org/10.1080/00221309.1979.9920074

Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition, 39*(7), 1198–1210. https://doi.org/10.3758/s13421-011-0094-z

Jarmulowicz, L., Taran, V. L., & Seek, J. (2012). Metalinguistics, Stress Accuracy, and Word Reading: Does Dialect Matter? *Language, Speech & Hearing Services in Schools, 43*(4), 410–423. https://doi.org/10.1044/0161-1461(2012/11-0060)

Jarrold, C., Baddeley, A. D., & Hewes, A. K. (2000). Verbal Short-term Memory Deficits in Down Syndrome: A Consequence of Problems in Rehearsal? *Journal of Child Psychology and Psychiatry, 41*(2), 233–244. https://doi.org/10.1017/S0021963099005120

Kargas, N., López, B., Morris, P., & Reddy, V. (2016). Relations among Detection of Syllable Stress, Speech Abnormalities, and Communicative Ability in Adults with Autism Spectrum Disorders. *Journal of Speech, Language, and Hearing Research, 59*(2), 206–215. https://doi.org/10.1044/2015_JSLHR-S-14-0237

Kim, S., Kim, J., & Cho, T. (2018). Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics, 71*, 65–80. https://doi.org/10.1016/j.wocn.2018.07.004

Kimball, A. E., Yiu, L. K., & Watson, D. G. (2020). Word recall is affected by surrounding metrical context. *Language, Cognition and Neuroscience, 35*(3), 383–392. https://doi.org/10.1080/23273798.2019.1665190

Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2022). The semantic relatedness effect in serial recall: Deconfounding encoding and recall order. *Journal of Memory and Language, 127*(104377). https://doi.org/10.1016/j.jml.2022.104377

Landry, É. R., Guitard, D., & Saint-Aubin, J. (2022). Arousal affects short-term serial recall. *Canadian Journal of Experimental Psychology, 76*(2), 99–110. https://doi.org/10.1037/cep0000272

Large, E. W., & Jones, M. R. (1999). The Dynamics of Attending: How People Track Time-Varying Events. *Psychological Review, 106*(1), 119–159. https://doi.org/10.1037/0033-295X.106.1.119

Levy, B. A. (1971). Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10*(2), 123–132. https://doi.org/10.1016/S0022-5371(71)80003-8

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The Word-length Effect and Disyllabic Words. Quarterly Journal of Experimental Psychology (2006), 53(1), 1–22. https://doi.org/10.1080/713755877

Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112*, 122–133. https://doi.org/10.1016/j.cortex.2018.04.016

Markanday, S., Brennan, S. L., Gould, H., & Pasco, J. A. (2013). Sex-differences in reasons for non-participation at recruitment: Geelong Osteoporosis Study. *BMC Research Notes, 6*(1), 104–104. https://doi.org/10.1186/1756-0500-6-104

Martin, K. I., Lawson, E., Carpenter, K., & Hummer, E. (2020). English Word and Pseudoword Spellings and Phonological Awareness: Detailed Comparisons From Three L1 Writing Systems. *Frontiers in Psychology, 11*, 1309–1309. https://doi.org/10.3389/fpsyg.2020.01309

Mateu, V., & Sundara, M. (2022). Spanish input accelerates bilingual infants' segmentation of English words. *Cognition, 218*, 104936–104936. https://doi.org/10.1016/j.cognition.2021.104936

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and Prosodic Effects on Word Segmentation in Infants. *Cognitive Psychology, 38*(4), 465–494. https://doi.org/10.1006/cogp.1999.0721

McAuley, J. D., & Jones, M. R. (2003). Modeling Effects of Rhythmic Context on Perceived Duration: A Comparison of Interval and Entrainment Approaches to Short-Interval Timing. *Journal of Experimental Psychology. Human Perception and Performance, 29*(6), 1102–1125. https://doi.org/10.1037/0096-1523.29.6.1102

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders, 38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

Michel, M., Kormos, J., Brunfaut, T., & Ratajczak, M. (2019). The role of working memory in young second language learners' written performances. *Journal of Second Language Writing, 45*(31–45). https://doi.org/10.1016/j.jslw.2019.03.002

Mohajer, L., & Jan., J. B. M. (2019). Gender Differences: Factors Influencing Men And Women's Participation In Gender Research. *The European Proceedings of*

Multidisciplinary Sciences, 786-796. https://doi.org/10.15405/epms.2019.12.80

Morgan, J. L., Edwards, S., & Wheeldon, L. R. (2014). The relationship between language production and verbal short-term memory: The role of stress grouping. *Quarterly Journal of Experimental Psychology, 67*(2), 220–246. https://doi.org/10.1080/17470218.2013.799216

