교육학박사학위논문

Working Memory in English Speech Production: Native Speakers vs. Korean EFL Learners

영어 발화에서의 작업기억 연구:
원어민과 한국인 학습자 비교

2020년 8월

서울대학교 대학원
외국어교육과 영어전공
이 옥 영

Working Memory in English Speech Production: Native Speakers vs. Korean EFL Learners

by

Ogyoung Lee

A Dissertation
Submitted to the Department of Foreign Language Education
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in English Language Education

At the Graduate School of
Seoul National University

August 2020

Working Memory in English Speech Production: Native Speakers vs. Korean EFL Learners

영어 발화에서의 작업기억 연구:
원어민과 한국인 학습자 비교

지도교수 안 현 기

이 논문을 교육학박사 학위논문으로 제출함
2020년 5월

서울대학교 대학원
외국어교육과 영어전공
이 옥 영

이옥영의 박사학위논문을 인준함
2020년 7월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Working Memory in English Speech Production: Native Speakers vs. Korean EFL Learners

by

Ogyoung Lee

A Dissertation Presented to the Department of Foreign Language Education
and the Graduate School of Seoul National University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in English Language Education
at the Graduate School of Seoul National University

July 2020

APPROVED BY DISSERTATION COMMITTEE:

_____
KITAEK KIM, Chair

_____
HO-YOUNG LEE, Vice Chair

_____
JIN-HYUNG KIM, Member

_____
SUNHEE KIM, Member

_____
HYUNKEE AHN, Member

ABSTRACT

WORKING MEMORY IN ENGLISH SPEECH PRODUCTION:
NATIVE SPEAKERS VS. KOREAN EFL LEARNERS

Ogyoung Lee
Department of Foreign Language Education (English Major)
The Graduate School
Seoul National University

This dissertation investigates how speech sounds may change as a function of varying working memory load during speech. While cognitive factors have long been suspected to influence speech sounds, being responsible for unexplained variability in speech prosody, the hypothesized connection between cognition and speech production has yet to be systematically theorized with empirical data. Despite the widely accepted assumption in psycholinguistics that working memory involves speech production for temporary maintenance and active manipulation of speech information, the sound planning process remains an open question. Production models not only acknowledge that they only partially account for the phonological-phonetic encoding processes, but production models also adopt direct retrieval of memorized articulatory routines from long-term memory that may allow speakers to bypass the working memory process. Moreover, a comprehensive review from a memory perspective requires direct evidence and reexamination of the psycholinguistic assumption, suggesting that there is no working memory used in speech production. Furthermore, if the key lies between active ongoing planning and automatic routined response, then native speakers and nonnative speakers of a language may reveal different production patterns. These speaker types and their production patterns have not yet been directly compared in the literature.

To address this gap, this study asks whether working memory is involved in English speech production and whether the involvement differs between native (first-language, L1) and nonnative (second-language, L2) speakers of English. To explore how cognitive processing overload derails typical speech sounds, in this study I experimentally manipulated the amount and the type of cognitive load in working memory by having participants engage in multitasking while speaking. Native speakers of American English and Korean EFL learners of English produced a set of syntactically complex sentences under two working memory load conditions and two control conditions. In the load conditions, speakers engaged in an additional task that taxed either verbal or spatial working memory while speaking. In the control no-load conditions, they solved an equation before speaking.

The results show different working memory influences on L1 vs. L2 speech production. Both the L1 and the L2 speech had more errors and were faster during multitasking. However, the effect of load type was different. The L1 speech remained similarly intact, regardless of whether the task was verbal or spatial. By contrast, the L2 speech was impaired only by the verbal task; speech produced during a concurrent verbal task became more erroneous, faster, less variable in word durations, and less distinctive in vowel quality. The spatial task did not impact L2 prosody.

I interpret these results as suggesting that different cognitive processes underly L1 and L2 speech production. This dissertation proposes a tentative model for L1 and L2 speech production with direct reference to working memory versus long-term memory. L1 speech production is usually automatic and spontaneous via retrieval from long-term memory and not impacted by (verbal) working memory. Thanks to sufficient exposure to and practice of form-sound units, speech production directly accesses long-term memory and instantly executes the articulatory gestural routines that are already stored along with the lexical forms. As L1 speech production draws from long-term memory, working memory is not overloaded by a (verbal or spatial) working memory task, and speech becomes only more erroneous and faster due to divided attention within a set time frame. By contrast, L2 speech sounds mainly emerge on the go via phonological-phonetic encoding in (verbal) working memory. Due to limited language experience in L2, articulatory templates are often not fully developed or readily retrievable for automatic execution. As L2 speech production draws from verbal working memory, a concurrent verbal task, but not a spatial task, overloads the capacity-limited verbal working memory component. While a general effect of divided attention can be observed just as in most multitasking, a verbal task disrupts speech more significantly than a non-verbal task.

The study contributes direct evidence regarding the relationship between working memory and speech production. I examine some speculated effects of cognitive load on prosodic variations and move toward resolving the controversy over the encoding and the retrieval processes by referring to language experience. Beyond the scope of this dissertation and as a guidepost for research, we need to investigate how speech production processes may develop as a function of L1 language experience and L2 proficiency.

Keywords: working memory, working memory capacity, speech production, prosody, phonological-phonetic encoding, retrieval, L1 versus L2
Student Number: 2006-30916

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER I.

INTRODUCTION

1.1 Context and Purpose of the Study

Speech-language production is a highly complex, communicative-intent-driven, linguistic activity that involves continuous concerted interaction and manipulation of cognitive mechanisms. It has been modelled as a multi-staged process in psycholinguistics (e.g., Dell & Reich, 1981; Fromkin, 1971; Garrett, 1975; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). In order for speech to be ultimately produced, linguistic information is actively retrieved and manipulated in multiple cognitive stages. A speaker's message has to be prepared, corresponding words are selected, the morphosyntactic forms are activated, the prosodic and phonemic properties are encoded, the articulatory routines are specified, and finally articulatory gestures produce sound waves (Levelt, Roelofs, & Meyer, 1999). The output of a preceding stage becomes the input of the following stage, where the output from each stage is maintained in a buffer storage before being transferred to the following stage. The process from selecting a metrical shape and sequencing phonological forms through to syllabification is referred to as phonological encoding. The process from selecting articulatory routines through to compiling prosodic words is referred to as phonetic encoding. These are the processes that constitute speech production. The interest is in an entailment of the phonological-phonetic encoding hypothesis; namely, that it predicts working memory involvement in the speech production process.

Insofar as phonological-phonetic encoding refers to the selection and manipulation of phonological material (see, e.g., Gathercole & Baddeley, 1993; Swets, Jakovina, & Gerrig, 2014), it requires working memory involvement in speech-language production. Working memory refers to the brain system for active maintenance, manipulation, and retrieval of relevant information to perform ongoing cognitive tasks such as language comprehension and learning (Baddeley, 1992; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). This system is capacity limited, which means we can only effectively process a set amount of information at any given time and overloading working memory results in impaired performance (Engle, 2002). Impaired performance has been observed across cognitive domains where it is required, including decision making (Caplan, Alpert, Waters, & Olivieri, 2000), event reconstruction (Hambrick & Engle, 2002), language comprehension (Daneman & Carpenter, 1980; King & Just, 1991; Leikin & Bouskila, 2004), and language production (Gathercole, Willis, Baddeley, & Emslie, 1994), inter alia.

In spite of this, we are usually able to complete two unrelated cognitive tasks at the same time; for example, listening to the news while solving a jigsaw puzzle. It is much more challenging to complete two related tasks at the same time; for example, telling a story while listening to and retaining the news. According to Baddeley and Hitch's (1974) multi-component model, this is because different working memory components serve different types of information processing. Verbal working memory serves language-related tasks (e.g., listening to the news); spatial working memory serves tasks that involve the retrieval and manipulation of

spatial relations (e.g., solving a jigsaw puzzle). According to another influential theory, Cowan's (1988) embedded-processes model, such interference in task performance is explained to arise from "similar coding of subsequent stimuli" (Cowan, 1999, p. 71). Working memory in this model is a single storage and processing unit for activation of all types of information into 'focus of attention'. Poor performance is predicted when additional information is similar to the concurrently activated memory at the time of processing (Cowan, 1999).

No matter what the theory is, the point is that the activation and manipulation of information in one domain is predicted to impede the activation and manipulation of other information in the same domain. This prediction is supported in studies of language production in both children and adults (e.g., Seigneuric, Ehrlich, Oakhill, & Yuill, 2000; Kellogg, Olive, & Piolat, 2007). For example, Kellogg and colleagues (2007) used a dual task paradigm to examine the extent to which a written language production task—writing definitions for concrete and abstract nouns—disrupted performance in a concurrent verbal, visual, and spatial working memory tasks. The findings were that writing disrupted performance in the verbal working memory task, but not in the spatial working memory task. When the nouns were picturable (i.e., concrete), performance in the visual working memory task was also disrupted. In the present study, the capacity-limited nature of working memory is used to assess its involvement in speech production processes.

Although the hypothesis of phonological-phonetic encoding also predicts verbal working memory involvement in speech planning and production, there is no

direct evidence to support this prediction. Moreover, the indirect evidence suggests the opposite; namely, that working memory is not relevant to speech planning and production (see, e.g., Gathercole & Baddeley, 1993, chapter 4). This evidence is consistent with phonetically-informed theories of production that hypothesize that planning is based on the activation of word-sized chunks stored in long-term memory (Fowler, Rubin, Remez, & Turvey, 1980; Browman & Goldstein, 1992; Nam, Goldstein, Saltzman, & Byrd, 2004; Redford, 2015). These chunks abstractly encode relative timing information that guides articulatory movements over the course of word production, and are stored in association with the lexical concept. This association allows for their direct access, obviating the need for a phonological-phonetic encoding stage in speech production where abstract, segment-sized units are sequenced and then translated into production routines.

The present study tested the hypothesis of phonological-phonetic encoding against an alternative retrieval hypothesis in an elicitation task. We used a working memory load manipulation to directly investigate the effects of verbal working memory on speech planning and production by comparing it to the effects of spatial working memory on the number of disfluencies and segmental errors that speakers produced as they read complex sentences. We also investigated the effects of verbal and spatial working memory load on the prosody and segmental articulation of sentences that speakers had produced correctly and fluently. We addressed the questions for both first and second language speech production. It is plausible that working memory involvement differs for first-language (L1) speech production

from second- or foreign- language (L2) speech production. L1 speakers tend to retrieve memorized articulatory routines that they have stored in long-term memory through over-practice (e.g., Browman & Goldstein, 1992). L2 speakers, who may have a limited number of stored articulatory chunks for L2 in their long-term memory, might or might not display similar patterns as observed in L1 production.

1.2 Research Questions and Predictions

The present dissertation investigated whether working memory is associated with speech production. The research questions are as follows:

Question 1: Is working memory used in speech production?

Question 2: Is the process different between native speakers and nonnative speakers of a language?

To address these questions, it had both native (first language, L1) and nonnative (second language, L2) speakers of English produce English sentences during a verbal and a spatial working memory task and also in associated no-load conditions. L1 speakers were native American English speakers. L2 speakers were Korean EFL (English as a foreign language) learners. Two experiments were completed, one to examine the questions for the native speakers of English and the other for the Korean learners of English. Each experiment is presented in two parts of analyses. The first focuses on disfluencies and segmental errors, the second on

acoustic patterns. The division follows from the phonological-phonetic encoding hypothesis, which assumes a planning stage during which phonemes are sequenced, followed by a stage where a more detailed phonetic plan is generated to guide speech output.

The primary interest is in the working-memory type effect during speech production. During a verbal task, speakers are supposed to engage in two verbal tasks; thus, their verbal working memory is taxed. During a spatial task, speaking uses verbal working memory and the spatial task uses spatial working memory; thus, without any potential working memory overload. During the control, no-load condition, speech should be normal. Given the fact that working memory is capacity limited by information type, the following hypotheses were formulated:

Hypothesis 1: If working memory process is used in speech production, speech during a verbal task should be different from speech during a spatial task.

Hypothesis 2: Working memory effects on L1 speech and L2 speech should be different.

The collected data was analyzed of whether the speech was phonologically and phonetically well encoded and executed. According to theory, phonological encoding specifies the metrical shape, phoneme sequences, and syllabification; phonetic encoding specifies the articulatory gestural scores, which determine multiple different gestures of our biological articulators (e.g., tongue, lips). In speech,

planning properties are manifested in prosody, which is the intonation and rhythm, the suprasegmental features (Shattuck-Hufnagel, 1979). Following these theoretical definitions, the analysis was in two parts: phonological encoding and phonetic encoding. Relevant assumptions are:

Assumption 1: If phonological encoding fails, we should find errors in sequencing phonemes and/or in syllabification. Phonological errors should include omitting, adding, or substituting phonemes or words.

Assumption 2: If phonetic encoding fails, even though the speech shows correct sequences of phonemes and words, we should find disfluencies in the articulatory execution of biological articulators. Phonetic disfluencies include intonational, rhythmic, and articulation clarity.

The encoding hypothesis and the retrieval hypothesis predict different working-memory type effects on speech.

Prediction 1: The encoding hypothesis predicts only verbal working memory effects on speech output. Disrupted speech is predicted under verbal load condition, distracted speech under spatial load due to divided attention, and normal speech under no-load.

Prediction 2: An alternative, retrieval hypothesis predicts no verbal vs. spatial type effect. Because pre-stored phonological-phonetic chunks and the relevant

articulatory movements are directly retrieved from long-term memory, and no active processing is conducted, no active manipulation of information is involved in the working memory system and thus no type effect.

1.3 Organization of the Current Dissertation

This dissertation consists of six chapters. Chapter 1 introduces the context that motivated the study and the purpose of the current study. It also presents relevant research questions and predictions based on the literature. Chapter 2 reviews the theoretical framework of speech planning and production and working memory. Chapter 3 and 4 report the methods and the results of the two experiments. Chapter 3 describes the experiment on speech production in first language, with native American English speakers. Chapter 4 describes the experiment on speech production in a second or foreign language, with Korean learners of English as a foreign language. Chapter 5 discusses the findings suggested in the data and answers the research questions of the current study. It also proposes a tentative model of speech planning and production for first-language production and second- or foreign-language production. Chapter 6 concludes the study with pedagogical implications and suggestions for future research.

CHAPTER II.

THEORETICAL BACKGROUND

This chapter provides relevant literature pertaining to the present dissertation. Section 2.1 presents linguistic and psycholinguistic background on speech production. Section 2.2 reviews psychological findings on working memory, the brain system that handles such cognitive processes as in speech production. Section 2.1.1 defines what speech production is. Section 2.1.2 describes how speech is produced. It details the staged cognitive processes leading to ultimate speech sounds in section 2.1.2.1. An alternative view to the staged models of speech production is given of the retrieval model in section 2.1.2.2. Section 2.2.1 defines working memory. Section 2.2.2 outlines models of working memory. Section 2.2.2.1 summarizes a brief historical development to the current working memory models. It describes the two most widely accepted working memory models: the multi-component model of working memory in section 2.2.2.2; an alternative and the second-most accepted model of working memory, the embedded-processes model in section 2.2.2.3. Section 2.2.3 discusses the processing limitation of the working memory system. It describes capacity and time limit in section 2.2.3.1, controversial type effect in section 2.2.3.2, and measurement method in section 2.2.3.3.

2.1 Speech Production

2.1.1 Definition of speech production

Speech production, as a subset of the entire speaking process, refers to the phonological and phonetic transformation of a pre-planned sequence of morphosyntactic lexical forms that represent a speaker's intended concepts. Speaking is a communicative intentional activity that delivers a speaker's intention, thought, and feeling via overtly articulated speech (Levelt, 1989). In order to produce speech, a speaker goes through multiple complex cognitive processes from preparing a conceptual message through selecting morphosyntactic lexical forms to articulating speech sounds (Dell & Reich, 1981; Fromkin, 1971; Garrett, 1975; Levelt, 1989). While the entire production processes that embrace all the planning and implementation from forming meaning through to producing speech output is called language production, the subset of language production that covers only the phonological-phonetic planning and implementation translating a prepared language form to speech output is called speech production (Fowler, 2007, 2010). Speech production implements a planned sequence of language forms as concurrent vocal tract activity of multiple articulators that generates an acoustic speech output (Fowler, 2010). This is done by specifying how the forms are pronounced via multiple controlled processes of an articulatory motor program, where the program instructs our biological articulators to move and produce speech output (Crompton, 1982).

While speech production may seem to be used interchangeably with

language production and/or sentence production, a clear distinction is made with respect to scope and modality. On the one hand, the three terms often refer to the same concept of spoken language production. Spoken language production has been expressed as speech production (e.g., Caramazza, Costa, Miozzo, & Bi, 2001; Costa & Caramazza, 1999; Fowler & Saltzman, 1993; Jescheniak & Levelt, 1994; Kormos, 2006; Levelt, 1995; Levelt et al., 1999; Roelofs, 1997; Shattuck-Hufnagel, 1983; Wheeldon & Lahiri, 1997), language production (e.g., Bock, 1986; Dell & O'Seaghdha, 1992; Ferreira & Swets, 2002; Meyer, 1990, 1991; Poulisse, 1999; Stemberger, 1989; Watson & Gibson, 2004), or sentence production (e.g., Bock, 1987; Dell, 1986; Dell & Reich, 1981; Dell, Oppenheim, & Kittredge, 2008; Ferreira, 1993; Garrett, 1975; Kempen & Hoenkamp, 1982, 1987; Shattuck-Hufnagel, 1979). Two or more of these terms were used interchangeably even within a single paper: speech production and language production (e.g., Dell, Reed, Adams, & Meyer, 2000; Poulisse, 1999; Stemberger, 1989), speech production and sentence production (e.g., Shattuck-Hufnagel, 1983), language production and sentence production (e.g., Ferreira, 1993; Garrett, 1975), or all three terms (e.g., Martin, Crowther, Knight, Tamborello II, & Yang, 2010).

On the other hand, however, speech production is distinguished from other terms by its scope and modality. As for the scope, speech production covers a subset of language production. Speech production accounts only for "planning for and implementation of language forms," whereas language production involves the overall "planning for and implementation of meaningful utterances" (Fowler, 2007,

p. 489). Speech production starts, in the middle of language production process, from a ready-planned language form and ends with an acoustic speech signal, whereas language production covers the entire set of processes from a concept through to the acoustic signal (Fowler, 2010). Morphosyntactically, it can be of any morphosyntactic scopes. If the scope of the description is a word, it can be a word production; if it is a phrase, a phrase production; if it is a sentence, a sentence production. From the perspective of communicative modality, speech production is the spoken part of language production, i.e., language produced in a spoken modality. Language can be produced in a spoken, written, or signed (gestured) modality. If the communicative message is produced as a spoken form, it is called spoken language production, thus speech production or speaking. If the message is produced as a written form, it is a written language production or writing. If the message is produced as a gestured form, it is a sign(ed) language production or signing. It is simply that researchers, knowing such clear distinction, distinguish the terms only to the extent that the research purpose requires or they select terms to highlight the scope of the argument. For instance, Caramazza (1997) contrasted speech production with written language production in order to describe phonological vs. orthographic, modality-specific lexical access models of language production, i.e., phonological mediation hypothesis vs. orthographic autonomy hypothesis.

2.1.2 Models of speech production

This section reviews how speech is planned from concepts to ultimately articulated

speech output. Most research traditions agree that speech is produced through multiple stages of incremental encoding. The staged models of speech production are reviewed in section 2.1.2.1. An alternative approach, retrieval model is described in section 2.1.2.2. In Levelt, Roelofs, and Meyer's (1999) influential model, the production process begins with conceptual preparation and the activation of a lexical concept to be expressed; for example, the concept PRODUCE (X, Y). The concept then triggers lexical selection, which provides syntactic information in the form of a lemma; for example, *produce* is retrieved from the mental lexicon, encoded with the transitive property having two argument positions and with a third-person singular present-tense specification of the verb. Next, the relevant morphemes are retrieved; for example, <produce> and <s>. At the same time, the metrical and segmental properties of the morphemes are retrieved; for example, <produce> is retrieved as an iambic foot (i.e., a weak-strong sequence of syllables = σσˈ) and <s> as an extrametrical suffix; segments are laid out in sequence (i.e., /p/, /r/, /ə/, /d/, /j/, /u/, /s/, /ə/, /z/) and the spelled-out word form is syllabified (i.e., /prə.dju.səz/) according to phonological rules, which also transform the underlying phonemes into context-dependent allophones. The syllabification procedure allows for the selection of appropriate articulatory routines from a mental syllabary. Once selected, the routines are passed to the articulatory buffer where they are held until a prosodic word is compiled for execution. The process from the retrieved morphosyntactic word form through to syllabification is referred to as phonological encoding. The selection of articulatory routines through to the compilation of prosodic words is referred to as

phonetic encoding. These are the processes that interest us in the present dissertation. In particular, we are interested in an entailment of the phonological-phonetic encoding hypothesis; namely, that it predicts working memory involvement in the speech production process.

Semantically accessed and selected lexical items (excluding semantically activated but unselected items) are phonologically activated and encoded (Levelt et al., 1991) and also phonetically encoded as articulatory gestures in the motor program. Given a selected word or lemma, its abstract phonological form, i.e., the metrical shape and segmental makeup, is accessed from the mental lexicon and the properties are activated and encoded as a phonological word; the phonological word is then specified of appropriate articulatory gestural scores in the prosodic context, i.e., of how it is phonetically uttered using the biological articulatory apparatus; the articulators finally execute the gestural scores so as to result in acoustic speech sounds (Levelt et al., 1999). The set of phonological and phonetic routine procedures are automatic and require little attentional effort, where a speaker's attention is mostly on planning and elaborating the communicative illocutionary concepts and expressions (Levelt, 1989, chapter 4).

2.1.2.1 Encoding in the staged models of speech production

Speech production has been modelled as multi-staged incremental-encoding processes in psycholinguistics (e.g., Bock, 1987; Boomer & Laver, 1968; Caramazza, 1997; Dell, 1986, 1994; Dell & Reich, 1981; Fromkin, 1971; Fry, 1969; Garrett,

1975; Kempen & Hoenkamp, 1987; Levelt, 1989, 1995, 1999; Levelt, Roelofs, & Meyer, 1999; Shattuck-Hufnagel, 1979; Vigliocco, Vinson, Martin, & Garrett, 1999). While some details differ across various models[1], the consensus in the literature is that speech production is decomposed into multiple distinct stages that proceed from conceptualization to articulation, where each stage produces its own level of representation and the output of a preceding stage becomes the input of the immediately following stage.

In an influential and recently updated model of speech production, Levelt, Roelofs, and Meyer (1999) describe speech production as involving five processing stages, after which it finally initiates articulation: conceptual preparation, lexical selection, morphological encoding, phonological encoding, and phonetic encoding. Each stage outputs a corresponding representation of lexical concepts, lemmas, morphemes, phonological words, and phonetic gestural scores, respectively. The final output, phonetic gestural scores are then executed in the articulation stage, and sound waves are generated. Throughout all these stages, the component processes continuously interact with and are dependent on each other, while self-monitoring in parallel may repair the internal and external speech output. The speech production process, encompassing both language planning and speech planning stages, is represented in Figure 2.1.

[1] Differences include the number of levels or stages (e.g., six in Fromkin, 1971; four in Garrett, 1975; three in Dell, 1986, and in Levelt, 1995) and their labels (e.g., functionally the same stage was called differently 'functional level' in Garrett, 1975, or 'syntactic level' in Dell, 1986).

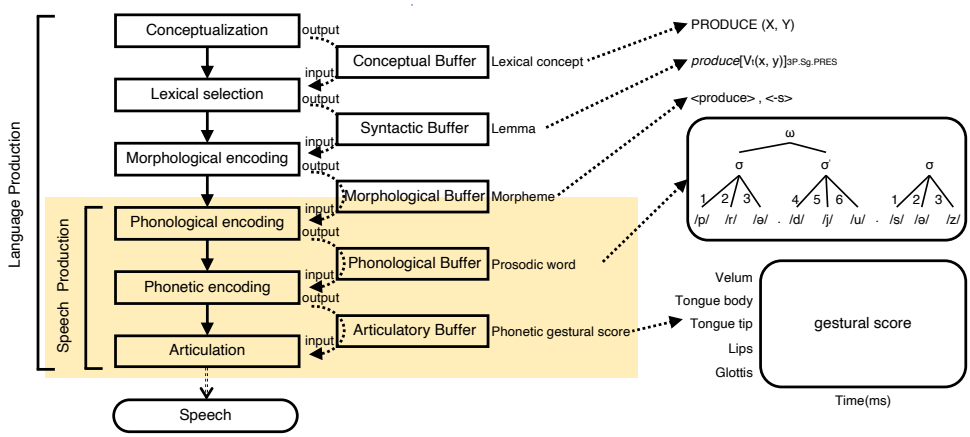


Figure 2.1 Speech production processes in the staged models
Multiple processing stages theorized by Levelt, Roelofs, and Meyer (1999, adapted from Figure 1, p. 3). Speech production is highlighted in yellow background.

Conceptual preparation activates a lexical concept to be expressed. For example, PRODUCE (X, Y), the meaning of the verb *produce*, is activated, where the concept is linked to other concept nodes such as GENERATE (X, Y) (Levelt, Roelofs, & Meyer, 1999, referring to the conceptual network in Roelofs, 1992). Lexical selection stage retrieves a lemma from the mental lexicon, making its syntactic properties available. Lemma *escort* is selected, activating its properties of a transitive verb with two argument positions. Morphological encoding accesses the grammatical/morphological form through morphological code retrieval (Levelt, 1999). For *produce*, morphological codes for the two relevant morphemes, <produce> and <s>, are accessed.

Phonological encoding specifies the metrical structure (i.e., stress pattern) and phonemes for each morpheme, labeling the phonemes for their correct ordering. The morpheme <produce> is linked to the metrical information iambic $\sigma\sigma'$ (i.e.,

disyllabic and stress-final, being a phonological word ω), and <-s> to *σ* (i.e., monosyllabic and unstressed, not being an independent phonological word). The phonemic segments are laid out as /p/, /r/, /ə/, /d/, /j/, /u/, /s/ and as /ə/, /z/, with strict order of 1, 2, 3, 4, 5, 6, 7 and 1, 2, respectively for each morpheme.

Whether phonemes are syllabified at this stage or not is controversial. Fromkin (1971) and Levelt (1995) describe syllabification is done along with phonological encoding. On the other hand, Levelt, Roelofs, and Meyer (1999) view syllables are not yet formed due to possible changes in context (e.g., *e-scort* or *e-scor-ting*). In their model, syllabification is done when generating the words, projecting its domain over prosodic words (e.g., *escort*, *escort us*, *escorting*). The spelled-out metrical structure for each morpheme may stay as it is or be modified according to the phonological context. For the phonological word *escorting*, the metrical structures for <escort> and <ing> will merge to a trisyllabic template *σσˈσ*. The spelled-out segments are successively inserted into the current metrical template, forming phonological syllables on the fly, i.e., *e-scor-ting*. The phonological syllables, *e*, *scor*, and *ting*, with a lexical accent on *scor*, activate the phonetic syllable scores [ə], [skɔr], and [tɪŋ].

Phonetic encoding, according to Levelt, Roelofs, and Meyer (1999), computes and accesses the articulatory gestural scores for phonological syllables or words in a mental syllabary. The gestural score is a specification of articulatory gestures to be executed during articulation to produce the word in the correct order of the articulatory movements of the vocal apparatus (as described in Articulatory

Phonology in Browman & Goldstein, 1992). The articulatory gestures in the gestural score are executed at various articulatory tiers of velic tier, glottal tier, and oral tiers (i.e., lips, tongue body, tongue tip tier, Browman & Goldstein, 1992).

This level of phonetic encoding seems to require further research especially, as also noted in Levelt, Roelofs, and Meyer (1999). They said their account for phonetic encoding was only partially done (p. 5). Moreover, the encoding process was described somewhat differently among different models.

Finally arriving at articulation, the phonological words' gestural scores are executed by the articulatory apparatuses. The articulatory motor control involves complicated concerted interaction and coordination between a neural control system and the relevant physical muscle controls (see Levelt, 1989).

Self-monitoring accompanies the whole production processes, based on our normal perceptual system. It checks whether the produced speech is what was intended, and repairs the speech output if inappropriate.

Such modelling of discrete stages for speech production was initiated to account for speech error data (e.g., Crompton, 1982; Dell & Reich, 1981; del Viso, Igoa, & García-Albea, 1991; Fromkin, 1971; Fry, 1969; Garrett, 1975; Meyer, 1992; Peterson & Savoy, 1998; Shattuck-Hufnagel, 1979; Stemberger, 1989). As a pioneer, Fromkin (1971) argued speech errors are not random but instead highly systematic and informative of speech performance. In order to account for at least eight error patterns evidenced in her data, she suggested a model of actual generation of an utterance (i.e., the utterance generator model) by means of five stages. She explained

transposition errors of switching noun for noun and verb for verb (but not a noun for a verb) occur when syntactic properties are already specified for each slot within the syntactic structure (e.g., concept *ball* tagged for [-animate, +noun, +count] etc. at Stage 2 in her model). Lexical selection errors occur (at Stage 4) when specifying an incorrect address in the lexicon due to mismatching incorrect semantic features (e.g., hate for like), or when substituting/selecting an incorrect address for the correctly obtained address (e.g., present for pressure, due to phonological similarity of their first three segments /p/, /r/, and /ɛ/). Some types of phonological errors may also occur at the same stage (at Stage 4) when the selected phonemes for the morphemes/words are assigned to ordered syllables. Such errors include segmental substitution (of phoneme 1 of syllable 1 with phoneme 1 of syllable 3, but without changing the syllabic ordering) and transposition or misplacement of parts of or whole syllables. Phonological errors such as *a* for *an* or *s* for *z* (plural) are after the phonological forms of the morphemes are specified following the language's morphophonemic constraints (at Stage 5).

The tip of the tongue phenomenon has also been examined to understand speech production processes (e.g., Burke, MacKay, Worthley, & Wade, 1991; Vigliocco, Antonini, & Garrett, 1997; Vigliocco, Vinson, Martin, & Garrett, 1999). It evidenced phonological encoding follows syntactic encoding, in the sense that English speakers in a tip of the tongue state can still indicate correct syntactic features like count nouns (e.g., *a gondola*) or mass nouns (e.g., *some asparagus*) (Vigliocco, Vinson, Martin, & Garrett, 1999) and that Italian speakers can access the

grammatical gender of the words for which they cannot produce or provide any phonological information (metrical or segmental) about the target (Vigliocco, Antonini, & Garrett, 1997). Burke, MacKay, Worthley, and Wade (1991) argued it occurs when the connections between lexical and phonological nodes become weakened, resulting from infrequent use, nonrecent use, and aging.

Given these distinct levels of representation, an agreement has not yet been reached as to how the information flows across the stages: in a serial manner or in parallel. In the former set of serial processing models (e.g., Fromkin, 1971; Garrett, 1975; Levelt, 1995; Caramazza, 1997), speech proceeds from one level of representation to the next lower level in a hierarchically sequential manner. The output representation feeds forward to the next lower level until the information is ready to instruct articulators. As in Levelt's (1995), at the highest level, a conceptual message is prepared. At the intermediate levels, syntactic structures are framed, then lexical items are selected, and morphological properties are specified. At the lowest level, phonological and phonetic details are tagged. In one extreme, Caramazza (1997) claimed the information flows only one way, from semantic representation to (orthographic or phonological) lexeme stage.

In the latter parallel processing models (e.g., Dell, 1986; Harley, 1984), the levels of representation interact and process information in parallel. In Dell's (1986) spreading activation model of speech production, information of a higher level of representation translates to the next lower level through spreading activation in the lexicon (Dell, 1986). An utterance is planned at each level of representation, which

consists of an ordered set of items in the lexicon. Encoding refers to constructing the representation at each level. Syntactic, morphological, and phonological representations are sequentially ordered sets of words, morphemes, and phonemes, respectively. Representations at all levels are formed in the same manner. The nodes of a higher representation activate nodes of the immediately lower representation through a spreading-activation mechanism. Representation at a level is constructed as an ordered set of selected items, which are done in the following sequential processes: generative rules build a sequence of categorically defined slots (e.g., noun at syntactic level, plural at morphological level, onset at phonological level); decision rules select an item with the highest activation level for each category; insertion rules fill in the slots with the selected items.

Moreover, some models suggested continuous and simultaneous interactions between the stages. Levelt, Roelofs, & Meyer, (1999) posited self-monitoring feeding back throughout the stages. Dennett (1991, cited in Levelt, Roelofs, Meyer, 1999:4) also suggested the products from morphophonological encoding could feed back up to activate the corresponding lexical concepts as internal speech.

Although we are not yet sure whether the vertical information flow is serial or parallel, what is consistent in the literature is that the horizontal encoding is strictly incrementally sequential from left to right, i.e., incremental encoding. For example, syllabification process is argued to be strictly sequential from left to right: within a syllable (from onset to rhyme, Meyer, 1991), within a word (from the first syllable to the next, Meyer, 1990), in a phrase and in a sentence (Griffin, 2001).

According to Roelofs (1997, p. 259), syllabification proceeds from the first segment to the second, and so forth, whereby syllable positions (onset, nucleus, coda) are assigned following the syllabification rules of the language. Essentially, each vowel or diphthong is assigned to a different syllable node and consonants are treated as onsets unless phonotactically illegal onset clusters arise. In the encoding of <hamer>, the /h/ is made syllable onset and the /a/ nucleus of the first syllable, and the /m/ onset, the /ə/ nucleus, and the /r/ coda of the second syllable. The online assignment of syllable positions provides for cross-morpheme and cross-word syllabification. In planning polymorphemic words or connected speech, adjacent morphemes or words may be syllabified together. For example, WEAVER may group the segments of the stem <hamer> and the suffix <en>. Together for the infinitive of the verb *hameren*, or may join the segments of <hamer> and <in> for the cliticization *hamerin*. Then, applying the syllabification rules, /r/ will be made onset of the third syllable instead of coda of the second syllable, yielding $(ha)_{\sigma}(mə)_{\sigma}(rə)_{\sigma}$ and $(ha)_{\sigma}(mə)_{\sigma}(rɪn)_{\sigma}$"

Despite the vast effort to argue for a staged process in speech production, problems and incompleteness have been noted in the literature. The problem of positing strict stages has been evidenced. Dell and Reich (1981) showed misordered-phoneme errors tend to be similarly sounding real words that share both semantic and phonological properties with the target words. They argued that information could leak between stages by way of mental lexicon, and thus that information from other stages (not from the modeled stages) can influence the production process. In addition, beyond morpheme-by-morpheme, the whole word form is prosodified

forming phonological word. Here, if we do not know the lexical stress information (i.e., metrical frame) of the whole word, we cannot prepare the first syllable (Roelofs & Meyer, 1998). More critically, Blanken, Dittmann, and Wallesch (2002) supported the parallel architecture in lexical activation and selection of lexical form in speech production. They counter-evidenced the serial modelling that locates word selection (and mis-selection) and lexical form access (and blockings) at separate serially posited stages, by showing inverse and compensatory relationship between semantic errors (e.g., *lion* for *tiger*) and omission errors (i.e., lexical deletion due to word finding blockings). In another study, some intermediate levels were removed: Caramazza (1997) argued that lexical retrieval is directly from the conceptual level to word form (dual-stage access model). The evidence was from modality-specific speech errors. That is, some speech errors were only in speaking or in writing, i.e., not always in both modalities. The paper models that phonological lexemes and orthographic lexemes are independently represented, and such representations are directly retrieved from lemma to word form.

Arguing against the suggestion that phonetic information is rather directly retrieved from the mental syllabary, Levelt, Roelofs, and Meyer (1999) explain phonetic information is encoded as speakers compose entirely new syllables.

2.1.2.2 Retrieval in the gestural models of speech production

Some phonetically-based theories of speech production suggest that the phonological and phonetic information, instead of being planned online, is directly

retrieved from the mental syllabary. They hypothesize that phonological and phonetic planning is based on the activation of word-sized chunks stored in long-term memory (Browman & Goldstein, 1992; Redford, 2015). These chunks abstractly encode relative timing information that guides articulatory movements over the course of word production and are stored in association with the lexical concept. This association allows for their direct access, obviating the need for a phonological-phonetic encoding stage in production. The phonological-phonetic information is retrieved from articulatory templates in long-term memory and executed automatically as overly practiced articulatory behavior.

A gesture is defined using a set of vocal tract variables, which are lip protrusion, lip aperture, tongue tip constrict location, tongue tip constrict degree, tongue body constrict location, tongue body degree, velic aperture, and glottal aperture (Browman & Goldstein, 1992). As in Figure 2.2, word-sized gestural scores are stored in the mental syllabary; they are retrieved as a whole and are executed as a consorted articulatory action to produce words such as *pan*.

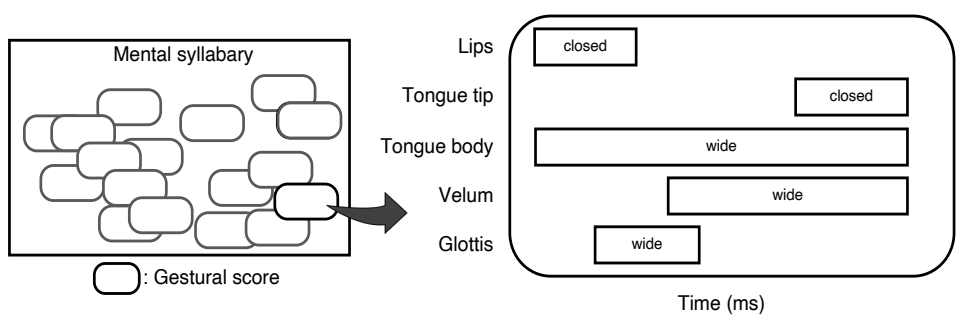


Figure 2.2 Retrieval of a phonological-phonetic information chunk
Gestural action score as a whole is directly retrieved from the mental syllabary (Adapted from Figure 2 in Browman & Goldstein, 1992, p. 157)

2.2 Working Memory

2.2.1 Definition of working memory

Working memory is the brain system that actively maintains a limited set of information (of about 3 to 5 items or chunks) temporarily (for about 10 to 20 seconds) in our controlled, conscious attention and at the same time manipulates the information in order to handle complex ongoing concurrent cognitive tasks such as language comprehension, production, learning, reasoning, and problem-solving. Amongst various definitions in various theories (e.g., see Miyake & Shah, 1999; Anderson, 2000), foundational definitions can be as follows. Working memory refers to the brain system for temporary storage and manipulation of information, necessary for the performance of such complex cognitive activities as language comprehension, learning, and reasoning (Baddeley, 1992). It is a cognitive system responsible for the active maintenance, manipulation, and retrieval of relevant information required for ongoing cognition (Unsworth, Redick, Heitz, Broadway, & Engle, 2009). With an emphasis on the role of controlled and conscious attention in working memory, it can be stated as the collection of mental processes that allow us to hold limited information in a temporarily accessible state in service of cognitive tasks (Cowan, 1999; Cowan et al., 2005).

