2018

Investigating Hybrid Models Of Speech Perception

Robert Jonathan Wilder
University of Pennsylvania, rjwilder4@gmail.com

Recommended Citation

Wilder, Robert Jonathan, "Investigating Hybrid Models Of Speech Perception" (2018). *Publicly Accessible Penn Dissertations*. 3202.
https://repository.upenn.edu/edissertations/3202

Investigating Hybrid Models Of Speech Perception

Abstract
The ability to perceive sounds as words involves a transformation from detailed speech signals to invariant meanings, which are separate from information about the speaker of a particular word. The nature of this transformation is a central issue in the field of speech perception. A particular focus of ongoing debate concerns talker-specific details: are they causally relevant to lexical perception, or are they useful only for tasks like speaker recognition?

One common way to investigate the impact of voice information is to examine the time-course of its effects on future perceptual events. Early research reported no consistent long-lasting effects, implying that speech representations do not contain talker-specific detail (Jackson & Morton, 1984). However, subsequent work reported long-lasting effects, leading to a focus on modelling speech representations as abstractions over detail-rich episodic memories (Goldinger, 1996). Current hybrid models (Church & Schacter, 1994; McLennan & Luce, 2005; Goldinger, 2007) incorporate abstract and detail-rich speech representations but differ in the relative importance assigned each.

Two types of hybrid models are differentiated: a) models with combined representations, where abstraction occurs over detailed memories of speech episodes; versus b) models with separate representations, where different processing paths exist from the speech signal to word and speaker recognition. To investigate these models, this thesis reports multiple experiments investigating the time-course of the decay patterns of voice effects in repetition priming. Results from auditory lexical decision indicate that voice information only affects the speed of future perceptual processes within a short time window: until around three items intervene between prime and target. This finding clarifies previous results, which found no long-lasting effects, by providing an exact time-course of voice information's impact. Nevertheless, the results reported here differ from the predictions of studies investigating recognition accuracy, where long-lasting effects are commonly found. To address these differences, additional experiments using continuous and blocked word recognition paradigms were conducted. Again, talker-specific effects only persist within the same short time window, while abstract repetition priming effects persist much longer. By de-emphasizing the contribution of voice information, these findings assert the importance of abstract linguistic representations in hybrid models with separate representations.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Linguistics

First Advisor
David Embick

Keywords
Abstract representations, Episodic representations, Hybrid models, Lexical decision, Speech perception, Word recognition

Subject Categories
Cognitive Psychology | Linguistics | Psychology

INVESTIGATING HYBRID MODELS OF SPEECH PERCEPTION

Robert J. Wilder

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

____

David Embick
Professor of Linguistics

Graduate Group Chairperson

____

Eugene Buckley
Associate Professor of Linguistics

Dissertation Committee:

Meredith Tamminga, Assistant Professor of Linguistics
Florian Schwarz, Associate Professor of Linguistics

INVESTIGATING HYBRID MODELS OF SPEECH PERCEPTION

To my wonderful family and friends, for all your care and support over the years.

ACKNOWLEDGMENTS

From the dawn of time in ancient Greece, language has been important (credit to Debra Wilder for the first sentence of my dissertation). Time is truly remarkable. Not only does time apparently diminish memory representations of speech, but seven years of it also somehow results in a completed dissertation. Instead of focusing on the time however, I'd like to focus on the intervening people who made the experience memorable.

Starting first with my dissertation committee David Embick, Meredith Tamminga, and Florian Schwarz, who provided academic support throughout this process and graciously put up with my lack of writing ability. I am truly thankful for their patience and generosity.

Individually, I would first and foremost like to thank my advisor, David Embick, for his effort on this project. His unique ability to construct arguments and see connections has taught me much about how to conduct a research project. Without his input, I would not have been able to connect the experimental dots together. He also has great taste in music, science fiction, and video games; a true role-model for life.

Meredith Tamminga has been a constant through my entire program here at the University of Pennsylvania as a senior student, a faculty mentor, and a supervisor. Her care for the students and the department at large has been inspiring, along with her incredible ability to deal with unexpected experimental results.

Lastly, I cannot thank Florian Schwarz enough for the many years of guidance. He took a naïve student interested in experimental linguistics and developed him into who I am today. He has always demonstrated his caring both academically and personally, for which I will always be grateful.

I would also like to thank Jon Brennan, whose incredible energy and excitement for neuro-linguistics inspired me from day one. My summer working on neuro-linguistic research at the University of Michigan ranks as one of my favorite academic experiences.

I would not have gotten to this point without the help of many other mentors. Specifically, Peter Cole, who somehow saw potential in a young computer science undergraduate, Tony Kroch, for many enjoyable conversations, Satoshi Tomioka, for the years of instruction and friendship, Mutsuko Sato, for the years of caring advice, Arild Hestvik, for helpful comments and professional advice, and my first advisor Jeffrey Heinz, whose infectious love of computational phonology convinced me to pursue linguistics. He taught me many things about language and theoretical computer science but also about being a scholar, which I (attempt to) put into practice even today. A true superhero of linguistics and life.

I also thank the fellow students and post-docs who provided their advice and expertise to this project. Specifically, Constantine Lignos, for early help in stimuli creation and for the *lingtools* package, Akiva Bacovcin, for the R, Python, Psychopy, and statistical know-how, Jérémy 'Jerome' Zehr, for his work on the Ibex *PennController* and for programming assistance (along with indulging my attempts at French), and Kobey Shwayder, Chris

Ahern, and Helen Jeoung for the hours of recordings (and friendship!). Our department administrator Amy Forsyth deserves thanks as well for her tireless work dealing with the bureaucracy on my behalf. Also, I'd like to thank Gene Buckley, for his longtime service as Graduate Group Chair and for all the instruction and conversations.

Moving on to my personal support team, I must first thank my counselor Matt Leroy and the rest of the Counseling and Psychological Services at Penn. Above all, Matt has been a truly stabilizing and supporting influence during the turbulent times of graduate school. I will never be able to thank him enough for everything. Much thanks also goes to my beverage support team Sharon, Mike, Jackie, and Brian at the New Deck Tavern, along with fellow patrons Dorian and especially Huner, for believing in me despite my self-doubt. Graduate school can be a solitary and often-depressing venture, but I was fortunate to have a place to go where everybody knew my name and was glad I stopped by.

And now for my physical support team, I am very grateful for the community provided by the Penn Taekwondo Club. Specifically, the instructors Felice Macera, Joke Bradt, and Malcolm Jones for their selfless donation of time and energy to lead our classes. There is nothing like a strenuous taekwondo workout after a long day of dissertation writing. Additionally, my laughter/gym buddies Richard Zimmermann and Rui Castro deserve much credit in keeping me fit enough to write!

To all of my friends, I am truly blessed to have such great people around me. I could write another 100 pages about you all. I'll spare my readers, but apologies in advance for not writing more. First, I thank my cohort and members of Phuntax for the lasting friendships and shared struggles; namely Eric Doty, Haitao Cai, Betsy Sneller, Anton Karl 'Anty' Ingason, Einar Freyr 'Einy' Sigurðsson, Sabriya 'Sabrielle' Fisher, Duna 'Dina' Gylfadóttir, Amy 'Lady Schmameford' Goodwin Davies. Special thanks goes to Einar and Hlíf (and kids) for providing a home for all of us in the early years and for the many conversations and laughter over delicious Pabst. Also to Duna (#teamBaton, formally #teamDissy, #teamSlides, future #teamCelebration?) for being the best long-distance (and when lucky, local) dissertation buddy ever. Additionally, pretty much everything I've done at Penn has involved Amy as a collaborator. I am so thankful for her earnest care for all of GradLingS and specifically for encouraging me throughout the dissertation writing process. I will miss the musings over /ŋʊll/, berry theory, grandparents, fake dissertations, etc. Thank you to the best counterpart anyone could ever ask for! Lastly, I'd like to thank Sabriya for the comments on drafts and presentations, and especially for always putting everything into proper focus when graduate school seemed insurmountable. The many laughs, explorations around Philly, and the delicious dinners made the whole process much more fun than it would have otherwise been.

Additionally, I'd like to thank my current office-mates Afa Creemers, for helping me believe in myself, and Pleng Chanchaochai, for her generosity and thoughtfulness. You both, along with past office-mates Amy and Duna, certainly put the 'fun' in experimental design. My fellow colleagues in Penn linguistics deserve thanks as well, specifically Milena Šereikaitė for the friendship and support, Ava Irani for the happiness she always brings, Aaron Freeman for sharing the dissertation struggles, David Wilson for imparting his massive amounts of knowledge on Afro-Asiatic and baseball, and fellow colleagues Luke Adamson, Kajsa Djarv, Andrea Ceolin, Faruk Akkuş, and Domonique Roberts-Mack. The members of the Language Variation & Cognition and Experimental Morphology Labs, specifically Wei Lai, Ruaridh Purse, and Yosiane White deserve much recognition for all the times they sat through my meandering presentations about experimental results. Also for the epic *Sharknado*-themed lab party. Much thanks goes to Hongzhi Xu, who amazingly provided me with a place to stay during the final year of graduate school. I've very much enjoyed the conversations, dinners, games, and friendship this past year.

My friends outside of linguistics also deserve much recognition. I'd like to thank Larry, Asher, and the Larragas for being my second family in DE. My church family at City Church Philadelphia and St. John's in DE deserves recognition, especially Joel Bassett & Kirsten Jensen for leading my community group. Additionally, Josh 'Dagles' Louder deserves recognition for the excellent advice of "you got to get to one (chapter) before two", Jana Reifegerste for the dissertation support, Steven & Erik Taylor and Aaron Siegel for the wonderful procrasti-distractions on Google chat, and Jennifer Lim for the many conversations. Additionally, Pau Pereira deserves much appreciation for the amazing friendship and support, as well as the long-night discussions about history and board game theories; not to mention the cigars! Finally, Ryan 'Charles' Gehrmann for being a life-long friend and a true linguist, as well as for motivating me to always continue forward.

Without the time given to the experiments in this thesis (and the other pilot studies not presented here) by the 630 nameless participants recruited from Penn and Prolific, this project would obviously not be much! I hope the final product makes you all proud of your contribution to science.

I've been also blessed to have an amazing family. My sister-in-law Kimberly 'Slaw' Wilder deserves much credit for all the support she's given and my brother Timothy 'Tito' Wilder for always being ready to remind me of the good old days. I'd like to thank my little nephew Elisha 'Snugglet' Wilder for always cheering me up and helping me as best as a two-year old could. My parents deserve the majority of credit for everything; specifically my father Robert G. Wilder for early on fostering a love of learning and science, and last but not least, my mother Debra Wilder, for the countless things she has done to support me. Her care for others has constantly inspired me to never neglect the people beside us.

At the end, I give praise and thanks to God who brought me through to the end and blessed me through all of these wonderful people.

ABSTRACT

INVESTIGATING HYBRID MODELS OF SPEECH PERCEPTION

Robert J. Wilder

David Embick

The ability to perceive sounds as words involves a transformation from detailed speech signals to invariant meanings, which are separate from information about the speaker of a particular word. The nature of this transformation is a central issue in the field of speech perception. A particular focus of ongoing debate concerns talker-specific details: are they causally relevant to lexical perception, or are they useful only for tasks like speaker recognition?

One common way to investigate the impact of voice information is to examine the time-course of its effects on future perceptual events. Early research reported no consistent long-lasting effects, implying that speech representations do not contain talker-specific detail (Jackson & Morton, 1984). However, subsequent work reported long-lasting effects, leading to a focus on modelling speech representations as abstractions over detail-rich episodic memories (Goldinger, 1996). Current hybrid models (Church & Schacter, 1994; McLennan & Luce, 2005; Goldinger, 2007) incorporate abstract and detail-rich speech representations but differ in the relative importance assigned each.

Two types of hybrid models are differentiated: a) models with combined representations, where abstraction occurs over detailed memories of speech episodes; versus b) models with separate representations, where different processing paths exist from the speech signal to word and speaker recognition. To investigate these models, this thesis reports multiple experiments investigating the time-course of the decay patterns of voice effects in repetition priming. Results from auditory lexical decision indicate that voice information only affects the speed of future perceptual processes within a short time window: until around three items intervene between prime and target. This finding clarifies previous results, which found no long-lasting effects, by providing an exact time-course of voice information’s impact. Nevertheless, the results reported here differ from the predictions of studies investigating recognition accuracy, where long-lasting effects are commonly found. To address these differences, additional experiments using continuous and blocked word recognition paradigms were conducted. Again, talker-specific effects only persist within the same short time window, while abstract repetition priming effects persist much longer. By de-emphasizing the contribution of voice information, these findings assert the importance of abstract linguistic representations in hybrid models with separate representations.

TABLE OF CONTENTS

ACKNOWLEDGMENTS .... **IV**

ABSTRACT .... **VII**

TABLE OF CONTENTS .... **IX**

LIST OF TABLES .... **XI**

LIST OF FIGURES .... **XII**

1 CHAPTER 1: INTRODUCTION **1**

1.1 Introduction **1**
- 1.1.1 Importance of abstract and token-specific information 3
- 1.1.2 Abstract vs. episodic models 5
- 1.1.3 Separate vs. combined representations 9
- 1.1.4 Evidence for speech perception models 14
- 1.1.5 Summary 17

1.2 Overview **18**

2 CHAPTER 2: EXPERIMENTAL METHODS **22**

2.1 Introduction **22**

2.2 Experimental Methods **22**
- 2.2.1 Participants 23
- 2.2.2 Stimuli 24
- 2.2.3 Design 26
- 2.2.4 Procedure 28
- 2.2.5 Analysis 29
- 2.2.6 Results 32
- 2.2.7 Accuracy 32
- 2.2.8 Reaction Time 33

2.3 Experiment 1 .... **36**
- 2.3.1 Method 36
- 2.3.2 Results 40
- 2.3.3 Discussion 47

2.4 Conclusion **48**

3 CHAPTER 3: TIME-COURSE IN IMPLICIT TASKS **49**

3.1 Introduction **49**

3.2 Background **54**
- 3.2.1 Models of Memory 54
- 3.2.2 Perceptual Identification 58
- 3.2.3 Indirect Tasks (Shadowing & Lexical Decision) 63
- 3.2.4 Repetition Priming Decay 68
- 3.2.5 Summary 73

3.3 Experiment 2 .... **73**
- 3.3.1 Method 75
- 3.3.2 Results 77

3.3.3 Discussion 89

3.4 Experiment 3..90
3.4.1 Method 91
3.4.2 Results..94
3.4.3 Discussion 105

3.5 Experiment 4..106
3.5.1 Method 107
3.5.2 Results..112
3.5.3 Discussion 125

3.6 Conclusion..127

4 CHAPTER 4: TIME-COURSE IN EXPLICIT TASKS 129

4.1 Introduction..129

4.2 Background..131
4.2.1 Motivating Implicit vs. Explicit Memory 132
4.2.2 Blocked Word Recognition..136
4.2.3 Continuous Word Recognition 143
4.2.4 Summary 147

4.3 Experiment 5..148
4.3.1 Method 149
4.3.2 Results..151
4.3.3 Discussion 162

4.4 Experiment 6..163
4.4.1 Method 164
4.4.2 Results..171
4.4.3 Discussion 174

4.5 Conclusion..175

5 CHAPTER 5: GENERAL DISCUSSION 177

5.1 Experimental Summary..177
5.1.1 Frequency Effects 178
5.1.2 Talker-specific representations 180
5.1.3 Lexicality Effects..181

5.2 Conclusion..184

APPENDIX I: EXPERIMENTAL STIMULI 188

AI.1 Words..188

AI.2 Non-Words 196

BIBLIOGRAPHY 201

LIST OF TABLES

Table 1: Summary of stimuli durations per speaker 25
Table 2: Summary statistics for the stimuli in this thesis 26
Table 3: Experiment 1 removal summary 40
Table 4: Experiment 1 data summary 40
Table 5: Experiment 1 combined accuracy model 42
Table 6: Experiment 1 target accuracy model 43
Table 7: Experiment 1 combined RT model 45
Table 8: Experiment 1 target RT model 47
Table 9: Experiment 2 removal summary 77
Table 10: Experiment 2 data summary 78
Table 11: Experiment 2 combined accuracy model 80
Table 12: Experiment 2 target accuracy model 81
Table 13: Experiment 2 combined RT model 84
Table 14: Experiment 2 target RT model 86
Table 15: Experiment 3 removal summary 94
Table 16: Experiment 3 data summary 95
Table 17: Experiment 3 combined accuracy model 97
Table 18: Experiment 3 target accuracy model 98
Table 19: Experiment 3 combined RT model 100
Table 20: Experiment 3 target RT model 102
Table 21: Experiment 4 removal summary 111
Table 22: Experiment 4 data summary 112
Table 23: Experiment 4 combined accuracy model 114
Table 24: Experiment 4 target accuracy model 116
Table 25: Experiment 4 combined RT model 118
Table 26: Experiment 4 combined RT model effect sizes 120
Table 27: Experiment 4 target RT model 122
Table 28: Experiment 4 target RT model effect sizes 123
Table 29: Experiment 5 removal summary 151
Table 30: Experiment 5 data summary 152
Table 31: Experiment 5 combined accuracy model 154
Table 32: Experiment 5 target accuracy model 156
Table 33: Experiment 5 combined RT model 158
Table 34: Experiment 5 target RT model 161
Table 35: Experiment 6 design 168
Table 36: Experiment 6 removal summary 170
Table 37: Experiment 6 data summary 171
Table 38: Experiment 6 target accuracy model 172

LIST OF FIGURES

Figure 1: The word 'cat' 5
Figure 2: Schematization of abstract models 6
Figure 3: Schematization of episodic models 7
Figure 4: Schematization of hybrid models with separate representations 10
Figure 5: Schematization of hybrid models with combined representations 11
Figure 6: The continuous lexical decision task (from Goodwin Davies, 2018) 27
Figure 7: Example reaction time distribution (pre-trimming) 31
Figure 8: Properties of Experiments 1-3 word stimuli 37
Figure 9: Phonotactic prob. of words (left) & non-words (right) in Experiments 1-3 38
Figure 10: Experiment 1 accuracy data 41
Figure 11: Experiment 1 reaction time data 44
Figure 12: Experiment 1 trimmed reaction time 46
Figure 13: Experiment 2 accuracy data 79
Figure 14: Experiment 2 reaction time data 82
Figure 15: Experiment 2 trimmed reaction time data 85
Figure 16: Experiment 2 effects of time vs. interveners 88
Figure 17: Experiment 3 accuracy data 96
Figure 18: Experiment 3 reaction time data 99
Figure 19: Experiment 3 trimmed reaction time 101
Figure 20: Experiment 3 effects of time vs. interveners 104
Figure 21: Properties of Experiments 4-5 word stimuli 108
Figure 22: Phonotactic prob. of words (left) & non-words (right) in Experiments 4-5 109
Figure 23: Experiment 4 accuracy data 113
Figure 24: Experiment 4 reaction time data 117
Figure 25: Experiment 4 trimmed reaction time 120
Figure 26: Experiment 4 target RT model predicted values 123
Figure 27: Experiment 4 effects of time vs. interveners 125
Figure 28: Experiment 5 accuracy data 153
Figure 29: Experiment 5 reaction time data 157
Figure 30: Experiment 5 trimmed RT data 160
Figure 31: Properties of Experiment 6 word stimuli 166
Figure 32: Experiment 6 accuracy over distance 173
Figure 33: Experiment 6 RT over distance 174
Figure 34: Experiment 6 accuracy by frequency 178
Figure 35: Accuracy by frequency in a lexical decision task 179
Figure 36: Experiment 4 non-word accuracy 181
Figure 37: Experiment 4 non-word RT 182
Figure 38: Experiment 5 non-word accuracy 183
Figure 39: Experiment 5 non-word RT 184

CHAPTER 1: Introduction

1.1 Introduction

Speech is highly variable. People’s ability to understand one another despite the challenges that this variability presents is quite remarkable. When two people from the same dialect converse with each other, they are not typically misled by changes in voice, speech rate, amplitude, or any of the other properties that differ between each speaker and each instance of speech. Instead, they successfully perceive the linguistic content of the speech they hear. And yet, people are still familiar with others’ voices and can recognize them quite easily. Additionally, extra-linguistic properties, like speech rate, can be interpreted by listeners as meaningful along prosodic or emotional dimensions. We can tell when the people we know sound sick, happy, or tired. So, at one level, we clearly perceive properties of speech episodes in people’s different voices, manners, and styles of speaking from instance to instance, and yet, at another level, all of this is ignored to get at the abstract meaning of what someone is saying. How this occurs has been one of the central questions driving work on speech perception and remains an unsolved puzzle, fundamental to a comprehensive understanding of how we perceive language.

The mental representation of speech is the focus of this thesis. To begin, we make a distinction between two types of information: invariant, abstract information that does not depend on any one utterance and specific, detailed information that comes from a unique utterance. The first type is most often referred to as *abstract* information, a term which we will adopt here as well. Examples of abstract information are semantic meaning, lexical status, the frequency of use, and the phonemic representation of words. The second type of information has been called *episodic*, *indexical*, or *token-specific* information. In this thesis, we will adopt the latter and refer to information about a specific speech episode, normally an instance of a spoken word, as token-specific information.

With these terms defined, we now clarify the object of study. We aim to determine to what extent abstract and token-specific information is stored in the mental representation of speech. This is accomplished by synthesizing the current body of research, with its different assumptions, terms, and findings, into a set of predictions concerning the time-course and memory status of token-specific information. These predictions are then tested with a set of experimental studies. These experimental manipulations are designed to reveal the presence of token-specific effects (TSEs, terminology adopted from Brown & Gaskell, 2014). If TSEs are indeed found, we would conclude that the evidence supports the hypothesis that the representation of speech is sensitive to token-specific information.

The importance of this evidence becomes clear when we consider how to model speech perception. Linguistics has been primarily concerned with abstract representations; for example, phonemes (Sneller, 2018), morphemes (Goodwin Davies, 2018), or syntactic features (Sigurðsson, 2017). Linguistic operations are defined with a certain set of these primitives available. The conclusions of this thesis inform how we can model the transformation of the speech signal into early mental representations. These representations have been hypothesized to be either abstract, mental units (like phonemes) or to be mapping operations which find similarities between detailed memories of previous events. This thesis addresses these two hypotheses and concludes, along with other recent research supporting hybrid models of speech perception, that there is evidence for both.

Specifically, the research presented here advocates for separate representations of abstract and token-specific information. Contrary to well-known studies conducted a few decades ago finding long-lasting TSEs (e.g., Goldinger, 1996), we find evidence that TSEs appear early and then disappear quite rapidly. This evidence calls into question the hypothesis that speech perception is built from detailed memories of previous speech instances. Instead, token-specific information appears to be causally separated from the word recognition process. This finding motivates a return to modelling speech perception as primarily an abstraction operation. Episodic, detail-rich properties do impact early perception but are better modelled as separate from the relationship between sound signals and mental representations of language. The rest of this introduction looks in more detail at the tension between abstract and token-specific information in the literature. Additionally, we introduce and define some of the key debates that will be addressed later in this thesis.

1.1.1 Importance of abstract and token-specific information

We now consider in more depth why the importance of abstract and token-specific information has been such a central debate in speech perception. On the one hand, there are classic perception studies showing that we perceive phonemes as categorical entities, not as gradient phenomena (Liberman et al., 1957). A corollary of this principle is that token-specific information is lost during the process of perception. For example, participants categorize a continuum of sounds within a specific VOT range (voice onset time) as /b/. On the other hand, the well-known phenomenon of *perceptual learning* (Norris et al., 2003; Eisner & McQueen, 2006; Kraljic & Samuel, 2005; *inter alia*) stands in direct contrast to the obligatory removal of token-specific detail. Perceptual learning is exhibited when token-specific information is remembered upon hearing sounds embedded in words which are ambiguous between two phonemes. These sounds later impact perceptual events by shifting a listener's categorical perception boundaries. Additionally, research has shown that phonetic details which are relevant for phonemic perception are in fact imitated in shadowing experiments (Nielsen, 2011). Categorical perception and perceptual learning and imitation therefore comprise well-known phenomena that separately suggest that token-specific information is crucially discarded to perceive categories and that it is crucially retained to imitate and shift category boundaries.

These contrary predictions about the effects of token-specific information are echoed throughout the speech perception literature. Some studies fail to find any TSEs while others find robust, long-lasting TSEs with multiple different information types (e.g., voice, speech rate, and emotional connotation). Speech perception models asserting the primacy of both abstract and token-specific information have been proposed, as we briefly review in the next section. The focus of this thesis then is not to completely discard one of these two sources of speech information but rather to determine how they are both packaged into representations following the perception of a speech stimulus.

To clarify, consider that all words we hear are sound waves like the following:

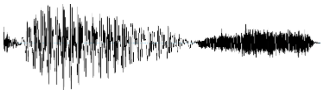

Figure 1: The word 'cat'

Embedded in this signal is token-specific information about the speaker, speech rate, and emotional content for example, along with information about phonemes, syllables, and words. Perceiving this signal as a word (meaning an invariant, morpho-syntactic object linked to a meaning) entails the removal of the token-specific information to get to the abstract representation; a process called *speech* normalization (Mullennix et al., 1989). Since the token-specific detail does matter for processes like speaker recognition however, the perception of this signal also must generate representations of speakers. These two different types of representations might exist simultaneously as separate entities or be combined into a unitary representation. How these representations interact is a foundational question in the field of speech perception.

1.1.2 Abstract vs. episodic models

Building on the theory of *speech normalization*, early models (specifically, the *Logogen* model: Murrell & Morton, 1974; Jackson & Morton, 1984) proposed that word recognition occurred after all token-specific information is removed. Some of the early speech perception studies supported this finding, as we review later in Section 3.2.2. In the auditory domain, this model proposes separate analysis and activation steps. After enough incoming speech information is analyzed (i.e., through the removal of detail to perceive sub-lexical representations), a word is then activated and available for future linguistic processing steps. This model's separation of speech detail and abstract word representations is the crucial aspect that we are concerned with. The schema below visually illustrates the proposal of abstract models:

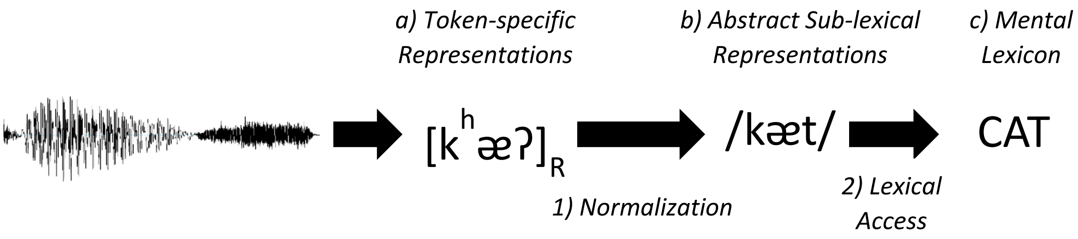

Figure 2: Schematization of abstract models

First, the speech signal is perceived with a) all allophonic and indexical information (the sub-script *'R'* indicating the speaker of the word). Then the process of 1) normalization occurs, which removes information irrelevant for word recognition from the representation. This results in b), the abstract sub-lexical representation needed for recognition. Models differ on what constitutes these sub-lexical representations, but proposals include phoneme combinations or syllables, for example. Then, the c) mental lexicon is accessed using these sub-lexical representations and the relevant word is perceived.

Many of the current well-known models of word recognition build upon these ideas of abstract lexical activation from the *Logogen* model and do not attempt to address token-specific information. Models like Cohort (Marslen-Wilson, 1987), TRACE (McClelland & Elman, 1986), and Shortlist (Norris, 1994; for an excellent review of these, see Weber & Scharenborg, 2012) are instead concerned with what sub-lexical abstract units may be underlying the activation of abstract word representations. They posit that incoming abstract phonemes introduce a candidate set of potentially perceived abstract words. As these models are mainly concerned with explaining lexical competition effects and determining the relative importance of bottom-up, perceptual information and top-down, predictive information, we will not be directly concerned with them in this thesis.