Neath, I., & Quinlan, P. T. (2021). The item/order account of word frequency effects: Evidence from serial order tests. *Memory & Cognition, 49*(6), 1188–1203. https://doi.org/10.3758/s13421-021-01144-7

Neuert, C. E. (2021). The Effect of Question Positioning on Data Quality in Web Surveys. *Sociological Methods & Research, 0*(0). https://doi-org.libaccess.lib.mcmaster.ca/10.1177/0049124120986207

Oberauer, K. (2019). Is Rehearsal an Effective Maintenance Strategy for Working Memory? *Trends in Cognitive Sciences, 23*(9), 798–809. https://doi.org/10.1016/j.tics.2019.06.002

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews, 36*(2), 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008

Ordin, M., & Nespor, M. (2016). Native Language Influence in the Segmentation of a Novel Language. *Language Learning and Development, 12*(4), 461–481. https://doi.org/10.1080/15475441.2016.1154858

Parker, A., Wilding, E., & Akerman, C. (1998). The von Restorff Effect in Visual Object Recognition Memory in Humans and Monkeys: The Role of Frontal/Perirhinal Interaction. *Journal of Cognitive Neuroscience, 10*(6), 691–703. https://doi.org/10.1162/089892998563103

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Papagno, C., Comi, A., Riva, M., Bizzi, A., Vernice, M., Casarotti, A., Fava, E., & Bello, L. (2017). Mapping the brain network of the phonological loop. *Human Brain Mapping, 38*(6), 3011–3024. https://doi.org/10.1002/hbm.23569

Peeters, R., Premchand, A., & Tops, W. (2023). Neuropsychological profile of children with Autism Spectrum Disorder and children with Developmental Language Disorder and its relationship with social communication. *Applied Neuropsychology: Child*, 1–11. https://doi.org/10.1080/21622965.2023.2211703

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and

Expressive Prosodic Ability in Children With High-Functioning Autism. *Journal of Speech, Language, and Hearing Research, 50*(4), 1015–1028. https://doi.org/10.1044/1092-4388(2007/071)

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 564-573. https://doi.org/10.1037/0096-1523.16.3.564

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43*(3), 489–499. https://doi.org/10.3758/s13421-014-0470-6

Preis, S., Schittler, P., Richter-Werkle, R., Sterzel, U., & Lenard, H.-C. (1997). Typical Pattern of the Kaufman-Assessment Battery in Children with Developmental Language Disorder. *Neuropediatrics, 28*(6), 328–332. https://doi.org/10.1055/s-2007-973725

Quené, H., & Port, R. (2005). Effects of Timing Regularity and Metrical Expectancy on Spoken-Word Perception. *Phonetica, 62*, 1–13. https://doi.org/10.1159/000087222

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. Nature Reviews. *Neuroscience, 4*(3), 193–202. https://doi.org/10.1038/nrn1052

Rassili, O., & Ordin, M. (2022). The effect of regular rhythm on the perception of linguistic and non-linguistic auditory input. *The European Journal of Neuroscience, 55*(11-12), 3365–3372. https://doi.org/10.1111/ejn.15029

Redford, M. A., & Oh, G. E. (2016). Children's abstraction and generalization of English lexical stress patterns. *Journal of Child Language, 43*(2), 338–365. https://doi.org/10.1017/S0305000915000215

Reichardt, R., Polner, B., & Simor, P. (2023). The graded novelty encoding task: Novelty gradually improves recognition of visual stimuli under incidental learning conditions. *Behavior Research Methods, 55*(4), 1587–1600. https://doi.org/10.3758/s13428-022-01891-8

Révész, A., Michel, M., & Lee, M. (2017). Investigating IELTS Academic Writing Task 2: Relationships between cognitive writing processes, text quality, and working memory. In *IELTS Research Reports Online Series* (pp. 1-44). IELTS Partners: British Council, Cambridge English Language Assessment and IDP: IELTS Australia. https://search.informit.org/doi/10.3316/informit.840774005843694

Revilla, M., & Ochoa, C. (2017). Ideal and Maximum Length for a Web Survey. *International Journal of Market Research, 59*(5), 557–565. https://doi-org.libaccess.lib.mcmaster.ca/10.2501/IJMR-2017-039