Working memory system enables us to perform a cognitive task. Information that is not consciously attended in working memory disappears (or decays) within about 2 to 10 seconds from our (sensory) memory. The incoming input needs to be

actively maintained via attention in working memory and coordinated with the already-stored information in long-term memory via working memory. For instance, if we see a sequence of words (i.e., environmental input), the visually presented linguistic input has to first be consciously attended and maintained in our memory (i.e., in the phonological loop in working memory), where it is translated into phonological code (via articulatory control process in working memory); the code is then sent to and coordinated with the stored linguistic knowledge in long-term memory (via episodic buffer in working memory); the coordinated information is then retrieved back to working memory (via episodic buffer) for linguistic output such as comprehension or production; the entire process is attentionally controlled by the commander unit (i.e., the central executive of working memory).

While the concept working memory was initially proposed by Baddeley and Hitch (1974), the construct of working memory was not entirely new but rather a revised version of the previously developed short-term memory store in the modal model theorized by Atkinson and Shiffrin (1968). Providing evidence from ten experiments to point out the simplicity of the short-term store, Baddeley and Hitch (1974) added dynamic feature to it in order to handle complex simultaneous tasks such as learning, reasoning, and language processing. They replaced the single-unit short-term memory store with dual-unit stores that hold different types (spatial or verbal) of information. Accordingly, the terms short-term memory and working memory are often used interchangeably (e.g., Engle & Kane, 2004; Jonides et al., 2008), where additional features (e.g., attentional, Engle & Kane, 2004; brain

processes, Jonides et al., 2008) are described to reflect specific views on working memory architecture and processes.

2.2.2 Models of working memory

A brief summary of the literature that has led to the current working memory models comes in section 2.2.2.1. Two most widely accepted working memory models are described. The multi-component model of working memory (Baddeley & Hitch, 1974; Baddeley, 2000; Baddeley, Allen, & Hitch, 2011) is in section 2.2.2.2. An alternative and the second-most accepted model of working memory, the embedded-processes model (Cowan, 1988, 1999, 2005), follows in section 2.2.2.3.

2.2.2.1 Development to current working memory models

The current models of working memory are the results of accumulated research efforts throughout history. Human memory, stretching its history back to antiquity (e.g., inter alia, Plato's memory as a wax tablet in 'Theaetetus,' circa 369 B.C.; Aristotle's 'On memory and reminiscence,' circa 350 B.C.), has been continuously defined and redefined in multiple concepts. Pioneered by Ebbinghaus (1885)'s probably first scientific attempt to study memory, it has been extensively experimented and explained as representation (e.g., iconic memory, Sperling, 1960; echoic memory, Crowder & Morton, 1969), as storehouse (e.g., a library, Broadbent, 1971; a house, James, 1890), and as process (e.g., encoding-storage-retrieval, Melton, 1963; Baddeley & Hitch, 1974; Cowan, 1988).

Its constructs have been elaborated mainly in the number of sub-components and the specific functions each component assumes. Earlier studies dealt mostly with the sensory nature of memory, focusing on a single aspect or modality. Assuming sensory memory, where transient sensory traces exist after a stimulus is gone, iconic memory described visual sensory memories faded rapidly subject to visual interference (Sperling, 1960); echoic memory explained auditory sensory memories faded rapidly subject to auditory interference (Crowder & Morton, 1969).

The dual-component distinction was initially proposed by James (1890). He distinguished primary memory and secondary memory: the former deals with the set of things we are currently aware of, including the recent past; the latter deals with the set of things we can remember if we want to. The distinction in capacity limitation and consciousness was already explained from this account. Primary memory is capacity limited and stores information through conscious attention for a brief amount of time, similar to modern short-term memory store; secondary memory has unlimited capacity and stores information permanently, absent consciousness, similar to modern long-term memory store.

The most influential dual-component[2] system, and the basis of the current working memory models, was proposed as the modal model by Atkinson and

[2] Atkinson and Shiffrin's (1968) model can be considered a 2-component or 3-component model, depending on how we view its sensory store. The original model had three types of stores: sensory stores, short-term store, and long-term store. However, the first type, sensory stores are in fact sensory registers that take all different environmental inputs including visual, auditory, haptic, etc. Thus, many researchers view it as register than store, different from the short-term or the long-term store (e.g., Baddeley & Hitch, 1974; Endestad, 2005; Haque, Al-Ameen, Wright, & Scielzo, 2017).

Shiffrin (1968). It has drawn extensive scholarly attention and has over fifty years so far been widely cited. It was the most comprehensive and coherent theory of the structuralists' notions and at the same time took the perspective of information processing theory. The modal model of memory explains how memory processes work. As with James (1890), the short-term store is capacity limited and the long-term store is capacity unlimited. However, it introduced information-controlling processes to two storages; it also introduced the idea that information is not lost from the long-term store but fails to be retrieved due to interference or search failures. It accounted for memory as processes of information rather than static and unchangeable unit. Via control processes that operate on short-term memory, we can change memory under our direct control.

The processes work as follows: sensory input that is consciously attended enters the short-term store through sensory registers/buffers/stores (visual, auditory, haptic, etc.); otherwise, unintended information immediately decays; about 4 to 7 items can be stored in the short-term store for about 20 seconds; if the information is rehearsed or encoded with some encoding strategies, it is sent to the long-term memory store; otherwise, the temporarily stored information is forgotten from the short-term store (without being encoded or stored in the long-term store); later, using retrieval strategies, stored information in the long-term store can be retrieved to the short-term store; information is rapidly lost from the limited-capacity short-term store, and thus needs to be rehearsed in order to be remembered (or encoded) in the long-term memory store.

To illustrate this, when we walk on a street, we in fact pass through an uncountable number of information, such as hundreds of stores, thousands of people, millions of details about all the objects around us. However, we do not remember (or encode) all of them, but instead only a few pieces of information come to our memory while all others immediately decay. The pieces that come to our attention is what we remember as what we just saw, heard, etc. Most are temporary and are gone in a few seconds or minutes from our short-term memory store, unless we rehearse or try to remember (or encode) them. If we try to remember some information using whatever methods we choose, we can remember (or store) it longer or forever in our long-term memory store. If we use the right cue, we can recall (or retrieve) the information from the long-term store.

Built upon this multi-store model of memory by Atkinson and Shiffrin (1968), Baddeley and Hitch (1974) proposed the multi-component model of working memory. They, providing a critical review especially of the short-term store, split the single short-term store into dual stores and added dynamic interaction and attention control. In Atkinson and Shiffrin's (1968) modal model, the short-term store is crucial to get the information into and out of the long-term store. Information needs to be rehearsed to get into the long-term store and needs to be retrieved from the long-term store back into the short-term store for report (output). The short-term store performs a storage function, lacking a more interactive processor to handle complex and simultaneous cognitive tasks. Baddeley and Hitch (1974) discussed results from ten experiments including simultaneous tasks such as concurrent

multitasks and complex tasks such as language comprehension. They first emphasized the need for a dynamic and interactive processor for active maintenance and attentional executive control to handle complex and often simultaneous cognitive activities. Some of the problems of the modal model that they pointed out included: short-term memory holds information via repetition without a dynamic buffer system; recency effects are not exclusively tided into short-term memory but witnessed in long-term memory. Second, they separated the single storage unit into two storages, one for visuo-spatial information processing and the other for linguistic information processing. Later, this initial model was revised to add an episodic buffer in Baddeley (2000), and further revised to elaborate on the binding functions of the episodic buffer in Baddeley et al. (2011), which is the current version of the multi-component model of working memory and the most widely accepted model of working memory.

Alternative working memory models have also been proposed, where some specifically focused on working memory system (e.g., Cowan, 1999; Engle, Kane, & Tuholski, 1999) and others extended general cognition models to incorporate working memory processes (e.g., Anderson & Lebiere, 1998; Jonides et al., 2008). They include embedded-processes model (Cowan, 1988, 1995, 1999, 2005), capacity-based executive attentional model (Engle et al., 1999; Engle & Kane, 2004), long-term working memory (Ericsson & Kintsch, 1995), computation-based Adaptive Control of Thought-Rational model (ACT-R, Anderson, 1996; Anderson & Lebiere, 1998; Anderson, Reder, & Lebiere, 1996; Anderson et al., 2004; Borst &

Anderson, 2013), neuroscience-based cognitive-neural process model (Jonides et al., 2008), and more (e.g., see Miyake & Shah, 1999).

2.2.2.2 Multi-component model of working memory

The current version of the multi-component model of working memory (proposed and revised primarily in Baddeley & Hitch, 1974; Baddeley, 1986, 2000; Baddeley et al., 2011) consists of 4 components: one processor, one buffer, and two storages. The model is schematically represented in Figure 2.3. The processor is the central executive. It is the commander of the entire working memory system, responsible for all the attentional control. As attentionally-based central controller (Baddeley & Logie, 1999), it does all the manipulation and reasoning by determining all the processes for the other components. It coordinates the activities of the other components, controlling all the processes that the other working memory components perform; it coordinates retrieval strategies, selective attention, temporary activation of long-term memory, suppression of habitual responses, etc.

The buffer is called the episodic buffer. It is a passive limited-capacity system that temporarily stores and integrates information across all different types and sources (e.g., modality) into unitized episodes or chunks: e.g., it binds information from the two working-memory storages into a single multi-dimensional code; it feeds information into and retrieves information from long-term memory. Conscious access to the two storages is granted via this buffer.

The two storages are the visuo-spatial sketchpad and the phonological loop. Visuo-spatial sketchpad holds visual, spatial, and haptic type of information.

Phonological loop holds language- and sound-related information. Phonological loop (aka articulatory loop) comprises two sub-components: phonological store and articulatory control process. Phonological loop itself is called the storage, phonological store. Articulatory control process does the subvocal rehearsal, inner speech, keeping information active in memory. Auditorily presented linguistic information automatically goes into phonological store, i.e., your ear has privileged access to the phonological store. Visually presented linguistic information is translated into phonological code via articulatory control process, to gain access to the phonological store. Articulatory control process does the re-coding through subvocal rehearsal of articulation. The encoding of smell and taste has been speculated but not yet been evidenced.

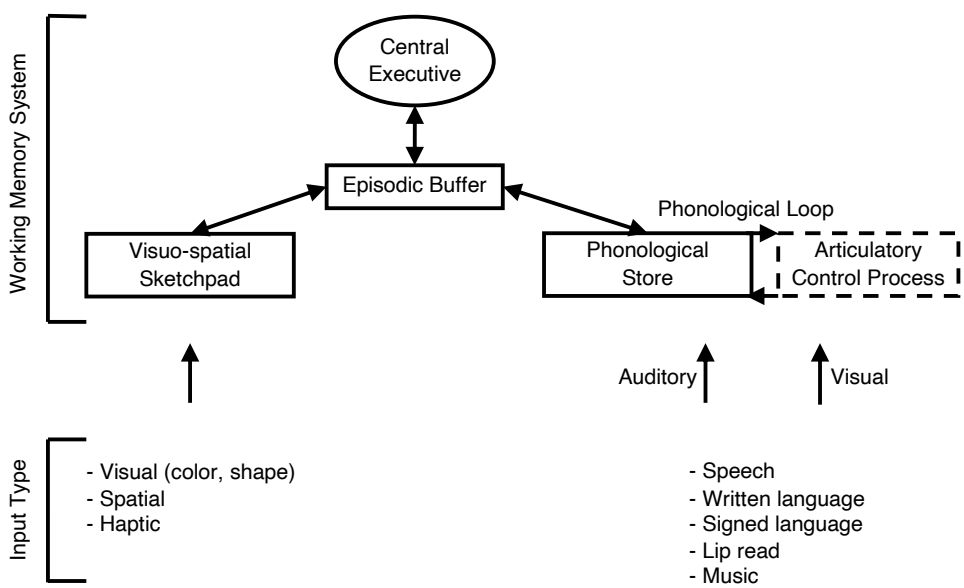


Figure 2.3 Multi-component model of working memory

The current version of multi-component model of working memory. Adapted from Baddeley, Allen, and Hitch (2011, p. 1399, Figure 8). The boxes represent memory stores. The circle represents the processor.

Postulating two separate storage components, this model explains how humans handle dual tasks (Baddeley & Hitch, 1974). Under modal model, which has a single short-term store, when we do one task, that task should take up our short-term memory capacity, making it unable to perform another task. Under multi-component working memory model, when we do one task, that task should take up one store, but we are still left with another store and thus able to use that store to perform another task.

Baddeley and Logie (1999, p. 28) gave five central features of the multi-component model in a book where different working memory models are explained and compared. The features are repeated here:

> (1) According to our view, working memory comprises multiple specialized components of cognition that allow humans to comprehend and mentally represent their immediate environment, to retain information about their immediate past experience, to support the acquisition of new knowledge, to solve problems, and to formulate, relate, and act on current goals.
> (2) These specialized components include both a supervisory system (the central executive) and specialized temporary memory systems, including a phonologically based store (the phonological loop) and a visuospatial store (the visuospatial sketchpad).
> (3) The two specialized, temporary memory systems are used to actively maintain memory traces that overlap with those involved in perception via rehearsal mechanisms involved in speech production for the phonological loop and, possibly, preparations for action or image generation for the visuospatial sketchpad.
> (4) The central executive is involved in the control and regulation of the working memory system. It is considered to play various executive functions, such as coordinating the two slave systems, focusing and switching attention, and activating representations within long-term memory, but it is not involved in temporary storage. The central executive in principle may not be a unitary construct, and this issue is a main focus of current research within this framework.
> (5) This model is derived empirically from studies of healthy adults and children and of brain-damaged individuals, using a range of experimental methodologies. The model offers a useful framework to account for a wide range of empirical findings on working memory. (p. 28)

2.2.2.3 Embedded-processes model of working memory

Cowan's (1988, 1995, 1999, 2005) embedded-processes model of working memory, which comes from a neuroscience approach, consists of 2 components with 3 activation levels: one processor and one storage, with information assigned one of three activation levels. The model is schematically represented in Figure 2.4. The processor is the central executive. It directs our attention and controls voluntary attentional processing. Our controlled, voluntary, and conscious attention, via central executive, assigns a level of activation to incoming information. The storage is in a single-storage system. (i.e., one box). All memory items are embedded within the single store, long-term memory store, where the distinction in memory items is made in the activation levels. Working memory is the activated (both activated and focused) part of memories within the long-term store.

The levels or states of activation are distinguished in three: inactive, activated, or focused. The activation levels are actively, attentionally, and voluntarily controlled and determined by the central executive. Inactive information is the inactive but stored contents or memory items in the long-term memory store. Activated information is maintained active for ready access. Whenever it is called for processing, it can readily be brought to our focus of attention by the central executive; it is otherwise inactivated if unattended. Focused information is items that are in the highest state of activation and subject to conscious processing within our focus of attention.

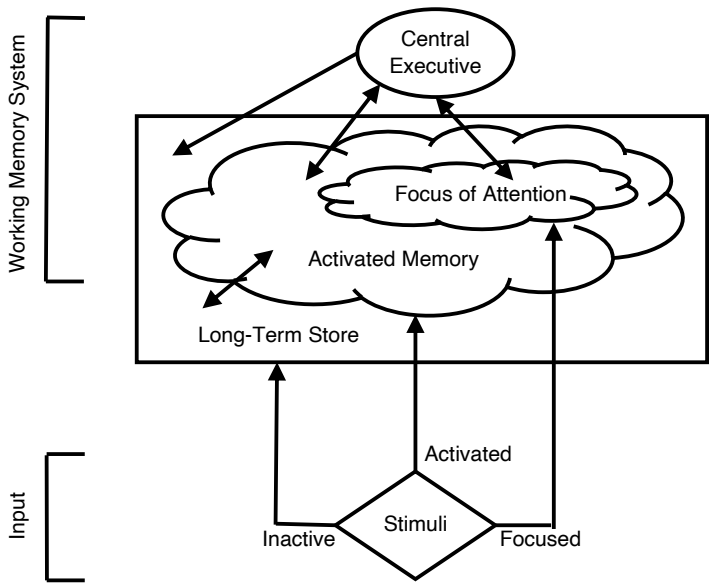


Figure 2.4 Embedded-processes model of working memory
The current version of the embedded-processes model of working memory. Adapted from Cowan (1988, p. 180, Figure 1; 2008, p. 326, Figure 1).

This model defines working memory system as the set of cognitive processes, which hold a limited number of items (about 3 to 5 items) for a limited amount of time (for 10 to 20 seconds unless reactivated) in our focus of attention (Cowan, 1999, 2001; Cowan et al., 2005). Memories that are currently activated and focused may fall into the contents of current working processes. All information items are stored and processed the same way within the same single storage medium, where the activation level of an item constantly changes in accordance with our cognitive goal or task. The attentional processes guide the focus of attention.

This model emphasizes the role of our controlled attention and its interaction with memory. It relies on the attentional activation levels rather than the box metaphors. The information that you care about and you pay attention to is better

stored in the memory system and is better retrieved later. This attentive processing results in more elaborate encoding, which is critical for voluntary retrieval and episodic storage. By controlling our cognitive focus, it can explain how our goal-directed behaviors work.

Control of focus is controlled in two ways: it is controlled voluntarily by the central executive; it is controlled automatically by environmental stimuli. The automatic environmental control is done by the attentional orienting system. For example, when you walk in the dark wood, your focus of attention is voluntarily on the ground, so you do not fall down. While walking, if some twigs grab your shirt, you might think someone is grabbing you to hurt you. Then, your focus of attention is automatically drawn to the twigs away from the ground. Here, you can actively control your attention: the central executive can voluntarily redirect your focus of attention (away from automatic response or distraction, like the twigs, and) back to walking or other goals.

Different from the multi-component model, it does not distinguish modality-specific stores (like visuo-spatial vs. phonological). The focus of attention is somewhat similar to the episodic buffer in the multi-component model, in that it takes all different types of information.

Cowan (1999, p. 62) highlighted five core principles of the embedded-processes model of working memory, where the principles link memory and attention. These were in a chapter of the same book as the five core features of the multi-component model were listed. To give detailed points and compare the two

models that this dissertation is most relevant to, the principles are repeated here:

> (1) Working memory information comes from hierarchically arranged faculties comprising: (a) long-term memory, (b) the subset of long-term memory that is currently activated, and (c) the subset of activated memory that is in the focus of attention and awareness.
> (2) Different processing limits apply to different faculties. The focus of attention is basically capacity limited, whereas activation is time limited. The various limits are especially important under nonoptimal conditions, such as interference between items with similar features.
> (3) The focus of attention is controlled conjointly by voluntary processes (a central executive system) and involuntary processes (the attentional orienting system).
> (4) Stimuli with physical features that have remained relatively unchanged over time and are of no key importance to the individual still activate some features in memory, but they do not elicit awareness (i.e., there is habituation of orienting).
> (5) Awareness influences processing. In perception it increases the number of features encoded, and in memory it allows new episodic representations to be available for explicit recall. (p. 62).

2.2.3 Working memory capacity

Working memory is a capacity-limited system (Engle, 2002; Kane & Engle, 2000, 2003). Overloading the working memory system results in poor performance (Keller, Cowan, & Saults, 1995; Oakhill, Cain, & Bryant, 2003; Seigneuric et al., 2000).

2.2.3.1 Capacity and time limit

Its processing ability is limited by capacity and by time, i.e., by the number of items it can hold at a time and by the temporal duration during which it can hold the items. It can hold about 3 to 5 items (or chunks) for about 10 to 20 seconds at a time, unless rehearsed or reactivated. This helps us to focus on our goal by not being distracted by continuously incoming input from the environment (Barrs & Gage, 2010). While

the capacity limit suggested in the literature ranged from 2 to 9 items, it seems 3 to 5 is supported the best. For example, Paas, Renkle, and Sweller (2003) suggested, if it is new information elements, the real capacity limit is 2 to 3 items. Broadbent (1975) suggested it is 3 items when strategic memorization techniques are eliminated. Cowan, Nugent, Elliott, Ponomarev, and Saults (1999) found that adults have an average limit of 3.5 items, up to which the limit increases as a function of age in childhood. Miller (1956), which is probably the most famous paper on this topic, suggested it is 5 to 9 items (i.e., what he called the magical number 7, plus or minus 2). He maintained that 7 plus or minus 2 is 'the span of absolute judgment' (i.e., the magnitude that we can identify with absolute confidence and accuracy) or 'the span of immediate memory' (i.e., what we can consider the capacity limit) for unidimensional judgments. To an extreme, Baddeley (1986) did not restrict capacity limit but proposed only a time limit for which information can be maintained for ongoing processing. However, experiments have evidenced the existence of a clear capacity limit, where the capacity was smaller than 7 items. Miller (1956) himself noted he would "withhold judgment" (p. 96); many other follow-up studies supported smaller numbers (e.g., Broadbent, 1975; Cowan, 2001, 2005, 2010; Cowan et al., 1999; Cowan, Rouder, Blume, & Saults, 2012; Paas et al., 2003; Sperling, 1960). In a direct answer to the controversy over capacity limit, Cowan (2001) argued a single central capacity limit is 3 to 5 chunks (i.e., what he called the magical number 4, plus or minus 1), by providing a comprehensive review of literature on the capacity limit and direct experimental evidence. For example, he

gave a table of 17 types of evidence (with source studies associated with the types) in the literature (see Table 1 in Cowan, 2001, p. 90). His analysis controlled for conditions that can influence the capacity limit of working memory, e.g., rehearsal, interaction with long-term memory, or refill from sensory memory.

The capacity limit of 3 to 5 works for individual items or chunks (or episodes). A chunk, which contains more than one item, can be defined as "a collection of concepts that have strong associations to one another and much weaker associations to other chunks currently in use" (Cowan, 2001, p. 89). 2 and 4 are two items, but can be made into 24, which is now one item or chunk. To illustrate this better, let us try to remember 14461009050505080515. It may feel challenging to remember a sequence of 20 numbers. However, it can be made easy and doable if you break it into 4 meaningful chunks as follows: (i) Hangeul Day, October 09, 1446, which is 14461009; (ii) Children's Day, 05/05, i.e., 0505; (iii) Parents' Day, 05/08, i.e., 0508; (iv) Teachers' Day, 05/15, i.e., 0515. When the 20 digits are chunked into 4 units, it comes within our capacity limit of 4 and we can successfully remember and recall the 20 digits.

Additionally, the capacity limit works for integrated objects, not sub-features of an object. Cowan (1998) explained the results from Luck and Vogel (1997) that participants' task performance was the same no matter whether the visual discrepancy was in one feature, two features, or four features of each object. He suggested the participants treated the entire object as one information chunk, recognizing the whole as the same or different, regardless of the number of different

sub-features of that object.

It should, however, be noted that the capacity limit can be actively expanded using various mnemonic strategies. Chunking can increase the recall rate by grouping the digits (Jacobs, 1887). Mnemonic strategies help. For example, Miller (1956) listed three strategies to get around the limit, such as giving an approximate answer, recategorizing the items into multiple dimensions and thus multiple questions, and rearranging the task to break it into a series of tasks.

As for the time (or duration) limit, about 20 seconds is suggested to be the time frame within which information can remain active in working memory (Baddeley, 1986; Cowan, 1995). Under the multi-component model of working memory, it was suggested that an item can remain active in working memory for 20 to 30 seconds unless rehearsed (Baddeley, 1986). Rehearsal helps maintain the item longer in working memory (see Baddeley, 1986, for a review). Under the embedded-processes model, it was suggested an item is maintained in the focus of attention for 10 to 20 seconds unless reactivated (Cowan, 1995; Cowan et al., 2005).

There have been reports that other factors can influence the capacity limit. For example, the amount of cognitive load (or the difficulty of the task) can affect the capacity limit. Barrouillet and Camos (2001), suggesting we need a more sophisticated model of working memory capacity than a time limit, showed that (9- and 11-year-old) children performed more poorly when they simultaneously engaged in a more difficult task. They compared children's performance on consonant recall while varying cognitive load/cost but maintaining the same task

duration. They asked 6-year-olds, 8-year-olds, and 11-year-olds to recall consonants, while counting red dots (as the more difficult task) or simply saying baba (as the easier task). They found that children's recall did not differ between counting and baba tasks. By contrast, they found poorer consonant recall performance when the 9-year-olds and the 11-year-olds performed operation tasks (i.e., three-operand additions and two-operand additions) compared to when simply saying baba. They explained it supports the limited resource hypothesis.

The length of the items can also affect the limit. Studies have supported for word-length effect, which states that longer words are harder to remember than shorter words. This is in fact an interaction effect with the duration. Recalling longer words were more difficult than recalling shorter words (Baddeley, Thomson, & Buchanan, 1975). While Caplan, Rochon, and Waters (1992) reported no significant effect of word length in immediate serial recall tasks, Baddeley and Andrade (1994) in a reply refuted that the material design was problematic because the minimal length difference was 1.9% or 2.3% in the experiments.

2.2.3.2 Controversy over information type

Despite tremendous scholarly efforts, there still seems to be controversy over the effect of the load/information/task type. On the one hand, there is a large body of literature that suggested verbal and spatial information take independent storage systems and thus independent capacity limits (e.g., Baddeley, 1986; Baddeley & Logie, 1999; Cocchini, Logie, Sala, MacPherson, & Baddeley, 2002; Luck & Vogel,

2013; Oakhill et al., 2003; Oberauer, 2009; Seigneuric et al., 2000). This view sides with the multi-component model. Verbal working memory involvement in language was clearly implicated (Oakhill et al., 2003), while direct contributions from spatial working memory have been ruled out (Seigneuric et al., 2000). A verbal task, but not a spatial task, is impeded by another verbal task (Baddeley, 1986; Baddeley & Logie, 1999). A verbal task impeded another verbal task (Cocchini et al., 2002). Phonological overlap impaired memory but similarity had a beneficial effect (Oberauer, 2009).

On the other hand, there is another body of literature suggesting that the capacity limit applies within a single unit (i.e., focus of attention), and thus does not distinguish the type of information. This view is in line with the embedded-processes model. The type of a distractor task did not influence performance in delayed tone comparison (Keller, Cowan, & Saults, 1995). The visually presented digits (the verbal distractor) and auditorily presented tones (the auditory distractor) similarly impacted the tone comparison performance. In the verbal condition, participants were asked to remember the numbers covertly rehearsing the names of the numbers during the inter-tone intervals, i.e., they were covertly saying four three seven, and so on, thus a verbal distractor. In the auditory condition, they were asked to covertly rehearse the pitch of the tone, thus a nonverbal auditory distractor). While the participants' performance was superior when rehearsing was allowed than when rehearsal was suppressed, the type of the distractor did not influence performance.

Of the components of working memory, it has been argued that the overload

results from the limited attentional resources in the central executive component of working memory. Engle and colleagues, who have focused their research interests particularly on working memory capacity, suggested that the central executive is specifically predictive of the relationship between measures of working memory and higher-order cognitive functions (evidenced e.g., in the results from a stroop task, Kane & Engle, 2003). They showed that the central executive is especially important to maintain the goal-directed behavior in the presence of interference, in order to neglect the distraction and to override automatic responses (Kane & Engle, 2003). This core role of the central executive, i.e., the function of maintaining the goal in the presence of distractors, has also been supported in neuroscience by means of experiments on the functions of the prefrontal cortex in delayed response tasks. For instance, Caplan, Alpert, Waters, and Olivieri (2000) and Stromswold, Caplan, Alpert, and Rauch (1996) evidenced, through increased blood flow in Broca's area measured with positron emission tomography (PET) in making a syntactic judgment, that the frontal cortex area is activated for cognitive processing. The frontal cortex area was also suggested to embrace the parts of the brain that are responsible for planning and articulation in speech production. Ackermann and Riecker (2004) demonstrated, based on previous work with functional imaging data and their analysis on functional magnetic resonance imaging (fMRI) data, the left anterior part of the brain contributes to the actual coordination of the motor aspects of speech production, controlling about 100 muscles that engage in articulation and phonation.

2.2.3.3 Measurement and manipulation of working memory

Recent researches have been using complex span tasks to measure working memory capacity, replacing the more traditional method of simple span tasks. It has been largely supported that complex span tasks predict better than simple span tasks and are consistently highly correlated with higher-order cognition. Waters and Caplan (1996a) concluded from the results of their experiments that working memory measures would be unreliable without measuring multiple components of processes, such as measures of sentence processing and recall component, and that the predictive power would depend on the overlap of the operations. Waters and Caplan (2003) also noted that a single measure of memory span does not consistently predict performance on different testing sessions and tasks, but a composite measure that reflects performance on multiple tasks improved the test-retest reliability and stability of participant classification for characterizing individuals' working memory performance. Widely used complex span tasks are operation span task, reading span task, and symmetry span task (Kane & Engle, 2003; Unsworth et al., 2005). A thorough analysis of the reliability of the tasks, correlation among different span tasks and with higher-order cognitive processes, and the predictive power of the span task results for the performance in cognitive activities is given in Unsworth et al. (2005), where they presented a demonstration that an automated version of the operation span (Aospan) task, as a sub-type of complex span task, can be a good measure of working memory capacity by having good reliability and validity. Automated versions of the span tasks (Turner & Engle, 1989; Unsworth et al., 2005)

are developed to have an easy-to-administer type of task that requires minimal intervention on the part of the experimenter (Unsworth et al., 2005). It then would be reasonable to expect that using one of the complex span tasks as a measure of working memory capacity will predict and correlate with prosody production, assuming that prosody production is a high cognitive activity that requires comprehension of the sentence and planning of the articulatory motors.

Different from simple span tasks, where people perform only the main task, complex span tasks ask participants to engage in a secondary task (distractor task) in addition to the main/primary task. For instance, in an operation span task, participants solve a series of math problems while trying to remember a set of unrelated consonant letters (such as a sequence of 3 to7 letters of F, H, J, K, L, N, P, Q, R, S, T, and Y, in Unsworth et al., 2005, or B, F, H, J, L, M, Q, R, and X, in the prescreening in Unsworth et al., 2007). In an automated version of the task, each participant will first perform a few practice lists, see a sequence of letters, each letter appearing on the screen (see Unsworth et al., 2009) for 800 or 1000 ms, then solve a math problem as presented in an operation-word string (such as "(9 / 3) + 2 = 5?", from Kane & Engle, 2003), and at recall be asked to click on the box next to the appropriate letters in the correct order (serial recall). After the recall, feedback is given for the number of letters correctly answered. The Ospan score was the sum of recalled words for all of the sets in which the entire set was recalled in the correct order. Scores ranged from 0 to 75. A more detailed description of complex span tasks is provided in Unsworth et al. (2005, 2009) for each sub-type of tasks with the

correlation results.

If one is to examine how people with different working memory capacity perform differently, i.e., the difference between high vs. low working memory capacity, it is recommended to select the upper and the lower quartile of all the participants in a participant-screening task. Waters and Caplan (2003) warned that it would be highly unreliable to predict other cognitive performances if we include all participants and classify them into high-, medium-, and low-span groups without leaving anyone uncategorized or setting a cutoff score for grouping. They suggested having upper and lower quartiles of the distribution to have high- and low-span individuals, as it would be a better and reliable way to examine individual differences in working memory capacity. We found other researchers also prescreened participants to have the top and bottom quartile of the distribution, instead of including all participants. For example, Kane and Engle (2003) and Unsworth et al. (2007) first screened individuals for working memory capacity with the ospan task. On the basis of the ospan scores, they invited the top and the bottom quartile of the ospan distribution, which they call the high-span individuals and low-span individuals, respectively. All of the screened participants are those who correctly answered at least 85% of the ospan operations.

CHAPTER III.

WORKING MEMORY IN L1 SPEECH PRODUCTION[3]

3.1 Introduction

The purpose of the present study reported in this chapter is to investigate the effects of working memory on speech produced by native speakers of English. Native speakers of American English produced syntactically complex sentences under two cognitive load conditions and under two control no-load conditions. Syntactically complex sentences were used both to maximize sentence length, and therefore the number of potential errors, and also to ensure different prosodification of the sentences, and therefore the length of planning chunks (see, e.g., Shattuck-Hufnagel, 2015). The working memory load tasks taxed either spatial or verbal working memory. Speakers solved equations before speaking in the control conditions.

The study is presented in two parts: one for phonological encoding and the other for phonetic encoding. The first part of this chapter (section 3.2) describes an analysis relevant to phonological encoding. The hypothesis of phonological encoding predicts an increase in segmental errors under verbal working memory load relative to the control condition, but not under spatial working memory load.

The second part of this chapter (section 3.3) focuses on the acoustic patterns

[3] Part of the work in this chapter but with the results from the initial analysis was previously reported in the peer-reviewed proceedings from the 18th International Congress of Phonetic Sciences: Lee, O., & Redford, M. A. (2015). Verbal and spatial working memory load have similarly minimal effects on speech production. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (Paper number 0798, pp. 1-5). Glasgow, UK: University of Glasgow. ISBN 978-0-85261-941-4.

of speech output in the load/no-load matched sentences that were fluently and correctly produced by the same speaker. The dependent measures were global speech patterns, including rhythm and pitch measures, and a specific acoustic correlate of segmental production, namely, vowel articulation. The phonetic encoding hypothesis predicts effects of verbal working memory load on acoustic patterns of produced speech.

3.2 Phonological Encoding

3.2.1 Methods

3.2.1.1 Participants

Participants were twenty (13 males and 7 females) college-aged adult native speakers of American English, recruited from the Psychology and Linguistics Human Subjects Pool at the University of Oregon. All reported normal hearing and speaking and no history of speech-language therapy. All were compensated with course credit for their time.

3.2.1.2 Speech materials

The speech materials manipulated sentence structure in order to elicit differently prosodified sentences while controlling for phrase length. 32 sentences (Appendix I) were designed around 4 structures (* 8 sentences) manipulating the dependent

relative clause (RC): the RC was either subject-extracted or object-extracted (e.g., *the smart shy boy that liked the quiet girl cut the cake* versus *the fat black cat that the mad dog hurt climbed the tree*); and, the RC was either embedded in the middle of the matrix clause or appended to the end of it (e.g., *the sly gray wolf bit the sheep that wore the gold bell* and *the swank rich man bought the paint that the young girl chose*). Each sentence consisted of 12 monosyllabic words: 3 occurrences of the definite determiner *the*, 3 adjectives, 3 nouns, 2 verbs, and 1 relativizer *that*.

3.2.1.3 Working memory manipulation

The task manipulated the type and the condition of working memory load during speaking using a modified complex working memory span task (adapted from Kane & Engle, 2003; Unsworth, Heitz, Schrock, & Engle, 2005). Load type was manipulated to tax either verbal or spatial working memory. In the load condition, a speaker was required to hold onto a sequence of 4 letters (verbal) or 4 spatial locations (spatial) while speaking aloud one of the stimulus sentences, which was displayed on the computer monitor. After speaking the sentence (primary task), the speaker completed the load task by choosing either the correct letter sequence or the correct spatial combination from among a set of 8 options (distractor task). In the control, no-load condition, participants were presented with a sequence of 4 numbers and asked to choose the correct sum from among 8 options before speaking the stimulus sentence. Because the participants completed answering the sum before speaking, no additional memory load was given during speaking other than reading

aloud the sentence. From a procedural perspective, this meant that the primary production task (P in Figure 3.1) came either between the serial presentation of to-be-remembered items and recall (R in Figure 3.1, load conditions) or after recall (no-load condition). Figure 3.1 illustrates the different tasks.

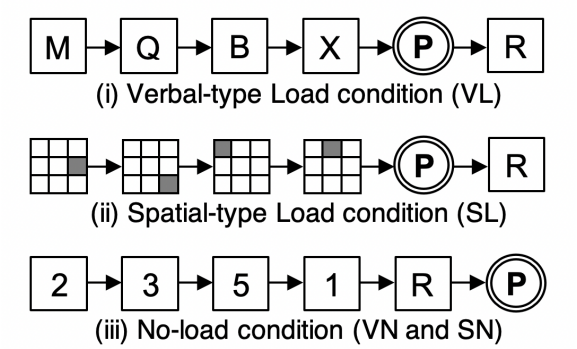


Figure 3.1 Working memory manipulation during speaking
The task manipulated the type and the condition of working memory taxed. The four consonants, grids, or numbers were serially presented one at a time in the order in which they were to be remembered (or collated or summed). The sentence to be produced (P) was presented either before or after the 8 options used to test recall (R).

During presentation, each letter / spatial location / number serially remained on a computer monitor for 800 milliseconds. Each sentence to be produced was displayed on a single line for 8 seconds, as were the 8 response options. Across all conditions (VL, SL, VN, SN), the list length was 4 (letters, locations, and numbers). All letters in the verbal load condition were out of 9 consonants, i.e., B, F, H, J, L, M, Q, R, and X (the same as used in Unsworth et al., 2005), in non-permissible sequences according to English phonotactics; none of the sequences formed acronyms. The spatial load condition presented non-overlapping locations of a gray cell within a 3 x 3 white grid. The no-load condition displayed natural numbers out

of 1 to 9, where the sum ranged from 9 to 16. The response options given during recall were all highly confusable in that every option repeated a part of the correct answer for the load conditions and that the incorrect options were close numbers to the correct answer for the control conditions. The difficulty of the task ensured that working memory would in fact be taxed. Most participants reported thinking that the primary goal of the experiment was to assess working memory and that sentence elicitation was secondary to this goal (i.e., used as a distractor task). The response scores were consistent with this feedback from participants, who managed 57.3% (M = 9.16, SD = 3.52) correct responses in the verbal load condition, 68.1% (M = 10.89, SD = 3.51) in the spatial load condition, and 93.3% (M = 14.92, SD = 1.19) in the no-load conditions. All scores were well above the chance performance of 12.5%.

3.2.1.4 Elicitation procedure

Data was collected in a within-subjects design, with two fixed factors: Load Type (verbal, spatial) and Load Condition (load, no-load). All participants produced all the sentences and engaged in all levels of Load Type and Load Condition. All participated in the study with the same researcher throughout the entire experiment.