Other research has contradicted the predictions of the normalization hypothesis by finding evidence for long-lasting TSEs in word recognition, discussed in Chapters 3 and 4. These findings have led to the proposal that word recognition is based upon *perceptual similarity comparisons* between the incoming speech signal and stored memories of other detail-rich speech episodes. Following this proposal, the episodic memory model MINERVA2 has been adapted to speech recognition by Goldinger (1998). Below, the operation of these types of models is schematized:

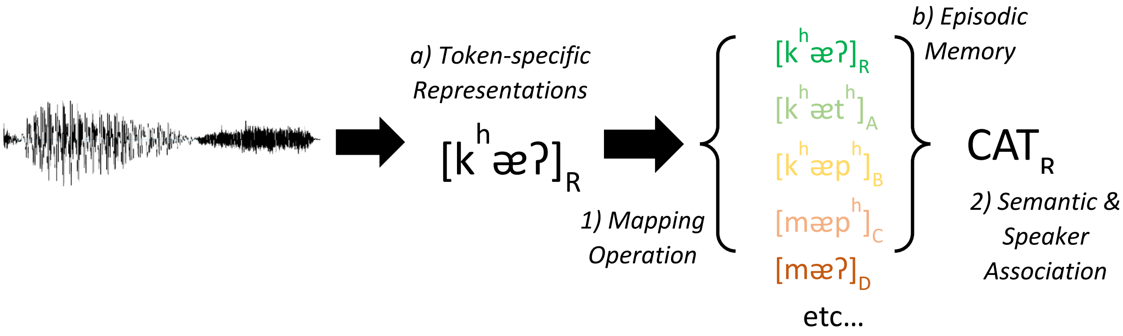

Figure 3: Schematization of episodic models

Again, the speech signal is perceived with a) all token-specific detail. This representation is then 1) mapped onto existing episodic traces of previous perceptual events in b) episodic memory. The mapping operation activates the most similar episodic traces (indicated by color). These traces are then averaged over to constitute the final representation of what was perceived, which is then 2) associated with meaning and any existing abstract indexical components.

These (necessarily vague) cartoon representations illustrate the potential steps a listener proceeds through from hearing a word's sound to perceiving a word's meaning. In evaluating these two types of models, we first discuss the predictions of each. While both can be modified to handle properties like word or phoneme frequency, they make different predictions concerning TSEs as revealed by priming patterns. To clarify, the term *priming*, as used in this thesis, refers to any effect (either facilitatory or inhibitory) that one item (called the *prime*) has on a later item (the *target*). First, fully abstraction-based models straightforwardly predict that:

- The normalization process is cognitively demanding; perceiving speech from multiple sources should be more difficult than from a single source
- Only representations accessible in the mental lexicon should be activated; non-words shouldn't participate in priming effects
- All priming effects should be based only on abstract content

Fully episodic-based models predict that:

- The mapping operation is an automatic process; every perceptually similar item is activated upon perception
- The speech signal is mapped to the nearest episodic trace; non-words should activate similar-sounding words and be quickly available to affect future perceptual events
- Priming effects are completely defined by token-specific detail; all types of detail may *a priori* cause priming

Naturally, while modifications and caveats exist for each of these, much of the early literature investigated these predictions and provided support for one of the two possible model types.

With the accumulation of evidence supporting both views on word recognition, most researchers now advocate for hybrid models. As Weber & Scharenborg (2012:396) state, "*With respect to form of representations, it has become obvious that both purely abstract models and purely episodic models are incomplete*, *and the challenge for the future is to develop a hybrid approach that combines both abstract and episodic representations."* To begin to address the challenge of combining abstract and token-specific information, we now turn to discussing some existing hybrid models in the literature.

1.1.3 Separate vs. combined representations

Unlike the unitary models discussed in the previous section, hybrid models acknowledge the contributing effects of both abstract and token-specific information on word recognition. The main distinction between these types of models concerns how the two types of information are packaged together into mental representations. In this thesis, we will make a distinction between what we term *separate* and *combined* representations. This distinction hinges around the causal access to words (i.e., what information is necessary and sufficient to activate a word), which in our view separates hybrid models into these two types. One way to think about this distinction is by analogy to the field of visual perception. Much of the classic research has advocated for dual-route pathways in perceiving objects, the *what* and *where* pathways, rather than a single-route pathway from simple to complex information. Hybrid models of speech can be divided along analogous lines into those in which speech information proceeds through a dual-route pathway (we can think of these as *what* and *who* pathways), and those with a single route from a detail-rich to an abstract representation. Again, this subtle distinction relates to the relative importance assigned to token-specific information, which makes this binary distinction somewhat fluid when examining the proposed hybrid models.

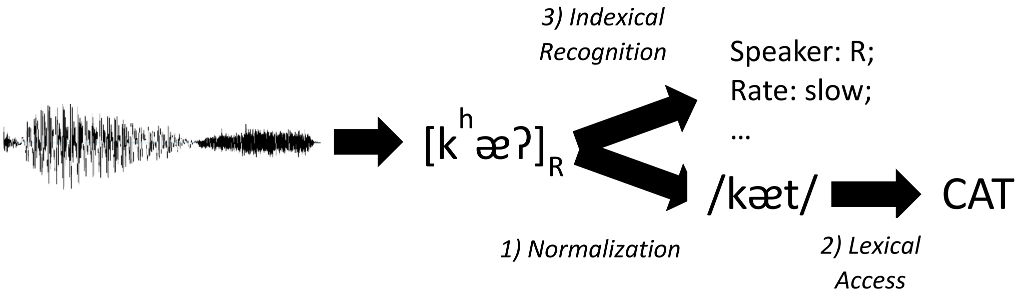

Figure 4: Schematization of hybrid models with separate representations

Hybrid models advocating separate representations (i.e., a dual-route representational system) posit that representations of token-specific information exist apart from abstract representations of words. After the initial stage of perception, as seen in the schematization above, two separate pathways exist to recognize the word and other indexical properties (like speaker or speech rate, for instance). This can be conceived of in (at least) two ways: 1) at some point in the perceptual process, separate representations of both are created from the same early processing of the speech signal or 2) abstract and token-specific information are dealt with by different parts of the brain (e.g., different hemispheres) altogether. An example of such a separate representational theory can be found in the work of Church & Schacter (1994). Their work synthesizing neural lesion studies with behavioral and neural speech perception experiments suggests that the left hemisphere deals with abstract representations while the right hemisphere deals with token-specific ones. Any interaction between the two is due to very early similarities in processing or perhaps to information sharing between the hemispheres. A separate, dual-route representational account is also present in Orfanidou et al. (2011), who do not find TSEs with word stimuli but do find repetition priming with non-word stimuli. Non-word repetition priming is theorized to access a different type of detail-rich representation that does not occur when an abstract representation is available. Finally, the examples cited in Pierrehumbert (2016) may be taken as support for a type of separate representational account which posits a large amount of information sharing between two separate representational modules of token-specific and abstract information.

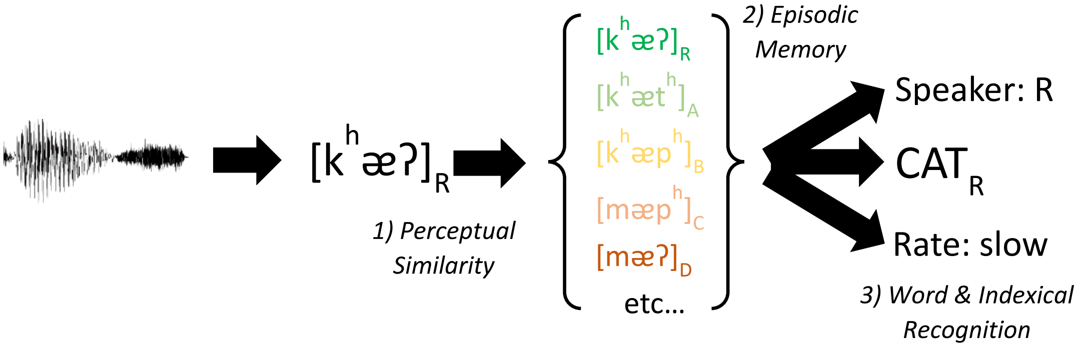

Figure 5: Schematization of hybrid models with combined representations

Other hybrid models are built on the idea of combined representations, also termed single-route models above. These types involve detail-rich representations being necessarily created upon hearing speech, with abstract representations being dependent on these detailed representations. In the schematization above, the early stage mirrors the episodic models from Section 1.1.2. After comparing the perceptual similarity between incoming sounds and stored episodic memory traces, different abstract representations can be accessed depending on the task. An example of this is Goldinger (2007) and Brown (2011) who propose the use of the *complementary learning systems* model (see e.g., Davis

& Gaskell, 2009) to tie abstract and episodic properties together. The idea is essentially this: upon hearing a word, the detail-rich representation is stored as an episodic memory trace in a hippocampal network. A cortical network summarizes similarities in this hippocampal network to form abstract representations from detail-rich representations. If enough input has been perceived for some property to be abstracted away from these detail-rich representations, an additional invariant representation is created and then is available to cause priming effects on future perceptual events. The detail-rich representations however should always be activated by primes and should therefore always impact priming effects. Work showing that *'atypical'* stimuli show TSEs while *'typical'* ones do not (Nygaard et al., 2000) has been taken to support this type of model. Atypical, rare instances of a word presumably do not have abstract representations to draw on and are only activated through similarity comparisons to other detail-rich memory traces. The *time-course hypothesis* of McLennan et al. (2003, 2005), which is built on *adaptive resonance theory* (see e.g., Grossberg, 1986), could also be thought of in this way. They also claim that abstract representations are formed from detail-rich ones; the main difference in their account lies in their added notion of frequency of activation. Since abstract representations are more frequently encountered and subsequently activated than detail-rich ones (i.e., one hears the word *cat* more often than one hears a single person saying *cat*), abstract representations resonate stronger with matching input. The similarity matching process linking incoming speech to detail-rich representations is correspondingly a slower process than the one linking speech to abstract representations, which fits their findings of early abstract effects and late-arising TSEs.

So, how should we determine whether separate or combined representations are appropriate for models of word recognition? This thesis does so by re-examining the time-course of TSEs and how they are affected by different tasks. Much evidence in this field comes from studies which examined the specific predictions of the two unitary abstract and episodic models we discussed in the previous section. These studies looked at the decay properties of TSEs, how TSEs were influenced by processing demands, and how implicit and explicit types of memory affected TSEs. Since these past studies often addressed tangential questions and started from different assumptions, the results in the literature are hard to reconcile. By re-visiting these questions with minimally contrasting experiments using modern analysis methods, we aim to provide clear results informing current hybrid models of speech perception.

The experiments in this thesis all focus on talker-specific information, which is one of the most often-investigated token-specific properties in the literature. The logic of the experimental investigations presented here is based on attempting to tease apart word recognition from speaker recognition. If talker-specific information, which we know can last for quite a long time in memory given our ability to remember the voices of those we met, yields different effects on perception than abstract lexical properties, we conclude that the two are represented separately in the cognitive system. From this conclusion, we would then support hybrid models with separate representations. If, on the other hand, talker-specific effects on perception are found to last for long-periods of time, then perhaps a unitary episodic memory system underlies all speech perception; a finding which would support hybrid models with combined representations. The next section highlights the types of evidence investigated in the literature, setting the stage for a larger investigation of each in the subsequent chapters of this thesis.

1.1.4 Evidence for speech perception models

1.1.4.1 Short- vs. long-distance effects

Some of the main evidence supporting various models of speech perception comes from the examination of the decay profiles of TSEs and abstract repetition priming. The reasoning (found in McKone, 1995; McLennan & Luce, 2005; Brown & Gaskell, 2014; and elsewhere throughout the literature) goes as follows: if TSEs and abstract repetition priming are found together at long-distances, then we can assume that the two of them stem from the same representation. If, however, the decay profiles of these two types of information are distinct, then we have found evidence that different representations must be causing the priming effects.

Other research compares the relative decay patterns of TSEs from different types of information (e.g., voice and speech rate) to hypothesize about the processing stages involved in word recognition. Most notably, the work of McLennan & Luce (2005) compares talker-specific effects with those of speech rate. Finding differences in the two decay patterns, they conclude that allophonic effects (i.e., speech rate, which for them is accomplished through phonetic reduction) are seen early and disappear before talker-specific effects appear.

Finally, studies investigating the pattern of repetition priming in the absence of token-specific manipulations have shown that two patterns of priming effects exist (discussed in Section 3.2.4). Strong, short-term priming effects are super-imposed on long-lasting but weaker priming effects. These two patterns combine to form a logarithmic decay profile for speech information. This is important for the purposes of this thesis as TSEs could theoretically impact either or both of these two patterns of priming.

1.1.4.2 Processing effort and processing time

The contributions of processing effort and processing time are another main source of theorizing for models of speech processing. The idea is that the additional attention due to increased processing effort may serve to highlight different aspects of what one hears. For example, focusing a listener's attention on how a word sounds may cause a more perceptual, detail-rich representation of that word to develop, while focusing attention on what a word means may result in more of an abstract representation. By manipulating these types of features, researchers have attempted to determine whether the pattern of TSEs depends on the level of processing.

Another often-manipulated experimental feature is the amount of time allowed for processing. In the literature, processing time ranges anywhere from less than 40ms in subliminal masked priming, in which prime items are presented below the level of conscious awareness, to over one second, in the case of items presented with a long inter-stimulus interval (ISI). The reasoning is that the longer a participant is allowed to process an item, the more detail may be stored in the representation of the item.

1.1.4.3 Implicit vs. explicit tasks

Comparisons of the types of task used to find TSEs is also frequently found in the literature. Following the terminology of Graf & Schacter (1985), we distinguish two types of tasks: *implicit* and *explicit*. Implicit tasks present items without requiring participants to access their conscious memories of what has previously occurred. Explicit tasks, on the other hand, ask participants whether they remember having heard an item before in the context of the experiment. Schacter (1987) discusses several reasons to separate the two into different memory types, which we review in Section 4.2.1. Assuming these reasons are sound, then it is quite likely that tasks tapping into implicit and explicit memory will show different patterns of TSEs. Below we summarize the tasks used in the literature.

- ***Implicit tasks***
    - **Lexical Decision:** Indicate whether an item is a word or not
    - **Shadowing:** Repeat the item as accurately and quickly after hearing it
    - **Stem Completion:** Given a few phonemes (or letters), complete a word
    - **Perceptual Identification:** Identify an item presented within noise
    - **Semantic/Perceptual Classification:** Identify a semantic/perceptual attribute of an item
- ***Explicit tasks***
    - **Blocked Word Recognition:** Indicate whether the item has been heard before; primes presented separately with an implicit classification task
    - **Continuous Word Recognition:** Indicate whether the item has been heard before; primes and targets mixed together
    - **Perceptual Discrimination:** Identify whether some property of the perceived item (e.g., voice) matches a previous encounter with the item
    - **Cued-Recall:** Upon presentation of some part of an item, fill in the rest to match what was previously perceived

The novel experiments presented in this thesis make use of lexical decision (Chapter 3), continuous word recognition (Section 4.3), and blocked word recognition tasks (Section 4.4). By contrasting the three, we investigate whether the resulting decay pattern of TSEs differs between the memory representations accessed by implicit and explicit tasks.

1.1.5 Summary

One of the enduring mysteries in the field of speech perception involves the mental representations of abstract lexical and token-specific information. Listeners clearly perceive both but, since the mental representations involved proceed from the same auditory input stream, it remains unclear how to model the perceptual process. Unitary models of speech perception focusing on either abstract or token-specific information have both been proposed. As the resulting investigations yielded equivocal findings, these simplistic unitary models were abandoned in favor of hybrid models, which assume that representations of both abstract and token-specific information exist.

In this thesis, we distinguish between two broad types of hybrid models. The first type posits an early separation of representations, such that combined, detail-rich representations and separate, abstract representations of speech are distinct. The second type consists of late abstract representations being built from earlier detail-rich representations. The interesting difference between these types of hybrid models concerns the interaction between separate and combined representations of word and voice information in speech processing. The former highlights the causal relationship between abstract properties of the speech signal and lexical representations while the latter predicts that all token-specific detail should influence perceptual events.

The goal of this thesis is to distinguish between these two types of models by determining whether the speech recognition processes in the brain represent abstract and token-specific information as separate or combined representations. To do so, this thesis presents a set of experiments which investigate the decay patterns of TSEs in both implicit and explicit tasks. We find evidence that TSEs only affect future perceptual events within a very short time window. We further find that this effect operates similarly in both implicit and explicit tasks. These results conflict with the established literature that TSEs are long-lasting in both implicit and explicit tasks. In light of these findings, we recommend a renewed focus on speech perception models which emphasize the abstract properties of speech. While we do still find TSEs and therefore must account for them in speech perception models, our findings de-emphasize them as the necessary representational source leading to word recognition.

1.2 Overview

The goal of this thesis is to investigate the nature of speech perception and how to model it using both abstract lexical information and token-specific detail. In this chapter, we proposed a division of the current hybrid models into two types. This division separates hybrid models built using combined representations from those built using separate representations. The former emphasize the early importance of detail-rich representations of the speech signal while the latter separate representations of abstract, invariant lexical properties, which are causally necessary for word recognition, from representations of token-specific properties, which are needed for processes like speaker recognition. The investigation in this thesis is accomplished in two broad steps. The first is to summarize the previous findings of the literature and re-cast them using the theoretical framework proposed in this chapter. The second is to conduct a set of minimally contrasting experiments which look for the presence of TSEs using modern analysis techniques. The results are then applied to evaluate between the predictions of these hybrid models.

In doing so, CHAPTER 2: Experimental Methods sets up the common set of methodological steps which will be used in this experimental thesis. Through the use of linear mixed-effects models, experimental designs with sufficient statistical power, and a carefully curated set of stimuli, we hope to remove some potential confounds and unnecessary noise that might have been affecting previous experimental findings. Additionally, CHAPTER 2: Experimental Methods describes the results of an initial experiment verifying the finding that TSEs exist upon immediately contiguous presentation of prime and target. Specifically, switching the voice of stimuli between two different male speakers reduces the priming facilitation on reaction time from cases where the same speaker produced both members of the prime/target pair.

CHAPTER 3: Time-Course in Implicit Tasks builds upon these findings by exploring and re-visiting previous experimental findings in the literature using tests of implicit memory. First, a discussion of the theoretical work on auditory and visual short-term memory is provided as background. This discussion sets up the basic concepts which we use throughout the thesis while also refining the focus of study. Then, experimental findings testing response accuracy (Section 3.2.2) and reaction time (Section 3.2.3) are summarized, leading to the conclusion that equivocal results supporting both abstract and episodic models have been found. The final background section describes the expected dual nature of repetition priming effects from the literature; that is, TSEs could be seen in both short- and long-term priming patterns. Three experimental studies using the lexical decision task are then presented which show a concrete decay pattern in the time-course of TSEs. Specifically, effects of voice are shown to persist until up to three items intervene between prime and target. This result matches well with studies previously testing reaction time but does not conform to the predictions from those testing recognition accuracy.

To follow up on these results, CHAPTER 4: Time-Course in Explicit Tasks focuses on explicit memory tasks which look at the effect of TSEs on recall accuracy. First, the distinction between implicit and explicit tasks is motivated from previous behavioral studies in Section 4.2.1. This proves to be important both in understanding the motivation for episodic models in the literature and in dealing with the potential cross-contamination of implicit and explicit effects of memory on experimental results. In the next sections, studies using the blocked (Section 4.2.2) and continuous (Section 4.2.3) word recognition paradigms are summarized. The broad conclusion from these is that long-lasting TSEs are found, normally with same- versus different-voice manipulations. The facilitatory effect of perceptual similarity however, which is a hallmark of models based on token-specific detail, turns out to be not as conclusive as has been claimed. The first experiment presented in Section 4.3 slightly modifies the design of the experiment in Section 3.5 to use the continuous word recognition paradigm. Similar results are found, with TSEs only marginally present until around three items intervene. The final experiment in Section 4.4 uses a new set of stimuli equally spanning the word frequency range and also does not find long-lasting TSEs. These results stand in direct conflict to what has been reported in the literature to date using explicit tasks of memory.

Finally, CHAPTER 5: General Discussion summarizes and concludes this investigation. The review of the literature indicates that both abstract properties and token-specific detail are relevant in word recognition. For this reason, hybrid models combining the two have been proposed. Our results stand in conflict with the findings in general and, if they withstand future replications, motivate a return to models of word recognition that focus primarily on the abstract properties present in the speech signal.

CHAPTER 2: Experimental Methods

2.1 Introduction

This chapter introduces the experimental methods that were developed for the experiments presented in this thesis. In creating these methods, care was taken to ensure that minimal noise would be introduced into the experimental results through the creation of a well-controlled set of auditory stimuli. Data trimming procedures were also implemented to remove potentially impactful outliers from the analysis (Baayen & Milin, 2010). Additionally, using mixed-effects models with parsimonious random effects structures (Bates et al., 2015a) improves the generalizability of the statistical findings by avoiding the *language as a fixed effect* fallacy (Clark, 1973). It also presents the results of an initial experiment illustrating that TSEs due to voice information are found in reaction time for immediate prime/target pairs. Switching the speaker of a word from one male to another increases the reaction time compared to switching between word tokens from the same speaker. Future studies in this thesis build upon this design to investigate the time-course of TSEs.

2.2 Experimental Methods

As the studies reported in this thesis all use the same methodology developed through collaborations between the Language Variation & Cognition Lab (PI: Meredith Tamminga) and the Experimental Morphology Lab (PI: David Embick), the common methods will be presented once here and modified as needed in the individual experimental sections.

2.2.1 Participants

Participants were either recruited from the experimental subject pool managed by the Department of Psychology at the University of Pennsylvania or through the online experimental hosting platform Prolific (Damer and colleagues, 2018). For the former, self-identified native speakers of English volunteered and were administered studies under Meredith Tamminga's IRB by our lab manager Elisha Cooper or by the author. For their participation, they were given course credit commensurate with the time taken to complete the study including any potential travel time. All experiments were described as "*This experiment involves listening to a number of sounds and indicating whether the sounds are words by pressing a button.*" No participants participated in more than one experiment and all were assigned sequentially to one of the experimental lists when appropriate.

For the participants recruited from Prolific, participants were restricted to be born in either the United States or Canada with English as their first language. They were encouraged to use headphones and the Google Chrome browser to complete the experiment in order to better standardize response timing. The instructions and recruitment materials differed slightly depending on the study and will be mentioned in the appropriate sections. For their participation, they were given $2.50. Upon accepting, participants had 90 minutes to complete the study before their participation was ended. None participated in more than one study across those hosted on Prolific and all were assigned sequentially to one of the experimental lists when appropriate.

2.2.2 Stimuli

A total of 5279 unique recordings of 983 stimuli were used in the experiments presented in this thesis (see APPENDIX I: Experimental stimuli for the full list). These consisted of 595 words and 388 non-words. The actual items for a given experiment were a subset of the total number, which are elaborated on in each experimental section. Three male speakers of American English from the regions of Philadelphia (MA1), Chicago (MA2), and Detroit (MA3) and one female speaker (FM1) from the Philadelphia region recorded all stimuli. The dialects of each speaker were slightly different, with speakers MA1 and FM1 from the North Midland dialect region while MA2 and MA3 were from the Upper Midwestern dialect region. Although all speakers were trained linguistics and aware of local dialect features, any potential differences between their pronunciations should only serve to highlight the voice-specific properties tested in the experiments and, we trust, did not introduce a confound in our studies. Further work is needed to look at the differences in lexical access between speakers with different accents, but for an excellent discussion of this, see Gylfadóttir (2018).

All speakers recorded multiple tokens of the relevant words and non-words in different recording sessions using a Blue Snowball microphone in a soundproof booth. Blank space at the beginning and end of each recording was trimmed to leave only the acoustic signal of the stimuli remaining. Using either a custom Praat script (Boersma & Weenink, 2016) or a Python script, the average amplitude of each file was normalized to 70 dB SPL, which served to reduce loudness differences between the items. The following table presents summary statistics of the durations for each speaker.

Table 1: Summary of stimuli durations per speaker

Speaker	**Means & Standard Deviations**
MA1	*538ms (97)*
MA2	*479ms (89)*
MA3	*487ms (79)*
FM1	*504ms (75)*

The average properties of the stimuli are outlined below (see APPENDIX I: Experimental stimuli for the individual values). Individual experimental sections present the relevant subset of the data. Stimuli were selected over the course of the author's graduate student career, starting from the stimuli list found in Brennan et al. (2014). Age of acquisition measures (AOA) come from Kuperman et al. (2012), neighborhood density measures from the *English Lexicon Project* (ELP: Balota et al., 2007), frequency measures from the *lg10CD* measure of *SUBTLEX-US* (Brysbaert & New, 2009), and concreteness measures from Brysbaert et al. (2014). Bi-gram phoneme frequency measures, along with other syllabic frequency measures for onsets, vowel nuclei, and codas, were calculated by the author using the partial list of the ELP and the CMUDict Pronunciation Dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). Syllabification information came from the CMUDict when available or was generated from Constantine Lignos' *lingtools* package (https://github.com/lingtools/lingtools). First, the set of lexemes from the ELP was collated and pronunciations, complete with syllabification, were obtained. The phonological frequency measures were then calculated over the partial ELP list by calculating the number of occurrences of a given feature over the total count of other possible features (e.g., the total count of bi-gram *'R AA'* over all other bi-grams). These measures were calculated from the counts only and were not weighted by frequency of occurrence. These phonological frequency measures were only used in non-word creation to insure that the non-words used in these experiments were phonotactically licit and valid potential words; they do not appear in any of the statistical comparisons. The non-words were designed to be as similar to real words as possible; optimally changing only one phoneme from a real word. All non-words in this thesis are transcribed using the ARPABET phonetic alphabet used by the CMUDict Pronunciation Dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict). The following table presents summary statistics for the lexical properties we consider in this thesis.

Table 2: Summary statistics for the stimuli in this thesis

	Means & Standard Deviations
Frequency (*lg10CD*)	*2.41 (0.77)*
Age of acquisition (years)	*6.85 (2.60)*
Phonological Neighborhood Density	*14.75 (9.89)*
Concreteness	*4.24 (1.02)*
Number of phonemes (words)	*3.53 (0.64)*
… (non-words)	*3.56 (0.54)*
Average bi-gram (words)	*0.0059 (0.0025)*
… (non-words)	*0.0058 (0.0027)*

2.2.3 Design

The experiments in this thesis were designed using three tasks although the majority of the studies used the continuous lexical decision task. In this task, which is illustrated below, participants are instructed to decide as quickly and as accurately as possible whether the sounds they hear are words or not. Preceding the experimental section, participants made responses to ten practice items, which were chosen to reflect the distribution of speakers' voices in the experiment. Feedback was, for the most part, never given during the experiment. However, after the practice session for in-lab studies, participants were given the opportunity to ask the experimenter for any clarification. In all of the lexical decision experiments presented here, the same number of words and non-words are presented. The experimental items were presented in three or four blocks, with mandatory breaks for the participant in between. No critical manipulations occurred across blocks (i.e., all repetitions occurred within a block). When possible, randomizations of the items occurred at the participant level in order to obviate any effects of presentation order. The inter-stimulus interval (ISI) was, unless otherwise indicated, randomly determined from a range of 400 to 600ms on a by-item basis to avoid participants getting into a response rhythm.

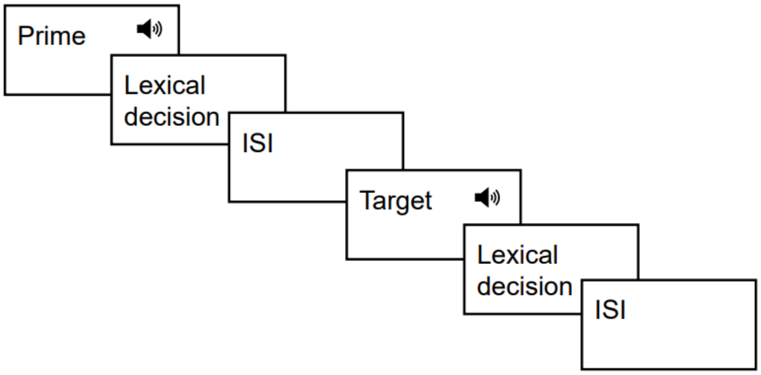

Figure 6: The continuous lexical decision task (from Goodwin Davies, 2018)

In CHAPTER 4: Time-Course in Explicit Tasks, two additional tasks are used to examine implicit vs. explicit task effects. These are the continuous word recognition and blocked word recognition tasks, which will be described in the relevant experimental sub-sections.

