CROSS-MODAL REDUCTION: REPETITION OF WORDS AND GESTURES

by

PRAKAIWAN VAJRABHAYA

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2016

DISSERTATION APPROVAL PAGE

Student: Prakaiwan Vajrabhaya

Title: Cross-Modal Reduction: Repetition of Words and Gestures

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Eric Pederson	Chairperson
Vsevolod Kapatsinski	Core Member
Melissa Baese-Berk	Core Member
Lucien Brown	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2016

DISSERTATION ABSTRACT

Prakaiwan Vajrabhaya

Doctor of Philosophy

Department of Linguistics

December 2016

Title: Cross-Modal Reduction: Repetition of Words and Gestures

This dissertation examines speakers' production of speech and representational gesture. It utilizes the Repetition Effect as the investigative tool. The Repetition Effect appears to vary by the tendency for some items to shorten when repeating, at least under the condition that speakers can primarily operate by their assumption of the state of knowledge of the listener.

In speech, a highly conventionalized form of performance, word duration reduces within the same stretch of coherent discourse; then, it resets in the first mention of a new stretch of coherent discourse regardless of the state of knowledge to the speaker or the listener. Therefore, the Repetition Effect in speech is best analyzed as an automatic behavior triggered by discourse structure, rather than reflecting online changes in word accessibility for either interlocutor, be it for the speaker (Listener-neutral explanation) or for the listener (Listener-modeling explanation). The Repetition Effect in speech production in this dissertation will be accounted for within an exemplar model of the perception/production loop.

However, in representational gestures, a much less conventionalized form of performance compared to speech, the Repetition Effect shows a different pattern. When speakers only operate by their assumption of the state of knowledge of the listener,

without dynamic, appreciable listener feedback, they steadily reduce most types of representational gesture across tellings. Based on these results, it can be argued that representational gestures primarily serve as a part of speech production, rather than as communicative acts. That is, they are produced without regard to the novelty of the information to the listener, thus, consistent with the Listener-neutral explanation.

CURRICULUM VITAE

NAME OF AUTHOR: Prakaiwan Vajrabhaya

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, USA
Old Dominion University, Norfolk, USA
Mahidol University, Bangkok, Thailand

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2016, University of Oregon
Master of Arts, Applied Linguistics, 2007, Old Dominion University
Master of Public Administration, 2005, Old Dominion University
Bachelor of Arts, Travel Industry Management, 2002, Mahidol University

AREAS OF SPECIAL INTEREST:

Psycholinguistics, co-speech gesture, and semantics

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, American English Institute, 2010-2016
University of Oregon, Eugene

Graduate Teaching Fellow, Department of Linguistics, 2009-2010
University of Oregon, Eugene

GRANTS, AWARDS, AND HONORS:

Graduate Teaching Fellowship, University of Oregon, 2009-2015

Travel Grants, University of Oregon, 2009-2016

The Jaqueline Schachter Award, University of Oregon, 2015

Student Transportation Subsidy, The Acoustical Society of America, 2013

Southeast Asian Studies Award, University of Oregon, 2011

ISGS Summer School Scholarship, The Volkswagen Foundation, 2010

Promising Scholars Award, University of Oregon, 2009

Tuition Grant, Old Dominion University, 2006

Graduate Teaching Assistantship, Old Dominion University, 2005-2006

PUBLICATIONS:

Vajrabhaya, P., & Kapatsinski ,V. (under review). First time's the charm: First-mention lengthening as automatic behavior.

Vajrabhaya, P., & Kapatsinski, V. (2013). Modeling the listener? What resets acoustic durations of repeated English words. *The Journal of the Acoustical Society of America*, *134*(5), 4203-4203.

Vajrabhaya, P., & Kapatsinski, V. (2011). There is more to the story: First-mention lengthening in Thai interactive discourse. In *Proceedings of the International Congress of Phonetic Sciences* (Vol. 17, pp. 2050-2053).

Vajrabhaya, P. & Pederson, E. (under review). Teasing apart listener-sensitivity: The role of interaction.

Vajrabhaya, P., & Pederson, E. (2016, July). *Teasing apart listener-sensitivity: The role of interaction.* In E. Pederson & Vajrabhaya, P., Gesture, gesture: Tracking gestures across repetitions. Symposium conducted at the meeting of the International Society for Gesture Studies, Paris, France.

Vajrabhaya, P., & Pederson, E. (2015). "The baking stick thing": Automatization of co-speech gesture during lexical access. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.) (2015). *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (p. 3007)*. Austin, TX: Cognitive Science Society.

Vajrabhaya, P., & Pederson, E. (2014). *When anyone's listening: Automatization and gesture reduction*. Talk given at the meeting of the International Society for Gesture Studies, San Diego, CA.

Vajrabhaya, P., & Pederson, E. (2013, October). Repetition vs. listener accommodation: A case study of co-speech gesture in retellings. Poster presented at the meeting of New Ways of Analyzing Variation, Pittsburgh, PA.

Vajrabhaya, P., & Pederson, E. (2012, July). *Have I told you this before?: Reduction in magnitude over repetition in co-speech gesture*. Poster presented at the meeting of International Society for Gesture Studies, Lund, Sweden.

ACKNOWLEDGMENTS

I could not have written this dissertation without guidance from Dr. Eric Pederson, my advisor. He has gone above and beyond in making me a better researcher, shaping the way I ask questions and the way I should go about answering those questions. I am forever in debt to him for his patience, kindness, and dedication. I will miss his one-of-a-kind sense of humor, an important factor that kept me afloat during the long process of dissertating. Although he might not admit it, he is, without a doubt, a superhero in disguise.

Dr. Vsevolod Kapatsinski became my mentor and a good friend from the very first day in graduate school. He encouraged new ideas that eventually became a chapter of this dissertation. In addition, I would like to thank Dr. Melissa Baese-Berk, Dr. Lisa Redford, and Dr. Lucien Brown for their time and insightful comments on multiple drafts of this dissertation. I would also like to thank Dr. Tyler Kendall for allowing me to use the Language Variation and Computation Lab to run experiments.

The American English Institute (AEI) provided me with financial support throughout the years; I am very grateful for it. While employed at the AEI, Dr. Tom Delaney, my supervisor, never hesitated to work with my busy academic schedule: I thank him for his understanding. The Department of Linguistics provided generous funding, which allowed me to travel to conferences, both domestically and internationally. I would like to thank Linda Campbell and Meg Taggart who assisted me with the paperwork and constantly reminded me to smile.

I thank my friends for putting up with my complaining about life in general and often times (un)necessary anger towards the data. I could not have survived this process without their wit, humor, and emotional support: Alex Abrams, Ted Bell, Stefano Cazzago, Wook Kyung Choe, Palita Chunsaengchan, Alan Cienki, Chris Doty, Ellen Gillooly-Kress, Charlie Farrington, Paul Harvey, Adam D.I. Kramer, Anne Mannering, Matilda Mannering-Bell, Jason McLarty, Keaton Miller, Sorakom Ngamdecho, Unchalee Pattarakijsophon, Ukrit Petpichetchian, Brandon Reich, Silavat Somsiri, Shahar Shirtz, Joanne Scheibman, Nate Severence, David Silvis, Matt Stave, Manuel Widmer, Thai students in Eugene, research assistants who worked with me, and everyone who regularly attended the Cognitive Linguistics Workgroup. Importantly, I would like to thank Amos Teo for his willingness to listen to anything and everything I said in any language, much of which did not make much sense, especially towards the end of the dissertating process. Amos, you are a true friend. I am thankful for the music performed by Joshua Bell, Evgeny Kissin, Mischa Maisky, Itzhak Perlman, and Arthur Rubinstein. Their music kept me company throughout the writing process.

Lastly, I thank my family for their love and their understanding. They have always understood why it is important for me to pursue this degree. I may not have been home much, but I felt like I was home. They made sure that I ate right, slept well, and laughed plenty. Somehow, they were able to accomplish that all the way from Bangkok. I thank you: mom, dad, and Tat.

For H.M. King Bhumibol Adulyadej, in memoriam

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION AND REVIEW OF THE LITERATURE

1. *Introduction*

When we enter a conversation, we enter with a set of presuppositions about the world, and importantly, about the listener. The presupposition we make about our listener, i.e., the evaluation we make about who they are, what they already know, and the context in which the conversation takes place, influences the production we make. The way I describe my research to others depends on my evaluation of the audience. Although I am essentially talking about the same general topic, what my work is about, the way I explain it to my parents and the way I explain it to colleagues at a conference are vastly distinctive. The assumption I make about how much my parents already know about psycholinguistics and the assumption I make about how much my colleagues already know about psycholinguistics drive the overall structure of the conversation; it drives my decision of what to include or exclude from the conversation and how the information should be presented. This is a simple example of large scale planning, by way of modeling the listener's state of knowledge, and how it plays a role in speech production.

Speakers also show listener-sensitivity via listener modeling in local execution. Production choices, such as lexical item and grammatical item, can reflect speakers' assumption of the state of knowledge of the listener. For instance, speakers' choice to refer to a document as a *W-2* as opposed to a *Wage and Tax Statement* shows an assumption that the listener has prior knowledge of the document. In grammar, when speakers use a pronoun to refer to a referent (e.g., *she* vs. *Pat*), it is a reflection of the

speakers' assumption of the state of knowledge of the listener, the assumption that the listener knows *she* and *Pat* are the same person. In addition to lexical choices and grammar, speakers may show listener-sensitivity via word duration. That is, word duration may vary depending on the assumption speakers make about what the listener knows. In sum, speakers can show their sensitivity to the listener in large scale planning or small scale planning.

In addition, it is evident that speakers of various ages, genders, and cultures also produce representational gestures in concert with speech; as a result, there has been an academic movement that is dedicated to the study of representational gestures in relations to speech in recent decades. Like speech, past literature have suggested that speakers' production of representational gestures are also a product of speakers' modeling the state of knowledge of the listener. Briefly speaking, the way people produce representational gestures differs based on their assumption of what the listener has already heard or already knows. Although speaking and gesturing are of two distinct modalities, they appear to demonstrate that speakers take into consideration the state of knowledge of the listener.

In recognition that the two modalities go hand-in-hand, the overarching goal of this dissertation is to examine speakers' production holistically, investigating both speech production and representational gesture production via the same data. By doing so, it allows documentation of both types of production under a specific condition, which highlights the similarities and differences between them.

Importantly, this dissertation bridges a gap in the literature. Literature in both speech production and representational gesture production that argue for listener-

sensitivity via modeling the state of knowledge of the listener have not provided sufficient control in the experiments. Examining the literature, most studies have allowed interlocutors to interact freely, making it problematic to determine the source of listener-sensitivity, namely, whether listener-sensitivity arises from modeling the state of knowledge of the listener (i.e., purely a result of the speakers' assumption) or listener-sensitivity arises from an on-going interaction between interlocutors. Essentially, this dissertation tests the notion of listener-modeling, to see whether the speakers' modeling of the listener without appreciable input from the listener will affect speaker production. Specifically, do speakers emphasize or de-emphasize productions of material presumed to be novel or familiar to the listener?

This dissertation utilizes the Repetition Effect as the investigative tool. I will demonstrate that the Repetition Effect is specific to modality, particularly under the condition that speakers can only operate by their assumption of the state of knowledge of the listener. In speech, a highly conventionalized form of performance, word duration reduces within the same stretch of coherent discourse; then, it resets in the first mention of a new stretch of coherent discourse regardless of the state of knowledge to the speaker or the listener. Therefore, the Repetition Effect in speech is best analyzed as an automatic behavior triggered by discourse structure, rather than reflecting online changes in word accessibility for either interlocutor, be it for the speaker (Listener-neutral explanation) or for the listener (Listener-modeling explanation). The Repetition Effect in speech production in this dissertation will be accounted for within an exemplar model of the perception/production loop.

However, in representational gestures, a much less conventionalized form of performance compared to speech, the Repetition Effect shows a different pattern. Under the condition that speakers can only operate by their assumption of the state of knowledge of the listener, without dynamic, appreciable listener feedback, speakers steadily reduce most types of representational gesture across tellings. Based on these results, it can be argued that most types representational gestures primarily serve as a part of speech production, rather than a communicative act. That is, they are produced without regard to the novelty of the information to the listener, thus, consistent with the Listener-neutral explanation.

In this chapter, multiple bodies of literature across modalities will be grouped by models posited by various authors. Section two of this chapter will be dedicated to grounding and defining key terms used throughout the dissertation. Section three will discuss literature on the Listener-neutral explanation; section four will discuss literature on the Listener-modeling explanation. Within the discussion of each explanation, literature on speech will precede literature on representational gesture. The literature will be primarily situated in studies that utilize the repetition paradigm, which results in the Repetition Effect, to examine speakers' production.

In section five, I will highlight problematic interpretations of the Repetition Effect that are anchored in the notion of information status (new vs. old/given) and problematic interpretations of the Repetition Effect that are accounted for with the Listener-modeling explanation. I will, then, propose an experimental method that eliminates a confound that exists in past literature.

2. *Definitions*

To avoid confusion of terminology, this section is dedicated to defining key terms that are indispensible to the dissertation. Although there may be terms in previous literature that are similar or overlapping with the definitions below, the dissertation will be operationalized by the definitions stated in this section.

2.1. *Definition: A Listener-neutral explanation*

A *Listener-neutral explanation* is defined as any online processing explanation of speakers' production that does *not* explicitly indicate an inherent reference to the listener. In this explanation, speakers' production is taken to be driven by internal cognitive processes; however, speakers' internal cognitive processes may not the only component that influences production. Indeed, speakers may be taking the listener into account (modeling of the listener) under this explanation; however, this explanation takes internal cognitive processes as the primary component that influences production; other factors are considered secondary.

In addition, this definition has a limited scope; it only accounts for specific aspects of speakers' production, i.e., duration of a repeated word and motoric commitment and effort used in a repeated representational gesture. In sum, this definition is not intended be extended beyond the current scope to account for other aspects of speaker production. For instance, it should not be construed to account for speakers' choice of a repeated lexical item.

2.2. *Definition: Common ground*

The term *common ground* is often used to explain speakers' sensitivity to the listener, an audience design of sort (e.g., Clark & Carlson, 1981; Clark & Haviland, 1977;

Clark & Marshall, 2002; Clark, Schreuder, & Buttrick, 1983; Stalnaker, 2002; Stalnaker, 1978). According to Clark (1992), common ground is defined as "the sum of [interlocutors'] mutual knowledge, mutual beliefs, and mutual suppositions" (p. 3). The term is widely used in linguistics to demonstrate that speakers' linguistic choices are made based on common ground they believe is shared with the listener; for example, Clark & Wilkes-Gibbs (1986) reported that speakers switch to names rather than a description of a referent when speakers believe that common ground is shared between them and the listener (see also Brennan & Clark, 1996 on *lexical entrainment* and *conceptual pact* and Lockridge & Brennan, 2002 on syntactic choices based on common ground). Although common ground has been widely used in studies of spoken language, it has been applied to studies in representational gesture in recent years (e.g., Gerwing & Bavelas, 2004; Holler, 2009; Holler & Stevens, 2007; Holler & Wilkin, 2009) to explain variation in production of representational gesture as a function of speakers' sensitivity to the listener.

Horton & Kaysar (1995), however, raised an important point that the classic definition of common ground does not clearly and sufficiently indicate when common ground comes into play in language production; in their words "we do know at what point in the utterance construction the production system incorporates the speaker's knowledge of common ground. We do not know whether common ground guides the initial planning utterances or whether the incorporation of common ground occurs rather late in the process" (p. 93). Because of the uncertainty in the timing common ground interplays with production, Horton & Kaysar (1995) conducted an experiment which imposed a time constraint, a way to increase a cognitive load; they showed that when

there is no time constraint, speakers operate by the *Initial Design* model, which entails that speakers employ common ground in the planning stage of an utterance; however, when a time constraint is imposed, speakers operate by the *Monitoring and Adjustment* model, which entails that speakers do not employ common ground in the planning stage of the utterance and that sensitivity to the state of knowledge of the listener is absent. In this model, linguistic output is likely post hoc rather than ad hoc; the output may constantly change, via monitoring, over the course of speech production. Based on Horton & Kaysar's (1995) work, it could be inferred that common ground may only occur in a specific context and perhaps should not be treated as necessarily involved when interlocutors share the same information with the speaker.

For the purpose of grounding the term in this dissertation, common ground will be strictly defined as a form of listener modeling by the speaker; that is, it refers to speakers' belief about what knowledge is shared between them and the listener *prior to* the current production (the Initial Design model). Such belief leads to the assumption and evaluation they make about what the listener already knows, which drives their choices during production (see section 2.3 in this chapter).

2.3. *Definition: A Listener-modeling explanation*

A *Listener-modeling explanation* is defined as any processing explanation of speakers' production that explicitly indicates an inherent reference to the listener. In this explanation, speakers' production is assumed to be influenced by their sensitivity to the listener. As a result of the speakers' assumption of the state of knowledge of the listener, they may modify or adapt their production in order to fulfill communicative needs for the listener. Such assumption of the state of knowledge of the listener is what the speakers

possess prior to engaging in an interaction with the listener. The definition is closely tied to the notion of common ground, as defined in section 2.2 in this chapter. It is important to emphasize that, as contradictory to the name as it may seem, the Listener-modeling explanation is nevertheless, very much about the speaker's internal cognitive processes. This explanation is not concerned with what speakers might do in response to listener behavior, nor is it about the interaction between interlocutors.

2.4. *Definition: A Listener-interactive explanation*

A *Listener-interactive explanation* is defined as any processing explanation of speakers' production that explicitly references the listener in response to listener's verbal and/or non-verbal behavior. In contrast to the Listener-modeling explanation, the Listener-interactive explanation is grounded in the dynamic and rapid exchanges of verbal and non-verbal cues, which interlocutors exchange and negotiate throughout the course of an active interaction. Importantly, listener-interaction is distinct from the speakers' assumption of the listener's state of knowledge, which the speakers hold prior to the current interaction (in contrast to the Listener-modeling explanation, see section 2.3 in this chapter).

Several bodies of literature across disciplines situate themselves in this explanation. Two well-recognized works by Clark & Schaefer (1989) and Goodwin (1981) discussed in depth how interlocutors manage a conversation through exchanging linguistic and non-linguistic means; for example: how interlocutors respond to one another utilizing a variety of linguistic structure, how gaze affects turn-taking, and how nodding plays a role in an interaction. In this explanation, speech, gesture, and bodily movements are factored into the mechanism that drives production.

2.4. *Definition: The Repetition Effect*

In speech, words vary in duration. Duration of a word may vary depending on whether it is being said for the first time (initial mention) or it is being repeated (subsequent mention). The phenomenon has been well-documented in speech literature (e.g., Anderson & Howarth, 2002; Aylett & Turk, 2006 and 2004; Baker & Bradlow, 2009; Bard & Aylette, 2005; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Fisher & Tokura; 1995; Freeman, 2014; Fowler, 1988; Fowler, Levy, & Brown, 1997; Fowler & Housum, 1987; Galati & Brennan, 2010; Lam & Watson, 2010; Pluymaekers, Ernestus, & Baayen, 2005; Vajrabhaya & Kapatsinski, 2011). Similarly, representational gestures have been found to vary in multiple aspects, depending on whether it is being performed for the first time or is being performed in repetition (e.g., Galati & Brennan, 2013; Gerwing & Bavelas, 2004; Hoetjes, Koolen, Goodbeek, Krahmer, & Swerts, 2011; Jacobs & Garnham, 2007; Masson-Carro, Goudbeek, & Krahmer, 2014).

In this dissertation, the phenomenon will be referred to as the *Repetition Effect*. Despite the mounting evidence for the Repetition Effect in both modalities, the reason why it exists remains elusive. Linguists and gesture researchers have long been interested in the Repetition Effect, for it provides insight into processes in speech production and processes in production of representational gesture (see section 3.2.2 and section 4.1 in this chapter).

2.5. *Definition: Representational gesture*

Despite the fact that gestures are tightly integrated with speech and are virtually impossible to suppress during speech production (Kendon, 2004; Kendon, 1994; McNeill, 2005; McNeill, 1992), the reason why people gesture, however, remains

elusive. Unlike spoken language, which at least synchronically-speaking, has a stable bond between form and meaning, the act of gesturing varies in the degree of conventionalization. "Kendon's continuum" (McNeill, 2005; McNeill, 1992) categorizes conventionality into four stages. Figure 1 is a replica of McNeill's (2005) continuum of conventionalization (p. 4).

Figure 1: McNeill's (2005) Continuum 3: relationship to conventions

Gesticulation not conventionalized	→	Pantomime ditto	→	Emblems partly conventionalized	→	Sign Language fully conventionalized

Although all four acts of gesturing on the continuum involve the use of hands and forearms, they vary significantly in the magnitude of conventionalization.

The leftmost on the continuum is *gesticulation*, gestures that are performed along side speech, where the "lack of convention is an attribute *sine qua non*" (p. 5, emphasis is original). That is, there is no standard form and meaning counterpart between gesturing and speech; for example, there is no standardized hand shape for COFFEE in gesticulation. In contrast, the rightmost on the continuum is sign language, where full-fledged conventionalized gestures are used to express meaning. For instance, in American Sign Language, COFFEE is gestured with two hands that are formed in a fist-like shape, moving in a circular motion. People who have knowledge of American Sign

Language can encode and decode COFFEE by using such gesture (hand sign) due to its highly conventionalized form and meaning pairing.

In this dissertation, the only kind of gesturing that is of interest is gesticulation, the least conventionalized kind of gesturing. The term *representational gesture*, which is a kind of iconic gesticulation is defined by McNeill (2005) as "[presenting] images of concrete entities and/or actions […] its manner of execution embodies picturable aspects of semantic content (aspects of which are also present in speech)" (p. 39) will be used in this dissertation. Unless stated otherwise, all references to gestures in this dissertation refer to representational gesture.

Figure 2 exemplifies a representational gesture. In this example, while the speaker says *the rabbit is hopping out of the room,* she also raises her slightly bent index and middle finger and performs a hopping away motion from her body.

Figure 2: An example of a representational gesture

This is a representational gesture since the hand shape represents the rabbit, the arch motion represents the rabbit hopping, and the away from the body trajectory represents the direction the rabbit is going. In short, the entity (the rabbit), the motion (hopping), and the trajectory (out of the room) are presented in speech and representational gesture, coinciding with McNeill's definition. Note that the semantic content of the representational gesture is unfolded when it is accompanied by speech, in contrast to hand signs which do not require speech for the semantic content of the message to unfold. In short, a representational gesture per se does not carry a concrete semantic content; speech is obligatory.

Although representational gestures have been found to bear an informative value when accompanying speech in the sense that listeners can gain a better comprehension of the message (Alibali, Flevares, & Goldin-Meadow, 1997; Beattie & Shovelton, 2005; Singer & Goldin-Meadow, 2005), Hostetter & Alibali (2008) have suggested that the communicative value of representational gesture, perhaps, should not be taken at face value. Listeners resorting to a representational gesture's semantic content may be largely dependent on the lack of redundancy of the message conveyed. In other words, listeners may make use of gestural information when the information in speech, the primary mode of communication, is absent or lacking. Whether representational gesture encodes semantic content for the listener or how much information the listener decodes from gesture is beyond the scope of the dissertation. The dissertation will only be concerned with the production aspect.

3. *The Listener-neutral explanation in repetition paradigm studies*

In this section, literature on the Repetition Effect in speech and in gesture that coincides with the Listener-neutral explanation will be discussed. Theories discussed in this section may differ in their components; however, they are unified with one element: these explanations are grounded in the framework that the Repetition Effect is driven by the speakers' production system and does not have an inherent reference to the listener (see a full definition, section 2.1 in this chapter).