Richardson, J. T. E., & Baddeley, A. D. (1975). The effect of articulatory suppression in free recall. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 623–629. https://doi.org/10.1016/S0022-5371(75)80049-1

Roesch, A. D., & Chondrogianni, V. (2021). Comprehension of Welcher-Questions in German-Speaking Children With and Without Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research, 64*(5), 1683–1695. https://doi.org/10.1044/2021_JSLHR-20-00121

Romani, C., McAlpine, S., Olson, A., Tsouknida, E., & Martin, R. (2005). Length, lexicality, and articulatory suppression in immediate recall: Evidence against the articulatory loop. *Journal of Memory and Language, 52*(3), 398–415. https://doi.org/10.1016/j.jml.2005.01.005

Ryan, J., Lopian, L., Le, B., Edney, S., Van Kessel, G., Plotnikoff, R., Vandelanotte, C., Olds, T., & Maher, C. (2019). It's not raining men: a mixed-methods study investigating methods of improving male recruitment to health behaviour research. *BMC Public Health, 19*(1), 814–814. https://doi.org/10.1186/s12889-019-7087-4

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical Learning by 8-Month-Old Infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Saint-Aubin, J., & LeBlanc, J. (2005). Word Frequency Effects in Immediate Serial Recall of Pure and Mixed Lists: Tests of the Associative Link Hypothesis. *Canadian Journal of Experimental Psychology, 59*(4), 219–227. https://doi.org/10.1037/h0087477

Saksida, A., Flo, A., Guedes, B., Nespor, M., & Peña Garay, M. (2021). Prosody facilitates learning the word order in a new language. *Cognition, 213*, 104686–104686. https://doi.org/10.1016/j.cognition.2021.104686

Schomaker, J., & Meeter, M. (2012). Novelty Enhances Visual Perception. *PloS One, 7*(12), e50599–e50599. https://doi.org/10.1371/journal.pone.0050599

Service, E. (1998). The Effect of Word Length on Immediate Serial Recall Depends on Phonological Complexity, Not Articulatory Duration. *The Quarterly Journal of Experimental Psychology, 51A*(2), 283–304. https://doi.org/10.1080/713755759

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology, 102*(2), 250–255. https://doi.org/10.1037/h0035855

Smith, W. G. (2008). Does gender influence online survey participation?: A record-linkage analysis of university faculty online survey response behavior. *ERIC Document Reproduction Service, 501717*. https://scholarworks.sjsu.edu/elementary_ed_pub/4/

Suppanen, E., Huotilainen, M., & Ylinen, S. (2019). Rhythmic structure facilitates learning from auditory input in newborn infants. *Infant Behavior & Development, 57*, 101346–. https://doi.org/10.1016/j.infbeh.2019.101346

Swank, L. K. (1999). Specific Developmental Disorders: The Language-Learning Continuum. *Child and Adolescent Psychiatric Clinics of North America, 8*(1), 89–112. https://doi.org/10.1016/S1056-4993(18)30198-6

Tanida, Y., Ueno, T., Lambon Ralph, M. A., & Saito, S. (2015). The roles of long-term phonotactic and lexical prosodic knowledge in phonological short-term memory. *Memory & Cognition, 43*(3), 500–519. https://doi.org/10.3758/s13421-014-0482-2

Tyler, M., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America, 126*(1), 367–376. https://doi.org/10.1121/1.3129127

Vallar, G., & Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. *Handbook of memory disorders*, 249-270.

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences, 35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology, 141*, 65–82. https://doi.org/10.1016/j.jecp.2015.06.015

Watanabe, T., Nanez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*(6858), 844-848. https://doi.org/10.1038/35101601

Appendix A: Sample of Demographic Questionnaire

Demographic Questionnaire

Please answer the following questions to the best of your knowledge. You may skip any questions that you do not wish to answer.

A. Demographic Information

1. Age: ____ years old
2. Gender: □ Female □ Male □ Prefer to self-identify as _________________
3. Highest **completed** education: □ Highschool □ Undergraduate □ Graduate
4. Length of residence in an English-speaking country: ____ years

B. Language Acquisition & Proficiency

1. Native (first) language: ______________
2. Second language: ______________ 3. Additional language: ____________________

Please describe:

Language(s)	Age you first learned to…			Ways you learned this language (e.g. family members, grade school, university courses, TV shows, etc.)	Total # of years spent knowing or learning this language
	Speak	Read	Write		
English					

Rate your **current** language abilities using the number scale below:

1	2	3	4	5	6
very low	*low*	*intermediate*	*advanced*	*near-native*	*native*

Language(s)	Reading	Writing	Speaking	Listening	Overall
English					

Page 1 of 3

C. Language Use

1. What language(s) do you usually speak to your family/roommates at home?
____

2. Estimate, in terms of percentage, how often you use English and other language(s) **in a typical day**. Total percentage should sum to 100%.