The most common manipulations involve the factors VOICE, DISTANCE, and SPEAKER. TSEs are investigated through comparing same-voice pairs to different-voice pairs. Any differences between these two constitute TSEs. Additionally, comparing the TSEs over different distances illustrate the effect of token-specific detail on speech representations. Throughout the background sections and some experiments, the factor GENDER is also manipulated. When used in this thesis, the word *gender* refers to the biological differences between male and female voices, primarily due to differences in vocal tract length, and is not used to imply any distinction between gender-identity.

2.2.4 Procedure

Experiments were either conducted in the lab or online. For the former, stimuli were binaurally presented in pairs through headphones to the participants on iMac computers in the lab spaces of either the Language Variation and Cognition lab or the Experimental Morphology lab. The continuous lexical decision tasks were presented using PsychoPy (Peirce, 2007) with the pairing of stimuli unknown to the participants. Additional tasks include continuous and blocked word recognition, which will be described when appropriate. Participants responded to stimuli using an Empirisoft Rotary Controller, with one button indicating a ‘*Word*’ response and the other a ‘*Not a word*’ response. The ISI between each item was measured from the end of the sound file or the participant’s response, whichever was later. Response times were calculated from the start of each item. There was no time-limit given on how long a participant could take to respond to an item; the next item would be presented only after a response was obtained. Therefore, outlier removal needed to be conducted on the resulting data.

For the studies conducted online, we used the experimental presentation platform Ibex (an acronym for *Internet Based Experiments*: Drummond, 2017). The tasks were constructed using various iterations of the *PennController* library (Zehr & Schwarz, 2018), which is a custom JavaScript library for psycho-linguistic experimentation created in the Experimental Study of Meaning lab at the University of Pennsylvania (PI: Florian Schwarz). The ISI between each item was again measured from the end of the sound file or the participant's response, whichever was later, and response times were calculated from the start of each item. Studies linked from the experimental subject pool to Ibex were described with "*This experiment involves listening to a number of sounds and indicating whether or not the sounds are words by pressing either the 'F' or 'J' keys. A brief demographic questionnaire follows the experiment.*" Again, participants were not limited in how long they were given to respond to items; subsequent items were presented after a response was obtained.

2.2.5 Analysis

In the analysis section of each experiment, the post-collection data preparation methods are summarized. For the continuous lexical decision tasks, these include participant and item removal, global reaction time cutoffs, and finally by-participant and by-item trimming. For the continuous and blocked word recognition tasks, slightly different procedures needed to be considered. The same over-arching mindset behind this process however was to insure that the data analyzed with our statistical tests was free of data from participants who were not performing the task. These data points have a higher probability of becoming influential outliers, which could have introduced errors into our analyses. Especially since the statistical tools used in this thesis assumed that responses were distributed in a Gaussian distribution, we feel these procedures are justified.

Beginning with the overall participant removal process, we sought to remove participants who were objectively not performing the task. This was an important step, especially for the data collected online. One of three criteria had to be met for a participant to be removed from the statistical analyses. Either they had 1) a global accuracy score less than 70% correct (quite reasonable for any native speaker paying somewhat close attention), 2) more than twenty absurdly fast responses (i.e., less than 300ms, indicative of participants holding down buttons to speed through parts of the experiment), or 3) were indicated as outliers in a boxplot of the Hodges-Lehmann estimates (described in the next section) of the central tendency of all participants' reaction time distributions (indicative of non-standard participation). These criteria were established to provide objective measures of determining whether a participant faithfully participated in all parts of the experiment. Item removal consisted of a global accuracy calculation, where items which were responded to with less than 50% accuracy were removed from consideration for all participants. Viewing the experiments chronologically, these items were eventually replaced with others to create a more balanced, recognizable set of stimuli.

The inspiration for the next steps of individual response trimming is found in Baayen & Milin (2010), who advocate for a three-step process in reaction time analyses. First, a global reaction time cutoff is applied to the data. Individual responses less than 300ms or greater than 3000ms were removed. These numbers were used in all of the lexical decision experiments and were determined to be adequate for responses to the monosyllabic words used in this thesis. Second, by-participant and by-item trimming was performed by visualizing the reaction time distribution of each and creating specific low and high cutoff points to remove obvious outlying responses. Third, the initial fit of the model to the data was visualized to determine if it was distressed (i.e., the residuals plotted against fitted data indicated a poor fit at the left and right ends of the distribution; a pattern present in all of our experiments testing reaction time). Data with residuals greater than 2.5 standard deviations from the mean were then removed and the model was re-fit to the resulting data.

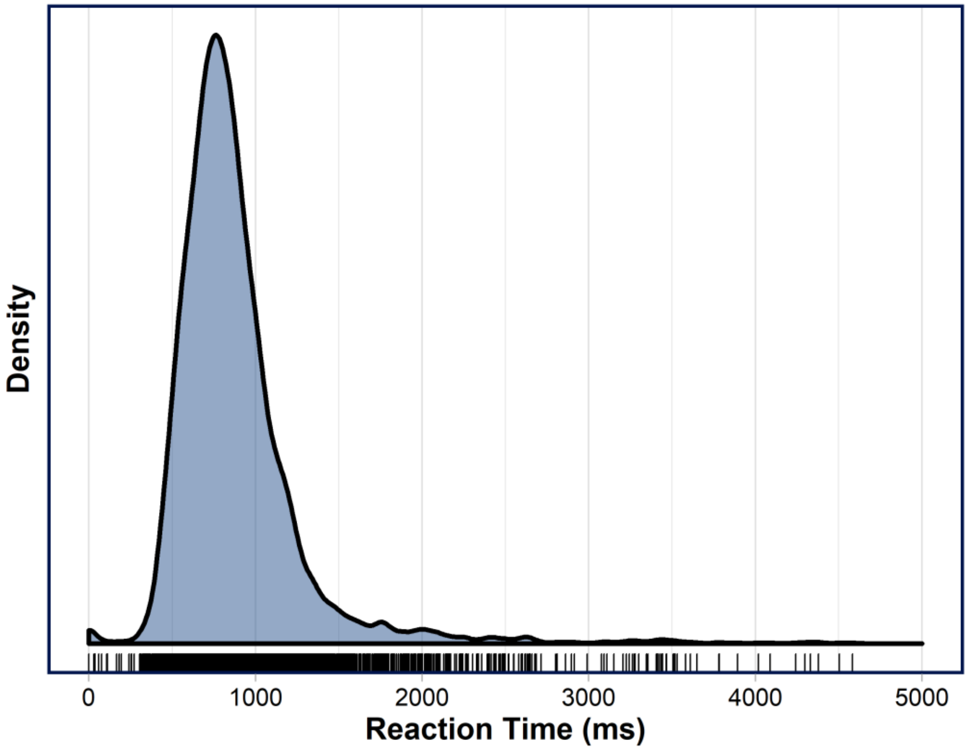

Figure 7: Example reaction time distribution (pre-trimming)

In conclusion, we feel that these methods are justified in creating analyzable datasets of reaction time data. All reaction time data is described visually as a Gaussian curve with a long right tail (sometimes described as an ex(ponential)-Gaussian distribution) seen above in Figure 7, as no restrictions on a maximum response time were imposed. All observations are either valid or invalid, depending on what the participant was responding to at the time (or whether they were checking their smartphone). Those observations in the long right tail have a higher probability of being invalid data, as normal lexical decision responses do not take multiple seconds. Our data trimming procedures basically serve to monitor the long right tail, with the hope of obtaining more generalizable results.

2.2.6 Results

Before the results of the reaction time and accuracy analyses are reported, the raw data is provided in tabular format. These tables include a break-down of the reaction time, accuracy, and counts of the target data in each combination of experimental conditions, along with the relevant data from the primes and the fillers. The summarized data is not trimmed except for the global reaction time cutoffs of 300 and 3000ms described above. For that reason, the reaction time reports are not means but Hodges-Lehmann estimates of location, which is a non-parametric method of finding the central tendency of a group of data. This method is especially valuable when the assumption of normality is not met, which is definitely true in raw reaction time data. Accuracy is reported as hit rates (i.e., the percentage of correct responses).

2.2.7 Accuracy

For studies of reaction time in lexical decision experiments, the accuracy data is of less interest. Priming effects are occasionally seen in accuracy measures, with small shifts occurring between prime and target. Overall, accuracy rates are quite high however, hovering in the 85% to 95% range, which does not provide much room for effects to be seen. The most concerning aspect to be considered in accuracy measures is the *speed-accuracy tradeoff* (see Bogacz et al., 2009 for a recent investigation), where slower responses are additionally more accurate than faster responses. For this reason, there have been attempts to improve the modeling of reaction time data by unifying the two types of information into one model (cf. Ratcliff et al., 2004 and Wagenmakers et al., 2007).

In this thesis however, we have chosen to model accuracy of lexical decision tasks using generalized logistic mixed-effects models of hit rates, which have the benefit of being directly comparable to the models of reaction time. These were implemented using the *lme4* package (Bates et al., 2015b) in the R statistical package (R Development Core Team, 2008). The *bobyqa* optimizer was used throughout to help with convergence issues. Additional methods, packages, and the included predictors are similar to those described in the next section. Finally, graphs of accuracy data seen at the beginning of each relevant section include the mean hit rates and the 95% confidence intervals, after the removal of between-subject variability (Cousineau, 2005; Morey, 2008).

2.2.8 Reaction Time

The statistical analysis of the experiments was influenced by the description of reaction time analyses provided by Baayen & Milin (2010) along with the analyses described in Bates et al. (2015a). Linear mixed-effect models of the word and non-word results were separately constructed using the *lme4* package (Bates et al., 2015b) in the R statistical package (R Development Core Team, 2008). The dependent variable in the models was $\log_2$-transformed reaction time to the target items. The critical independent variables along with their coding schema are discussed in each specific experimental section. Control predictors were included in the model to control for potential effects due to properties of the stimuli or participants. These were centered and scaled by standard deviation measures (i.e., z-scored) or were included as categorical factors which were sum-coded (also called *simple coding*) so the intercept reflects the grand mean across all of the levels. Starting with the z-scored variables, we included predictors for frequency, phonological neighborhood density, age of acquisition, duration, trial number, $\log_2$-transformed reaction time to the previous item (following Baayen & Milin, 2010), and age (when recorded). These are described in more depth in Section 2.2.2. Categorical predictors included participant gender, participant handedness (when recorded), and participant group.

Following the procedure to find the most parsimonious random effect structure in Bates et al. (2015a), models with random intercepts and the set of critical predictors for both subject and item random effects was initially evaluated. On a per-model basis, the random effects structure was iteratively simplified by removing elements which accounted for the least variance until a chi-squared test indicated a significant change in model fit. Each experimental subsection reports the random effect structure but for the most part we will not be concerned with it further in this thesis.

Significance of effects was evaluated by obtaining estimates of p-values using the Satterthwaithe (1946) approximation to degrees of freedom found in the package *lmerTest* (Kuznetsova et al., 2015). The marginal and condition R^2 terms describing how well the models fit the data were determined from the *sjPlot* package (Lüdecke, 2018) using the *MuMIn* package's implementation (Bartoń, 2018) of the Nakagawa and Schielzeth (2013) method. Other packages used in the analyses include *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015) to facilitate the construction of fixed-effects structures, *sjPlot* to generate additional plots and tables (Lüdecke 2018), *ICSNP* for non-parametric calculations (Oja et al., 2006; Tyler et al., 2009), *ggplot2* (Wickham, 2009) to generate graphs, and *plyr* (Wickham, 2011), *dplyr* (Wickham & Francois, 2015), and *reshape2* (Wickham, 2007) for data manipulation.

The initial model in each experiment looks at the main effect of abstract repetition priming in each of the experimental conditions by dummy-coding a factor with levels corresponding to each experimental manipulation group. This model is used to verify that there are significant differences between the primes and targets. Additionally, this model is used to verify that the priming effects survive a multiple comparisons correction using Holm's method from the *lsmeans* package (Lenth, 2016). The figure preceding the discussion of the initial model plots the Hodges-Lehman estimate of the participants' reaction times along with the 95% confidence interval after the removal of between-subject variability (Cousineau, 2005; Morey, 2008).

The main model presented afterwards is the one examining the main effects and interactions between the experimental predictors. When needed, other follow-up models are reported to clarify the reaction time priming patterns in subsets of the data. Visualizations of the trimmed reaction time data are provided with the model summaries. Effect sizes are obtained from the model by calculating the percentage change in reaction time due to a given coefficient, as is appropriate given the fact that the dependent variable was $\log_2$-transformed. This percentage change is applied to a word of average duration to give the average effect size in milliseconds. The figure accompanying the main models indicates the median prime reaction time and the distribution of the reaction time from each relevant condition. Finally, any post-hoc variables are evaluated by conducting chi-squared tests of the models with and without the post-hoc variable in question.

2.3 Experiment 1

This first experiment verifies that TSEs due to switching the voice of stimuli exist when primes and targets are presented contiguously. Two male voices were used in this study to avoid the added confound of switching genders. Given the understanding from the literature, we strongly predict that even the small perceptual difference between two male voices should result in TSEs in the priming pattern.

2.3.1 Method

2.3.1.1 Participants

A total of 34 (age range 18-22, mean 19.8; 11 male) undergraduate participants were run in the fall semester of 2015. They were recruited from the experimental subject pool at the University of Pennsylvania and voluntarily completed the study in person using a custom PsychoPy implementation of a continuous lexical decision task at the Language Variation and Cognition Lab.

2.3.1.2 Stimuli

In total, 600 stimuli were used in this experiment (300 each of words and non-words), recorded by speakers MA1 and MA2 (seen in APPENDIX I: Experimental stimuli). These stimuli will also be used in Experiment 2 and Experiment 3. 100 items were repeated in word and non-word conditions. The remaining 200 words and 200 non-words served as fillers to hide the manipulation of repetition. Figure 8 below illustrates the relationship between frequency (mean = 2.94, standard deviation = 0.48, range = [1.40, 3.92]), age of acquisition (mean = 5.24, standard deviation = 1.24, range = [2.5, 7.9]), and phonological neighborhood density (mean = 17.95, standard deviation = 10.30, range = [1, 47]) for the words in this experiment. As can be seen, the words used in the experiment were primarily selected from a relatively high frequency range. Slight trends can be seen for age of acquisition and phonological neighborhood density, which justifies including them in our statistical models.

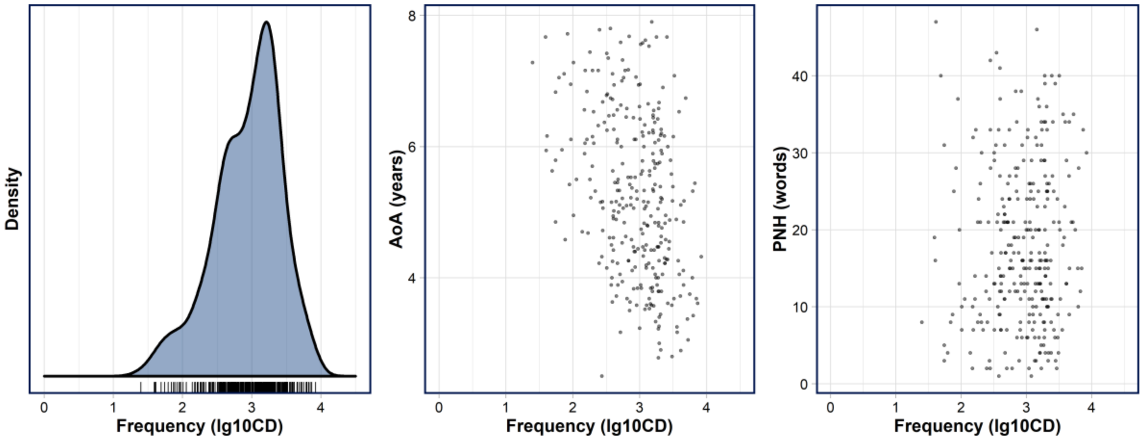

Figure 8: Properties of Experiments 1-3 word stimuli

Figure 9 shows the phonotactic probability of both the words on the left and the non-words on the right using the calculated bi-gram metric described in Section 2.2.2. The first facet plots the mean bi-gram frequency and the second plots the standard deviation. The third and fourth facets plot the minimum and maximum bi-gram frequency from each stimulus. Overall, the non-words are slightly phonotactically less licit than the words, with lower means, higher deviations, and lower minimums and maximums. We do not expect these small differences to drastically impact participants' task strategies in the experiments however.

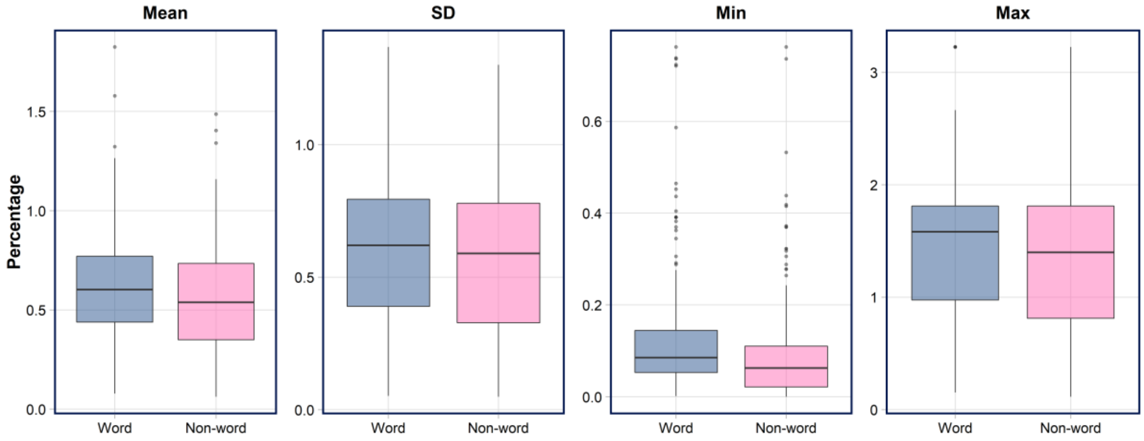

Figure 9: Phonotactic prob. of words (left) & non-words (right) in Experiments 1-3

2.3.1.3 Design

The 300 words and 300 non-words were divided into two sets, the repetition (100 items each) and non-repetition sets (200 items each). Within the word and non-word repetition sets, eight lists were created by varying the factors VOICE$_{\text{[Same vs. Different Voice]}}$ × SPEAKER$_{\text{[MA1-prime vs. MA2-prime]}}$ × DIRECTION$_{\text{[1-2 vs. 2-1]}}$. The DIRECTION factor details which of two different sound-file tokens appeared as the prime or the target; we will not discuss this factor further. Therefore, 25 words and non-words existed in each possible grouping of these two conditions. Non-repetition (filler) trials were randomly grouped together into pairs. All of these pairs were then randomly presented to subjects, giving a different trial order per each participant. The experiment was presented in a total of four experimental blocks. The number of word and non-word stimuli, and additionally the number of repeated vs. non-repeated items from each, were equal across all blocks, resulting in a repetition rate of 25% across the experiment.

2.3.1.4 Procedure

Participants completed ten practice lexical decisions as practice before the experiment. These consisted of the following five sets chosen from the non-repetition filler stimuli sets, randomly presented per subject (with subscript representing different speakers and bold face indicating non-words): *mark*$_1$ - *lamp*$_2$, *guard*$_2$ - ***geyk***$_2$, ***trowz***$_2$ - *rag*$_1$, ***vaebd***$_1$ - ***kuhg***$_1$, *school*$_1$ - ***wahng***$_2$. Following the practice session, four blocks of 200 items each were presented (800 total lexical decisions) using the custom PsychoPy continuous lexical decision task implementation described in this chapter. All experiments included in the analysis were completed on average within 23 minutes (range: 19-30 minutes). As mentioned, each participant had a random trial order, which serves to eliminate any effects due to trial order.

2.3.1.5 Analysis

Of the original 34 participants in the experiment, no participants were removed. They all had accuracy scores of over 70% on the experimental items. Additionally, none were indicated as outliers a Hodges-Lehmann estimate of the reaction time distribution or had multiple (i.e., greater than twenty) absurdly fast responses. The overall results per each item were examined. The non-words *suhls*, *daek*, and *theyz* from the filler list in addition to *traak* from the experimental list had accuracy scores of less than 50% correct. These items were removed and will not be reported further in the analyses. The table below describes the amount of data trimmed to create the dataset for the models of reaction time.

Table 3: Experiment 1 removal summary

	Observations	**Percentage**
Inaccurate trials	*230*	*6.8*
RT trimming (300 > RT < 3000)	*47*	*1.4*
By-participant trimming	*121*	*3.6*
By-item trimming	*83*	*2.4*
Total removed	*481*	*14.2*
Total remaining	*2919*	

2.3.2 Results

The following table reports the distribution of the data per the factors VOICE and SPEAKER. These data are reported after only the minimal global trimming procedure of unreasonable reaction times was applied. For that reason, the reaction time reports are central tendency measures from the non-parametric Hodges-Lehmann estimates. The accuracy scores given indicate the amount of correct responses out of the total, after all global participant and item removal was conducted.

Table 4: Experiment 1 data summary

VOICE			**SPEAKER**			**Total**
Same	*682*	*98.4*	**MA1**	*679*	*98*	***833***
			MA2	*686*	*98.8*	***830***
Diff.	*726*	*98.5*	**MA1**	*725*	*98.3*	***831***
			MA2	*726*	*98.7*	***832***
Filler	*948*	*91.8*	**MA1**	*944*	*90.7*	***3291***
			MA2	*952*	*93*	***3296***

2.3.2.1 Accuracy

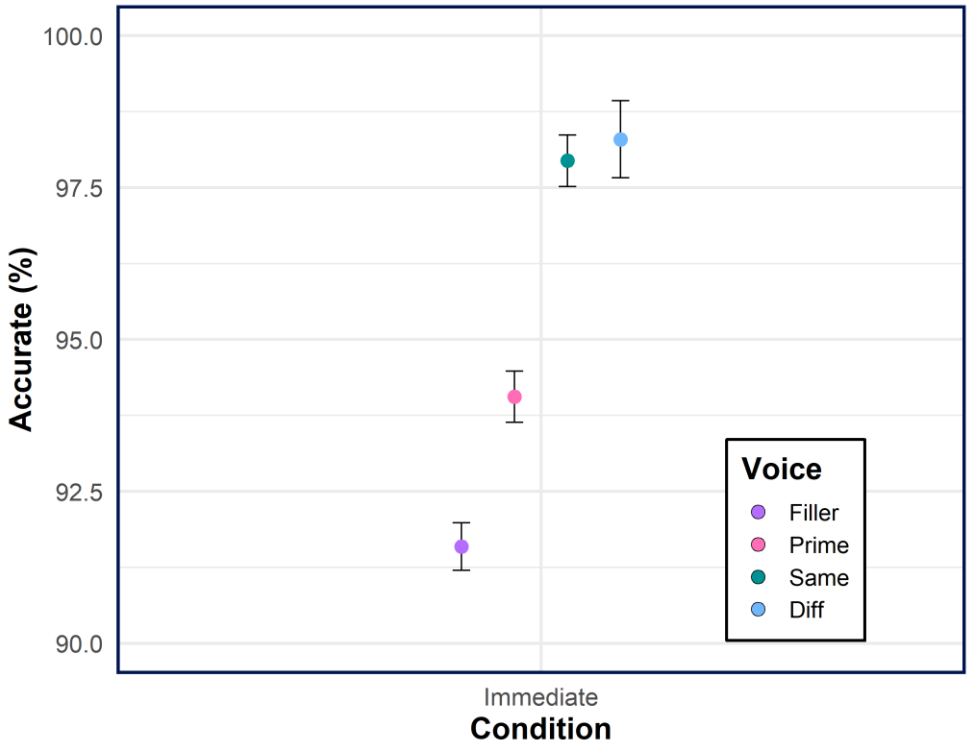

Figure 10: Experiment 1 accuracy data

A generalized linear mixed-effects model was fit to the accuracy data. This model combines primes and fillers together as the reference level for the condition factor, with targets in each combination of VOICE and SPEAKER dummy-coded. Random effects were set as intercepts for participants and items. The outcome of this model is seen in Table 5.

This model shows whether any priming was seen when items were repeated. Repeated targets were identified significantly more accurately (same voice: $p = 0.001$, different voice: $p < 0.001$). There was also a significant effect of speaker, such that words spoken by speaker MA2 were identified more accurately, but this did not significantly interact with the VOICE condition.

Table 5: Experiment 1 combined accuracy model

	Accuracy (combined)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	20.96	16.18 – 27.15	**<0.001**
VOICE			
Diff.	4.38	2.48 – 7.75	**<0.001**
Same	3.76	2.24 – 6.33	**<0.001**
SPEAKER			
MA2	1.30	1.02 – 1.66	**0.031**
Item (z-scored)			
Frequency	1.16	1.03 – 1.30	**0.013**
PNH	0.85	0.76 – 0.95	**0.004**
AoA	0.77	0.68 – 0.87	**<0.001**
Duration	1.08	0.96 – 1.22	0.204
Trial	0.80	0.73 – 0.88	**<0.001**
Participant			
zAge	1.11	0.89 – 1.38	0.350
Male	0.90	0.57 – 1.44	0.670
Group			
2	0.57	0.28 – 1.16	0.122
3	1.32	0.60 – 2.87	0.492
4	0.44	0.22 – 0.89	**0.022**
5	0.82	0.40 – 1.68	0.580
6	0.89	0.42 – 1.84	0.745
7	0.56	0.27 – 1.15	0.113
8	1.10	0.50 – 2.45	0.807
VOICE (Diff.) × SPEAKER (MA2)	1.20	0.51 – 2.83	0.678
VOICE (Same) × SPEAKER (MA2)	1.45	0.63 – 3.34	0.386
Observations	13203		
Marginal R^2 / Conditional R^2	0.158 / 0.379		

Turning now to the model examining the interaction of the experimental predictors of VOICE, Table 6 shows the results of the data considering only targets. VOICE and SPEAKER were dummy-coded, with the reference levels set to same voice pairs for speaker MA1.

Here we see the additional information that, while abstract repetition priming existed for accuracy in both conditions, no difference was found in the accuracy of same- and different-voice pairs. This indicates that no TSEs were found in the accuracy data for targets immediately presented after primes. With the conclusion that switching voices did not impact accuracy data, we now turn to effects on reaction time.