Prior to the main experimental task, participants were given as much time as they needed to read through the 32 sentences. This familiarization procedure was intended to control for effects of language planning and comprehension, by bypassing the language-planning process that prepares meaning and the associated forms and giving participants enough time to comprehend the sentences. Each

sentence was presented in exactly the same format as to be shown in the main production phase., i.e., a single sentence was displayed in a single line located in the middle of a slide. Each sentence was shown in two consecutive slides. The second slide highlighted the sense of the matrix clause, which was to encourage participants to pay attention to and remember the meaning. See a sample in Appendix III. We emphasized that participants have confidence in their comprehension of all sentences before they began the main task.

Once participants verbally expressed confidence in their comprehension of all the sentences, they proceeded to the main task, which was blocked by Load Type and Load Condition. Between the four experimental blocks, participants were given opportunity to take a water/bathroom break. They were allowed to take the microphone off during breaks. At the beginning of each block, participants were provided with practice, which was comprised of both the span and elicitation tasks. This practice session used a simple sentence, *a happy fish was swimming in the river*. Once participants had completed the practice session, they clicked on the computer screen to proceed to the main task for that block.

For the main task, the 32 sentences were divided into two sets of 16 sentences with four sentences from each of the four syntactic structures. Sentence assignment to a particular set was randomized for each participant. Each set was then assigned to a Load Type (verbal or spatial) and elicited in random order during the associated load condition and no-load conditions. The order in which participants completed elicitation under Load Type and Load Condition was also randomized. The fixed

sequences of letters / locations / numbers were also randomly paired with the 16 sentences. Accordingly, each speaker produced a total of 64 sentences.

Participants' speech was digitally recorded for later analysis using a Marantz PMD660 and Shure ULXS4 standard wireless receiver and lavaliere microphone. The microphone was attached to a hat that participants were given to wear. The entire experiment took no more than 90 minutes to complete.

3.2.1.5 Speech error determination

Data from one participant was excluded at the outset because the majority of his productions were not recorded due to technical difficulties. The author listened to the remaining 1,216 sentences (= 19 participants * 32 sentences * 2 productions), and determined that 56 additional items should be excluded from analysis due to a non-linguistic disruption during production (e.g., yawning or coughing). The remaining 1,160 sentences were then measured and coded in the following way:

Each utterance was first transcribed and then coded as correct or incorrect productions. Incorrect productions included disfluencies and/or speech articulation errors. Disfluencies were defined (following three out of four hesitation types in Maclay & Osgood, 1959) as (i) filled pause (e.g., *whipped the poor <uh>*), (ii) false start (e.g., *<the wa-> the wild bad guy* for *the wild bad guy*), and (iii) repeat (e.g., *the nice large cow <cow>*). Speech (articulation) errors were categorized (referring to Shriberg, 1994) as either phonemic or lexical and defined (following Shattuck-Hufnagel, 1979) as (i) addition (e.g., *we<f>t* for *wet* or *aunt cleaned <up>* for *aunt*

cleaned), (ii) omission (e.g., *spot*() for *spots* or *the smart* () *boy* for *the smart shy boy*), (iii) substitution (e.g., {*th*}*ick* for *sick* or {*girl*} for *friend*), (iv) exchange (e.g., *g*{*lu*}*ped* for *gulped* or *that* {*the had*} for *that had the*), and (v) shift (e.g., *the great cooked bake* for *the great cook baked*; from Shattuck-Hufnagel, 1979, *myn ow way* for *my own way* or *give youing* for *giving you*).

Note that unfilled pauses (the fourth type in Maclay & Osgood, 1959) and lengthening (without semantic change, Shriberg, 1994, e.g., [*s*~]*ick aunt, th*[*e*~]), which are traditionally disfluencies, were here coded as correct. That is, as long as the segment-wise transcript was correct as targeted, suprasegmental variations were treated as correct and included in the acoustic analyses in section 3.3. This was (i) to avoid subjectivity bias in determining the "unusual" length of pauses or phonemes, (ii) to examine any durational changes due to working memory load (as in section 3.3), and (iii) to compare the L1 speech with the L2 using the same criteria. First, as pointed out in Maclay and Osgood (1959, p. 24), the judgment on unusual silence and on unusual non-phonemic lengthening of phonemes may vary with different listeners and different speakers, in that different listeners are familiar with different pace and speech style and that the same length can be judged fluent or disfluent depending on the speech rate. Likewise, all errors in Shattuck-Hufnagel (1979) resulted in a non-target segment-wise transcript. Second, given a part of the current purposes is to analyze durational changes due to different working memory loads during production, it should be crucial to look into durational differences, even disfluencies, in different conditions. Third, especially for L2 speakers, and partly for

L1, it may be controversial to set a boundary as to usual vs. unusual length.

3.2.1.6 Statistical analyses

Generalized linear mixed-effects logistic regression models (with logit-link function, implemented in the package lme4, Bates, Maechler, Bolker, & Walker, 2015, of software R, version 3.6.3, R Core Team, 2020) were used in combination with multimodel inference (implemented in the R package MuMIn, Bartoń, 2019) and likelihood ratio tests (operationalized with anova function in R). For significant interactions, simple effects coefficients were computed, where alpha was adjusted for the number of simple effects tests for each factor, e.g., .05/2 = .025, to maintain the probability of Type I error at .05.

The following variables were tested in candidate models to justify the best model fit: the dependent variable was presence or absence of speech error in a sentence, henceforth speech error (1 = with one or more speech error in a sentence, 0 = no speech error in a sentence); fixed factors (or predictors) included Load Type (verbal, spatial), Load Condition (load, no-load), RC Type (subject-extracted, object-extracted), and RC Location (middle, end); random intercepts were for speakers (Speaker), sentences (Sentence), and relative order (RelOrder, 1 = first production of a sentence, 2 = second production of the same sentence by the same speaker). Within-unit random slopes justified by the design were included and tested, i.e., of Load Type within speakers, Load Condition within speakers, Load Type within sentences, Load Condition within sentences, RelOrder within speakers, and

RelOrder within sentences, following the recommendation to maintain maximal design-driven random effects structure in order to minimize Type I error and have power advantage especially for within-unit designs and for confirmatory hypothesis testing (Barr, Levy, Scheepers, & Tily, 2013). Between-unit random slopes were excluded, which anyway was not identified in the current experiments, because a random intercept can be sufficient for a between-unit factor (Barr et al., 2013, p. 275).

To avoid fitting overly complex models and to balance Type I error and power (as advised in Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), we started off with the maximal random effects structure and progressively simplified the complexity until convergence was reached: we removed only the random effects required to allow convergence and non-singular fit (Barr et al., 2013) and used model selection criterion (Matuschek et al., 2017) via multimodel inference and likelihood ratio tests. Through this, the optimal maximal structure was justified by the data: while the design-driven maximal structure is recommended for any kind of data, the data-driven structure would be more justifiable for categorical data such as the current ones (Barr et al., 2013, p. 274-5).

Multimodel inference was used to deal with model selection uncertainty (Burnham & Anderson, 2002; Kuperman & Bresnan, 2012; Barth & Kapatsinski, 2014). With multimodel inference, instead of selecting the single best regression model, we made formal and reliable inferences based on the entire set of models, ranking the fitted models and estimating the relative importance of variables

(Burnham & Anderson, 2002, Ch.4). We derived regression coefficients by averaging them over all regression models that can be derived by selecting a subset of our predictors (including the full set), with each model weighted by its predictiveness. The predictiveness was measured using the Akaike Information Criterion (AIC, Akaike, 1973), adjusted with small-sample bias correction (AIC_c, derived as second-order bias-adjustment variant of AIC for small samples by Sugiura, 1978, and criterion developed by Hurvich & Tsai, 1989). The use of AIC_c is advocated when the ratio n/K (where n is the sample size and K is the number of estimable parameters) is smaller than 40 (Burnham & Anderson, 2002, pp. 66). The best (or most plausible) candidate model has the highest Akaike weight (w_i), ranging from 0 to 1 (Burnham & Anderson, 2002, p. 75).

The best (and most simple) model fit for the data was additionally justified by likelihood ratio tests (see description and examples in Baayen, Davidson, & Bates, 2008). It checks if adding a predictor improves the predictability of the model. The null hypothesis is that the two compared models, where one model adds one predictor to the other, are not different in the predictability. A p-value higher than .05 justifies that adding a predictor does not contribute to better predictability and thus that the simpler model should be preferred to make the model parsimonious; a p-value lower than .05 rejects the null hypothesis and supports the additional predictor be included in the analysis.

To present variance explained (R^2) for a generalized mixed-effects model, two types of R^2 were obtained, marginal R^2 ($R^2_{GLMM(m)}$) and conditional R^2 ($R^2_{GLMM(c)}$)

(as proposed in Nakagawa & Schielzeth, 2013), where each R^2 uses both observation-level variance (σ^2_ε) and distribution-specific theoretical variance (σ^2_d) (as recommended in Nakagawa, Johnson, & Schielzeth, 2017). As described in Nakagawa and Schielzeth (2013, p. 137), marginal R^2_{GLMM} represents the variance explained by the fixed factors; conditional R^2_{GLMM} represents the variance explained by the entire model, including both fixed and random factors; the difference between the two values indicates the variability in random effects (p. 140). While these values can be derived with a few different methods, e.g., theoretical method, the delta method, lognormal approximation, and the trigamma function, Nakagawa et al. (2017) recommended using observation-level variance calculated with the delta method in addition to, if not replacing, distribution-specific variance calculated with the theoretical method, especially for data in binomial distributions and/or with logit-link function (such as the current data). They argued, for example, R^2_{GLMM} calculated using the delta method via observation-level variance approximated to R^2 on the original observation scale and that its distinction from the distribution-specific variance was strictly appropriate for binomial distributions because of the often-found large difference between the two R^2 values. All values reported here were obtained in the implemented R version using the revised statistics following Nakagawa et al. (2017, noted in Bartoń, 2019, p. 53).

3.2.2 Results

In total, 24.3% (N = 282) out of the 1,160 sentences were incorrectly produced with at least one speech disfluency and/or error. There was a total number of 40 disfluencies (an average of 0.14 per incorrectly produced sentence) and of 430 errors (1.67 per sentence). False start (N = 17, 42.5%) and repeat (N = 17, 42.5%) were equally more frequent types of disfluency than filled pause (N = 6, 15.0%). Lexical errors were less frequent than phoneme errors (N = 146, 34.0% vs. N = 284, 66.0%). For lexical errors, omission (N = 63, 43.2%) was most frequent, followed by substitution (N = 61, 41.8%), addition (N = 17, 11.6%), exchange (N = 4, 2.7%), and shift (N = 1, 0.7%). For phonemic errors, addition (N = 161, 56.7%) was most frequent, followed by omission (N = 66, 23.2%), substitution (N = 57, 20.1%). None of observed of exchange or shift. Table 3.1 summarizes the distribution of sentences produced with speech errors by Load Type, Load Condition, RC Type, and RC Location. Numbers of errors are indicated in the numerator; the total numbers of sentences are indicated in the denominator. In total, the proportion of sentences with one or more speech error was: verbal load M = 32.4%, 94/290; verbal no-load M = 21.1%, 61/289; spatial load M = 28.2%, 81/287; spatial no-load M = 15.6%, 46/294; subject middle M = 27.5%, 83/302; subject end M = 19.8%, 60/303; object middle M = 21.1%, 54/256; object end M = 28.4%, 85/299.

Table 3.1 Proportion of sentences with speech error in L1 speech

Proportion and [cumulative number] of sentences with speech error (as numerators) out of total sentences (as denominators) in L1 speech: by Load Type (verbal, spatial), Load Condition (load, no-load), Relative-Clause Type (subject-extracted, object-extracted), and Relative-Clause Location with respect to matrix clause (middle, end).

		Verbal		Spatial		
Relative clause		Load	No-load	Load	No-load	Total
Subject-extracted	Middle	32.0% [24/75]	26.7% [20/75]	32.9% [25/76]	18.4% [14/76]	27.5% [83/302]
	End	25.0% [19/76]	16.0% [12/75]	23.7% [18/76]	14.5% [11/76]	19.8% [60/303]
Object-extracted	Middle	27.0% [17/63]	17.5% [11/63]	25.4% [16/63]	14.9% [10/67]	21.1% [54/256]
	End	44.7% [34/76]	23.7% [18/76]	30.6% [22/72]	14.7% [11/75]	28.4% [85/299]

A generalized linear mixed-effects logistic regression was used to predict the presence of speech disfluency or error in a sentence (henceforth speech error) based on working memory load type and load condition. The model included: speech error as the dependent variable; Load Type, Load Condition, and their interaction as fixed factors; speakers, sentences, and relative order as random factors. The total number of observations was 1,160. The results are in Table 3.2.

The results indicated that only Load Condition was a statistically significant predictor of speech error, $b = -.91$, $se(b) = .22$, $p < .001$. Neither Load Type nor the interaction between the two working memory factors were statistically significant, Load Type, $b = .20$, $se(b) = .19$, $p = .299 > .05$, and the interaction, $b = .19$, $se(b) = .29$, $p = .510 > .05$. The AIC was 1223.28. The BIC was 1258.67. The log-likelihood was -604.64. The working memory factors explained about 3.1% of the variance in speech error. Variance explained by the fixed factors (marginal R^2_{GLMM}, $R^2_{GLMM(m)}$) was 3.1% (observation-level variance σ^2_ε) or 4.6% (distribution-specific

theoretical variance σ^2_d). Variance explained by the entire model (conditional R^2_{GLMM}, $R^2_{GLMM(c)}$), including both the fixed and the random effects, was 12.7% (σ^2_ε) or 19.3% (σ^2_d).

Table 3.2 Generalized linear mixed effects logistic regression results for working memory effects on L1 speech error

Generalized linear mixed effects logistic regression results for working memory effects on L1 speech error (N = 1,160).

Fixed effects:	***b* [95% CI]**	***se(b)***	***z***	***p***
(Intercept)	-.99 [-1.87, -.11]	.31	-3.23	.001 **
Load Type	.20 [-.18, .58]	.19	1.04	.299
Load Condition	-.91 [-1.34, -.48]	.22	-4.17	.000 ***
Load Type * Load Condition	.19 [-.39, .77]	.29	0.66	.510
Random effects:	***σ²***	***SD***		
Speaker (Intercept)	.22	.47		
Sentence (Intercept)	.27	.52		
RelOrder (Intercept)	.11	.33		
AIC	1223.28			
BIC	1258.67			
Log*L*	-604.64			
$R^2_{GLMM(m)}$_σ^2_ε; $R^2_{GLMM(m)}$_σ^2_d	.03; .05			
$R^2_{GLMM(c)}$_σ^2_ε; $R^2_{GLMM(c)}$_σ^2_d	.13; .19			

Note. Estimates of fixed-effects regression coefficient (*b*), 95% confidence intervals (CI) calculated by the confint function, standard error of regression coefficient (*se*(*b*)), *z*-value (*z*), *p*-value (*p*), variance (σ^2), standard deviation (*SD*), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (Log*L*), variance explained by the fixed effects (marginal R^2, $R^2_{GLMM(m)}$) obtained using observation-level variance (σ^2_ε) and distribution-specific theoretical variance (σ^2_d), and variance explained by the entire model (conditional R^2, $R^2_{GLMM(c)}$) using σ^2_ε and σ^2_d. Significance: *** < .001, ** < .01, * < .05.

Speech error rate statistically significantly increased by 11.9% when the L1 speakers were multitasking while speaking. More sentences were produced incorrectly at an error rate of 30.3% (= 175/577) during the load conditions compared to 18.4% (= 107/583) in the no-load conditions. Load type had no effect on the distribution of errors in the analyses that controlled for individual differences

between speakers, sentences, and the order in which sentences were spoken: speakers produced statistically the same number of sentences incorrectly in the verbal working memory load condition (26.8% = 155/579) as they did in the spatial working memory condition (21.9% = 127/581); the data indicated large speaker- and sentence-variance (Appendix V). Figure 3.2 plots the higher probability of a sentence to be produced with speech error in the load conditions than in the no-load conditions, regardless of the load type. The shaded area around the regression line represents the standard error of the regression coefficient, which was .07 for verbal load, .05 for verbal no-load, .06 for spatial load, and .04 for spatial no-load condition.

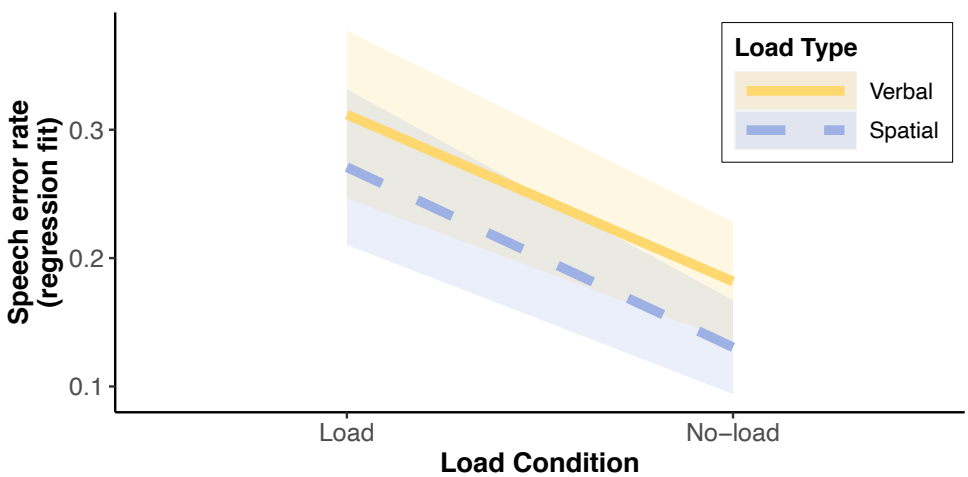


Figure 3.2 L1 speech error rate by working memory load type and condition
L1 speech error rate (in means of regression model fit) predicted by working memory load type and load condition. The colored area around the regression lines denotes standard error of regression coefficient.

The above regression model was selected as the best to account for the presence of speech error in the sentences produced by the native speakers of English. It added an interaction term to the best model suggested by multimodel inference

and likelihood ratio tests. First, multimodel inference and likelihood ratio tests suggested Load Type and Load Condition as the best predictor set. By means of dredge function in the package MuMIn, Multimodel inference automatically selected the best model comparing all possible combinations of the fixed factors Load Type, Load Condition, RC Type, and RC Location, including all two-way, three-way, and four-way interactions. By comparing a total of 32,767 different models, it examined which were the most important fixed factors to predict the variation in speech error. As shown in Table 3.3, it inferred that the best predictor set was the Load Type and Load Condition with a weight of .098. It was 2.33 (= .098/.042) times better to explain the speech-error variation in the data than having Load Condition only; it was 7 (= .098/.014) times more likely so than including both the linguistic factors (i.e., RC Type and RC Location) and the cognitive factors (i.e., Load Type and Load Condition); it was 7.54 (= .098/.013) times more likely so than including the linguistic factors with their interaction and the cognitive factors with their interaction. The three-way or four-way interactions were not recommended, at least not in the top 47 models. The same best predictor set and the same ranks maintained when the interactions between the predictors were excluded from the comparison. The same best model was selected with a weight of .369; including all four tested predictors had a weight of .052. The ranks were exactly the same as in Table 3.3 if we read only the lines with the main effects.

Table 3.3 Model selection table that analyzes relative predictiveness of predictors for L1 speech error

Model selection table that analyzes relative predictiveness of predictors (or fixed factors) for L1 speech error: from an automated multimodel inference analysis (using dredge function) comparing all possible combinations of the predictors of Load Type (LT), Load Condition (LC), RC Type (RT), and RC Location (RL). Predictor variables, number of parameters (K), log-likelihood (Log$L$), Akaike's Information Criterion with small-sample bias correction (AIC_c), AIC_c differences (Δ_i), and Akaike weights (w_i) for a complete set of candidate models for predicting speech error presence.

Predictor variables (predictors)	K	Log$L$	AIC_c	Δ_i	w_i
LC, LT	6	-604.85	1221.8	0.00	.098
LC, LT, LC*LT	7	-604.64	1223.4	1.60	.044
LC	5	-606.71	1223.5	1.69	.042
LC, LT, RT	7	-604.80	1223.7	1.91	.038
LC, LT, RL	7	-604.85	1223.8	2.01	.036
LC, LT, RL, RT, RL*RT	9	-603.03	1224.2	2.44	.029
LC, LT, RL, LC*RL	8	-604.21	1224.5	2.76	.025
LC, LT, RT, LT*RT	8	-604.24	1224.6	2.83	.024
LC, LT, RT, LC*RT	8	-604.34	1224.8	3.02	.022
LC, LT, RL, RT, LC*RL, RL*RT	10	-602.38	1224.9	3.17	.020
LC, LT, RL, RT, LT*RT, RL*RT	10	-602.50	1225.2	3.41	.018
LC, LT, RL, LT*RL	8	-604.54	1225.2	3.42	.018
LC, LT, RT, LC*LT	8	-604.58	1225.3	3.51	.017
LC, LT, RL, RT, LC*RT, RL*RT	10	-602.57	1225.3	3.55	.017
LC, RT	6	-606.65	1225.4	3.58	.016
LC, LT, RL, LC*LT	8	-604.64	1225.4	3.62	.016
LC, RL	6	-606.70	1225.5	3.70	.015
LC, LT, RL, RT, LT*RL, RL*RT	10	-602.73	1225.6	3.86	.014
LC, LT, RL, RT	8	-604.79	1225.7	3.92	.014
LC, LT, RT, LC*RT, LT*RT	9	-603.77	1225.7	3.92	.014
LC, LT, RL, RT, LC*LT, RL*RT	10	-602.82	1225.8	4.06	.013

Note. The models are sorted from best (top) to worst (bottom), ranked by AIC_c. Random factors were Speaker, Sentence, and RelOrder across the candidate models. Random slopes were excluded, e.g., Load Condition within speakers was excluded from the models, because there were more than 50 warnings, mostly nonconvergence and singular fit warnings. The comparison included all possible combinations of two-way, three-way, and four-way interactions (*). The bottom 32,746 rows are omitted.

Likelihood ratio tests also justified removing the linguistic predictors: RC Type, $\chi^2(1) = 0.12$, $p = .732 > .05$, and RC Location, $\chi^2(1) = 0.01$, $p = .921 > .05$. They did not improve the predictability of the model. A regression analysis supported that the linguistic factors did not contribute to predicting the presence of speech error in a sentence. As shown in Table 3.4b, none of the effects were

statistically significant: RC Type, b = -.50, $se(b)$ = .31, p = .103 > .05; RC Location, b = -.41, $se(b)$ = .32, p = .187 > .05; their interaction, b = .86, $se(b)$ = .44, p = .051 > .05.

Finalizing the selection of the fixed factors, we added the interaction term between the best predictor set of Load Type and Load Condition. We interpreted that the interaction term was considered less important in multimodel selection because there was no significant interaction effect, as shown in Table 3.2. However, it would be important to examine the interaction and report that there was no significant interaction in the L1 data. It would also help to have the same model for both L1 and L2 in making a direct comparison of the two; as to be discussed in section 4.2.2, we found a significant interaction between the cognitive predictors in the L2 data. Moreover, including the interaction term as in Table 3.2 to the best model as in Table 3.4a should not change the conclusion of the current hypothesis testing, because both the models drew the same the statistical conclusion. The model without the term, like the one with the term, supported a significant effect of Load Condition, b = -.80, $se(b)$ = -.51, p < .001, but not of Load Type, b = .28, $se(b)$ = .15, p = .052 > .05. Without interaction term, the variance explained decreased but negligibly by 0.1 to 0.2%: variance explained by the fixed factors was 2.9% (σ^2_ε) or 4.5% (σ^2_d); variance explained by the entire model was 12.6% (σ^2_ε) or 19.1% (σ^2_d).

Table 3.4 L1 speech error predicted by different regression models

L1 speech error predicted by different regression models with various predictors: (a) the best model justified by multimodel inference and likelihood ratio tests and (b) a model with both the linguistic and the cognitive predictors.

(a) the best model (N = 1,160)

Fixed effects:	***b* [95% CI]**	***se(b)***	***z***	***p***
(Intercept)	-1.03 [-1.91, -.16]	.30	-3.44	.001 ***
Load Type	.28 [-.01, .57]	.15	1.94	.052
Load Condition	-.80 [-1.10, -.40]	-.51	-5.37	.000 ***
Random effects:	***σ²***	***SD***		
Speaker (Intercept)	.22	.47		
Sentence (Intercept)	.27	.52		
RelOrder (Intercept)	.11	.33		
AIC	1221.71			
BIC	1252.04			
Log*L*	-604.85			
$R^2_{GLMM(m)}$_$\sigma^2\varepsilon$; $R^2_{GLMM(m)}$_σ^2_d	.03; .04			
$R^2_{GLMM(c)}$_$\sigma^2\varepsilon$; $R^2_{GLMM(c)}$_σ^2_d	.13; .19			

(b) a model with both linguistic and cognitive fixed factors (N = 1,160)

Fixed effects:	***b* [95% CI]**	***se(b)***	***z***	***p***
(Intercept)	-.75 [-1.66, .14]	.35	-2.13	.033 *
Load Type	.20 [-.18, .58]	.19	1.04	.295
Load Condition	-.90 [-1.34, -.48]	.22	-4.16	.000 ***
Load Type * Load Condition	.19 [-.39, .77]	.29	0.65	.514
RC Type	-.50 [-1.12, .13]	.31	-1.63	.103
RC Location	-.41 [-1.06, .23]	.32	-1.32	.187
RC Type * RC Location	.86 [-.04, 1.76]	.44	1.96	.051
Random effects:	***σ²***	***SD***		
Speaker (Intercept)	.22	.47		
Sentence (Intercept)	.21	.46		
RelOrder (Intercept)	.11	.33		
AIC	1225.64			
BIC	1276.21			
Log*L*	-602.82			
$R^2_{GLMM(m)}$_$\sigma^2\varepsilon$; $R^2_{GLMM(m)}$_σ^2_d	.04; .06			
$R^2_{GLMM(c)}$_$\sigma^2\varepsilon$; $R^2_{GLMM(c)}$_σ^2_d	.12; .19			

Note. Estimates of fixed-effects regression coefficient (*b*), 95% confidence intervals (CI) calculated by confint function, standard error of regression coefficient (*se*(*b*)), *z*-value (*z*), *p*-value (*p*), variance (σ^2), standard deviation (*SD*), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (Log*L*), variance explained by the fixed effects (marginal R^2, $R^2_{GLMM(m)}$) obtained using observation-level variance ($\sigma^2\varepsilon$) and distribution-specific theoretical variance (σ^2_d), variance explained by the entire model (conditional R^2, $R^2_{GLMM(c)}$) using $\sigma^2\varepsilon$ and σ^2_d. Significance: *** < .001, ** < .01, * < .05.

Second, the data justified controlling for differences in speakers, sentences, and the relative order of the production of a sentence by the same speaker, when predicting the effects of the cognitive factors on producing a sentence with speech error. We started off with the maximal random effects structure justified by the design and examined whether progressively simplified models reached convergence and avoided singular fit. Having Load Type, Load Condition, and their interaction as predictors, we found that the maximal structure justified by the data was the one having Speaker, Sentence, and Relative Order as random intercepts and Load Type within Speaker as random slope. Table 3.5 summarizes the model and its results. As the model without the slope (in Table 3.2), maintaining the maximal random effects structure drew the same statistical conclusion about the fixed effects: a significant effect of Load Condition, b = -.90, $se(b)$ = .22, $p < .001$; insignificant Load Type, b = .17, $se(b)$ = .21, (a higher p-value,) $p = .424 > .05$; insignificant interaction, b = .16, $se(b)$ = .29, $p = .590 > .05$. The amount of the explained variance appeared equivalent: the fixed effects explained 0.2% to 0.1% less variance, explaining 2.9% (σ^2_ε) or 4.5% (σ^2_d); the entire model explained 12.6% (σ^2_ε, 0.1% less) or 19.7% (σ^2_d, 0.4% more) of the total variance in speech error.

Table 3.5 Regression results of a model with maximal random effects structure for L1

Regression results of a model including the maximal random effects structure justified in the L1 data (N = 1,160).

Fixed effects:		***b* [95% CI]**	***se(b)***	***z***	***p***
(Intercept)		-.97 [-1.85, -.09]	.30	-3.26	.001 **
Load Type		.17 [-.26, .59]	.21	0.80	.424
Load Condition		-.90 [-1.33, -.48]	.22	-4.17	.000 ***
Load Type * Load Condition		.16 [-.42, .74]	.29	0.54	.590
Random effects:		**σ^2**	***SD***	***r***	
Speaker	(Intercept)	.02	.16		
Speaker	(Intercept)	.08	.28		
	Load Type	.11	.34	.89	
Sentence	(Intercept)	.26	.51		
RelOrder	(Intercept)	.11	.33		
AIC		1226.55			
BIC		1277.11			
Log$L$		-603.28			
$R^2_{GLMM(m)}$_$\sigma^2\varepsilon$; $R^2_{GLMM(m)}$_σ^2_d		.03; .05			
$R^2_{GLMM(c)}$_$\sigma^2\varepsilon$; $R^2_{GLMM(c)}$_σ^2_d		.13; .20			

Note. Estimates of fixed-effects regression coefficient (b), 95% confidence intervals (CI) calculated by the confint function, standard error of regression coefficient ($se(b)$), z-value (z), p-value (p), variance (σ^2), standard deviation (SD), correlation (r), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (Log$L$), variance explained by the fixed effects (marginal R^2, $R^2_{GLMM(m)}$) obtained using observation-level variance ($\sigma^2\varepsilon$) and distribution-specific theoretical variance (σ^2_d), variance explained by the entire model (conditional R^2, $R^2_{GLMM(c)}$) using $\sigma^2\varepsilon$ and σ^2_d. Significance: *** < .001, ** < .01, * < .05.

However, despite our intention to include the maximal random effects structure justified by the L1 data, we determined to remove the random slopes, mainly to match the model with the one for the L2 analysis. The L2 data did not allow the slope Load Type within Speaker due to nonconvergence. Then, the model in Table 3.2 should be the one with the maximal random effects structure justified in both the L1 and the L2 data. Luckily, in addition to the fact that the slope did not change the statistical conclusion, the finally selected random structure (without any slopes) was what was recommended by multimodel inference and likelihood ratio

tests. Holding the fixed factors as Load Type, Load Condition, and their interaction, we compared (using model.sel function) models with one of the possible combinations of the three random intercepts. As summarized in Table 3.6[4], including all the three of Speaker, Sentence, and Relative Order was suggested to be the best with an extremely high weight of .997. Likelihood ratio tests supported the same: adding Speaker to Sentence and Relative Order improved the predictability of the model, $\chi^2(1) = 19.90$, $p < .001$; so did adding Sentence to the other two, $\chi^2(1) = 19.47$, $p < .001$; so did adding Relative Order to the other two, $\chi^2(1) = 13.85$, $p < .001$. Likewise, the same best random-factor set was selected as the best when the intercepts-only models were compared with models with (a) slope(s). The analysis compared a total of 71 models: 1 with no random factor, 7 with one or more random intercepts (only), and 63 (= number of all possible combinations of the 6 random slopes for the predictors within speakers and within sentences) with one or more random slopes that was/were added to the three random intercepts. Some candidates are listed in the model selection results in Table 3.7. The best model was again the one with the three random intercepts without any slopes, with a weight of .606. The model in Table 3.5 was ranked second-best with a weight of .113. Some other models

[4] We controlled for the random effects as suggested in multimodel inference and likelihood ratio tests. However, even without one or all random factors, the same statistical suggestion for the fixed effects is maintained in all of the conducted regression analyses. For example, in a total of four analyses with the same fixed factors and varying random effects, i.e., in a regression model with no random factors and in models with one of the three random intercepts, Load Condition was always the only significant predictor of speech error, $b = -.85 \sim -.75$, $se(b) = .21$, $p < .001$; Load Type was not significant, $b = .19 \sim .21$, $se(b) = .18 \sim .19$, $p = .262 \sim .294 > .05$; there was no significant interaction, $b = .17 \sim .18$, $se(b) = .28 \sim .29$, $p = .531 \sim .552 > .05$.

including a slope were ranked high, but convergence was not reached anyway. All models with four or more slopes had a weight of .000, considered poor models. The ranking of the set of slopes was exactly the same as in Table 3.7, when the intercepts were excluded, i.e., when only the slopes were compared. Lastly, a likelihood ratio test indicated Load Type within Speaker did not improve the model predictability, $\chi^2(3) = 2.73$, $p = .435 > .05$.

Table 3.6 Model selection table that analyzes effects of random intercepts in predicting L1 speech error

Model selection table that analyzes effects of random intercepts in predicting L1 speech error: from a multimodel inference analysis (using model.sel function) comparing models including possible combinations of random factors of Speaker (Sp), Sentence (Sn), and Relative Order (Ro). Predictor variables, number of parameters (K), log-likelihood (Log$L$), Akaike's Information Criterion with small-sample bias correction (AIC_c), AIC_c differences (Δ_i), and Akaike weights (w_i) for a set of candidate models for predicting speech error presence.

Random intercepts	K	Log$L$	AIC_c	Δ_i	w_i
Sp, Sn, Ro	7	-604.64	1223.4	0.00	.997
Sp, Sn	6	-611.56	1235.2	11.82	.003
Sp, Ro	6	-614.38	1240.8	17.45	.000
Sn, Ro	6	-614.59	1241.3	17.87	.000
Sp	5	-620.86	1251.8	28.40	.000
Sn	5	-621.57	1253.2	29.82	.000
Ro	5	-623.55	1257.2	33.78	.000
(None)	4	-629.93	1267.9	44.52	.000

Note. The models are sorted from best (top) to worst (bottom), ranked by AIC_c. Fixed factors are Load Type, Load Condition, and Load Type * Load Condition across all candidate models.

Table 3.7 Model selection table for random effects to predict L1 speech error

Model selection table that recommends random effects structure including random slopes: from a multimodel inference analysis (using model.sel function) comparing models with handpicked combinations of random factors for Speaker (Sp), Sentence (Sn), and Relative Order (Ro) and random slopes for Load Type (LT), Load Condition (LC), and Relative Order within Speaker (|Sp) and within Sentence (|Sn). Predictor variables, number of parameters (*K*), log-likelihood (Log*L*), Akaike's Information Criterion with small-sample bias correction (AIC_c), AIC_c differences (Δ_i), and Akaike weights (w_i) for a set of candidate models for predicting L1 speech error.

Random factors	*K*	Log*L*	AIC_c	Δ_i	w_i
Sp, Sn, Ro	7	-604.64	1223.4	0.00	.606
Sp, Sn, Ro, LT|Sp	10	-603.28	1226.7	3.36	.113
Sp, Sn, Ro, LT|Sn	10	-603.97	1228.1	4.76	.056
Sp, Sn, Ro, Ro|Sp	10	-604.03	1228.3	4.88	.053
Sp, Sn, Ro, Ro|Sn	10	-604.43	1229.0	5.66	.036
Sp, Sn, Ro, LC|Sp	10	-604.56	1229.3	5.93	.031
Sp, Sn, Ro, LC|Sn	10	-604.64	1229.5	6.09	.029
Sp, Sn, Ro, LT|Sp, LT|Sn	13	-602.28	1230.9	7.50	.014
Sp, Sn, Ro, LT|Sp, Ro|Sp	13	-602.85	1232.0	8.64	.008
Sp, Sn, Ro, LT|Sp, Ro|Sn	13	-603.10	1232.5	9.13	.006
Sp, Sn, Ro, LT|Sn, Ro|Sp	13	-603.17	1232.7	9.28	.006
Sp, Sn, Ro, LT|Sp, LC|Sp	13	-603.23	1232.8	9.40	.005
Sp, Sn, Ro, LT|Sp, LC|Sn	13	-603.27	1232.9	9.48	.005
Sp, Sn, Ro, Ro|Sp, Ro|Sn	13	-603.71	1233.7	10.36	.003
Sp, Sn, Ro, LC|Sp, LT|Sn	13	-603.75	1233.8	10.44	.003
Sp, Sn, Ro, LT|Sn, Ro|Sn	13	-603.82	1233.9	10.57	.003
Sp, Sn, Ro, LT|Sn, LC|Sn	13	-603.86	1234.0	10.65	.003
Sp, Sn, Ro, LC|Sp, Ro|Sp	13	-603.92	1234.1	10.77	.003
Sp, Sn, Ro, LC|Sn, Ro|Sp	13	-604.03	1234.4	11.01	.002
Sp, Sn, Ro, LC|Sp, Ro|Sn	13	-604.31	1234.9	11.55	.002
Sp, Sn, Ro, LC|Sn, Ro|Sn	13	-604.43	1235.2	11.79	.002
Sp, Sn	6	-611.56	1235.2	11.82	.002
(19 rows are omitted for simplicity.)					
Sp, Sn, Ro, LC|Sp, LC|Sn, Ro|Sp	16	-603.91	1240.3	16.93	.000
Sp, Ro	6	-614.38	1240.8	17.45	.000
Sp, Sn, Ro, LC|Sp, LC|Sn, Ro|Sn	16	-604.31	1241.1	17.71	.000
Sn, Ro	6	-614.59	1241.3	17.87	.000
Sp, Sn, Ro, LT|Sp, LT|Sn, Ro|Sp, Ro|Sn	19	-601.81	1242.3	18.90	.000
(19 rows are omitted for simplicity.)					
Sp, Sn, Ro, LC|Sp, LT|Sn, LC|Sn, Ro|Sp, Ro|Sn	22	-603.04	1251.0	27.60	.000
Sp	5	-620.86	1251.8	28.40	.000
Sn	5	-621.57	1253.2	29.82	.000
Sp,Sn,Ro,LT|Sp,LC|Sp,LT|Sn,LC|Sn,Ro|Sp,Ro|Sn	25	-601.79	1254.7	31.36	.000
Ro	5	-623.55	1257.2	33.78	.000
(None)	4	-629.93	1267.9	44.52	.000

Note. The models are sorted from best (top) to worst (bottom), ranked by AIC_c. Fixed factors are Load Type, Load Condition, and Load Type * Load Condition across all candidate models. 36 rows in the middle are omitted for simplicity.