English: ____%
(Specify which other language) _______________: ____%
(Specify which other language) _______________: ____%

Please rate your language use for the following activities, using the number scale below:

1	2	3	4	5	6
never	*rarely*	*sometimes*	*often*	*very often*	*always*

Language(s)	Arithmetic (add, multiply, counting)	Remembering or telling numbers (phone #, student #)	Dream	Think	Talk to yourself	Express anger or affection
English						

Among the language(s) you know, which one do you prefer to use in the following scenarios?
At home _______________ At school _______________ At work _______________
At a party _______________ In general _______________

3. If you have any other information to share about your language history, add them here:

Page 2 of 3

D. Music Experience

1. Do you have any experience with any musical instruments (including voice and percussion)? If yes, please continue. If not, disregard the remainder of questions.

Please list the instrument(s) you have experience with, and describe:

Instrument(s)	Age you first learned this instrument	Ways you learned this instrument (e.g. professional lessons, band class, self-taught, etc.)	Total # of years learning this instrument

Please describe your **highest ability level achieved to date** in each instrument. Ability levels can be described with tests that you have passed, competitions/awards that you have won, etc.; or you can use the number scale below:

1	*2*	*3*	*4*	*5*
complete beginner	*beginner*	*intermediate*	*advanced*	*professional*

Instrument(s)	Highest ability level ever achieved

2. Do you **currently** practice any of the instrument(s) once a month or more often? If yes, please continue. If not, disregard the remainder of questions.

Please describe how often you are currently practicing the instrument(s):

Instrument(s)	Frequency of practice (e.g. twice a week, once a day, etc.)

3. If you have any other information to share about your music experience, add them here:

Thank you for completing the Demographic Questionnaire.
Please notify the researcher that you are done with this part of the experiment.

Page 3 of 3

Appendix B: Copy of Stimuli in Experiment II

list_ID	list_stressPattern	list_subStressPattern	wordListVA	wordListVB
WLR_1	Mixed	112121	poison, organ, combine, steamer, parade, athlete	steamer, athlete, parade, poison, combine, organ
WLR_2	Mixed	112121	mantle, vapour, debate, capsule, defeat, tractor	capsule, tractor, defeat, mantle, debate, vapour
WLR_3	Mixed	112121	segment, weapon, produce[V], woodland, suspect[V], dairy	woodland, dairy, suspect[V], segment, produce[V], weapon
WLR_4	Mixed	112121	harbour, office, retreat, trumpet, balloon, balance	trumpet, balance, balloon, harbour, retreat, office
WLR_5	Mixed	112121	selling, student, revolt, hardware, fatigue, reading	hardware, reading, fatigue, selling, revolt, student
WLR_6	Mixed	112121	spider, statement, disease, temper, manure, acre	temper, acre, manure, spider, disease, statement
WLR_7	Mixed	112121	bargain, message, resort, gospel, command, blanket	gospel, blanket, command, bargain, resort, message
WLR_8	Mixed	112121	tempest, prairie, cigar, servant, attack, otter	servant, otter, attack, tempest, cigar, prairie
WLR_9	Mixed	121211	sunshine, beloved, cradle, receipt, hostage, lantern	hostage, receipt, lantern, beloved, sunshine, cradle
WLR_10	Mixed	121211	candy, exchange, ceiling, canoe, honey, hour	honey, canoe, hour, exchange, candy, ceiling
WLR_11	Mixed	121211	wire, refuse, boulder, destroy, velvet, shiny	velvet, destroy, shiny refuse, wire, boulder
WLR_12	Mixed	121211	liquor, salute, circus, exhaust, barrel, swallow	barrel, exhaust, swallow, salute, liquor, circus
WLR_13	Mixed	121211	eating, obese, cocktail, delay, needle, quiet	needle, delay, quiet, obese, eating, cocktail
WLR_14	Mixed	121211	flavour, alone, ocean, discharge[V], harness, subject	harness, discharge[V], subject, alone, flavour, ocean
WLR_15	Mixed	121211	prosper, embrace, rigid, design, moisture, silence	moisture, design, silence, embrace, prosper, rigid
WLR_16	Mixed	121211	iris, device, nickel, assault, hockey, twilight	hockey, assault, twilight, device, iris, nickel
WLR_17	Mixed	121121	swimming, cement, onion, congress, recruit, banker	onion, recruit, swimming, banker, cement, congress
WLR_18	Mixed	121121	summer, report, platform, maiden, adult, tumble	platform, adult, summer, tumble, report, maiden
WLR_19	Mixed	121121	daisy, canal, butcher, vulgar, reward, parish	butcher, reward, daisy, parish, canal, vulgar