Table 6: Experiment 1 target accuracy model

	Accuracy (targets)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	2558.75	388.97 – 16832.29	**<0.001**
VOICE			
Diff.	1.25	0.50 – 3.11	0.629
SPEAKER			
MA2	1.79	0.41 – 7.82	0.438
Item (z-scored)			
Frequency	1.02	0.52 – 1.99	0.960
PNH	0.86	0.45 – 1.63	0.641
AoA	0.87	0.45 – 1.69	0.678
Duration	1.28	0.64 – 2.57	0.484
Trial	1.21	0.82 – 1.76	0.335
Participant			
zAge	0.95	0.60 – 1.49	0.812
Male	1.05	0.40 – 2.79	0.915
Group			
2	0.38	0.05 – 3.00	0.359
3	0.64	0.07 – 6.19	0.699
4	0.44	0.08 – 2.58	0.365
5	0.75	0.09 – 6.11	0.788
6	0.45	0.06 – 3.63	0.454
7	0.32	0.04 – 2.44	0.270
8	0.41	0.06 – 2.77	0.359
VOICE (Diff.) × SPEAKER (MA2)	0.67	0.17 – 2.72	0.579
Observations	3326		
Marginal R^2 / Conditional R^2	0.012 / 0.868		

2.3.2.2 Reaction Time

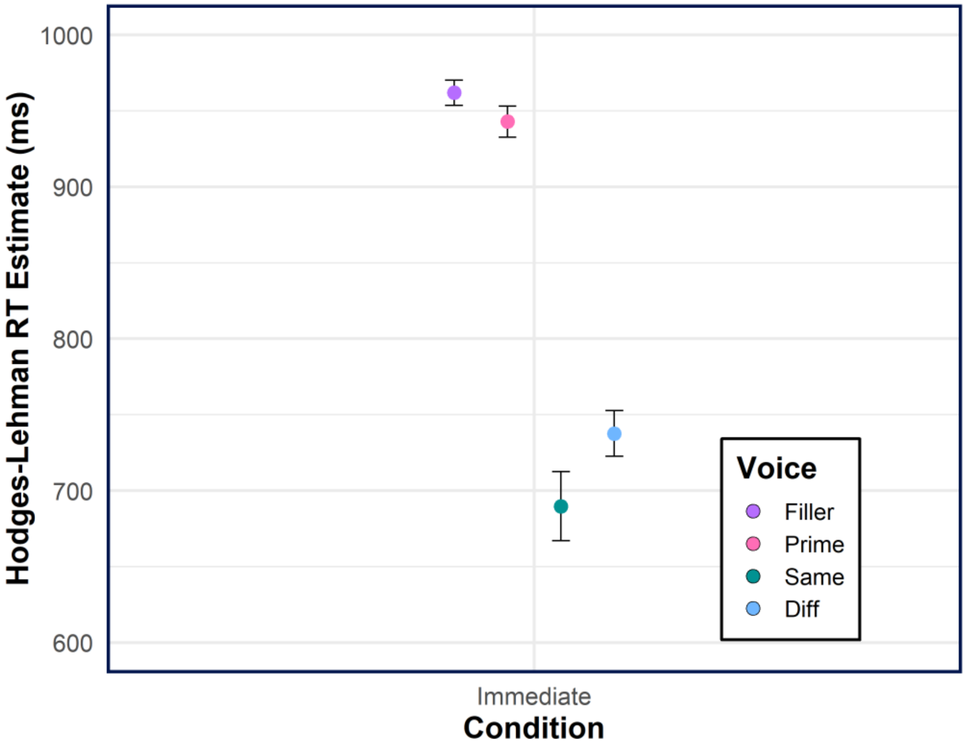

Figure 11: Experiment 1 reaction time data

The first analysis of reaction time comes from the large model investigating whether repetition priming existed in all experimental conditions. The responses to primes and targets were combined into one dataset and a linear mixed-effect model was run examining the $\log_2$-transformed response time. Random effects included by-subject and by-item intercepts. Fixed effects of interest were a dummy-coded variable with the baseline being responses to the primes and factor levels indicating each of the two VOICE conditions which interacted with a dummy-coded variable indicating the SPEAKER of the sound-file. Model criticism resulted in 174 additional observations being removed; a total of 2.98% of the remaining data.

Table 7: Experiment 1 combined RT model

	Log$_2$-transformed RT (combined)		
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.84	9.82 – 9.87	**<0.001**
Voice			
Diff.	-0.41	-0.43 – -0.39	**<0.001**
Same	-0.49	-0.51 – -0.47	**<0.001**
Speaker			
MA2	0.00	-0.01 – 0.02	0.772
Item (z-scored)			
Frequency	-0.00	-0.01 – 0.01	0.940
PNH	0.00	-0.00 – 0.01	0.560
AoA	-0.00	-0.01 – 0.00	0.198
Duration	0.03	0.03 – 0.04	**<0.001**
Trial	-0.02	-0.03 – -0.02	**<0.001**
Previous RTLog	0.26	0.25 – 0.26	**<0.001**
Participant			
zAge	0.04	0.01 – 0.06	**0.011**
Male	-0.01	-0.07 – 0.05	0.718
Group			
2	0.08	-0.01 – 0.17	0.090
3	0.01	-0.08 – 0.10	0.832
4	-0.05	-0.14 – 0.04	0.260
5	-0.01	-0.10 – 0.08	0.812
6	0.01	-0.07 – 0.10	0.742
7	0.01	-0.07 – 0.10	0.760
8	-0.02	-0.11 – 0.08	0.737
Voice (Diff.) × Speaker (MA2)	0.00	-0.03 – 0.03	0.983
Voice (Same) × Speaker (MA2)	0.04	0.02 – 0.07	**<0.001**
Random Effects			
σ^2	0.04		
τ_{00} Item	0.00		
τ_{00} Participant	0.00		
Observations	5664		
Marginal R^2 / Conditional R^2	0.754 / 0.783		

For the comparisons of interest, all factor levels came out significant (all $p < 0.001$), indicating that in each Voice condition, significant abstract priming was found (same-voice = 28.7% / 153ms, different-voice = 24.6% / 131ms). These comparisons remained highly significant after doing a multiple comparisons correction using Holm’s method. One additional interaction term came out significant: in the same-voice prime/target condition, speaker MA2 trials were recognized significantly slower than speaker MA1 ($p = 0.005$), but this significant term will not be interpreted further here. Overall, this model, indicates that significant abstract repetition effects were found.

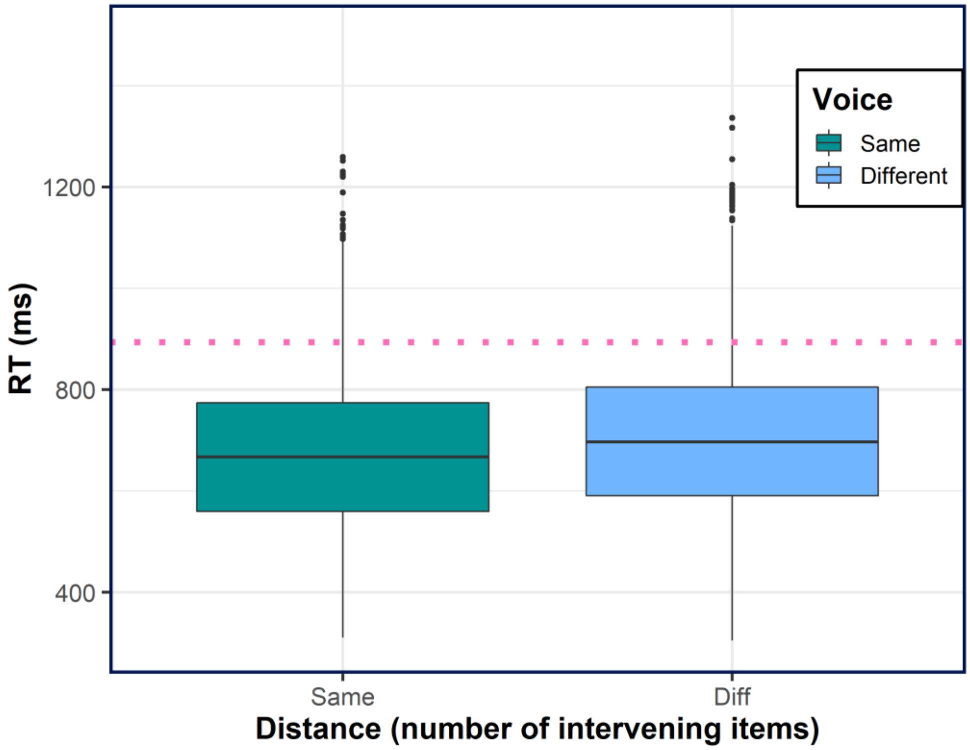

Figure 12: Experiment 1 trimmed reaction time

Next, we examine the model testing the differences between same- and different-voice conditions. The random effects for this model consisted of by-subject and by-item intercepts. The fixed effects of interest were formed by the interactions between two terms: the dummy-coded factors of VOICE (baseline = same-voice) and SPEAKER (baseline = speaker MA1). After fitting the model, 57 additional values (1.95% of the remaining data) with residuals > 2.5 SDs from the mean were removed.

Table 8: Experiment 1 target RT model

Log$_2$-transformed RT (targets)			
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.33	9.27 – 9.39	**<0.001**
VOICE			
Diff.	0.08	0.06 – 0.10	**<0.001**
SPEAKER			
MA2	0.07	0.04 – 0.10	**<0.001**
Item (z-scored)			
Frequency	0.00	-0.01 – 0.01	0.717
PNH	0.01	-0.00 – 0.02	0.246
AoA	-0.01	-0.02 – 0.01	0.343
Duration	0.09	0.08 – 0.10	**<0.001**
Trial	-0.06	-0.07 – -0.05	**<0.001**
Previous RTLog	0.09	0.08 – 0.10	**<0.001**
Participant			
zAge	0.08	0.02 – 0.14	**0.017**
Male	-0.03	-0.17 – 0.10	0.618
Group			
2	0.18	-0.03 – 0.39	0.100
3	0.05	-0.17 – 0.27	0.647
4	-0.11	-0.31 – 0.10	0.309
5	-0.03	-0.23 – 0.17	0.779
6	0.04	-0.17 – 0.24	0.728
7	0.04	-0.16 – 0.24	0.704
8	-0.03	-0.25 – 0.19	0.798
VOICE (Diff.) × SPEAKER (MA2)	-0.04	-0.08 – -0.01	**0.010**
Random Effects			
σ^2	0.04		
τ_{00} Item	0.00		
τ_{00} Participant	0.02		
Observations	2862		
Marginal R^2 / Conditional R^2	0.329 / 0.581		

The main effects for the factors of VOICE and SPEAKER of interest came out significant, indicating reaction time reductions from same- to different voice pairs ($p < 0.001$, 5.7% / -30ms) and from speaker MA1 to MA2 ($p < 0.001$). The interaction term between VOICE and SPEAKER however was not significant, indicating a general slow-down for words spoken by MA2.

2.3.3 Discussion

Overall, this experiment finds the following results.

- **Abstract repetition priming**
    - *Accuracy:*
        - All targets recognized more accurately than primes
        - Speaker MA2 recognized better than MA1
    - *Reaction time:*
        - All repeated targets recognized faster than primes
- **Talker-specific effects**
    - *Accuracy:*
        - No accuracy differences between same- and different-voice targets
    - *Reaction time:*
        - Different-voice pairs recognized *slower* than same-voice pairs

The analysis confirms the prediction that TSEs due to voice switches between two male speakers are found at immediate prime/target presentation. Later on in this thesis, we will expand on this simple design by introducing more distances between prime and target (as measured both with raw time and the number of intervening items).

2.4 Conclusion

In this chapter, we defined the experimental methods that the rest of this thesis will use. The methods generally apply to all of the studies, as the majority are conducted using the lexical decision task. By specifying the same methodological structure and subjecting each study to similar design practices, we aim to reduce experimental noise in the results. Later, in CHAPTER 4: Time-Course in Explicit Tasks, two additional tasks will be introduced, which will be discussed in the relevant experimental sub-sections. Finally, we established that these analysis methods, designs, and stimuli show the predicted effects of TSEs in a continuous lexical decision experiment looking at immediate repetition priming.

CHAPTER 3: Time-Course in Implicit Tasks

3.1 Introduction

This chapter focuses on examining the representations of speech by investigating the time-course of the impact of token-specific detail using implicit memory tasks, with a specific focus on voice information with lexical decision tasks. The goals of this chapter are two-fold: (1) we aim to synthesize the existing relevant literature relevant and (2) we present three experiments which, by comparing the decay patterns of TSEs, inform us of the nature of the mental representation of speech. The first goal is achieved through summarizing four separate lines of research: (a) theories of short-term memory, (b) studies investigating the perceptual identification of words, (c) studies using indirect tasks (lexical decision and shadowing in particular), and (d) decay profiles of abstract repetition priming. This examination of the literature reveals a missing component; namely, it remains unknown exactly how long TSEs last. At the end of this chapter, three experiments using the lexical decision task reveal that they only persist until around three intervening items have been perceived. This novel finding contradicts some previous research finding long-lasting effects of voice, specifically studies using the perceptual identification task, while corroborating other studies using lexical decision and shadowing tasks. The chapter concludes by emphasizing the importance of these results in modelling speech perception.

The rest of this section summarizes the findings of this chapter, starting with some thoughts about the nature of the literature which will be reviewed. Many studies of word recognition and memory decay exist but since the specific researchers were interested in separate questions and started from different assumptions, it is not easy to understand what has already been discovered. A brief discussion of these different threads will make the presented background more straightforward. The main questions that researchers have been interested in are (1) how short-term memory should be modelled, (2) how short-term representations actually fade, (3) what effect does the level of processing (indexed through either depth or length) have on the formation of speech representations, and (4) how do certain properties of a task influence TSEs.

First, the background section begins with a cursory introduction of foundational concepts from the literature on short-term memory. Most research assumes that representations are first activated and then decay over time in the absence of rehearsal. Research in this field has been mainly conducted using list memorization tasks, where properties of a list are modified and the corresponding effects on recall ability are measured. Recent research has found evidence that conflicts with the standard assumptions of memory, leading to a re-formulation of the structure of memory by highlighting processes such as cue-based recall. However, for our purposes, the standard set of assumptions will serve us well. As the focus of this thesis is on mental representations of speech and not on modelling domain-general memory structures, the terms used in the memory literature will be slightly modified here. Discussing the differences between the goals of this thesis and those of the short-term memory literature serves to motivate our discussion and base our terms on those in the literature.

Other research is interested solely in how these short-term speech representations persist in memory. Using mainly visual (with some auditory) repetition priming studies, a dual-route source of priming effects has been hypothesized. Early, strong priming effects exist which persist for a short duration after which a weaker, long-term priming effect remains. One question that often arises in this field is whether these priming effects decrease through the influence of time alone or because of intermediate processing of new events. For our purposes, this discussion is useful as it indicates what type of priming patterns we expect to find. TSEs could influence either one or both sources of priming.

Moving on to the complicated notion of level of processing, we first note that it is a broad idea with deep roots in psychology. As we all know, the salience of any item impacts perception. For example, this led people in the early days of the internet to design websites with flashing bright yellow text. The improved salience directs attention to the item, which in turn impacts the processing of the item. Turning to auditory properties, we first consider word frequency as a basic way of manipulating salience. An infrequent word (e.g., *frond*) will certainly draw more attention than a frequent word (e.g., *tree*), leading to increased attention and corresponding processing. Exactly how attention, processing, and perception are linked in the auditory mental system is unclear however. Many researchers investigate this question through changing either (a) the method or (b) the length of processing items.

Starting with the former, the general idea is that the mental representation of a word depends on which task was used to present the word. The main distinction lies between *perceptual* and *semantic* processing tasks (also called *encoding* tasks, indicating the way the participant interacts with an item to encode it in memory). Perceptual processing tasks direct participants to respond to the superficial form of an item, whether it is written in upper-case, spoken in a higher tone or different voice, or enunciated clearly for example.

Semantic tasks on the other hand ask participants to engage with deeper aspects, like the meaning, syntactic category, or other invariant property of an abstract word. If TSEs are only consistently seen when superficial properties of an item are highlighted, then we have learned that the mental representation of speech depends on how people are engaging with the items. The length of processing has been investigated in a similar manner. If more processing time is given a participant, then perhaps more properties of the words will be present in their mental representation. As we will see, many researchers approach the nature of auditory representations through the lens of levels of processing. While we are not concerned as much with this distinction, we do use the insights gained from these discussions in setting up the existing theories about mental representations of speech.

A similar idea is concerned with broader types of task manipulations. For example, other researchers hold the level of processing constant while they manipulate more global properties; e.g., the task length or the amount of repetition present. This is a different way of looking at speech representations that focuses more on *predictive* aspects of processing. By manipulating the task, this research aims to demonstrate a participant's changing expectations about what they will perceive. These expectations then influence the information a participant retains from a given stimulus. Clearly, this is a similar idea to manipulating the level of processing, but the method in generating conclusions remains different. A lack of targeted investigations of the two ideas causes unfortunate difficulty in understanding the predictions of both. However, knowing the assumptions behind each idea helps to synthesize the current theories of speech perception.

The background discussion in this chapter serves to motivate the experimental goal of determining the exact decay profile of TSEs in speech processing. It is still an open question about whether two separate representations exist consisting of separate abstract and episodic components with different decay profiles, or whether only detail-rich representations exist immediately following perception (as motivated in Section 1.1.3). Following the common methodology found in McKone (1995), McLennan & Luce (2005), and Brown & Gaskell (2014) for example, if effects of two types of information decay at different rates, we obtain indirect evidence that multiple representations are at play. If all effects decay at the same rate, then we have no evidence to confirm or deny the existence of multiple representations following perception. To forecast the experimental results, we find evidence that effects of talker-specific information persists for only a short time, disappearing after around three items have been processed, while abstract lexical properties persist throughout the length of the experiments.

As a final note, this chapter is concerned with studies of implicit memory. Recalling the introduction to these concepts found in Section 1.1.4, we are using the terms *implicit* and *explicit* memory to indicate the level that certain tasks refer to a participant's conscious memory. Explicit memory tasks require participants to respond using their conscious recollections of previously experiencing an item while implicit tasks are designed to not access conscious memory in this way. In these studies, effects of token-specific properties are seen only indirectly through, for example, reaction time differences. This division is one that is present in the literature on speech perception (see Graf & Schacter, 1985) and therefore we adopt it in this thesis. This chapter therefore makes only passing mention of studies of explicit memory; the focus of the presented background literature and the experiments is on implicit memory. CHAPTER 4: Time-Course in Explicit Tasks takes up the potential differences between implicit and explicit tasks, and we will postpone more discussion of explicit tasks until then.

3.2 Background

The general consensus is that token-specific, episodic information is available following perception, at least for a short time. Since token-specific information has been shown to impact repetition priming, it therefore needs to be somehow represented along with the abstract, semantic, and lexical content of a word. As we will soon see, there has been a great amount of equivocation about the importance of token-specific information, with some researchers finding TSEs at long distances and others at only short distances. This background section therefore highlights the lack of a targeted investigation of the decay profile of TSEs in implicit memory tasks. This investigation is needed to adequately model abstract and episodic properties of speech.

3.2.1 Models of Memory

A brief look at the theoretical insights from the literature on human memory starts the discussion which concludes by clarifying the object of study and defining the terms used in this thesis. Adopting the terminology from Nairne (2002), the *Standard Model* is a general description of several specific formulations of short-term/working memory. It is made up of three concepts: *activation*, *decay*, and *rehearsal*. Nairne (1996, 2002) starts by defining *activation* broadly as the initial creation of a memory trace resulting from some cognitive process, like speech perception. This trace is a transformation of the raw input signal which decreases in strength, over either the passage of time or the accumulation of mental representations from other cognitive processes (or both). This process is defined as *decay*. Keeping representations active requires *rehearsal*, which is an active process refreshing or reactivating the memory trace. *Short-term memory* is then defined as the set of stimulus representations that are active enough to be available for further processing.

This *Standard Model*, as Nairne (2002) states, is the foundation for many of the models of short-term/working memory. Focusing specifically on auditory memory, the working-memory system proposed by Baddeley & Hitch (1974) and Baddeley (1992, 2000), for example, contains a *phonological loop* component. This component is built from the *phonological store* and the *articulatory control process*. This control process is responsible for keeping memory representations of speech active in the phonological store past the two seconds they are hypothesized to last; an example of *rehearsal* needing to counteract the *decay* of *activation*. Other formulations exist, which attempt to tie short-term memory into a unitary system with long-term memory (Cowan, 1995) or emphasize the importance of cue-based recall (Nairne, 1990)**.** Beginning with the Logogen model of Morton (1969), specific models of speech perception adopt the concepts of *activation* and *decay* from the memory literature to model lexical access. As Weber & Sharenborg (2012) state, most speech perception models begin with the activation of a set of words competing for lexical access as the acoustic signal progresses. These potential percepts are either strengthened with incoming congruent phonetic information, inhibited with incongruent information, or decay over time until one winner remains and is perceived. Having discussed speech perception models in more depth in Section 1.1.2, we now turn to contrasting the goals of the auditory memory and speech perception literatures in order to specifically motivate why decay patterns are useful in examining representations of speech.

The important difference between the focus of psycholinguistic models of auditory memory and the focus of this thesis is the nature of the mental representations themselves. Whereas memory models take representations as a given and then theorize about how they persist or decay, the goal of this thesis is to determine what information is contained in the representation generated by the initial cognitive process of perception. The phenomenon of *memory chunking* (Miller, 1956) is a good example of this point. It is a well-known fact that grouping representations together into units improves recall; for example, we can memorize more digits overall if they are presented as a sequence of years than if they are presented as single digits in a different order. Understanding the nature of the representation is a necessary precursor to investigating questions of memory. Without this step, we are left with a collection of disparate results that are hard to reconcile into a unified model. For example, attempting to model something like digit recall ability without knowing that people can represent digits as units of years, we would be faced with the confusing fact that sometimes people can remember around 7 digits (the common assumption concerning short-term memory capabilities) and sometimes around 28 (i.e., seven representations of years). Turning to speech, without understanding what information (e.g., voice, speech rate, amplitude, emotional connotation, etc.) is available in the representation which exists after perception, we have no way of unifying all the diverse findings into one sufficient model.

At first glance however, since this chapter is specifically concerned with the time-course of representations, the results from the literature and the studies presented below resemble the discussion of short-term memory models. They both present stimuli and see how long some piece of information persists in memory. The difference is that the studies this thesis is concerned with look at a stimuli's effect on future perception, not memory recall of the stimulus itself. If a certain type of information (e.g., a speaker's speech rate or voice) impacts later perceptual events, we have evidence to conclude that it is present in the representation generated by perception.

To that end, we adopt the terms of *activation* and *decay* from the memory literature but define them in specific ways so as to not imply that we are describing a model of domain-general short-term memory. In this thesis, the term *activation* is used as a short-hand term for the creation of a mental representation of a word upon perception. When the acoustic signal ceases, whatever mental representation of what was perceived we will say has been activated. We also adopt the term *decay* but narrow the definition to only signify the lessening impact of a piece of information on future perception. For example, if a property of a stimulus (e.g., speech rate) affects the processing of later stimuli for only a set amount of time (indexed through either reaction time or accuracy), we will say that the property decays over that time span. Crucially, we intend this term to be agnostic about the (possibly combined) effects of decay and interference discussed in the memory literature. In this chapter, we will occasionally present distance as both time and number of interveners, without making a firm stance on whether the representations we are concerned with decay naturally or require the interference of additional accumulating representations to diminish.

We will not be concerned with the term *rehearsal* in this thesis, as it is unlikely that people mentally rehearse exactly what they heard to perceive future speech.

3.2.2 Perceptual Identification

We now turn to previous results in the literature. A diverse range of hypotheses have been proposed concerning how long token-specific properties impact perception. This is actually quite a contentious question, as the answers found in the literature range from forever (as predicted by strong Exemplar Theory models; cf. Goldinger, 1998) to at least a week (effects of voice similarity in Goldinger, 1996) to a distance of less than 64 intervening words (effects of speech rate in Hanique et al., 2013 and Pisoni, 1993) to a small effect disappearing by around ten intervening items (Orfanidou et al., 2011). In addition, researchers disagree about when the relative contributions of abstract and token-specific properties should emerge. For example, the *time-course hypothesis* of McLennan & Luce (2005) predicts that only tasks which are harder and take longer should show talker-specific effects. Exactly opposite to this view, Otgaar et al. (2012) hypothesize that token-specific information is only available quickly after the presentation of a stimulus with abstract effects taking over with time as episodic information consolidates.

In one of the earliest studies looking at TSEs by manipulating the time-course of word recognition (a test of explicit memory), Craik & Kirsner (1974) found effects of switching voices (between different genders) on recognition accuracy at all lags tested, up to a maximum of 31 intervening items between prime and target. This early conclusion that voice properties persist in memory for at least up to two minutes (the average time distance between prime and target with the maximum number of interveners) contradicted the existing assumptions that token-specific properties decay quickly. However, the fact that this word recognition task examined explicit memory left open the possibility that what was being tested was specifically the participants' episodic memory system; that is, their memory for previous events. Without any linking assumptions that this type of memory matters for speech perception, it remained open how long token-specific properties can influence perception.

Building on this discussion, Jackson & Morton (1984) tested effects of switching voices in a between-subjects blocked word identification task; a test of implicit memory. Primes were presented in the first block with a semantic classification task (i.e., whether the prime was an animate noun) and then targets were presented in the second block with a perceptual identification task. The results indicated that switching the voice (by switching genders) between the prime and target had no effect on perceptual identification accuracy compared to a group with the same voice in both blocks. These results were interpreted in the *Logogen* model of Morton (1969) which hypothesized an abstract representation for words, separate both from modality (visual and auditory) and token-specific properties.

However, citing the existence of early phonological priming, Schacter & Church (1992) and Church & Schacter (1994) proposed that a pre-semantic perceptual representation system mediates between the speech signal and abstract lexical access. They hypothesized that the semantic encoding task of animacy classification was the reason that Jackson & Morton (1984) failed to find results. In multiple between-subjects long-distance repetition priming studies, they manipulated both implicit/explicit tasks and semantic/perceptual encoding of the primes (e.g., animacy vs. enunciation or pitch judgements) to attempt to find TSEs. They presented a list of 24 primes and then a target block of 48 items, half of which had already been presented, resulting in a large variation of the number of intervening words between prime and target. Summarizing and simplifying, they found results of switching voices between prime and target (by switching genders) only with the implicit task of perceptual identification and only when the items were presented without background noise. Manipulating the level of processing did not in fact impact the results of the implicit task, although it did so with the explicit task. They also tested other token-specific properties, namely ‘emotional connotation’ (angry/happy and question/statement intonations) and F0 frequency measures and found evidence that these impacted accuracy in implicit tasks as well. With the variable number of interveners and the long ISIs used in these tasks (normally around 7 seconds), it is difficult to generalize these results to other studies, but we tentatively conclude that TSEs of voice, F0, and intonation information impact accuracy on implicit memory tasks and that these effects occur in implicit tasks regardless of the level of processing involved.

These results were expanded on by Goldinger (1996) who used a similar perceptual identification task to examine the effect of switching voices between study and test blocks. While he also included a manipulation comparing implicit with explicit memory, we focus only on the implicit perceptual identification task here. In a between-subjects design (around 30 participants in each condition), experimental conditions were created using three different delays between study and test blocks (5 minutes, 1 day, and 1 week) and three different numbers of speakers (2, 6, and 10; split between genders). Altogether, 300 words were spoken in each block with the test block consisting of around one half same-voice trials and the rest divided evenly among the other voices. While effects of voice on perceptual identification accuracy did predictably decline over the three delays used, Goldinger reported significant effects of switching voice between prime and target at all delays, even up to 1 week. Using a perceptual similarity matrix calculated from a previous experiment, he also showed that increased voice similarity resulted in increased accuracy all the way up to delays of a week between the study and test block. An additional experiment comparing the effects of the level of processing of the primes found that the deeper the processing (i.e., the more semantic), the greater the priming effects on accuracy. With higher levels of processing however, TSEs were reduced. Altogether, these surprising results were taken as indicating that voice information plays a large role in the memory storage and future perception of speech.

These effects were partly replicated in Sheffert (1998), who further investigated the importance of similarity. Advocating what she termed as the *transfer appropriate processing view*, which is basically a formation of an episodic memory-based lexicon, she hypothesized generally that the more similar a pair of stimuli are, the more accurate perceptual identification responses would be; an indication of facilitatory priming. She conducted two implicit perceptual identification studies to evaluate this, specifically challenging the interpretation found in Schacter & Church (1992) that pre-semantic, perceptual representations intervene between the acoustic signal and the abstract lexeme. In these between-subject studies, within-gender voice switches (rare for the studies in this literature) were tested both with and without background noise and filtering. Only when the prime and target stimuli matched (i.e., both presented with noise) did voice-specific effects on perceptual identification accuracy emerge. This result supports the tentative conclusions from Goldinger (1996) that increased perceptual similarity between voices increased the priming relationships. However, these voice-specific effects were only seen when the stimuli were harder to perceive; when words were presented without background noise or filtering, no voice-specific effects were found. This result, indicative of a potential ceiling effect, runs counter to the results of Goldinger (1996) that increased processing of stimuli decreased voice-specific effects.

To briefly summarize the presented studies thus far, multiple authors, except for Jackson & Morton (1984), have found effects of voice (and other token-specific properties) on perceptual identification accuracy. Presenting repeated words in a different voice as was encountered in the study block reduced identification accuracy. The following two interpretations of these results have been discussed: (1) voice-specific effects stem from activated representations in a pre-semantic (for our purposes, pre-lexical) perceptual system or (2) the incoming speech signal directly activates lexical representations. For the former to adequately account for the results, the representations must be relatively long-lasting, as matching perceptual representations appear to improve identification accuracy at distances greater than a couple of minutes separating prime and target. This account additionally would attribute the priming effects common to both same and different voice pairs as stemming from separate abstract representations, resembling separate hybrid models discussed in Section 1.1.3. The direct-access (a combined representation account, also discussed in Section 1.1.3) predicts that all repetition priming effects are due to raw similarity between representations that are faithful to the incoming speech signal. The more token-specific properties match, the more repetition priming effects are found.