3.3 Phonetic Encoding

The second part of the analysis is on the phonetic encoding. Based on the consensus that prosody is the primary measure of speech planning units (Shattuck-Hufnagel & Turk, 1996, p. 194),

3.3.1 Methods

3.3.1.1 Speech materials

All matched sentential pairs that were fluently and correctly produced by the same speaker were selected for acoustic measurement and analyses in order to test for effects of working memory load and load type on production. A total of 680 sentences were identified. There were 320 such sentences produced in the verbal load condition and 360 in the spatial load condition. These represented an average of 35.8 (*SD* = 9.31) sentences per speaker (or 17.9 sentential pairs). See Appendix VI for per speaker details.

3.3.1.2 Acoustic measurements

The 680 matched sentences were acoustically segmented first into spoken chunks, then into vocalic intervals. Boundaries for spoken chunks that start or end with a consonant were marked at the start or the end of visual speech energy in the waveform and in the spectrogram. Boundaries for spoken chunks that start or end with a vowel were based on periodic waveforms and clear band energy of the second

formant in the spectrogram. (Unfilled) pause boundaries were segmented based on the absence of visual speech energy in the waveform and in the spectrogram. There was no minimum pause duration in environments between sonorants, fricatives, and vowels. In the context of a [+stop] __ [+stop] sequence (e.g., *bird __ pecked*), pauses were given if the silent interval (preceding a burst) exceeded 150 milliseconds[5]. Vocalic intervals were identified as in Low, Grabe, and Nolan (2000). These excluded nasals, glides, and prevocalic /l/s, but included postvocalic alveolar approximants (/ɹ/ and /ɬ/ as in *dark* and *hole*).

The following nine measures were obtained based on automatically extracted durations and F0s across the segmented sentences: (i) sentence duration, in seconds including unfilled pauses; (ii) articulation rate, calculated as syllables per second excluding pauses; (iii) duration variability, in word durations using the normalized pairwise variability index (nPVI; Low et al., 2000); (iv) duration range, minimum word duration subtracted from maximum of the same sentence; (v) pitch initial, sentence initial pitch in Hertz represented by the median F0 of vocalic interval in the first content word (or the second word in the current stimuli) per sentence; (vi) pitch mean, in Hertz of medians across the vocalic intervals for each sentence; (vii) pitch variability, calculated using the nPVI formula; (viii) pitch range, minimum median

[5] This was based on the following findings: the perceptual threshold of a silent interval between two stop consonants (i.e., p-p, as in *top pick* as opposed to *topic*) ranged from 140 to 320 milliseconds depending on the speech rates of 8 to 2 syllables per second (Pickett & Decker, 1960); the silent interval between two stops required longer duration than that preceding one stop consonant (Pickett & Decker, 1960); the silent intervals preceding the burst of a stop consonant ranged from an average of 68.6 to 117.6 milliseconds, from /d/ to /p/ among all 6 English stops (Suen & Beddoes, 1974).

pitch of vocalic intervals subtracted from maximum of the same sentence; (ix) articulation clarity, in vowel space area[6] (in F1 x F2) as a correlate of articulation clarity.

The last measure articulation clarity was calculated as follows. The first three formant-values at the midpoint of content-word monothongal vowels were first extracted automatically from 85 words and 9 vowels (/i, ɪ, ɛ, æ, ɑ, ʌ, ɔ, ʊ, u/) per speaker. The default maximum 5th formant setting was 5,000 Hz for males and 5,500 Hz for females. Tracking errors were hand-corrected and formants remeasured. The frequency values in Hz were then converted to Bark using the formula $Z = [26.81/(1 + 1960/f)] - 0.53$, where any Z values lower than 2 Bark were corrected using $Z' = Z + 0.15 (2 - Z)$ (as proposed in Traunmüller, 1990). Formant values were then normalized for vocal tract length using a modified Bark Difference Metric (Syrdal & Gopal, 1986). Bark-transformed F0 was subtracted from bark-transformed F1 (i.e.,

[6] Area was selected over Euclidean distance, the latter of which is a method used in literature as a correlate of articulatory clarity, e.g., Redford, 2014. The distance may be useful to measure how close a produced vowel is to the target centroid of normal and fluent productions of the same vowel by the same speaker or by a group of (native) speakers: if an instantiated vowel is closer to the target, it should mean the production is more representative of the vowel quality than one that is farther from the target. However, after the purposes of the current research were taken into careful consideration and after initial results from the Euclidean distances were examined, we have determined that the distances do not reflect how clearly the vowels were produced in that the distance measure is a scalar but not a vector, i.e., it lacks direction information. For instance, when vowels from a speaker are far apart from a point/centroid, it could mean they are all merged to one direction becoming less discernible or that they are all towards the extreme vertices of a space. The magnitude information seemed limited. Instead, also taking directional information into account, we tried to plot the vowel space of each person by the four working memory blocks and compared the spaces in the four blocks. While we acknowledge that the currently selected measure, area, is not "the" correlate of articulation clarity, we for now report the results of our current analyses in this dissertation.

Z1-Z0) to model vowel height[7]; bark-transformed F2 was subtracted from bark-transformed F3 to model tongue advancement (Z3-Z2). Mean height and tongue advancement for each vowel type produced by each speaker gave us per-vowel centroids for each speaker, blocked by load type and load condition. The same vowel types were used across all blocks per speaker. Then, MATLAB (version R2019b 9.7.0.1319299) calculated the area of each F1-F2 vowel space by creating a boundary around the points on a coordinate plane using a shrink factor of 0.1 and getting the area of the polygon. We assumed that a larger vowel space area corresponded to relatively clearer vowel articulation in that the per-vowel centroids that constituted the vertices of a polygon were farther apart and potentially more distinctive.

3.3.1.3 Statistical analyses

A factorial multivariate analysis of covariance (MANCOVA, with manova syntax in SPSS Mac version 25.0.0.1) evaluated working memory load differences

[7] We chose Z1-Z0 over Z3-Z1 to model vowel height. We found some other researchers like Thomas and Kendall (2007) substituted Z3-Z1 for Z1-Z0 (i.e., what Syrdal & Gopal, 1986, originally proposed); they stated that F0 might be problematic in that it is influenced by factors such as intonation, tone, consonantal context, creakiness, and especially aging. However, we concluded we could circumvent the issue because the current study compared within-speaker and within-item productions, i.e., the same vowel pairs in the same word produced by the same speaker. Moreover, it was reported that the distance between F1 and F0 was decisive for vowel openness, whereas higher formants (like F3) contributed a marginal role (Traunmüller, 1981). Z1-Z0 explained 84% of the variance in the perception of vowel openness, and the higher variance of 90.4% was explained when the weight changed to Z1-0.6Z0 (Traunmüller, Ericksson, & Menard, 2003). Al-Tamimi (2017) also used Z1-Z0 to model vowel height, summarizing literature (p. 7) that Z1-Z0, using bark difference formant frequencies, correlates well with openness dimension, with more close or high vowels located lower than 3 Bark and more open vowels around 5 Bark.

associated with various acoustic measures of produced speech, after the composite scores were adjusted by different sentences. Covariate (CV) was the 32 sentences (Sentence, coded as 1 to 32) as the various sentences were not the theoretical interest of the current study but to elicit various prosodic hierarchical structures. Independent variables (IVs[8]) were Load Type (verbal, spatial) and Load Condition (load, no-load). CV and IVs were manipulated, qualitative, within-subjects effects. Pillai's Trace (*P*) was adopted as test statistics for multivariate significance, following Finch and French's (2013) findings that *P* statistics was most robust to assumption violations, especially nonnormality, as is the case for our current data, in terms of maintaining normal Type I error and optimal power[9].

Dependent variables (DVs) were the nine production measures as detailed in the immediately preceding section 3.3.1.2: (i) sentence duration (DrSn), a higher number denotes a longer duration in speaking the sentence; (ii) articulation rate (AtRt), a higher number denotes a faster speech; (iii) duration variability (DrVr), a higher number denotes more variation in word durations; (iv) duration range (DrRn), a higher number denotes a larger range in word durations; (v) pitch initial (F0In), a higher number denotes a higher pitch at the beginning of a sentence; (vi) pitch mean

[8] We did not test the effects of relative-clause type and location in this section, in that it should be well expected that sentence types influence acoustic patterns. As expected, sentence structure characteristics influenced overall speech patterns: RC Type, $P = 0.11$, $F(9, 667) = 9.07$, $p < .001$; RC Location, $P = 0.05$, $F(9, 667) = 3.82$, $p < .001$; interaction, $P = 0.04$, $F(9, 667) = 2.93$, $p = .002 < .05$. Instead of examining the effects of the syntactic structures, we controlled for linguistic factors by including sentences as covariate.

[9] For all the conducted analyses, Wilks' Lambda, which is a more commonly used test statistics, gave identical test results (i.e., identical *F*s and *p*s). General guidelines were followed as in Tabachnick and Fidell (2007, Ch. 7).

(F0Mn), a higher number denotes a higher mean pitch across a sentence; (vii) pitch variability (F0Vr), a higher number denotes more variation in pitch; (viii) pitch range (F0Rn), a higher number denotes a larger pitch range; (ix) articulation clarity (AtCl), a higher number denotes a clearer articulation. These nine measures were defined as scores that constitute the speech production composite. All durations were in seconds and all pitches in Hertz. All DVs were quantitative continuous.

Importance of controlling for the CV sentences for current analyses was supported in significant multivariate and univariate effects of the CV on the DVs. In a preliminary correlation check, sentences were significantly correlated with four out of the nine dependent measures (i.e., sentence duration, duration variability, duration range, and pitch variability). Pillai's Trace test of multivariate significance indicated that this CV was statistically significantly associated with the multivariate composite, $P = 1.56$, $F(270, 5841) = 4.52$, $p < .001$. Univariate tests of the relationship between the CV and each of the individual DV scores that made up the composite, conducted using a corrected alpha based on the Bonferroni's procedure $p = .05/9 = .006$, revealed that the CV was significantly related to four of the nine measures: sentence duration, articulation rate, duration variability, and duration range, $p < .001$.

Multivariate analyses of the relationship between the CV and the IVs indicated that the data were robust to the assumption of homogeneity of the regression plane and thus MANCOVA analysis can be appropriately interpreted, as the interaction between the sentences (CV) and working memory load (IVs) was not

statistically significant on the composite DV: CV and Load Type, $P = 0.02$, $F(9, 668) = 1.57$, $p = .120$; CV and Load Condition, $P = 0.01$, $F(9, 668) = 0.69$, $p = .719$.

To complement MANCOVA results, two relevant follow-ups were conducted: discriminant function analysis and univariate analyses of variance (ANOVAs). Discriminant function analysis investigated which individual DVs contributed the most in driving significant effects of the IV group differences on the composite. This involved examining (i) associated standardized discriminant function coefficients (*SDFC*) used to weight the multivariate composite and (ii) structure coefficients (r) indicating the correlation of the DVs with one another and with the composite. A high absolute value of *SDFC* indicates the changes in a univariate measure primarily impact the changes in the composite score. Even when a measure with a low absolute value of *SDFC* has some effects on the composite, the net gain in the composite should be relatively small (such as variance explained), in that the composite is primarily controlled by measures with high absolute *SDFCs*. A large structure coefficient tells us the univariate measure is highly correlated with the changes in the entire composite scores. If a significant effect on the multivariate composite is supported, the effect should have been drawn from that measure. If a measure has a low structure coefficient, the extent to which the measure affects the entire composite property should be minimal. Then, when a measure has a high *SDFC* but has a low r, the extent to which the measure accounts for the entire composite should be minimal. Next, univariate ANOVAs assessed effects on each of the nine measures comprising the multivariate composite. Alpha was adjusted for

multiple group mean tests on each subtest (i.e., .05/9 = .006, following Bonferroni's procedure) to maintain the probability of Type I error at .05.

3.3.2 Results

Effects of working memory load (Load Type and Load Condition) were examined on weighted multivariate composite of speech production. Descriptive statistics are given in Table 3.8 for each dependent variable by working memory load. Across all working memory groups, we observed the following raw values: Sentence duration was $M = 3.34$, $SD = 0.50$; articulation rate $M = 3.71$, $SD = 0.47$; duration variability $M = 69.91$, SD = 10.27; duration range $M = 0.42$, $SD = 0.09$; pitch initial $M = 149.06$, $SD = 52.99$; pitch mean $M = 133.87$, $SD = 45.47$; pitch variability $M = 5.65$, $SD = 2.69$; pitch range $M = 36.76$, $SD = 21.80$; articulation clarity $M = 7.12$, $SD = 3.01$.

Table 3.8 Means and standard deviations for nine acoustic measures of L1 speech production

Means (*M*) and standard deviations (*SD*) for the nine acoustic measures of speech produced by L1 speakers of English, by working memory Load Type and Load Condition

WM			DrSn	AtRt	DrVr	DrRn	F0In	F0Mn	F0Vr	F0Rn	AtCl
V	L	*M*	3.24	3.81	70.17	0.41	150.58	135.04	5.62	36.93	6.75
		SD	0.47	0.50	10.87	0.09	53.30	46.39	2.49	20.26	2.85
	N	*M*	3.39	3.68	69.96	0.43	149.67	134.97	5.83	37.70	7.39
		SD	0.58	0.49	11.13	0.10	53.54	46.39	2.48	21.38	3.06
S	L	*M*	3.30	3.76	69.59	0.42	148.49	133.07	5.51	35.95	6.97
		SD	0.45	0.45	9.43	0.09	52.41	44.74	2.62	22.14	3.03
	N	*M*	3.43	3.61	69.95	0.43	147.76	132.67	5.65	36.59	7.37
		SD	0.49	0.43	9.81	0.10	53.21	44.86	3.11	23.23	3.08

Note. WM = working memory; V = verbal; S = spatial; L = load; N = no-load; DrSn = sentence duration, in seconds; AtRt = articulation rate, in syllables per second; DrVr = duration variability, in nPVI of word durations; DrRn = duration range, in seconds; F0In = pitch initial, in Hertz; F0Mn = pitch mean, in Hertz; F0Vr = pitch variability, in nPVI of median F0s from rhymes; F0Rn = pitch range, in Hertz; AtCl = articulation clarity, in vowel space area). Verbal load $N = 160$ sentences; verbal no-load $N = 160$; spatial load $N = 180$; spatial no-load $N = 180$.

Speakers' overall speech production patterns, as weighted multivariate combination of the nine dependent measures, varied systematically with Load Condition, $P = 0.05$, $F(9, 667) = 3.67$, $p < .001$, $\eta^2 = .05$, but not with Load Type, $P = 0.01$, $F(9, 667) = 0.83$, $p = .587$. The effect of working memory condition accounted for 4.72% of the production composite variance ($\Lambda = .95$, 100% of the single root identified). The interaction was not statistically significant, $P = 0.00$, $F(9, 667) = 0.37$, $p = .949$. Figure 3.3 shows the native English speakers' produced speech, as weighted composite, changed when speaking in the middle of engaging in a working memory task, compared to when the same sentences were produced without an additional task. However, the type of task did not differently impact speech acoustics. Given the insignificant type or interaction effects, speaking different sentences did not change the speech patterns. Note that the same sentences were used for both load and no-load groups within each load type. The difference between the two blocks of no-load condition, i.e., verbal no-load and spatial no-load, was different sentences. In the two blocks/levels of the load condition, i.e., verbal load and spatial load, the changed direction or the magnitude was not different between the two types of load.

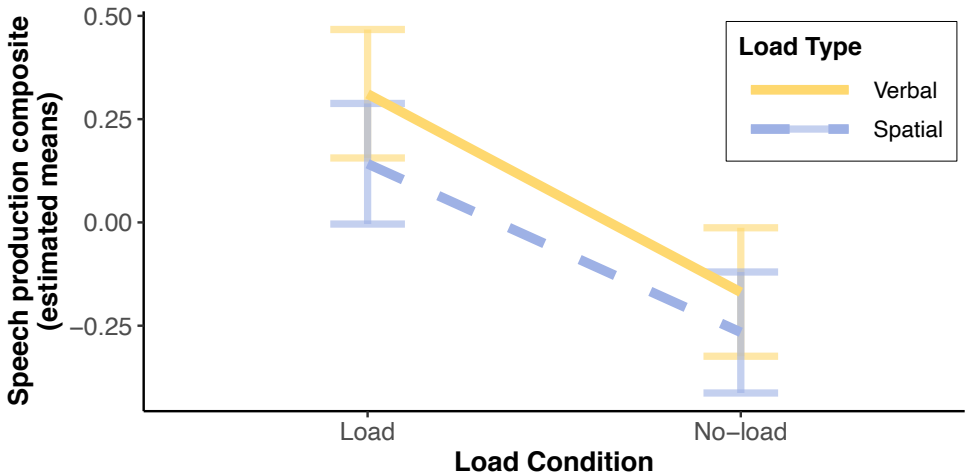


Figure 3.3 L1 speech production multivariate composite influenced by working memory load, irrespective of load type

L1 speech production composite (in estimated marginal means of multivariate composite) was influenced by working memory load condition but not by load type. Significant effect of load condition (load vs. no-load) was supported, but the difference in load type (verbal vs. spatial) was not statistically significant. The native English speakers' speech acoustics, analyzed as weighted composite of nine acoustic measures of the produced speech, was disrupted by the additional working memory task during speaking, but the verbal task did not differently influence speech production from the spatial type of task. Error bars indicate 95% CI.

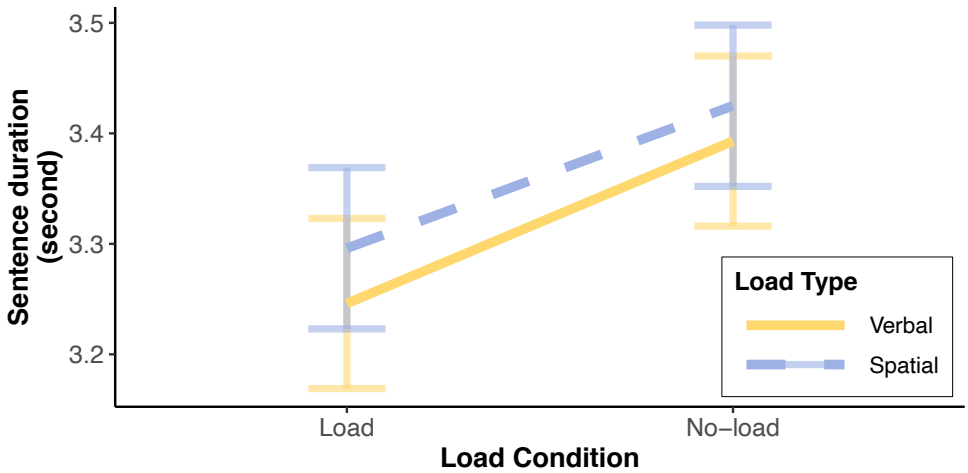
An analysis of the composition and correlation of the composite with respect to the individual measures implied that the significant multivariate effects were associated mainly with articulation rate and possibly with articulation clarity. Associated standardized discriminant function coefficients (*SDFC*) revealed that pitch initial (*SDFC* = 1.42) and articulation clarity (*SDFC* = -1.00) contributed primarily to the composite. Articulation rate (*SDFC* = 0.64) and pitch range (*SDFC* = -0.56) contributed to medium extent. These were most important in forming the function that distinguished the working memory load groups. Inspection of the structure coefficients (*r*) indicated that sentence duration (*r* = -.62) and articulation rate (*r* = .70), duration range (*r* = -.38), and articulation clarity (*r* = -.39) were largely

correlated with the composite. Given the results from these two coefficient criteria, pitch initial (r = .03) and pitch range (r = -.07), despite their high *SDFC*s, were hardly correlated with the composite and thus the significant effects of the working memory load should not be on these measures[10]; sentence duration (*SDFC* = -0.19) and duration range (*SDFC* = 0.16), despite the high correlations, had minimal unique contribution and the variance explained for these measures were already explained by other measures.

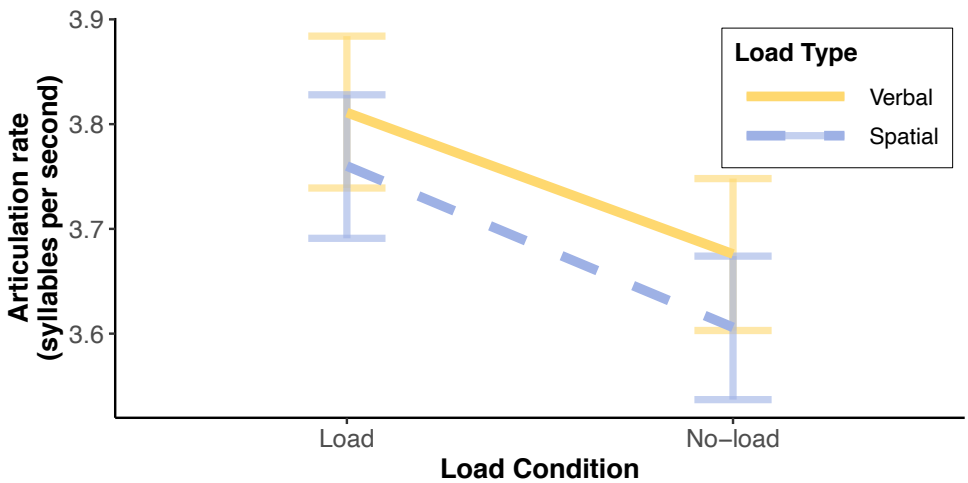
Univariate analysis of variance (ANOVAs) on each of the nine dependent measures that made up the multivariate composite indicated that sentence duration and articulation rate were the only measures influenced by Load Condition: on sentence duration, $F(1, 675) = 13.04$, $p < .001 < .006 = .05/9$, $\eta^2 = .02$; on articulation rate, $F(1, 675) = 16.25$, $p < .001$, $\eta^2 = .02$. As in the overall analysis, this effect did not interact with Load Type. Load Type was not significant in any of the univariate analyses. Neither were any of the interactions. Figure 3.4 graphs the faster speech under load. Speakers completed speaking a sentence in an average of about 4.1% (or 0.14 second) faster under load (M = 3.27 seconds, SD = 0.46) compared to when they read a sentence in the control no-load conditions (M = 3.41 seconds, SD = 0.54). The same was true when we excluded the pauses. Taking only the speech portions, they produced an average of 3.8% (or 0.14) more syllables or words per second

[10] Univariate ANOVA supported the same. Pitch initial and pitch range, despite the large SDFC, were not consistently affected by the working memory load factors. The effect of Load Condition was not significant on pitch initial, $F(1, 675) = 0.04$, p = 843, or on pitch range, $F(1, 675) = 0.18$, p = .670. The effect of Load Type was also insignificant, $F(1, 675) = 0.22$, p = .639, and $F(1, 675) = 0.33$, p = 567. Interactions were insignificant ($p > .05$).

under load (M = 3.78 syllables per second, SD = 0.47) compared to speech in the no-load conditions (M = 3.64 syllables per second, SD = 0.46).



(a) Sentence duration



(b) Articulation rate

Figure 3.4 Faster L1 speech under working memory load during speaking

Faster L1 speech under working memory load during speaking: (a) shorter sentence duration (in seconds, including pauses) and (b) higher articulation rate (in syllables per second, excluding pauses), in L1 speech during a working memory task compared to speech in the no-load condition. The type of load, whether it was verbal or spatial, did not change the condition effect on speed. Error bars indicate 95% Confidence Interval.

3.4 Discussion

The results indicate significant effects of working memory load on speech errors and on speech rate. Native speakers of English produced more speech errors under load than when not under load. They were also "in a hurry" to complete speaking the sentences in order to shift their attention to the other load task. Yet, the effects are not sensitive to the type of working memory taxed. The results are thus inconsistent with the phonological-phonetic encoding model, and instead suggest a generalized effect of load on speech production processes, consistent both with attentional models of working memory (Cowan, 1999; Engle, 2002) and with retrieval models of speech production (e.g., Browman & Goldstein, 1992; Redford, 2015).

The results suggest that speech is not phonologically-phonetically encoded in verbal working memory. If we assume phonological-phonetic encoding during speech production, as proposed in the staged models (Fromkin, 1971; Levelt et al., 1999), metrical and segmental properties are selected and sequenced during phonological encoding, and articulatory routines are selected and manipulated in the articulatory buffer during phonetic encoding. These linguistic materials should occupy the processing capacity in the verbal working memory component (Baddeley & Hitch, 1974). Then, if we overload the capacity-limited (Baddeley, 2000; Gathercole & Baddeley, 1993) verbal working memory component by having speakers engage in one more verbal task at the same time, as in our verbal load condition (i.e., speech + verbal task), the performance in (either or both) verbal tasks should be impaired. By contrast, speech should not be impaired when we ask

speakers to do a concurrent but non-verbal task, as in our spatial load condition (i.e., speech + spatial task). This is because the capacity of working memory is dissociated by the content domain (Baddeley & Hitch, 1974; Seigneuric et al., 2000; Kellogg et al., 2007). The two different types of information are separately processed in two different working memory subcomponents, i.e., speech in the verbal component and a spatial task in the spatial component.

Contrary to predictions, however, the current data seem as counterevidence of the hypothesis of phonological-phonetic encoding. There was no load-type effect. Their verbal, speaking behavior was disrupted the same way by the spatial task or another verbal task. The type of load did not affect either the number of sentences with disfluencies and errors or the prosody and segmental articulation clarity of the fluently-correctly produced sentences. The only statistically significant factor was the load condition. Compared to when only speaking sentences, when the speakers had to perform one more task in addition to speaking, they simply made more errors and spoke faster.

The generality of this (condition) effect across load types undercuts the idea that verbal working memory is relevant to speech production, as already suggested in Gathercole and Baddeley's (1993) comprehensive review on that relationship. Instead, the results remind us of the general phenomena observed also in non-language domains. We are less successful in doing multiple things at the same time than when we do one thing at a time. People are better, for example, at riding a bike only than riding a bike and thinking about a shopping list at the same time. In trying

to do multiple jobs well, people struggle and tend to make more errors in some, if not all, tasks. Air-traffic controllers are also more prone to error when multitasking, independent of general intelligence (Colom, Martínez-Molina, Shih, & Santacreu, 2010).

The current findings, i.e., a generalized effect of cognitive load on speech production processes, are consistent with attentional models of working memory (Cowan, 1999; Engle, 2002) and with retrieval models of speech production (e.g., Browman & Goldstein, 1992; Redford, 2015). In attentional models of working memory, impaired speech is simply an instance of a wide range of real-world cognitive tasks (Engle, 2002). According to Engle (2002), humans have a domain-free limitation in their ability to control attention. Working memory system focuses its attention on some information and holds it activated so it can be readily retrievable. However, its attentional capacity is limited, varied to individual people. Thus, when the system focuses on multiple different tasks, it can result in impaired or inappropriate performance due to interference. Likewise, the present load effect also seems to be attentional in nature. Whatever the additional task is, e.g., whether spatial or verbal in addition to speaking, the added cognitive load onto the domain-free executive attention working memory system makes it difficult to control attention due to its limited capacity. Our speakers may have cycled their attention between speaking and the span task. Cycling back and forth this way would have reduced the overall amount of attention (and time) the speakers devoted to production.

Based on speech production literature, the results support retrieval models (e.g., Browman & Goldstein, 1992; Redford, 2015). Instead of the to-be-articulated sounds being planned online, speakers simply retrieve from the memorized articulatory routines that they have already stored in the long-term memory system through over-practice (Fowler et al., 1980; Browman & Goldstein, 1992). Retrieval thus does not engage in ongoing working memory processes or articulatory buffer, and is resultantly free of cognitive capacity limitation. This can account for an insignificant type effect. None of the speech material is actively processed in verbal working memory online. Phonological and phonetic information is simply retrieved as word-sized chunks from the remembered action templates (Browman & Goldstein, 1992). As the chunks are already encoded of the articulatory movements for the associated lexical concepts, the phonological-phonetic encoding stage can be bypassed during production. This makes possible the argument that speech production is an automatic, remembered action (Redford, 2015).

The current argument against phonological-phonetic encoding may be further tested in follow-up studies. For example, the current sentence structure with only single-syllable words was not appropriate to examine part of the processes in phonological encoding: metrical planning and syllabification. Including multi-syllable words with various stress patterns may partially resolve this issue.

With regards to load task(s), ensuring a moderate task difficulty would be critical to properly manipulate, especially verbal, working memory overload (Turner & Engle, 1989). It is important to confirm whether the combined amount of load

from the speech production and the additional verbal task overloads a speaker's working memory capacity (Engle, 2002; Swets et al., 2014). It was interpreted that the present study's distractor task was difficult enough to resource and overload a speaker's working memory capacity. The number of correct-fluent sentences decreased from 78.9% in no-load conditions to 67.6% in the verbal load condition. The speech was impaired during multitasking, showing significant condition effect. Still, one might argue that this study's distractor task was too easy for some speakers so that their working memory was not moderately resourced, thus possibly not overloading working memory capacity for some speakers enough to affect speech. Some solutions can be: allowing a shorter time; having a more difficult verbal task; measuring individual working memory capacity. The last method may also account for our unexplained variance across speakers (Appendix V). As there implied, individuals with low working memory capacity may perform differently from individuals with high working memory capacity (Colom et al., 2010; Swets et al., 2014).

The argument against phonological-phonetic encoding can further be supported if we find the same in non-native English speech or with speakers of other languages. Those speakers may reveal different types of interference in the same contexts. Different aspects of speech can be disrupted or in the opposite direction. Such results can help conclude whether phonological-phonetic encoding is not evidenced regardless of language backgrounds.

CHAPTER IV.

WORKING MEMORY IN L2 SPEECH PRODUCTION[11]

4.1 Introduction

This chapter examines the effects of working memory on speech produced by nonnative speakers of English. Korean learners of English as a foreign language spoke aloud 32 English sentences. They produced (randomly assigned) 16 sentences during a verbal working memory load task and the same sentences again in a control no-load condition. They produced the other set of 16 sentences during a spatial task and again in another block of no-load condition. The same experimental methods were used for this L2 study as used for the L1 study, unless otherwise noted in this chapter, in methods section 4.2.1 or 4.3.1. Modifications are specified to highlight the differences between the L1 and the L2 study. See more details in section 3.2.1 and 3.3.1.

As with the L1 study, the L2 study is presented in two parts. Section 4.2, to address phonological encoding, presents an analysis of sentences produced with speech disfluency or error. Section 4.3, to address phonetic encoding, presents an analysis of prosodic variations in 9 acoustic measures as observed in the correctly produced sentences. The same for the L1 study, the hypothesis of phonological

[11] Part of the work in this chapter is to appear in the peer-reviewed proceedings from the Pronunciation in Second Language Learning and Teaching: Lee, O., & Ahn, H. (2020). Faster and less clear L2 speech with more errors during a verbal working memory task but not during a spatial task. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Annual Pronunciation in Second Language Learning and Teaching Conference* (pp. 141-153). Flagstaff, AZ: Northern Arizona University. ISSN 2380-9566.

encoding predicts an increase in segmental errors under verbal working memory load relative to the control condition, but not under spatial working memory load; the phonetic encoding hypothesis predicts effects of verbal working memory load but not of spatial load on acoustic patterns of produced speech.

4.2 Phonological Encoding

4.2.1 Methods

4.2.1.1 Participants

Participants were twenty (10 males and 10 females) college-aged adult native speakers of Korean who speak English as a foreign language. They were recruited at Seoul National University, Seoul, Korea. All reported normal hearing and speaking and no history of speech-language therapy. All were compensated with a coffeehouse gift-card for their time.

4.2.1.2 Speech materials

The same 32 sentences were used as in the L1 study. To repeat, the sentences were organized by dependent relative clause type (subject-extracted, object-extracted) and its location relative to the matrix clause (middle, end). See Appendix I for complete sentence list.

4.2.1.3 Working memory manipulation

The same as in the L1 study, the task was designed to manipulate the type and the condition of working memory load during speaking. Working memory load type manipulated the type of load as either verbal or spatial. A load task was either loaded on top of speaking or not loaded. During the two types of load conditions, a speaker was required to remember a sequence of 4 letters (verbal load condition) or 4 spatial locations (spatial load condition) while speaking aloud a given English sentence. During the two blocks of no-load conditions, no additional task was given while a speaker was speaking aloud the English sentence displayed on the computer monitor.

To ensure each English consonant be pronounced as single syllable in Korean, F, H, J, and X were replaced with D, G, P, and T in the verbal load condition. The resultant consonants were B, D, G, L, M, P, Q, R, and T.

In the distractor tasks, the participants managed 57.2% (M = 9.15, SD = 2.64) correct responses in the verbal load condition, 90.6% (M = 14.5, SD = 1.91) in the spatial load condition, and 96.9% (M = 15.5, SD = 0.82) in the no-load condition. All scores were well above the chance performance of 12.5%.

4.2.1.4 Elicitation procedure

The same as in the L1 study, data was collected in a within-subjects design, with two fixed factors of Load Type (verbal, spatial) and Load Condition (load, no-load). Participants first read through the 32 sentences to understand the meaning and the structure of the sentences. They were given as much time as they needed to go back

and forth the sentences and ask questions while and/or after reading the sentences. They were informed, with a sample slide shown, that the sentence line at the center of the slide was to be shown the same in the main production phase. In addition to the information given for L1 speakers, Korean L2 learners of English were provided with translation in Korean; each relative clause was marked within parentheses. See Appendix III for illustration.

After this familiarization stage, they proceeded to the main part of the experiment, which was blocked by Load Type and Load Condition. Each participant produced a sentence in either verbal or spatial load condition, and once again in the associated no-load condition. Accordingly, each participant produced a total of 64 sentences (= 32 sentences * 2 productions). All participants were assisted by the same researcher throughout the entire experiment.

Participants' speech was digitally recorded for later acoustic analyses using a Tascam DR-100MKIII. The wireless recorder, with a built-in microphone, was put on a desk to maintain a relatively consistent distance from the speaker. The entire experiment took no more than 60 minutes to complete.

4.2.1.5 Speech error determination

A total of 1,280 sentential productions were collected from 20 speakers * 32 sentences * 2 productions. 34 excluded due to non-linguistic disruption during production (e.g., coughing or chair-dragging), the remaining 1,246 sentential productions were measured and coded for analysis.

The same as the error coding for the L1 data, each sentential production was transcribed and coded as correct or incorrect. Incorrect productions included one or more disfluencies and/or speech errors. Disfluencies were (i) filled pause, (ii) false start, and (iii) repeat. Speech errors were either phonemic or lexical and categorized as (i) addition, (ii) omission, (iii) substitution, (iv) exchange, and (v) shift.

In addition to unfilled pauses and lengthening without a semantic change being coded as correct, as we did for the L1 data, we coded a part of non-native phonemic productions as correct for the L2 data: as long as a consonant or vowel production may be categorized as a foreign pronunciation of a target phonemic inventory, we considered it correct for L2. For example, if a word *fat* was pronounced as /pæt/, we coded it as correct. We interpreted it as a case that the EFL speaker aimed to pronounce *fat* but failed to produce a native English phoneme. The same applied to other consonant and vowel productions such as /lid/ for *read*, /lʌbd/ for *loved*, /big/ for *big*, /mɛd/ for *mad*, /dog/ for *dog*, and /kuk/ for *cook*. Note, however, that we coded a production as incorrect if the produced consonant or vowel may not fall into the target phoneme. For example, if a word *ball* was pronounced as /bel/, we coded it as incorrect. All such observed instances were /kʊk/ for *took*, /maʊf/ for *mouse*, /kit/ and /kjut/ for *cut*, /belt/ for *built*, /mɛn/ for *mean*, /pɛkd/ for *picked*, and /caʊtʃ/ for *caught*.

4.2.1.6 Statistical analysis

Exactly the same statistical procedure was followed for the L2 analysis as used for

the L1. Generalized linear mixed-effects logistic regression models with logit-link function were used in combination with multimodel inference and likelihood ratio tests. For significant interactions, simple effects coefficients were computed with adjusted alpha to maintain the probability of Type I error at .05. The dependent variable was presence of speech error in a sentence. The fixed factors included Load Type (verbal, spatial), Load Condition (load, no-load), RC Type (subject-extracted, object-extracted), and RC Location (middle, end). Random intercepts were for speakers (Speaker), sentences (Sentence), and relative order (RelOrder). Within-unit random slopes justified by the design were tested of Load Type within speakers, Load Condition within speakers, Load Type within sentences, Load Condition within sentences, RelOrder within speakers, and RelOrder within sentences. To secure objectivity in selecting the fixed and the random factors, the best model and fit was justified by the data by means of multimodel inference and likelihood ratio tests. To explain variance explained (R^2) separately for the fixed factors and for the entire model, we report two types of R^2: marginal R^2 for the variance explained by the fixed factors and conditional R^2 for the variance explained by the fixed and the random factors.

4.2.2 Results

32.0% (N = 399) out of 1,246 sentences were produced with at least one speech disfluency and/or error. There was a total number of 84 disfluencies (an average of 0.30 per incorrectly produced sentence) and of 460 errors (1.15 per sentence). The

most frequent disfluency was repeat (N = 37, 44.0%), followed by false start (N = 26, 31.0%), and then by filled pause (N = 21, 25.0%). Lexical errors were less frequent than phoneme errors (N = 131, 28.5% vs. N = 329, 71.5%). For lexical errors, substitution (N = 69, 52.7%) was most frequent, followed by addition and omission (for each N = 29, 22.1%), and then by exchange and shift (for each N = 2, 1.5%). For phonemic errors, addition (N = 159, 48.3%) was most frequent, followed by substitution (N = 114, 34.7%) and omission (N = 56, 17.0%). None were observed of exchange or shift. Table 4.1 summarizes the distribution of sentences produced with speech errors by Load Type, Load Condition, RC Type, and RC Location. Numbers of errors are indicated in the numerator; the total numbers of sentences are indicated in the denominator. The total proportions of sentences produced with one or more speech disfluency or error were as follows: verbal load M = 47.1%, 147/312; verbal no-load M = 21.4%, 67/313; spatial load M = 36.5%, 113/310; spatial no-load M = 23.2%, 72/311; subject middle M = 34.6%, 110/318; subject end M = 25.3%, 81/320; object middle M = 32.6%, 94/288; object end M = 35.6%, 114/320.

Table 4.1 Proportion of sentences with speech error in L2 speech

Proportion and [cumulative number] of sentences with speech error (as numerators) out of total sentences (as denominators) in L2 speech: by Load Type (verbal, spatial), Load Condition (load, no-load), Relative-Clause Type (subject-extracted, object-extracted), and Relative-Clause Location with respect to matrix clause (middle, end).