3.1. *The Repetition Effect in speech*

Linguistically speaking, the Repetition Effect could result in a fewer number of words used (Clark & Wilkes-Gibbs, 1986), or it could result in subsequent mentions being shorter in duration when compared to the first mention of the same word. (e.g., Anderson & Howarth, 2002; Aylett & Turk, 2006 and 2004; Baker & Bradlow, 2009; Bell et al., 2009; Fisher & Tokura; 1995; Freeman, 2014; Fowler, 1988; Fowler et al., 1997; Fowler & Housum, 1987; Galati & Brennan, 2010; Lam & Watson, 2010; Pluymaekers et al., 2005; Vajrabhaya & Kapatsinski, 2011). Several prominent explanations of the Repetition Effect in speech that coincide with the Listener-neutral model can be generally categorized into three areas: (1) automatization in articulatory planning; (2) speed of lexical access; and (3) bidirectional adjustment prosodic prominence.

3.1.1. *The Repetition Effect in speech as a result of automatization in articulatory planning*

To reiterate, when a motoric action is repeated, it reduces. Speech, being a kind of motoric action, operates by this concept. Bybee (2002a, 2001) proposed that repetition

causes automatization in articulatory sequences, thereby causing reduction. Simply put, words are automatically reduced when they are repeated due to the characteristics of the human's speech production system. Consistent with Bybee's proposal, Kapatsinski (2010) believes that upcoming phonological gestural targets of a repeated word (a word that has been practiced) are anticipated and activated faster as the word is being executed (relative to preceding phonological gestural targets within the same word). The rapid activation of anticipated phonological gestural targets (after word execution has started) results in shorter word duration (see also Dell, Burger, & Svec, 1997).

Figure 3: Repetition priming

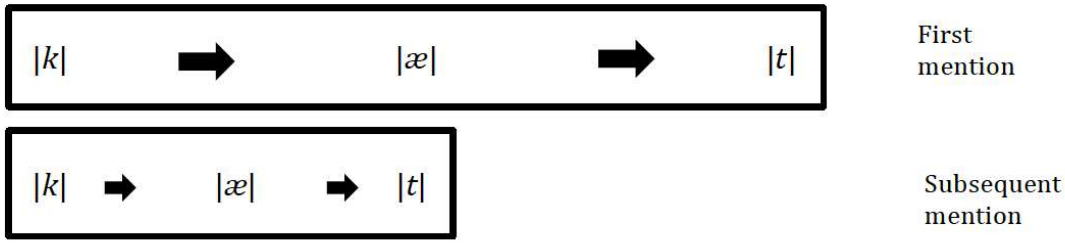


Figure 3 is a simplified demonstration of Kapatsinski's (2010) explanation of repetition priming, which results in hypoarticulation in subsequent mention. In this explanation, once a word has been produced (in this case, the first mention), the word, as well as its segmental phonetic details, becomes activated. The activation lessens pressure of having to retrieve segments within the word; therefore, speakers can look-ahead at a faster rate to the next segment, resulting in a more rapid execution of motoric production,

and essentially leading to hypoarticulation. Therefore, it is repetition priming that is primarily responsible for the Repetition Effect.

3.1.2. *The Repetition Effect as a result of speed of lexical access*

Another explanation that is consistent with the Listener-neutral explanation is by Bell et al. (2009), who argued that it is not necessarily a reduction of word duration in subsequent mention (as explained in section 3.1.1.); rather it is a lexical access difficulty that speakers encounter when retrieving the word that results in lengthening the first mention. Bell et al.'s (2009) proposal is highly plausible; however, priming is known to affect infrequent words more than it affects frequent words (the frequency attenuation effect, Forster & Davis, 1984 et seq.). Over time, priming results in the effect of word frequency on lexical access speed being logarithmic rather than linear (e.g. Broadbent, 1967; Goldiamond & Hawkins, 1958; Howes & Solomon, 1951; Norris & McQueen, 2008; Oldfield & Wingfield, 1965).

Given the frequency attenuation effect, lexical access should be facilitated by mentioning a word to a greater extent when that word is infrequent. And since the frequency attenuation effect holds for both production and perception (frequency effects are logarithmic in both, cf. Howes & Solomon, 1951; Oldfield & Wingfield, 1965), it becomes problematic to interpret whether the Repetition Effect is caused by accessibility of the word to the speaker or accessibility of the word to the listener. Furthermore, since repetition should affect low-frequency words (a production that has not yet been automatized) more than it should affect high-frequency words (a production that has been automatized) (e.g. Kapatsinski, 2010) in accordance with the Power Law of Practice (Logan, 1988; Seibel 1963), it leaves room for an alternative explanation.

Although Bell et al. (2009) discussed first-mention lengthening being driven *specifically* by lexical access difficulties, it is worthwhile to note that word lengthening could be driven by *any* kind of processing difficulty, as long as that processing difficulty competes for resources with the speaker's articulatory planning or the listener's utterance comprehension processes. For example, Vajrabhaya & Kapatsinski (2011) conducted an interactive map-task in Thai, where the speaker directs the listener where each animal picture (target word) should be located on the board. Speakers told and retold two different "stories" involving these animal pictures to each listener, where each "story" consisted of the same set of animal pictures, but the pictures were arranged differently on the speaker's board. In other words, the *only* difference between stories was the arrangement of animal pictures. Results showed word duration to systematically lengthens in the first mention of each story, regardless of story or listener, which appears to coincide with Bell et al.'s (2009) proposal of first-mention lengthening, rather than subsequent-mention reduction.

It is possible that the greater length at first mention found in this study is due to a broader version of the processing difficulty. When an animal picture (target word) was first mentioned in a story, the speaker may have been in the midst of planning the spatial description of the picture in relation to other pictures on the board. It could be argued that speakers could delay production instead of lengthening the word while engaged in the process of planning the utterance. This, however, does not seem plausible since all the target words examined were in the medial position, thus, first-mention lengthening is likely to have derived from a processing difficulty. In addition, the listener may also needed time to locate the corresponding animal picture to match the speaker's spatial

description. Therefore, first-mention lengthening observed in this study could have been due to a task effect. In sum, any kind of processing difficulties could be influencing word duration, especially when it is completing with the speaker's articulatory planning or the listener's comprehension.

Furthermore, Fowler et al. (1997) reported first-mention lengthening and suggested that it is the result of a conventionalized discourse structure. The argument in their study is problematic since there exists a confound in their stimuli which elicit a heavier processing load in the first mention. In other words, it is difficult to interpret whether first-mention lengthening in their study was driven by discourse structure as they suggested or it was driven by difficulties in processing on the speaker's part. Because of this, first-mention lengthening observed in Fowler et al.'s data could have been a task effect, similar to in Vajrabhaya & Kapatsinski (2011), where the task demanded that speakers utilize higher cognitive processing in the first mention.

To elaborate on their work, in Experiment I, Fowler et al. had six American English speakers narrate an episode of a TV series to a listener they had just met. The narration was coded for "episode boundaries" using a variety of discourse and content cues to narrative discontinuity. They observed significant shortening *within* the resulting episodes but not across episode boundaries, concluding that "an episode boundary blocks shortening" (p.21).

In Experiment II, Fowler et al. asked 12 participants to read stories in which some paragraphs began with cues to a change of scene. Again, shortening was blocked when mentions were separated by an explicit cue to change of scene ("but in another scene") and was not blocked when the scenes formed a single narrative, the change marked by

“but weeks later”. Fowler et al. referred to the effect as blocking shortening rather than lengthening since they did not find significant lengthening across episode boundaries (p.21, fn.2), although they did observe numerical trends towards lengthening across episode boundaries.

Based on these results, they raised the possibility that the blocking of shortening across stories (first-mention lengthening in this case) as “the way you tell a story” in English. However, they discarded this explanation in favor of an online accessibility-based explanation. The “more fundamental” reason to reject the conventionalized explanation for them was that “Some recent findings of our own and of other researchers lead us to doubt that durational shortening and blocking of shortening are markers provided deliberately for listeners by speakers” (Fowler et al., 1997, p.38).

Fowler et al. (1997, p.38) also mentioned another “reason for doubting that our speakers used blocking of shortening as an optional or alternative marker of a discontinuity in a narrative is simply that it is not obvious why references to a [change in] film or scene should not have served the same purpose as the time/location markers”. To respond to their comment, it is possible that a change in time/location does not signal an end to one narrative/story and the beginning of another, while a change in the film being narrated or explicit switch of scene does.

Consider the excerpts from Fowler et al.’s (1997, p.33) stimuli in (1)-(2). In (1), the narrative continuity is entirely broken by the scene-changing “episode transition”: the description of the next scene has no logical connection to the preceding scene (see [Boundary inserted below]. In contrast, the narrative storyline continues fluidly through the time/location episode transition in (2). It is, therefore, suggested that Fowler et al.’s

data seem consistent with the proposal that story/narrative boundaries trigger the speaker to lengthen words when they are mentioned for the first time within the new story.

(1) …The old woman was very happy to see the boys and asked them to stay for tea. They couldn't stay, but [Boundary] in another scene, you see the nanny regaling Martin with stories about Aloysius when he was a little boy at Maidenhead.

(2) "…The old woman was very happy to see the boys and asked them to stay for tea. They couldn't stay, but [Boundary] weeks later, the nanny visited them and regaled Martin with stories about Aloysius when he was a little boy at Maidenhead.

A major limitation of Fowler et al.'s stimulus design is that first mention in their study is confounded with being located in the episode-initial sentence. While the first mention of a word within a story or episode does not, in general, occur in the first sentence of the story or episode, they do in Fowler et al.'s stimuli, as in (1)-(2). Episode-initial sentences have been found to elicit a higher processing load (e.g. Haberlandt, Berian, & Sandson, 1980; den Uyl & van Oostendorp, 1980); therefore, first-mention lengthening in their work could be an artifact of heavier processing load on the speaker's part and not necessarily driven by discourse structure; this is a similar concern to Vajrabhaya & Kapatsinski (2011).

3.1.3. *The Repetition Effect as a result of prosodic prominence*

One of the functions of the indefinite article *a*/*an* in English is to code a referent that is entering discourse, signifying that the referent is new. Once that referent has entered into discourse, the definite article *the* in English functions to code that such referent is old/given (See Du Bois, 1980; Givón, 1984, for example). In speech production, coding of information status can be found prosodically in addition to syntactic coding.

The notion of information status could be construed to be consistent with the Listener-modeling explanation (see section 4 in this chapter). In other words, speakers could be modeling the state of knowledge of the listener, treating a certain referent as new or old/given, by factoring in whether the listener has been exposed to the referent before vs. not. The decision for speakers to code a referent as new or old/given via grammar or prosody could be the result of the speakers' modeling of the state of knowledge of the listener. This dissertation, however, categorizes coding of information status under the Listener-neutral explanation, rather than the Listener-modeling explanation, for the following reasons.

Firstly, it is unclear *to whom* information status applies to; that is, whether the information is new to the speaker or whether the information is new to the listener. In the same direction, it is unclear whether the information is old/given to the speaker or the whether the information is old/given to the listener. Although studies in this section anchored themselves within the realm of information status, they do not explicitly address this particular confound between information status to the speaker vs. listener.

In addition, the definition of Listener-modeling (see section 2.3 in this chapter) ascribes that there must be an *explicit* inherent reference to the listener. Since studies in

this section do not make an explicit inherent reference to the listener, it is more conservative, in my opinion, to categorize such work to be consistent with the Listener-neutral explanation. In this dissertation, I will be operating under the premise that information status of a referent, new or old/given, applies to the speaker, not the listener due to lack thereof explicit inherent reference to the listener.

Fundamental frequency (f_0) has been argued to code whether a referent is new or old/given in discourse (Bolinger, 1972; Brown, 1983; Halliday 1967; Chafe 1976; Selkrik, 1996; Yule, 1980). Proponents of this view suggested that in the initial mention, when speakers introduce a new referent into discourse, a higher f_0 is used. In contrast, when the same information repeated in the same discourse, it becomes old/given information; hence, a lower f_0 is used. Furthermore, Swerts & Geluykens (1994) provided experimental evidence to support the relationship between pitch peaks, the maxima f_0, to coincide with newness of information status; on the other hand, when a referent becomes old/given information, for instance through repetition, lower f_0 is found (Sridhar, Nenkova, Narayanan, & Jurafsky, 2008).

In particular, Aylett & Turk (2004) proposed that the effect of repetition on duration is mediated by online assignment of prosodic prominence. Under this view, words mentioned for the first time undergo accentual lengthening by being assigned prosodic focus, which is marked by high f_0 and high intensity, as well as increased duration; indeed, conveying that the focused element is new information (in the first mention) has been argued to be a major function of prosodic focus. That is, new information, (the first mention) receives prosodic prominence; whereas, old/given

information (repeated mention) of the same referent, does not receive prosodic prominence.

The claim that f_0 is highly associated with information status is nevertheless, controversial. Fowler & Housum (1987) showed that repeated words, which are considered old/given, exhibit a significantly shorter duration and lower intensity; however, f_0 was not found to differ. Along the same line, Issacs & Watson (2010) reported that it is duration and intensity (loudness measured in decibel) that are reflected in new information, not f_0. Furthermore, Terken & Hirschberg (1994) report that high f_0 is not necessarily restricted to new information: old/given information that is in a certain syntactic role in a coherent stretch of discourse can also bear high f_0 just like new information. Hence, there is not a transparent relationship between high f_0 and newness information.

Work that has systematically examined the relations between duration, f_0 and intensity questions the notion that the three variables should be collapsed together under the notion of prosodic prominence. In particular, Baker & Bradlow (2009), Fowler & Housum (1987), Isaacs & Watson (2010), and Lam & Watson (2010) suggested that repetition most robustly affects duration, rather than f_0 or intensity. In particular, Fowler & Housum (1987) and Isaacs & Watson (2010) measured both f_0 and duration in the same tokens and found that first and subsequent mentions reliably differed in duration but not in f_0.

In addition, Baker & Bradlow (2009) asked naïve phoneticians who were blind to the first vs. second-mention condition to code their recordings using the ToBI prosodic annotation system. As predicted by Aylett & Turk (2004), Baker & Bradlow (2009)

found that second mentions, the repetition of the same word, were less likely to be accented than first mentions. However, a significant effect of mention remained when both first and second mentions were accented or both were unaccented. Similarly, Aylett & Turk (2004) found a significant effect of repetition on duration after controlling for prosodic factors. While Aylett & Turk (2004) claimed that this was likely due to imperfections in the prosodic coding, they acknowledged that it could also mean that the Repetition Effect on duration is not fully mediated by prosodic prominence. Based on this work, it is possible that repetition will affect f_0 in addition to duration, but it is very likely that duration would be affected to a greater extent.

Intensity is an additional cue to prosodic prominence. Lam & Watson (2010) found that repetition had a stronger effect on duration than on intensity (while predictability mainly influenced intensity and had negligible effects on duration). This finding, again, suggests that prosodic prominence, which is strongly correlated with high f_0 and intensity, is unlikely to fully mediate the effect of repetition on duration.

In sum, the Repetition Effect appears to be robustly reflected in word duration, but not necessarily in f_0. In other words, when a referent/word is new in discourse, presumably via first mention, its duration is consistently longer than its repeated counterparts. However, there does not appear to be a systematic pattern of f_0 as a function of information status. Perhaps, word duration is the primary cue in conveying information status in repeated referents, and f_0 is a secondary cue.

3.2. *The Listener-neutral explanation in representational gesture*

The primary function of speech is to communicate. The function of a representational gesture, however, remains debatable. This section will begin with

prominent models of representational gesture production that are compatible with the Listener-neutral explanation; then, it will be followed by recent work on the Repetition Effect in representational gesture that coincides with the Listener-neutral explanation.

3.2.1. *The Listener-neutral explanation: Models of representational gesture production*

In this view, representational gestures are performed to benefit the speaker cognitively for the purpose of speech planning; thus, the direct function of a representational gesture is *for the speaker*. One prima facie case that representational gestures are for the speaker is that congenitally blind children engaging in a face-to face conversation performed representational gestures comparably to sighted children and children who are blind-folded (Iverson & Goldin-Meadow, 2001; Iverson & Goldin-Meadow 1997). Furthermore, Jalec & Jaworska (2014) reported that blind children employ representational gesture when they express complex concepts, entailing that gesturing helps them in thinking and speech production. If children who are blind from birth behave similarly to children in the other two groups when expressing complex thoughts, then it is reasonable to deduce that gestures are rooted in the speaker's internal production system rather than a learned behavior or a behavior that has inherent reference to the listener, which possibly has an intention to communicate.

Several models of representational gesture production have been proposed to spell out the Listener-neutral explanation. These models attempt to account for the role representational gestures play in various stages of the production system. The Lexical Retrieval Hypothesis (Krauss, Chen, & Gottesman, 2000; Krauss, Chen, & Chawla, 1996) asserts that the primary function of representational gestures is specifically for retrieving a lexeme. Based on Levelt's (1992) model of speech production, which is a

cascading model that begins with the conceptualizer and ends with the articulator, representational gestures facilitate speech production at the lemma level, where the stage where grammatical encoding occurs. The facilitation is argued to be carried out via a *kinesic monitor*, which is a cross-model priming, to aid speakers in lexical retrieval. Once the word is retrieved and the speaker utters the word, representational gestures are terminated since the goal of lexical retrieval has been achieved. Empirical evidence to support the Lexical Retrieval Hypothesis includes a higher number words recalled using representational gesture (Frick-Horbury & Guttentag, 1998; Krauss et al., 2000; Kruss et al., 1996), resolution of the Tip of the Tongue state (Beattie & Coughlan, 1999), and temporal asynchrony of representational gesture onset and speech onset in unfamiliar words (Morrel-Samuels & Krauss, 1992). To proponents of the Lexical Retrieval Hypothesis, representational gestures do not have a communicative intention, for their primary function is to serve a very restricted and local purpose of lexical retrieval (Krauss, Dushay, Chen, & Rauscher 1995; Krauss, Morrel-Samuels, & Colasante, 1991).

Representational gestures, however, have been argued to originate above and beyond the lemma level and exhibit functions other than lexical retrieval. The following models can be construed as an extension of, as well as compatible with, McNeill's (1992) Growth Point model, a hypothesis that posits a primitive and fundamental unit where language and representational gestures emerge and organization of thoughts commence. Although these various models may differ in their components, they share a common core assumption: representational gestures help speakers in the thought process, originating at the conceptualization level of thoughts, a level that precedes the lemma level in Levelt's speech production model.

Kita (2000) proposed the Information Packaging Hypothesis, which argues that representational gestures help organize thoughts pre-verbally. The model is grounded in the assumption that people have a repertoire of semantically-rich spatio-motoric information; for such information to be transferred to the verbal form, representational gestures assist in breaking down the information into smaller chunks so that it can be "packaged" appropriately in the verbal output. Offering empirical evidence that is in contrast to the Lexical Retrieval Hypothesis, Alibali, Kita, & Young (2000) showed that when there is comparable demand for lexical retrieval but a higher demand for conceptual packaging of information, children utilize more representational gestures in their description. Such evidence makes a strong case that representational gestures occur at the conceptualization level, rather than at the lemma level. Results in Kita (1993), exhibiting data from Japanese and English, also support the argument that representational gestures function in shaping the linguistic output, thus, suggesting that representational gestures lend a hand in organizing rich information in the speaker's production system. Also compatible with the Information Packaging Hypothesis is the Interface model (Kita & Özyürek, 2003), which argues that representational gestures facilitate the organization of spatial information pre-verbally; importantly, such organization and packaging of information is language specific and may vary depending on one's spatial and verbal skill (Hostetter & Alibali, 2007).

Along the same line of how representational gestures help the thinking process, the Gesture as Simulated Action framework (GSA), proposed by Hostetter & Alibali (2008), contends that representational gestures are a mentally simulated action; both linguistic input and embodiment act in concert as a unified system in the speaker's

production system. Hostetter & Alibali (2008) reported a much higher representational gesture rate when speakers describe how to wrap a package than when they describe a cartoon, signifying that a motoric action elicits more representational gestural execution. Perhaps, the GSA framework best accounts for a specific type of representational gesture; in other words, as Wagner, Malisz, & Kopp (2014) commented, "[the GSA model] focuses on action-related (pantomimic) gestures that may directly reflect internal simulations during the moment of speaking" (p. 217). The GSA framework is essential to this dissertation since the representational gestures data in chapter IV are motorically mimetic in nature.

In contrast to the Lexical Retrieval Hypothesis, the Growth Point model, the Information Packaging Hypothesis, the Interface model, and the GSA framework, posit that representational gestures are a part of a communicative intention, for the origin of representational gestures is at the conceptualization stage, where some messages are to be executed via representational gestures while others are to be executed via speech. In other words, semantic content which speakers intend to convey can be reflected multi-modally. Although Melinger & Levelt (2003) provided evidence to support that representational gestures bear a communicative intention, it is plausible that some representational gestures may demonstrate more communicative-intention than others, depending on circumstance (e.g., the task in hand). It is possible that some representational gestures have little to no communicative intention and are simply residue of one's production system.

It is important to think of a representational gesture as a function-specific act. That is, representational gestures should not be viewed as a one-size-fits-all phenomenon;

they can serve multiple functions appearing in an array of unconventionalized forms. For example, if a speaker says to her friend that she caught a fish and that the fish is *this big* while holding out both hands, performing a representational gesture to show the fish's size, then the representational gesture functions to carry semantically-rich content that is necessary for encoding and decoding absolute size information. In this case, it is evident that the representational gesture is meant to be communicative. Although this kind of representational gesture carries rich semantic content, it is, nevertheless, a form of gesticulation, and not an emblem (see McNeill's continuum, section 2.5 in this chapter). The representational gesture itself is incomplete and likely to be incomprehensible without the verbal counterpart of *this big*; whereas, an emblem requires no accompanying speech to encode and decode the semantic content. An extensive discussion of this type of representational gesture is in chapter IV of the dissertation.

On the other hand, a representational gesture can lack communicative intention. Recall the bunny hopping out of the room example (see section 2.5 in this chapter). Unlike the absolute size representational gesture, the communicative value of representational gesture in this case is unclear. The semantic content of *the rabbit is hopping out of the room* is packaged in speech, the primary mode of communication. In this case, the representational gesture carries highly redundant semantic content; thus, the communicative function of such representational gesture is presumably low. Indeed, these two examples indicate two kinds of representational gesture on the extremes in terms of the function they serve and how communicatively oriented they are. Much of representational gesture, however, falls in between the two, and it is more difficult to determine their function.

Since representational gestures vary in function, any simple model of representational gesture production would only be able to account for a specific function they interplays with, be it retrieving a lexeme, organizing thoughts, or expressing mental simulation. Operating under the assumption that representational gestures serve multiple functions, these models may be just as equally plausible in their explanation, depending on the phenomenon being examined. In short, no one particular representational gesture production model can account for all representational gestures.

These models are only some of the prominent models that account for the role representational gestures play in speech production. Despite the specific details on how they assist speakers in speech production or whether they are communicatively-intended, one consensus these models share is that they emphasize the function of representational gestures in the speaker's production system without an inherent reference to the listener. Hence, these models support the argument that representational gestures are *for the speaker*, which is in line with the Listener-neutral model.

3.2.2. *The Listener-neutral explanation: The Repetition Effect in gesture*

The Repetition Effect found in representational gesture is unlike the Repetition Effect found in sign language (Hoetjes et al., 2011). When hand signs in the Sign Language of the Netherlands (NGT) are repeated during signing, the proportion of signs ("number of signs divided by duration") performed remains similar in the first signing and subsequent signing. Note that although they are proportional, the subsequent signing appears to be less precise; such decrease in precision is perceptually detectable. On the other hand, when representational gestures are repeated, they decrease in number, size, and number of hands used (frequently one-handed instead of two). Precision of the

repeated representational gesture, similar to NGT, is also lower. In sum, unlike hand signs in NGT, which only reduces qualitatively when repeated, representational gestures were found to reduce qualitatively and quantitatively when repeated. This preliminary work is of great importance since it is possibly the first experimental work that emphasizes quantification of reduction in repeated representational gesture.