The effects of the level of processing on the occurrence of token-specific effects is less clear. The empirical results appear to be sometimes sensitive to the level of processing of the words in the study block. Sheffert (1998) finds that increasing the demands of the task (by making stimuli harder to perceive) also increases the reliance on "data-driven processing" (i.e., increases the effect of token-specific detail). Goldinger (1996) finds the opposite, such that deeper processing and increased attention comes with an increase in the overall abstract repetition accuracy effect but a decrease in the difference in TSEs found between same and different voice pairs. Schacter & Church (1992) find no effect of level of processing on the voice-effects found in implicit representations. These equivocal results make it difficult to model the effects of token-specific detail, as the interactions between such detail and the level of processing involved in the task may obscure or highlight effects that are not reliable.

3.2.3 Indirect Tasks (Shadowing & Lexical Decision)

We now turn to a different examination of TSEs that attempts to remove at least part of the complications involved in studies relying on perceptual identification accuracy. By looking at results using tasks like shadowing and lexical decision, we can remove the complicating aspect of level of processing differences between prime and target, as the participant encounters both using the same task. Instead of testing accuracy, the following studies test the reaction time distributions of primes and targets to find potential differences in the processing and recognition speed of words.

McLennan et al. (2003) used both shadowing and lexical decision to investigate another type of token-specific information; namely speech rate. They questioned whether words with different speech rates (operationalized as the presence or absence of flaps in intervocalic alveolar stop environments and duration in non-alveolar environments) would prime each other. The rationale for this investigation was to find data disambiguating the predictions of *mediated* models, which predict a gradual abstraction from the speech signal through various sub-lexical, abstract representations (e.g., phonetic, phonemic, and syllabic units) from *direct* models, which predict a direct sound to lexeme mapping. The former is a linguistically-informed model related to the perceptual representation system of Schacter & Church (1992) and relevant to what we term *separate representations* in Section 1.1.3, with the speech recognition process depending on abstract properties of the speech signal. The latter describes models similar to the episodic-based processing systems of Goldinger (2007) and Sheffert (1998) which we have previously discussed, as well as to the *LAFF* (Lexical Access From Features: Stevens, 2002) and *LAFS* (Lexical Access From Spectra: Klatt, 1989) models, which all postulate no intermediate abstract representations. The data revealed differences such that alveolar stimuli primed each other regardless of speech rate whereas non-alveolar stimuli were shadowed faster if they matched previously presented speech rates. To continue the common theme, they also found that the level of processing of the primes impacted whether priming occurred, with deeper processing (either by more word-like non-words or speeded shadowing) eliminating TSEs compared to shallower processing.

Using this reasoning again in McLennan & Luce (2005), they compared the TSEs of switching the gender of the speaker with switching the speech rate. Unlike the results of speech rate in McLennan et al. (2003), increasing the difficulty of the lexical decision and shadowing tasks (by changing the non-words from non-licit to word-like or delaying the shadowing response) *caused* TSEs to arise. In the easier tasks however, both same-voice and different-voice pairs primed each other equally well. As they found different results for speech rate and voice information, they separated what they term *allophonic effects* (phonetic reduction due to speech rate) from *indexical effects* (switching the voice). To explain this, they appealed to the time-course of token-specific information, saying that allophonic effects are found first while indexical effects arise later in speech processing, a hypothesis later work termed the *time-course hypothesis*.

This may seem counter-intuitive, as we would normally consider both speech rate and speaker as detail-rich properties that are available early. However, they position this finding in the framework of *Adaptive Resonance Theory* (ART: Grossberg, 1986 and Grossberg et al., 1997) which postulates that different layers of processing exist beginning from the acoustic input stream. Their *time-course hypothesis* can be restated such that the initial layer of abstraction involves the mapping of different allophonic properties onto an abstract phonemic representation. The next layer is the hypothesized location of TSEs due to voice information. Assuming then that progressing through these layers takes time (and that time is dissociable with priming tasks), we would expect to see allophonic effects before indexical ones. For our purposes, this interpretation is interesting as it uses the decay patterns of different types of token-specific information to investigate the nature of lexical representations. If borne out, this implicates separate representations consisting of allophonic information, necessary for lexical access, and of indexical (e.g., speaker's voice) information, which may influence processing depending on the task. As McLennan & Luce (2005:14) state, "*... we believe that information associated with linguistic and indexical variability may potentially map onto qualitatively distinct types of representations.*" However, the low number of test items (from 12 to 24 items), the variability of the interveners (from immediate to 32 intervening items), and the fact that tokens were not switched when speaker was held constant again makes these experiments difficult to generalize. For this section, the main contribution is the finding that both gender-switch and speech rate have effects on speech processing and that the time-course of these sources of information may differ, indicating separate representations for voice and other token-specific properties.

Finding opposite results to McLennan & Luce (2005), Orfanidou et al. (2011) conclude that abstract effects, and not episodic effects of gender-switch, dominate auditory processing. Their studies used long-distance repetition priming with a relatively long ISI of 1.5 seconds. Crossing stimuli from male and female speakers had no significant effect on reaction time with an intervening lag reported in Orfanidou et al. (2011:101) as occurring "*after approximately 12 intervening items.*" Additionally, increasing the difficulty of the task, as McLennan & Luce (2005) suggested, by embedding all stimuli in noise resulted in no differences in the effects. Long-distance repetition priming with non-words was found however (again not modulated by changing speakers), a fact not straightforwardly predicted by pure abstraction models as no abstract lexical code should exist for non-words. Post-hoc tests examined two additional hypotheses; one concerning individual differences and the other concerning the disambiguation points of non-words, which we will not discuss here. If increased processing time facilitates the appearance of TSEs, as McLennan & Luce (2005) claim, we would predict that slower participants rely more heavily on late-available token-specific information, while faster participants rely more on quicker, abstract representations. The distribution of the data appears to support this claim. To explain these two findings, they conclude that hybrid accounts combining episodic detail with abstract information are the only models able to handle the data. These data provided a replication of the behavioral results of the fMRI study of Orfanidou et al. (2006) with an equivalent design, where neuro-imaging differences were found between prime and target (again around 12 items intervening), but no differences were found between same-voice and different-voice pairs.

The group of studies conducted by Hanique et al. (2013) examined the robustness of TSEs on future speech processing. Their studies tested first speech rate and then the interaction of speech rate with speaker's voice (switching genders). Each of these investigations was conducted in both a short and a long form. The short forms consisted of 288 total trials with 34% of the trials being repeated while the long forms were around 800 trials long with a repetition rate less than 20%. TSEs were only found in the short form experiment testing speech rate alone. These results stand in conflict with the numerous other studies exhibiting TSEs using perceptual identification accuracy, specifically the week-long effects found in Goldinger (1996). To explain these differences, the authors conclude that TSEs only arise if properties of tokens are highlighted through frequent repetitions. Since TSEs arose in one of the experiments, they cite hybrid models combining abstract and token-specific properties as being necessary to model the results. As a final note concerning the time-courses in these experiments, the short form experiments had distances of on average 67 interveners (19 to 100 items) while the long form experiments had on average 405 interveners (79 to 765 items). The authors note that the vast distances at play may have been hiding TSEs but they cited evidence to the contrary, specifically a non-significant control predictor of intervening distance in their models.

These results conflict with those presented above in Section 3.2.2 in that, with the exception of McLennan & Luce (2005), switching the voice of the speaker between prime and target did not impact reaction time distributions on implicit tasks at long distances. There does exist converging evidence however that increasing the difficulty of the task induces slower responses and a greater reliance on episodic properties of speech. Global properties of the task, like the number of trials and the percentage of repetition, should be considered when comparing studies of this type with the studies of perceptual identification in Section 3.2.2, which tend to have much fewer items and a higher percentage of repetition. This may have caused those studies to find abnormally high TSEs.

3.2.4 Repetition Priming Decay

The studies presented above for the most part compare the presence of TSEs between immediate and long-distance conditions. The long-distance conditions are commonly between blocks, or after a variable number of intervening items. This gives us a broad picture of the effects of voice on immediate and long-distance priming, but it does not help to firmly identify how long TSEs are active in perception. The studies in this section look at establishing the decay profiles of repetition priming. Note that these studies do not compare conditions with different token-specific characteristics; they just set up the expected nature of repetition priming in general. They are useful for this thesis though as they set up an expectation of what to look for when comparing conditions with potentially different decay profiles.

Beginning with the work of McKone (1995) and McKone & Dennis (2000), we discuss repetition priming effects using visual, auditory, and cross-modal lexical decision priming paradigms. McKone (1995) advanced a dual-route theory of lexical priming for visual stimuli. A short-term, stronger priming effect decaying after around four intervening items (or 9.3 seconds) is superimposed on a long-term, weaker priming effect. The short-term effect size began as an initial 100ms boost from prime to target. The long-term effect however remained quite strong at around a 50ms boost from prime to target up to when 23 items intervened (and she even found an effect of around 20ms at 1050 interveners, which was around 45 minutes). McKone & Dennis (2000) extended these findings to both auditory and cross-modal presentations. Focusing on the differences between visual and auditory modalities, they tentatively concluded that the same overall, dual-route pattern held in both. In the auditory modality however, the short-term effect persisted longer than it did in the visual modality. They found immediate priming with an approximate 200ms effect size for auditory repetition pairs which decayed to around 40ms by six intervening items. Without including distances longer than six intervening items, they were unable to speak to the exact strength of the auditory long-term priming effect, but they suggested it is comparable to that in the visual modality. These results in the auditory modality match those of Mimura et al. (1997), who compared word and non-word stimuli in an auditory, long-distance repetition priming lexical decision experiment. Focusing on the pattern found with the lexical stimuli, they find a logarithmic decay pattern from immediate presentation to up to eight items intervening in two experiments. Over their long-distance conditions, these priming effects decayed in size from around 170ms to around 90ms; quite comparable to the results of McKone & Dennis (2000), with the caveat that greater priming effects were found at slightly longer lags by Mimura et al. (1997: 90ms) than McKone & Dennis (2000: 40ms).

Explicitly examining the potential dual effects of intervening items and raw time, Berman et al. (2009) set out to answer this question using a paradigm named the *visual recent probes paradigm*. In short, this paradigm tests explicit recognition of probe words from sets of four study words previously visually presented. By varying the ISI of the study and probe words (and holding the intervening task constant), they tested whether memory decay operates over raw time. They consistently found a non-existent effect of raw time decay, as summarized by the disheartening quote, "*What we have in our first five experiments is null results, replicated over and over.*" (Berman et al., 2009: 326). When regressing over all seven of the experiments in this study, they did however find a very small but significant effect of raw time such that each additional second of delay resulted in a 1.8ms decrease in the observed priming effects.

The previous study tested for the presence of decay effects due solely to time in an explicit memory paradigm. The responses participants made were about whether something had explicitly occurred before or not, which may have introduced subtle rehearsal strategies even with the specific use of a task designed to eliminate them. McKone (1998) attempted to find the same types of effects but using an implicit memory paradigm, namely lexical decision. Using a four second ISI, she crossed four numbers of interveners with five raw time distances. Interference due to intervening items yielded the strongest effect, but weaker effects did exist for the passage of time in the absence of any intervening items. These combined effects are described using terms from the *Standard Model*, described above in Section 3.2.1, such that the initial activation of a lexeme decays over time and is partially overwritten with each interfering item. These results stand in conflict with those of Berman et al. (2009), but the differences in the types of tasks prevent direct comparisons.

Another study relevant to the investigation of interveners or time as the impetus for activation decay is found in Lee & Zhang (2017). They designed a study to test the *time-course hypothesis* of McLennan & Luce (2005) which crossed ISIs of 50ms and 250ms in a paired auditory lexical decision task. If varying speakers only impacts perception after time elapses, as speaker representations are hypothesized to be established slower, the prediction is that greater TSEs would be seen in the longer ISI than in the shorter. Additionally, they compared these results to an auditory semantic priming study; examining whether switching talkers between prime and target impacts semantic processing (hypothesized to be a late occurrence in the overall lexical access process). While they found no effects of switching voices (within gender) in the semantic priming condition (and therefore concluded that voice properties are established before semantic properties), they do find TSEs with repetition priming. These effects pattern as expected, with different-voice pairs having a mean priming effect of 138 and 117ms at the 50 and 250ms ISI respectively, while the same-voice pairs' priming effects stand at 211 and 132ms respectively. This significant difference, both between speakers and between ISIs, shows that raw time does have an effect on token-specific properties.

These studies of repetition priming lead us to expect a logarithmically decaying function of priming facilitation on reaction time measures. While it is possible that this pattern could reflect a single source of decay, the fact that early and late priming appear to be quite different lends support for a theory breaking repetition priming into two parts, a short-term and a long-term component. The short-term component is expected to be quite strong (even up to a 200ms priming effect) but is only expected to persist until distances of four to eight intervening items. After the short-term component disappears, long-term priming effects (with a strength of 20 to 40ms) are the sole remaining source of priming. They are relatively stable for quite some time (shockingly even up to 1050 interveners). Looking at the cause of this decay, evidence suggests that the intervening items causes the most amount of interference, due to either an over-writing effect of additional mental representations or the act of perception itself. Some tentative evidence also exists that priming effects decay with only the passage of time. In examining the influence of token-specific information, it is therefore useful to determine whether short- or long-term priming components (or both) are affected. Some studies emphasizing the perceptual nature of token-specific information (e.g., Schacter & Church, 1992) would predict an influence of token-specific information only on the short-term priming component, while the theories of Goldinger (1996) and

Sheffert (1998) would predict that token-specific information affects both short- and long-term components of priming.

3.2.5 Summary

Summarizing so far, the question of the time-course of episodic information, even just of voice-specific information, remains contentious. There is suggestive evidence showing that episodic effects can last quite a long time, whereas other evidence exists showing that the same effects decay rapidly. The level of processing in implicit tasks appears to matter (depending on which studies one is looking at), with increased processing potentially increasing or decreasing the presence of TSEs, depending on whether the task taps into implicit or explicit memory. The experiments presented in this chapter are designed to shed some light on these questions by establishing a concrete time-course of the effect of switching talkers on implicit, lexical decision tasks. Additionally, we examine the impact of the number of interveners and the raw time between prime and target to separate out these different loci of priming effects. This investigation proves crucial to a further examination of the nature of speech processing, and, to our knowledge, hasn't been explicitly established to date.

3.3 Experiment 2

Building on Experiment 1, which revealed a significant reduction in repetition priming when speakers were switched between an immediate presentation of prime and target, this study tests the same stimuli both immediately and at a distance of ten intervening words. This distance was chosen to resemble the similar designs found in Orfanidou et al. (2006, 2011). However while they averaged over a number of distances, this study compares only two discrete distances. This is also a two-speaker (both male) repetition priming study as, following the studies of Sheffert (1998) and Goldinger (1996) who find both between- and within-gender effects, we aimed to look specifically at effects of different voices without the added confound of gender-specific effects arising (cf. Geiselman & Bellezza, 1976; 1977). We additionally manipulated lexicality to investigate the possibility of an interaction between episodic properties of language and lexicality. The lexicality manipulation will not be discussed in this chapter, as we focus on the effects seen with words alone. As the difficulty of the task may prove important, we note that the goal of the experiment was to create a difficult lexical decision task through the use of phonotactically licit non-words and a relatively short ISI of between 400 and 600ms.

We expect a replication of the results of Experiment 1 (i.e., significantly slower repetition priming when speakers are switched) upon immediate presentation of the target. With ten intervening words however, given the *time-course hypothesis* of McLennan & Luce (2005), we would expect a greater influence of voice switch at a distance, since these effects are reported to emerge later in difficult implicit tasks. However, Orfanidou et al. (2006, 2011) found no significant differences between same- and different-voice conditions at a comparable average distance. If our data also show no voice effects in the long distance condition, we can draw tentative conclusions that TSEs only have an impact on the short-term component of repetition priming. Lastly, given the discussion of the repetition priming decay profile, we also predict to find small but significant, additive effects of raw time distance on top of the number of intervening items between prime and target. Having found that our stimuli and task shows TSEs at an immediate distance between prime and target, this experiment adds more detail about how long such effects persist in influencing auditory perception.

3.3.1 Method

3.3.1.1 Participants

A total of 50 (age range not recorded but presumably 18-22; 15 male) undergraduate participants were run in the fall semester of 2015. They were recruited from the experimental subject pool at the University of Pennsylvania and voluntarily completed the study in person using a custom PsychoPy (Pierce, 2007) implementation of a continuous lexical decision task at the Language Variation and Cognition Lab (described in Section 2.2.4).

3.3.1.2 Stimuli

The same 600 stimuli from Experiment 1 were used in this experiment, recorded by speakers MA1 and MA2 (see APPENDIX I: Experimental stimuli for more details). For the counterbalancing purposes outlined in the next section, 96 items were repeated in the word and non-word conditions (compared to 100 repeated words in Experiment 1). Four words (*dish*, *throat*, *land*, and *game*) and four non-words (*nayd*, *bays*, *jaelk*, and *vowz*) were taken from the repetition groups and added to the non-repetition groups.

3.3.1.3 Design

The 300 words and non-words were divided into two sets, the repetition (96 items) and non-repetition sets (204 items). Within the word and non-word repetition sets, eight lists were created by varying the factors VOICE$_{\text{[Same vs. Different Voice]}}$ × DISTANCE$_{\text{[Immediate vs. Long-distance]}}$ × SPEAKER$_{\text{[MA1-prime vs. MA2-prime]}}$. Therefore, twelve words and non-words existed in each possible grouping of these three conditions. Immediate, long-distance, and non-repetition (filler) trials were randomly interleaved together such that a different random order existed for each of the eight lists. Due to the distance manipulation, randomization had to occur at the list level and not at the participant level; however, the eight different trial orders should obviate any potential confounds to due trial order. The experiment was presented in a total of four experimental blocks with no repetitions occurring between blocks. The number of word and non-word stimuli, and additionally the number of repeated vs. non-repeated items from each, were equal across all blocks, resulting in a repetition rate of 32% across the experiment, similar to the short-form experiments of Hanique et al. (2013).

3.3.1.4 Procedure

Participants completed ten practice lexical decisions as practice before the experiment. These consisted of the following five sets chosen from the non-repetition filler stimuli sets, randomly presented per subject (with subscript representing different speakers and bold face indicating non-words): *mark*$_1$ - *lamp*$_2$, *guard*$_2$ - ***geyk***$_2$, ***trowz***$_2$ - *rag*$_1$, ***vaebd***$_1$ - ***kuhg***$_1$, *school*$_1$ - ***wahng***$_2$. Following the practice session, four blocks of 198 items each were presented (792 total lexical decisions) using the custom PsychoPy continuous lexical decision task implementation described in Section 2.2.4. All experiments included in the analysis were completed on average within 22 minutes (range: 18-27 minutes). As mentioned, each of the eight lists had their own specific trial order due to the distance manipulation; stimuli within lists were not randomized when presented to participants.

3.3.1.5 Analysis

Of the original 50 participants in the experiment, four participants were removed due to overall poor performance. From the four, one was removed due to an overall accuracy score of less than 70% correct, one from overall slow responses as indicated by the Hodges-Lehmann estimate of the reaction time distribution, and two from having multiple (i.e., greater than twenty) absurdly fast responses. After this global participant removal, the overall results per each item were examined. The non-words *daask*, *neyn*, and *theyz* from the filler list in addition to *faht* and *traak* from the experimental list had accuracy scores of less than 50% correct. These items were removed and will not be reported further in the analyses. The table below describes the amount of data trimmed to create the dataset for the models of reaction time.

Table 9: Experiment 2 removal summary

	Observations	**Percentage**
Inaccurate trials	*309*	*7.0*
RT trimming (300 > RT < 3000)	*70*	*1.6*
By-participant trimming	*222*	*5.0*
By-item trimming	*106*	*2.4*
Total removed	*707*	*16.0*
Total remaining	*3709*	

3.3.2 Results

The following table reports the distribution of the data per the factors VOICE, DISTANCE, and SPEAKER. These data are reported after only the minimal global trimming procedure of unreasonable reaction times was applied. For that reason, the reaction time reports are central tendency measures from the non-parametric Hodges-Lehmann estimates. The accuracy scores given indicate the amount of correct responses out of the total, after all global participant and item removal was conducted.

Table 10: Experiment 2 data summary

DISTANCE			**VOICE**			**SPEAKER**			**Total**
0 Interveners	*702*	*98.1*	**Same**	*677*	*98.2*	**MA1**	*680*	*97.4*	***547***
						MA2	*674*	*98.9*	***544***
			Diff.	*726*	*98.1*	**MA1**	*724*	*98.4*	***547***
						MA2	*730*	*97.8*	***544***
10 Interveners	*847*	*95.9*	**Same**	*847*	*95.6*	**MA1**	*855*	*95.3*	***534***
						MA2	*840*	*95.9*	***541***
			Diff.	*847*	*96.2*	**MA1**	*852*	*94.4*	***535***
						MA2	*843*	*97.9*	***535***
Fillers	*919*	*93.1*				**MA1**	*916*	*92.3*	***4565***
						MA2	*922*	*94.0*	***4596***
Primes	*896*	*95.1*				**MA1**	*896*	*95.1*	***2144***
						MA2	*895*	*95.1*	***2128***

3.3.2.1 *Accuracy*

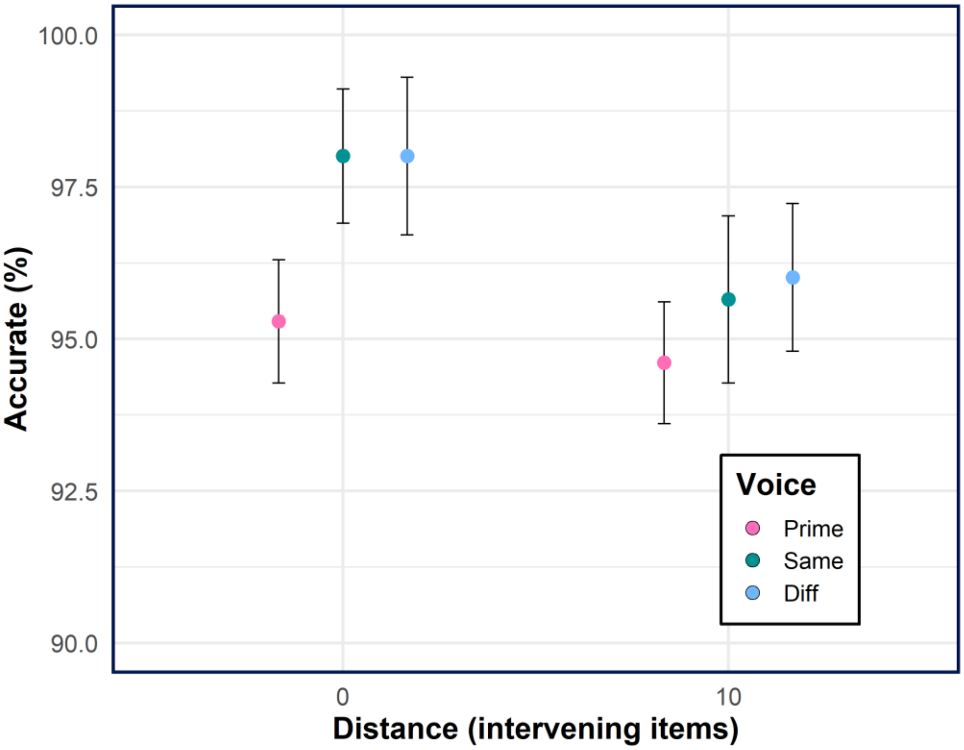

Figure 13: Experiment 2 accuracy data

A generalized linear mixed-effects model was fit to the accuracy data. This model combines primes and fillers together as the reference level for the Condition factor, with targets in each combination of VOICE, DISTANCE, and SPEAKER dummy-coded. Random effects were set as intercepts for participants and items. The outcome of this model is seen in Table 11.

This model shows whether any priming was seen when items were repeated. Immediately repeated targets were identified significantly more accurately (same voice: $p = 0.001$, different voice: $p < 0.001$) whereas non-significant effects were found in the long-distance conditions (same voice: $p = 0.157$, different voice: $p = 0.815$). There was a significant interaction effect indicating that for speaker MA2, the long-distance different-voice condition was significantly more accurate, however we will not remark on this further.

Table 11: Experiment 2 combined accuracy model

	Accuracy (combined)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	27.31	21.10 – 35.35	**<0.001**
VOICE × DISTANCE			
Diff. at 0	4.03	2.00 – 8.12	**<0.001**
Diff. at 10	0.95	0.62 – 1.45	0.815
Same at 0	3.03	1.60 – 5.73	**0.001**
Same at 10	1.45	0.87 – 2.43	0.157
SPEAKER			
MA2	1.24	0.98 – 1.57	0.079
Item (z-scored)			
Frequency	1.12	1.01 – 1.25	**0.037**
PNH	0.86	0.78 – 0.96	**0.007**
AoA	0.76	0.68 – 0.85	**<0.001**
Duration	1.01	0.90 – 1.13	0.860
Trial	0.77	0.71 – 0.83	**<0.001**
Participant			
Male	1.03	0.68 – 1.55	0.896
Group			
2	0.78	0.40 – 1.54	0.478
3	0.89	0.46 – 1.74	0.738
4	1.93	0.93 – 4.00	0.078
5	1.38	0.67 – 2.83	0.376
6	1.45	0.73 – 2.87	0.286
7	1.34	0.67 – 2.67	0.414
8	0.73	0.37 – 1.43	0.360
VOICE × DIST. (Diff.-0) × SPKR (MA2)	0.65	0.25 – 1.64	0.360
VOICE × DIST. (Diff.-10) × SPKR (MA2)	2.70	1.25 – 5.80	**0.011**
VOICE × DIST. (Same-0) × SPKR (MA2)	1.76	0.60 – 5.14	0.305
VOICE × DIST. (Same-10) × SPKR (MA2)	0.88	0.42 – 1.83	0.731
Observations	17760		
Marginal R^2 / Conditional R^2	0.107 / 0.355		

Turning now to the model examining the interaction of the experimental predictors of VOICE and DISTANCE, the table below shows the results of the data considering only targets.

VOICE, DISTANCE, and SPEAKER were dummy-coded, with the reference levels set to same voice pairs at the immediate distance for speaker MA1.

Table 12: Experiment 2 target accuracy model

	Accuracy (targets)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	117.38	51.28 – 268.67	**<0.001**
VOICE			
Diff.	1.30	0.48 – 3.56	0.605
DISTANCE			
10 interveners	0.46	0.22 – 0.96	**0.039**
SPEAKER			
MA2	2.74	0.89 – 8.49	0.080
Item (z-scored)			
Frequency	0.80	0.61 – 1.04	0.089
PNH	0.88	0.68 – 1.13	0.312
AoA	0.94	0.72 – 1.23	0.665
Duration	1.41	1.06 – 1.86	**0.017**
Trial	0.67	0.54 – 0.84	**<0.001**
Participant			
Male	1.14	0.61 – 2.15	0.676
Group			
2	0.80	0.28 – 2.32	0.682
3	0.49	0.16 – 1.48	0.205
4	0.73	0.22 – 2.39	0.606
5	0.79	0.24 – 2.64	0.706
6	1.47	0.43 – 5.05	0.544
7	0.96	0.30 – 3.09	0.940
8	0.59	0.19 – 1.84	0.367
VOICE (Diff.) × DIST. (10)	0.57	0.19 – 1.68	0.309
VOICE (Diff.) × SPEAKER (MA2)	0.35	0.08 – 1.59	0.173
DIST. (10) × SPEAKER (MA2)	0.49	0.15 – 1.62	0.243
V. (Diff.) × D. (10) × S. (MA2)	8.21	1.53 – 44.10	**0.014**
Observations	4327		
Marginal R^2 / Conditional R^2	0.137 / 0.459		

Here we see the additional information that while the DISTANCE manipulation yielded significant inhibitory effects on accuracy (p = 0.039), the VOICE manipulation did not significantly affect accuracy. The main effect (p = 0.605) indicates a non-significant effect at the immediate distance and the non-significant interaction terms (p = 0.309, 0.173) indicates no significant difference from accuracy in same-voice pairs at a long-distance.