		Verbal		Spatial		
Relative clause		Load	No-load	Load	No-load	Total
Subject-extracted	Middle	53.8% [43/80]	22.5% [18/80]	39.7% [31/78]	22.5% [18/80]	34.6% [110/318]
	End	40.0% [32/80]	16.3% [13/80]	30.0% [24/80]	15.0% [12/80]	25.3% [81/320]
Object-extracted	Middle	52.8% [38/72]	16.4% [12/73]	33.3% [24/72]	28.2% [20/71]	32.6% [94/288]
	End	42.5% [34/80]	30.0% [24/80]	42.5% [34/80]	27.5% [22/80]	35.6% [114/320]

A generalized linear mixed-effects logistic regression was calculated to investigate whether working memory load type and condition can predict the presence of speech disfluency or error in a sentence (henceforth speech error). The model included: speech error as dependent variable; Load Type, Load Condition, and Load Type * Load Condition as fixed factors; speakers, sentences, and relative order as random factors. See at the latter part of this section 4.2.2 for how this model was selected as the best. The total number of observations was 1,246. The results are given in Table 4.2.

The results indicated both working memory factors were significant predictors of speech error. Speech error was statistically significantly predicted by Load Type, $b = .48$, $se(b) = .17$, $p = .005 < .05$, and by Load Condition, $b = -.68$, $se(b) = .24$, $p = .005 < .05$. There was also a significant interaction between the two working memory factors, $b = -.58$, $se(b) = .26$, $p = .025 < .05$. The AIC was 1475.64.

The BIC was 1511.54. The log-likelihood was -730.82. The working memory factors explained about 5.1% of the variance in speech error. Variance explained by the fixed factors (marginal R^2_{GLMM}, $R^2_{GLMM(m)}$) was 5.1% (observation-level variance σ^2_ε) or 6.8% (distribution-specific theoretical variance σ^2_d). Variance explained by the entire model (conditional R^2_{GLMM}, $R^2_{GLMM(c)}$), including both the fixed and the random effects, was 12.5% (σ^2_ε) or 16.6% (σ^2_d).

Table 4.2 Generalized linear mixed effects logistic regression results for working memory effects on L2 speech error

Generalized linear mixed effects logistic regression results for working memory effects on L2 speech error (N = 1,246).

Fixed effects:	***b* [95% CI]**	***se(b)***	***z***	***p***
(Intercept)	-.60 [-1.05, -.27]	.19	-3.21	.001 **
Load Type	.48 [.14, .81]	.17	2.80	.005 **
Load Condition	-.68 [-1.06, -.24]	.24	-2.84	.005 **
Load Type * Load Condition	-.58 [-1.10, -.07]	.26	-2.24	.025 *
Random effects:	***σ²***	***SD***		
Speaker (Intercept)	.07	.26		
Sentence (Intercept)	.32	.56		
RelOrder (Intercept)	.00	.03		
AIC	1475.64			
BIC	1511.54			
Log*L*	-730.82			
$R^2_{GLMM(m)}$_σ^2_ε; $R^2_{GLMM(m)}$_σ^2_d	.05; .07			
$R^2_{GLMM(c)}$_σ^2_ε; $R^2_{GLMM(c)}$_σ^2_d	.12; .17			

Note. Estimates of fixed-effects regression coefficient (*b*), 95% confidence intervals (CI) calculated by the confint function, standard error of regression coefficient (*se*(*b*)), *z*-value (*z*), *p*-value (*p*), variance (σ^2), standard deviation (*SD*), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (Log*L*), variance explained by the fixed effects (marginal R^2, $R^2_{GLMM(m)}$) obtained using observation-level variance (σ^2_ε) and distribution-specific theoretical variance (σ^2_d), and variance explained by the entire model (conditional R^2, $R^2_{GLMM(c)}$) using σ^2_ε and σ^2_d. Significance: *** < .001, ** < .01, * < .05.

More sentences were produced incorrectly during the load conditions (41.8% = 260/622) than the no-load conditions (22.3% = 139/624), and more during the

verbal type (34.2% = 214/625) than the spatial type (29.8% = 185/621). However, given significant interaction, these main effects cannot always be true. Figure 4.1 graphs the interaction between load type and load condition in predicting the probability of sentences produced with speech error. The shaded area around the regression line represents the standard error of regression coefficient, which was .05 for VL, .03 for VN, .04 for SL, and .03 for SN.

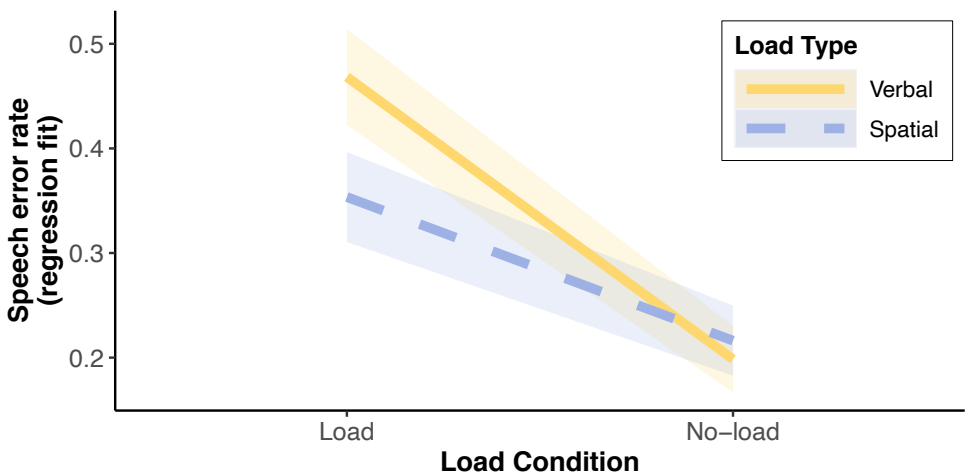


Figure 4.1 L2 speech error rate predicted by working memory load type and condition
L2 speech error rate (in means of regression model fit) predicted by working memory load type and load condition. The colored area around the regression lines denotes standard error of regression coefficient.

To probe into the interaction, simple effects coefficients were computed. First, it was supported that engaging in a working memory task during speaking, regardless of whether the task was verbal or spatial, statistically significantly increased the L2 speakers' speech error rate. When the sentences were produced during a working memory task, for either of the load types, more sentences were

produced with error compared to when there was no additional task other than speaking. Verbal load (VL, 47.1% = 147/312) increased the error rate by 25.7% compared to when the same sentences were produced in the associated no-load condition, i.e., verbal no-load (VN, 21.4% = 67/313), b = -1.26, $se(b)$ = .19, p < .001. Spatial load (SL, 36.5% = 113/310) also induced 13.3% higher error rate compared to when the same sentences were produced in the associated no-load condition, i.e., spatial no-load (SN, 23.2% = 72/311), b = -.69, $se(b)$ = .19, p < .001. Second, comparing the types of load, verbal type of load (VL, 47.1%) led the error rate by 10.6% higher than did spatial type of load (SL, 36.5%), b = .48, $se(b)$ = .17, p = .005 < .025 = .05/2. However, as implied in the significant interaction effect, the two no-load conditions, i.e., VN (21.4%) and SN (23.2%), which differed only in the sentences, were not statistically different in speech error rate, b = -.11, $se(b)$ = .20, p = .573 > .025.

The aforereported model in Table 4.2 was determined to be the best to explain the variation in the response variable speech error. It was the model with the fixed factors of the best goodness-of-fit and the random factors in maximal random effects structure justified by the design and the data. It added a random intercept 'relative order' to the best model (as in Table 4.3) justified by multiple multimodel inference tests and likelihood ratio tests. The more complex model (Table 4.2) was ranked second best to the simpler model (Table 4.3) in multimodel inference tests (e.g., see support in Table 4.6). The simpler model was preferred by definition in that the two models were statistically equivalent in the predictability ($\chi^2(1)$ = .00, p

= .967 > .05) in a likelihood ratio test. We nevertheless selected the more complex model so that we have the maximal random effects structure (see more discussion in section 3.2.1.6) that at the same time achieved convergence and nonsingular fit; we can also compare the L2 data directly with the L1 using the same regression model. Moreover, as will be discussed below, the one in Table 4.2 was the most complex model (or maximal random effects structure) justified by the data. Furthermore, regardless of which model we chose, the statistical conclusion was the same[12]. The fixed effects remained significant also in the simpler model, Load Type, b = .48, $se(b)$ = .17, p = .005 < .05, Load Condition, b = -.69, $se(b)$ = .19, p < .001, their interaction, b = -.58, $se(b)$ = .26, p = .025 < .05. The variance explained by the model is maintained 12.5%. The working memory factors explained 5.2% (marginal R^2_{GLMM}, σ^2_ε, or .6.9%, σ^2_d) of the variance in speech error; the entire model including both the fixed and the random factors explained 12.5% (conditional R^2_{GLMM}, σ^2_ε, or 16.6%, σ^2_d). The simple effects tests with either random structures gave exactly the same estimate values. Marginal changes due to the added random intercept included a higher p-value for Load Condition from .000 to .005 and the variance explained by the fixed factors from 5.2% to 5.1%.

[12] As in the L2 results, the L1 analyses also gave the same statistical suggestion as to the fixed effects regardless of whether the random-factor set included Relative Order in addition to Speaker and Sentence (as in Table 3.2) or not. Excluding Relative Order from the model shown in Table 3.2 did not impact the statistical significance for the fixed factors: only Load Condition was a significant predictor of speech error, b = -.81, $se(b)$ = .21, p < .001. Neither Load Type nor the interaction impacted speech error, b = .20, $se(b)$ = .19, p = .298 > .05, and b = .18, $se(b)$ = .29, p = .537 > .05. The AIC was 1235.13. The BIC was 1265.46. The log-likelihood was -611.56. Variance explained by the entire model was 10.2% (σ^2_ε) or 15.7% (σ^2_d). Variance explained by the fixed factors was 2.6% (σ^2_ε) or 4.0% (σ^2_d).

Table 4.3 Regression results of the best model to predict L2 speech error

Regression results of the best model to predict L2 speech error, justified by multimodel inference and likelihood ratio tests (N = 1,246).

Fixed effects:	***b* [95% CI]**	***se(b)***	***z***	***p***
(Intercept)	-.60 [-.94, -.27]	.17	-3.55	.000 ***
Load Type	.48 [.14, .81]	.17	2.80	.005 **
Load Condition	-.69 [-1.06, -.33]	.19	-3.73	.000 ***
Load Type * Load Condition	-.58 [-1.10, -.07]	.26	-2.24	.025 *
Random effects:	**σ^2**	***SD***		
Speaker (Intercept)	.07	.26		
Sentence (Intercept)	.32	.56		
AIC	1473.65			
BIC	1504.41			
Log*L*	-730.82			
$R^2_{GLMM(m)}$_$\sigma^2\varepsilon$; $R^2_{GLMM(m)}$_σ^2_d	.05; .07			
$R^2_{GLMM(c)}$_$\sigma^2\varepsilon$; $R^2_{GLMM(c)}$_σ^2_d	.13; .17			

Note. Estimates of fixed-effects regression coefficient (*b*), 95% confidence intervals (CI) calculated by the confint function, standard error of regression coefficient (*se*(*b*)), *z*-value (*z*), *p*-value (*p*), variance (σ^2), standard deviation (*SD*), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (Log*L*), variance explained by the fixed effects (marginal R^2, $R^2_{GLMM(m)}$) obtained using observation-level variance ($\sigma^2\varepsilon$) and distribution-specific theoretical variance (σ^2_d), and variance explained by the entire model (conditional R^2, $R^2_{GLMM(c)}$) using $\sigma^2\varepsilon$ and σ^2_d. Significance: *** < .001, ** < .01, * < .05.

Below describes how multimodel inference tests and likelihood ratio tests supported the model in Table 4.3 as the best regression model and fit to predict the presence of speech error in a sentence for the L2 data. First, regarding fixed factors, multimodel inference recommended as the best model the one with two fixed factors of Load Type and Load Condition and their interaction. The model selection table is given in Table 4.4. Using dredge and get.models function in the package MuMIn, the best model was selected automatically comparing all possible combinations of fixed factors of Load Type, Load Condition, RC Type, and RC Location, including all two-way, three-way, and four-way interactions. The best model had a weight of .092. It was 1.64 (= .092/.056) time more likely to be the best to explain the speech-error variation in the data than the second-best model (Table 4.5a). It was

also 3.83 (= .092/.024) times more likely so than the model including the linguistic factors of RC Type, RC Location, and their interaction in addition to the cognitive factors (Table 4.5b). Random factors across the candidate models had little impact on the ranked order of fixed factors in model selection. For example, the one having speakers, sentences, and relative order as random intercepts, as in Table 4.4a, and the one dropping the last factor, as in Table 4.4b, gave the same selection with exactly the same weights (at least for the first 20+ rows). All other multimodel inference analyses, with different numbers and combinations of fixed and random factors, selected the model in Table 4.3 as the best (with weight ratios almost equivalent to aforementioned 1.64 and 3.83: respectively, 1.64 ~ 1.65, e.g., from .279/.170 or .268/.162; 3.88 ~ 3.94, e.g., from .279/.072 or .280/.071).

Likewise, likelihood ratio tests via anova function in R justified removing RC Type, $\chi^2(1) = 1.03$, $p = .311 > .05$, and RC Location, $\chi^2(1) = 0.55$, $p = .459 > .05$, for the best fit of the data. In fact, neither RC Type nor RC Location statistically significantly predicted speech error (Table 4.5), $b = -.54$, $se(b) = .32$, $p = .088 > .05$ and $b = -.12$, $se(b) = .32$, $p = .701 > .05$. They did not interact with each other, $b = .61$, $se(b) = .45$, $p = .180 > .05$, or with the cognitive predictors, $p > .05$ in all other conducted regression analyses. They did not improve predictability of the model, $p > .05$ in likelihood ratio tests. The same was true in models with various combinations of fixed and random factors, e.g., a model with speakers and sentences (without relative order) did not change any of the estimates for the fixed effects giving the same chi-, coefficient-, and *p*-values.

Table 4.4 Model selection tables that analyze relative predictiveness of predictors for L2 speech error

Model selection tables that analyze relative predictiveness of predictors (or fixed factors) for L2 speech error: from an automated multimodel inference analysis (using dredge function) comparing all possible combinations of the predictors of Load Type (LT), Load Condition (LC), RC Type (RT), and RC Location (RL). Predictor variables, number of parameters (K), log-likelihood (Log$L$), Akaike's Information Criterion with small-sample bias correction (AIC_c), AIC_c differences (Δ_i), and Akaike weights (w_i) for a complete set of candidate models for predicting speech error presence.

(a) Speaker, Sentence, and RelOrder as random factors

Predictor variables (predictors)	K	Log$L$	AIC_c	Δ_i	w_i
LC, LT, LC*LT	7	-730.82	1475.7	.00	.092
LC, LT, RT, LC*LT	8	-730.31	1476.7	1.00	.056
LC, LT, RT, LC*LT, LC*RT	9	-729.47	1477.1	1.35	.047
LC, LT, RL, LC*LT	8	-730.55	1477.2	1.48	.044
LC, LT, RT, LC*LT, LT*RT	9	-729.94	1478.0	2.29	.029
LC, LT, RL, RT, LC*LT	9	-730.01	1478.2	2.43	.027
LC, LT, RL, LC*LT, LC*RL	9	-730.05	1478.2	2.51	.026
LC, LT, RT, LC*LT, LC*RT, LT*RT	10	-729.13	1478.4	2.70	.024
LC, LT, RL, RT, LC*LT, RL*RT	10	-729.14	1478.5	2.73	.024

(b) Speaker and Sentence as random factors

Predictor variables (predictors)	K	Log$L$	AIC_c	Δ_i	w_i
LC, LT, LC*LT	6	-730.82	1473.7	.00	.092
LC, LT, RT, LC*LT	7	-730.31	1474.7	1.00	.056
LC, LT, RT, LC*LT, LC*RT	8	-729.47	1475.1	1.34	.047
LC, LT, RL, LC*LT	7	-730.55	1475.2	1.47	.044
LC, LT, RT, LC*LT, LT*RT	8	-729.94	1476.0	2.28	.029
LC, LT, RL, RT, LC*LT	8	-730.01	1476.1	2.42	.027
LC, LT, RL, LC*LT, LC*RL	8	-730.05	1476.2	2.51	.026
LC, LT, RT, LC*LT, LC*RT, LT*RT	9	-729.13	1476.4	2.69	.024
LC, LT, RL, RT, LC*LT, RL*RT	9	-729.14	1476.4	2.72	.024

Note. The models are sorted from best (top) to worst (bottom), ranked by AIC_c. The comparison included all possible combinations of two-way, three-way, and four-way interactions (*). Random factors are Speaker, Sentence, and RelOrder in analysis (a) and Speaker and Sentence in (b) across all candidate models. From each of (a) and (b), the bottom 32,758 rows are omitted.

Table 4.5 L2 speech error predicted by different regression models

L2 speech error predicted by different regression models: (a) the second-best model in a multimodel inference and (b) a model with both the linguistic and the cognitive predictors.

(a) the second-best model (N = 1,246)

Fixed effects:	***b* [95% CI]**	***se(b)***	***z***	***p***
(Intercept)	-.48 [-.96, -.07]	.22	-2.18	.029 *
Load Type	.48 [.14, .81]	.17	2.80	.005 **
Load Condition	-.68 [-1.06, -.25]	.24	-2.83	.005 **
Load Type * Load Condition	-.59 [-1.10, -.07]	.26	-2.24	.025 *
RC Type	-.24 [-.72, .23]	.24	-1.02	.308
Random effects:	**σ^2**	***SD***		
Speaker (Intercept)	.07	.26		
Sentence (Intercept)	.30	.55		
RelOrder (Intercept)	.00	.02		
AIC	1476.62			
BIC	1517.64			
Log*L*	-730.31			
$R^2_{GLMM(m)}$_σ^2_ε; $R^2_{GLMM(m)}$_σ^2_d	.05; .07			
$R^2_{GLMM(c)}$_σ^2_ε; $R^2_{GLMM(c)}$_σ^2_d	.13; .17			

(b) a model with both linguistic and cognitive fixed factors (N = 1,246)

Fixed effects:	***b* [95% CI]**	***se(b)***	***z***	***p***
(Intercept)	-.42 [-.98, .09]	.27	-1.58	.114
Load Type	.48 [.14, .81]	.17	2.79	.005 **
Load Condition	-.69 [-1.06, -.25]	.24	-2.83	.005 **
Load Type * Load Condition	-.58 [-1.10, -.07]	.26	-2.24	.025 *
RC Type	-.54 [-1.20, .10]	.32	-1.70	.088
RC Location	-.12 [-.77, .53]	.32	-.38	.701
RC Type * RC Location	.61 [-.31,1.53]	.45	1.34	.180
Random effects:	**σ^2**	***SD***		
Speaker (Intercept)	.07	.26		
Sentence (Intercept)	.27	.52		
RelOrder (Intercept)	.00	.02		
AIC	1478.28			
BIC	1529.56			
Log*L*	-729.14			
$R^2_{GLMM(m)}$_σ^2_ε; $R^2_{GLMM(m)}$_σ^2_d	.06; .08			
$R^2_{GLMM(c)}$_σ^2_ε; $R^2_{GLMM(c)}$_σ^2_d	.13; .17			

Note. Estimates of fixed-effects regression coefficient (*b*), 95% confidence intervals (CI) calculated by the confint function, standard error of regression coefficient (*se*(*b*)), *z*-value (*z*), *p*-value (*p*), variance (σ^2), standard deviation (*SD*), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), log-likelihood (Log*L*), variance explained by the fixed effects (marginal R^2, $R^2_{GLMM(m)}$) obtained using observation-level variance (σ^2_ε) and distribution-specific theoretical variance (σ^2_d), and variance explained by the entire model (conditional R^2, $R^2_{GLMM(c)}$) using σ^2_ε and σ^2_d. Significance: *** < .001, ** < .01, * < .05.

Second, regarding random factors, including two random factors of Speaker

and Sentence was supported to be the best for the current data. Multiple multimodel inference analyses with different sets of fixed factors and random slopes were conducted to select the best model with respect to random factors out of Speaker, Sentence, and Relative Order. Results always suggested including Speaker and Sentence together be the best. For instance, in a multimodel inference analysis using the recommended fixed factors of Load Type, Load Condition, and the interaction between the two, having Speaker and Sentence together as random intercepts was the best having a weight .540, as shown in Table 4.6. It was 2.74 (= .540/.197) times better than the second-best model, which included all three, and 2.80 (= .540/.193) times better than having Sentence alone. Including Speaker only, Relative Order only, or having no random intercepts had weights of .000. Excluding the no-intercept candidate from the comparison did not affect the selection, giving exactly the same ranks with exactly the same weights. Tests with different fixed factors and tests with random slopes recommended the same best random factors, and if applicable, with the same ordering and similar weight-ratios of the candidate models. For example, the analysis as in Table 4.7 selected the same best random factors, with a weight of .483; the best model was 2.74 (= .483/.176) times better than the second-best model, 2.79 (= .483/.173) times better than having Sentence alone, and at least 21.95 (= .483/.022) times better than additionally including any random slopes. In addition, likelihood ratio tests suggested that adding Relative Order to the model would not significantly impact the predictability of the regression model ($\chi^2(1) = 0.00$, $p = .967 > .05$). Across multiple analyses, adding Relative Order to Speaker and Sentence

resulted in merely increasing the *p*-values for Load Condition from .000 to .005, while maintaining the significance level and other estimates the same.

Table 4.6 Model selection table that analyzes effects of random intercepts in predicting L2 speech error

Model selection table that analyzes effects of random intercepts in predicting L2 speech error: from a multimodel inference analysis (using model.sel function) comparing models including possible combinations of random factors of Speaker (Sp), Sentence (Sn), and Relative Order (Ro). Predictor variables, number of parameters (*K*), log-likelihood (Log*L*), Akaike's Information Criterion with small-sample bias correction (AIC_c), AIC_c differences (Δ_i), and Akaike weights (w_i) for a set of candidate models for predicting speech error presence.

Random intercepts	*K*	Log*L*	AIC_c	Δ_i	w_i
Sp, Sn	6	-730.82	1473.7	.00	.540
Sp, Sn, Ro	7	-730.82	1475.7	2.02	.197
Sn	5	-732.86	1475.8	2.06	.193
Sn, Ro	6	-732.86	1477.8	4.08	.070
Sp	5	-748.45	1506.9	33.24	.000
(None)	4	-749.91	1507.9	34.14	.000
Sp, Ro	6	-748.45	1509.0	35.26	.000
Ro	5	-749.91	1509.9	36.16	.000

Note. The models are sorted from best (top) to worst (bottom), ranked by AIC_c. Fixed factors are Load Type, Load Condition, and Load Type * Load Condition across all candidate models.

Third, although we wished to retain within-unit random slopes justified by the design, as recommended by Barr et al. (2013, p. 263), we had to remove all the random slopes because convergence and nonsingular fit was not reached in any of the models with one or more random slopes, as is often the case as discussed in Barr et al., 2013, pp. 262-263. Moreover, according to multimodel inference results, none of the random slopes for the predictors of interest within speakers and/or within sentences were supported to improve the plausibility of the model fit. For example, models with one or multiple random slopes had low weights of almost .000 and up to .022 when the slopes were added to the model in Table 4.2 (i.e., to the three random intercepts) and were compared together with no-slope candidates in Table

4.6. The model selection result is in Table 4.7. The best candidate had no slope, with a weight .483. A model with any random slopes had a weight of .022 at best. Any model that included more than two slopes had a weight .000. The best model, which included Speaker and Sentence as random intercepts but without a slope, improved the model fit at least 21.95 (= .483/.022) times better than any model with 1 or more slopes. The model in Table 4.2 was at least 8.00 (= .176/.022) times better than any model with 1 or more slopes. The ordering and relation among the random intercepts and slopes remained the same in all other multimodel inference analyses; e.g., the same best model with a weight .408 and the best slope model with a weight of .052 when the slopes were added to Speaker and Sentence.

Likewise, likelihood ratio tests supported removing all the random slopes, in that models including any random slopes for the predictors within speakers and within sentences were not different from the one without them ($p > .05$). We ran more than 63 different models including one or more within-unit random slopes. See some examples in Table 4.7. All pairwise likelihood ratio tests suggested removing the slopes from the models. The lowest *p*-value found was .550 when the model with relative order within speakers and Load Type within speakers was compared to the model with Load Type within speakers, $\chi^2(3) = 2.11$, $p = .550 > .05$. Compared to the model in Table 4.2, the removal was justified of the parameters for within-unit random slopes: Load Type within Speaker, $\chi^2(3) = 0.00$, $p = 1.000 > .05$; Load Condition within Speaker, $\chi^2(3) = 1.78$, $p = .620 > .05$; Load Type within Sentence, $\chi^2(3) = 0.17$, $p = .983 > .05$; Load Condition within Sentence, $\chi^2(3) = 0.31$, $p = .958$

> .05; Relative Order within Speaker, $\chi^2(3) = 1.97$, $p = .578 > .05$; Relative order within Sentence, $\chi^2(3) = 1.44$, $p = .697 > .05$. By excluding the random slopes, we assume that the fixed effects are invariant across speakers and across sentences in the population (reported as advised in Barr et al., 2013).

Table 4.7 Model selection table that recommends random effects structure for L2 speech error

Model selection table that recommends random effects structure including random slopes: from a multimodel inference analysis (using model.sel function) comparing models with handpicked combinations of random factors for Speaker (Sp), Sentence (Sn), and Relative Order (Ro) and random slopes for Load Type (LT), Load Condition (LC), and Relative Order within Speaker (|Sp) and within Sentence (|Sn). Predictor variables, number of parameters (*K*), log-likelihood (Log*L*), Akaike's Information Criterion with small-sample bias correction (AIC_c), AIC_c differences (Δ_i), and Akaike weights (w_i) for a set of candidate models for predicting L2 speech error.

Random factors	*K*	Log*L*	AIC_c	Δ_i	w_i
Sp, Sn	6	-730.82	1473.7	.00	.483
Sp, Sn, Ro	7	-730.82	1475.7	2.02	.176
Sn	5	-732.86	1475.8	2.06	.173
Sn, Ro	6	-732.86	1477.8	4.08	.063
Sp, Sn, Ro, Ro|Sp	10	-729.84	1479.8	6.13	.022
Sp, Sn, Ro, LC|Sp	10	-729.93	1480.0	6.33	.020
Sp, Sn, Ro, Ro|Sn	10	-730.10	1480.4	6.67	.017
Sp, Sn, Ro, LC|Sn	10	-730.67	1481.5	7.80	.010
Sp, Sn, Ro, LT|Sn	10	-730.74	1481.7	7.94	.009
Sp, Sn, Ro, LT|Sp	10	-730.82	1481.8	8.10	.008
Sp, Sn, Ro, Ro|Sp, Ro|Sn	13	-729.10	1484.5	10.79	.002
Sp, Sn, Ro, LC|Sp, Ro|Sn	13	-729.31	1484.9	11.21	.002
Sp, Sn, Ro, LC|Sp, Ro|Sp	13	-729.38	1485.0	11.33	.002
Sp, Sn, Ro, LC|Sn, Ro|Sp	13	-729.65	1485.6	11.89	.001
Sp, Sn, Ro, LC|Sp, LC|Sn	13	-729.75	1485.8	12.07	.001
Sp, Sn, Ro, LT|Sn, Ro|Sp	13	-729.75	1485.8	12.08	.001
Sp, Sn, Ro, LT|Sp, Ro|Sp	13	-729.77	1485.8	12.11	.001
Sp, Sn, Ro, LC|Sn, Ro|Sn	13	-729.79	1485.9	12.16	.001
Sp, Sn, Ro, LC|Sp, LT|Sn	13	-729.85	1486.0	12.29	.001
Sp, Sn, Ro, LT|Sn, Ro|Sn	13	-729.93	1486.2	12.44	.001
(47 rows are omitted for simplicity)					
Sp,Sn,Ro,LT|Sp,LC|Sp,LT|Sn,LC|Sn,Ro|Sp,Ro|Sn	25	-728.25	1507.6	33.86	.000
(None)	4	-749.91	1507.9	34.14	.000
Sp, Ro	6	-748.45	1509.0	35.26	.000
Ro	5	-749.91	1509.9	36.16	.000

Note. The models are sorted from best (top) to worst (bottom), ranked by AIC_c. Fixed factors are Load Type (LT), Load Condition (LC), and Load Type * Load Condition across all candidate models. 47 rows in the middle are omitted for simplicity.

4.3 Phonetic Encoding

4.3.1 Methods

4.3.1.1 Speech materials

Of 847 correct sentential productions, all matched sentential pairs that were fluently and correctly produced by the same speaker were selected for acoustic measurement and analyses in order to test for effects of working memory load and load type on production. There were 280 such sentences produced in the verbal load condition and 320 in the spatial load condition. These represented an average of 30.0 (*SD* = 6.55) sentences per speaker (or 15.0 sentential pairs). See Appendix VI for per speaker details.

4.3.1.2 Acoustic measurements

The 600 matched sentences were segmented and annotated into spoken parts, words, and rhymes. The same nine acoustic measures as in the L1 study were obtained: sentence duration, articulation rate, duration variability, duration range, pitch initial, pitch mean, pitch variability, pitch range, articulation clarity. See section 3.3.1.2 for more details.

4.3.1.3 Statistical analysis

A factorial multivariate analysis of covariance (MANCOVA) evaluated the effects

of working memory load (Load Type and Load Condition, independent variables, IVs) on weighted multivariate composite of speech production (dependent variable, DV) after the composite was adjusted by the sentences (covariate, CV). The nine dependent acoustic measures, as described in 4.3.1.2, constitute the speech production composite. Pillai's Trace (*P*) was adopted for multivariate significance.

To investigate which individual DV measures contributed to induce significant (if there is) working memory (IV) effects, discriminant function analysis and simple effects tests were conducted. Discriminant function analysis (see 3.3.1.3 for more description) examined how individual DV measures constituted and were correlated with the speech production composite. Simple effects tests replaced univariate analysis of variance tests to examine interaction effect, which we will find in the results section (4.3.2) of this part of the study. For significant interactions, simple effects tests are recommended because the procedures maintain the essential structure (e.g., 2 x 2, not 4 groups) of the experimental design. The common alpha is obtained by dividing the alpha level (e.g., .05) by the number of simple effects (see more about simple effects tests in Pedhazur & Schmelkin, 1991, Ch. 20). For the current design, alpha was adjusted (i.e., .05/9 = .006) to maintain the probability of Type I error at .05.

Significant multivariate and univariate effects of the CV on the DVs supported the importance of controlling for the CV sentences to examine working memory load effects on speech production. The CV had significant influence on the multivariate composite ($P = 1.54$, $F(279, 5112) = 3.77$, $p < .001$) and on six out of

the nine univariate dependent measures (i.e., sentence duration, articulation rate, duration variability, duration range, pitch variability, and pitch range, each $p < .001 < .006 = .05/9$ following Bonferroni's procedure). A supplement correlation test indicated the CV was significantly correlated with three of the nine dependent measures (i.e., sentence duration, duration variability, and duration range).

The CV did not significantly interact with the IVs on the composite DV, which thus assumes the homogeneity of the regression plane for appropriate MANCOVA interpretation. The interaction between the CV sentences and Load Type was insignificant, $P = 0.01$, $F(9, 588) = 0.62$, $p = .779$. The interaction between the CV and Load Condition was also insignificant, $P = 0.02$, $F(9, 588) = 1.50$, $p = .145$.

4.3.2 Results

Prior to interpreting the working memory effects on L2 speech production, descriptive statistics were generated for each dependent variable by working memory load. Table 4.8 summarizes the results. For a total of 600 sentences, sentence duration had $M = 4.06$, $SD = 0.66$; articulation rate $M = 3.22$, $SD = 0.41$; duration variability $M = 67.35$, $SD = 11.00$; duration range $M = 0.47$, $SD = 0.11$; pitch initial $M = 193.13$, $SD = 67.32$; pitch mean $M = 163.59$, $SD = 53.84$; pitch variability $M = 10.31$, $SD = 3.82$; pitch range $M = 62.30$, $SD = 31.83$; articulation clarity $M = 9.15$, $SD = 2.54$.

Table 4.8 Means and standard deviations for nine acoustic measures of L2 speech production

Means (*M*) and standard deviations (*SD*) for the nine acoustic measures of speech produced by L2 speakers of English, by working memory Load Type and Load Condition

WM			DrSn	AtRt	DrVr	DrRn	F0In	F0Mn	F0Vr	F0Rn	AtCl
V	L	*M*	3.88	3.34	64.99	0.46	194.49	164.21	10.07	62.41	8.15
		SD	0.63	0.42	11.50	0.12	66.09	52.86	3.64	27.82	2.62
	N	*M*	4.07	3.15	68.48	0.48	194.28	164.68	10.76	63.42	9.83
		SD	0.60	0.41	11.49	0.10	67.90	54.86	3.80	32.40	2.33
S	L	*M*	4.12	3.21	66.98	0.47	192.34	162.71	9.92	60.64	9.07
		SD	0.73	0.41	10.73	0.11	66.64	53.54	3.76	31.47	2.62
	N	*M*	4.13	3.17	68.81	0.48	191.74	162.99	10.51	62.87	9.50
		SD	0.67	0.38	10.07	0.10	69.14	54.56	4.03	35.04	2.31

Note. WM = working memory; V = verbal; S = spatial; L = load; N = no-load; DrSn = sentence duration, in seconds; AtRt = articulation rate, in syllables per second; DrVr = duration variability, in nPVI of word durations; DrRn = duration range, in seconds; F0In = pitch initial, in Hertz; F0Mn = pitch mean, in Hertz; F0Vr = pitch variability, in nPVI of median F0s from rhymes; F0Rn = pitch range, in Hertz; AtCl = articulation clarity, in vowel space area). Verbal load $N = 140$ sentences; verbal no-load $N = 140$; spatial load $N = 160$; spatial no-load $N = 160$.

Korean EFL speakers' overall English speech production, as weighted combination of the nine acoustic measures, was statistically significantly affected by working memory Load Type, $P = 0.03$, $F(9, 587) = 2.14$, $p = .025 < .05$, $\eta^2 = .01$, and Load Condition, $P = 0.09$, $F(9, 587) = 6.53$, $p < .001$, $\eta^2 = .08$, with significant interaction, $P = 0.03$, $F(9, 587) = 2.25$, $p = .018 < .05$, $\eta^2 = .03$. The interacted effects accounted for 3.3% of the production composite variance ($\Lambda = .97$, 100% of the single root identified), Load Type 3.2% and Load Condition 9.1%. Figure 4.2 represents the significant interaction between the two working memory factors. In order to probe into the interaction effect, i.e., to identify which working memory group differences led to statistically significant effects on the overall speech production, multivariate simple effects tests were followed up.

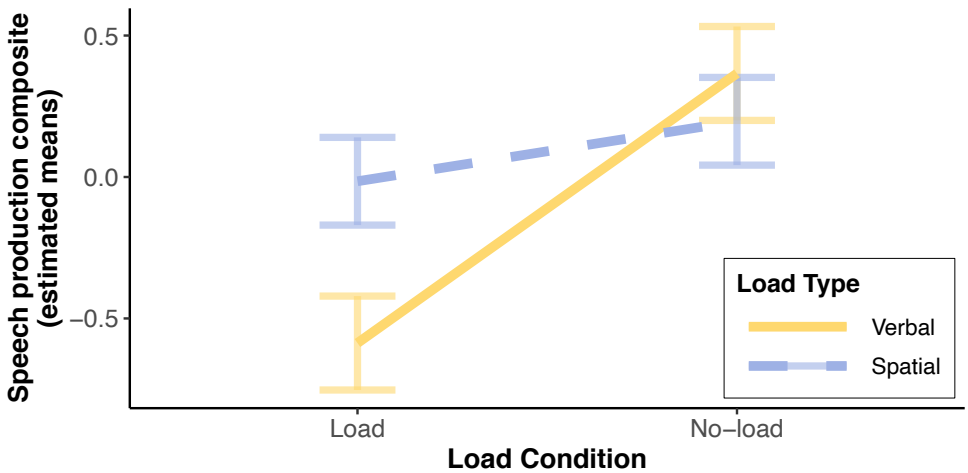


Figure 4.2 L2 speech production influenced only by verbal working memory load
L2 speech production composite (in estimated marginal means of multivariate composite) influenced only by verbal working memory load. Significant interaction is represented by the nonparallel lines. Error bars indicate 95% Confidence Interval.

Multivariate simple effects tests supported only the verbal task statistically significantly changed the acoustic composite patterns of the produced speech. Speech produced during a verbal task exhibited statistically significantly different composite scores from speech during a spatial task and from speech produced without additional processing load: verbal load was different from spatial load (VL ≠ SL), $F(1, 597) = 19.68$, $p < .001$; verbal load was different from verbal no-load (VL ≠ VN), $F(1, 597) = 62.86$, $p < .001$. By contrast, sentences spoken during a spatial task (spatial load, SL) were not acoustically different from the same sentences produced without additional load (spatial no-load, SN), SL = SN, $F(1, 597) = 3.23$, $p = .073 > .025 = .05/2$. Likewise, different sentences did not show systematic acoustic differences, VN = SN, $F(1, 597) = 1.16$, $p = .282$. Figure 4.3 plots verbal load had statistically significantly different speech patterns from spatial load, verbal

no-load, or spatial no-load.