Although Hoetjes et al. (2011) did not position the Repetition Effect in the Listener-neutral framework, the results could be interpreted to coincide with the Listener-neutral framework. Indeed, such difference in the magnitude of reduction is to be expected, considering the degree of conventionalization of NGT and representational gesture varies (see McNeill's continuum, section 2.5 in this chapter). Since hand signs are highly conventionalized and carry a rich semantic content for the primary purpose of communication, they should be more resistant to reduction in repetition in order to retain transparent semantic content; in contrast, representational gestures, which are not at all conventionalized, are more prone to reduction in repetition. To elaborate, a hand sign carries highly conventionalized meaning, which is parallel to spoken words; therefore, a substantial reduction of a sign might have an effect on how a sign is perceived (cf. Battinson, 1974 for asymmetry in ASL, although see Sumner, Kim, King, & McGowan, 2013 on linguistic and external factors that contribute to successful speech perception). In contrast, a representational gesture does not carry a conventionalized meaning; therefore, reduction is much less likely to affect the meaning, considering the meaning is primarily carried out in speech.

It is possible that in the initial reference, representational gestures benefit speakers in speech production and are of great value; however, in the repeated reference when the

cognitive demand is lower, possibly due to greater accessibility (see Ariel, 1990; Ariel, 1988) or rehearsal (see Chawla & Krauss, 1994), the value of gesturing diminishes, resulting in reduction (see also discussion in Masson-Carro, Goudbeek, & Krahmer, 2014). Therefore, based on this interpretation, reduction in repeated gestures could be taken to align with the Listener-neutral explanation since gestures are performed for the benefit of the speaker and has no inherent reference to the listener.

Recent work by Masson-Carro et al. (2014) has shown that in a director-matcher task, when speakers are faced with a high cognitive demand due to a time-constraint, the Repetition Effect in the production of representational gesture appears to occur less than when speakers are not faced with a high cognitive demand. In other words, the magnitude of reduction in subsequent mention varies depending on how much cognitive load is imposed on the speaker. When speakers require assistance from a representational gesture, they will be more resistant to reduction. Masson-Carro et al.'s (2014) position is consistent with the Listener-neutral explanation; that is, gestures are intended to benefit the speaker internal cognitive processes and have no inherent reference to the listener. In particular, they situated their results with Kita's (2000) Information Packaging Hypothesis, a gesture production model that is in line with the Listener-neutral explanation.

4. *The Listener-modeling explanation*

In this section, literature on the Repetition Effect in speech and in gesture that coincides with the Listener-modeling explanation (see section 2.3 in this chapter for a full definition) will be discussed. In contrast to the Listener-neutral explanation, speakers take into consideration the state of knowledge of their listener; their performance is, thus, a

production modulated by their sensitivity to the state of knowledge of the listener. In this view, there is an inherent reference to the listener and that speakers' performance is influenced by it.

4.1. *Listener-modeling in speech: The Repetition Effect*

At least on a coarse-grained level, speakers are assumed to track the listener's state of mind and adjust their production accordingly, making words longer (hyperarticulation) when the listener might have trouble understanding them and shorter (hypoarticulation) when no such difficulty is expected (see also Frank & Jaeger, 2008; Jager, 2010; Lindblom, 1990). In this view a repeated referent is hypoarticulated, making word duration shorter, because speakers are assumed to be sensitive to what the listener already knows or what the listener has heard before.

Initially, Fowler and Housum (1987) provided evidence that repeated content words undergo reduction when they are used in a spontaneous conversation. Fowler (1988) further provided evidence to support the Listener-modeling view with data showing that repeated words are reduced when a listener is present; in contrast, repeated words do not undergo reduction when they are read from a context-less list into a microphone. The early work by Fowler & Housum (1987) and Fowler (1988) suggested that the presence of a listener and context play an important role in the existence of the Repetition Effect in speech.

The notion of listener-modeling was additionally supported by Galati & Brennan (2010) who have shown that attenuation occurs when speakers retell a story to someone who has heard it; however, attenuation does not occur when speakers retell a story to someone who has not heard it. In particular, in the first experiment, they found that

duration of target words produced by the speakers, measured acoustically, did not differ as a function of the state of knowledge of the listener, nor did it differ as a function of state of knowledge of the speakers themselves. In the second experiment, however, they found words that are repeated to the same listener show lower intelligibility (in a clarity rating); therefore, Galati & Brennan (2010) argued that although repeated words to different listeners may not differ in duration, they differ in how intelligible they are to the listener, thus, exhibiting listener-sensitivity.

It is important to point out that word duration and intelligibility often overlap. Generally speaking, a word that is hyperarticulated (often longer in duration) has higher intelligibility, and a word that is hypoarticulated (often shorter in duration) has lower intelligibility (cf. Lindblom, 1990). Indeed, the motoric commitment speakers invest in articulation of these words differ; hyperarticulation entails more effort than hypoarticulation. That Galati & Brennan (2010) found word duration to not differ as a function of state of knowledge of the listener but found intelligibility to differ as a function of state of knowledge of the listener indicates less commitment and effort used in articulation. In short, when speakers produce words that are shorter in duration or speakers produce words that exhibit lower intelligibility, both cases imply less commitment and effort used in articulation.

While the idea that the speaker is running a full-scale simulation of the listener's brain in real time may appear implausible, Galati & Brennan (2010) argued that the computational costs associated with modeling the listener can be fairly minimal, a single bit of information encoding whether the speaker remembers saying the current word to the current listener. Thus, they proposed that speakers operate by a "one-bit" model

where speakers take into consideration whether the listener has been exposed to the information before. Their "one-bit" model is consistent with the Listener-modeling explanation, for the model explicitly takes into account the state of knowledge of the listener.

One criticism of Galati & Brennan's (2010) work is that they positioned their results in the Listener-modeling framework; in essence, they argue that speakers' speech is driven by modeling the state of knowledge of the listener. However, naïve speakers and listeners were allowed to interact throughout the experiment, which was likely to lead to constant exchanges of verbal and non-verbal cues. Since an interaction may have occurred, it is difficult to determine whether the results in their study are purely a product of listener modeling or whether the results in their study are a product of the interaction between interlocutors. This is a question that is core to the dissertation and will be vigorously discussed the general discussion.

4.2. *Listener-modeling in gesture*

Beattie & Aboudan (1994) reported that speakers gesture the least in a monologue and the most in a dialogue, which suggests that listeners are likely to play a role that influences speakers' gesture production. Indeed, a great bulk of literature is dedicated to demonstrating that gesturing is a listener-sensitive act, rather than driven by an internal mechanism that assists speakers in the production system (see section 3.1.2 in this chapter). In this view, speakers may modify and adapt their representational gestures in order to fulfill communicative needs; hence, they are performed *for the listener* as a mean of social interaction.

4.1.1. *Sources of listener-sensitivity in representational gesture production*

Visibility is a common source of evidence that is used to support the notion of listener-sensitivity. Many studies have investigated the effect of visibility on gesture, albeit yielding mixed results (see Bavelas & Healing, 2013 for an extensive list and discussion of these studies). While it is true that speakers gesture when there is a lack of visibility, for instance, on the telephone (Rimé, 1982; de Ruiter, 1995), speakers gesture at a higher rate in the visibility condition (Cohen & Harrison, 1973; Krauss et al. 1995; Mol, Krahmer, Maes, & Swerts, 2009a; Mol, Krahmer, Maes, & Swerts 2009b). Note that Alibali et al. (2001) reported speakers gesture at a comparable rate in both visibility and lack of visibility condition. They explained this as the result of imagining an interlocutor being present.

Moreover, Bavelas et al. (2008) reported a qualitative difference in the visibility condition, where speakers' representational gestures contain more semantically-rich information (such as larger in size), and speakers explicitly draw the listener's attention to their representational gestures. Furthermore, visibility can lead speakers to adapt their representational gesture to the listener's representational gesture. Kimbara (2008) and Kimbara (2006) reported that when speakers can see the listeners' representational gesture, they adapt and mimic their representational gesture to the listener's. Note that the act of mimicking is not restricted to between the interlocutors: Kuhlen & Seyfeddinipur (2007) showed that speakers mimic representational gesture of those who are not the paired interlocutor, and even when those representational gestures were viewed briefly. Despite these studies showing that mimicking of representational gesture occurring during an interaction, it is unclear, however, whether the speaker is adapting to

the listener's representational gesture or they are adapting to listener behavior in general. These two factors have not been yet teased apart in the literature and require further examination.

Other than the effect of visibility, the notion of common ground, the speaker's assumption and beliefs of shared knowledge between interlocutors, is often invoked to demonstrate speakers' sensitivity to the listener (see section 2.2 for a full definition). When speakers believe that there is an established common ground between themselves and the listener, it results in a decrease in various aspects of representational gesture such as precision, complexity, and informativeness (Gerwing & Bavelas, 2004; Holler & Stevens, 2007). Similarly, when there is a decrease in salience along with an increase in common ground, there appears to be less gesturing (Parill, 2010). Conversely, when speakers know that common ground is lacking between themselves and the listener, they provide more detailed representational gesture by utilizing higher space in gesturing and emphasizing the trajectory of an action for the listener, while they did not alter their speech as a function of the state of common ground with the listener (Hillard & Cook, 2015).

In terms of a specific type of gesture, Holler & Stevens (2007) showed that representational gestures encoding size are expressed differently as a function of common ground; that is, when common ground is lacking between interlocutors, size is expressed strictly via representational gesture or both speech and representational gesture; in contrast, when common ground exists, size information is predominantly conveyed via speech and is not expressed via representational gesture. In other words, a modality shift away from gesture occurs as a function of common ground.

In sum, these studies have shown the production of representational gesture can be listener-sensitive, rather than purely planned via an internal mechanism that serves the speaker. It appears to be listener-driven, possibly to achieve the goal of communicating information, resulting in different gesture rate, size, and form. Thus, representational gestures have also been argued to serve a communicative function to the listener.

4.1.2. *Listener-modeling and the Repetition Effect in Listener-modeling and the Repetition Effect*

Over the course of a dynamic conversation, a representational gesture may reduce when it is repeated, demonstrating the Repetition Effect. Gerwing & Bavelas (2004) showed that when a new referent is introduced, speakers produce more representational gesture than during subsequent mentions of the same event, both qualitatively and quantitatively. When speakers introduce new information into discourse, representational gestures used to accompany the new information have high saliency (i.e., larger in size and clearer); in contrast, when speakers refer to information that has been established in prior discourse (old information), the repeated representational gestures have low saliency (smaller in size and less clear). Gerwing & Bavelas (2004) account for the Repetition Effect with the notion of information status and common ground where new information generates more prominent representational gestures while old information does not.

While the idea that new information requires greater communicative weight, resulting in hypergesticulation, and old information requires less communicative weight, resulting in hypogesticulation, is a plausible one, it might be premature to conclude that the new vs. old/given information dichotomy is causing the Repetition Effect. As with

Bybee's (2001, 2002a,) speech production model, gesture reduction could merely be due to automatization of a repeated action. That is, any motoric action, be it speech or gesture, seems prone to reduction with repetition and not necessarily driven by a change in information status. Hence, there appears to be a confound between reduction due to information status (new *vs*. old/given) and reduction due to automatization of a repeated action. Unless this confound is disentangled, possibly with a sequence of listeners, the motivation for the reduction remains uncertain. Moreover, Gerwing & Bavelas' (2004) interpretation of information status still poses a problem identical to the discussion in section 3.1.3 in this chapter; that is, it is impossible to determine whether the information status of new vs. old/given is relevant for the speaker, for the listener, or for both interlocutors.

Furthermore, the foundation of Gerwing & Bavelas's (2004) and work is based on the notion of common ground, a notion that is consistent with the Listener-modeling explanation, for it focuses on the speakers' assumption of knowledge between them and the listener. However, the data in their study were elicited from an interactive situation where interlocutors were free to exchange verbal and non-verbal cues, which is consistent with the Listener-interactive explanation. Thus, it is quite possible that it something in the interaction that the drove variation in representational gesture found in their data, and not common ground per se. In sum, speakers may not have been modeling the listener through common ground; rather, they may have been responding to the listener's behavior, which is more consistent with the Listener-interactive explanation.

A joint-effort task has also been reported to demonstrate the Repetition Effect. Naïve speakers in Hoetjes et al. (2011) were engaged in a director-matcher task of novel

GREEBLE creatures, where they find that the subsequent reference to the GREEBLE creature is reduced in the rate of representational representational gesture produced. Note that Hoetjes, Koolen, Goudbeek, Krahmer, & Swerts (2015) later reported that repeated gestures within the same discourse show a U-shaped distribution where the first reference and the third reference of the same referent appear to have the same rate of gesturing, and only the second referent shows a reduction. They explain the U-shaped distribution with the task effect, where speakers gradually learn to effectively and efficiently gesture certain information.

In their work, they account for the Repetition Effect with the Uniform Density Hypothesis (Jager, 2010; Frank & Jager, 2008), which asserts that the expression of predictable information, in this case, information that is repeated, is likely to be reduced. Again, it is unclear whether the information is presumed predictable to the speaker alone, is presumed predictable to the listener by the speaker, or presumed predictable to both interlocutors (see section 3.1.3 in this chapter).

4.1.3. *Listener-modeling and the Repetition Effect in gesture with multiple listeners*

The Repetition Effect has also been found when there is a switch in listeners. Jacobs & Garnham (2007) found that the rate and the number of representational gestures in repeated referents reduce as a function of the listener's previous exposure to the information. In one of the conditions of their experiment, they asked speakers to tell and retell the same comic strip three times to the same listener; and in another condition, they asked speakers to tell and retell the same comic strip to three different listeners. There was a significant decrease in gesture rate and number of representational gestures only when the comic strip was repeated to the same listener.

Elaborating on the work of Jacobs & Garnham (2007), Galati & Brennan (2013) also used multiple listeners, more importantly, in a sequence of new and old listeners. Their experimental design allows a more substantive comparison between conditions since the design creates a mismatch of information between the speaker and the listener (see section 5.1 in this chapter for a full discussion). Confirming Jacobs & Garnham's (2007) results on frequency of representational gestures as a function of the listener having been exposed to the information before vs. not, Galati & Brennan (2013) found that when a story is being retold to a listener who has heard it before, the rate of representational gesture decreases. Additionally, speakers perform representational gestures that are smaller in size and with less precision. They also noted that, "after the first telling, speakers attenuated both gesture size and iconic precision, but less so if the retelling was to a new addressee than to an old addressee" (p. 447). In other words, repeated representational gestures reduce as expected; however, the magnitude of reduction is dependent of the listener's previous exposure to the information. They interpret the results with the "one-bit" model, which entails listener-modeling (see section 4.1 in this chapter for the explanation in speech).

Once again, the results found in Galati & Brennan (2013) and Jacobs & Garnham (2007) could be interpreted as having derived from an interaction between interlocutors (the Listener-interaction explanation) rather than the speaker's assumption about the listener's state of knowledge (the Listener-modeling explanation, the "one-bit" model). In Galati & Brennan (2013), naïve speakers were able to freely interact with the listeners in a narrative task; in fact, participants who played the listener role were told that they could "freely comment or ask questions for clarification during the narration" (p. 440).

Galati & Brennan (2013) were aware of the possible effects feedback could have on production of representational gesture. They examined the effect of listener feedback in their study and did not find an effect; hence, they suggested that the speakers' gestures were unlikely to be influenced by the listener's responses. However, their analysis may not be fine-grained enough to make this conclusion. Firstly, all audible feedback examined was collapsed into a single category, where listener comments, questions, back channeling, and laugher were treated the same. This single category may not be sufficient to evaluate the effects of the various kinds of feedback found in their data. For example, Holler & Wilkin (2011) reported that the rate of representational gesture decreases when the verbal feedback is confirmatory; however, the rate remains constant when the verbal feedback is clarification, elaboration, or correction. They also reported that not only the rate of representational gesture is dependent of verbal feedback, but the size may vary as well. Therefore, representational gestures may vary quantitatively (in rate) or qualitatively (in form) depending on the type of verbal feedback.

Next, feedback coded in Galati & Brennan (2013) was only auditory (linguistic or non-linguistic), leaving all non-verbal feedback unaccounted for. This poses a problem since there is evidence that non-verbal feedback plays a significant role during an interaction. For example, Chovil (1991) reported that facial displays accompanying speech have multiple functions in discourse (in a conversation) and carry semantic content. Allwood & Carrato (2003) highlighted the relationship between non-verbal feedback and verbal feedback in Swedish conversation; importantly, they stress that these kinds of feedback have multiple functions in discourse (e.g., reinforcing or negative/rejection). Along the same line, Stave & Pederson (2015) reported that multiple

non-verbal cues contribute to triggering feedback in English, and Pederson (2014) emphasized that listener head gestures significantly contribute to the structure of conversation in Tamil. In sum, the nature of such fluid exchanges of non-verbal feedback should not be overlooked when interlocutors are engaged in an interaction.

While Jacobs & Garnham (2007) used confederate listeners in their study, they were instructed to "listen to the descriptions of the comic strips, but were given no further instructions" (p. 295), from which it could be reasonably inferred that interaction between interlocutors took place, though it seems less likely that the confederates pretending to hear the story afresh will have behaved like genuinely naïve listeners. Although Jacobs & Garnham (2007) found that the Repetition Effect in representational gesture only occurs when speakers retell the same story to the same listener, they noted that since their confederates were allowed to interact with the speaker, the variation in gesturing could have come from differences in individual confederate behavior; in their words, since "[the confederates] were not controlled in any way, it is difficult to know if the reactions of the [confederates] had any notable effect on the number of gestures produced by the speaker" (p. 298). Eventually, Jacobs & Garnham (2007) concluded based on their data that speakers are sensitive to the listener. It would be more precise to interpret that such listener-sensitivity to have likely derived from an interaction between interlocutors, rather than sensitivity deriving from the speaker modeling the listener's state of knowledge.

5. *Addressing potential issues in past literature*

As stated in the beginning of this chapter, the goal of the dissertation is to examine speakers' production across modalities. In particular, I aim to use the Repetition Effect to investigate speakers' production in speech and in representational gesture, two

modes of production that greatly differ in the degree of conventionalization, and determine the model that speakers operate by. The investigation of two modalities allows a holistic examination of speakers' production processes, possibly highlighting features of the Repetition Effect that may be modal specific.

In this chapter, potential issues in the interpretation of the Repetition Effect in both modalities were discussed in detail. The issues can be summarized and grouped into two areas: (1) the confound between new and old/given information in the interpretation of the Repetition Effect; (2) the confound between the Listener-modeling explanation and the Listener-interactive explanation in the interpretation of the Repetition Effect.

5.1. *Addressing the confound between new and old/given information*

Previous studies that anchored themselves in the notion of information status, both in speech and gesture, utilized a single listener in the data (see section 3.1.3. in this chapter on speech and 4.1.2. in this chapter on representational gesture). Having a single listener is problematic because it does not allow teasing apart of information status to the speaker and the listener. Table 1 demonstrates that by having only one listener, information status overlaps between the speaker and the listener.

Table 1: A confound of information status to the speaker and listener

Assumed state of Information status	*Speaker* production	*Listener* exposure
New	First production	First hearing
Old/given	Subsequent production	Subsequent hearing

In the first production, the information is presumed new to both the speaker and the listener since the referent is being produced for the first time, as well as being heard for the first time. In the subsequent production, the information is presumed old/given to both the speaker and the listener since the referent is being produced for the second time, as well has being heard for the second time. Therefore, there exists a confound of to *whom* the state of information applies to: whether it applies to the speaker or to the listener. The more important question, however, is which is responsible for guiding speaker behavior?

To disentangle such confound, the experiments in this dissertation use multiple listeners in a sequence to create a mismatch of information status between speaker production and listener exposure. The sequence used in this dissertation, shown in Table 2, originally appeared in Galati & Brennan (2013, 2010), where speakers repeat the same information three times, twice to listener 1 and once to listener 2. This sequence is highly favored since it teases apart the confound between reduction due to automatization and

reduction due to assumed information status, which was discussed in detail in section 4.1.2. in this chapter .

Table 2: Design of the experiment and the assumed state of knowledge

<table>
<tr><th>Condition</th><th>Speaker production</th><th>Listener exposure</th></tr>
<tr><td>A
Listener 1</td><td>First production</td><td rowspan="2">First hearing</td></tr>
<tr><td>B
Listener 2</td><td rowspan="2">Subsequent production</td></tr>
<tr><td>C
Listener 1</td><td>Subsequent hearing</td></tr>
</table>

In condition A, speakers present the information for the first time, and the listener hears the information for the first time. Since information is new to both interlocutors, condition A to is the baseline. In condition B, speakers repeat the same information for the second time; hence, the information is assumed old/given to the speaker. However, the same information is assumed new to listener 2 since it is the first time it is being heard. In condition C, the speakers repeat the story to listener 1 again. The information is assumed old/given to both interlocutors. Because of the mismatch of information status to the speaker and the mismatch of information status to the listener, it can be better determined what exactly is guiding speaker behavior.

5.2. *Addressing the confound between Listener-modeling and Listener-interaction*

The Repetition Effect has been accounted for with speakers demonstrating sensitivity to the listener (see section 4 in this chapter). To elaborate, as a result of modeling the state of knowledge of the listener, speakers' production, both in speech and gesture, may reduce in repeated referents because the speaker is sensitive to the fact that the listener has heard the information before. In a sense, if the listener has heard it before, then the information is considered old/given to the listener and such information is conveyed differently from information that is new to the listener.

Although Galati & Brennan (2013, 2010) used multiple listeners in a sequence to rid the information status confound addressed in section 5.1 and proposed the "one-bit" model, which is consistent with the Listener-modeling explanation, their study lacked control for the interactive aspect in the experiment (see section 4.1.3. in this chapter). To conclude that speakers operate by modeling the state of knowledge in their production while using data elicited from an interactive setting is a problematic conclusion.

To tease apart the issue of listener-modeling and listener-interaction, experiments in this dissertation use confederate listeners, as opposed to naïve listeners in most of the other studies. In this dissertation, in addition to presenting the speaker with multiple listeners in a sequence (see section 5.1. in this chapter), speakers were presented with confederate research assistants who provided minimal interaction with the speaker. Using controlled confederate listeners can shed light on whether speakers' production is strictly the result of modeling the state of knowledge of the listener and not the result of responding to dynamic listener feedback.

The confederates are native speakers of American English, one male and one female, in the same age range as the speakers, who were recruited from an introductory

psycholinguistic class for extra credit. Prior to the experiment, they were given explicit instruction and training on giving limited feedback to the speakers. Namely, they followed Mol et al's (2009b) description of a limited feedback, which states that, "We prevented true dialogue from happening by instructing [confederates] not to interrupt the speaker but to act naturally otherwise. Thus, [confederates] were looking at the speaker and gave occasional non-verbal feedback, [confederates] tried to avoid speaking themselves" (p. 1571). Based on this description, confederate listeners provided linguistic back channeling, for example saying "uh-huh", maintained eye contact with the speakers, and nodded while listening. Furthermore, confederates were asked not to gesture while engaging with the speakers.

Gerwing & Bavelas (2013), Bavelas & Healing (2013), and Kuhlen & Brennan (2013) have cautioned against the use of confederates in experiments due to the issue of ecological validity. It is true that the use of confederates may alter speakers' performance, thus, providing results that may be less representative; however, confederates offer a stricter control that allows teasing apart listener-sensitivity deriving from modeling the listener vs. listener-sensitivity deriving from an interaction.

In addition, the confederates have heard the story many times prior to coming in contact in this study, for they were confederates in the pilot data, which have been excluded from the analysis. Hence, condition A, B, and C (see section 5.1 in this chapter) are all equally familiar content to the confederates, further minimizing variation in their response. To ensure that there was no apparent difference between condition A and B, the videos of the confederates were reviewed; no discernable difference was found in their behavior in each condition.

CHAPTER II

THE REPETITION EFFECT ON WORD DURATION

1. *Introduction*

When people talk, some words are used more than once. The duration of a word is longer when it is mentioned for the first time within a discourse than in subsequent mentions (Anderson & Howarth, 2002; Aylett & Turk, 2004; Baker & Bradlow, 2009; Bell et al., 2009; Fisher & Tokura; 1995; Freeman, 2014; Fowler, 1988; Fowler et al., 1997; Fowler & Housum, 1987 Galati & Brennan, 2010; Lam & Watson, 2010; Pluymaekers et al., 2005; Vajrabhaya & Kapatsinski, 2011). While the existence of this effect of repetition on word duration is by now indisputable, the explanation for it remains elusive. In particular, researchers disagree on whether the effect is best thought of as subsequent-mention reduction (Bybee, 2001, Bybee, 2002a; Fowler & Housum, 1987; Galati & Brennan, 2010); first-mention lengthening (Bell et al., 2009); or bidirectional durational adjustment (Aylett & Turk, 2006).