Altogether, the accuracy results pattern as we would expect in a long-distance, repetition priming lexical decision task. With the conclusion that voice switches did not impact accuracy data, we now turn to effects on reaction time.

3.3.2.2 Reaction Time

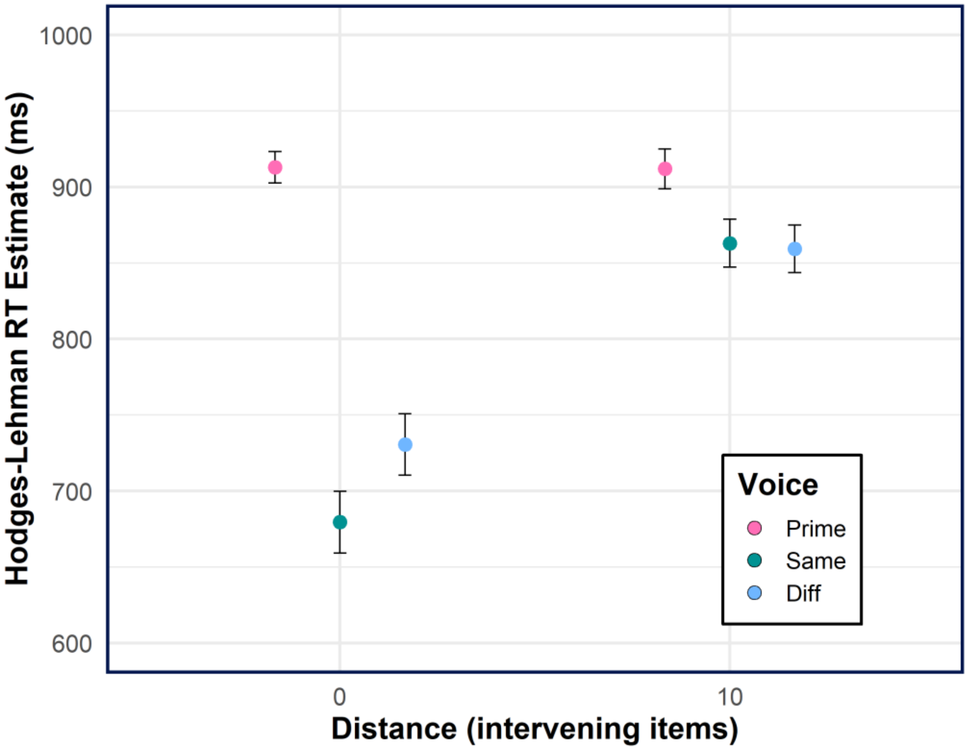

Figure 14: Experiment 2 reaction time data

The first analysis of reaction time comes from the large model investigating whether repetition priming existed in all experimental conditions. The responses to primes and targets were combined into one dataset and a linear mixed-effect model was run examining the $\log_2$-transformed RTs. Random effects included by-subject and by-item intercepts along with random slopes for the factors of VOICE$_{\text{[Same vs. Different Voice]}}$ and DISTANCE$_{\text{[Immediate vs. Long-distance]}}$. Fixed effects of interest were a dummy-coded variable with the baseline being responses to the primes and factor levels indicating each of the four VOICE by DISTANCE conditions, which interacted with a dummy-coded variable indicating the SPEAKER of the sound-file. Model criticism resulted in 180 additional observations being removed, a total of 2.43% of the remaining data.

For the comparisons of interest, all factor levels came out significant (all $p < 0.001$; same-voice at 0 = 24% / 127ms; diff-voice at 0 = 19% / 101ms; same-voice at 10 = 7% / 40ms; diff-voice at 10 = 8% / 43ms), indicating that in each condition formed by the interaction of the factors VOICE and DISTANCE, significant priming was found. These comparisons remained highly significant after doing a multiple comparisons correction using Holm's method. One additional interaction term came out significant: in the same-voice prime/target condition at the immediate distance, speaker MA2 trials were recognized significantly slower than speaker MA1 ($p = 0.039$), but this significant term will not be interpreted further here. Overall, this model, indicates that significant abstract repetition effects were found.

Table 13: Experiment 2 combined RT model

Log$_2$-transformed RT (combined)			
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.77	9.74 – 9.80	**<0.001**
VOICE × DISTANCE			
Diff. at 0	-0.30	-0.33 – -0.27	**<0.001**
Diff. at 10	-0.12	-0.15 – -0.09	**<0.001**
Same at 0	-0.39	-0.43 – -0.35	**<0.001**
Same at 10	-0.11	-0.15 – -0.08	**<0.001**
SPEAKER			
MA2	0.01	-0.01 – 0.02	0.293
Item (z-scored)			
Frequency	-0.01	-0.02 – 0.00	0.110
PNH	0.02	0.01 – 0.03	**0.001**
AoA	0.01	-0.00 – 0.02	0.214
Duration	0.09	0.08 – 0.10	**<0.001**
Trial	-0.03	-0.03 – -0.02	**<0.001**
Previous RTLog	0.06	0.05 – 0.06	**<0.001**
Participant			
Male	-0.01	-0.06 – 0.05	0.804
Group			
2	0.04	-0.05 – 0.13	0.378
3	-0.02	-0.11 – 0.07	0.650
4	-0.01	-0.11 – 0.08	0.803
5	-0.03	-0.12 – 0.07	0.574
6	-0.02	-0.11 – 0.07	0.675
7	0.04	-0.05 – 0.13	0.397
8	-0.02	-0.11 – 0.07	0.599
VOICE × DIST. (Diff.-0) × SPKR (MA2)	-0.01	-0.04 – 0.03	0.765
VOICE × DIST. (Diff.-10) × SPKR (MA2)	0.00	-0.03 – 0.04	0.841
VOICE × DIST. (Same-0) × SPKR (MA2)	0.04	0.00 – 0.09	**0.039**
VOICE × DIST. (Same-10) × SPKR (MA2)	0.02	-0.02 – 0.06	0.348
Random Effects			
σ^2	0.06		
τ_{00} Item	0.01		
τ_{00} Participant	0.01		
τ_{11} Participant × VOICE (Diff.)	0.01		
τ_{11} Participant × VOICE (Same)	0.00		
τ_{11} Participant × DISTANCE (10)	0.00		
ρ_{01} Participant × VOICE (Diff.)	-0.11		
ρ_{01} Participant × VOICE (Same)	0.03		
ρ_{01} Participant × DISTANCE (10)	-0.18		
Observations	7238		
Marginal R^2 / Conditional R^2	0.334 / 0.474		

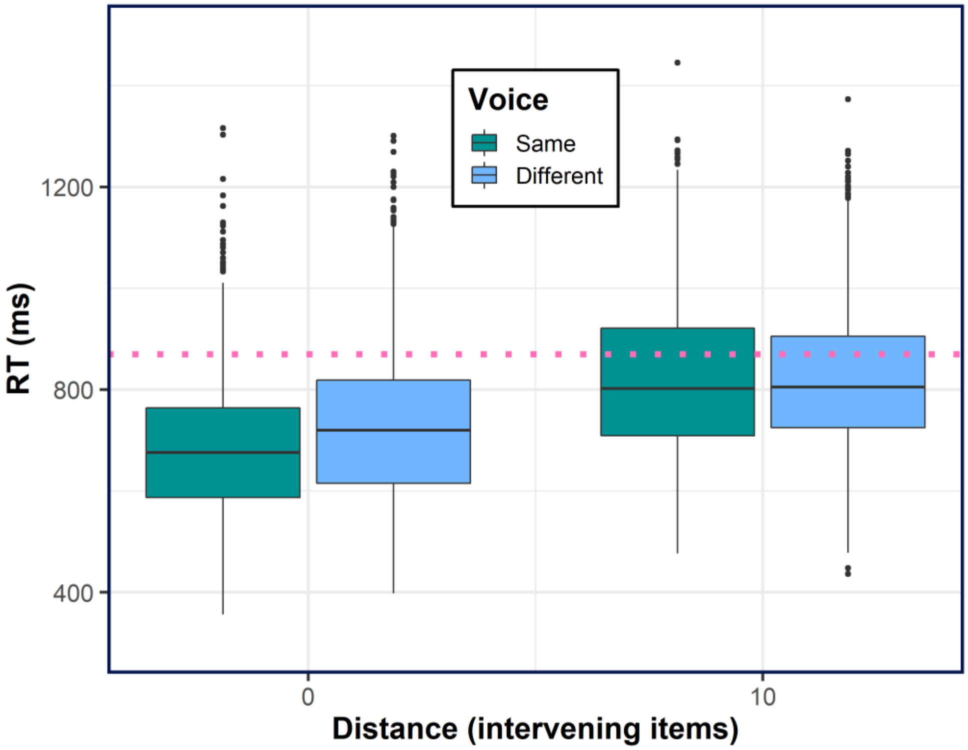

Figure 15: Experiment 2 trimmed reaction time data

Next, we examine the model testing the differences between same- and different-voice conditions at immediate and long distances. Figure 15 plots the distribution of each condition in this model, with the median prime reaction time indicated with the dotted line. The random effects for this model consisted of by-subject and by-item intercepts along with random slopes for the factor of DISTANCE. The fixed effects of interest were formed by the interactions between three terms: the dummy-coded factors of VOICE (baseline = same-voice), DISTANCE (baseline = immediate repetition), and SPEAKER (baseline = speaker MA1). After fitting the model, 60 additional values (1.62% of the remaining data) with residuals > 2.5 SDs from the mean were removed.

Table 14: Experiment 2 target RT model

	Log$_2$-transformed RT (targets)		
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.36	9.31 – 9.41	**<0.001**
VOICE			
Diff.	0.09	0.05 – 0.12	**<0.001**
DISTANCE			
10 interveners	0.29	0.25 – 0.33	**<0.001**
SPEAKER			
MA2	0.07	0.03 – 0.10	**<0.001**
Item (z-scored)			
Frequency	-0.01	-0.02 – 0.00	0.215
PNH	0.02	0.01 – 0.03	**0.002**
AoA	0.00	-0.01 – 0.01	0.621
Duration	0.09	0.08 – 0.11	**<0.001**
Trial	-0.04	-0.05 – -0.03	**<0.001**
Previous RTLog	0.05	0.04 – 0.06	**<0.001**
Participant			
Male	-0.00	-0.06 – 0.05	0.884
Group			
2	0.05	-0.04 – 0.15	0.301
3	-0.01	-0.11 – 0.08	0.775
4	0.03	-0.07 – 0.13	0.575
5	-0.02	-0.12 – 0.08	0.728
6	0.02	-0.07 – 0.12	0.637
7	0.05	-0.04 – 0.15	0.282
8	-0.01	-0.10 – 0.09	0.877
VOICE (Diff.) × DIST. (10)	-0.11	-0.15 – -0.07	**<0.001**
VOICE (Diff.) × SPEAKER (MA2)	-0.04	-0.09 – 0.01	0.154
DIST. (10) × SPEAKER (MA2)	-0.04	-0.08 – -0.01	**0.023**
V. (Diff.) × D. (10) × S. (MA2)	0.05	0.00 – 0.11	**0.047**
Random Effects			
σ^2	0.04		
τ_{00} Item	0.01		
τ_{00} Participant	0.02		
τ_{11} Participant × DISTANCE (10)	0.01		
ρ_{01} Participant	-0.82		
Observations	3649		
Marginal R^2 / Conditional R^2	0.317 / 0.527		

The main effects for each of the factors of interest came out significant, indicating reaction time slow-downs from same- to different voice pairs ($p < 0.001$, 6.3% / -34ms), from immediate to long-distance ($p < 0.001$, 22.7% / -120ms), and from speaker MA1 to speaker MA2 (p < 0.001). Multiple significant interaction terms implicated different effects with the combination of the three interacting factors of interest.

We now turn to two additional models separated by the DISTANCE condition to interpret these interaction effects. These models were the same as the experimental model, except for the removal of the DISTANCE factor from the variables of experimental interest and the fact that the random effect structure for each ended up including only by-subject and by-item intercepts. The first model of priming effects at the immediate distance (with 38 responses removed totaling 1.96% of the data) indicated a significant slow-down in the different-voice compared to the same-voice condition (p < 0.001, 7.0% / -37ms). Additionally, stimuli from speaker MA2 were recognized slower than those from speaker MA1 (p = 0.001), although the interaction between the two did not come out significant. The second model looking at the long-distance condition (21 responses removed totaling 1.19% of the data), indicated no significant effects of interest either between the same- and different-voice conditions (p = 0.15, 1.8% / -10ms) or between speakers (p = 0.13) or in the interaction between the two (p = 0.54).

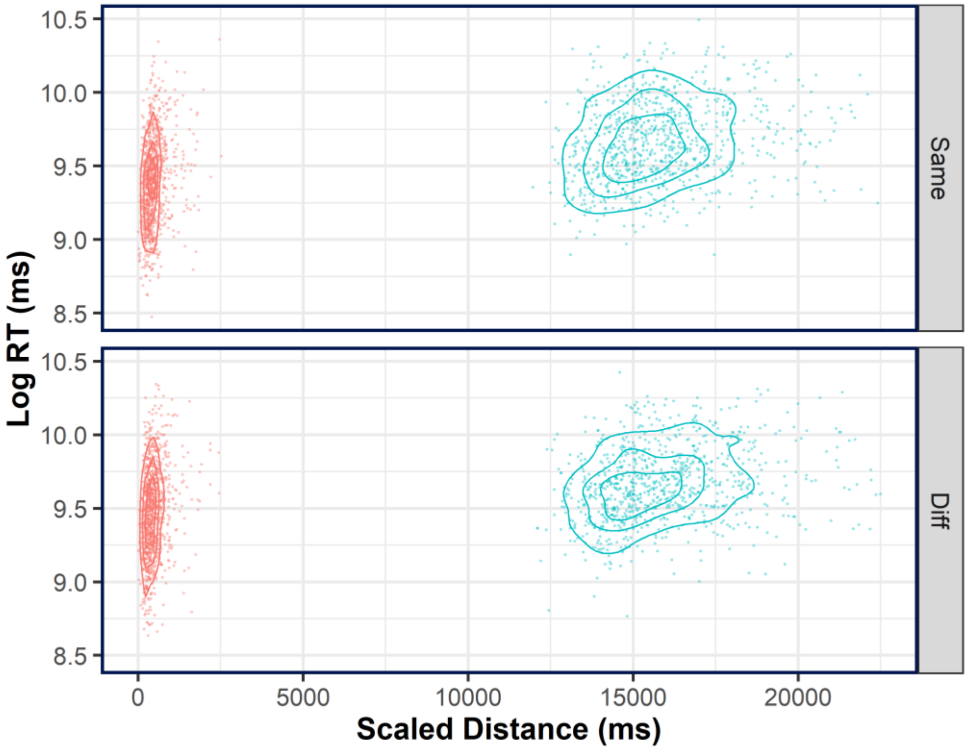

Figure 16: Experiment 2 effects of time vs. interveners

Using these two models separated by distance, we additionally tested the impact that the passage of raw time has over the interference due to intervening items. To do so, we used the chi-squared test to compare models with and without an additional factor of z-scored, $\log_2$-transformed time between the end of the prime and the beginning of the target. At immediate distances, the raw time estimate significantly improved the model ($p = 0.035$), and indicated that with increased time between prime and target, reaction time was faster; potentially indicating better processing with more time. In the long-distance condition, the raw time estimate also significantly improved the model ($p < 0.001$) but indicated that increased raw time slowed reaction times. The different pattern in the predicted values of the raw time variable may possibly indicate different stages of processing at immediate and long-distance conditions. The data is plotted in Figure 16, with color indicating the number of intervening items.

3.3.3 Discussion

The analysis of Experiment 2 indicates the following:

- **Abstract repetition priming**
    - *Accuracy:*
        - All immediately repeated targets recognized more accurately than primes
        - At long-distance, no targets recognized more accurately than primes
    - *Reaction time:*
        - All repeated targets in both distances recognized faster than primes
- **Talker-specific effects**
    - *Accuracy:*
        - Increased distance reduced accuracy
        - No difference between same- and different-voice targets
    - *Reaction time:*
        - Different-voice targets at immediate distance recognized slower than same-voice targets
        - No difference between same- and different-voice targets at the long-distance condition
- Increasing raw time between prime and target speeds reaction time in the immediate condition and slows reaction time in the long-distance condition

Altogether, these models indicate that while TSEs existed with the immediately subsequent presentation of prime and target, these effects disappeared by the time ten words intervened. Abstract repetition was still found in both same- and different-voice conditions and both distances however. These results both replicate the results from

Experiment 1 in the immediate condition and echo the results found in Orfanidou et al. (2006, 2011). Contradictory effects of raw time over and above that of intervening items was also found, which we should keep in mind after examining the results from the other experiments.

3.4 Experiment 3

Having found no talker-specific effects between two male voices at a long distance in Experiment 2, this study asks whether the more salient effect of switching voices between male and female will show long-distance TSEs. This is a relevant question as many of the previous results described as talker effects actually came from speakers of different genders. Differences in vocal tract length between men and women, along with other perceptual differences, clearly result in easily distinguishable voices. As discussed in the background, some studies do indicate that an increase in similarity also increases TSEs. In fact, this is a necessary component of combined representational models built on detail-rich speech representations. The findings of Goldinger (1996) are especially relevant in that between-gender voice switches accounted for most of the observed voice effects. Within-gender switches still often resulted in significant effects, interestingly correlated with the perceptual similarity between voices. The experiment presented here also more closely resembles the studies of Orfanidou et al. (2006, 2011) than Experiment 2 in that the voice manipulation is achieved by switching the gender of the speaker.

To that end, this study compares the effect of switching the gender of the speaker both immediately and at a distance of ten intervening items. The same-gender condition is the same as the different-voice condition in Experiment 2 while the same-voice condition is replaced by a different-gender condition. We expect a further replication of the results from Experiment 1 and Experiment 2 at the immediate distance. Since increased perceptual similarity is said to lead to greater priming effects, the decreased perceptual similarity due to switching genders on top of switching voices should lead to reduced priming effects at both distances. Again, lexicality was also experimentally manipulated, but we will not discuss this additional manipulation here. The post-hoc test investigating the effect of time vs. interveners is also conducted, but we again expect similar patterns to Experiment 2.

If our data show no voice effects in the long-distance condition for either voice switches, we again must conclude that talker-specific information does not influence the long-term source of repetition priming. This would give further support to the idea that separate representations of abstract and token-specific properties exist upon perception. Also, if voice effects are reduced when the gender of the speaker is switched in the immediate distance, we would conclude along with Goldinger (1996), that the more perceptually similar items are, the more priming occurs.

3.4.1 Method

3.4.1.1 Participants

A total of 33 (age range 18-21, mean 19.3; 10 male) undergraduate participants were run at the end of the fall semester of 2015 and the beginning of the spring semester in 2016. They were recruited from the experimental subject pool at the University of Pennsylvania and voluntarily completed the study in person using a custom PsychoPy implementation of a continuous lexical decision task at the Language Variation and Cognition Lab.

3.4.1.2 Stimuli

The set of 600 stimuli from Experiment 1 and Experiment 2 were used in this experiment, recorded by speakers MA1, MA2, and FM1 (see APPENDIX I: Experimental stimuli for details). Unlike Experiment 2 however, 100 items each were repeated in the word and non-word conditions, while the rest constituted the non-repeated filler conditions. Items from speaker MA1 were included in each pair as either the prime or the target. Speakers MA2 and FM1 provided the other member of each pair. This resulted in more experimental items produced by MA1 than the other two speakers. The non-repeating filler items asymmetrically were taken from speakers MA2 and FM1 to roughly balance the number of items produced by each speaker throughout the experiment (83 each from MA2 and FM1 and 34 from MA1).

3.4.1.3 Design

The 300 words and non-words were divided into two sets, the repetition (100 items) and non-repetition sets (200 items). Within the word and non-word repetition sets, four lists were created by varying the factors GENDER$_{\text{[Male-Male vs. Male-Female]}}$ × DISTANCE$_{\text{[Immediate vs. Long-distance]}}$. Unlike Experiment 2, the factor indicating speaker, termed DIRECTION$_{\text{[MA1-prime vs. MA1-target]}}$ here, was balanced between-subjects, not within-subjects. This resulted in a total of eight lists, but for analysis purposes, the factor of DIRECTION will be treated as a fixed effect, collapsing the number of lists back down to four. Overall, twenty-five words and non-words existed in each possible grouping of the two conditions GENDER and DISTANCE. A custom Python script was written to randomly interleave immediate distance, long-distance, and non-repetition (filler) trials together such that a different random order existed for each of the original eight lists. Again, randomization had to occur at the list level and not the participant level due to the distance manipulation. The experiment was presented in a total of four experimental blocks with no repetitions occurring between blocks. The number of word and non-word stimuli, and additionally the number of repeated vs. non-repeated items from each, were equal across all blocks, resulting in a repetition rate of 33% across the experiment, similar to the short-form experiments of Hanique et al. (2013).

3.4.1.4 Procedure

Participants completed ten practice lexical decisions as practice before the experiment. These consisted of the following five sets chosen from the non-repetition filler stimuli sets, randomly presented per subject (with subscript representing different speakers and bold face indicating non-words): *mark*$_1$ - *lamp*$_2$, *guard*$_2$ - ***geyk***$_2$, ***trowz***$_3$ - *rag*$_3$, ***vaebd***$_1$ - ***kuhg***$_1$, *school*$_3$ - ***wahng***$_2$. Following the practice session, four blocks of 200 items each were presented (800 total lexical decisions) using the custom PsychoPy continuous lexical decision task implementation described in Section 2.2.4. All experiments included in the analysis were completed on average within 26 minutes (range: 18-33 minutes). Each of the original eight lists had their own specific trial order due to the distance manipulation; stimuli within lists were not randomized when presented to participants.

3.4.1.5 Analysis

Of the original 33 participants in the experiment, four participants were removed due to overall poor performance. From the four, two were removed due to an overall accuracy score of less than 70% correct and the other two from overall slow responses as indicated by the Hodges-Lehmann estimate of the reaction time distribution. After this global participant removal, the overall results per each item were examined. The non-words *daask* from the filler list and *traak* from the experimental list had accuracy scores of less than 50% correct. These items were removed and will not be reported further in the analyses. The table below describes the amount of data trimmed using the common procedures.

Table 15: Experiment 3 removal summary

	Observations	**Percentage**
Inaccurate trials	*223*	*7.7*
RT trimming (300 > RT < 3000)	*32*	*1.1*
By-participant trimming	*63*	*2.3*
By-item trimming	*84*	*2.9*
Total removed	*406*	*14.0*
Total remaining	*2494*	

3.4.2 Results

Table 16 reports the distribution of the data per the factors GENDER, DISTANCE, and SPEAKER (an easier visualization than the DIRECTION factor). These data are reported after only the minimal global trimming procedure of unreasonable reaction times was applied. Reaction time reports are central tendency measures from the non-parametric Hodges-Lehmann estimates and accuracy scores indicate the amount of correct responses out of the total, after all global participant and item removal was conducted.

Table 16: Experiment 3 data summary

Distance			**Gender**			**Speaker**			**Total**
0 Interveners	*764*	*98.5*	**Same**	*768*	*98.5*	**MA1**	*782*	*98.2*	***397***
						MA2	*751*	*98.8*	***323***
			Diff.	*760*	*98.5*	**MA1**	*762*	*98.2*	***400***
						FM1	*757*	*98.8*	***320***
10 Interveners	*870*	*95.9*	**Same**	*876*	*94.9*	**MA1**	*872*	*95.7*	***391***
						MA2	*880*	*94.1*	***320***
			Diff.	*865*	*96.8*	**MA1**	*868*	*95.2*	***397***
						FM1	*860*	*98.7*	***314***
Fillers	*932*	*93.4*				**MA1**	*948*	*91.7*	***971***
						MA2	*964*	*92.2*	***2357***
						FM1	*894*	*95.2*	***2357***
Primes	*912*	*94.4*				**MA1**	*930*	*92.6*	***1269***
						MA2	*921*	*96.1*	***787***
						FM1	*874*	*95.7*	***787***

3.4.2.1 *Accuracy*

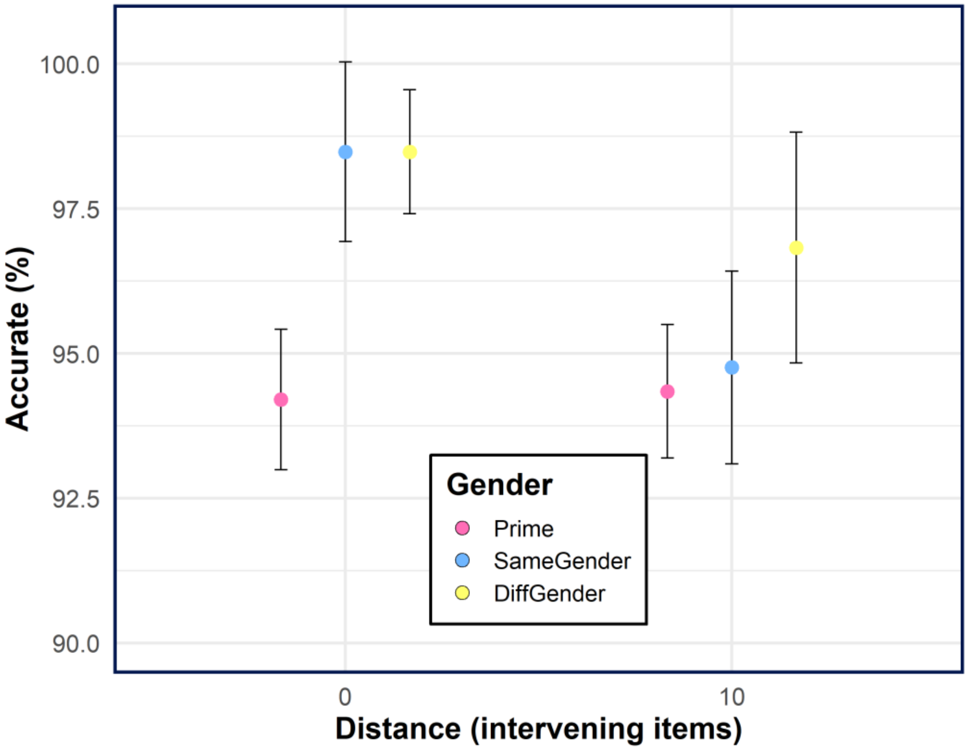

Figure 17: Experiment 3 accuracy data

A generalized linear mixed-effects model was fit to the accuracy data. This model combines primes and fillers together as the reference level for the condition factor, with targets in each combination of GENDER, DISTANCE, and DIRECTION dummy-coded. Random effects were set as intercepts for participants and items. The outcome of this model is seen in Table 17.

Table 17: Experiment 3 combined accuracy model

	Accuracy (combined)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	31.97	18.82 – 54.29	**<0.001**
GENDER × DISTANCE			
Diff. at 0	6.82	2.30 – 20.21	**0.001**
Diff. at 10	7.25	2.44 – 21.57	**<0.001**
Same at 0	7.73	2.58 – 23.15	**<0.001**
Same at 10	1.25	0.66 – 2.36	0.495
DIRECTION			
MA1-target	1.41	0.76 – 2.62	0.275
Item (z-scored)			
Frequency	0.83	0.62 – 1.12	0.226
PNH	0.70	0.56 – 0.88	**0.002**
AoA	0.78	0.62 – 1.00	**0.047**
Duration	1.03	0.80 – 1.31	0.838
Trial	0.78	0.66 – 0.91	**0.002**
Participant			
zAge	0.74	0.55 – 0.98	**0.037**
Male	1.60	0.83 – 3.07	0.157
Group			
2	1.22	0.60 – 2.50	0.580
3	1.52	0.77 – 2.97	0.226
4	1.10	0.57 – 2.15	0.773
GEN × DIST. (Diff.-0) × DRCTN (MA1-t)	0.37	0.09 – 1.59	0.182
GEN × DIST. (Diff.-10) × DRCTN (MA1-t)	0.13	0.04 – 0.50	**0.003**
GEN × DIST. (Same-0) × DRCTN (MA1-t)	0.34	0.08 – 1.46	0.145
GEN × DIST. (Same-10) × DRCTN (MA1-t)	0.83	0.30 – 2.28	0.722
Observations	5705		
Marginal R^2 / Conditional R^2	0.146 / 0.400		

This model shows whether any priming was seen when items were repeated. Immediately repeated targets were identified significantly more accurately (same-gender: $p < 0.001$, different-gender: $p = 0.001$). Stark differences exist in the long-distance data however. For same-gender pairs, no significant priming in accuracy was found ($p = 0.495$). Robust accuracy priming was found with different-gender pairs ($p < 0.001$).