WLR_20	Mixed	121121	pollen, present[V], colour, outfit, antique, distance	colour, antique, pollen, distance, present[V], outfit
WLR_21	Mixed	121121	mortal, express, union, pulpit, concern, maple	union, concern, mortal, maple, express, pulpit
WLR_22	Mixed	121121	fire, display, story, candle, degree, garden	story, degree, fire, garden, display, candle
WLR_23	Mixed	121121	ally, dispute, crooked, riot, machine, angle	crooked, machine, ally, angle, dispute, riot
WLR_24	Mixed	121121	essay, account, cherry, supper, alert, journal	journal, alert, supper, cherry, account, essay
WLR_25	Uniform	111111	clothing, orange, annex, salad, thunder, mirror	thunder, annex, mirror, clothing, salad, orange
WLR_26	Uniform	111111	planet, labour, lighter, carbon, mattress, temple	mattress, lighter, temple, planet, carbon, labour
WLR_27	Uniform	111111	meeting, victim, duchess, ration, newborn, basement	newborn, duchess, basement, meeting, ration, victim
WLR_28	Uniform	111111	carpet, lesson, outbreak, column, hammer, liar	hammer, outbreak, liar, carpet, column, lesson
WLR_29	Uniform	111111	measles, table, fencing, garbage, patient, window	patient, fencing, window, measles, garbage, table
WLR_30	Uniform	111111	diet, buckle, comfort, pattern, ankle, monarch	ankle, comfort, monarch, diet, pattern, buckle
WLR_31	Uniform	111111	landscape, stranger, polish, linen, impact, ferry	impact, polish, ferry, landscape, linen, stranger
WLR_32	Uniform	111111	lettuce, polo, grammar, yellow, stocking, menu	stocking, grammar, menu, lettuce, yellow, polo
WLR_33	Uniform	111111	pastor, lotion, dresser, brutal, sugar, fashion	sugar, dresser, fashion, pastor, brutal, lotion
WLR_34	Uniform	111111	baby, witness, jewel, donor, paper, body	paper, jewel, body, baby, donor, witness
WLR_35	Uniform	111111	echo, woman, battle, leader, manor, rabbit	manor, battle, rabbit, echo, leader, woman
WLR_36	Uniform	111111	soda, derby, magnet, reflex, mammal, stable	mammal, magnet, stable, soda, reflex, derby
WLR_37	Uniform	111111	human, olive, concrete[N], sunset, leather, damage	leather, concrete[N], damage, human, sunset, olive
WLR_38	Uniform	111111	fabric, vessel, willow, contents, session, captive	session, willow, captive, fabric, contents, vessel
WLR_39	Uniform	111111	builder, hearing, carrot, forearm, drama, vacuum	drama, carrot, vacuum, builder, forearm, hearing
WLR_40	Uniform	111111	profit, whistle, junior, berry,	herring, junior, tailor, profit,

			herring, tailor	berry, whistle
WLR_41	Uniform	111111	cedar, bitter, powder, topic, wallet, hero	wallet, powder, hero, cedar, topic, bitter
WLR_42	Uniform	111111	basket, liver, alley, country, diamond, hurdle	diamond, alley, hurdle, basket, country, liver
WLR_43	Uniform	111111	brandy, kettle, fellow, keeper, diving, ruby	diving, fellow, ruby, brandy, keeper, kettle
WLR_44	Uniform	111111	speaker, wedding, cellar, motion, nephew, pudding	nephew, cellar, pudding, speaker, motion, wedding
WLR_45	Uniform	111111	princess, armour, volume, chicken, kingdom, harvest	kingdom, volume, harvest, princess, chicken, armour
WLR_46	Uniform	111111	driver, utter, marble, chloride, rattle, finance	rattle, marble, finance, driver, chloride, utter
WLR_47	Uniform	111111	pasture, butter, open, weather, poet, motor	poet, open, motor, pasture, weather, butter
WLR_48	Uniform	111111	private, morning, sergeant, margin, breakfast, buffer	breakfast, sergeant, buffer, private, margin, morning