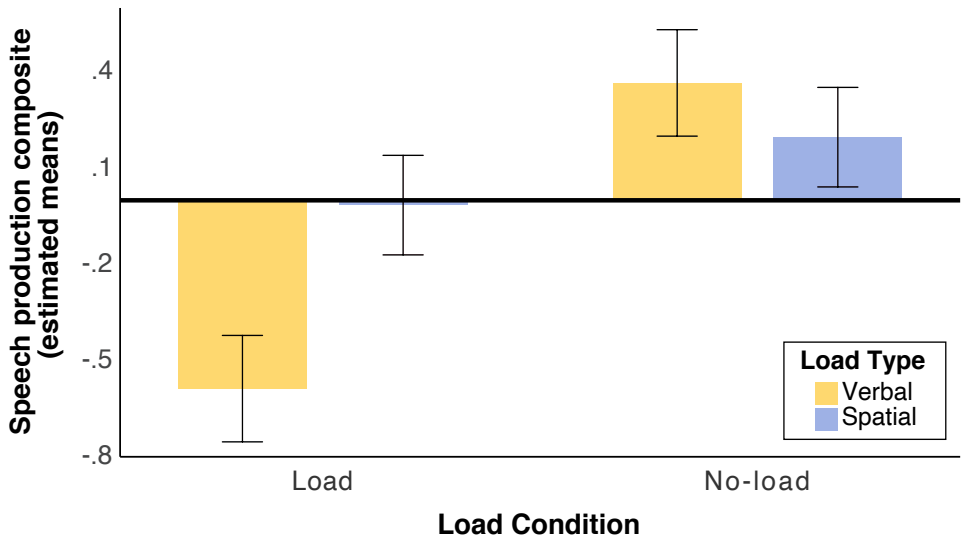
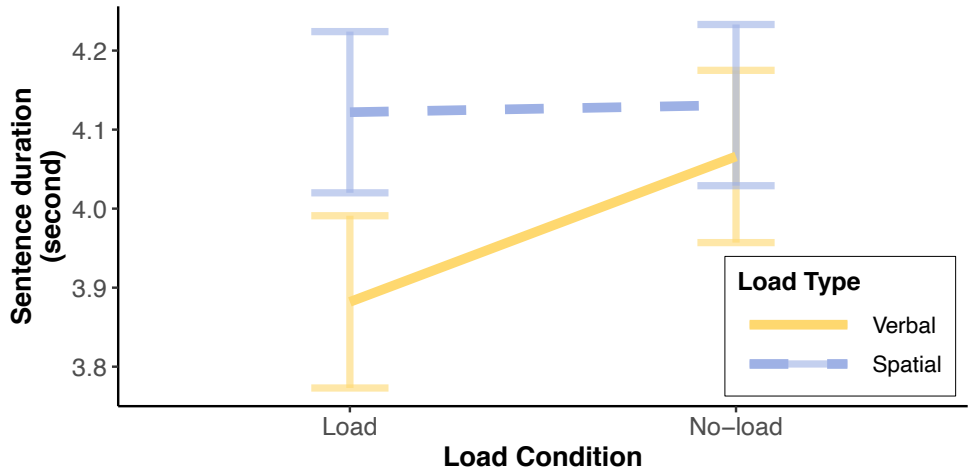


Figure 4.3 Significant effect of verbal load on L2 speech production
Only verbal load statistically significantly changed the overall L2 speech patterns, indicated in multivariate simple effects tests. Verbal load is different from the other three working memory groups. At alpha level of .025, VL ≠ SL; VL ≠ VN; SL = SN; VN = SN. VL = verbal load; VN = verbal no-load; SL = spatial load; SN = spatial no-load. Error bars indicate 95% Confidence Interval.

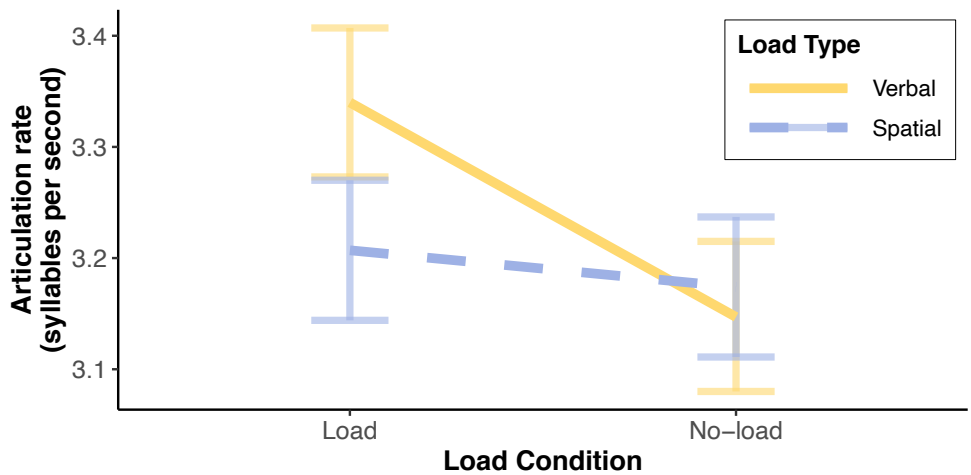
Now that we found a significant effect of verbal load, we proceeded to investigate which speech measures were impacted by the manipulated working memory load. A discriminant function analysis indicated that articulation rate (*SDFC* = -.94, *r* = -.53) and articulation clarity (*SDFC* =.79, *r* = .69) contributed primarily to distinguishing the working memory groups. Univariate simple effects tests indicated that sentence duration, articulation rate, duration variability, and articulation clarity were statistically significantly influenced by the verbal working memory load, but not by the spatial load. None of the pitch patterns were influenced

by working memory load. Sentence duration during a verbal task (VL) was different from that during a spatial task (SL), $F(1, 597) = 9.44$, $p = .002 < .025 = .05/2$, and from that when the same sentences were produced without additional load (VN), $F(1, 597) = 5.41$, $p = .020$. Articulation rate in VL was different from that in SL, $F(1, 597) = 7.11$, $p = .008$, and from that in VN, $F(1, 597) = 15.82$, $p < .001$. Duration variability in VL was different from that in VN, $F(1, 597) = 8.24$, $p = .004$. Articulation clarity in VL was different from that in SL, $F(1, 597) = 8.55$, $p = .004$, and from that in VN, $F(1, 597) = 32.28$, $p < .001$. All other differences were statistically insignificant. None of the individual measures were different between SL and SN, duration variability having the lowest *p*-value of .110; none between VN and SN, sentence duration having the lowest *p*-value of .355.

Speakers completed speaking a sentence about 5.8% (or 0.24 second) faster during a verbal working memory task ($M = 3.88$ seconds, $SD = 0.63$) than during a spatial task ($M = 4.12$, $SD = 0.73$) and 4.7% (or 0.19 second) faster than when without additional task ($M = 4.07$, $SD = 0.60$). They articulated 4.1% (or 0.13) more word (or syllable) per second during a verbal task ($M = 3.34$ syllables or words, $SD = 0.42$) than during a spatial task ($M = 3.21$, $SD = 0.41$) and 6.0% (or 0.19) more word (or syllable) per second than when without additional task other than speaking ($M = 3.15$, $SD = 0.41$). Figure 4.4 shows speech was faster during a verbal task than during a spatial or in the control no-load conditions.



(a) Sentence duration



(b) Articulation rate

Figure 4.4 Faster L2 speech under verbal working memory load during speaking
Faster L2 speech under working memory load during speaking: (a) shorter sentence duration (in seconds, including pauses) and (b) higher articulation rate (in syllables per second, excluding pauses), in L2 speech during a verbal working memory task. Speech was faster during a verbal task, but not during a spatial task. The speed during a spatial task was statistically the same as that during speaking without additional task. Error bars indicate 95% Confidence Interval.

L2 speakers spoke with less variable word durations during a verbal task (durational variability $M = 64.99$ in nPVI, $SD = 11.50$) compared to when speaking the same sentences without additional task ($M = 68.48$, $SD = 11.49$). Speech during a spatial task gave a higher but statistically insignificant mean variability difference (SL $M = 66.98$, $SD = 10.73$; SN $M = 68.81$, $SD = 10.07$). Figure 4.5 plots the relations. Word-duration variability[13] is illustrated in Figure 4.6 for two sentence structures (i.e., structure A and G, see section 3.2.1.2 and Appendix I for a description of the sentence structures). For example, word-duration variability decreased during a verbal task, e.g., from an average of 70.25 in VN to 63.81 in VL in sentence structure A or from 60.67 to 53.93 in sentence structure G. Variability was statistically the same in SL and SN, where the raw mean values either slightly increased, e.g., from 65.22 in SN to 67.34 in SL in sentence structure A and from 60.39 to 60.67 in G, or slightly decreased, e.g., from 69.56 to 68.15 in sentence structure B. The other 6 sentence structures are illustrated in Appendix IV.

[13] As the sentences in V and S can be different, sentences were controlled for as covariate. V and S may include different numbers of the same sentences. For example, the same four sentences in Type A in Appendix I were included in V and S, but the total number of each sentence may not be equal. Note, however, that VL and VN or SL and SN included the same numbers of the same sentences produced by the same speakers.

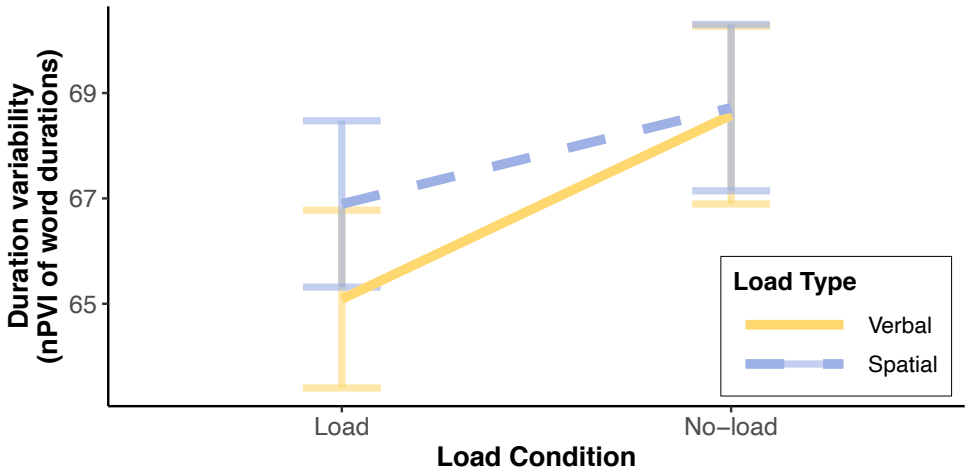
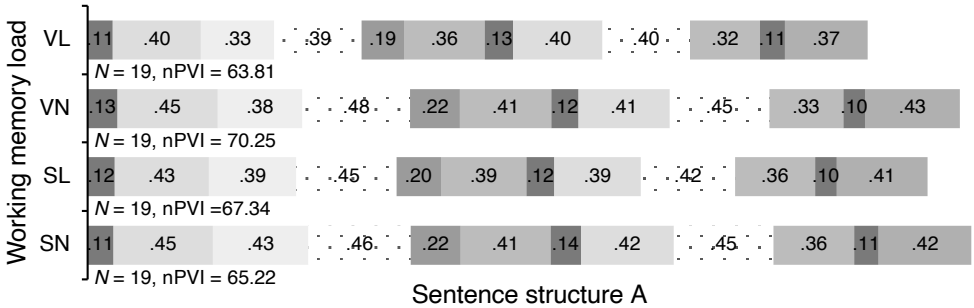
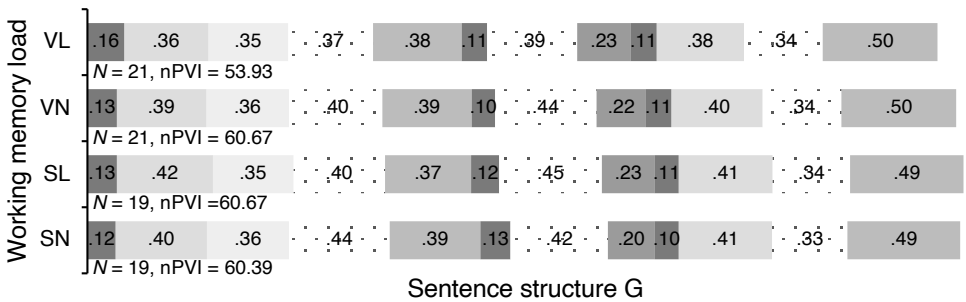


Figure 4.5 Duration variability: less variable L2 word durations during a verbal task
Speech produced during a verbal task exhibited less variable word durations than speech produced in the other three conditions. The other three groups were not statistically different in word duration variability. nPVI = normalized pairwise variability index, where higher numbers correlate with higher variability. Error bars indicate 95% Confidence Interval.



(a) L2 duration variability in sentence structure A (Subject Middle)



(b) L2 duration variability in sentence structure G (Object End)

Figure 4.6 Illustration of word-duration variability by working memory load type and condition

L2 word durations were less variable during a verbal task compared to when the sentences were produced without additional load (VL ≠ VN). Spatial load did not decrease the variability statistically significantly (SL = SN). Intervals represent the mean proportion of 12 words, where the mean was by VL, VN, SL, and SN. Numbers inside the intervals give raw word duration in seconds. The syntactic categories of the words are noted at the bottom of each sentence structure (as D1 A1 and so on, see Appendix I for full structure).

Abbreviations: D = determiner; A = adjective; N = noun; V = verb; R = relativizer; numbers following D, A, and V = the order of each syntactic category within a sentence, e.g., N2 being the second noun in the sentence; *N* = number of sentences; nPVI = normalized pairwise variability index of word durations across a sentence, a higher number denotes higher variation in word durations; VL = verbal load; VN = verbal no-load; SL = spatial load; SN = spatial no-load.

In addition, the L2 speakers' produced vowels were acoustically closer together (having a smaller vowel space, area M = 8.15, SD = 2.62) during a verbal task than during a spatial task (M = 9.07, SD = 2.62) or than when the same sentences were produced without an additional task (M = 9.83, SD = 2.33). The vowel space area in the verbal load was smaller than the areas in the other three conditions. Figure 4.7 demonstrates how speakers' vowel space got smaller due to verbal load during speaking. The 9 vertices were 9 means of all 20 speakers' by-vowel centroids. The vowel types were /i, ɪ, ɛ, æ, ɑ, ʌ, ɔ, ʊ, u/.

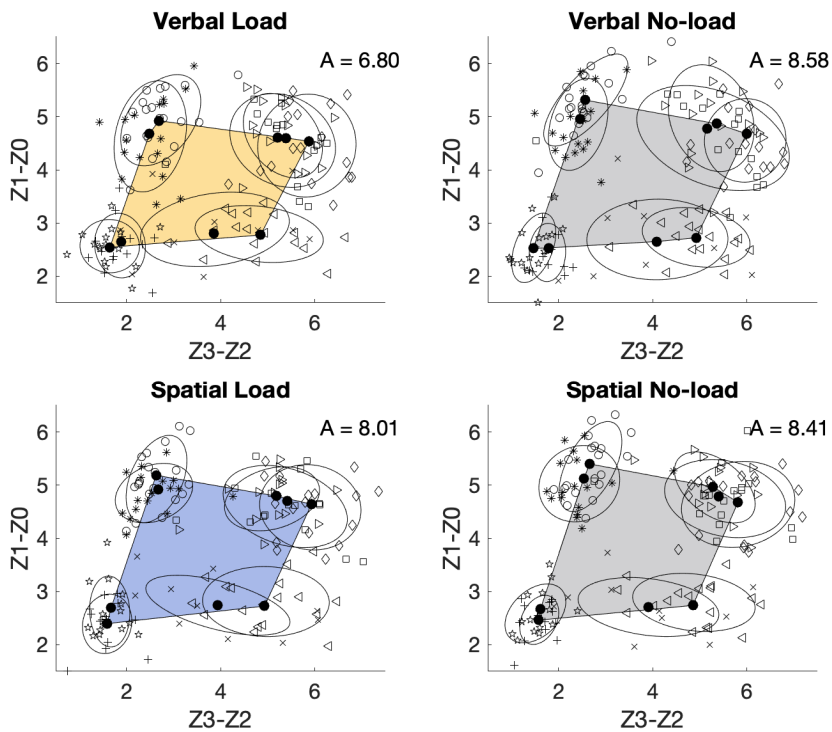


Figure 4.7 Articulation clarity: smaller vowel space (area, A) during a verbal task in speech produced by Korean EFL speakers

4.4 Discussion

The data supported significant effect of working memory load and load type on L2 speech. Korean learners of English as a foreign language spoke aloud English sentences displayed on a computer monitor. When multitasking during speaking, the L2 speakers made more speech errors and spoke faster, compared to when speaking only without an additional task. So far is not surprising and can be considered a general effect of multitasking that we can often observe across various cognitive activities. We usually find it easier to do only one thing at a time than doing multiple different things simultaneously. We usually error more when multi-tasking, if not all tasks done equally well as mono-tasked. When we need to complete multiple tasks within a given set of short limited time, we tend to hurry up and/or allocate a part of time to one task and another part to another task. Then, we would expedite the process for each individual task, and naturally rushing leads to more errors.

These patterns are not unique to language but may apply to almost all other cognitive behaviors. As Cowan (1988) and Engle (2002) explained, our brain is capacity limited in the amount of information that it can pay attention to at a time (about 3 to 5 unrelated information items). Of course, the speaking task in the current study was not a letter-span or a digit-span task, where we can directly measure quantitively how many information items are loaded onto a task, and thus we cannot be sure whether the speakers were cognitively "overloaded." We just assume that the speakers were cognitively "overloaded" while multitasking. Part of support, though, can be the decreased number of correctly produced sentences from 77.7%

in the no-load conditions to 63.5% during the spatial task or to 52.9% during the verbal task (see Appendix V for details). Referring to the embedded-processes models of working memory (Cowan, 1988, 1999; Engle, 2002), we may explain the observed patterns as follows. The speakers were either overloaded in their working memory system during the load conditions. They failed to pay good simultaneous attention to both the consonant letters or spatial locations and the speech materials. Or they were voluntarily shifting their focus of attention to one task and then to the other task, possibly shifting back and forth a few times. In doing so, they did not give enough time and cognitive process to do each task well.

What is important is, in fact, the significant effect of load type. A verbal task disrupted the L2 speech production task significantly more than the spatial task. While it is true that the L2 speakers produced more errors during multitasking, regardless of whether the additional task was verbal or spatial, they produced statistically significantly even more errors when the task was verbal than when it was spatial. Furthermore, only the verbal task changed the speech acoustics. The spatial task did not change the statistical significance of speech patterns. While trying to memorize a sequence of English consonants, Korean L2 speakers articulated the English sentences faster. The same was found when the sentence-internal pauses were included in the total sentence duration to complete an entire sentence and when only the speaking rates were examined excluding pauses as in the articulation rate measure. Their produced words were durationally less variable across sentences. The articulated vowels were acoustically closer to one another

indicating a possibility that the speakers failed to pay more attention to articulate the vowels more clearly and distinctively. These differences were only significant between the verbal load condition and the associated no-load condition where the same sentences were produced by the same speaker. This significant dissociation between spatial and verbal domain is consistent with literature findings that the capacity of working memory is dissociated by the content domain (Baddeley & Hitch, 1974; Seigneuric et al., 2000; Kellogg et al., 2007).

These production differences may well be associated with faster speech, as partly indicated in the significant correlation results among these measures. Faster speech may have resulted in more errors, smaller durational ratios, and less time to move articulators to hit articulatory targets. Speedup due to added processing load was reported for perception in Dronjic (2013), although load types were shown to be irrelevant. The L1 and L2 readers of English speeded up reading when they had to remember the result of a math calculation and concurrently judge morphological grammaticality.

A significant type effect in L2 production is consistent with the predictions of the encoding hypothesis (e.g., Fromkin, 1971; Levelt et al., 1999) that the phonological and phonetic materials are planned and processed online. The morphosyntactic forms, displayed on a monitor, become the input for phonological-phonetic encoding, when verbal working memory allows for planning the metrical and segmental specifications and executing the prosodic words. Then, it may tax a speaker's verbal working memory resources to engage in another linguistic task of

remembering a letter sequence on top of speaking. The overloaded system results in malfunctioning. A spatial task should not overload verbal working memory because it is processed separately in the spatial component of working memory.

By contrast, if the phonological-phonetic information is retrieved from articulatory templates in long-term memory and executed automatically as overly practiced articulatory behavior, we should not find significant type effects. Embedded-processes model does not account for type effect because focus of attention does not distinguish the type of information.

In relation to the working memory models, we suggest the L2 results are consistent with the multi-component model than with the embedded-processes model. We interpret from the significant load type effect that the L2 speakers went through verbal working memory processing to plan and produce L2 speech. Different working memory capacity limit by content domain was implicated in the multi-component model of working memory (Baddeley & Hitch, 1974; Baddeley, 2000; Baddeley et al., 2011). The model postulates two separate storage components, one to maintain and manipulate spatial information and the other for linguistic information. The significant spatial vs. verbal type effect can be explained by the separate storage units. By contrast, the embedded-processes model of working memory (Cowan, 1988, 1999) does not distinguish the type of information for processing, but explains poor performance arises from the limited capacity that our focus of attention can hold up to about four activated chunks for about 20 seconds (Cowan, 1999). The model explains similarly coded items can degrade the

representation of another similarly coded information that is concurrently activated within focus of attention (Cowan, 1999). However, by theory, all information types are processed the same way within the same space of the brain. The capacity limit applies to the number of items and the time limit, but the model does not specify that some types are more difficult than others to activate or manipulate via our attentional processes.

CHAPTER V.

GENERAL DISCUSSION

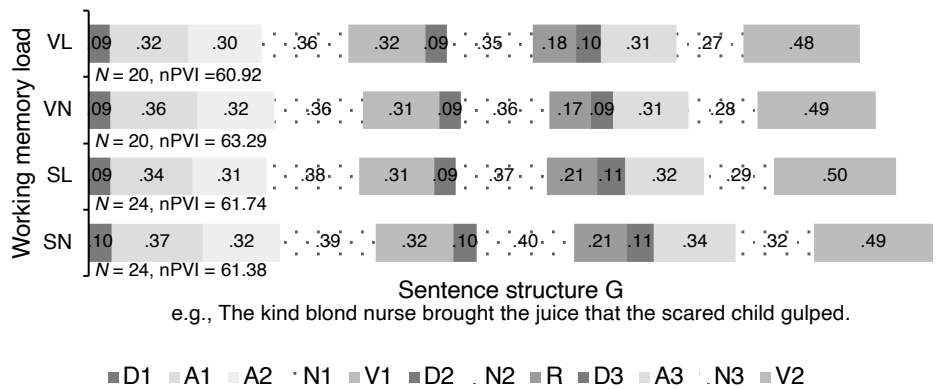
5.1 Working Memory Involvement in L1 and L2 Speech Production

It was suggested that L1 and L2 speakers of English were influenced differently by the manipulated working memory load. The L1 study suggests that L1 speech is not phonologically-phonetically encoded in verbal working memory. The results undercut the hypothesis of phonological-phonetic encoding during speech production and are thus more consistent with a model based on the retrieval of preplanned speech. The results indicate generalized effects of cognitive load on both phonological-type errors and on the rate at which speech is produced. Together, these results suggest that, as with other cognitive activities, when we multitask, we make more errors and we might need to finish one thing faster so we can turn our attention to (the) other tasks.

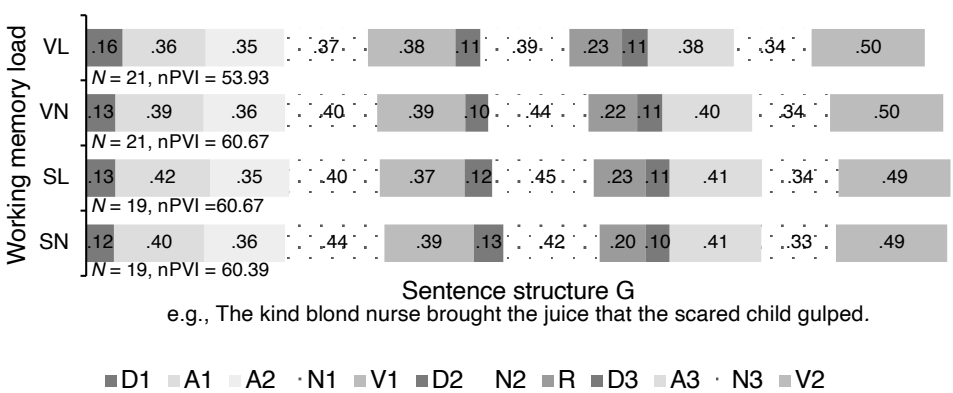
By contrast, the L2 study suggests that L2 speech is phonologically-phonetically encoded via verbal working memory. A significant type effect in L2 production supports the predictions by the encoding hypothesis (e.g., Fromkin, 1971; Levelt et al., 1999) that the phonological and phonetic materials are planned and processed online. The type effect is also in line with the multi-component model than with the embedded-processes model.

Figure 5.1 and Figure 5.2 demonstrates two acoustic differences between L1 and L2 speech. Figure 5.1 shows the native speakers' word duration variability was unaffected by the working memory load or type, while the non-native word durations

became less variable during a verbal working memory task while speaking but not during a spatial task. The same influence from the working memory factors was shown in Figure 5.2. Native speakers' vowel articulation clarity, measured as vowel space area, was intact regardless of the additional cognitive load during speaking. Non-native speakers' articulation was indicated to be potentially less clear during a verbal task than the other three working memory load conditions.

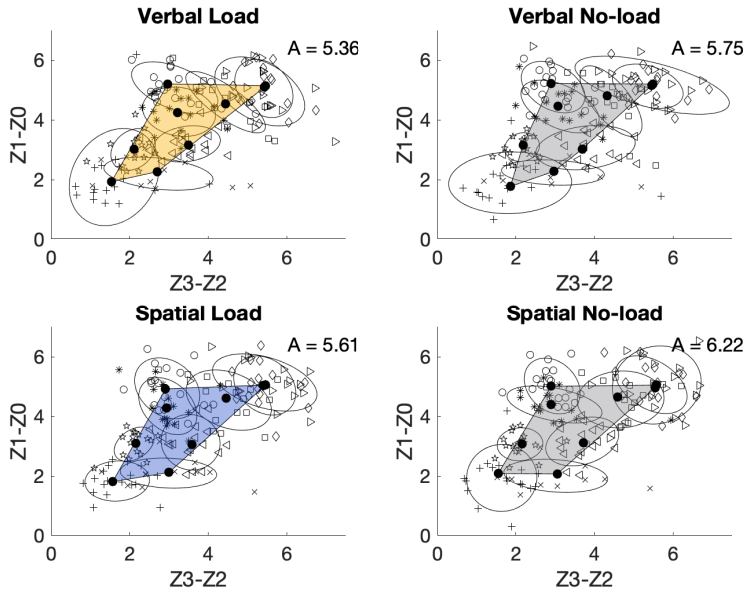


(a) L1 duration variability in sentence structure G (Object End)

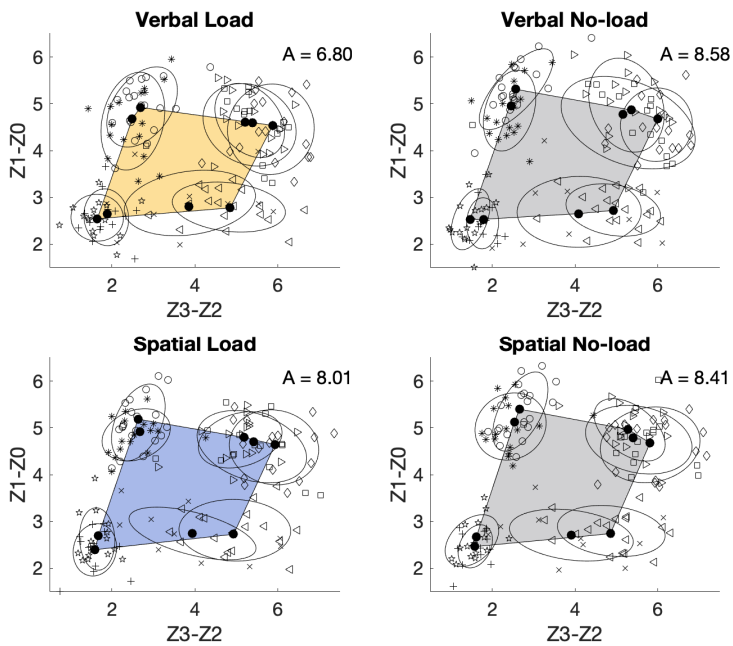


(b) L2 duration variability in sentence structure G (Object End)

Figure 5.1 Word-duration variability in L1 and L2 speech production, in sentence structure G



(a) Vowel space of English L1 speakers, by working memory load type and load condition



(b) Vowel space of English L2 speakers, by working memory load type and load condition

Figure 5.2 Vowel space (area = A) of English L1 and L2 speakers by working memory load type and load condition

5.2 Encoding vs. Retrieval of Phonological and Phonetic Information

Based on the different working memory effects on L1 and L2 speech production processes, I suggest both the encoding and the retrieval hypothesis hold. The phonological and phonetic encoding hypothesis proposed in the staged models of speech production applies to nonnative speech production. The retrieval process applies to native speech production.

Figure 5.3 illustrates the current findings on differential L1 vs. L2 speech production processes, with reference to the processes proposed by the staged models. It adapts the Levelt et al.'s (1999) model of speech production. L1 speakers bypass the phonological and phonetic encoding stages by directly retrieve the already stored articulatory template from long-term memory. L2 speakers go through the speech planning stages as predicted by the staged models.

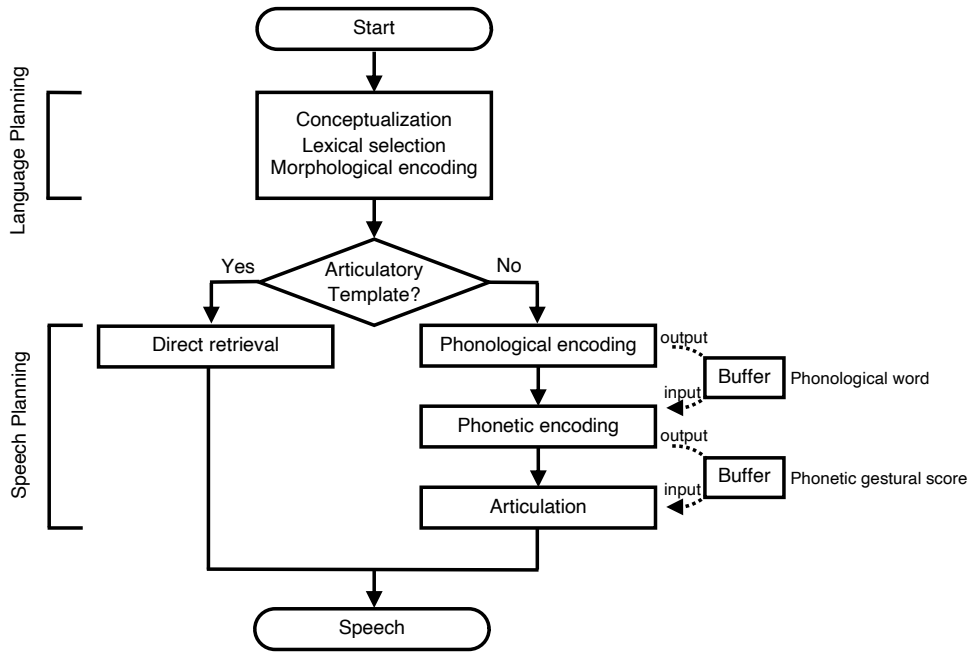


Figure 5.3 Speech production process for L1 and L2
L1 speakers bypass the phonological and phonetic encoding stages by directly retrieve the already stored articulatory template from long-term memory. L2 speakers go through the speech planning stages as predicted by the staged models.

5.3 Proposed Model for L1 and L2 Speech Production

Based on the current findings, we propose a tentative model of speech production that differentiates between L1 speech production and L2 speech production. The model is illustrated in Figure 5.4. It may be true that L1 speech production is automatic and working memory is irrelevant to L1 speech production (Gathercole & Baddeley, 1993). We found a significant effect of load condition, i.e., more planning errors and faster production during a working memory task. However, different from the L2 speakers, the native English speakers produced similar speech patterns regardless of the type of load. L2 speakers use verbal working memory processes to plan and produce L2 speech sounds as well as any other verbal tasks (Baddeley &

Hitch, 1974). Their verbal working memory may be overloaded when engaging in two verbal tasks, and thus we may observe disrupted performance in speech and/or verbal tasks. This dissociation between L1 and L2 implies a dissociation of psycholinguistic processes underlying L1 and L2 speech planning and production.

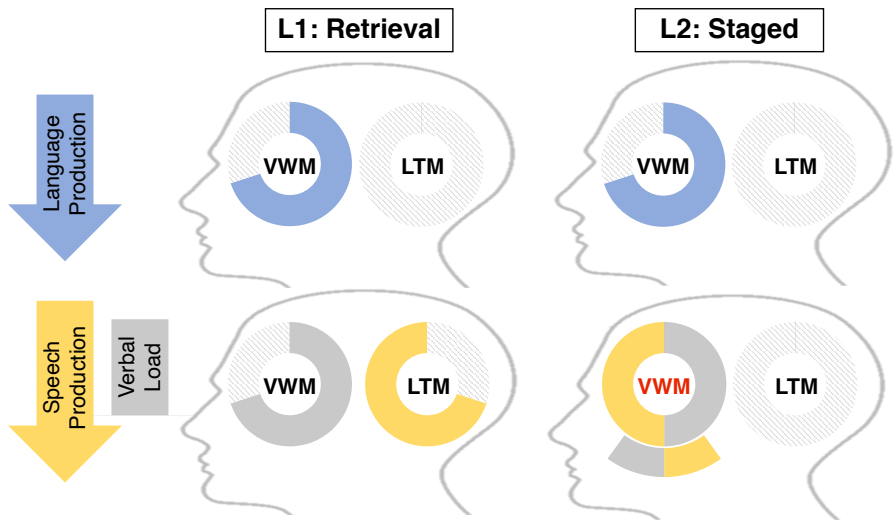


Figure 5.4 Proposed model of L1 and L2 speech production
L2 speech production requires active ongoing speech planning via verbal working memory, different from L1 production via retrieval from memory.

CHAPTER VI.

CONCLUSION

6.1 Pedagogical Implications

The results may have some pedagogical implications. For instance, we now know that even in what we often consider an automatic process, e.g., speaking aloud off of a given sentence, L2 speakers, different from L1 speakers, can be significantly disturbed by one additional task. It was interpreted that L2 speakers undergo active manipulation and ongoing computational process of information for (what may be considered a low-level task of) sound production and articulatory activities. Because they tend to run out their cognitive resources in working memory system by paying attention also to sound generation, it is more likely that L2 speakers produce more errors and/or poor performance in speaking. This is contrasted to L1 speakers, who could automatically perform the same job with little attentional effort, as evidenced in our first experiment and noted in Levelt (1989). It is not just that L1 and L2 speakers are shown to process speech materials via different psycho-cognitive processes, but it is rather that L2 speakers are readily interfered with by external stimuli. This should be taken into consideration in works relevant to L2 populations, including material designs, performance evaluations, etc.

Language teachers are encouraged to help L2 learners develop good articulatory routine procedures for phonological and phonetic information processes.

6.2 Future Direction

This dissertation contributes only a preliminary step to understanding the effects of working memory in speech planning and production. It, however, started an initial step toward the long-term goal of accounting for the cognitive and psychological processes underlying the first- and the second- or foreign-language speech planning and production. Further evidence is required to fully address the questions we asked in the beginning. Some of the many ideas for follow-up studies are laid out, from methodological perspectives and from the topic-wise perspectives.

The first methodological modification we thought of was to make sure participants really engage in simultaneous multitasking Instead of potentially shifting their attention between tasks, we are going to display the letters or locations while they speak. In the current experiment, we followed a format of working memory capacity span tests used in psychology literature. However, in a follow-up experiment, we plan to split the computer screen horizontally in half and display letters or locations in the upper half and the sentence to be produced in the lower half. The time intervals between the letters and between the locations are going to be maintained as in the current experiment.

The display time for a sentence to be produced can be reduced to 4 seconds. One might ask why we would not reduce the time frame of 8 seconds to, say, 5 seconds, given the mean sentence duration was about 3 and 4 seconds. However, including speech errors or some lengthened sentences, we found quite a few sentences produced in up to 7 seconds. If we are not interested in speech error, maybe

yes. However, if we are interested in the disfluent patterns, it is recommended to look at a longer time frame than 7 seconds or longer. We think as long as we present the letters or locations on the same screen while speaking, the total amount of time allowed for each sentential production should not matter much.

Another modification we are currently considering is to provide immediate feedback as to whether participants got the correct answer or not following each question. This is to have participants really do speak only without thinking of anything else. (Of course, we cannot prevent them from daydreaming but we do not think we need to worry about it too much as the daydreaming effects should be the same across all experimental conditions as long as we randomize them.) Especially during the no-load condition, we do not want participants to think about the response that they gave for the previous questions. However, again, the feedback (i.e., correct or incorrect) should not matter much. We guess different individuals might react differently to this modification. Some participants may forget about the previous questions when they get feedback on whether they got them correct or not. Other participants may worry about the results or get dismayed to know they did not do well.

Moving onto future research topics, we have found probably way too many sub-topics related to the current subject matter, and should select some later on. First, it would be interesting to look at spontaneous speech and examine how speakers' segmental and suprasegmental properties change. We wished to first figure out what we should look for in highly controlled experiments, prior to starting to look at

spontaneous speech, which includes probably too much of variability for us to draw sound inferences from. Participants are going to produce different words in various phrasal- and sentential-constructions. We can then collect multiple instances of all or selected types of consonants and vowels produced in load and no-load conditions. We can measure intonational phrases to analyze the changes in their pitch accent, the length of intonational phrases, in-clause and across-sentence pauses, etc.

Second, it can be intriguing to take into account how the currently suggested speech patterns interact with individual differences in working memory capacity. Would speakers with high working memory capacity show completely different patterns or just a lesser amount of disturbance compared to speakers with low working memory capacity? In order to have two groups of individuals with different working memory capacity (i.e., individuals with low working memory capacity and those with high working memory capacity), 80 adults will first be prescreened for working memory capacity with an automated version of the operation span (Aospan) task (provided by Attention and Working Memory Lab at Georgia Institute of Technology, see Unsworth et al., 2005). All will be healthy adult native speakers of Korean. Individuals will be excluded who self-report hearing, speaking or reading problems. Of the 80 participants who participated in the Aospan task, the top quartile of the working-memory-capacity-score distribution will be considered a high working memory capacity group (high-span) while the bottom quartile will be a low capacity group (low-span). We will re-invite the individuals who fall into these groups and who at the same time correctly answered at least 85% of the Aospan

operations. We will try to invite as many and as a balanced-number of people for both groups as possible. During the production experiment, the order of experimental conditions will be counterbalanced by having half the participants engaging in the no-load condition first and the other half in the with-load condition first. The task will be administered by means of e-prime software package 2.0. 5- to 10-minute break will be given in-between the experimental blocks. E-prime software 2.0 will randomly select the sentences to be produced.

Third, it sounds interesting to explore cross-linguistic differences and other population-group differences. Beyond English and Korean, for instance, how would speakers of a language react when the language requires a certain pitch or tone, say Chinese? How about French speakers? Or are different groups of speakers (e.g., L1 children, L1 adults, L2 children, L2 adults) differently hampered by added processing load during speaking?