This difference was only partially borne out in the model examining the interaction of the experimental predictors of GENDER and DISTANCE. In the table below, the results of the data considering only targets are displayed. GENDER, DISTANCE, and DIRECTION were dummy-coded, with the reference levels set to same-gender pairs at the immediate distance with speaker MA1 as the prime.

Table 18: Experiment 3 target accuracy model

	Accuracy (targets)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	185.98	56.34 – 613.91	**<0.001**
GENDER			
Diff.	0.98	0.22 – 4.40	0.983
DISTANCE			
10 interveners	0.18	0.06 – 0.56	**0.003**
DIRECTION			
MA1-target	0.69	0.17 – 2.75	0.599
Item (z-scored)			
Frequency	0.78	0.57 – 1.07	0.128
PNH	0.68	0.50 – 0.92	**0.012**
AoA	0.91	0.66 – 1.25	0.560
Duration	1.00	0.72 – 1.38	1.000
Trial	0.86	0.66 – 1.12	0.256
Participant			
zAge	0.86	0.62 – 1.19	0.358
Male	1.12	0.54 – 2.32	0.751
Group			
2	1.07	0.47 – 2.42	0.871
3	1.39	0.63 – 3.04	0.411
4	1.00	0.46 – 2.16	0.991
GEN (Diff.) × DIST. (10)	5.58	0.92 – 33.97	0.062
GEN (Diff.) × DRCTN (MA1-t)	0.99	0.16 – 6.31	0.995
DIST. (10) × DRCTN (MA1-t)	1.97	0.46 – 8.46	0.361
V. (Diff.) × DI. (10) × DR. (MA1-t)	0.17	0.02 – 1.54	0.114
Observations	2862		
Marginal R^2 / Conditional R^2	0.130 / 0.405		

Here we see that while the DISTANCE manipulation yielded significant inhibitory effects on accuracy (p = 0.003), the GENDER manipulation did not significantly affect accuracy. The main effect (p = 0.983) indicates a non-significant effect at the immediate distance. However, the marginally significant interaction term illustrates the different-gender asymmetry in the long-distance condition that was found in the previous model (p = 0.062). Altogether, the accuracy results hint that different-gender repetitions were recognized *more* *accurately* than within-gender (but different-voice) repetitions. This finding, which contradicts the hypothesis that perceptual similarity increases priming effects, should be kept in mind in examining the reaction time data.

3.4.2.2 Reaction Time

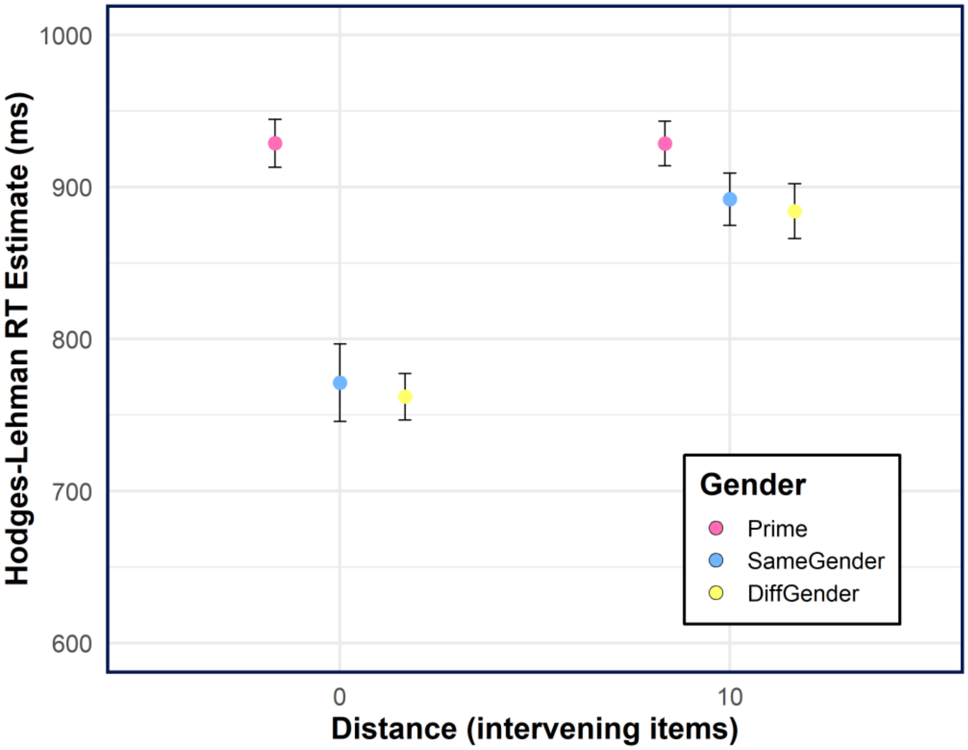

Figure 18: Experiment 3 reaction time data

The first analysis of reaction time comes from the large model investigating whether repetition priming existed in all experimental conditions. The responses to primes and targets were combined into one dataset and a linear mixed-effect model was run examining the $\log_2$–transformed RTs. Random effects included by-subject and by-item intercepts along with random slopes for the factors of GENDER$_{\text{[Same vs. Different]}}$. Fixed effects of interest were a dummy-coded variable with the baseline being responses to the primes and factor levels indicating each of the four GENDER by DISTANCE conditions, which interacted with a dummy-coded variable indicating the $\text{DIRECTION}_{\text{[MA1-prime vs. MA1-target]}}$ of the item spoken by MA1. Model criticism resulted in 120 additional observations being removed, a total of 2.41% of the remaining data.

Table 19: Experiment 3 combined RT model

	Log$_2$-transformed RT (combined)		
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.82	9.76 – 9.88	**<0.001**
GENDER × DISTANCE			
Diff. at 0	-0.30	-0.35 – -0.25	**<0.001**
Diff. at 10	-0.13	-0.18 – -0.09	**<0.001**
Same at 0	-0.31	-0.36 – -0.26	**<0.001**
Same at 10	-0.11	-0.16 – -0.06	**<0.001**
DIRECTION			
MA1-target	-0.04	-0.11 – 0.04	0.350
Item (z-scored)			
Frequency	0.00	-0.01 – 0.01	0.870
PNH	0.02	0.01 – 0.03	**0.001**
AoA	0.02	0.01 – 0.03	**0.001**
Duration	0.07	0.06 – 0.08	**<0.001**
Trial	-0.02	-0.03 – -0.02	**<0.001**
Previous RTLog	0.04	0.04 – 0.05	**<0.001**
Participant			
zAge	-0.00	-0.04 – 0.03	0.807
Male	-0.04	-0.13 – 0.04	0.309
Group			
2	-0.03	-0.13 – 0.06	0.502
3	-0.13	-0.22 – -0.04	**0.010**
4	-0.10	-0.19 – -0.01	**0.035**
GEN × DIST. (Diff.-0) × DRCTN (MA1-t)	0.07	0.00 – 0.14	**0.048**
GEN × DIST. (Diff.-10) × DRCTN (MA1-t)	0.08	0.01 – 0.15	**0.020**
GEN × DIST. (Same-0) × DRCTN (MA1-t)	0.10	0.03 – 0.17	**0.009**
GEN × DIST. (Same-10) × DRCTN (MA1-t)	0.05	-0.02 – 0.12	0.204
Random Effects			
σ^2	0.05		
τ_{00} Item	0.01		
τ_{00} Participant	0.01		
τ_{11} Participant × GENDER (Diff.)	0.00		
τ_{11} Participant × GENDER (Same)	0.00		
ρ_{01} Participant × GENDER (Diff.)	-0.13		
ρ_{01} Participant × GENDER (Same)	-0.40		
Observations	4868		
Marginal R^2 / Conditional R^2	0.276 / 0.435		

For the comparisons of interest, all factor levels came out significant (all $p < 0.001$; same-gender at 0 = 19% / 106ms; diff- gender at 0 = 19% / 103ms; same- gender at 10 = 7% / 40ms; diff- gender at 10 = 9% / 48ms), indicating that in each condition formed by the interaction of the factors GENDER and DISTANCE, significant priming was found. These comparisons remained highly significant after doing a multiple comparisons correction using Holm's method. Three additional interaction terms came out significant each indicating a small reduction in priming when speaker MA1 produced the target instead of the prime, but these will not be discussed further here. Overall, this model, indicates that significant abstract repetition effects were found.

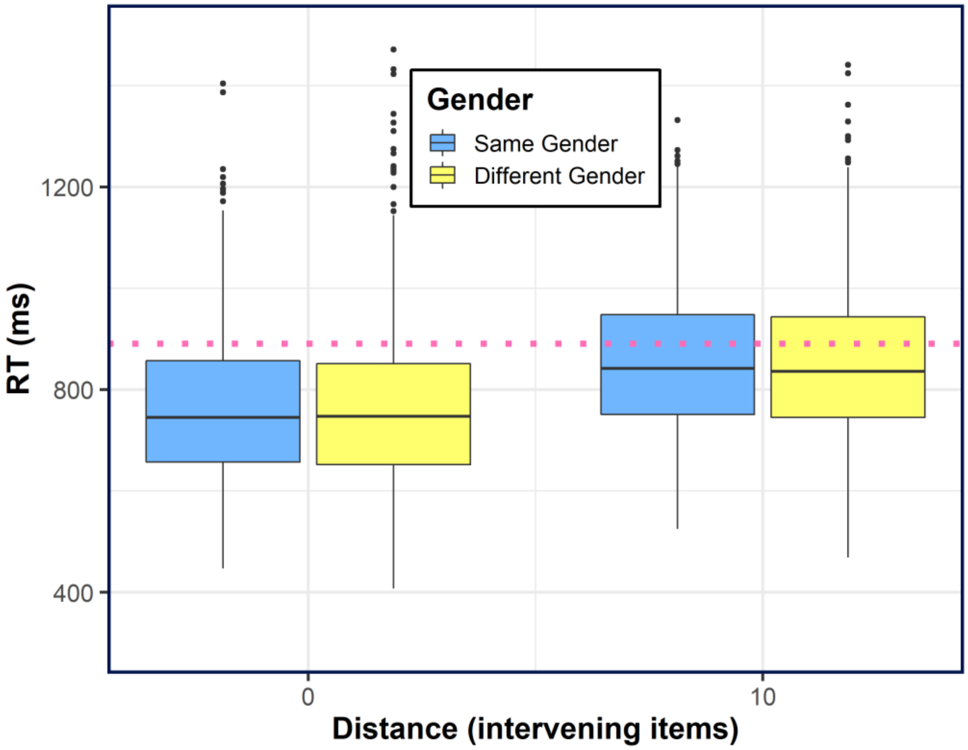

Figure 19: Experiment 3 trimmed reaction time

Next, we examine the model testing the differences between same- and different-gender conditions at immediate and long distances. The random effects for this model consisted of by-subject and by-item intercepts only. The fixed effects of interest were formed by the interactions between three terms: the dummy-coded factors of GENDER (baseline = same-gender), DISTANCE (baseline = immediate repetition), and DIRECTION (baseline = speaker MA1 producing the prime). After fitting the model, 46 additional values (1.84% of the remaining data) with residuals > 2.5 SDs from the mean were removed.

Table 20: Experiment 3 target RT model

	Log$_2$-transformed RT (targets)		
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.55	9.49 – 9.61	**<0.001**
GENDER			
Diff.	-0.07	-0.11 – -0.03	**0.001**
DISTANCE			
10 interveners	0.19	0.16 – 0.23	**<0.001**
DIRECTION			
MA1-target	0.02	-0.05 – 0.10	0.547
Item (z-scored)			
Frequency	0.00	-0.01 – 0.02	0.618
PNH	0.02	0.00 – 0.03	**0.016**
AoA	0.01	-0.01 – 0.02	0.294
Duration	0.08	0.07 – 0.10	**<0.001**
Trial	-0.04	-0.05 – -0.03	**<0.001**
Previous RTLog	0.05	0.04 – 0.05	**<0.001**
Participant			
zAge	-0.00	-0.04 – 0.04	0.855
Male	-0.02	-0.11 – 0.07	0.644
Group			
2	-0.04	-0.14 – 0.07	0.488
3	-0.14	-0.24 – -0.05	**0.008**
4	-0.08	-0.18 – 0.02	0.130
GENDER (Diff.) × DIST. (10)	-0.02	-0.07 – 0.03	0.390
GENDER (Diff.) × DRCTN (MA1-t)	0.05	-0.00 – 0.10	0.073
DIST. (10) × DRCTN (MA1-t)	-0.04	-0.09 – 0.00	0.069
G. (Diff.) × DI. (10) × DR. (MA1-t)	0.05	-0.01 – 0.12	0.113
Random Effects			
σ^2	0.04		
τ_{00} Item	0.01		
τ_{00} Participant	0.02		
Observations	2448		
Marginal R^2 / Conditional R^2	0.291 / 0.467		

The main effects for the factors of interest came out significant, surprisingly indicating *faster* reaction times in different-gender pairs compared to same-gender pairs (p = 0.001, 5% / 26ms) and a decrease from immediate to long-distance (p < 0.001, 14% / -78ms). No significant interaction terms were implicated by the model.

We now turn to two additional models separated by the DISTANCE condition to interpret these interaction effects. These models were the same as the experimental model, except for the removal of the DISTANCE factor from the variables of experimental interest. The first model of priming effects at the immediate distance (with 24 responses removed totaling 1.86% of the data) indicated a significant speed-up in the different-gender condition compared to the same-voice condition (p = 0.018, 3.5% / 19ms). The second model looking at the long-distance condition (12 responses removed totaling 0.99% of the data), indicated the same speed-up in the different-gender condition (p < 0.001, 5.6% / 30ms). In the second model only, the interaction effect of GENDER and DIRECTION came out significant, indicating a slow-down in the different-gender condition when speaker MA1 produced the target (p < 0.001, compared to p = 0.14 in the immediate model).

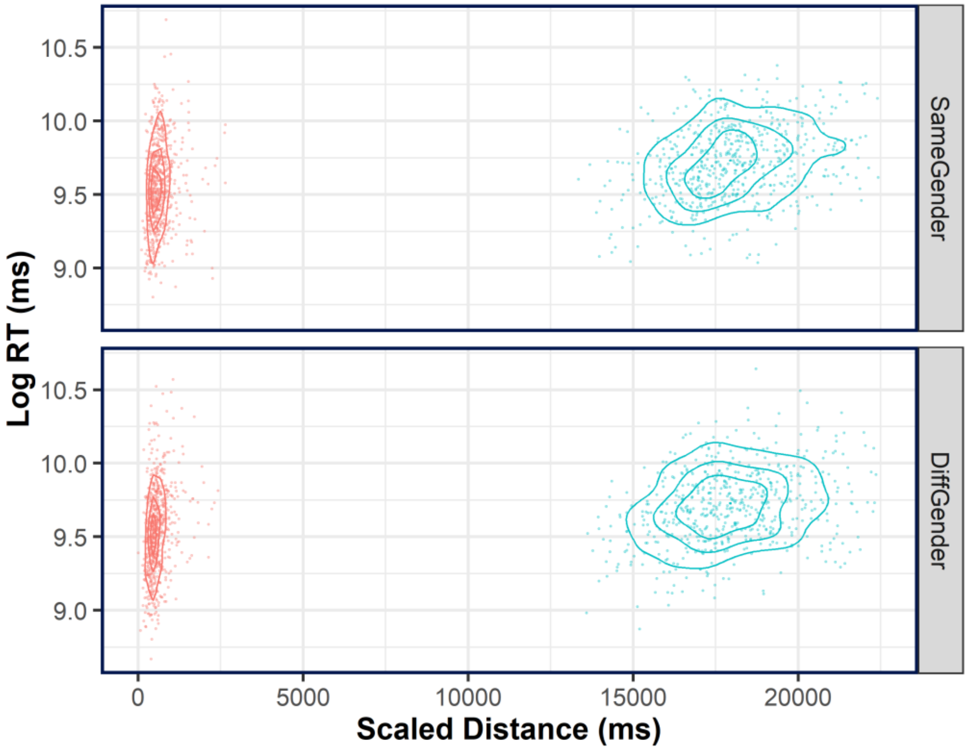

Figure 20: Experiment 3 effects of time vs. interveners

Using these two models separated by distance, we then tested the impact that the passage of raw time has over the interference due to intervening items. To do so, we used the chi-squared test to compare models with and without an additional factor of z-scored, $\log_2$-transformed time between the end of the prime and the beginning of the target. At immediate distances, the raw time estimate did not significantly improve the model ($p = 0.357$) unlike in Experiment 2. In the long-distance condition, the raw time estimate did significantly improve the model ($p < 0.001$) and indicated that increased raw time slowed reaction times, mirroring Experiment 2. In this experiment, raw time only significantly mattered in the long-distance condition. The data is plotted in Figure 20, with color indicating the number of intervening items.

3.4.3 Discussion

The analysis of Experiment 3 indicates the following:

- **Abstract repetition priming**
    - *Accuracy:*
        - All immediately repeated targets recognized more accurately than primes
        - At long-distance, only different-gender targets recognized more accurately than primes
    - *Reaction time:*
        - All repeated targets in both distances recognized faster than primes
        - Speaker MA1 targets recognized slower than those from MA2/FM1
- **Talker-specific effects**
    - *Accuracy:*
        - Increased distance reduced accuracy
        - No difference between same- and different-gender targets
    - *Reaction time:*
        - Increased distance increased reaction time
        - Different-gender pairs recognized *faster* than same-gender pairs
- Increasing raw time between prime and target has no effect in the immediate condition and slows reaction time in the long-distance condition
- Confluence of (sometimes marginally) significant interaction terms indicates speaker MA1 targets recognized slower and less accurately than those from MA2/FM1

These models indicate the surprising result that different-gender pairs are recognized more accurately and faster than same-gender pairs. This results directly contradict the findings of Orfanidou et al. (2006, 2011), who found no facilitatory priming effects at distances of on average ten intervening items. Additionally, they contradict the results finding that increased perceptual similarity yields facilitatory priming (cf. Goldinger, 1996; Sheffert, 1998). Considered with the results of Experiment 2, the interesting conclusion emerges that the switch between male and female voices yields qualitatively different priming patterns than switching between two male voices. Therefore, calling effects of switching gender ‘talker-specific effects’ complicates the picture we see from the literature, as talker-specific effects need to be separately investigated by switching between voices of the same gender.

3.5 Experiment 4

The final experiment presented in this chapter investigates a discrete time-course of TSEs at distances of 0, 1, 2, 3, 4, 5, 7, and 10 intervening items. Previous research, including our previous experiments, has failed to reveal long-distance TSEs from switching voices between two male speakers. Indeed, although we do find long-distance TSEs in Experiment 3, the effects run in the opposite direction from the predictions. Assuming that switching gender reduces perceptual similarity between voices, the prediction would have been that same-gender pairs would have been recognized faster and more accurately than different-gender pairs. This was not the case, which contradicts models of speech perception based upon the perceptual comparison of detail-rich representations.

Considering the long-distance null result found in Experiment 2, we now focus on determining the decay pattern of the time-course of TSEs due to same-gender voice switch. A concrete investigation of this has, to our knowledge, not been conducted; many studies cite the lack of long-distance TSEs (reviewed in Section 3.2.3) but it remains unclear how long they actually last. Additionally, as a reminder, we are interested in determining whether TSEs affect the short- or long-distance sources of repetition priming effects, discussed in Section 3.2.4. The prediction is that TSEs will impact the short-distance source of priming, as Experiment 2 revealed no differences between same- and different-voice targets at ten intervening items. Starting from this as a maximum lag, this experiment additionally tests seven shorter lags with the intention of determining exactly how long TSEs due to voice persist.

3.5.1 Method

3.5.1.1 Participants

A total of 137 (age range 18-74, mean 35.3; 88 male) participants were run at the end of the fall semester of 2017. They were recruited from the experimental platform Prolific and voluntarily completed the study online using a custom Ibex implementation of a continuous lexical decision task.

3.5.1.2 Stimuli

This experiment consisted of a different set of 288 total stimuli (split halfway between words and non-words). Two tokens of each stimulus were recorded and used in the experiment by speakers MA1 and MA3. For a complete list of the properties of the stimuli in this experiment, see APPENDIX I: Experimental stimuli. This stimuli list was used also for Experiment 5. The figure below illustrates the relationship between frequency (mean = 3.07, standard deviation = 0.34, range = [2.34, 3.87]), age of acquisition (mean = 5.05, standard deviation = 1.07, range = [3.52, 7.80]), and phonological neighborhood density (mean = 16.80, standard deviation = 9.23, range = [1, 40]) for the words in this experiment. In constructing this set, we attempted to minimize the variation along these three lexical properties, resulting in a more tightly-controlled set of stimuli than in the previous experiments.

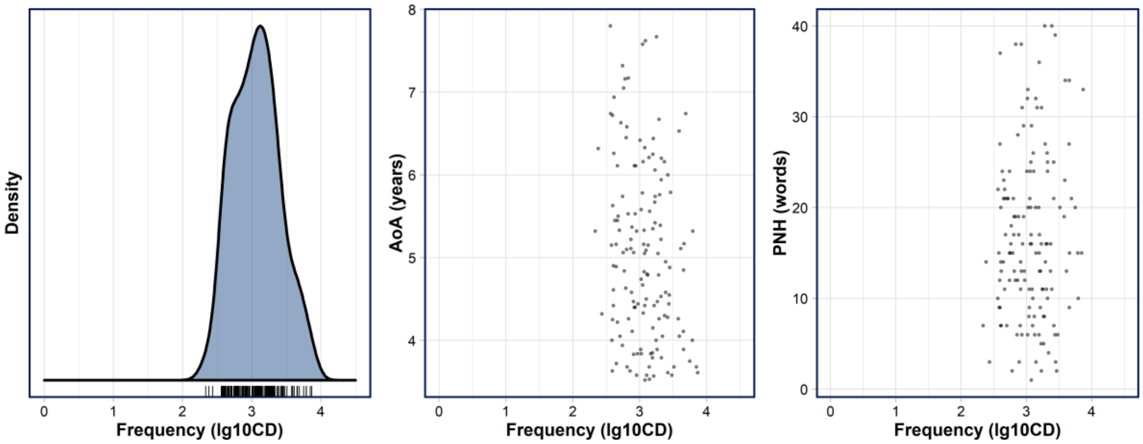

Figure 21: Properties of Experiments 4-5 word stimuli

Similarly, the non-word stimuli in this set were chosen to remove the problematic non-words from previous experiments and to maximize the phonotactic probability. Using the calculated bi-gram metric described in Section 2.2.2, the values from both words and non-words were calculated. The first facet plots the mean bi-gram frequency and the second plots the standard deviation. The third and fourth facets plot the minimum and maximum bi-gram frequency from each stimulus. Indeed, in this set, the non-words are slightly phonotactically *more* licit than the words, with higher means, lower deviations, and higher minimums. The differences are however not numerically much greater than the previous experiments, so we do not expect any difference in participant's task strategies.

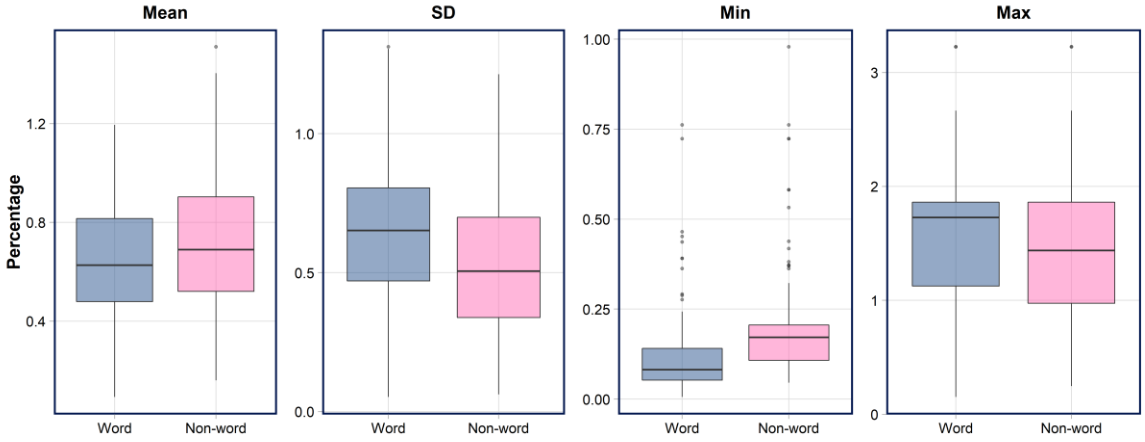

Figure 22: Phonotactic prob. of words (left) & non-words (right) in Experiments 4-5

3.5.1.3 Design

Unlike the previous experiments, each of the 288 words and non-words were repeated; no non-repeated fillers were included. An initial four lists were created by varying the factors VOICE[Same vs. Different Voice] and SPEAKER[MA1-prime vs. MA3-prime]. These four lists insured that each word in the experiment would be seen between-subjects in all four possible combinations. The experiment design was complicated by the fact that a by-subjects randomization was added. To do so, a templatic interleaving pattern was created, inspired by McKone (1995) and implemented by the author with assistance from Jérémy Zehr. A total of twelve unique patterns were created, corresponding to four groups of three blocks. Each of these patterns included slots for primes and targets with distances randomly interspersed throughout. Each of the twelve patterns had 192 slots for items (i.e., prime/target slots for 24 of each of the eight distances). Distances were matched between each pattern. When a participant took the experiment, their list of items were randomly inserted into these prime/target slots within three patterns, creating a different random list for each subject. These items were balanced by condition and lexicality status, resulting in three pairs each of words and non-words per combination of VOICE$_{\text{[Same vs. Different Voice]}}$, SPEAKER$_{\text{[MA1-prime vs. MA3-prime]}}$, and DISTANCE$_{[0,1,2,3,4,5,7,10]}$ per block. Over the whole experiment, three blocks were presented, resulting in nine items per experimental manipulation. The repetition rate over the whole experiment was 50%, however the amount of immediate repetition was quite low. The increased amount of repetition, given the results of Hanique et al. (2013), is predicted to increase the presence of TSEs; a prediction we need to bear in mind for the conclusions.

3.5.1.4 Procedure

The experiment, as presented on Prolific, was titled *Word Recognition* and had the description *"In this study, you will listen to three groups of sounds and indicate with the keyboard whether each sound was a word or not. You will also respond to a very brief demographic questionnaire. Completion time depends on your internet connection speed, as downloading the stimuli may take a while."* Participants completed ten practice lexical decisions as practice before the experiment. These consisted of the following items, randomly presented per subject with an 800ms ISI (with subscript representing different speakers and bold face indicating non-words): *fluff*$_1$, ***fiyk***$_1$, ***gaech***$_2$, *lump*$_2$, ***skrown***$_2$, *plum*$_1$, *smell*$_2$, ***kwaanch***$_1$, and ***nowk***$_1$. Feedback was given for these practice items, as the online nature of the experiment prevented participants from asking for clarification before beginning the experiment. Following the practice session, three blocks of 192 items each were presented (576 total lexical decisions) using the custom Ibex continuous lexical decision task implementation described in Section 2.2.4. Participants were encouraged to take breaks only at the end of a block. All experiments included in the analysis were completed on average within 24 minutes (range: 15-67 minutes). As mentioned, each participant had their own specific trial order due to the distance manipulation.

3.5.1.5 Analysis

Of the original 137 participants in the experiment, 26 participants were removed due to overall poor performance. All of these were removed due to an overall accuracy score of less than 70% correct, although subsets of the 26 showed poor performance by the other measures (Hodges-Lehmann estimated RT distribution outliers and/or over 20 near-immediate responses). After this global participant removal, the overall results per each item were examined. With this experiment, no items had accuracy scores of less than 50% correct. The table below describes the amount of data trimmed using the common procedures after the global removal steps were taken. Overall, a greater percentage of data was removed, specifically from inaccurate trials. We suspect that this was an unfortunate by-product of conducting the study online as opposed to in the lab.