2. *Processing accounts of the Repetition Effect in speech*

Most popular explanations for the Repetition Effect have appealed to processing difficulty. Broadly speaking, these processing explanations for the Repetition Effect can be categorized into two types: the Listener-modeling explanation and the Listener-neutral explanation. According to Listener-modeling explanation, repetition influences duration because speakers are assumed to be sensitive to what the listener already knows or has heard before (see section 4.1 in chapter I). In other words, speakers are assumed to track the listener's state of mind and adjust word durations accordingly, making words longer when the listener might have trouble understanding them and shorter when no such

difficulty is expected (see Lindblom, 1990; Fowler, 1988; Frank & Jaeger, 2008; Galati & Brennan, 2010). Note that Baese-Berk & Goldrick (2009), experiment 2, have provided evidence that suggests otherwise; in other words, they proposed that modeling the listener is unlikely to be the motivation for hyper/hypo articulation of VOT. Although the scope of this chapter is on word duration, VOT is a vital phonetic cue that is closely related to word duration and should be noted here.

While the idea that the speaker is running a full-scale simulation of the listener's brain in real time appears implausible, Galati & Brennan (2010) argued that the computational costs associated with modeling the listener can be fairly minimal, a single bit of information encoding whether the speaker remembers saying the current word to the current listener (see Galati & Brennan's "one-bit" model, section 4 in chapter I).

Proponents of the Listener-neutral explanation, on the other hand, are skeptical of the speaker's ability to model the listener, or at least of the idea that low-level dynamics of articulatory planning and execution, which manifest themselves in word duration differences, are affected by such modeling (e.g. Bard et al., 2000; Bard & Aylette, 2005, Bell et al., 2009; Bybee, 2002; Bybee 2001). They, therefore, attribute duration differences associated with repetition to the effect of repetition on automatization of articulatory sequences (Bybee 2002; Bybee, 2001) and repetition priming, where previous mention of a word is assumed to result in a sustained increase of the word's activation level (repetition priming), making the word easier to retrieve and pronounce, and also reducing its duration by increasing gestural overlap due to greater anticipatory activation of upcoming targets (Kapatsinski, 2010; see also Dell et al., 1997). It should be highlighted that I do not claim that proponents of the Listener-neutral explanation to

reject that idea that speakers are sensitive to the listener (via modeling of the listener to be precise). Recall that the notion of listener-neutrality, as defined in section 2.1 in chapter I), only applies to a limited scope; in this case, it applies to word duration. Indeed, speakers may show their sensitivity to the listener via other phonetic cues; for example, VOT length or intensity, features that are beyond the scope of this dissertation.

Rather than viewing the Repetition Effect as subsequent mention reduction, Bell et al. (2009) argued that it is rather first-mention lengthening, deriving from processing difficulties in the speaker's production system (see section 3 in chapter I). Despite their differences, processing explanations of the Repetition Effect share the assumption that duration differences among words are to be explained by how accessible those words are in the moment of production, either *to the listener* or *to the speaker*. It is this assumption that will be questioned and tested in this chapter.

In this chapter, data suggesting that word duration does not strictly track word accessibility for either interlocutor will be presented. Namely, a word that is repeatedly mentioned by a speaker to the same listener can lengthen (relative to preceding productions) when the word is mentioned for the first time within a re-telling of a story. Based on the data in this study and prior data from Vajrabhaya & Kapatsinski (2011) and Fowler et al. (1997) (see section 3.1.2. in I), it is strongly suggested that words are automatically lengthened when they are mentioned for the first time within a story. An exemplar-theoretic account for the results will be provided in the discussion section of this chapter.

3. *Methods*

3.1. *Participants*

Data for this study was obtained from fifteen native speakers of American English, consisting of eleven female speakers and four male speakers, who self-reported to have normal hearing and speech. All participants were undergraduate students at the University of Oregon who took an introductory psychology or linguistics class that required participation in the Psychology/Linguistics Human Subjects Pool (or an alternative assignment). Participants earned research credit for their classes by taking part in the experiment.

3.2. *Procedures*

Participants were asked to watch a short video clip of a man making a pizza dough. They were allowed to watch the clip as many times as they wished until they were certain that they were able to tell other people how a pizza dough is made. Participants were also explicitly told that the duration of time they spent watching the clip was not being recorded, hence, allowing them to comfortably become familiar to the stimulus without time pressure. All participants watched the clip at least twice before they informed the researcher that there were ready to describe the process.

3.3. *Listener type*

Two confederate listeners, one male and one female, were present during the experiment (see section 5.2 in chapter I for use of confederates). The confederates were undergraduate research assistants who were in the same age range as the participants. Generally following Mol et al. (2009b)'s description of limited feedback, confederates' primary task was to provide non-verbal feedback (non-reactive response), which includes

nodding and maintaining eye contact with the participants. Confederates limited their verbal response to occasionally saying ‘uh-huh’ or ‘yeah’. The order in which the listeners were presented to the speakers was counterbalanced across participants. Confederates were not informed about the hypothesis until data collection has completed.

3.4. *The sequence of listeners*

The sequence of listeners in this study was originally used by Galati & Brennan (2013, 2010). Participants were asked to describe how a pizza dough is made to the first confederate, followed by a ten-minute break, during which participants were asked to play a game of Tetris. When the ten-minute break was over, participants were asked to describe the same story to the second confederate, followed by another ten-minute break which participants continued the game of Tetris. Lastly, participants were asked to once again tell the same story to the first confederate.

In order to avoid the implication that the first listener did not understand or follow the story on the first telling (which could result in hyperarticulation for clarity), the need to retell the story was explained to the participants as resulting from a recording equipment malfunction. Specifically, the researcher told the participant that the equipment did not record properly the first time around. After the third telling, participants were debriefed and assigned research credit for their participation. In sum, participants told the same story three times, twice to the first confederate listener and once to the second confederate listener.

In addition, participants were asked what they thought about the purpose of the experiment during debriefing. The majority reported believing that it was a memory task. The next most popular response was that the experiment was about how people talk to

men *vs*. women (because the confederates were of different genders). No participant reported catching on to the fact that the experiment was about how words are pronounced across repetitions (of the words or the story).

3.5. *Predictions*

Table 3 summarizes the experimental design and corresponding hypotheses. In condition (A), the information is presented and heard for the first time by listener 1; hence, the information is new to both the speaker and the listener. Thus, words should be as long as they ever are when mentioned for the first time in condition (A). This condition is considered to be the baseline.

In condition (B), the information is presented for the second time by the speaker, but it is heard for the first time by the listener (listener 2); hence, the information is old/given to the speaker but new to the listener. The Listener-neutral explanation of the Repetition Effect predicts that first mention of target words will be reduced in condition (B) compared to condition (A). In contrast, the Listener-modeling explanation hypothesis predicts no reduction at first mention relative to first mention in condition (A).

In condition (C), the information is presented for the third time by the speaker and is heard for the second time by the listener; hence, the information is old/given to both interlocutors. Reduction is expected to occur in this condition, relative to condition (A) according to both explanations.

Table 3: Design of the experiment (narrative task in American English) and the information status of target words to the speaker and the listener

<table>
<tr><th>Condition</th><th>Speaker production</th><th>Listener exposure</th></tr>
<tr><td>A
Listener 1</td><td>First production</td><td rowspan="2">First hearing</td></tr>
<tr><td>B
Listener 2</td><td rowspan="2">Subsequent production</td></tr>
<tr><td>C
Listener 1</td><td>Subsequent hearing</td></tr>
</table>

3.6. *Recording*

Participants' speech was recorded with a Marantz PMD 671 digital recorder. The WAV sound files were analyzed using Praat (Boersma & Weenink, 2010). Participants were also video-recorded for analysis of representational gesture in chapter III to V.

3.7. *Measurement*

Speech data was segmented into utterance units: a pause that is equal to or longer than 100 milliseconds demarcates a new utterance unit (cf. Kendall, 2013). Due to the nature of the video stimulus described in the speakers' narrations, which is a man making a pizza dough, several content words were used repeatedly across participants and conditions. The target words obtained for analysis include: *circle*, *circular*, *dough*, *flour*, *knead*, *knuckle*, *pizza*, *pizza dough*, *rotate*, *rip, spread*, *stretch*, and *surface*. If a target word was adjacent to disfluencies or formed an utterance on its own, the target word was excluded from the analysis. A total of 1,190 tokens were hand measured.

3.8. *Dependent variables*

There were two dependent variables: duration and f_0. To determine the initial position of a target word, each target word was hand-measured from the release of the syllable initial-consonant until the end of vowel formants or at the final-consonantal constriction (cf. Abramson, 2002). Generally, I followed Olive, Greenwood, & Coleman's (1993) methods and guidelines on segmenting static speech sounds (see chapter VII in particular). Since many of the target words include /r/, which is difficult to measure due to varying formant value in Hz, F3 was used as a cue used to decide where the phoneme begins or ends.

For f_0 measurement, the spectrogram range for females was set at 80-250 Hz, while the spectrogram range for males was set at 50-180 Hz. Target words were categorized by number of syllables, from one to three. For target words that are multisyllabic, the stressed syllable was distinguished from the unstressed syllable; for example, the target word *CIRcle* was segmented into two parts where *CIR* was coded as stressed and *cle* was coded as unstressed. The stressed syllable, where the consonant in the initial position was excluded, was measured for average f_0. Lastly, the average f_0 of the utterance was also measured.

3.9. *Analysis*

Data were analyzed using linear mixed effects models in R (R Development Core Team, 2008) with the lme4 package (Bates, 2013). Random intercepts for speakers and words and random slopes for all fixed-effects predictors of interest (excluding the speech rate covariates) within speakers and words (following Barr et al., 2013) were included. To generate p values, the z approximation was used. Note that since p values-based on

the *z*-statistics has been criticized for being anti-conservative, a model comparison was conducted, both with and without the predictors. The *z*-based estimates reported below show do not indicate a quantitative difference. Note that results in Barr (2013), Table 5, indicate very minimal difference between the two models.

For the duration analysis, the dependent variable was log word duration. The fixed effects predictors were: (1) mention within story (first mention vs. second mention) Note that Bell et al. (2009) observed no difference between second and subsequent mentions but a significant difference between first and section mention; (2) condition, with condition (A) used as the baseline level; and (3) speech rate measured as syllables per second within each utterance containing the word of interest (with speech rate before the word and speech rate after the word entered as separate predictors in analyzing utterance-medial durations).

The f_0 analysis consists of two dependent variables: mean f_0 of the stressed syllable of the target word and the difference between maximum and minimum f_0 within the target word (f_0 range within the target word). Control predictors were: (1) Mean f_0 within the utterance containing the target word and (2) Speaker gender. Mention and condition were also entered as fixed-effects predictors, as in the duration analysis.

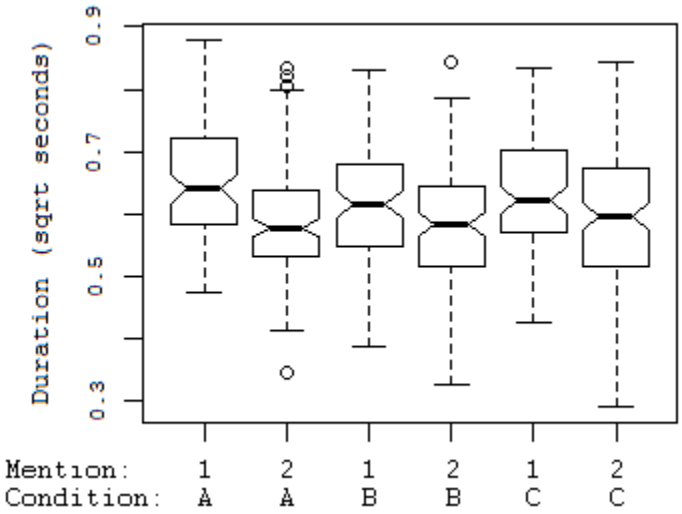
4. *Results*

4.1. *Word duration*

Figure 4 shows word duration as a function of repetition within story ('Mention') across the three story repetitions ('Condition': First telling of the story to the first listener = 'A'; First telling of the story to the second listener = 'B'; Second telling of the story to the first listener = 'C') for utterance-medial word tokens. There is a clear effect of

repetition within story (confirmed to be significant using mixed-effects modeling, as shown in Table 5 below) with no effect of repetition across stories. In addition, there is an effect of speech rate in the expected direction.

Figure 4: Word duration in medial position boxplot. Notches show 95% confidence intervals for the median.



Results remain unchanged if boundary-adjacent instances are included. Including final occurrences, there is still a significant effect of Repetition within story (b = -.021, $se(b)$ = .007, t = -2.86, p = .004) and Speech rate (b = -.003, $se(b)$ = .001, t = -2.68, p = .007) in addition to a significant effect of position within utterance (b = -.05, $se(b)$ = .006, t = -8.10, p<.0001).

Table 4: Fixed-effects coefficient estimates from a mixed-effects model with maximal random effects structure examining the effect of repetition on duration.

	b	*se(b)*	*t*	*p*
(Intercept)	.650348	.023968	27.134	<.0001
Repetition within story	-.02719	.008065	-3.372	.0007
Condition=B	-.01211	.008454	-1.432	.15
Condition=C	-.01158	0.0108	-1.072	.28
Speech rate before	-.0051	.001364	-3.741	.0002
Speech rate after	-.00058	.000584	-0.996	.32

This effect of position is responsible for the apparently lower median for initial mentions in Condition B in Figure 5: if position is not included in the model, words in Condition B are significantly shorter than those in Condition A. When initial instances are included, there is again a significant effect of Repetition within story (b = -.026, *se(b)* = .008, t = -3.36, p = .0008), with all other effects failing to reach significance ($|t| < 1$).

Comparison of subsequent mentions of a word in Story_{N} to the first mentions of the same word in Story_{N+1} reveals a significant lengthening effect across the story boundary, as shown in Table 6. When words that have previously mentioned are mentioned for the first time within a story, they are longer than at previous mention. The effect of story boundary remains significant if word occurrences adjacent to utterance boundaries are included: b = .022, *se(b)* = .008, t = 2.80, p = .005 including utterance-initial words, b = .021, *se(b)* = .008, t = 2.60, p = .009 including utterance-final ones).

Table 5: Fixed-effects coefficient estimates from a mixed-effects model with maximal random effects structure examining the effect of story boundaries on duration.

	b	*se(b)*	*t*	*p*
(Intercept)	.631083	.026738	23.602	<.0001
Repetition across story boundary	.019365	.008024	2.414	.016
Speech rate before	-.00351	.001598	-2.199	.028
Speech rate after	-.0055	.00239	-2.302	.021

Replicating Bell et al., (2009)'s corpus data, there was no significant duration difference between second and subsequent mentions: $|t|<1$ for medial words, non-final words and non-initial words. There is also no significant effect of mention if mention is entered as a continuous predictor if first mentions are excluded ($|t| < 1$).

4.2. *Fundamental frequency (f_0)*

As shown in Table 6 and Table 7, there were no significant effects of repetition within story or condition on either of the two variables. The pitch of the stressed syllable of the target word is very well predict ed by average pitch across the utterance. While it is slightly (~24 Hz) higher than average for the utterance, consistent with the fact that this syllable may sometimes bear an accent, pitch height is not affected by repetition within or across stories. The within-word pitch range measure (*M*=54 Hz) appears to be affected only by speaker gender, with men having lower pitch ranges.

Table 6: Fixed-effects coefficient estimates from a mixed-effects model with maximal random effects structure examining the effect of repetition on mean f_0 of the stressed syllable of the target word

	b	*se(b)*	*t*	*p*
(Intercept)	24.3987	10.477	2.329	.020
Repetition within story	3.1269	3.6184	.864	.387
Condition = B	4.5383	3.0796	1.474	.141
Condition = C	2.5152	11.706	.215	.830
Gender = Male	-7.3266	4.8922	-1.498	.134
Mean f_0	.8243	.0521	15.823	<.05

Table 7: Fixed-effects coefficient estimates from a mixed-effects model with maximal random effects structure examining the effect of repetition on f_0 range within the target word.

	b	*se(b)*	*t*	*p*
(Intercept)	53.73806	17.05215	3.151	.002
Repetition within story	-7.2339	8.36464	-.865	.387
Condition = B	-4.60099	4.44303	-1.036	.300
Condition = C	-3.5883	7.467373	-.481	.631
Gender = Male	-22.9008	6.752733	-3.391	.0007
Mean utterance f_0	-.00838	.07625	-.11	0.912

5. *Preliminary discussion*

Results of the experiment show initial mentions of target word are reliably longer than subsequent mentions within the same telling of the same story. Furthermore, second

mentions do not differ in duration from subsequent mentions, suggesting that it is the first mention that is unique in duration (Bell et al., 2009). When a new story starts, regardless of whether the listener has heard the story before, word duration resets. In sum, word duration reduces within story but lengthens back to its original, first-mention duration when speakers retell the story. Fundamental frequency (f_0), on the other hand, shows no statistically significant results.

Lack of duration reduction across stories accompanied by significant reduction within stories is consistent with prior work. Despite differing conclusions, neither Fowler et al. (1997), nor Bard et al., (2000, Experiment 1), nor Galati & Brennan (2010) found a significant decrease in word durations across stories (for words mentioned for the first time within a story). While Galati & Brennan (2010) concluded that words are reduced across stories when a story is repeated to the same listener, there was no significant effect of repetition across stories on duration in their data (*minF'* = .01 for story repetition within listener in Table 2). Thus, their conclusion is based entirely on a numerical effect of story repetition within listener on clarity ratings of words mentioned for the first time within a story, which is not conventionally significant (*minF'* (1,24) = 3.15, p = .08) (Galati & Brennan, 2010, p.12). The effect of story repetition across listeners is in the unexpected direction with p = .12.

In contrast to the lack of significant *across*-story duration differences, Bard et al. (2000) in experiments 2-4, Bell et al. (2009), Fowler & Housum (1987), and Fowler et al. (1997) found a significant difference in duration between first and subsequent repetitions *within* coherent stretches of discourse. Note that Galati & Brennan (2010) did not report

durations of subsequent repetitions within a story. The results here are not qualitatively changed if only first mentions within a story are analyzed.

In addition to replicating within-story reductions in duration, this study documents that words *lengthen* relative to preceding productions by the same speaker when they are mentioned for the first time within a new story. These results hold true even though words mentioned for the first time within a story generally do not occur in story-initial sentences that have been argued to be harder to process (Haberlandt et al., 1980; den Uyl & van Oostendorp, 1980) (see section 3.1.2. in chapter I for comparison). In fact, the first mention of a word in the data occurs on average 42 seconds into the story.

Researchers in the Listener-neutral tradition have been concerned that modeling of the listener is a resource-demanding process that, if carried out in full detail in real time, should make conversation impossible. To alleviate this burden, several researchers in the Speaker-neutral tradition have proposed that the speaker can default to assuming that the listener's state of mind is the same as the speaker's (e.g. Bard et al., 2000; Clark & Marshall, 1981; Horton & Keysar, 1996). In addition to this kind of egocentric processing, the cognitive burden of listener modeling can also be reduced by gradual incorporation of the 'generic' behavior of the listener into the conventions of the language, either over historical time at the level of the speech community or over the course of first language development within the individual. An account of this conventionalization process for first-mention lengthening is highly plausible.

The Exemplar-based models of the lexicon ascribe differences to storage of phonetically detailed production exemplars of words and phrases (e.g. Bybee 2002b;

Pierrehumbert, 2001, 2006). The category of possible realizations of a word is represented by the entire set of previously experienced realizations; each is associated with information about the discourse context in which it was produced (Pierrehumbert, 2001, 2006).

Joint activation of a word and a discourse context leads to activation of all exemplars of the activated word that were experienced perceptually or produced articulatorily in the activated discourse context, and in addition, weaker activation of other exemplars of the same word. As such, the activated exemplars influence the production target to the extent that they are activated (Pierrehumbert, 2001). In the current case, when a word is said for the first time, it falls under one exemplar category; whereas, when the same word is repeated within the same discourse, it falls under another exemplar category. The exemplar categories (first vs. subsequent mention) are independent; importantly, they are argued to be driven by structure of the discourse, not by accessibility of the lexical item.

The cornerstone of the Exemplar Model is rooted in one's experience. It could be the case that initially, speakers may lengthen the first mention of a word (hyperarticulate) to avoid misunderstanding (Lindblom, 1990); however, if the goal usually arises, and thus a strategy is used in a specific linguistic environment (here, the first time a word is said), the strategy *eventually* becomes associated with that environment (see also Bybee, 2006).

It is possible that over time, the strategy becomes part of the convention, a part of the grammar, starting from something that speakers once upon a time *chose to do* in order to accomplish a specific goal (i.e., of avoiding misunderstanding): it becomes something

that the speakers *cannot help but do* in a specific discourse/linguistic environment. Production of the erstwhile strategy in the context becomes automatized for the speaker; thus, at this point, it is an automatized and a conventionalized act.

The conventionalization of goal-driven production strategies into discourse structure is analogous to the generalization of phonetic pressures into the conventionalized patterns of phonology. For example, there is a phonetic motivation for devoicing of stops like /b/, /d/, /g/ in utterance-final, pre-pausal position: vocal folds, which need to be brought together to produce voicing, are open during pauses. Hence, the speaker preparing to pause may stop voicing early in anticipation. At the same time, subglottal pressure decreases throughout the utterance, making voicing (vibration of the vocal folds due to the passage of air between them) hard to maintain at the ends of long utterances (see Myers & Padgett, in press, for a review of the literature). There is no phonetic motivation for devoicing at the ends of utterance-initial words before vowels (as in /be<u>d</u> of/), yet in many languages (e.g., German and Russian), final devoicing occurs at the ends of *all* words, even before vowels. This is a prime case of a generalization that extends beyond the context where it was originally motivated.

Experimental work by Myers & Padgett (2014) shows that language learners are predisposed to generalize patterns from the utterance-final context to all word-final contexts. Once generalization occurs, the process is said to be phonological, and therefore conventionalized (phonology being conventionalized phonetics, e.g. Pierrehumbert, 2001). In the same way, that first mention is always longer than the latter could very well have been generalized beyond the context in which it is motivated, in this case, to avoid misunderstanding. This process of domain generalization (Myers & Padgett, 2014)

appears to be a sign of conventionalization: as a linguistic behavior becomes conventionalized, it becomes associated with smaller and smaller units (or domains).

It is profitable to distinguish between lexical and postlexical 'phonetic implementation' and the influences on word duration (see also Plug, 2005). Postlexical influences can be described as 'true' lengthening or shortening which can in turn be categorized into those influences affecting planning and those affecting execution. Planning-stage influences operate on a previously selected form and stretch or compress it by parametric adjustment. These include at least lengthening for the purposes of buying time for planning or accessing upcoming material (e.g. Schnadt, 2009), shortening of replacements in repair (Plug & Carter 2013; O'Shaughnessy, 1995), durational adjustments based on contextual predictability (e.g. Baker & Bradlow 2009; Jurafsky, Bell, Gregory, & Raymond, 2001), gestural alignment changes associated with differences in speech rate (Browman & Goldstein, 1990; Davidson, 2006), speech style (Smiljanić & Bradlow, 2009; Uchanski, 2008), and phrase-or utterance-level prosodic prominence (Aylett & Turk, 2006; Aylett & Turk, 2004). Execution-stage influences include automatization of execution of frequent words (Kapatsinski, 2010) as well as premature anticipatory activation of upcoming gestural targets resulting in unplanned anticipatory coarticulation (Farnetani & Recasens, 1999). Previous accounts of the Repetition Effect have assumed a postlexical locus for the psychological process that leads to the observed differences in duration.