Fourth, lexical or phrasal frequency effect in relation to speakers' speaking proficiency seems the most interesting and relevant to the current findings. If the currently suggested differences between L1 and L2 indeed resulted from practice or training in the perception and production experience, via sound-embedded mental lexicon and over-practiced articulators, would highly proficient L2 learners, or especially proficient speakers, approximate the L1 patterns? Would high-frequency words or phrases show different results from low-frequency words or phrases? If the retrieval model grounds on over-practice of the articulatory routines in order for the routines to be stored in long-term memory, are all routines stored? The purpose is to

support a view that over-practice helps the articulatory routines be stored and ultimately to support the idea that articulatory routines are stored and ready for retrieval as argued in the retrieval models.

Tentative assumptions and predictions are as follows. First, for high-frequency-words (HF) production, if we define frequency with respect to their use in spoken conversations, then we can argue that HF are articulatorily over practiced for adult native speakers. Then, following the retrieval models and Bybee (2001, 2002), these should be stored in our mental lexicon with their associated phonological-phonetic detail. Second, for low-frequency-words (LF) production, (even when we speak our native language, if we are asked to pronounce LF,) LF are not articulatorily well practiced enough to be stored in our lexicon to the extent that their articulatory gestures are ready for automatic retrieval.

The load can be manipulated as including both high-frequency words/expressions and low-frequency words/expressions in each level of the load conditions. Lexical or expression frequency can be defined using the frequency of use in spoken corpora; we should match the number of syllables and control for other possible properties; if we want to analyze potential effects on syllabification and stress-assignment, among other things, we can include multi-syllable words; we can include function words and content words, but we may want to include HF content words as most function words are short (syllables); LF should not be too low because then we might have disfluent productions in the no-load conditions (as well as in the load conditions), in which case the analysis can be difficult—we want to avoid too

much disfluent productions in the no-load conditions; some LF words can be foreign city names or animal names.

Possible conclusions may be that, if we find more disfluent/erroneous speech in low-frequency expressions (compared to high-frequency expressions), over-practice in producing HF words contributed to storing the articulatory gestures in long-term memory and thus help the routines be ready for direct retrieval as suggested in retrieval models; for LF, even native speakers have not fully practiced pronouncing those words. Thus, in order to produce low-frequency words, even native speakers of a language need to engage in phonological-phonetic encoding than retrieval.

Fifth and lastly for now, if we find some interesting results from some of the future experiments, we can look further into the patterns with purposes of devising pedagogical applications that we can practically make use of. Rather than simply giving out conclusions from in-lab research, we would like to apply the findings in actual learning and teaching fields.

REFERENCES

Abney, S., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research, 20*(3), 233-250.

Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language, 89*, 320-328.

Ackermann, H., Wildgruber, D., Daum, I., & Grodd, W. (1998). Does the cerebellum contribute to cognitive aspects of speech production? A functional magnetic resonance imaging (fMRI) study in humans. *Neuroscience Letters, 247*, 187-190.

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech, Language, and Hearing Research, 38*(2), 403-414.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267-281). Budapest, Hungary: Akademiai Kiado.

Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 8*(1), 1-40.

Anderson, J. R. (1996). A simple theory of complex cognition. *American Psychologist, 51*(4), 355-365.

Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York, NY: John Wiley & Sons, Inc.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Anderson, J. R., & Lebiere, C. J. (1998). *The atomic components of thought.* New York, NY: Psychology Press.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*(3), 221-256.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York, NY: Academic Press.

Baars, B. J., & Gage, N. M. (2010). *Cognition, brain, and consciousness: Introduction to cognitive neuroscience* (2nd ed.). San Diego, CA: Academic Press.

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

Baddeley, A. D. (1986). *Working memory.* New York, NY: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 255*(5044), 556-559.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. D. (2003a). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189-208.

Baddeley, A. D. (2003b). Working memory: Looking back and looking forward. *Nature Reviews, 4*, 829-839.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *The Annual Review of Psychology, 63*, 1-29.

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*, 1393-1400.

Baddeley, A. D., & Andrade, J. (1994). Reversing the word-length effect: A comment on Caplan, Rochon, and Waters. *The Quarterly Journal of Experimental Psychology, 47A*(4), 1047-1054.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York, NY: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York, NY: Cambridge University Press.

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.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65-94.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language, 45*, 1-20.

Barth, D., & Kapatsinski, V. (2014). A multimodel inference approach to categorical variant choice: Construction, priming and frequency effects on the choice between full and contracted forms of *am*, *are* and *is*. *Corpus Linguistics and Linguistic Theory, 13*(2), 203-260.

Bartoń, K. (2019). *MuMIn: Multi-Model Inference. R package version 1.43.15*. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Maechler, M., & Bolker, B. (2013). *lme4: Linear mixed-effects models using S4 classes. R package (version 0.999999-2).* Retrieved from http://CRAN.R-project.org/package=lme4

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S. Jun (Ed.), *Prosodic Typology* (pp. 9-54). Oxford, UK: Oxford University Press.

Blanken, G., Dittmann, J., & Wallesch, C. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters, 325*, 72-74.

Blevins, J. (2006). New perspectives on English sound patterns: Natural and unnatural in evolutionary phonology. *Journal of English Linguistics, 34*(1), 6-25.

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355-387.

Bock, K. (1987). Exploring levels of processing in sentence production. In G. Kempen (Ed.), *Natural language generation: New results in artificial intelligence, psychology and lingusitics* (pp. 351-363). Dordrecht, The Netherlands: Martinus Nijhoff.

Boomer, D. S., & Laver, J. D. (1968). Slips of the tongue. *British Journal of Disorders of Communication, 3*(1), 2-12.

Borst, J. P., & Anderson, J. R. (2013). Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proceedings of the National Academy of Sciences, 110*(5), 1628-1633.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English. *Phonology and Second Language Acquisition, 36*, 287-308.

Brazil, D. (1980). *Discourse intonation and language teaching*. New York, NY: Longman.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy, & A. Wilkes (Eds.), *Studies in long-term memory.* New York, NY: John Wiley & Sons.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*(3-4), 155-180.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30*, 542-579.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition, 42*(1-3), 261-286.

Caplan, D., Alpert, N., & Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience, 10*(4), 541-552.

Caplan, D., Alpert, N., Waters, G., & Olivieri, A. (2000). Activation of broca's area by syntactic processing under conditions of concurrent articulation. *Human Brain Mapping, 9*, 65-71.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *The Quarterly Journal of Experimental Psychology, 45A*(2), 177-192.

Caplan, D., & Waters, G. (1994). Articulatory length and phonological similarity in span tasks: A reply to Baddeley and Andrade. *The Quarterly Journal of Experimental Psychology, 47A*(4), 1055-1062.

Caplan, D., & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*, 77-126.

Caplan, D., & Waters, G. (2002). Working memory and connectionist models of parsing: A reply to MacDonald and Christiansen. *Psychological Review, 109*(1), 66-74.

Caplan, D., & Waters, G. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory, 13*(3-4), 403-413.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1), 177-208.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1430-1450.

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W., & Thulborn, K. (1999). Graded functional activation in the visuospatial system with the amount of task demand. *Journal of Cognitive Neuroscience, 11*(1), 9-24.

Carpenter, P. A., Miyake, A., & Just, M. A. (1995). Language comprehension: Sentence and discourse processing. *Annual Review of Psychology, 46*, 91-120.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge, UK: Cambridge University Press.

Chafe, W. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), *Coherence and grounding in discourse.* Amsterdam, The Netherlands: John Benjamins Publishing Company.

Chang, H. (2008). Effectiveness of pronunciation training on suprasegmentals. *Korean Journal of Applied Linguistics, 24*(1), 85-108.

Chein, J. M., Ravizza, S. M., & Fiez, J. A. (2003). Using neuroimaging to evaluate models of working memory and their implications for language processing. *Journal of Neurolinguistics, 16*, 315-339.

Cocchini, G., Logie, R. H., Sala, S. D., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory and Cognition, 30*(7), 1086-1095.

Colom, R., Martínez-Molina, A., Shih, P. C., & Santacreu, J. (2010). Intelligence, working memory, and multitasking performance. *Intelligence, 38*, 543-551.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language and Cognition, 2*(3), 231-244.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163-191.

Cowan, N. (1995). *Attention and memory: An integrated framework. Oxford Psychology Series, No. 26.* New York, NY: Oxford University Press. (Paperback edition 1997).

Cowan, N. (1998). Visual and auditory working memory capacity. *Trends in Cognitive Sciences, 2*(3), 77-78.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). New York, NY: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114.

Cowan, N. (2005). *Working memory capacity.* New York, NY: Psychology Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323-338.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51-57.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42-100.

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70*(5), 1082-1097.

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review, 119*(3), 480-499.

Croft, W. (1995). Intonation units and grammatical structure. *Linguistics, 33*, 839-882.

Crompton, A. (1982). *Syllables and segments in speech production.* In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 109-162). Amsterdam, The Netherlands: Mouton.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics, 5*(6), 365-373.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Dauer, R. M. (2005). The lingua franca core: A new model for pronunciation instruction? *TESOL Quarterly, 39*(3), 543-550.

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

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes, 23*(4), 583-608.

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

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1355-1367.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611-629.

del Viso, S., Igoa, J. M., & García-Albea, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20*(3), 161-185.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation training: A research-based approach. *TESOL Quarterly, 39*(3), 379-398.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*, 1-17.

Dronjic, V. (2013). *Concurrent memory load, working memory span, and morphological processing in L1 and L2 English* (Doctoral dissertation, University of Toronto). Retrieved from https://tspace.library.utoronto.ca

Dunkel, P. (1991). Listening in the native and second/foreign language: Toward an integration of research and practice. *TESOL Quarterly, 25*(3), 431-457.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York, NY: Teachers College, Columbia University. (Original work published 1885)

Endestad, T. (2005). Metaphors of memory: To reconstruct a dinosaur. *Synergies,* 83-85.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 44, pp. 145-200). San Diego, CA: Elsevier.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). New York, NY: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.

Fant, G., Kruckenberg, A., & Ferreira, J. B. (2003). Individual variations in pausing: A study of read speech. *PHONUM, 9*, 193-196.

Faw, B. (2003). Pre-frontal executive committee for perception, working memory, attention, long-term memory, motor control, and thinking: A tutorial review. *Consciousness and Cognition, 12*(1), 83-139.

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

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

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

Ferreira, F., & Engelhardt, P. E. (2006). Syntax and production. In M. J. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 61-91). London, UK: Academic Press.

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

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*(3), 399-423.

Fillmore, C. J. (2006). Frame semantics. In D. Geeraerts (Ed.), *Cognitive linguistics: Basic readings* (pp. 373-400). Berlin, Germany: Mouton de Gruyter.

Finch, H., & French, B. (2013). A monte carlo comparison of Robust MANOVA test statistics. *Journal of Modern Applied Statistical Methods, 12*(2), 4.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of Acoustical Society of America, 101*, 3728-3740.

Fowler, C. A. (2007). Speech production. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 489-501). Oxford, UK: Oxford University Press.

Fowler, C. A. (2010). Speech production. In I. B. Weiner, & W. E. Craighead (Eds.), *The corsini encyclopedia of psychology* (Vol. 4, 4th ed., pp. 1685-1687). Hoboken, NJ: John Wiley & Sons.

Fowler, C. A., Rubin, C. F., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 373-420). New York, NY: Academic Press.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, *36*(2-3), 171-195.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172-179.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*(1), 27-52.

Fry, D. (1969). The linguistic evidence of speech errors. *BRNO Studies of English, 8*, 69-74.

Fullana, N. (2006). The development of English (FL) perception and production skills. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 41-64). Clevedon, UK: Multilingual Matters Ltd.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 133-177). New York, NY: Academic Press.

Garrett, M. F. (1989). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey* (Vol. 3 Language: Psychological and biological aspects, pp. 69-96). Cambridge, UK: Cambridge University Press.

Gathercole, S. E., & Baddeley, A. D. (1993). Speech production. In *Working memory and language* (pp. 75-100). New York, NY: Psychology Press.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2*(2), 103-127.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*(4), 411-458.

Gee, J. P., & Grosjean, F. (2002). Performance structures: A psycholinguistic and linguistic appraisal. In Gerry T. M. Altmann (Ed.), *Psycholinguistics: Critical concepts in psychology* (Vol. 5, pp. 121-167). New York, NY: Routledge.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1-76.

Gile, D. (2008). Cognitive load in simultaneous interpreting and its implications for empirical research. *Forum, 6*(2), 59-77.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*(1), B1-B14.

Guion, S., Clark, J. J., Harada, T., & Wayland, R. P. (2003). Factors affecting stress placement for English nonwords include syllable structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech, 46*(4), 403-427.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*(2), 274-307.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-Is-Power Hypothesis. *Cognitive Psychology, 44*, 339-387.

Haque, S. M. T., Al-Ameen, M. N., Wright, M., & Scielzo, S. (2017). Learning system-assigned passwords (up to 56 bits) in a single registration session with the methods of cognitive psychology. *Proceedings of USEC, The Internet Society, February 2017*. San Diego, CA.

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Applied Psycholinguistics, 24*, 495–522.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8*, 191-219.

Harnesberger, J. D. (2001). On the relationship between identification and discrimination of nonnative nasal consonants. *Journal of the Acoustical Society of America, 110*, 489–503.

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech, 9*(4), 207-216.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews, 8*, 393-402.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of Acoustical Society of America, 97*(5), 3099-3111.

Hirata, Y. (2003). Learning to form new L2 phonetic categories in sentence contexts. In M. J. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 515–518). Barcelona, Spain.

Hirschberg, J. (1993). Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence, 63*(1-2), 305-340.

Hismanoglu, M. (2006). Current perspectives on pronunciation leaning and teaching. *Journal of Language and Linguistic Studies, 2*(1), 101-110.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297-307.

Hwang, E. (2008). Factors affecting Korean learners' English pronunciation and comprehensibility. *English Teaching, 63*(4), 3-28.

Jacobs, J. (1887). Experiments on prehension. *Mind*, *12*, 75-79.

James, W. (1950). *The principles of psychology* (Vol. 1). New York, NY: Dover. (Original work published 1890)

Jenkins, J. (2002). A sociolinguistically based, empirically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*(1), 83-103.

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

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*, 193-224.

Jun, S. (2005). Prosody in sentence processing: Korean vs. English. *UCLA Working Papers in Phonetics, 104*, 26-45.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kane, M. J., & Engle, R. W. (2000). Working memory capacity, proactive interference and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 336-358.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to stroop interference. *Journal of Experimental Psychology: General, 132*(1), 47-70.

Keller, T. A., Cowan, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 635-645.

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica, 124*, 382-397.

Kempen, G., & Hoenkamp, E. (1982). Incremental sentence generation: Implications for the structure of a syntactic processor. In J. Horecky (Ed.), *Proceedings of the 9th International Conference on Computational Linguistics, Prague, July 1982*. Amsterdam: North-Holland Publishing Company. 151-156.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11*(2), 201-258.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580-602.

Kleinow, J., & Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research, 43*, 548-559.

Kormos, J. (2006). *Speech production and second language acquisition*. New York, NY: Routledge.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics, 35*(2), 162-179.

Krivokapić, J. (2010). Speech planning and prosodic phrase length. In *Speech Prosody 2010, Chicago, IL, May 10-14, 2010.* Paper 311.

Kumpf, L. (1987). The use of pitch phenomena in the structuring of stories. In R. S. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 189-216). Amsterdam, The Netherlands: John Benjamins.

Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, *66*, 588-611.

Lee, O., & Ahn, H. (2020). Faster and less clear L2 speech with more errors during a verbal working memory task but not during a spatial task. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Annual Pronunciation in Second Language Learning and Teaching Conference* (pp. 141-153). Flagstaff, AZ: Northern Arizona University. ISSN 2380-9566.

Lee, O., & Redford, M. A. (2015). Verbal and spatial working memory load have similarly minimal effects on speech production. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (Paper number 0798, pp. 1-5). Glasgow, UK: University of Glasgow. ISBN 978-0-85261-941-4.

Leikin, M., & Assayag-Bouskila, O. (2004). Expression of syntactic complexity in sentence comprehension: A comparison between dyslexic and regular readers. *Reading and Writing: An Interdisciplinary Journal, 17*, 801-821.

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

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

Levelt, W. J. (1995). The ability to speak: From intentions to spoken words. *European Review, 3*(1), 13-23.

Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences, 3*(6), 223-232.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

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

Levis, J. M. (2005). Streaming speech: Listening and pronunciation for advanced learners of English. *TESOL Quarterly, 39*(3), 559-562.

Lively, S. E., Pisoni, D. B., Summers, W. V., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of Acoustical Society of America, 93*(5), 2962-2973.

London, J., & Jones, K. (2011). Rhythmic refinements to the nPVI measure: A reanalysis of Patel & Daniele (2003a). *Music Perception: An Interdisciplinary Journal, 29*(1), 115-120.

Low, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43*, 377-401.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391-400.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35-54.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181-209). Amsterdam, The Netherlands: John Benjamins.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*(1), 19-44.

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of Acoustic Society of America, 125*(6), 3962-3973.

Martin, R. C., Crowther, J. E., Knight, M., Tamborello II, F. P., & Yang, C. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition, 116*, 177-192.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*(1), 1-21.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*, 524-545.

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

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42*, 181-211.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63*(2), 81-97.

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology, 98*(4), 839-843.

Miller, S. F. (2001). Targeting pronunciation: The intonation, sounds, and rhythm of American English. Boston, MA: Houghton Mifflin.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Miyake, A., & Shah, P. (Eds.). (1997). *Models of working memory: Mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press

Morey, C. C., & Cowan, N. (2004). When visual and verbal memory compete: Evidence of corss-domain limits in working memory. *Psychonomic Bulletin & Review, 11*(2), 296-301.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly, 25*(3), 481-520.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology, 117*(3), 411–424.

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface, 14*, 20170213.

Nakagawa, S., & Schielzeth, H. (2013). A generalized and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution 2013, 4*, 133-142.

Nam, H., Goldstein, L., Saltzman, E., & Byrd, D. (2004). TADA: An enhanced, portable Task Dynamics model in MATLAB. *The Journal of the Acoustical Society of America, 115*(5), 2430-2430.

Norrelgen, F., Lacerda, F., & Forssberg, H. (1999). Speech discrimination and phonological working memory in children with ADHD. *Developmental Medicine and Child Neurology, 41*, 335-339.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of single-word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18*, 443-468.

Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory and Cognition, 37*(3), 346-357.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1997). A biologically-based computational model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375-411). New York, NY: Cambridge University Press.

Ormanci, E., Nikbay, U. H., Turk, O., & Arslan, L. M. (2002). Subjective assessment of frequency bands for perception of speaker identity. In J. Hansen, & B. Pellom (Eds.), *Proceedings of the 7th International Conference on Spoken Language Processing 2002*, 2581-2584. Denver, Colorado.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1-8.

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Pan, S., & McKeown, K. R. (1999). Word informativeness and automatic pitch accent modeling. *In Proceedings of EMNLP/VLC'99*, 148-157.

Paul, E. (2009). *ESL speakers' production of English lexical stress: The effect of variation in acoustic correlates on perceived intelligibility and nativeness* (Doctoral dissertation, University of New Mexico). Retrieved from https://digitalrepository.unm.edu

Pecher, D. (2013). No role for motor affordances in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 2-13.

Pederson, E., & Guion, S. (2010). Orienting attention during phonetic training facilitates learning. *Journal of Acoustic Society of America, 127*(2), EL54-59.

Pedhazur, E. J., & Schmelkin, L. P. (1991). Multiple categorical independent variables: Factorial designs. In *Measurement, design, and analysis: An integrated approach* (pp. 504-544). New York, NY: Psychology Press.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 539-557.

Petrone, C., Fuchs, S., & Krivokapić, J. (2011). Consequences of working memory differences and phrasal length on pause duration and fundamental frequency. In *Proceedings of the 9th International Seminar on Speech Production* (pp. 393-400). Montreal, Canada.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psycholinguistical Bulletin, 134*(3), 427-459.

Pickett, J. M., & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech, 3*(1), 11-17.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*(2), 191-215.

Redford, M. A. (2014). The perceived clarity of children's speech varies as a function of their default articulation rate. *The Journal of the Acoustical Society of America, 135*(5), 2952-2963.

Redford, M. A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics, 53*, 141-152.

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural bases of strategy and skill in sentence-picture verification. *Cognitive Psychology, 40*, 261-295.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfield and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). Amsterdam, The Netherlands: John Benjamins.

Rochon, E., Waters, G. S., & Caplan, D. (2000). The relationship between measures of working memory and sentence comprehension in patients with Alzheimer's disease. *Journal of Speech, Language, and Hearing Research, 43*, 395-413.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107-142.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249-284.

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

Ross, E. D. (2000). Affective prosody and the aprosodias. In M.-M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (pp. 316-331). Oxford, UK: Oxford University Press.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*(8), 2022-2039.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333-382.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knolwedge structures*. Hillsdale, NJ: Lawrence Erlbaum.

Schnotz, W., & Kurschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*, 469-508.

Schuetze-Coburn, S., Shapley, M., & Weber, E. G. (1991). Units of intonation in discourse: A comparison of acoustic and auditory analyses. *Language and Speech, 34*(3), 207-234.

Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing: An Interdisciplinary Journal, 13*, 81-103.

Selkirk, E. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The handbook of phonological theory* (2nd ed., pp. 435-485). Oxford, UK: Wiley-Blackwell.

Selkirk, E., & Katz, J. (2008). Contrastive focus, givenness and the unmarked status of 'discourse-new'. *Acta Hungarica Linguistica, 55*(3-4), 331-346.

Sereno, J., Lammers, L., & Jongman, A. (2016). The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics, 37*(2), 303-322.

Shah, P., & Miyake, A. (1999). Models of working memory: An introduction. In P. Shah, & A. Miyake (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 1-27). New York, NY: Cambridge University Press.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper, & E. C. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109-136). New York, NY: Springer-Verlag.

Shattuck-Hufnagel, S. (2015). Prosodic frames in speech production. In M. Redford (Ed.), *The handbook of speech production* (pp. 419-444). Malden, MA: Wiley-Blackwell.

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

Shriberg, E. E. (1994). *Preliminaries to a theory of speech disfluencies* (Doctoral dissertation, University of California, Berkeley). Retrieved from http://citeseerx.ist.psu.edu/index

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition, 11*(2), 245-260.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6*, 11-20.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1-29.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience, 4*, 232-243.

Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language, 28*, 164-188.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language, 52*, 452-473.

Suen, C. Y., & Beddoes, M. P. (1974). The silent interval of stop consonants. *Language and Speech, 17*(2), 126-134.

Sugiura, N. (1978). Further analysts of the data by akaike's information criterion and the finite corrections. *Communications in Statistics: Theory and Methods, 7*(1), 13-26.

Sweller, J. (1988). Cognitive load during problem solving. *Cognitive Science, 12*, 257–285.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2014). Individual differences in the scope of speech planning: Evidence from eye-movements. *Language and Cognition, 6*, 12-44.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*(4), 1086-1100.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (7th ed.). New York, NY: Pearson.

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. Retrieved from http://lingtools.uoregon.edu/norm/index.php

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of Acoustical Society of America, 69*(5), 1465-1475.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of Acoustical Society of America, 88*(1), 97-100.

Traunmüller, H., Eriksson, A., & Ménard, L. (2003). Perception of speaker age, sex and vowel quality investigated using stimuli produced with an articulatory model. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1739-1742). Barcelona, Spain: Causal Productions. ISBN 1-876346-49-3.

Turk, A., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1658), 20130395.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104-132.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 489-505.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*(6), 635-654.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science, 8*(4), 314-317.

Vigliocco, G., Vinson, D. P., Martin, R. C., & Garrett, M. F. (1999). Is "count" and "mass" information available when the noun is not? An investigation of tip of the tongue states and anomia. *Journal of Memory and Language, 40*(4), 534-558.

Warren, P. (1996). Prosody and parsing: An introduction. *Language and Cognitive Processes, 11*(1-2), 1-16.

Waters, G. S., & Caplan, D. (1996a). Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition, 24*(3), 342-355.

Waters, G. S., & Caplan, D. (1996b). The Capacity Theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*(4), 761-772.

Waters, G. S., & Caplan, D. (1996c). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology, 49A*(1), 51-79.

Waters, G. S., & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and Aging, 16*(1), 128-144.

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers, 35*(4), 550-564.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology, 57A*(1), 129-163.

Waters, G. S., Rochon, E., & Caplan, D. (1998). Task demands and sentence comprehension in patients with dementia of the Alzheimer's type. *Brain and Language, 62*, 361-397.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19*, 713-755.

Watson, D., & Gibson, E. (2005). Intonation phrasing and constituency in language production and comprehension. *Studia Linguistica, 59*(2-3), 279-300.

Weismer, S. E., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1249-1260.

Wen, Z. E. (2016). *Working memory and second language learning: Towards an integrated approach*. Bristol, England: Multilingual Matters.

Wheeldon, L. R., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356-381.

Winkler, I., & Cowan, N. (2004). From sensory to long-term memory: Evidence from auditory memory reactivation studies. *Experimental Psychology, 51*(3), 1-17.

APPENDICES

Appendix I Speech Materials

32 sentences by sentence structure with respect to relative-clause type and location

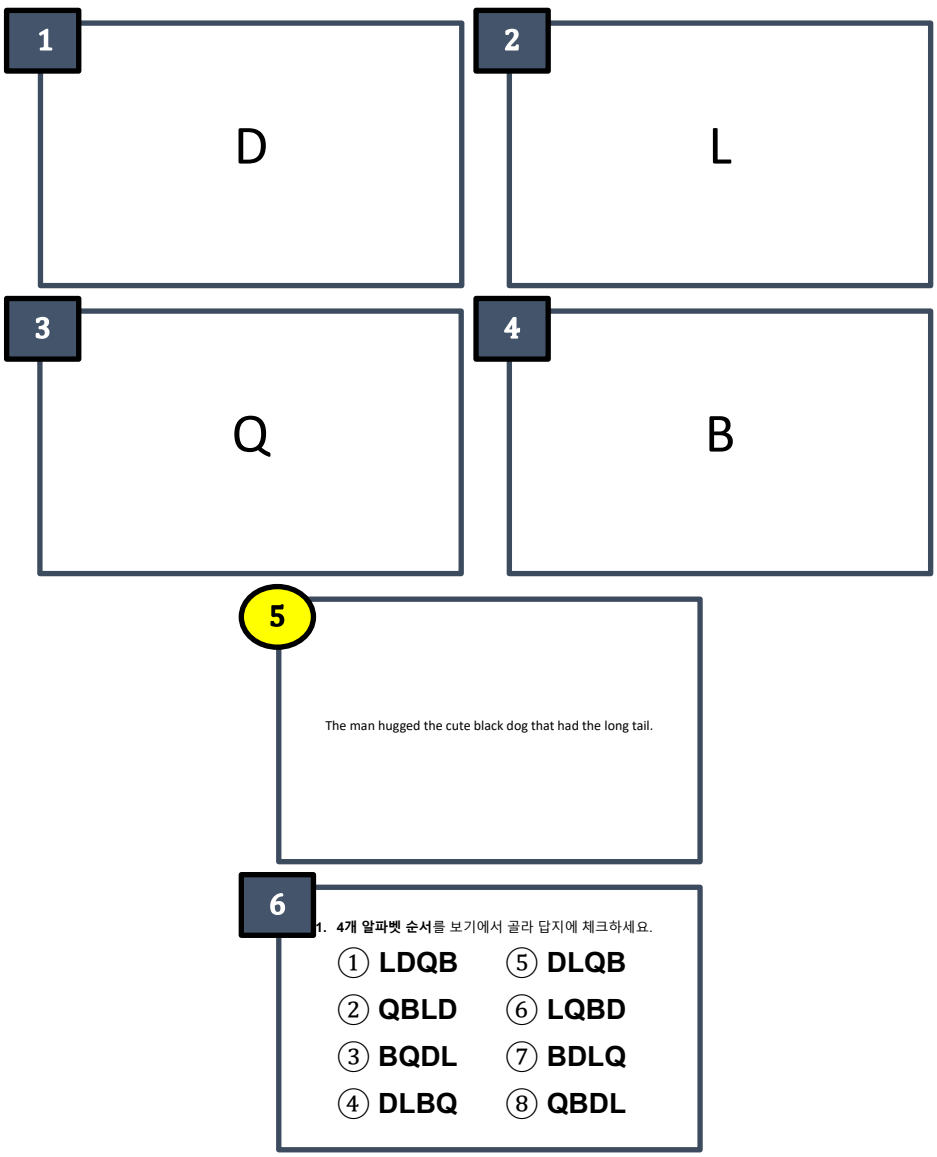
RC Type	Location	Sentences	*N*
Subject-extracted	Middle	**Structure A**: [D_1 A_1 A_2 N_1 [R V_1 D_2 A_3 N_2]$_{RC}$]$_{Sbj}$ V_2 [D_3 N_3]$_{Obj}$	
		The smart shy boy that liked the quiet girl cut the cake.	
		The loud brown bird that pecked the young child took the jam.	
		The wild bad guy that whipped the poor horse lost the race.	
		The swift bold lion that ate the soft sheep saw the trap.	
		Structure B: [D_1 N_1 [R V_1 D_2 A_1 N_2]$_{RC}$]$_{Sbj}$ V_2 [D_3 A_2 A_3 N_3]$_{Obj}$	
		The guy that saw the tall friend caught the slow green bus.	
		The cat that chased the fast mouse blocked the deep dark hole.	
		The maid that helped the sick aunt cleaned the old blue room.	
		The boy that built the weird stool dumped the odd wet trash.	8
	End	**Structure C**: [D_1 A_1 A_2 N_1]$_{Sbj}$ V_1 [D_2 N_2 [R V_2 D_3 A_3 N_3]$_{RC}$]$_{Obj}$	
		The fun cute girl loved the man that wrote the thick book.	
		The sly gray wolf bit the sheep that wore the gold bell.	
		The calm kind boy pet the dog that chewed the big bone.	
		The new young nurse met the child that drew the great house.	
		Structure D: [D_1 N_1]$_{Sbj}$ V_1 [D_2 A_1 A_2 N_2 [R V_2 D_3 A_3 N_3]$_{RC}$]$_{Obj}$	
		The cook fed the quick sharp thief that took the new cup.	
		The child stopped the small red fox that crossed the short fence.	
		The friend rode the nice large cow that had the white spots.	
		The man hugged the cute black dog that had the long tail.	8
Object-extracted	Middle	**Structure E**: [D_1 A_1 A_2 N_1 [R D_2 A_3 N_2 V_1]$_{RC}$]$_{Sbj}$ V_2 [D_3 N_3]$_{Obj}$	
		The fat black cat that the mad dog hurt climbed the tree.	
		The strong tall man that the wild horse kicked crushed the can.	
		The fun smart bird that the great aunt raised pecked the ball.	
		The short stout thief that the cool cop stalked stole the ring.	
		Structure F: [D_1 N_1 [R D_2 A_1 N_2 V_1]$_{RC}$]$_{Sbj}$ V_2 [D_3 A_2 A_3 N_3]$_{Obj}$	
		The hen that the wise pig watched left the small hot coop.	
		The girl that the mean bear shocked grabbed the large hard rock.	
		The maid that the bad guy pushed dropped the fine gold key.	
		The friend that the clean cook loved drank the fresh cold juice.	8
	End	**Structure G**: [D_1 A_1 A_2 N_1]$_{Sbj}$ V_1 [D_2 N_2 [R D_3 A_3 N_3 V_2]$_{RC}$]$_{Obj}$	
		The kind blond nurse brought the juice that the scared child gulped.	
		The big white cat tore the mat that the cool guy left.	
		The swank rich man bought the paint that the young girl chose.	
		The good gray dog found the book that the sad friend read.	
		Structure H: [D_1 N_1]$_{Sbj}$ V_1 [D_2 A_1 A_2 N_2 [R D_3 A_3 N_3 V_2]$_{RC}$]$_{Obj}$	
		The boy smelled the fresh sweet bread that the great cook baked.	
		The goat squashed the round brown box that the old maid hid.	
		The cook stirred the hot green soup that the good guy ate.	
		The child read the big blue card that the neat aunt sent.	8

Note: RC = relative clause; Sbj = subject; Obj = object; D = determiner; A = adjective; N = noun; V = verb; R = relativizer; Subscript numbers denote the order of each syntactic category within a sentence (e.g., D_2 is the second determiner in the sentence).

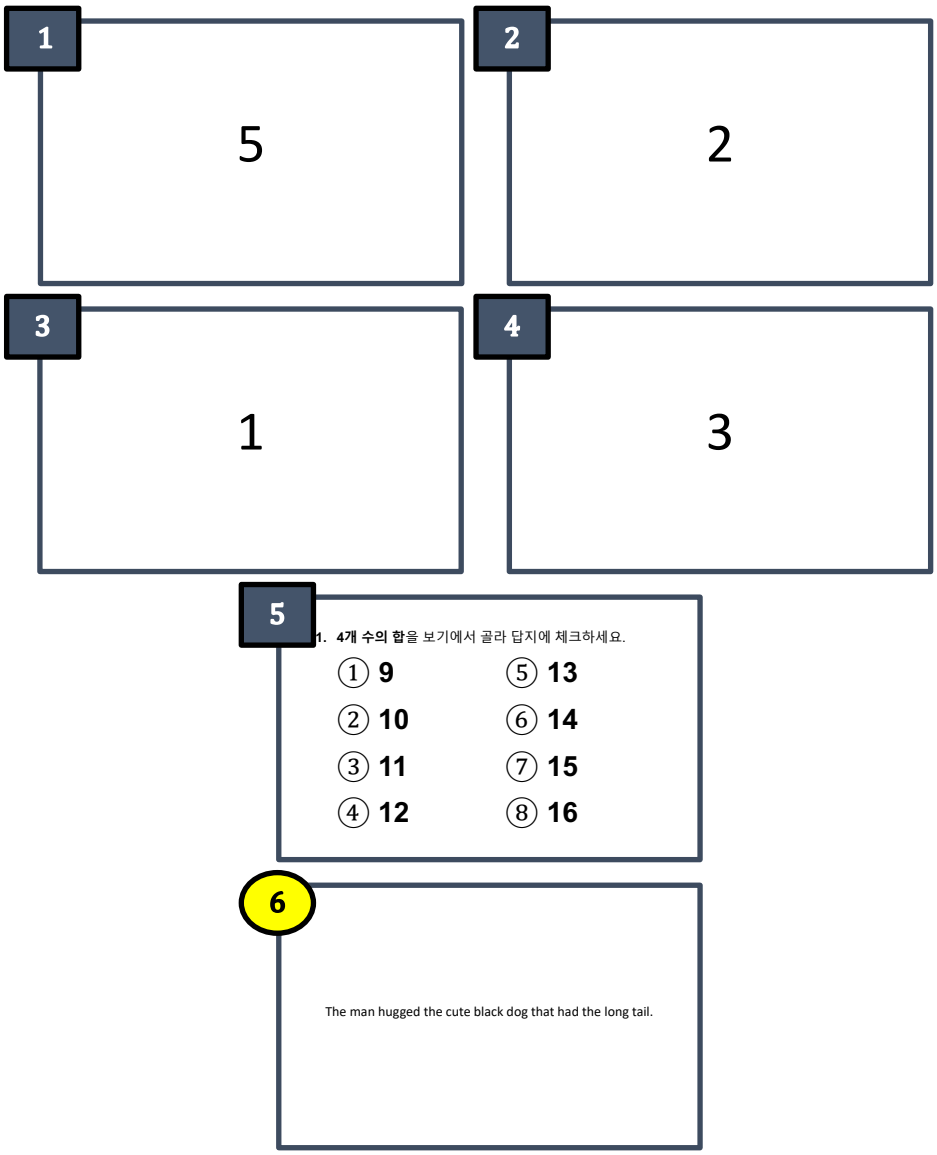
Appendix II Working Memory Manipulation

Illustration of the actual presentation of the four experimental blocks: (a) Verbal Load, (b) Verbal No-load, (c) Spatial Load, and (d) Spatial No-load. Six slides make one question. The order of the slides is here denoted at the upper left corner of each slide. Note the position of the sentence (production task, marked as circled number) relative to the working memory load task (distractor task, squared numbers).