In contrast, it could be argued that the effect should be at least partially attributed to lexical form selection, akin to selecting a full noun phrase vs. a pronoun to refer to a discourse referent (see Fowler et al., 1997 for a direct comparison of the two processes).

Additional evidence for the use of lexically-stored phonetic detail comes from studies documenting that words that tend to occur in reduction-favoring contexts are heavily prone to reduction even when used outside of these contexts (Bybee, 2002a; Kapatsinski, submitted; Raymond & Brown, 2012; Schweitzer, 2010; though cf. Cohen-Goldberg, 2015).

The viability of this kind of lexical selection mechanism for the Repetition Effect is suggested by Bell et al.'s (2009) and Baker & Bradlow's (2009) finding that the magnitude of the Repetition Effect is larger for frequent words. As proposed by Bybee (1988), because of their high individual frequency, each contextual variant of a frequent word can be processed relatively independently of the other variants. Their productions will then be driven more by their own exemplars than by the exemplars associated with the same word used in other contexts or those of the sublexical units composing the word. As a result, contextual variants of frequent words are expected to be less similar than contextual variants of rare words, resulting in greater variability among realizations of frequent words (see also Schweitzer & Möbius (2004) and Walsh, Möbius, Wade, & Schütze (2010)).

It is possible that when a word is mentioned for the first time within a discourse, it is much less accessible to both the speaker and the listener. Thus, words produced in this context are expected to be relatively long due to postlexical lengthening processes. Similarly to Galati & Brennan (2010), it can be argued that speakers keep track of whether a word is mentioned for the first time in a discourse episode, which then makes this 'first-mention' context available for being stored in memory with the produced exemplar of the word. As a result, the 'first-mention' context is, therefore, populated by

relatively long word exemplars. Joint activation of the 'first-mention' context and a word is then expected to result in activation of relatively long exemplars of that word even when the word is easily accessible to both the speaker and the hearer.

6. *Preliminary conclusions*

With the exception of Fowler et al. (1997), past literature has examined the effect of repetition on word duration as an online process of reduction or, less commonly, lengthening (Bell et al., 2009) driven by in-the-moment accessibility of the word's referent to either the speaker or the listener. Results in this chapter show that speakers' production of words that are mentioned for the first time within a coherent stretch of discourse ('story') are longer than subsequent mentions of the same word within that story or preceding mentions of the word within the preceding telling of the same story. This first-mention lengthening occurs even if the word has previously mentioned for the same listener and is therefore easily accessible to both speaker and listener; therefore, first-mention lengthening has a significant offline component, modeled here as storage and retrieval of richly-specified exemplars associated with various discourse contexts (cf. Pierrehumbert, 2006).

This account makes sense of the relative insensitivity of duration to online information status differences, as documented in Bard et al., (2000), Lam & Watson (2010), and the present results. It also successfully accounts for the finding that first-mention lengthening is greater in magnitude for high-frequency words (Baker & Bradlow, 2009; Bell et al., 2009) using mechanisms already well-established for describing diachronic lexical splits and 'special reduction' (Bybee, 1988; Gahl, 2008).

CHAPTER III

THE REPETITION EFFECT WITH ABSOLUTE SIZE REPRESENTATIONAL GESTURE

1. *Introduction*

In chapter II, the Repetition Effect in speech was examined. Unlike speech, which is highly conventionalized, representational gestures are not at all conventionalized (see McNeill's Continuum 3, section 2.5 in chapter I) and may behave differently from speech when they are repeated. Although they are unconventional by definition, they are not homogenous in the function they serve. Approaching the examination of representational gestures functionally allows a more fine-grained categorization, which could reveal the role they play in discourse.

In this chapter, representational gestures that depict an absolute size, which will be referred to as *absolute size representational gesture*, will be examined. I will demonstrate that although absolute size representational gestures may not be conventionalized by definition, they are resistant to reduction, regardless of the model that speakers operate by, due to the necessary semantic content they carry. In short, they function to encode size information, thus, cannot be spatially reduced in order to preserve the semantic content of size.

2. *The form and function of an absolute size representational gesture*

When speakers refer to an object's size, they can convey the message verbally, for instance, by saying that an object is one foot long or that an object is three inches wide. Speech alone, without a doubt, is sufficient to communicate size information. However, speech is not the only way speakers can convey size information; there are instances

when speakers perform an absolute size representational gesture co-verbally. In cases such as these, speech alone becomes insufficient to communicate size information because the essential semantic content (the size of the object) is *entirely* packaged in the absolute size representational gesture, not in speech. Strictly speaking, both modalities must operate in a synchronized fashion for the semantic content to unfold. Figure 5 and 6 demonstrate an example.

Figure 5: First example of an absolute size representational gesture

In Figure 5, while the speaker says *I caught a fish this big*, she also performs an absolute representational gesture that depicts the fish's size. The verbal description *this big* is incomplete without the accompanying absolute size representational gesture. That is, if she were to say *this big* without gesturing, the message is incomplete since size information has not been encoded (and of course, cannot not be decoded). Along the same rationale, if the speaker gestures without the verbal counterpart *this big*, the message is incomplete since the viewer cannot be certain that the representational gesture

does not have other interpretations. Note that any representational gesture is unconventionalized by definition and cannot standalone, unlike an emblem or a sign in ASL, for instance (see section 2.5 in chapter I); speech is always required for a complete message to unfold.

Figure 6: Second example of an absolute size representational gesture

Suppose that the speaker says *I caught a fish this big* again, but this time, the absolute size representational gesture appears as in Figure 6, instead of Figure 5. Despite the verbal description being identical, the message regarding size is different. In other words, the fish that the speaker referred to in Figure 6 is clearly not the same as the fish that was described in Figure 5; the fish in Figure 6 is a much smaller than the fish in Figure 5.

These two examples emphasize the fact size information is strictly encoded in the absolute size representational gesture: the representational gesture itself reflects the size of the object being described. Since size information is solely embedded in the gesture,

the Repetition Effect should have no impact when absolute size representational gestures are repeated, regardless of the speaker's assumption of the state of knowledge of the listener.

3. *Objective of this chapter*

The objective of this chapter is to demonstrate that absolute size representational gestures do not undergo reduction in repetition due to the essential and necessary semantic content they carry. When speakers make a reference to an object's size in the retellings, absolute size representational gestures will not reduce due to automatization (the Listener-neutral explanation, see section 3.2.2 in chapter I) nor will they reduce due to the speakers' assumed state of knowledge of the listener (Listener-modeling explanation, see section 4.1.2 and 4.1.3 in chapter I). In sum, absolute size representational gestures will remain constant in repetitions because of the function they serve in encoding size information.

3. *Methods*

3.1. *Participants*

Speakers in this study were recruited from the University of Oregon Psychology and Linguistics Human Subject Pool, where they earn one research credit for participating in the experiment. They were undergraduate students, consisting of seven males and sixteen females who self-reported American English as their native language. All twenty-three speakers reported having normal speech and hearing. These participants are the same participants in chapter II, with right additional participants who were recruited after analysis in of word duration had completed.

3.2. *The stimulus*

The stimulus is a short video clip, approximately two minutes in length, of a man making a pizza dough and explaining the process concurrently. As the man works the dough, the size of the dough expands into a larger size. See Figure 7 for an example.

Figure 7: The dough expanding in size

Note that in the stimulus, the man does not describe the size of the dough verbally. In other words, he does not refer to the size of the dough while performing the task (e.g., the dough is 6 inches in diameter). Thus, the stimulus has a potential in eliciting absolute size representational gestures if participants find that dough size is essential information to include in the description.

3.3. *Recording*

All recordings were performed without the presence of the researcher. Speakers were video-recorded during the experiment, where the original MTS files were converted to MP4 files before transcription and coding. Additionally, speakers were also audio-recorded with a Marantz PMD 671 since this is the same set of data used in chapter II.

3.4. *Procedures*

Speakers were told that they would be watching a short video clip of a cooking show. Their goal was to become familiar with the content of the clip so that they could explain it to other undergraduate students who were scheduled for a related experiment. They were told that the amount of time spent watching the stimulus was not being recorded (and it was not) and that they could replay the clip as many times as they wished. They watched the stimulus at least twice on a computer screen while alone. Once they felt that they have a good grasp of the material, they informed the researcher who was waiting in a room nearby.

Speakers explained how a pizza dough is made three times, twice to listener 1, and once to listener 2 (see the sequence of listeners in Table 2 and the rationale of this sequence in section 5.1 in chapter I). Between each telling, there was a ten-minute break where speakers were asked to play a game of Tetris on the computer by themselves. These extended breaks are a conservative choice since they are intended to minimize reduction due to repetition priming from the previous telling. All listeners were confederate listeners who were counter-balanced across speakers (see the use of confederates in section 5.2. in chapter I).

Speakers were debriefed and were also asked what they thought the experiment was about; two most common responses were either it was a memory task or it was on how people talk to men *vs.* women since the listeners were of both genders and the study was conducted in a sociolinguistics lab. They did not suspect it was a gesture study.

3.5. *Predictions for absolute size representational gestures*

All absolute size representational gestures across all conditions (A, B, and C) are predicted to be identical in size due to the rich-semantic content they carry. Reduction due to automatization and assumed state of knowledge of information for the speaker and/or for the listener should have no effect on absolute size representational gestures in repetition.

3.6. *Coding*

First, speech sound was transcribed and segmented into intonation phrases independent of the video. In this study, a pause that exceeds 100 milliseconds designates a new intonation phrase (*cf*. Kendall, 2013). Events of an absolute size representational gesture selected for analysis must occur across all three trials. This is because it is impossible to determine whether an omission is due to the speaker simply forgetting to mention the event or it could have been because the speaker omitted the event due to an assumption about the state of knowledge of the listener.

Once all data points of interests have been extracted, the conditions were omitted for trial-blindness as preparation for coding. The verbal description that coincides with an absolute size representational gesture includes but is not limited to: *this large*, *this size*, and *yea big*. A total of five triplets were converted into a still picture. I examined each triplet qualitatively, comparing the distance from the left hand to the right hand across all tellings. If the two hands are joined, then the space created in between two arms is examined.

4. *Results*

Most participants did not include size information in the description across three tellings; however, five participants did include size information in their description across three tellings. Although triplets of absolute size representational gesture are few and far between in the data, mostly because they did not consistently occur across all three conditions, the triplets can be examined qualitatively.

Results show that in all five triplets captured, when speakers refer to the size of the dough across three tellings, the spatial distance from the left hand to the right hand or the space between the two arms remains constant, thus, reflecting the same semantic content of size in repetitions. In the five cases are exhibited below, absolute size representational gestures do not reduce in size in repetitions, whether the information is being repeated to a new listener (condition B) or to a listener who has heard the description before (condition C). For each participant, the leftmost photo is the first telling (condition A, baseline), the middle photo is the second telling (condition B, to a new listener), and the rightmost photo is the third telling (condition C, to the first listener again).

Figure 8: Participant 4 demonstrating the size of the dough

In the first telling, participant 4 said *once it's about this big*; in the second telling, she said *once it's about, like, yea big*; in the third telling, she said *once it's about, like, dunno, like, this big.* Notice that the space between her hands was roughly the same across all three trials.

Figure 9: Participant 10 demonstrating the size of the dough

The verbal description across three tellings by participant 10 was *once it's, like, this big*. Again, notice that the space between her hands was identical across all three trials; thus, the size of the dough remained constant. It is interesting to note that for this participant, the verbal description in other parts (parts that do not refer to size) in the description across three trials varied, but the verbal description indicating size, along with the absolute size representational gestures, were identical.

Figure 10: Participant 15 demonstrating the size of the dough

In the first telling, participant 15 said *once it's gets to about this size*; in the second telling, she said *once it gets to a certain size*; in the third telling, she said *when it gets to about this size.* The size of the dough, as exhibited via absolute size representational gesture, remained constant.

Figure 11: Participant 18 demonstrating the size of the dough

In the first telling, participant 18 said *once it's gets to be about, I say, about this size*; in the second telling, she said *so maybe, like, about this size*; in the third telling, she

said *until you reach about, like, that size.* Unlike the two previous examples, both hands are joined together to exhibit the size of the dough. The area between her arms appeared identical across all three trials; hence, the size of the dough remained constant.

Figure 12: Participant 21 demonstrating the size of the dough

In the first telling, participant 21 said *pizza that's about this big*; in the second telling, she said *this, like, big of a blob, I guess*; in the third telling, she said *pizza dough about this big*. Again, the size of the dough, as exhibited via absolute size representational gesture, remained constant.

Figure 13: Participant 20 demonstrating the size of the dough

Unlike the previous five cases, participant 20, as shown in Figure 13, used a different kind verbal description. In the first telling, participant 20 said *he* [man in the video clip] *had it to, uh, plateish-size*; in the second telling, he said *he* [man in the video clip] *had it at about plate-size*; in the third telling, he said *he's* [man in the video clip] *gotten to about a plate-size.*

A case such as this could be construed as exhibiting an absolute size representational gesture, in the sense that the participant was gesturing the size of the dough, indicating how big *plateish size* and *plate-size* is. Although similar, this instance is not the same as the rest of the cases presented earlier (from Figure 8 to Figure 12) since the verbal description can stand alone without the absolute size representational gesture for the semantic content to unfold. That is, *plateish-size* and *plate-size* already pack some semantic content (i.e., a plate is assumed to be of a certain size). This is in contrast to saying *this big* or *that size* without the accompanying absolute size representational gesture, where size information is entirely absent. In sum, although this instance is not qualified as an absolute size representational gesture, it has a similar characteristic in being resistant to reduction in repetition. It remained constant throughout the repetitions, even when the verbal counterpart already exhibits some semantic content of size.

5. *Preliminary discussion*

Results in this chapter show that an absolute size representational gesture is resistant to reduction in repetition from automatization (Listener-neutral explanation) or speakers' assumption of the state of knowledge of the listener (Listener-modeling explanation). Such resistance to reduction, I argue, is due to the rich semantic content that

is packaged in the absolute size representational gesture itself. An absolute size representational gesture cannot be reduced, despite the number of repetitions or as a result of the speaker's assumed state of knowledge of the listener, if the speaker is being truthful and precise in communicating size information.

The qualitative findings in this chapter raise an important question on how representational gestures should be approached. Indeed, all representational gestures are unconventionalized by definition (section 2.5 in chapter I); however, they vary in the function they serve; thus, they should not be viewed and treated as a one-size-fits all phenomenon. I argue that *some* kinds of representational gestures have more communicative intention than others; in this case, absolute size representational gestures have a high communicative intention since size information is strictly packaged in gesturing rather than in speech. Since they bear such rich semantic content, they cannot not undergo reduction in repetition since any reduction will alter message. In contrast, other kinds of representational gestures may not be as communicatively-intended and may be more prone to reduction in repetition; these kinds will be discussed in detail in chapter IV and chapter V.

Other studies that gesturing plays a prominent role in conveying size information. Holler & Stevens (2007) reported that the modality, which encodes size information is dependent of the speaker's assumption of the listener's state of knowledge (i.e., common ground). That is, when size information is new information, it is carried out via representational gesture alone or it is carried out via representational gesture and speech; in contrast, if size information is old information, it is carried out via speech alone. Beattie & Shovelton (2005) revealed that when size is considered important information,

it is carried out via gesturing; conversely, when size information is not considered important information, it is carried out via speech.

Although these two studies did not examine absolute size representational gestures (i.e., the representational gestures were not accompanied by the same kind of verbal counterpart), they, too coincide with the findings in this chapter. That is, size information can be carried out via gesturing, be it new information that contains size or important information that contains size. In the current case of an absolute size representational gesture, the representational gesture itself depicts the size of the object and does not reduce spatially in repetition. Perhaps, gestures that convey size should be a class of gestures by itself, a class that has a tendency to be highly communicative, which consists of different sub-types that are categorized by form and function.

6. *Preliminary conclusions*

In this section, I have shown that absolute size representational gestures do not reduce in repetition. The primary function of absolute size representational gestures is to convey size message, thus, they cannot be reduced in repetition. Functionally speaking, absolute size representational gestures have a high communicative intention, possibly more than other kinds of representational gesture, especially those that are not concerned with size information.

CHAPTER IV

THE REPETITION EFFECT ON MIMETIC REPRESENTATIONAL GESTURE

1. *Introduction*

Chapter III discussed a particular kind of representational gesture that is resistant to reduction in repetition. That is, absolute size representational gestures do not reduce under repetition since they function to carry necessary semantic content. In this chapter, *mimetic representational gestures*, a representational gesture that exhibits an action will be examined. Although both kinds of representational gestures are unconventionalized, they behave differently when they are repeated.

The primary objective of this chapter is to demonstrate that unlike absolute size representational gestures, which do not undergo reduction in repetition, mimetic representational gestures undergo reduction in repetition. I will demonstrate that reduction in repetition in mimetic representational gestures is due to automatization (Listener-neutral explanation, section 3.2.2 in chapter I); it is not due to the speakers' sensitivity to the state of knowledge of the listener (Listener-modeling explanation, section 4.1.3 in chapter I). Therefore, I argue that mimetic representational gestures function to serve the speaker in cognitively, rather than having a communicative function.

2. *Function of mimetic representational gestures*

Recall the rabbit hopping out of the room example in section 2.5 (Figure 2) in chapter I. The speaker says *the rabbit is hopping out of the room*, while performing a representational gesture. Now, consider the speaker saying *the rabbit is hopping out of the room*, but the motoric commitment and effort used in gesturing is reduced from the original (as shown in Figure 2).

Figure 14: "The rabbit hopping out the room" gestured with reduced motoric commitment and effort

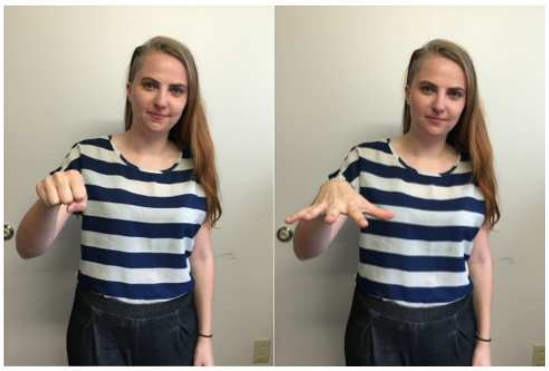
Regardless of the reduced motoric commitment and effort invested by the speaker, the semantic content is still intact, unlike the case of an absolute size representational gesture (see chapter III). To elaborate, if the index finger and the middle finger, representing the rabbit, are flimsy, the verbal counterpart of *rabbit* still functions to indicate the kind of animal being talked about. If the arch in hopping action is smaller, the verbal counterpart of *hopping* functions to indicate the kind of motion being performed. Lastly, if the trajectory of out of the room is not as extended, the verbal counterpart of *out of the room* functions to indicate the direction the rabbit is going. The redundancy of message in both modalities is high; there is much overlapping between content reflected in speech and content reflected in gesturing.

A mimetic representational gesture is a gesture that depicts an action, performed along side speech. This is not a pantomime, so speech is a vital component. In Figure 14, the mimetic representational gesture is the arch motion that represents the action of the rabbit hopping. Compare to Figure 2 and Figure 14, a pantomime would probably exhibit

a whole body movement of hopping, where the rabbit is embodied as the person performing the hopping action.

Another example of a mimetic representational gesture is demonstrated in Figure 15. In this example, only the action alone is depicted; hence, it purely demonstrates a mimetic representational gesture.

Figure 15: "I dropped something" mimetic representational gesture



In this example, the speaker says *I dropped something* while performing a mimetic representational gesture. Similar to the rabbit hopping out of the room example, there is high redundancy of semantic content in speech and in gesturing. Should the speaker perform a reduced mimetic representational gesture, using less motoric commitment and effort, of dropping an object or the speaker does not perform a mimetic representational gesture at all, the message should be comprehensible since all the semantic content necessary is already packaged in speech. As a result, the communicative intention of a mimetic representational gesture is presumably low due to the redundancy.

3. *Objective of this chapter*

The object of this chapter is to demonstrate that mimetic representational gestures undergo reduction in repetition. I will show that speakers invest in less and less motoric commitment and effort as a function of repetition, regardless of their assumed state of knowledge of the listener. Therefore, I argue that mimetic representational gestures are *for the speaker*, rather than bearing a communicative intention.

4. *Methods*

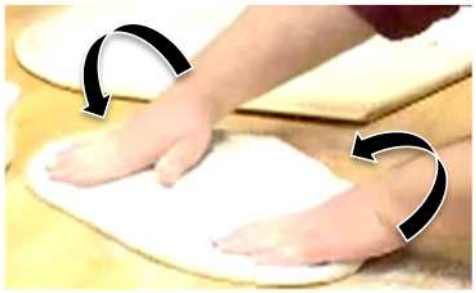
4.1. *Participants*

Speakers in this study were recruited from the University of Oregon Psychology and Linguistics Human Subject Pool, where they earn one research credit for participating in the experiment. They were undergraduate students, consisting of seven males and sixteen females who self-reported American English as their native language. All twenty-three speakers reported having normal speech and hearing. These are the same participants and same set of data presented in chapter II (Repetition Effect in speech) and chapter III (Repetition Effect in absolute size representational gesture).

4.2. *The stimulus*

The stimulus is a short video clip, approximately two minutes in length, of a man making a pizza dough and explaining the process concurrently. It is the same stimulus in chapter II and chapter III. Figure 16 shows an excerpt of the man in the video stretching the pizza dough.

Figure 16: The man stretching the dough



This particular stimulus was selected since it has a high potential in eliciting mimetic representational gesture since it is extremely action-oriented. In this example, the man in the clip is working the dough, stretching it in a circular motion. When speakers explain this process, they are more likely to demonstrate each step gesturally along with speech, demonstrating *how* something is done in detail. If these mimetic representational gestures carry semantic content intended for the listener (communicatively-intended), then they should be *less* prone to reduction in the retellings, especially in the retelling to a new listener (condition B).

4.3. *Recording*

All recordings were performed without the presence of the researcher. Speakers were video-recorded during the experiment, where the original MTS files were converted to MP4 files before transcription and coding. Additionally, speakers were also audio-recorded with a Marantz PMD 671 since this is the same set of data used in chapter II.

4.4. *Procedures*

Speakers were told that they would be watching a short video clip of a cooking show. Their goal was to become familiar with the content of the clip so that they could explain it to other undergraduate students who were scheduled for a related experiment. They were told that the amount of time spent watching the stimulus was not being recorded and that they could replay the clip as many times as they wished. They watched the stimulus at least twice on a computer screen while alone. Once they felt that they have a good grasp of the material, they informed the researcher who was waiting in a room nearby.

Speakers explained how a pizza dough is made three times, twice to listener 1, and once to listener 2. Both listeners were confederate listeners who were counter-balanced across speakers (see use of confederates in section 5.2. in chapter I to tease apart the confound between Listener-modeling and Listener interaction).
Listeners were counter-balanced across speakers. Between each telling, there was a ten-minute break where speakers were asked to play a game of Tetris on the computer by themselves. These extended breaks are a conservative choice since they are intended to minimize reduction due to repetition priming from the previous telling.

First, the speakers told the story to the listener 1, followed by a ten-minute Tetris break. Then, the speakers repeated the same story to listener 2 (a new listener), followed by a ten-minute Tetris break. Lastly, speakers, again, repeated the same story to listener 1. Prior to the last condition (repetition to listener 1 again), speakers were told that the recording equipment malfunctioned in the first telling and that the recording needed to be conducted again. The rationale for this scenario is so that speakers would not be under the

impression that listener 1 did not understand the content of what was told, which could lead to hypergesticulation in the repetition (see also Hoetjes et al., 2014 on the impact of a negative feedback).

Speakers were debriefed and were also asked what they thought the experiment was about; two most common responses were either it was a memory task or it was on how people talk to men *vs.* women since the listeners were of both genders and the study was conducted in a sociolinguistics lab. They did not suspect it was a gesture study.

4.5. *Experimental conditions and hypotheses*

The sequence of listeners used in this study originally appeared in Galati & Brennan (2010, 2013). This sequence of listeners is preferred since it allows teasing apart the confound between reduction due to automatization and reduction due to assumed information status (see section 5.1 in chapter I). A replica of table 2 in chapter I is presented below for convenience.