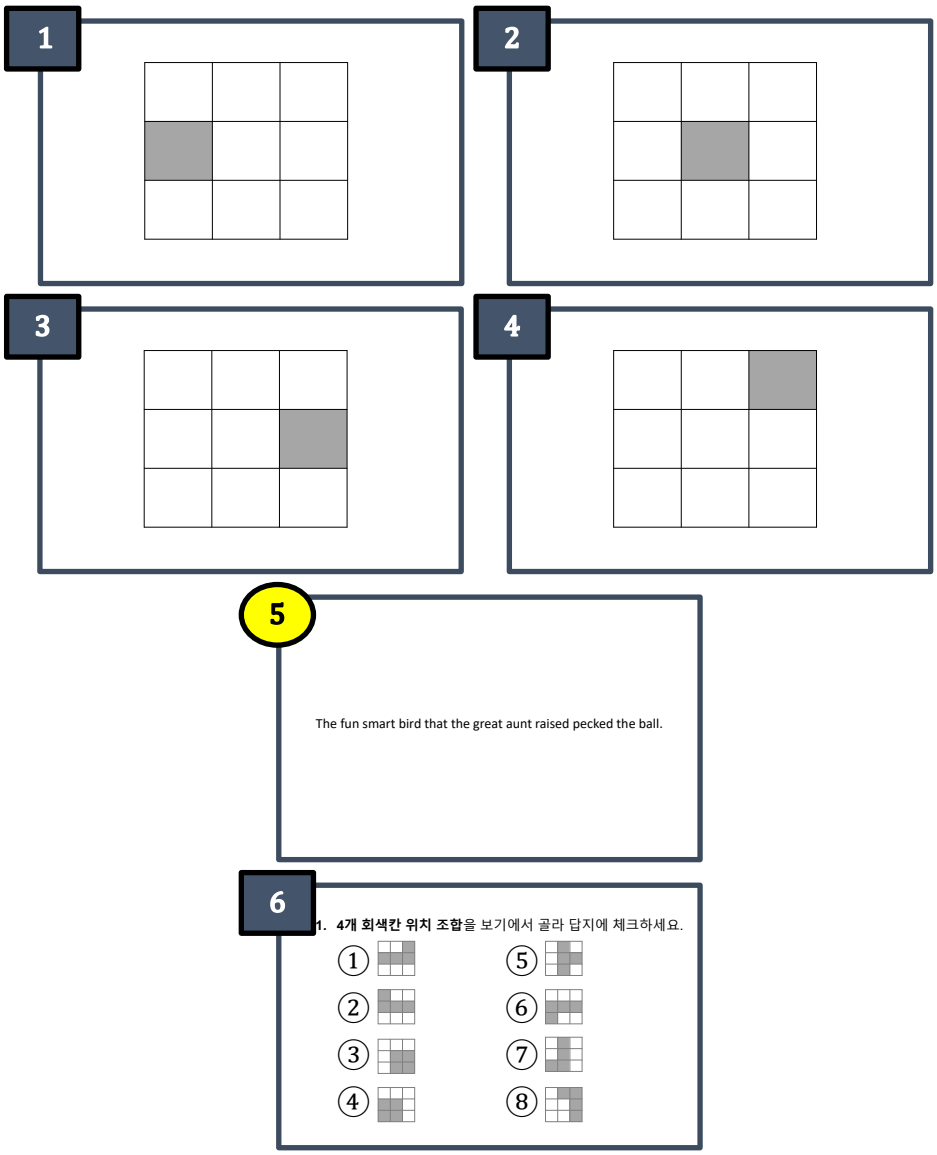
(a) Verbal Load



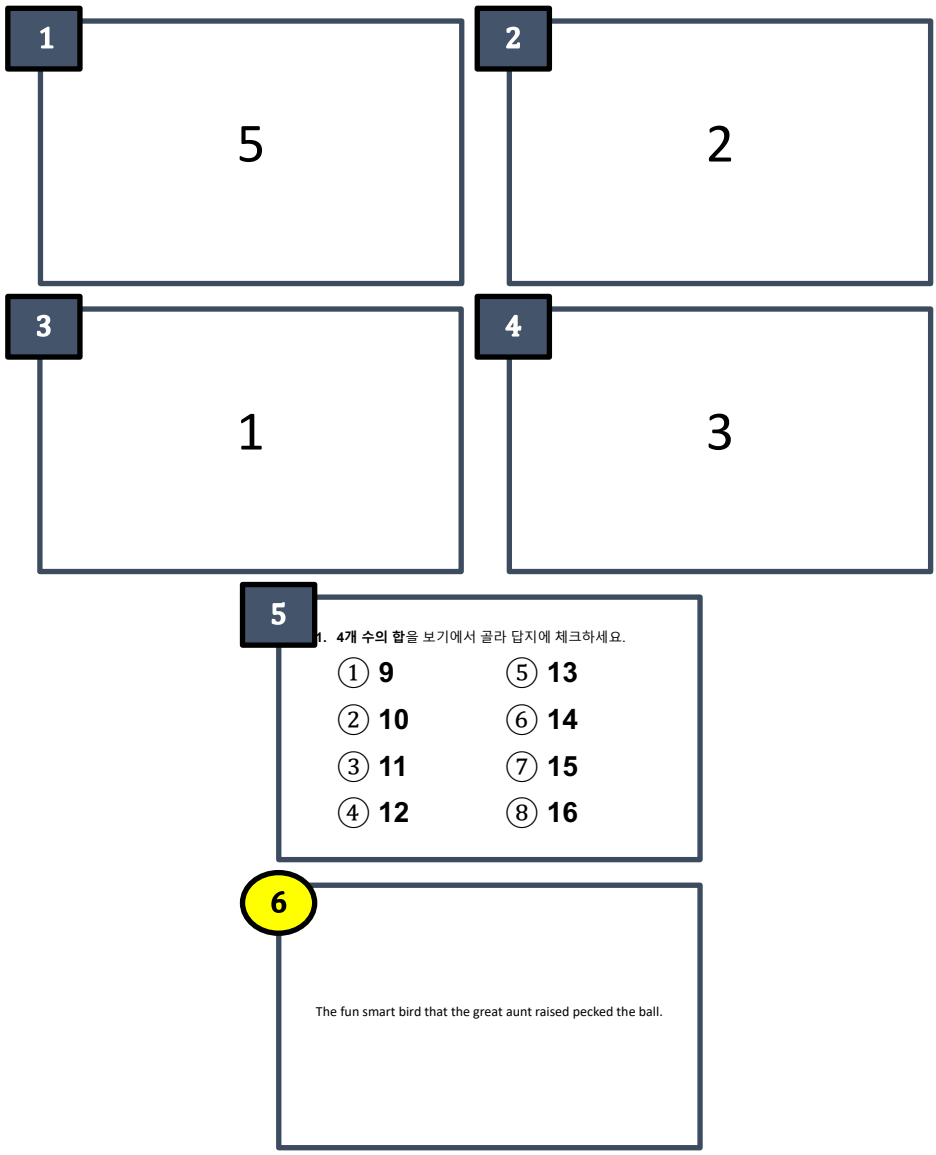
(b) (Verbal) No-load



(c) Spatial Load



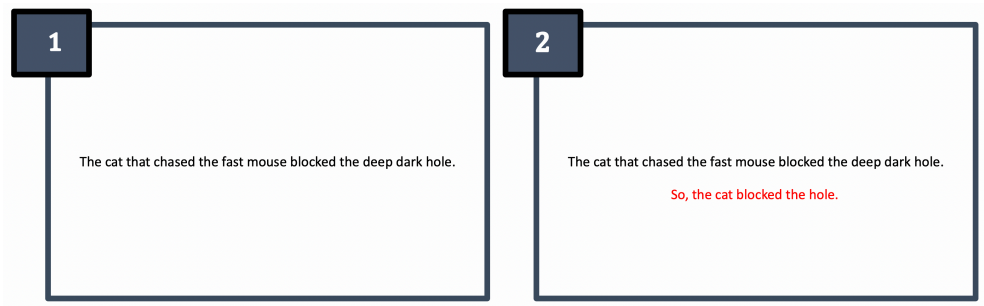
(d) (Spatial) No-load



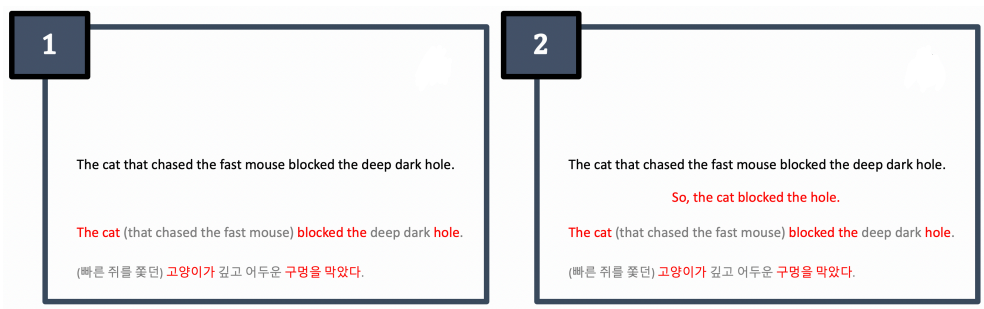
Appendix III Familiarization Procedure

Illustration of the run-through stage for familiarization of the sentences prior to the main production phase: (a) slides presented to the L1 speakers and (b) slides to L2. Two successive slides were allocated to each of the thirty-two sentences. The order of the slides, from 1 to 2, is (only here) marked at the upper left corner of each slide. The second slides highlighted the matrix clause to assist comprehension. For Korean L2 learners of English, additional information was provided for translation in Korean and the relative clause separated within parentheses.

(a) Presented to L1 speakers



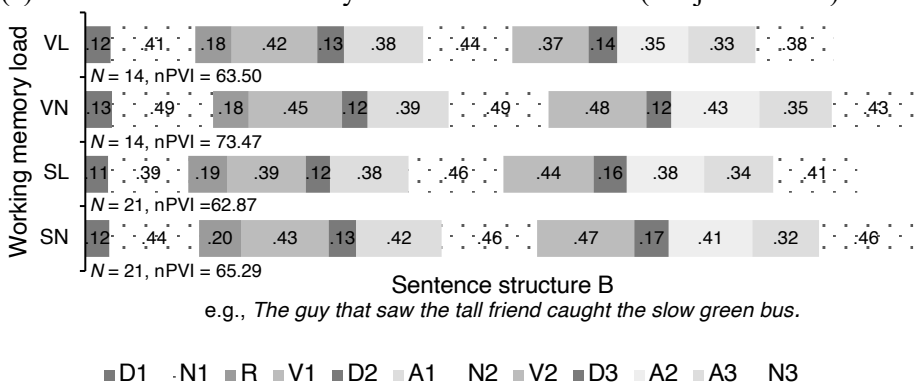
(b) Presented to L2 speakers



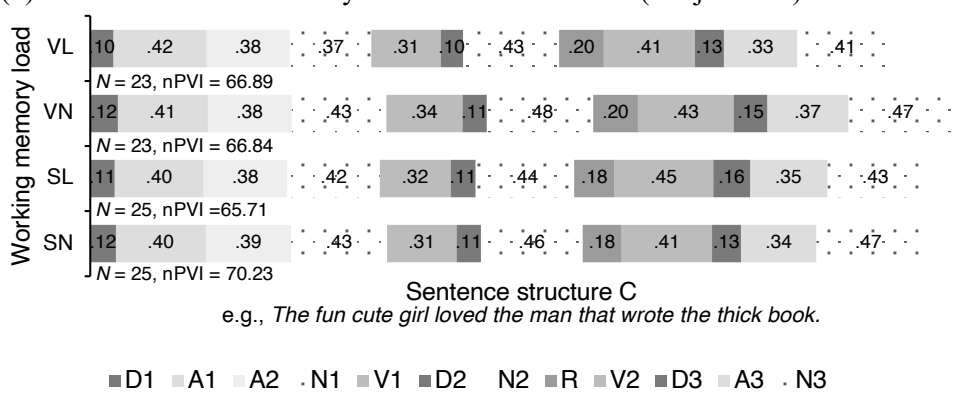
Appendix IV Word Duration Variability in L2 Speech

For 6 out of 8 sentence structures (see 3.2.1.2 for description and Appendix I for full list). The other two are in Figure 4.6. (*N* = number of sentences).

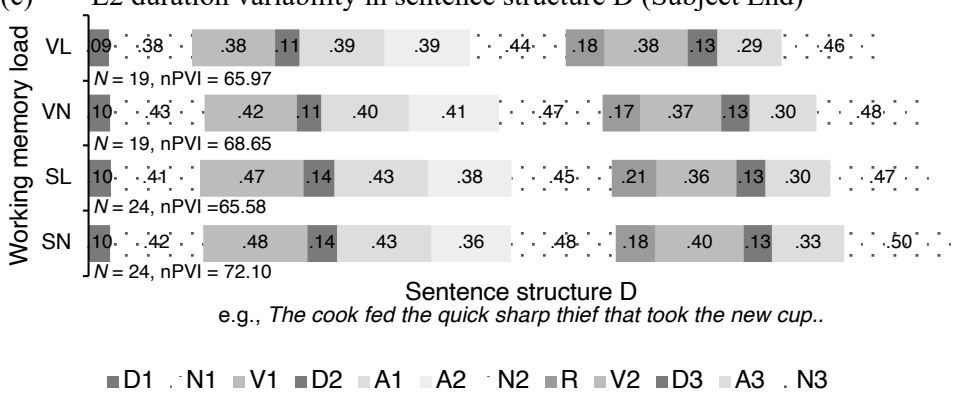
(a) L2 duration variability in sentence structure B (Subject Middle)



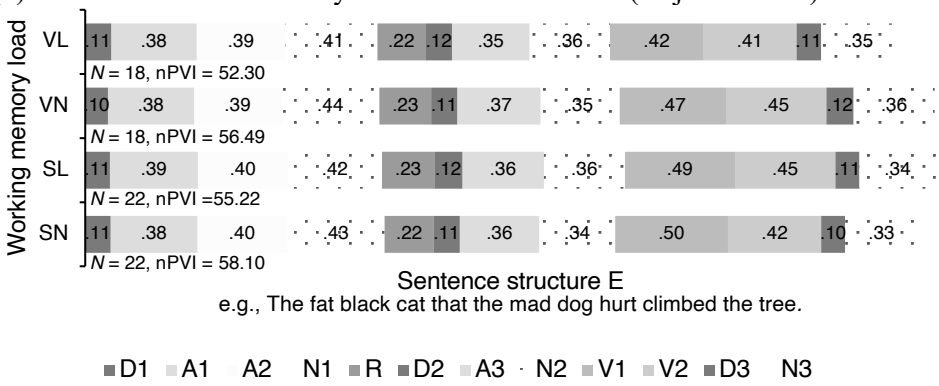
(b) L2 duration variability in sentence structure C (Subject End)



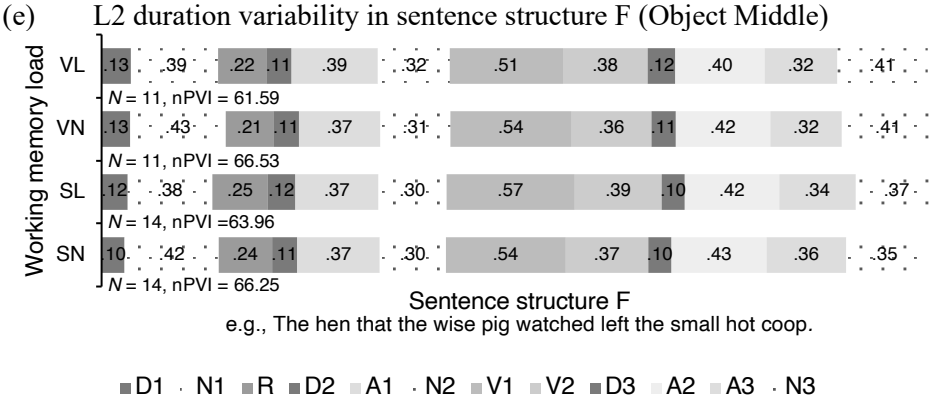
(c) L2 duration variability in sentence structure D (Subject End)



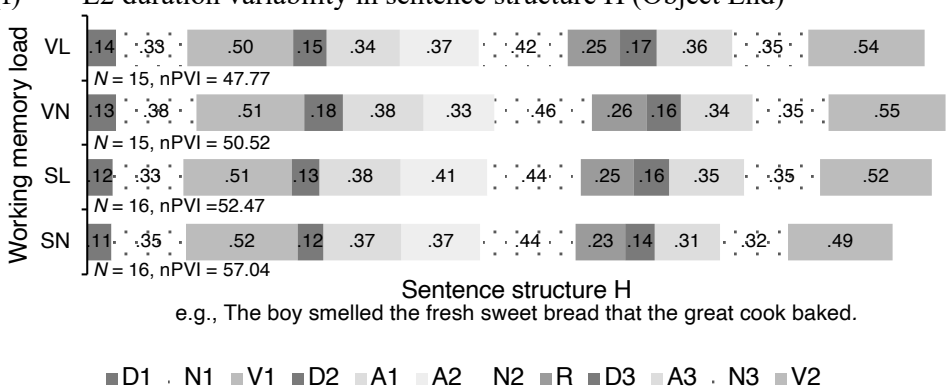
(d) L2 duration variability in sentence structure E (Object Middle)



(e) L2 duration variability in sentence structure F (Object Middle)



(f) L2 duration variability in sentence structure H (Object End)



Appendix V Incorrect Sentences by Working Memory Load Type and Load Condition

Distribution is tabled in the number of disfluent-incorrect sentences / total (error rate, %): per speaker (in a and c) and per sentence (in b and d) for L1 and L2 data.

(a) L1 per speaker

	Verbal		Spatial		
Speaker	Load	No-load	Load	No-load	Total
1	1/15 (6.7)	1/15 (6.7)	4/15 (26.7)	3/15 (20.0)	9/60 (15.0)
2	3/15 (20.0)	1/15 (6.7)	2/15 (13.3)	2/15 (13.3)	8/60 (13.3)
3	3/15 (20.0)	1/15 (6.7)	3/15 (20.0)	4/15 (26.7)	11/60 (18.3)
4	4/16 (25.0)	5/16 (31.3)	1/15 (6.7)	3/15 (20.0)	13/62 (21.0)
5	5/16 (31.3)	4/16 (25.0)	3/15 (20.0)	0/15 (0.0)	12/62 (19.4)
6	3/16 (18.8)	5/16 (31.3)	4/15 (26.7)	2/15 (13.3)	14/62 (22.6)
7	5/15 (33.3)	3/15 (20.0)	2/16 (12.5)	2/16 (12.5)	12/62 (19.4)
8	7/15 (46.7)	6/15 (40.0)	4/16 (25.0)	3/16 (18.8)	20/62 (32.3)
9	13/15 (86.7)	6/15 (40.0)	10/16 (62.5)	3/16 (18.8)	32/62 (51.6)
10	7/15 (46.7)	1/15 (6.7)	3/16 (18.8)	2/16 (12.5)	13/62 (21.0)
11	3/16 (18.8)	1/16 (6.3)	6/15 (40.0)	1/14 (7.1)	11/61 (18.0)
12	3/15 (20.0)	5/13 (38.5)	5/14 (35.7)	7/16 (43.8)	20/58 (34.5)
13	3/15 (20.0)	2/15 (13.3)	3/14 (21.4)	2/16 (12.5)	10/60 (16.7)
14	7/15 (46.7)	6/15 (40.0)	5/14 (35.7)	3/16 (18.8)	21/60 (35.0)
15	7/15 (46.7)	5/15 (33.3)	3/14 (21.4)	1/16 (6.3)	16/60 (26.7)
16	5/15 (33.3)	1/16 (6.3)	7/15 (46.7)	1/15 (6.7)	14/61 (23.0)
17	9/16 (56.3)	3/16 (18.8)	10/15 (66.7)	1/15 (6.7)	23/62 (37.1)
18	3/15 (20.0)	0/15 (0.0)	3/16 (18.8)	2/16 (12.5)	8/62 (12.9)
19	3/15 (20.0)	5/15 (33.3)	3/16 (18.8)	4/16 (25.0)	15/62 (24.2)
Total	94/290 (32.4)	61/289 (21.1)	81/287 (28.2)	46/294 (15.6)	282/1160 (24.3)

(b) L1 per sentence

	Verbal		Spatial		
Sentence	Load	No-load	Load	No-load	Total
1 (A1)	1/8 (12.5)	2/9 (22.2)	3/10 (30.0)	2/10 (20.0)	8/37 (21.6)
2 (A2)	2/9 (22.2)	1/9 (11.1)	1/10 (10.0)	0/10 (0.0)	4/38 (10.5)
3 (A3)	2/10 (20.0)	4/9 (44.4)	4/9 (44.4)	1/9 (11.1)	11/37 (29.7)
4 (A4)	3/10 (30.0)	2/10 (20.0)	3/9 (33.3)	1/9 (11.1)	9/38 (23.7)
5 (B1)	1/9 (11.1)	0/9 (0.0)	3/10 (30.0)	0/10 (0.0)	4/38 (10.5)
6 (B2)	3/9 (33.3)	1/9 (11.1)	4/10 (40.0)	5/10 (50.0)	13/38 (34.2)
7 (B3)	7/10 (70.0)	6/10 (60.0)	4/9 (44.4)	3/9 (33.3)	20/38 (52.6)
8 (B4)	5/10 (50.0)	4/10 (40.0)	3/9 (33.3)	2/9 (22.2)	14/38 (36.8)
9 (C1)	0/9 (0.0)	2/9 (22.2)	3/10 (30.0)	2/10 (20.0)	7/38 (18.4)
10 (C2)	0/9 (0.0)	2/9 (22.2)	3/10 (30.0)	1/10 (10.0)	6/38 (15.8)
11 (C3)	1/10 (10.0)	1/10 (10.0)	2/9 (22.2)	1/9 (11.1)	5/38 (13.2)
12 (C4)	2/10 (20.0)	2/10 (20.0)	3/9 (33.3)	1/9 (11.1)	8/38 (21.1)
13 (D1)	5/9 (55.6)	1/9 (11.1)	3/10 (30.0)	3/10 (30.0)	12/38 (31.6)
14 (D2)	2/9 (22.2)	0/9 (0.0)	1/10 (10.0)	2/10 (20.0)	5/38 (13.2)
15 (D3)	5/10 (50.0)	2/9 (22.2)	2/9 (22.2)	0/9 (0.0)	9/37 (24.3)
16 (D4)	4/10 (40.0)	2/10 (20.0)	1/9 (11.1)	1/9 (11.1)	8/38 (21.1)
17 (E1)	2/9 (22.2)	2/9 (22.2)	2/10 (20.0)	2/10 (20.0)	8/38 (21.1)
18 (E2)	2/9 (22.2)	1/9 (11.1)	2/10 (20.0)	2/10 (20.0)	7/38 (18.4)
19 (E3)	3/10 (30.0)	3/10 (30.0)	1/9 (11.1)	0/9 (0.0)	7/38 (18.4)
20 (E4)	3/10 (30.0)	3/10 (30.0)	5/9 (55.6)	1/9 (11.1)	12/38 (31.6)
21 (F1)	4/9 (44.4)	1/9 (11.1)	3/10 (30.0)	2/10 (20.0)	10/38 (26.3)
22 (F2)	1/6 (16.7)	1/6 (16.7)	2/6 (33.3)	2/10 (20.0)	6/28 (21.4)
23 (F3)	2/10 (20.0)	0/10 (0.0)	1/9 (11.1)	1/9 (11.1)	4/38 (10.5)
25 (G1)	6/9 (66.7)	2/9 (22.2)	2/6 (33.3)	1/10 (10.0)	11/34 (32.4)
26 (G2)	0/9 (0.0)	0/9 (0.0)	1/10 (10.0)	0/10 (0.0)	1/38 (2.6)
27 (G3)	3/10 (30.0)	4/10 (40.0)	2/9 (22.2)	1/9 (11.1)	10/38 (26.3)
28 (G4)	7/10 (70.0)	0/10 (0.0)	3/9 (33.3)	2/9 (22.2)	12/38 (31.6)
29 (H1)	3/9 (33.3)	1/9 (11.1)	4/10 (40.0)	3/10 (30.0)	11/38 (28.9)
30 (H2)	4/9 (44.4)	5/9 (55.6)	4/10 (40.0)	2/10 (20.0)	15/38 (39.5)
31 (H3)	4/10 (40.0)	2/10 (20.0)	2/9 (22.2)	0/9 (0.0)	8/38 (21.1)
32 (H4)	7/10 (70.0)	4/10 (40.0)	4/9 (44.4)	2/8 (25.0)	17/37 (45.9)
Total	94/290 (32.4)	61/289 (21.1)	81/287 (28.2)	46/294 (15.6)	282/1160 (24.3)

(c) L2 per speaker

	Verbal		Spatial		
Speaker	Load	No-load	Load	No-load	Total
1	6/16 (37.5)	9/16 (56.3)	4/15 (26.7)	5/15 (33.3)	24/62 (38.7)
2	11/16 (68.8)	4/16 (25.0)	6/15 (40.0)	2/15 (13.3)	23/62 (37.1)
3	10/15 (66.7)	1/15 (6.7)	5/16 (31.3)	3/16 (18.8)	19/62 (30.6)
4	6/15 (40.0)	1/15 (6.7)	7/15 (46.7)	3/16 (18.8)	17/61 (27.9)
5	8/15 (53.3)	4/15 (26.7)	5/16 (31.3)	3/16 (18.8)	20/62 (32.3)
6	8/15 (53.3)	0/15 (0.0)	5/15 (33.3)	1/16 (6.3)	14/61 (23.0)
7	9/15 (60.0)	4/15 (26.7)	5/16 (31.3)	5/16 (31.3)	23/62 (37.1)
8	6/16 (37.5)	3/16 (18.8)	4/15 (26.7)	0/15 (0.0)	13/62 (21.0)
9	10/16 (62.5)	4/16 (25.0)	10/15 (66.7)	6/15 (40.0)	30/62 (48.4)
10	5/16 (31.3)	1/16 (6.3)	6/15 (40.0)	5/15 (33.3)	17/62 (27.4)
11	8/15 (53.3)	2/15 (13.3)	6/16 (37.5)	3/15 (20.0)	19/61 (31.1)
12	2/15 (13.3)	3/15 (20.0)	3/16 (18.8)	3/16 (18.8)	11/62 (17.7)
13	6/15 (40.0)	1/16 (6.3)	3/15 (20.0)	4/15 (26.7)	14/61 (23.0)
14	10/16 (62.5)	2/16 (12.5)	6/16 (37.5)	4/16 (25.0)	22/64 (34.4)
15	7/16 (43.8)	4/16 (25.0)	3/15 (20.0)	2/15 (13.3)	16/62 (25.8)
16	8/16 (50.0)	5/16 (31.3)	9/15 (60.0)	4/15 (26.7)	26/62 (41.9)
17	5/16 (31.3)	3/16 (18.8)	6/16 (37.5)	8/16 (50.0)	22/64 (34.4)
18	8/16 (50.0)	6/16 (37.5)	7/16 (43.8)	4/16 (25.0)	25/64 (39.1)
19	7/16 (43.8)	6/16 (37.5)	7/16 (43.8)	2/16 (12.5)	22/64 (34.4)
20	7/16 (43.8)	4/16 (25.0)	6/16 (37.5)	5/16 (31.3)	22/64 (34.4)
Total	147/312 (47.1)	67/313 (21.4)	113/310 (36.5)	72/311 (23.2)	399/1246 (32.0)

(d) L2 per sentence

	Verbal		Spatial		
Sentence	Load	No-load	Load	No-load	Total
1 (A1)	2/10 (20.0)	2/10 (20.0)	2/10 (20.0)	3/10 (30.0)	9/40 (22.5)
2 (A2)	4/10 (40.0)	1/10 (10.0)	1/8 (12.5)	0/10 (0.0)	6/38 (15.8)
3 (A3)	7/10 (70.0)	3/10 (30.0)	6/10 (60.0)	3/10 (30.0)	19/40 (47.5)
4 (A4)	6/10 (60.0)	0/10 (0.0)	6/10 (60.0)	4/10 (40.0)	16/40 (40.0)
5 (B1)	4/10 (40.0)	2/10 (20.0)	1/10 (10.0)	1/10 (10.0)	8/40 (20.0)
6 (B2)	4/10 (40.0)	3/10 (30.0)	6/10 (60.0)	1/10 (10.0)	14/40 (35.0)
7 (B3)	8/10 (80.0)	1/10 (10.0)	3/10 (30.0)	0/10 (0.0)	12/40 (30.0)
8 (B4)	8/10 (80.0)	6/10 (60.0)	6/10 (60.0)	6/10 (60.0)	26/40 (65.0)
9 (C1)	3/10 (30.0)	2/10 (20.0)	4/10 (40.0)	0/10 (0.0)	9/40 (22.5)
10 (C2)	5/10 (50.0)	3/10 (30.0)	6/10 (60.0)	2/10 (20.0)	16/40 (40.0)
11 (C3)	2/10 (20.0)	2/10 (20.0)	1/10 (10.0)	1/10 (10.0)	6/40 (15.0)
12 (C4)	3/10 (30.0)	1/10 (10.0)	2/10 (20.0)	1/10 (10.0)	7/40 (17.5)
13 (D1)	5/10 (50.0)	2/10 (20.0)	2/10 (20.0)	4/10 (40.0)	13/40 (32.5)
14 (D2)	9/10 (90.0)	1/10 (10.0)	5/10 (50.0)	2/10 (20.0)	17/40 (42.5)
15 (D3)	2/10 (20.0)	0/10 (0.0)	4/10 (40.0)	1/10 (10.0)	7/40 (17.5)
16 (D4)	3/10 (30.0)	2/10 (20.0)	0/10 (0.0)	1/10 (10.0)	6/40 (15.0)
17 (E1)	2/9 (22.2)	1/10 (10.0)	1/10 (10.0)	0/9 (0.0)	4/38 (10.5)
18 (E2)	6/10 (60.0)	2/10 (20.0)	5/10 (50.0)	2/10 (20.0)	15/40 (37.5)
19 (E3)	3/10 (30.0)	1/10 (10.0)	5/10 (50.0)	4/10 (40.0)	13/40 (32.5)
20 (E4)	8/10 (80.0)	2/10 (20.0)	4/10 (40.0)	2/10 (20.0)	16/40 (40.0)
21 (F1)	5/10 (50.0)	2/10 (20.0)	2/10 (20.0)	4/10 (40.0)	13/40 (32.5)
22 (F2)	7/10 (70.0)	2/10 (20.0)	3/10 (30.0)	4/10 (40.0)	16/40 (40.0)
23 (F3)	5/10 (50.0)	2/10 (20.0)	3/10 (30.0)	2/10 (20.0)	12/40 (30.0)
24 (F4)	2/3 (66.7)	0/3 (0.0)	1/2 (50.0)	2/2 (100.0)	5/10 (50.0)
25 (G1)	5/10 (50.0)	3/10 (30.0)	5/10 (50.0)	1/10 (10.0)	14/40 (35.0)
26 (G2)	1/10 (10.0)	1/10 (10.0)	3/10 (30.0)	1/10 (10.0)	6/40 (15.0)
27 (G3)	6/10 (60.0)	1/10 (10.0)	3/10 (30.0)	6/10 (60.0)	16/40 (40.0)
28 (G4)	4/10 (40.0)	2/10 (20.0)	6/10 (60.0)	1/10 (10.0)	13/40 (32.5)
29 (H1)	6/10 (60.0)	2/10 (20.0)	4/10 (40.0)	3/10 (30.0)	15/40 (37.5)
30 (H2)	6/10 (60.0)	7/10 (70.0)	6/10 (60.0)	4/10 (40.0)	23/40 (57.5)
31 (H3)	2/10 (20.0)	3/10 (30.0)	2/10 (20.0)	1/10 (10.0)	8/40 (20.0)
32 (H4)	4/10 (40.0)	5/10 (50.0)	5/10 (50.0)	5/10 (50.0)	19/40 (47.5)
Total	147/312 (47.1)	67/313 (21.4)	113/310 (36.5)	72/311 (23.2)	399/1246 (32.0)

Appendix VI Correct Sentences by Working Memory Load Type and Load Condition

Balanced cumulative number of fluent-correct sentential pairs between Load and No-load condition within speakers: (a) L1 and (b) L2.

(a) L1

	Verbal		Spatial		
Speaker	Load	No-load	Load	No-load	Total
1	13	13	8	8	42
2	11	11	11	11	44
3	12	12	10	10	44
4	8	8	11	11	38
5	9	9	12	12	42
6	10	10	11	11	42
7	8	8	12	12	40
8	4	4	10	10	28
9	2	2	5	5	14
10	7	7	13	13	40
11	12	12	8	8	40
12	7	7	6	6	26
13	11	11	10	10	42
14	5	5	8	8	26
15	5	5	11	11	32
16	10	10	8	8	36
17	5	5	4	4	18
18	12	12	12	12	48
19	9	9	10	10	38
Total	160	160	180	180	680

(b) L2

	Verbal		Spatial		
Speaker	Load	No-load	Load	No-load	Total
1	5	5	8	8	26
2	5	5	9	9	28
3	4	4	10	10	28
4	9	9	7	7	32
5	6	6	9	9	30
6	7	7	10	10	34
7	6	6	7	7	26
8	9	9	11	11	40
9	5	5	2	2	14
10	10	10	8	8	36
11	6	6	7	7	26
12	10	10	11	11	42
13	8	8	11	11	38
14	6	6	7	7	26
15	6	6	11	11	34
16	7	7	3	3	20
17	10	10	6	6	32
18	7	7	8	8	30
19	6	6	8	8	28
20	8	8	7	7	30
Total	140	140	160	160	600

ABSTRACT IN KOREAN

국문초록

본 논문은 멀티태스킹을 하면서 말을 할 때 말소리가 어떻게 변화하는지 살펴봄으로써, 작업기억이 발화의 말소리 생성에 관여하는지를 모국어와 외국어로서의 영어 발화를 비교하여 연구하였다. 이는 선행 연구에서 발화시 말소리 생성이 두뇌의 작업기억 시스템을 통해 계획되고 처리된다고 설명함에도 불구하고, 직접적인 실증 근거가 없고 간접적인 자료들이 반증한다는 점에서 출발하였다. 우선, 인지적 요인들이 리듬과 억양으로 정의되는 운율의 설명되지 않는 변이와 관련이 있을 것이라는 추측이 제기되어 왔지만, 이런 인지와 발화 사이의 연관성은 여전히 추측일 뿐 실증적 자료는 아직 제시되지 않았다. 게다가, 심리언어학에서 널리 수용되는 발화 모델들에서, 소리 정보의 일시적 저장 및 능동적 조작이 작업기억을 통해 처리된다고 설명하고 있지만, 역시 직접적 실험 근거가 없고, 특히 음운-음성적 발화 단계에 대한 해당 이론의 설명이 부분적일 뿐임을 인정하였고, 장기기억에 저장된 조음 뭉치를 그대로 출력해서 사용한다는 설명을 혼합함으로써, 작업기억이 필수적인지 의문을 남겼다. 아울러, 기억 이론 관점에서의 선행연구 리뷰 역시 작업기억이 말소리 생성에는 관여하지 않을지도 모른다는 잠정적 결론에 이르렀다. 마지막으로, 만약 말소리 생성이 작업기억을 통해 이루어지는지 아닌지 결정 여부가, 실시간으로 능동적으로 계획하며 만들어 발음하는 것인지 아니면 이미 저장된 발음 루틴을 습관적이고 반사적으로 출력하는 것인지에 달려 있다면, 원어민들과 외국어 학습자들이 다른 발화 처리 과정을 거칠지도 모른다는 추측을 하게 되었다.

따라서, 본 연구는 영어 발화의 말소리 생성에 작업기억이 연관되어 있는지, 그리고 영어 모국어 화자와 외국어 학습자가 다른 패턴을 보이는 지의 두 질문을 직접적으로 실험하였다. 미국인들과 한국인들이 각각 의미와 통사형태적으로 친숙해진 영어 문장들을 언어 혹은 공간 과제를 동시에 수행하는 멀티태스킹 상황에서 소리내어 말했다. 이들은 또한 같은 영어 문장들을 오직 말만 하는 비-멀티태스킹 상황에서 대조군으로 발화했다.

실험 결과 미국인들과 한국인들은 작업기억에 부과되는 인지 부하에 따라 다른 발화 패턴을 보였다. 우선 미국인과 한국인 모두 말을 하면서 멀티태스킹을 하게 되면 말실수를 더 많이 하고 더 빨리 말했다. 미국인들은 말을 하면서 멀티태스킹한 과제가 언어 과제이든 공간 과제이든 상관없이 같은 말소리 패턴을 보였으며, 측정한 그 외 모든 운율적 특성에서 말만 할 때와 통계적으로 동일했다. 그러나 대조적으로, 한국인들은 말하면서 언어 과제를 동시에 했을 때 발음에 큰 어려움을 보였다. 언어 과제를 할 때는 말만 할 때 뿐만 아니라 공간 과제를 할 때와 비교해서도 말실수가

더 많았고 운율이 평소와 더 많이 달라졌다. 공간 과제와 말을 멀티태스킹할 때는 앞서 말한 더 많은 말실수와 더 빠른 말속도 외에 다른 어떤 말소리의 차이를 보이지 않은 반면, 언어 과제와 말을 멀티태스킹할 때는 말 실수도 훨씬 더 많이 하고, 훨씬 더 빨리 말하고, 명사나 동사와 관사 등의 단어들 길이가 더 비슷해지고, 모음을 좀 더 웅얼거리며 발음했다.

본 논문은 이 실험 결과를 다음과 같이 해석한다. 작업기억은 말소리 생성에 관여할 수도 하지 않을 수도 있다. 이는 화자가 해당 언어에 대해 가지는 언어 경험과 훈련의 양과 그에 따라 장기 기억에 저장된 정보의 양에 따라 결정된다. 이는 해당 실험에서는 이분법적으로 나타났으나 실제로는 점진적인 선형 관계일 수도 있다. 이분법적으로 설명하자면, 모국어 성인 화자는 다년간의 언어 경험과 발음 훈련을 통해 이미 단어 형태와 조음 정보가 쌍으로 엮여 장기 기억에 루틴화된 템플릿 형태로 존재함으로써, 실제 발음을 할 때는 시각적 혹은 개념적으로 주어진 단어 형태를 실시간으로 조작하여 발음을 계획하고 시행하지 않는다. 즉각적이고 자동적으로 장기기억에서 저장된 형태-조음을 직접 출력해서 반사적으로 발음하게 되기 때문에, 작업기억을 통한 정보 처리는 이루어지지 않는다. 반면, 외국어로 언어를 배우는 학습자는, 그 언어 경험이 상당히 제한적이고, 특히 많은 경우 말하기 경험에서 더욱 그러하기에, 언어 형태와 조음 정보가 단단하게 장기 기억에 묶음으로 저장되어 있지도 않고 조음기관 훈련도 반사적으로 즉각적으로 이루어질 만큼은 아닐 확률이 높다. 따라서 할 말이 이미 의미, 문법, 단어, 형태적으로 모두 준비된 상태라 하더라도, 이를 발음하는 일 조차 실시간으로 계획되고 조작된다. 작업기억이 필요한 것이다.

이를 바탕으로 본 논문은 두뇌 기억 시스템의 작업기억과 장기기억과 관련하여 모국어와 외국어 말소리 생성에 대한 모델을 제시한다. 또한, 오랫동안 실증적 자료 없이 상반되는 추측에 그친 작업기억과 발화 말소리 생성의 관계를 직접적인 실험을 통해 검증을 시도했다는 점, 언어 운율에 대한 변이 중 멀티태스킹 시 발생할 수 있는 운율 변화를 인지와 연결하여 설명한 점, 상반된 선행연구의 추측과 논의들이 실제로는 해당 언어에 대한 경험과 훈련에 바탕을 두면 일관적으로 설명될 수 있음을 제안한 점, 그리고 이것이 모국어 대 외국어 혹은 나아가 모국어 어린이와 외국어 초급 학습자 대 모국어 성인과 외국어 숙련자로 확대하여 말소리 생성의 정보 처리 모델 및 교육적 함의를 제시한다는 점 등에서, 본 논문은 의의를 갖는다.

주요어 : 작업기억, 작업기억 능력, 발화, 운율, 음운적-음성적 부호화, 검색, 모국어 대 외국어

학 번 : 2006-30916