<table>
<tr><th>Condition</th><th>Speaker production</th><th>Listener exposure</th></tr>
<tr><td>A
Listener 1</td><td>First production</td><td rowspan="2">First hearing</td></tr>
<tr><td>B
Listener 2</td><td rowspan="2">Subsequent production</td></tr>
<tr><td>C
Listener 1</td><td>Subsequent hearing</td></tr>
</table>

In condition A, speakers present the information for the first time, and the listener hears the information for the first time. Since information is new to both interlocutors, we consider condition A to be the baseline. In other words, speakers should show the most motoric commitment and effort in the explanation in this condition, resulting in the biggest gesture size used.

In condition B, speakers repeat the same information for the second time; hence, the information is assumed old to the speaker. However, the same information is assumed new to listener 2 since it is the first time he/she is hearing it. In this condition, the Listener-neutral explanation would predict that because of automatization, speakers will show less motoric commitment and effort in gesturing, resulting in smaller gesture size when compared to condition A. On the other hand, the Listener-modeling explanation would predict gesture size to remain comparable to condition A since speakers are sensitive to the fact that listener 2 has not heard the information before.

In condition C, the speakers repeat the story to listener 1 again. The information is assumed old to both interlocutors. The Listener-neutral explanation would predict the least motoric commitment and effort since the same information is repeated for the third time. In addition, the Listener-modeling explanation would also predict less motoric commitment and effort since listener 1 has already heard the information before in condition A.

4.6. *Coding*

As preparation for representational gesture coding, speech sound was transcribed and segmented into intonation phrases independent of the video. In this study, a pause that exceeds 100 milliseconds designates a new intonation phrase (*cf*. Kendall, 2013).

Navigated only by the transcription and the assistance of audio files, 230 triplets were located. In other words, events that were verbally described across all three conditions were located. Table 8 shows events in a chronological sequence of pizza making that are found in participants' verbal description of the stimuli.

Table 8: Events of a of pizza making verbally described by speakers

Event	Example of linguistic description
1	'You take the dough'
2	'You flour the surface/table'
3	'You put the dough rough side down [on the table]'
4	'You stretch the dough'
5	'Until the dough gets this big'
6	'Then you stretch the dough with your knuckle'
7	'If there's a tear in the dough, you can fix it'
8	'You can fix the dough by folding over'
9	'Press down on the dough with your fingertips'
10	'You stretch it again'
11	'Until it's about this size'
12	'Then you take the, uh, the [pizza peel]'
13	'You flour the pizza peel so the dough won't stick'
14	'Put the dough in the oven'

If a speaker explains that the surface of the table should be floured (event 2), then the verbal explanation of this event must occur in condition A, B, and C to be included in the analysis. The exclusion of data points such as this was because the motivation for omission cannot be determined; it could have been because the speaker simply forgot to mention the event or it could have been because the speaker omitted the event due to an assumption about the knowledge of the listener.

Note that speakers' descriptions of an event need not be identical linguistically across three conditions. The use of lexical items and/or syntactic structure may vary, and indeed they often did, but as long as the utterances describe the same event, they were selected for analysis. All data points of interest from the original files were turned into short video files that only represented the events, where the conditions were omitted for trial-blindness. All triplets were parsed into gesture units and gesture phrases (McNeill, 1992) in ELAN (Sloetjes & Wittenburg, 2008).

Only mimetic representational gestures were selected for analysis in this chapter to keep the data homogenous by function. Since absolute size representational gestures, in (5) and (11), are considered to serve a different function (see chapter III), they were excluded from the analysis. In addition, representational gestures that occur with speech dysfluencies while retrieving a lexeme in (12) were also excluded. These gestures have been argued to be inherently non-communicative; they have been argued to function to assist speakers with recalling the word (Krauss et al., 1995; Krauss et al., 1991). However, whether they successfully assist speakers with lexical retrieval is beyond the scope of this dissertation. In sum, the selection of data points is function-based.

In order to examine reduction in mimetic representational gestures, a way to code reduction magnitude was needed. Mimetic representational gestures were coded categorically on the basis of the set of body parts actively involved in the production of the gesture. The categorical coding decision includes: shoulder (the most motoric effort), elbow, wrist, fingers, or no movement (the least motoric effort). This category of coding is somewhat similar to that in Hoetjes et al. (2015) in the sense that the method here, too, examines body parts executed in gesturing; furthermore, no movement was included as a category. Note that self-adapters (e.g., scratching the face or playing with the hair) are also included in the no movement category as they are presumed to carry no semantic content and to not be generated by the speech production system.

All body parts are paired in all possible combinations of body movements with independent coding for each arm. Therefore, possible combinations range from: [shoulder/shoulder] as the most motoric effort and [no movement/no movement] as the least motoric effort. The order is arranged iconically from [left/right]. All other possible combinations fall in between the most motoric effort and least motoric effort. For instance, a combination of [elbow/shoulder], treated identically to [shoulder/elbow], are both ranked to indicate less motoric effort than [shoulder/shoulder]. When multiple body parts are in motion, the bigger body part, which entails more motoric effort, is always coded to be conservative in the coding decision; for example, if both the wrists and fingers are in motion, the data point is coded as wrists, not fingers.

Body parts in motion are presumed to be a measure of motoric commitment and effort in gesturing: it indicates how much motoric commitment and effort speakers are willing to invest gesturally in the description of the same event. Operating under the

assumption that gestures that are bigger in size require more motoric effort than gestures that are smaller in size, gesture size can be interpreted as revealing motoric commitment. The dependent variables are either: (1) *a decrease* in body parts combination in motion from the previous condition or (2) *a no decrease* in body parts combination in motion from the previous condition.

4.7. *Predictors for mimetic representational gesture size and analysis*

Data were analyzed using linear mixed effects models in the binomial family in R (R Development Core Team, 2014) with the lme4 package (Bates, 2013), with condition as a fixed effect, a random intercept for subject and a random slope for condition within subject. P values reported below were based on the *z* approximation. To address the concern that *p* values-based on the *z* statistics for mixed effects model may be anti-conservative, the *p* values were also checked by comparing models with and without the predictor of interest, using a log likelihood test. The results show no qualitative difference to the *z*-based estimates reported below.

5. *Results and preliminary discussion*

The binomial test showed a steady decline in body parts in motion as a function of repetition across condition A, B, and C. In the mosaic plots, the light gray area shows that a decrease has occurred from the previous condition; whereas, the dark gray area shows that a decrease has not occurred, meaning motoric commitment remained the same from the previous condition or increased from the previous condition. Bars in the mosaic plots are arranged according to size of body parts combination, from largest (to the left of the plot) to smallest (to the right of the plot), with the exclusion of no movement in both hands, for such category has reached floor in physical movement and is impossible to

reduce any further. Note that in the mosaic plots, *s* refers to shoulder, *e* refers to elbow, *w* refers to wrist, *f* refers to fingers, and *0* refers to no movement. Each combination pair is arranged from left to right; for example, *w0* refers to left wrist and no movement on the right.

Figure 17: Change in body parts in motion from condition A to condition B

Figure 17 shows a decline in body parts in motion from condition A to condition B. The width of each bar indicates the proportion of the combination of body parts in the data. Gestures in condition B involve less motoric commitment than in gestures in condition A 82 percent of the time. Overall, the light grey area is significantly greater than 50 percent of the total area: $b = 2.30$, $se(b) = 0.25$, $z = 9.40$, $p < .05$.

Figure 18: Change in body parts in motion from condition B to condition C

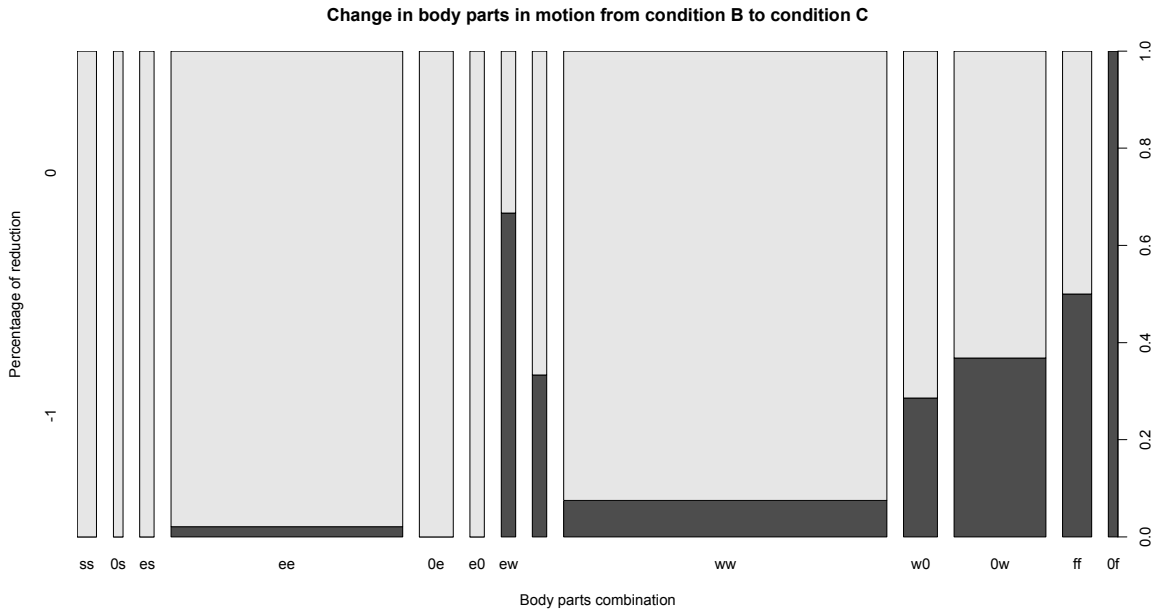
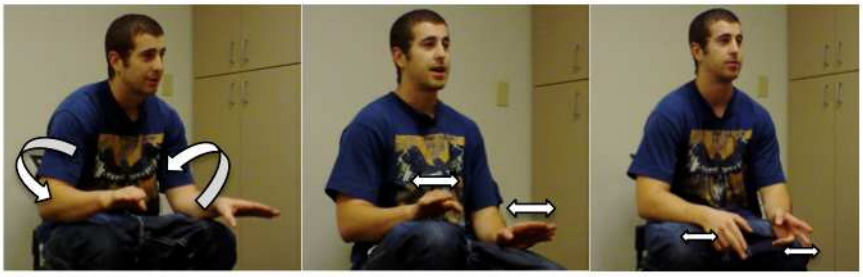


Figure 18 shows a decline in body parts in motion from condition B to condition C. The width of each bar indicates the proportion of the combination of body parts in the data. Gestures in condition C involve less motoric commitment than in gestures in condition B 71 percent of the time. Again, the light grey area is significantly greater than 50 percent of the total area: $b = 2.58$, $se(b) = 0.37$, $z = 6.99$, $p < .05$.

In both mosaic plots, notice that a decrease prominently occurs with bigger body part combinations (to the left of the mosaic plots), and no decrease prominently occurs with smaller body parts (to the right of the mosaic plots). This trend is to be expected as bigger body parts have more opportunity for a decrease; in contrast, smaller body parts have less opportunity for a decrease, for they are approaching floor in physical movement.

To better demonstrate the result, figure 19 exemplifies a case of a continued decrease in motoric commitment as a function of repetition. In this example, the speaker is explaining event 10 (see table 8) of the pizza-making process where the stretching of the dough takes place. The more times this speaker repeats the same information regardless of the assumed knowledge of the listener, the less motoric commitment is utilized, as shown by smaller body parts in motion.

Figure 19: An example of a decrease in motoric commitment as a function of repetition



Condition A Condition B Condition C

In condition A, the speaker used two elbows to demonstrate the action of stretching the dough, then in condition B, both elbows are still and only two wrists are used when re-telling stretching of the dough to a new listener. Lastly, in condition C, when the speaker was retelling to the first listener (the same person as in condition A), his gesturing has decreased to wiggling fingers on both hands.

In sum, speakers are largely insensitive to the listener's assumed state of knowledge. They continually invest in less and less motoric commitment and gestural effort regardless of the assumed state of knowledge of the listener. The results can be

interpreted that mimetic representational gestures are performed for the speaker, rather for the listener since they speakers do not appear to be modeling the listener's assumed state of knowledge.

6. *A perceptual judgment follow-up study*

In Galati & Brennan (2013), two coders, the first author and another researcher, rate the space in between two hands to determine a repeated gesture's size. In Hoetjes et al. (2014), however, the authors use body parts to determine a repeated gesture's size. Although the coding criteria used in this chapter are somewhat similar to Hoetjes et al. (2014) in the sense that body parts used in gesturing are observed, the current method of coding differs from theirs in several aspects (see section 4.6 in this chapter). Since the coding method is somewhat novel, a perceptual judgment follow-up by naïve raters allows justification of the coding decisions; importantly, it can demonstrate whether the coding parameters correspond to behavior observable and interpretable by the observer in real time.

6.1. *Participants and the stimuli*

A total of nine raters, seven males and two females, participated in rating of participant gestural behavior. Triplets of the data were chosen randomly and presented in E-Prime (Schneider, Eschman, & Zuccolotto, 2012). Importantly, only the speakers were visible to the raters; that is, the confederates were taken out of the frame. The stimuli were presented without audio.

6.2. *Procedures*

Raters judged 112 triplets, which are half of all triplets in the data. Without knowledge of the condition, raters were presented with pairs of videos and asked to judge

which of the two exhibits more gestural effort. Unbeknownst to the raters, the paired clips corresponded to gestures about the same event from two different adjacent tellings (A & B, B & A, B & C or C & B). Tellings were counterbalanced across pairs, so the video from the earlier telling appeared on the left side of the screen half the time. Order of video pairs was randomized. Raters were able to see the speakers only; confederate listeners were excluded from the frame. Raters were able to view each comparison once; reaction time was not recorded.

Participants were instructed to judge gestural effort on the basis of a set of cues for detecting more gestural effort. They are: body parts used in the action (bigger body parts suggest a more effortful action than smaller body parts), the width of the action (a wider gesture suggests a more effortful action than a narrower gesture), the height of the action (a higher gesture suggests a more effortful action than a lower gesture), and the force of the action (a more forceful gesture suggests a more effortful action than a less forceful gesture). Furthermore, all self-adapters (e.g., playing with hair or scratching the face) were asked to be judged as less effortful as they presumably contain no semantic content.

6.3. *Predictors for gestural effort and analysis*

The binary relative size judgment data were analyzed using linear mixed effects models in the binomial family in R (R Development Core Team, 2014) with the lme4 package (Bates, 2013), with raters and video triplets as random intercepts. Since the main interest is in whether the relative judgments are significantly different from choosing the two conditions with chance probabilities (50/50), there is no fixed effects predictor in the models reported.

6.4. *Results and preliminary discussion*

Raters judged gestures in condition A to exhibit more gestural effort than gestures in condition B 68 percent of the time: $b = 1.06$, $se(b) = 0.19$, $z = 5.42$, $p < .05$. In the same direction, they judged gestures from condition B to exhibit more effort than gestures from condition C 69 percent of the time: $b = 1.05$, $se(b) = 0.24$, $z = 4.45$, $p < .05$.

The results indicate that naïve raters can perceptually detect a reduced gestural effort as a function of repetition. That is, raters observed a constant decline in repeated gestures across triplets, from condition A to condition B and from condition B to condition C. This perceptual judgment provides additional support for the novel coding scheme used in this study. The coding provides details about the motoric behavior of the speaker, and these details align with the subjective impression of naïve viewers.

7. *Preliminary general discussion*

The goal of this chapter is to demonstrate that mimetic representational gestures decrease in size as a simple function of repetition, and not in response to the assumed state of knowledge of the listener. In addition, gesture reduction is perceptually detectable to naïve raters. Three plausible explanations that do not contradict are offered.

An appealing explanation for the pattern shown in the result is that automatization causes reduction; that is, reduction is caused by an automatized process in motor planning that is rooted in the speaker's internal production system (Bybee, 2001, 2002). The more times a motoric action is performed, the more it will reduce. Without a need for hypergesticulation (e.g., for clarification or emphasis), reduction could continue until it reaches floor. In the case of gesturing, floor could be a complete lack of gesturing.

Articulators in gesturing are larger in size than in speech and may reduce all the way to no movement; the scale of reduction is more observable and prominent.

Another explanation, which does not contradict the previous explanation, is that speakers require less assistance from gestures in the retellings due to a decrease in cognitive demand. Kita (2000) proposes that gestures can help speaking, that "speakers are cognitively compelled to produce more representational gestures if the spatio-motoric event to be verbally conveyed does not readily lend itself to linguistic structuring […] (p. 181). Operating under Kita's model, it is plausible that when the cognitive demand is high (in the first and novel telling), speakers utilize more gesturing to assist with speech production. However, in the retellings when the cognitive demand is lower due to greater accessibility (see Ariel, 1988, 1980) or rehearsal (see Chawla & Krauss, 1994), the value of gesturing diminishes, resulting in reduction. The relationship between cognitive demand and reduction in representational gesture has been made apparent by Masson-Carro et al. (2014), who report that when there is a higher cognitive demand, there appears to be less reduction in a repeated gesture. In this case, repeating the same information leads to less cognitive demand; thus, speakers' gestures reduce.

Unlike Kita's (2000) proposal, which focuses on how representational gestures assist speaker in packaging information and how representational gestures intertwine with speech, Hostetter & Alibali's (2008) Gesture as Simulated Action (GSA) framework focuses more on more general cognitive processes and how representational gestures are a product of such processes. In their framework, multiple factors influence the production of a representational gesture: one of which is the strength of activation of the simulated action. They explained that, "simulated action may be evoked when the task at

hand involved imagining […] when motion is a part of the original perception being stimulated" (p. 503).

Operating under this idea, after having been exposed to the stimuli which is heavily action-oriented (e.g., stretching, kneading, and folding), speakers may be imagining that they, themselves, are working on the dough, which results in mimetic representational gestures found in the data. In the first telling, the simulated action is the most evoked; then in the later tellings, at least 10 minutes in the second telling, and at least 20 minutes in the last telling (see section 4.4 in this chapter for experimental procedures), the simulation becomes much less evoked; hence, less motoric commitment is used in enactment of the action. In most cases, mimetic representational gesture do not disappear in the retellings, which could be inferred based on this framework that the mental simulation is still activated; it is a matter of the degree of activation.

The three explanations offered here do not contradict theoretically. Whether it is reduction due to automatization, reduction due to a lower cognitive load, or reduction due to a lesser degree of mental simulation, these explanations point to the fact that mimetic representational gestures are Listener-neutral; they are rooted in the speaker's production system, intended *for the speaker* without any inherent reference to the listener. If these mimetic representational gestures were intended to be communicative and produced *for the listener*, then there should not have been a reduction in motoric commitment and effort in condition B since the information is presumed new to the listener.

9. *Preliminary conclusions*

In this chapter, I have shown that mimetic representational gestures reduce as a function of repetition, not as a function of the speakers' assumed knowledge of the

listener. Results in this chapter are consistent with the Listener-neutral explanation; speakers utilize mimetic representational gestures to benefit their own cognitive processes. Such gestures reduce when they become less beneficial to the speakers themselves.

CHAPTER V

THE REPETITION EFFECT ON REPRESENTATIONAL GESTURES ASSOCIATED WITH LEXICAL ACCESS

1. *Introduction*

Two kinds of representational gesture were discussed in chapter III and chapter IV. I have shown that despite sharing the characteristic of being unconventionalized, absolute size representational gestures are resistant to reduction in repetition, but mimetic representational gestures are prone to reduction in repetition. Based on these results, I argued that the different patterns are motivated by the function each kind of representational gesture serves. Absolute size representational gestures function to convey size information; therefore, they must remain the same in repetition if the speaker wishes to accurately covey size information. Mimetic representational gestures, on the other hand, function to assist speakers cognitively; they reduce in repetition since they presumably provide less benefit to the speaker.

The third kind of gesture, which will be referred to as *representational gestures associated with lexical access*, co-occurs with dysfluent speech. The primary function, as Krauss et al. (1996) and Krauss et al. (2000) argued, is to assist speakers cognitively in the process of retrieving a lexeme. Since the function is solely intended for the speaker, for a very restricted purpose of lexical retrieval, they are inherently non-communicative. Until now, there has yet to be a study that examines under repetition; it has not been documented whether these gestures will undergo reduction when they are repeated (similar to mimetic representational gesture, see chapter IV); or they will be resistant to

reduction when they are repeated (similar to absolute size representational gesture, see chapter III).

2. *Selected literature on lexical retrieval and gestures*

The Listener-neutral explanation (see section 3.2. in chapter I) contends that representational gestures are meant *for the speaker* and function to serve the speaker cognitively, arguably to assist in speech production. Although Graham & Heywood (1995) found that prohibiting speakers from gesturing does not affect speech production, others have presented evidence to suggest otherwise. Rauscher, Krauss, & Chen (1996), for instance, showed that when speakers are prohibited from gesturing, their speech becomes less fluent; along the same line, Hostetter, Alibali, & Kita (2007) showed that linguistic description of an event is more semantically-rich when gestures are not restricted. These studies suggest that restricting speakers from gesturing leads to a poorer performance in speech; conversely, allowing speakers to gesture leads to a better performance in speech.

Gestures have been shown to lend a hand when speakers encounter difficulties in retrieving a lexeme. Morsella & Krauss (2004) showed that the rate of gesturing is a function of lexical access difficulty; that is, the more difficult the word, the higher rate of gesturing is found, and vice versa. The benefits of gesturing are noteworthy when speakers encounter the tip-of-the-tongue state (ToT), a state in which speakers fail to retrieve a word. For instance, prohibiting speakers from gesturing results in a poor outcome in resolving the ToT state (Frick-Hornsbury & Guttentag, 1998, although see also Beattie & Caughlan, 1999 for a contradicting result). Not only adults have shown to benefit from gesturing when they are in the ToT state; children, too, have been found to

better resolve the tip-of-the-tongue state when representational gestures are not prohibited (Pine, Bird, & Kirk, 2007).

The Lexical Retrieval Hypothesis, as proposed by Krauss et al. (2000), maintains that one of the functions of representational gestures is to assist speakers in retrieving a lexeme. The model spells out the cognitive architecture of how these representational gestures function to assist the speaker. Their model based on Levelt's (1992) model of speech production, which is a cascading model that begins with the conceptualizer and ends with the articulator. Representational gestures facilitate speech production at the lemma level, the stage where grammatical encoding occurs, through a *kinesic monitor*, a cross-model priming, to aid speakers in lexical retrieval. Once the word is retrieved and the speaker utters the word, representational gestures are terminated since the goal of lexical retrieval has been achieved (Krauss et al., 2000, see Figure 13.1, p. 267). It must be emphasized that according to Kraus et al. (2000) these gestures are not intended for the listener, for their sole function is a assist the speaker (see 3.2 in chapter I for a more detailed discussion).

3. *Objectives of this chapter*

In this chapter, I will demonstrate that representational gestures associated with lexical access, like mimetic representational gestures (see chapter IV), undergo reduction in repetition. Regardless of whether retrieval difficulties have been resolved, speakers' motoric commitment and effort reduces in repetition. Hence, I argue that reduction in representational gesture in repetition is an automatized process that operates across multiple functions.

4. *Methods*

4.1. *Participants & stimuli*

Eight naïve raters, five males and three females, were recruited for this rating task. They were students in an introductory psycholinguistic class, participating for extra credit. Navigated solely by the transcription, segments where participants verbally refer to *the pizza peel* (see Table 8, sequence 12 in chapter IV). Figure 20 is an example of the object. When participants refer to this object across three tellings, they were selected and converted into individual video files. The triplets were chosen randomly and presented to naïve raters in E-Prime without audio (Schneider, Eschman, & Zuccolotto, 2012). Note that participants appearing in the triplets need not refer to the object as the *pizza peel*. Any linguistic description that points to the object suffices.

Figure 20: The pizza peel in the stimuli

4.2. *Procedures*

Raters judged 21 triplets of the pizza peel data. Without knowledge of the condition, raters were presented with pairs of videos and asked to judge which of the two exhibits more gestural effort. Unbeknownst to the raters, the paired clips corresponded to gestures about the same event from two different adjacent tellings (A & B, B & A, B &

C, C & B, A & C, or C & A). Tellings were counterbalanced across pairs, so the video from the earlier telling appeared on the left side of the screen half the time. Order of video pairs was randomized. Raters were able to see the speakers only; confederate listeners were excluded from the frame. Raters were able to view each comparison once; reaction time was not recorded.

Participants were instructed to judge gestural effort on the basis of a set of cues for detecting more gestural effort. They are: body parts used in the action (bigger body parts suggest a more effortful action than smaller body parts), the width of the action (a wider gesture suggests a more effortful action than a narrower gesture), the height of the action (a higher gesture suggests a more effortful action than a lower gesture), and the force of the action (a more forceful gesture suggests a more effortful action than a less forceful gesture). Furthermore, all self-adapters (e.g., playing with hair or scratching the face) were asked to be judged as less effortful as they presumably contain no semantic content.

4.3. *Predictors for gestural effort and analysis*

The binary relative size judgment data were analyzed using linear mixed effects models in the binomial family in R (R Development Core Team, 2014) with the lme4 package (Bates, 2013), with raters and video triplets as random intercepts. Since the main interest is in whether the relative judgments are significantly different from choosing the two conditions with chance probabilities (50/50), there is no fixed effects predictor in the models reported.

4.4. *Results and preliminary discussion*

Raters judged gestures in condition A to exhibit more gestural effort than gestures in condition B 74 percent of the time: $b = 1.61$, $se(b) = 0.50$, $z = 3.22$, $p < .05$. In the same direction, they judged gestures from condition A to exhibit more effort than gestures from condition C 76 percent of the time: $b = 1.67$, $se(b) = 0.51$, $z = 3.28$, $p < .05$. However, gestures in condition B were not found to exhibit more gestural effort than gestures in condition C: $b = 0.52$, $se(b) = 0.43$, $z = 1.20$, $p > .05$.

4.4.1. *Examples of lexical retrieval difficulties and gesturing*

To better illustrate what raters observed in the stimuli, this section provides examples that show a decrease in motoric commitment and effort in the repetition, regardless of whether the lexical retrieval difficulties have been resolved in the retelling. Terms used to describe speakers' gestures are from Streeck (2008) depiction methods, where *bounding* is defined as "practices involving relative positioning […] of the hands […] Sometimes the hands are configured as if they were in contact with the object's boundaries (sides or edges) […] (p. 292). In addition, *handling* is defined as "objects are indirectly represented by a schematic act that "goes with them", in a way in which turning goes with keys and wielding with sticks […] (Streeck, 2008, p. 293).

The following three figures, from Figure 21 to Figure 23, demonstrate a case when speakers encounter lexical retrieval difficulties in the first telling, as indicated by disfluencies in speech; then, in the later tellings, the difficulties were resolved, as indicated by fluent speech.

Figure 21: An example of lexical retrieval difficulties resolved, first telling

In Figure 21, the first telling, the speaker said *flour the* <pause> *wooden pizza* <pause> *pan*. Lexical retrieval difficulties are indicated by the filled pauses. The handling depiction preceded the verbal counterpart *wooden*, where the speaker was depicting handling an object, as in holding the pizza peel. The bounding depiction coincides with the rest of the utterance, where his hands outline the edge of the pizza peel. Two depictions of representational gestures were present while the speaker was encountering lexical access difficulties, verbally struggling to name the object. These representational gestures are presumed to function in assisting the speaker while retrieving the lexeme.

Figure 22: An example of lexical retrieval difficulties resolved, second telling

In Figure 22, the second telling, he said *you got a wooden pizza pan*. The speaker did not exhibit any disfluency in speech; thus, there was not a lexical retrieval difficulty. The only representational gesture coinciding with speech is the bounding depiction, a representational gesture that was also used in the first telling.

Notice that when speech is fluent, there is a reduction, from handling and bounding to bounding only.

Figure 23: An example of lexical retrieval difficulties resolved, third telling

In Figure 23, the third telling, he said *put it on the wooden pizza pan*. Again, the speaker did not exhibit any disfluency in speech, and the representational gesture used was also bounding. The third telling is similar in motoric commitment and effort to the second telling, but it is different from the first telling.

In contrast to the last speaker, the next speaker encountered lexical retrieval difficulties repeatedly. In other words, every time he referred to the object, there was a clear disfluency in his speech, suggesting lexical retrieval difficulties. Figure 24 to Figure 26 demonstrates such case.

Figure 24: An example of repeated lexical retrieval difficulties, first telling

In Figure 24, the first telling, the speaker said *he had* <pause>, *pallet, uh, thing that thing where you put pizza in the oven*. Lexical retrieval difficulties were indicated by disfluencies in speech, consisting a filled pause preceding *pallet* and the filler *uh*. The bounding depiction (leftmost photo) coincides with the verbal counterpart *pallet.* The bounding depiction in this case is the speaker outlining the bottom of the pizza peel. The handling depiction (middle and rightmost photo) occurred repeatedly, in three strokes, until the end of the utterance.

Figure 25: An example of repeated lexical retrieval difficulties, second telling

In Figure 25, the second telling, the speaker said *on the* <pause> *wooden, um,* <pause> *thing that you take out of the oven*. In this retelling, speech disfluencies still appeared in the form of filled pauses before *wooden* and *thing*; the filler, *um*, was also used. The depiction that occurred alongside disfluent speech, however, has been reduced to handling only, despite the recurring retrieval difficulty.

Figure 26: An example of repeated lexical retrieval difficulties, third telling

In Figure 26, the third telling, the speaker said *on the wooden* <pause> *scoop* <pause> *thing*. Speech disfluencies still appeared in the form of filled pauses before

scoop and *thing*. The depiction that occurred alongside disfluent speech was handling, similar to the second telling but reduced from the first telling.

5. *Preliminary discussion*

The purpose of this chapter is not to evaluate the Lexical Retrieval Hypothesis (Krauss, 2000) nor is it to test whether representational gestures effectively facilitate lexical retrieval. Rather, it aims to observe representational gestures that have been argued to function in assisting speakers with lexical retrieval difficulties in the Repetition Effect. One might expect that like absolute size representational gestures (see chapter III), lexical access representational gestures would be resistant to reduction by virtue of their presumed function. However, it appears that regardless of their presumed function in helping speakers retrieve a lexeme, speakers show a decrease in motoric commitment and effort from the first telling to the second telling.

Figure 21 to Figure 23 show a pattern that is to be expected. When speakers are encountered with lexical retrieval difficulties (in this data is the first telling), they utilize representational gestures that are intended to aid lexical retrieval, as Krauss (2000) proposed. That there are more representational gestures when speakers have difficulties accessing a word is anticipated, considering the presumed function of the gesture. The finding in this chapter is similar to that of Morsella & Krauss (2004), where speakers gestured at a higher rate when the word they tried to access was more difficult. In sum, when speakers have trouble accessing a lexeme, representational gestures become prominent, which could be measured by means of motoric commitment and effort, as presented in this chapter, or in rate, as reported by Morsella & Krauss (2004).

Figure 24 to Figure 26, however, show an unexpected pattern. Operating under the assumption that representational gestures facilitate speakers when they encounter lexical retrieval difficulties, there should *not* be a reduction in repetition, considering speakers still require assistance from such gestures to retrieve the lexeme. In other words, there should not have been a decrease in representational gestures in the second telling, compared to the first, because lexical retrieval difficulties still persist.

Automatization is one explanation for the Repetition Effect found in lexical retrieval representational gestures. Bybee (2001, 2002) proposed that any repeated motoric action reduces when it is repeated. To Bybee, the automatized motoric action is articulation in speech (see chapter I, section 3.1.1.); in this case, the motoric action that is automatized is gesture. Importantly, reduction in gesture caused by automatization seems to occur across multiple gesture functions. That is, whether representational gestures are used to assist in conceptualization (see chapter IV) or whether they are used to assist in lexical retrieval (the current chapter), if their function is presumed to be *for the speaker to benefit the speaker*, then they are likely to undergo reduction in repetition.

6. *Preliminary conclusions*

In this chapter, I have exemplified and discussed the Repetition Effects in representational gestures associated with lexical access. Similar to mimetic gestures, gestures also reduce in motoric commitment and effort when they are repeated, even when speakers still face retrieval difficulties. The Repetition Effect, thus, appears to be a much more general phenomenon that occurs across representational gesture functions, especially when the function is presumed to serve the speaker cognitively.

CHAPTER VI

GENERAL DISCUSSION

1. *Introduction*

Although the Repetition Effect is observed in both speech and representational gesture, its behavior is specific to modality. In speech, the Repetition Effect only occurs within a coherent stretch of discourse, where first mention of a word is always longer than the subsequent mentions regardless of the speakers' assumption of the state of knowledge of the listener. That the first-mention of a word within a coherent stretch of discourse is always longer than its subsequent counterparts is accounted for by the Exemplar model and conventionalization of discourse structure (see chapter II).

The Repetition Effect in representational gestures, on the other hand, varies largely depending on the function the representational gesture is intended to serve. If a representational gesture is prone to benefit the speaker cognitively, then it will undergo reduction in repetition. Specifically, mimetic representational gestures and representational gestures associated with lexical access, two kinds of representational gesture that have been argued to benefit then speakers cognitively, undergo reduction in repetition (see chapter IV and chapter V). In contrast, if a representational gesture has a communicative function, then it may be resistant to the Repetition Effect, as shown with absolute size representational gesture (see chapter III).

Despite having demonstrated how and the extent to which the Repetition Effect manifests itself in two different modalities, the crux of the dissertation is not necessarily about the Repetition Effect. The Repetition Effect is essentially an indicator about the

nature of speakers' production; it is an indicator, which allows inference about speakers' cognitive mechanism and how speakers communicate information.

2. *Methodological improvements*

The experiments in this dissertation are a product of utilizing an existing design while improving it for a tighter control, which results in a more precise interpretation of results, both in this study and previous studies. The experiments consist of using a sequence of listeners to tease apart a confound between new and old/given information between the speaker and the listener; the sequence of listeners originally appeared in Galati & Brennan (2010, 2013) (see section 5.1 in chapter I). In addition to the sequence of listeners, these listeners were confederates who provided minimal linguistic and non-linguistic feedback to the speaker in order to tease apart a confound between listener-interaction from listener-modeling (see section 5.2 in chapter I). By combining these elements, two confounds were eliminated, yielding results that differ from past literature. This section elaborates the importance of the current experimental design and the contribution it has made to the literature.

2.1. *Multiple listeners required: the notion of information status disentangled*

Several studies in speech and gesture position their results on the Repetition Effect as motivated by information status (see, for example, section 3.1.3 on prosodic prominence and section 4.1.2. on representational gesture in chapter I). Indeed, information status plays a crucial role in language production; however, it is premature to conclude that information status is responsible for the Repetition Effect when only a single listener is present. Table 1 and section 5.1 in chapter I highlights the confound when a single listener is used to elicit the Repetition Effect. To restate, by having only

one listener, it cannot be determined whether the reduction in repeated referents is driven by automatization, something that occurs automatically due to human physiology, which is consistent with the Listener-neutral explanation (see section 3.1.1. in chapter I on speech and section 3.2.2 in chapter I on representational gesture) or the reduction in repeated referents is driven by speakers' modeling of the assumed state of knowledge of the listener, which is consistent with the Listener-modeling explanation (see Fowler & Housum, 1997 and 1988, for example).

Section 5.1. in chapter II spells out how Galati & Brennan's (2010, 2013) experimental design disentangles the confound between automatization and listener-modeling. Precisely, their design creates a mismatch between assumed state of information for the speaker and for the listener, allowing a precision in pin pointing *to whom* the state of information applies to and the underlying motivation that is responsible for guiding speaker behavior. Indisputably, their experimental design creates a condition that essentially teases apart the said confound. Any study that intends to investigate the role of assumed state of information in discourse should utilize this particular design to avoid conflating the Repetition Effect deriving from automatization and the Repetition Effect deriving from listener-modeling.

2.2. *Minimal feedback: the notion of listener-sensitivity re-examined*

Past literature argues that speakers show listener-sensitivity by means of modeling the state of knowledge of the listener. Such argument has been supported by evidence elicited from the Repetition Effect (see chapter I for a full review, section 4.1 for speech and section 4.2 for gesture). Without a doubt, I agree that speakers may model the state of knowledge of the listener and that their production could be a reflection of such

sensitivity. The concern here, however, is whether the data that were used in past literature are compatible with and appropriate for the argument. In this dissertation, I propose that the notion of listener-sensitivity should be construed in two mutually exclusive ways: listener-sensitivity deriving from an on-going and dynamic interaction between interlocutors (the Listener-interactive explanation, see section 2.4. in chapter I) and listener-sensitivity deriving from the speakers modeling the state of knowledge of the listener (the Listener-modeling explanation, see section 2.3. in chapter I).

The former interpretation, the Listener-interactive explanation, is rooted in interaction between two people, where both linguistic and non-linguistic cues act in concert, constantly changing (see Clark & Schaefer, 1989 on contribution in discourse and Goodwin, 1981 on interaction between speakers and hearers). For example, if a speaker says something that the listener fails to understand, it is likely that the listener will cue the speaker that a clarification is needed. The listener could demand for a clarification verbally (e.g., asking *what do you mean*?), or the listener could show non-verbal cues (e.g., tilting the head or making a puzzled face). Attentive speakers are sensitive to such cues and will adjust their performance, in speech and gesture, to meet the listener's needs (see Horton & Keysar's 1996 on the Monitoring and Adjustment model). The fluid exchange of verbal and/or non-verbal information, which drives the interaction, is key to this interpretation of listener-sensitivity.

On the other hand, the latter interpretation of listener-sensitivity, the Listener-modeling explanation, is rooted purely in the speaker's assumption of the listener's state of knowledge (see also Horton & Keysar's 1996 on the Initial Design model). A simple example would be one of the functions of the definite article *the* in English (Du Bois,

1980). For instance, a speaker may say to her friend *I will bring curry to the party*; in this case, by using the definite article *the*, she assumes that the listener shares the knowledge that there will be a party, and this presumption is reflected via the grammar of English. Indeed, if her assumption of the listener's state of knowledge is false, the listener is likely to respond with an utterance such as *What party?* or perhaps, the listener may make a puzzled expression as a response. Note that with anticipatory modeling, the decision to use the English definite article *the* can be made in the absence of feedback.

The act of responding to listener's linguistic and non-linguistic behavior (listener-interactive) and the act of modeling the state of knowledge of the listener (listener-modeling) are vastly different and should not be conflated, yet they were in past literature (see section 4.1.3 and section 5.2 in chapter I). In order to tease apart listener-interactive and listener-modeling, production data in this dissertation were elicited with confederate listeners in a more controlled environment, which in contrast to experimental conditions in past literature.

In theory, to rigorously test the notion of listener-modeling, there should be a complete absence of feedback from the listener, both linguistic and non-linguistic. However, in order to create such situation, ecological validity of a face-to-face environment would have to be sacrificed (i.e., it is rarely the case that the listener will show absolutely no feedback while engaging with the speaker). As a result, to maintain as much ecological validity as possible while aiming to test the notion of listener-modeling, the experiments in this dissertation were conducted with confederate listeners who provided minimal and unvarying feedback. Confederates were instructed to nod, maintain eye contact, and say "uh-huh" in all conditions; doing so allows minimization of

possible varying feedback that are likely to lead speakers to respond to "what the listener does", instead of speaker production driven by their assumption of "what the listener knows". Ultimately, the environment created in the experiments in this dissertation should be considered a reflection of speakers' behavior when an interaction goes smoothly, without varying input from the listener. When both linguistic and non-linguistic feedback is held constant, it results in behavior that is in contrast to past literature, which allowed varying feedback.

Results indicate that when speakers mostly rely on their assumption of the state of knowledge of the listener, their performance lacks sensitivity to the listener via modeling the state of knowledge of the listener. A pattern in speech production, word duration to be precise, emerges: hyperarticulation occurs in the first mention and hypoarticulation occurs in the subsequent mentions within the same stretch of coherent discourse. Likewise, a pattern in representational gesture production also emerges; motoric commitment and effort used in gesturing steadily decreases as a function of repetition. These findings are different from findings in past literature that allowed, if not encouraged, interlocutors to exchange feedback freely.

The findings in this dissertation can be thought of as the baseline of speakers' production. When variables contributed by the listener (i.e., linguistic and non-linguistic feedback) are held constant and kept to a minimum, speakers will behave in this default manner, which is interpretable as showing a general lack of listener-modeling. However, if an interaction is allowed, as past literature have shown, then speakers will respond to the interaction and show sensitivity to the listener by adjusting their performance accordingly. For instance, speakers may hyperarticulate or hypergesticulate repeated

mentions if they receive a verbal and/or non-verbal cue from the listener. Since confederate listeners in the experiments in this dissertation provided no such overt cues, there is no motivation for speakers to invest more energy than they are required to in production of repeated referents.

It should be emphasized that although results in this dissertation differ from results presented in past literature, they do not fundamentally contradict. In fact, results presented here help strengthen findings in past literature that speakers are, indeed, sensitive to the listener. However, I contend that speakers are sensitive to verbal and non-verbal cues signaled by the listener during an interaction, instead of modeling of the listener's state of knowledge.

Galati & Brennan (2010, 2013) proposed the "one-bit" model built on listener modeling, to account for listener-sensitivity found in their speech and gesture data (see section 4.1.3. in chapter I), while using data that were elicited from an interactive environment. Based on the findings in this dissertation, I strongly suggest a reinterpretation of their results; in other words, listener-sensitivity found in their data should be interpreted as having derived from an interaction, rather than the modeling of the listener's state of knowledge, the "one-bit" model. I predict that if they replicate their experiment but eliminate informative listener feedback, they, too, will find an absence in listener-sensitivity.

Similarly, the foundation of Gerwing & Bavelas's (2004) and Holler & Stevens' (2007) gesture work is based on the notion of common ground, a notion that is often interpreted to coincide with the Listener-modeling explanation in the sense that common ground is the speakers' assumption of knowledge between them and the listener *prior to*

the interaction (see section 2.2 in chapter I). However, an alternative interpretation of common ground is that it is something that is built during an interaction, an establishment of shared *specified* knowledge, which is negotiated and built throughout the course of an interaction via verbal and non-verbal behavior. This view is similar to that of Gerwing & Bavelas (2013) who point out that "[interlocutors'] accumulation of common ground over the course of the dialogue similarly influenced the form of their gestures" (p. 833). Therefore, the crucial element is not necessarily what interlocutors *assume prior* to the interaction, the classic sense of common ground, but it is what interlocutors *do during* the interaction that brings about listener-sensitivity. Thus, I argue that it was not common ground per se that drove listener-sensitivity; it was the interaction between two interlocutors, working together, that drove listener-sensitivity. In other words, listener-sensitivity critically depends on the listener behavior.

In sum, this dissertation has teased apart the confusion between listener-sensitivity deriving from an interaction and listener-sensitivity deriving from modeling the listener. I have demonstrated that modeling the state of knowledge of the listener alone is unlikely to result in listener-sensitivity, at least with respect to reduction behavior. Speakers may enter a conversation with a set of beliefs, but such beliefs become secondary and can change/updated once an interaction between interlocutors takes place. As a result, the interaction between interlocutors should be carefully considered when analyzing speech or gesture data since it plays a crucial role in influencing speaker production.

3. *Cross-modality reduction and automatization*

The Repetition Effect, also called the Reducing Effect (Bybee, 2002a), applies to all any kind of neuromotor behavior. The Repetition Effect has been vigorously discussed in the realm of speech production, especially in the area of phonetics (see section 3.1.1. in chapter I). The usage-based approach in linguistics has greatly profited from observing the Repetition Effect and convincingly accounted for the process of a diachronic sound change (see Hooper, 1976; Bybee & Scheibman, 1999; Bybee 2002a; Bybee, 2002b; Bybee 2000, for example). As Bybee (2001, 2002) argued, repetition of the same motoric sequences causes reduction in speech, where such automatization is strictly an internal mechanism on the speaker's part (the Listener-neutral explanation, see section 2.1. in chapter I and section 3.1.1. in chapter I).

Gesturing, a kind of motoric action, automatically reduces when repeated (see chapter IV for mimetic representational gestures and chapter V for lexical access representational gestures). This is consistent with Bybee's proposal that automatization applies to any kind of neuromotor behavior. As a matter of fact, automatization leading to reduction in repeated referents can be clearly observed in gesturing. Since gesturing is performed in a large spatial range (i.e., the area in which a gesture can occur is minimally at the rest position and can be maximally displaced at the speaker's extended arm length), when reduction occurs, it is easily observable. For instance, the reduction from using two shoulders to perform a gesture to using two wrists is an observable and prominent reduction. Second, gesturing may reduce to as little as no movement, where the zero case is a complete absence of gesturing. In other words, an extreme case of reduction in gesturing is synonymous with deletion. One might argue that speech could behave

comparably, for instance final t/d deletion in English (Bybee, 2002a), but this not the same case. While the final t/d segment in speech may be absent or deleted, the root still appears; whereas, speakers may reduce a gesture all the way to no movement in repetition.

In short, automatization is a common phenomenon that occurs across modalities. Despite having been mostly discussed in the domain of speech production to account for sound change, it can also be found in gesticulation, and importantly, more observable and prominent in gesturing. Since automatization is not modality specific, it is can be argued that any motoric action that is repeated is likely to undergo reduction as Bybee argued.

4. *Conventionalization and the Repetition Effect*

One element that sets speech and gesture apart is the degree of conventionalization. While gesturing is not at all conventionalized by definition (see section 2.5 in chapter I), speaking is a highly conventionalized act. Since the Repetition Effect manifests itself differently in the two modalities, one plausible explanation that accounts for the patterns is the degree of conventionalization.

Based on speech data in chapter II of this dissertation, the Repetition Effect on word duration occurs exclusively within a coherent stretch of discourse (within each telling). Regardless of the assumed state of knowledge of the listener, this pattern of the Repetition Effect holds true. I have accounted for this pattern of word duration with the Exemplar Model is responsible of an offline mechanism, which essentially becomes a conventionalized act (see section 5.1 in chapter II for a full discussion).

The conventionalization of word duration within a coherent stretch of discourse, what I am proposing here, is comparable to the use of the indefinite article *a* in English in

a certain construction. Generally speaking, the indefinite article *a* in English is not used with old/given information; it is often used with new information (cf. Du Bois, 1980). However, one almost always would use the indefinite article *a* in preference to the definite article *the* after *There is* in the presentational construction, even if the following noun is old/given information, e.g., *There is a town that's called Eugene.* In this example, regardless of whether the town of *Eugene* is assumed old/given to the listener, the indefinite article *a* is often the preferred form.

In this case, the definite article *a* is simply triggered by the presentational construction in English; competent speakers of English need not use it strategically or deliberately. It is used automatically since it is highly a conventionalized form that is used this particular syntactic construction, the presentational construction. While the scope of this example is at the sentence level, the scope of variation in word duration is at the discourse level, in each telling. Therefore, a structure-driven production, i.e., conventionalization, exists in multiple layers of language production.

In addition, conventionalization may put a limitation on how much reduction is allowed in repetition. Consider the previous example, *there is a town that's called Eugene*, which I have argued to be a conventionalized use of the indefinite article *a* in the presentational construction. Despite the number of times the phrase is repeated, the speaker may *not* reduce the indefinite article *a* to zero as in **there is town called Eugene.* Reducing the indefinite article a to zero, or an omission of the element, makes the sentence ungrammatical.

Representational gesture, on the other hand, lacks conventionalization by definition (see section 2.5. in chapter I for a full definition). Unlike words, there is no

shape or form that conventionally maps onto meaning. Since act itself lacks a concrete parameter, it would be difficult to be governed by discourse structure like word duration or the use of articles in English. In other words, an unconventionalized act is unlikely to conform to a pattern since the act itself does not have a solid form.

It would be interesting to observe emblems in repetition, perhaps using the paradigm utilized in this dissertation. Emblems share characteristics with speech, in the sense that it has conventionalized meaning/form pair. For example, holding up the index finger and the middle finger is equivalent to the word 'peace'. At the same time, emblems are produced via gesturing, which is a different modality than speech. Therefore, it would be interesting to observe emblems since they share characteristics that of speech and gesturing, which could provide more information on the effect of conventionalization on repeated referents.

5. *The functions of representational gesture and the Repetition Effect*

The purpose of speaking, without a doubt, is to communicate information to the listener. The purpose of representational gesture, on the other hand, is not as clear-cut. It seems that, generally speaking, the literature on representational gesture is divided into two broad views: gesturing as a speaker-oriented act (see section 3.2.1. in chapter I) or gesturing as a listener-oriented act (see section 4.2 in chapter I). Although representational gestures have been found to benefit listeners in comprehension of the message (e.g., Alibali et al., 1997; Beattie & Shovelton, 2005; Singer & Goldin-Meadow, 2005), Hostetter & Alibali (2008) made a convincing point that the information listeners gain from representational gesture is secondary to speech.

Representational gestures vary in the function they serve. Instead of arguing whether representational gestures are for the speaker or for the listener, it is more fruitful to view representational gestures as a function-driven act. Representational gestures that primarily benefit speakers cognitively are prone to reduction in repetition, while representational gestures that are primarily intended for communication are resistant to reduction in repetition. In other words, representational gestures should not be treated as a single-unified class since they vary depending on their function.

Chapter III of the dissertation demonstrates that absolute size representational gestures do not undergo reduction in repetition. I have argued that this kind of representational gesture is resistant to reduction in repetition because of the semantic content of size that is embedded in the representational gesture; should any reduction in gesturing take place, the semantic content of size will change. Therefore, I argue that the fundamental function of an absolute size representational gesture is for the purpose of communicating size information to the listener.

In contrast, chapter IV and chapter V of the dissertation demonstrate another kind of representational gesture. In chapter IV, mimetic representational gestures show a continual decline in motoric commitment and effort as a function of repetition; the more they are repeated, the more they are reduced. I argue that this kind of representational gesture primarily functions to assist speakers cognitively and is more prone to the Repetition Effect due to automatization.

Similarly, in chapter V, representational gestures associated with lexical access undergo reduction in repetition, even when speakers are still faced with lexical access difficulties in the repetition. This kind of representational gesture has been vigorously

argued in the literature to have no communicative value (e.g., Krauss, 2000) and is solely performed to aid the speaker in the process of retrieving a lexeme.

It should be highlighted that I am neither arguing for nor against the effectiveness of this kind of gesture; it is beyond the scope of this dissertation to evaluate the whether gestures that occur when speech dysfluencies occur help speakers resolve the problematic lexeme. In this dissertation, I am simply demonstrating that when speakers appear to have trouble recalling a word in speech, the motoric effort invested declines as a function of repetition. Operating under the assumption that these gestures assist speakers cognitively, for the purpose of retrieving a lexeme, they appear to become less and less prominent as a function of repetition.

In this dissertation, I have documented three different kinds of representational gesture, two of which can be argued to serve the speakers cognitively and behave comparably in the Repetition Effect; the other kind serves to communicate information for the speaker is resistant to the Repetition Effect. Based on the diverged results, I recommend treating representational gesture functionally, rather than treating them as a single class. Some representational gestures pack a semantic content that is intended for the listener, while others may not; as a result, it leads to a variation in communicative weight in each category.

Therefore, to ask whether representational gestures are *for the speaker* or *for the listener* would be an imprecise question, a question that is too broad to ask. The question, then, becomes *what kind* of representational gesture is intended for the speaker and *what kind* of representational gesture is intended for the listener? At the very least, based on the results presented in this dissertation, absolute size representational gestures

are intended for the listener, while mimetic representational gestures and representational gestures associate with lexical access are more likely to be intended for the speaker.

Breaking down representational gestures into sub-classes, navigated by function, is the most appropriate and fruitful approach. I believe that much of the controversy regarding whether a representational gesture is for the speaker or for the listener is rooted in the fact the specific functions of those representational gestures were not explicitly explicated.

CHAPTER VII

CONCLUSION AND FUTURE DIRECTIONS

Throughout the dissertation, I have demonstrated that the Repetition Effect manifests itself differently in each modality. In speech, the Repetition Effect is confined in the scope of a coherent stretch of discourse. In other words, when a word is repeated within a story, its duration will be shorter than in the first mention; this pattern of word duration persists regardless of the assumed state of the listener. On the other hand, two kinds of representational gestures, mimetic and representational gestures associated with lexical access, appear to undergo the Repetition Effect across a coherent stretch of discourse, while one kind of representational gesture, absolute size representational gesture, is resistant to the Repetition Effect.

While the dissertation captures the Repetition Effect in two modalities, the significance of the findings is not necessarily about the Repetition Effect. In this dissertation, I have made the following arguments. First, listener-modeling alone is insufficient to account for speakers' sensitivity to the listener. To be precise, it is not listener-modeling that is responsible for listener-sensitivity as the past literature claims; rather, it is the interaction between interlocutors that is responsible for listener-sensitivity. Second, automatization is a widespread phenomenon that occurs in any motoric action. As I have shown and argued, automatization occurs in production of representational gesture, especially when the representational gestures are intended to serve the speakers cognitively. In addition, the magnitude of reduction is, which is the result of automatization, also varies depending on modality. In other words, automatization leading to reduction in representational gesture is much more observable and prominent

than in speech and could reduce all the way to a complete absence of performance. Third, conventionalization has an effect on the Repetition Effect. Lastly, the notion of representational gesture should not be treated as a single class. Although all representational gestures are unconventionalized and lack a specific form, they may serve different functions for the speakers. Some are more prone to benefit the speaker cognitively, while others are more prone to serve a communicative purpose. Treating all representational gestures in the analysis as a single class would lead to an imprecise, and possibly a false conclusion.

Other than these arguments, I have also highlighted the importance of maintaining a tighter experimental control if one wishes to investigate how speakers model the state of knowledge of the listener. In other words, an interaction must be kept to a minimum, or better yet, be absent, to truly test speakers' assumption of the state of knowledge of the listener. Conversely, when interlocutors are able to freely interact, it is important to consider the interactional aspect, both verbal and non-verbal, in the analysis and not only analyze the production of the speakers, not the behavior of all interlocutors. This suggests that further studies should carefully manipulate or monitor interaction as an independent variable. Detailing the types of interaction and their effects on speech and/or gesture should provide a more complete understanding of the role interaction plays in conversation.

For future directions, other sequences of listeners, e.g., A-A-B-C, are encouraged in order to clearly tease apart possible fatigue effect; moreover, to observe what floor in reduction is in gesturing. It is possible that participants may stop gesturing all together, or a there could be a type shift from a representational gesture to a beat gesture. It would

also be interesting to observe how speakers refer to the referent (i.e., the dough) across tellings. That is, observing the use of definite and indefinite articles in the first and subsequent referents (possibly demonstratives) would yield insights to the extent to which speakers model the state of knowledge of the listener via grammar.

REFERENCES CITED

Abramson, A. (2002). The stability of distinctive vowel length in Thai. In K. Tingsabadh & A. Abramson. (Eds.), *Essays in Tai linguistics* (pp.13-26). Bangkok: Chulalongkorn University Press.

Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand?. *Journal of Educational Psychology*, *89*(1), 183.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*(2), 169-188.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and cognitive processes*, *15*(6), 593-613.

Allwood, J., & Cerrato, L. (2003, September). A study of gestural feedback expressions. In *first nordic symposium on multimodal communication* (pp. 7-22). Copenhagen, Denmark.

Anderson, A. H., & Howard, B. (2002). Referential form and word duration in video-mediated and face-to-face dialogues. In J. Bos, M. E. Foster, C. Matheson (Eds.), *Proceedings of EDILOG 2002*. Edinburgh: Cognitive Science Centre, University of Edinburgh.

Ariel, M. (1988). Referring and accessibility, *Journal of Linguistics*, 24, 65-87.

Ariel, M. (1990). *Accessing noun phrase antecedents*. London: Routledge.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31-56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*(5), 3048-3058.

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

Baker, R. E. & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech, 52,* 391-413.

Bard, E. G., and Aylett, M. (2005). Referential form, duration, and modelling the listener in spoken dialogue. In J. Trueswell and M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge: MIT Press. Pp. 173-191.

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

Bates, D. (2013). *Linear mixed model implementation in lme4*. University of Wisconsin Madison, Madison.

Battison, R. (1974). Phonological deletion in American sign language. *Sign language studies*, *5*(1), 1-19.

Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning an integrated message model of language in face-to-face dialogue. *Journal of Language and social Psychology*, *19*(2), 163-194.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*(2), 495-520.

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, *13*(1), 63-92.

Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, *99*(3-4), 239-272.

Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology*, *90*(1), 35-56.

Beattie, G., & Shovelton, H. (2005). Why the spontaneous images created by the hands during talk can help make TV advertisements more effective. *British Journal of Psychology*, *96*(1), 21-37.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92-111.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer (version 5.1.25)[Software]. Available from: http://www.praat.org/.

Bolinger, D. (1972). Accent is predictable (If you're a mind-reader). *Language*, *48*(3), 633-644.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review, 74*, 1-15.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341-376.

Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & D.R. Ladd (Eds), *Prosody: Models and measurements* (pp. 67-77). New York: Springer-Verlag.

Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language* 82(4): 711-733.

Bybee, J. (2002a). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*, 261-290.

Bybee, J. (2002b). Phonological Evidence for Exemplar Storage of Multiword Sequences. *SSLA* 24: 215-221.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (1988). Morphology as lexical organization. *Theoretical morphology*, *119141*.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics*, *37*(4), 575-596.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li (Ed.), *Subject and topic* (pp.25–55). New York: Academic Press.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. Journal of Experimental Social Psychology, 30, 580-601

Chovil, N. (1991). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction*, 25, 163-194.

Clark, H. H. (1992). *Arenas of language use*. University of Chicago Press.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, 332-373.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. *Discourse production and comprehension. Discourse processes: Advances in research and theory*, *1*, 1-40.

Clark, H. H., & Marshall, C. R. (2002). Definite reference and mutual knowledge. *Psycholinguistics: critical concepts in psychology*, *414*.

Clark, H.H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A.H. Joshe, B. Webber, & I.A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge, England: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive science*, *13*(2), 259-294.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior*, *22*(2), 245-258.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Cohen-Goldberg, A. M. (2015). Abstract and Lexically Specific Information in Sound Patterns: Evidence from/r/-sandhi in Rhotic and Non-rhotic Varieties of English. *Language and Speech*, 0023830914567168.

Cohen, A. A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, *28*(2), 276.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variances). *The Journal of the Acoustical Society of America*, *120*(1), 407-415.

DeLancey, S. (2011). Grammaticalization and syntax – A functional view. In H. Narrog & B. Heine (Eds.), *The Oxford handbook of grammaticalization* (pp.365-377). Oxford: Oxford University Press.

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

Du Bois, J.W. (1980) Beyond definiteness: The trace of identity in discourse. In W. Chafe (Ed.), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp.203-274). Norwood, NJ: Ablex.

Farnetani, E., & Recasens, D. (1999). Coarticulation models in recent speech production theories. *Coarticulation: Theory, data and techniques*, 31-65.

Fisher, C., & Tokura, H. (1995). The given-new contract in speech to infants. *Journal of Memory and Language*, *34*(3), 287-310.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680-698.

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

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

Fowler, C., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in Speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 62*, 35-51.

Frank, A., & Jaeger, F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 939-944). Austin, TX: Cognitive Science Society.

Freeman, V. (2014). Hyperarticulation as a signal of stance. *Journal of Phonetics*, *45*, 1-11.

Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American journal of psychology*, *111*(1), 43.

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

Galati, A., & Brennan, S. E. (2013). Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, *29*(4), 435-451.

Galati, A. & Brennan, S. E. (2010). Attenuating repeated information: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35-51.

Gerwing, J., & Bavelas J. (2013). The social interactive nature of gestures: Theory, Assumptions, methods, and findings. In Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., & Teßendorf, S. (Eds). *Body—Language—Communication: An international handbook on multimodality in human interaction (vol. 1)* (pp. 821-836). Berlin: Walter de Gruyter.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, *4*(2), 157-195.

Givón, T. (1979). From Discourse to Syntax: Grammar as a Processing Strategy in Discourse and Syntax. *Syntax and Semantics Ann Arbor, Mich.*, *12*, 81-112.

Givón, T. (1984). Syntax, vol. I. *Amsterdam: Benjamins*.

Goldiamond, I. & Hawkins W.F. (1958). Vexierversuch: The logarithmic relationship between word-frequency and recognition obtained in the absence of stimulus words. *Journal of Experimental Psychology*, *56*, 457–463.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. Academic Press.

Graham, Jean A. & Simon Heywood (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. European Journal of Social Psychology, 5, 189–195.

Haberlandt, K., Berian, C., & Sandson, J. (1980). The episode schema in story processing. *Journal of Verbal Learning and Verbal Behavior, 19*, 635-650.

Halliday, M. A. K. (1967). *Intonation and grammar in British English* (Vol. 48). Walter de Gruyter GmbH & Co KG.

Hilliard, C., & Cook, S. W. (2015, June 29). Bridging Gaps in Common Ground: Speakers Design Their Gestures for Their Listeners. Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication. http://dx.doi.org/10.1037/xlm000015.

Hoetjes, M., R. Koolen, M. Goudbeek, E. Krahmer, & M. Swerts (2011). GREEBLES Greeble greeb. On Reduction in Speech and Gesture in Repeated References. *Proceedings of the 33rd annual meeting of the Cognitive Science Society (CogSci)*. July 2011, Boston, Massachusetts, 3250-3255.

Hoetjes, M., R. Koolen, M. Goudbeek, E. Krahmer and M. Swerts (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language* 79-80, 1-19.

Hoetjes, M., E. Krahmer and M. Swerts (2014). On what happens in speech and gesture when communication is unsuccessful. *Proceedings of the 36th annual meeting of the Cognitive Science Society (CogSci)*, July 2014, Québec City, Canada, 2369-2374.

Holler, J. (2009, February). Speakers' use of interactive gestures as markers of common ground. In *International Gesture Workshop* (pp. 11-22). Springer Berlin Heidelberg.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*(1), 4-27.

Holler, J., & Wilkin, K. (2009). Communicating common ground: how mutually shared knowledge influences the representation of semantic information in speech and gesture in a narrative task. *Language and cognitive processes*, *24*, 267-289.

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, 43(14), 3522-3536.

Hooper, J. (1976). *An introduction to natural generative phonology*. Academic Press.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?. *Cognition*, *59*(1), 91-117.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review*, *15*(3), 495-514.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, *22*(3), 313-336.

Howes, D. H. & Solomon, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology, 41*, 401-410.

Isaacs, A. M., & Watson, D. G. (2010). Accent detection is a slippery slope: Direction and rate of F0 change drives listeners' comprehension. *Language and cognitive processes*, *25*(7-9), 1178-1200.

Iverson, J. M., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental psychology*, *33*(3), 453.

Iverson, J. M., & Goldin-Meadow, S. (2001). The resilience of gesture in talk: Gesture in blind speakers and listeners. *Developmental Science*, *4*(4), 416-422.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, *56*(2), 291-303.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, *61*(1), 23-62.

Jelec, A., & Jaworska, D. (2014, December). Thoughts on the table: Gesture as a tool for thinking in blind and visually impaired children. In *Yearbook of the Poznan Linguistic Meeting* (Vol. 1, No. 1, pp. 73-88).

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language*, *45*, 229-254.

Kapatsinski, V. (2013). Sound change and hierarchical inference. What is being inferred? Effects of words, phones and frequency. Manuscript submitted for publication.

Kapatsinski, V. (2010). Frequency of use leads to automaticity of production: Evidence from repair in conversation. *Language & Speech*, 53(1), 71-105.

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation: studies in corpus sociophonetics*. Palgrave Macmillan.

Kendon, A. (1994). Do gestures communicate? A review. *Research on language and social interaction*, *27*(3), 175-200.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Kimbara, I. (2006). On gestural mimicry. *Gesture*, *6*(1), 39-61.

Kimbara, I. (2008). Gesture form convergence in joint description. *Journal of Nonverbal Behavior*, *32*(2), 123-131.

Kita, S. (1993). Language and thought interface: A study of spontaneous gestures and Japanese mimetics. Unpublished Ph.D. dissertation, University of Chicago.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture*: *Window into thought and action* (pp. 162-185). Cambridge, UK: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, *48*(1), 16-32.

Krauss, R.M. (1998). Why do we gesture when we speak? *Current directions in psychological science*, 7, 54-60.

Krauss, R. M., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal, and visible factors in judgments of another's affect. *Journal of Personality and Social Psychology*, *40*(2), 312.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us?. *Advances in experimental social psychology*, *28*, 389-450.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process model. In D. McNeill (Ed.), *Language and gesture*: Window into thought and action (pp. 261-283). Cambridge, UK: Cambridge University Press.

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gesture. *Journal of experimental social psychology*, *31*(6), 533-552.

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate?. *Journal of personality and social psychology*, *61*(5), 743.

Kuhlen, A. K. & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin and Review, 20*, 54-72.

Kuhlen, A. & Seyfeddinipur, M. (2007). From speaker-to-speaker: Repeated gestures across speakers. Presentation held that *Berlin Gesture Center Colloquium*, Berlin.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & cognition*, *38*(8), 1137-1146.

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

Lindblom, B. (1990). Explaining phonetic variation: A sketch of H&H theory. In W. Hardcastle & A. Marchal. (Eds), *Speech production and speech modeling* (pp.403-439). Dordrecht: Kluwer.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic bulletin & review*, *9*(3), 550-557.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492.

Masson-Carro, I., M. Goudbeek, & E. Krahmer (2013). The influence of cognitive load on repeated references in speech and gesture. *Proceedings of the Tilburg Gesture Research Meeting (TiGeR 2013)*. June 2013, Tilburg, The Netherlands.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture & thought*. Chicago: Chicago University Press.

Melinger, A., & Levelt, W. J. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*(2), 119-141.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009a). Communicative gestures and memory load. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society. Austin, TX*.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009b). The communicative import of gestures: Evidence from a comparative analysis of human–human and human–machine interactions. *Gesture*, *9*(1), 97-126.

Mol, L., E. Krahmer, A. Maes, and M. Swerts (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language*, 66, 249-264.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 615.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American journal of psychology*, 411-424.

Myers, S., & Padgett, J. (2014). Domain generalisation in artificial language learning. *Phonology*, *31*(03), 399-433.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*, 357-395.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *4*, 272-281.

Olive, J. P., Greenwood, A., & Coleman, J. (1993). *Acoustics of American English speech: a dynamic approach*. Springer Science & Business Media.

O'Shaughnessy, D. (1995). Timing patterns in fluent and disfluent spontaneous speech. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on* (Vol. 1, pp. 600-603). IEEE.

Parrill, F. (2010). The hands are part of the package: Gesture, common ground, and information packaging. *Empirical and experimental methods in cognitive/functional research*, 285-302.

Pederson, E. (2014, July). *Listener head gestures and the co-construction of narrative timing*. Talk given at the meeting of the International Society for Gesture Studies, San Diego, CA.

Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, *16*, 691-698.

Pierrehumbert, J. (2006). Syllable structure and word structure: a study of tricomonantal clusters in English. *Phonological Structure and Phonetic Form*, *3*, 168.

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*(6), 747-754.

Plug, L., & Carter, P. (2013). Prosodic Marking, Pitch and Intensity in Spontaneous Lexical Self-Repair in Dutch. *Phonetica*, *70*(3), 155-181.

Pluymaekers, M., Ernestus, M., Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. Phonetica, 62, 146-159.

R Development Core Team. (2014). R: A language and environment for statistical Computing. Vienna: R Foundation for Statistical Computing. Available from: http://www.R-project.org.

R Development Core Team. (2009). R: A language and environment for statistical Computing. Vienna: R Foundation for Statistical Computing. Available from: http://www.R-project.org.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*(4), 226-231.

Raymond, W. D., & Brown, E. L. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. *Frequency effects in language learning and processing*, *1*, 35-52.

Rimé, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal variables. *European journal of social psychology*, *12*(2), 113-129.

de Ruiter, J. (1995). Why do people gesture at the telephone?. *Proceedings of the CLS opening Academic Year 1995-1996*.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.

Schweitzer, A. (2010). *Production and perception of prosodic events: evidence from corpus-based experiments*. (Doctoral dissertation, Stuttgart University).

Schweitzer, A., & Möbius, B. (2004). Exemplar-based production of prosody: Evidence from segment and syllable durations. In *Speech Prosody 2004, International Conference*.

Seibel, R. (1963). Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology*, *66*(3), 215.

Selkirk, E. (1996). The prosodic structure of function words. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, *187*, 214.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*(2), 85-89.

Sloetjes, H., & Wittenburg, P. (2008, May). Annotation by Category: ELAN and ISO DCR. In *LREC*.

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and linguistics compass*, *3*(1), 236-264.

Sridhar, V. K. R., Nenkova, A., Narayanan, S., & Jurafsky, D. (2008). Detecting prominence in conversational speech: pitch accent, givenness and focus.

Stalnaker, R. (1998). On the representation of context. *Journal of Logic, Language and Information*, *7*(1), 3-19.

Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, *25*(5), 701-721.

Streeck, J. (2008). Depicting by gesture. *Gesture*, *8*(3), 285-301.

Stave, M. & Pederson, E. (2015). Is the listener really listening? Exploring the effect of verbal and gestural speaker cues on backchanneling. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.) (2015). *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (p. 3088)*. Austin, TX: Cognitive Science Society.

Sumner, M., Kim, S. K., King, E., and McGowan, K. (2014). The socially-weighted encoding of spoken words: A dual-route approach to speech perception. Frontiers in Psychology, 4, 1 – 13.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and speech*, *37*(1), 21-43.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing 'given'information: Effects of persistence of grammatical function and surface position. *Language and Speech*, *37*(2), 125-145.

Uchanski, R. M. (2008). Clear Speech. *The handbook of speech perception*, 207.

den Uyl, M., & Oostendorp H.V. (1980). The use of scripts in text comprehension. *Poetics*, *9*, 275-294.

Vajrabhaya, P., & Kapatsinski, V. (2013). Modeling the listener? What resets acoustic durations of repeated English words. *The Journal of the Acoustical Society of America*, *134*(5), 4203-4203.

Vajrabhaya, P., & Kapatsinski, V. (2011). There is more to the story: First-mention lengthening in Thai interactive discourse. In *Proceedings of the International Congress of Phonetic Sciences* (Vol. 17, pp. 2050-2053).

Vajrabhaya, P., & Pederson, E. (2012, July). *Have I told you this before?: Reduction in magnitude over repetition in co-speech gesture*. Poster presented at the meeting of International Society for Gesture Studies, Lund, Sweden.

Vajrabhaya, P., & Pederson, E. (2013, October). Repetition vs. listener accommodation: A case study of co-speech gesture in retellings. Poster presented at the meeting of New Ways of Analyzing Variation, Pittsburgh, PA.

Vajrabhaya, P., & Pederson, E. (2014). *When anyone's listening: Automatization and gesture reduction*. Talk given at the meeting of the International Society for Gesture Studies, San Diego, CA.

Vajrabhaya, P., & Pederson, E. (2015). "The baking stick thing": Automatization of co-speech gesture during lexical access. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.) (2015). *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (p. 3007)*. Austin, TX: Cognitive Science Society.

Vajrabhaya, P., & Pederson, E. (2016, July). *Teasing apart listener-sensitivity: The role of interaction.* In E. Pederson & Vajrabhaya, P., Gesture, gesture: Tracking gestures across repetitions. Symposium conducted at the meeting of the International Society for Gesture Studies, Paris, France.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: an overview. *Speech Communication*, *57*, 209-232.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive science*, *34*(4), 537-582.

Yule, G. (1980). Speakers' topics and major paratones. *Lingua*, *52*(1-2), 33-47