Table 21: Experiment 4 removal summary

	Observations	**Percentage**
Inaccurate trials	*2167*	*13.6*
RT trimming (300 > RT < 3000)	*353*	*2.2*
By-participant trimming	*488*	*3.1*
By-item trimming	*394*	*2.5*
Total removed	*3402*	*21.3*
Total remaining	*12582*	

Table 22: Experiment 4 data summary

Distance			**Voice**			**Speaker**			**Total**
0 Interveners	*747*	*95.6*	**Same**	*712*	*95.2*	MA1	*728*	*93.8*	***497***
						MA3	*696*	*96.7*	***484***
			Diff	*782*	*96*	MA1	*790*	*96.9*	***508***
						MA3	*774*	*95.2*	***478***
1 Interveners	*829*	*93.6*	**Same**	*821*	*92.9*	MA1	*848*	*91.9*	***492***
						MA3	*794*	*93.9*	***477***
			Diff	*836*	*94.3*	MA1	*851*	*93.7*	***492***
						MA3	*822*	*95*	***479***
2 Interveners	*868*	*92.2*	**Same**	*865*	*92*	MA1	*882*	*90.8*	***476***
						MA3	*850*	*93.3*	***491***
			Diff	*872*	*92.4*	MA1	*886*	*90.9*	***483***
						MA3	*858*	*93.9*	***476***
3 Interveners	*888*	*93*	**Same**	*882*	*93.7*	MA1	*909*	*94*	***481***
						MA3	*856*	*93.4*	***483***
			Diff	*894*	*92.4*	MA1	*908*	*91.6*	***475***
						MA3	*882*	*93.1*	***494***
4 Interveners	*888*	*92.1*	**Same**	*886*	*91.6*	MA1	*901*	*90.1*	***485***
						MA3	*870*	*93.2*	***482***
			Diff	*890*	*92.6*	MA1	*902*	*91.2*	***477***
						MA3	*877*	*93.9*	***491***
5 Interveners	*890*	*93.5*	**Same**	*893*	*93.5*	MA1	*912*	*92.6*	***476***
						MA3	*874*	*94.4*	***485***
			Diff	*888*	*93.4*	MA1	*904*	*92.2*	***485***
						MA3	*872*	*94.7*	***474***
7 Interveners	*876*	*93.8*	**Same**	*884*	*93.6*	MA1	*906*	*93.4*	***484***
						MA3	*863*	*93.8*	***485***
			Diff	*870*	*94.1*	MA1	*893*	*93.6*	***471***
						MA3	*848*	*94.5*	***493***
10 Interveners	*884*	*93.2*	**Same**	*880*	*93.4*	MA1	*895*	*93.6*	***483***
						MA3	*864*	*93.2*	***487***
			Diff	*888*	*93*	MA1	*900*	*90.7*	***472***
						MA3	*875*	*95.2*	***483***
Primes	*932*	*90.9*				MA1	*949*	*90.3*	***7730***
						MA3	*914*	*91.5*	***7723***

3.5.2 Results

Table 22 reports the distribution of the data per the factors Voice, Distance, and Speaker. These data are reported after only the minimal global trimming procedure of unreasonable reaction times was applied. Reaction time reports are central tendency measures from the non-parametric Hodges-Lehmann estimates and accuracy scores indicate the amount of correct responses out of the total, after all global participant and item removal was conducted.

3.5.2.1 Accuracy

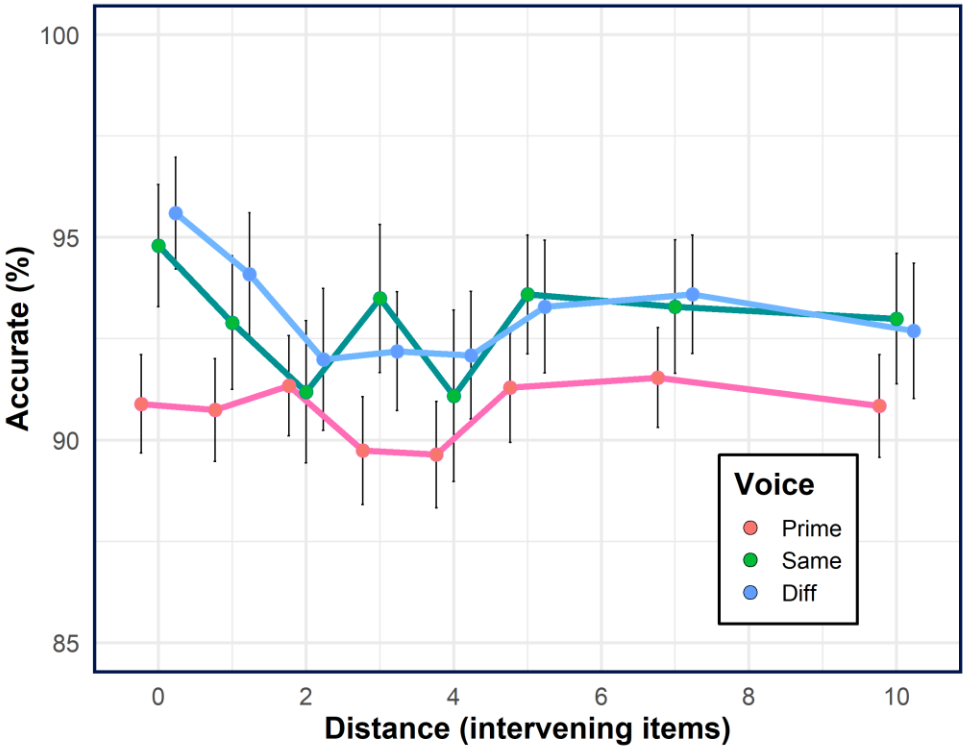

Figure 23: Experiment 4 accuracy data

A generalized linear mixed-effects model was fit to the accuracy data. This model sets the primes as the reference level for the condition factor, with targets in each combination of VOICE, DISTANCE, and SPEAKER dummy-coded. Random effects were set as intercepts for participants and items. The outcome of this model is seen below.

Table 23: Experiment 4 combined accuracy model

	Accuracy (combined)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	18.17	14.24 – 23.18	**<0.001**
Voice × Distance			
Same at 0	1.74	1.10 – 2.75	**0.018**
Diff. at 0	3.97	2.33 – 6.74	**<0.001**
Same at 1	1.27	0.83 – 1.95	0.264
Diff. at 1	1.45	0.98 – 2.15	0.061
Same at 2	1.02	0.67 – 1.55	0.928
Diff. at 2	1.13	0.80 – 1.61	0.482
Same at 3	1.64	1.03 – 2.61	**0.038**
Diff. at 3	1.06	0.74 – 1.51	0.759
Same at 4	1.01	0.67 – 1.53	0.950
Diff. at 4	1.28	0.89 – 1.82	0.181
Same at 5	1.53	0.98 – 2.41	0.063
Diff. at 5	1.34	0.93 – 1.94	0.117
Same at 7	1.59	1.01 – 2.52	**0.046**
Diff. at 7	1.56	1.04 – 2.32	**0.031**
Same at 10	1.63	1.03 – 2.59	**0.037**
Diff. at 10	1.19	0.84 – 1.69	0.331
Speaker			
MA3	1.16	0.89 – 1.52	0.280
Item (z-scored)			
Frequency	1.11	1.01 – 1.23	**0.037**
PNH	1.05	0.95 – 1.17	0.311
AoA	0.90	0.81 – 0.99	**0.038**
Duration	1.05	0.95 – 1.16	0.296
Trial	0.96	0.91 – 1.00	**0.049**
Participant			
zAge	1.02	0.87 – 1.20	0.816
Male	0.65	0.47 – 0.91	**0.011**
Group			
2	0.73	0.47 – 1.13	0.162
3	0.97	0.62 – 1.53	0.912
4	0.68	0.43 – 1.07	0.096
Voice × Dist. (Same-0) × Spkr (MA3)	1.72	0.83 – 3.56	0.142
Voice × Dist. (Diff.-0) × Spkr (MA3)	0.57	0.28 – 1.15	0.115
Voice × Dist. (Same-1) × Spkr (MA3)	1.12	0.60 – 2.09	0.733
Voice × Dist. (Diff.-1) × Spkr (MA3)	1.54	0.85 – 2.79	0.152
Voice × Dist. (Same-2) × Spkr (MA3)	1.27	0.69 – 2.33	0.437
Voice × Dist. (Diff.-2) × Spkr (MA3)	1.34	0.78 – 2.29	0.284
Voice × Dist. (Same-3) × Spkr (MA3)	0.79	0.42 – 1.52	0.486
Voice × Dist. (Diff.-3) × Spkr (MA3)	1.28	0.76 – 2.16	0.356
Voice × Dist. (Same-4) × Spkr (MA3)	1.22	0.66 – 2.23	0.526
Voice × Dist. (Diff.-4) × Spkr (MA3)	1.15	0.68 – 1.96	0.602
Voice × Dist. (Same-5) × Spkr (MA3)	1.00	0.52 – 1.91	0.997
Voice × Dist. (Diff.-5) × Spkr (MA3)	1.26	0.72 – 2.21	0.416
Voice × Dist. (Same-7) × Spkr (MA3)	0.81	0.43 – 1.55	0.534
Voice × Dist. (Diff.-7) × Spkr (MA3)	1.01	0.57 – 1.80	0.968
Voice × Dist. (Same-10) × Spkr (MA3)	0.83	0.44 – 1.57	0.563
Voice × Dist. (Diff.-10) × Spkr (MA3)	1.60	0.91 – 2.80	0.101
Observations	30932		
Marginal R^2 / Conditional R^2	0.039 / 0.356		

This model shows whether any priming was seen when items were repeated by comparing each repetition condition with the accuracy of the primes. In this large model, we do not see much of an interpretable pattern. Immediately repeated targets were identified significantly more accurately (same-voice: p = 0.018, different-voice: p < 0.001). After the immediate distance, a few other conditions come out with significant accuracy priming effects (same-voice at 3: p = 0.038, same-/different-voice at 7: p = 0.046/0.031, same-voice at 10: p = 0.037). The numerical trend however is for slight accuracy priming in each repeated condition with targets from speaker MA1. Targets from speaker MA3 are recognized less accurately, as indicated by the interaction effects.

The model examining the interaction of the experimental predictors of VOICE, DISTANCE, and SPEAKER on target accuracy is presented in the table below. The factors VOICE and SPEAKER dummy-coded as in the full model. The contrast coding for DISTANCE however was backwards-difference coded. This coding scheme compares each level of the DISTANCE condition with the prior level. For example, the first coefficient *1-0 interveners* compares the one word intervening condition with the immediate intervening condition. This was done to better test the experimental predictions this design allows for. We would not think for example that seven interveners would result in actual better performance than five interveners. The significant effects from the full model are hypothesized to be statistical anomalies given the large amount of comparisons performed in this model.

Table 24: Experiment 4 target accuracy model

	Accuracy (targets)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	25.85	19.67 – 33.98	**<0.001**
VOICE			
Diff.	1.01	0.74 – 1.36	0.970
DISTANCE			
1-0 interveners	0.74	0.45 – 1.23	0.245
2-1 interveners	0.80	0.50 – 1.28	0.357
3-2 interveners	1.59	0.96 – 2.64	0.074
4-3 interveners	0.64	0.38 – 1.05	0.078
5-4 interveners	1.51	0.92 – 2.47	0.101
7-5 interveners	1.04	0.61 – 1.76	0.896
10-7 interveners	1.00	0.58 – 1.71	0.991
SPEAKER			
MA3	1.21	0.88 – 1.65	0.237
Item (z-scored)			
Frequency	1.08	0.97 – 1.21	0.175
PNH	1.05	0.94 – 1.17	0.402
AoA	0.89	0.80 – 1.00	**0.043**
Duration	1.02	0.91 – 1.15	0.718
Trial	0.95	0.89 – 1.02	0.132
Participant			
zAge	1.00	0.84 – 1.19	0.989
Male	0.61	0.43 – 0.87	**0.007**
Group			
2	0.76	0.45 – 1.30	0.319
3	0.89	0.52 – 1.51	0.660
4	0.68	0.39 – 1.18	0.167
VOICE (Diff.) × DIST. (1-0)	0.57	0.26 – 1.28	0.173
VOICE (Diff.) × DIST. (2-1)	0.88	0.44 – 1.75	0.721
VOICE (Diff.) × DIST. (3-2)	0.60	0.30 – 1.19	0.141
VOICE (Diff.) × DIST. (4-3)	1.82	0.91 – 3.64	0.090
VOICE (Diff.) × DIST. (5-4)	0.74	0.37 – 1.48	0.393
VOICE (Diff.) × DIST. (7-5)	1.16	0.55 – 2.43	0.696
VOICE (Diff.) × DIST. (10-7)	0.68	0.33 – 1.42	0.303
DIST. (1-0) × SPEAKER (MA3)	1.14	0.74 – 1.76	0.550
DIST. (2-1) × SPEAKER (MA3)	0.64	0.29 – 1.44	0.281
DIST. (3-2) × SPEAKER (MA3)	1.16	0.57 – 2.34	0.689
DIST. (4-3) × SPEAKER (MA3)	0.63	0.31 – 1.31	0.217
DIST. (5-4) × SPEAKER (MA3)	1.46	0.71 – 3.01	0.300
DIST. (7-5) × SPEAKER (MA3)	0.83	0.40 – 1.72	0.621
DIST. (10-7) × SPEAKER (MA3)	0.82	0.38 – 1.74	0.599
V. (Diff.) × D. (1-0) × S. (MA3)	1.03	0.49 – 2.19	0.929
V. (Diff.) × D. (2-1) × S. (MA3)	3.60	1.10 – 11.71	**0.034**
V. (Diff.) × D. (3-2) × S. (MA3)	0.86	0.30 – 2.41	0.770
V. (Diff.) × D. (4-3) × S. (MA3)	1.45	0.53 – 3.98	0.470
V. (Diff.) × D. (5-4) × S. (MA3)	0.61	0.22 – 1.68	0.342
V. (Diff.) × D. (7-5) × S. (MA3)	1.30	0.46 – 3.64	0.621
V. (Diff.) × D. (10-7) × S. (MA3)	0.93	0.32 – 2.75	0.903
Observations	15479		
Marginal R^2 / Conditional R^2	0.038 / 0.354		

In this model, we again fail to find any discernable priming pattern in accuracy. The comparisons of each distance with the distance before does not reveal any significant effects (with the exception of the three-way interaction of different-voice, 1 intervener to immediate when speaker MA3 produced the target). This indicates that the increase in the number of interveners between prime and target did not cause any discernable effects on accuracy. Additionally, the VOICE manipulation did not appear to have any substantial effects as well. These results are useful to keep in mind when we compare them with the results of the recognition experiments presented in CHAPTER 4: Time-Course in Explicit Tasks.

3.5.2.2 Reaction Time

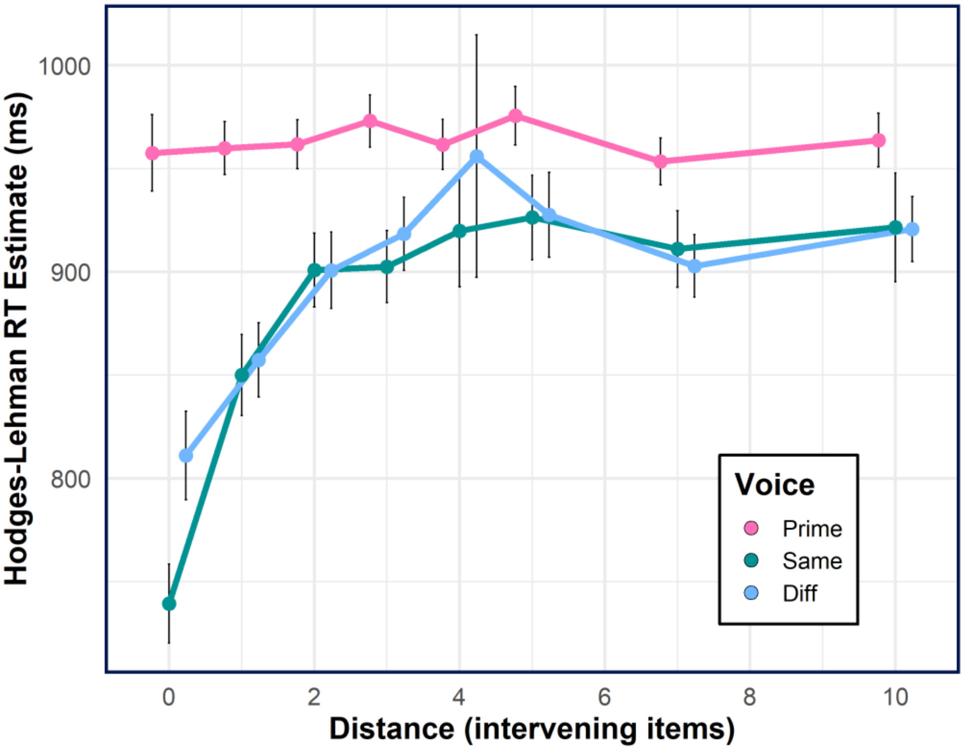

Figure 24: Experiment 4 reaction time data

The first analysis of reaction time comes from the large model investigating whether repetition priming existed in all experimental conditions. The responses to primes and targets were combined into one dataset and a linear mixed-effect model was run examining the $\log_2$-transformed RT. Random effects included by-subject and by-item intercepts along with random slopes for the factors of VOICE$_{\text{[Same vs. Different]}}$. Fixed effects of interest were a dummy-coded variable with the baseline being responses to the primes and factor levels indicating each of the four VOICE by DISTANCE conditions, which interacted with a dummy-coded variable indicating the SPEAKER$_{\text{[MA1 vs. MA3]}}$ of the target. Model criticism resulted in 588 additional observations being removed, a total of 2.34% of the remaining data.

For the comparisons of interest, all factor levels came out significant (all p < 0.001), indicating that in each condition formed by the interaction of the factors VOICE and DISTANCE, significant priming was found. These comparisons remained highly significant after doing a multiple comparisons correction using Holm's method. No other interaction terms came out significant, indicating no overall priming differences between speakers. Overall, this model, indicates that significant abstract repetition effects were found at all distances. Table 25 shows the output of the model and Table 26 shows the effects sizes of the abstract repetition priming found in this experiment

Table 25: Experiment 4 combined RT model

	Log_2-transformed RT (combined)		
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.84	9.82 – 9.87	**<0.001**
VOICE × DISTANCE			
Same at 0	-0.39	-0.41 – -0.36	**<0.001**
Diff. at 0	-0.28	-0.30 – -0.25	**<0.001**
Same at 1	-0.20	-0.23 – -0.18	**<0.001**
Diff. at 1	-0.17	-0.19 – -0.15	**<0.001**
Same at 2	-0.13	-0.15 – -0.11	**<0.001**
Diff. at 2	-0.12	-0.14 – -0.09	**<0.001**

Same at 3	-0.11	-0.13 – -0.09	**<0.001**
Diff. at 3	-0.09	-0.11 – -0.07	**<0.001**
Same at 4	-0.10	-0.13 – -0.08	**<0.001**
Diff. at 4	-0.09	-0.12 – -0.07	**<0.001**
Same at 5	-0.09	-0.11 – -0.07	**<0.001**
Diff. at 5	-0.08	-0.10 – -0.06	**<0.001**
Same at 7	-0.10	-0.12 – -0.07	**<0.001**
Diff. at 7	-0.10	-0.12 – -0.07	**<0.001**
Same at 10	-0.11	-0.14 – -0.09	**<0.001**
Diff. at 10	-0.08	-0.11 – -0.06	**<0.001**
SPEAKER			
MA3	-0.00	-0.02 – 0.01	0.830
Item (z-scored)			
Frequency	-0.01	-0.01 – 0.00	0.083
PNH	0.00	-0.00 – 0.01	0.159
AoA	0.02	0.01 – 0.03	**<0.001**
Duration	0.10	0.09 – 0.10	**<0.001**
Trial	-0.01	-0.01 – -0.01	**<0.001**
Previous RTLog	0.05	0.05 – 0.05	**<0.001**
Participant			
zAge	0.02	-0.01 – 0.04	0.179
Male	-0.01	-0.06 – 0.03	0.632
Group			
2	0.03	-0.03 – 0.10	0.291
3	-0.04	-0.11 – 0.02	0.187
4	0.05	-0.01 – 0.12	0.113
VOI × DIST. (Same-0) × SPKR (MA3)	0.01	-0.02 – 0.04	0.539
VOI × DIST. (Diff.-0) × SPKR (MA3)	0.00	-0.03 – 0.04	0.788
VOI × DIST. (Same-1) × SPKR (MA3)	-0.01	-0.05 – 0.02	0.391
VOI × DIST. (Diff.-1) × SPKR (MA3)	-0.01	-0.04 – 0.03	0.736
VOI × DIST. (Same-2) × SPKR (MA3)	0.01	-0.03 – 0.04	0.668
VOI × DIST. (Diff.-2) × SPKR (MA3)	0.01	-0.03 – 0.04	0.753
VOI × DIST. (Same-3) × SPKR (MA3)	-0.01	-0.04 – 0.03	0.632
VOI × DIST. (Diff.-3) × SPKR (MA3)	0.01	-0.03 – 0.04	0.688
VOI × DIST. (Same-4) × SPKR (MA3)	0.01	-0.03 – 0.04	0.766
VOI × DIST. (Diff.-4) × SPKR (MA3)	-0.01	-0.05 – 0.02	0.514
VOI × DIST. (Same-5) × SPKR (MA3)	-0.01	-0.04 – 0.03	0.670
VOI × DIST. (Diff.-5) × SPKR (MA3)	-0.02	-0.05 – 0.01	0.275
VOI × DIST. (Same-7) × SPKR (MA3)	-0.01	-0.04 – 0.03	0.754
VOI × DIST. (Diff.-7) × SPKR (MA3)	-0.00	-0.04 – 0.03	0.835
VOI × DIST. (Same-10) × SPKR (MA3)	0.02	-0.02 – 0.05	0.345
VOI × DIST. (Diff.-10) × SPKR (MA3)	0.00	-0.03 – 0.04	0.854
Random Effects			
σ^2	0.05		
τ_{00} Item	0.01		
τ_{00} Participant	0.01		
τ_{11} Participant × VOICE (Diff.)	0.00		
τ_{11} Participant × VOICE (Same)	0.00		
ρ_{01} Participant × VOICE (Diff.)	1.00		
ρ_{01} Participant × VOICE (Same)	0.25		
Observations	24576		
Marginal R^2 / Conditional R^2	0.267 / 0.479		

Table 26: Experiment 4 combined RT model effect sizes

	Same-Voice		**Different-Voice**	
DISTANCE	**Percentage**	**Effect size (ms)**	**Percentage**	**Effect size (ms)**
0	*23.5*	*131*	*17.4*	*97*
1	*13.0*	*73*	*11.1*	*62*
2	*8.6*	*48*	*7.9*	*44*
3	*7.3*	*41*	*6.1*	*34*
4	*6.9*	*38*	*6.2*	*34*
5	*6.0*	*34*	*5.4*	*30*
7	*6.6*	*37*	*6.4*	*36*
10	*7.5*	*42*	*5.7*	*32*

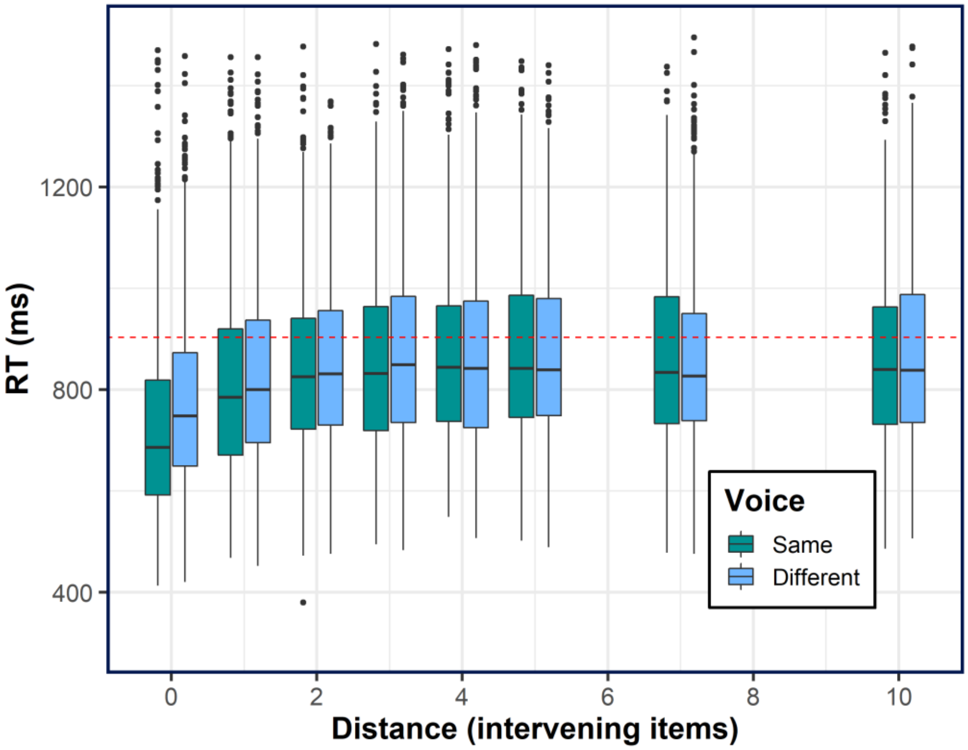

Figure 25: Experiment 4 trimmed reaction time

Next, we examine the model testing the differences between same- and different-voice conditions at immediate and long distances. The random effects for this model consisted of by-subject and by-item intercepts only. The fixed effects of interest were formed by the interactions between two terms: the dummy-coded factors of VOICE (baseline = same-voice) and the backwards-difference coded factor of DISTANCE. Again, backwards-difference coding compares each level of a factor to the level previous to it, allowing us to test more specifically the hypothesis that priming should decay over increasing interveners. The dummy-coded factor of SPEAKER (baseline = speaker MA1) was treated only as fixed effect without participating in any interactions, as the full model indicated no effect of speaker (and since including a three-way interaction dramatically expanded the running time of the models). After fitting the model, 239 additional values (1.90% of the remaining data) with residuals > 2.5 SDs from the mean were removed.

In this model, we see a main effect of VOICE introducing a slow-down from same to different voice pairs ($p < 0.001$). The backwards-difference coded DISTANCE factor only showed the main effects for one intervener differing from immediate priming ($p < 0.001$) and two interveners differing from one intervener ($p < 0.001$). Other main effects for DISTANCE were not significant, indicating a stable level of priming after two interveners in the same voice condition. The effect of DISTANCE was different with the different-voice pairs however, as the interaction effects for 1 to 0 ($p < 0.001$), 2 to 1 ($p = 0.024$), and 4 to 3 interveners ($p = 0.011$) were significant. The predicted values from the model are plotted in Figure 26. Table 27 shows the output of the model and Table 28 below shows the effect sizes of these comparisons.

Table 27: Experiment 4 target RT model

	Log$_2$-transformed RT (targets)		
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.68	9.66 – 9.71	**<0.001**
VOICE			
Diff.	0.03	0.01 – 0.04	**<0.001**
DISTANCE			
1-0 interveners	0.17	0.15 – 0.19	**<0.001**
2-1 interveners	0.09	0.07 – 0.11	**<0.001**
3-2 interveners	0.01	-0.01 – 0.03	0.368
4-3 interveners	0.02	-0.00 – 0.04	0.083
5-4 interveners	0.01	-0.01 – 0.03	0.567
7-5 interveners	-0.01	-0.03 – 0.01	0.186
10-7 interveners	0.00	-0.02 – 0.02	0.831
SPEAKER			
MA3	-0.00	-0.02 – 0.01	0.651
Item (z-scored)			
Frequency	-0.01	-0.01 – 0.00	0.157
PNH	0.00	-0.01 – 0.01	0.962
AoA	0.02	0.01 – 0.02	**<0.001**
Duration	0.10	0.09 – 0.11	**<0.001**
Trial	-0.01	-0.02 – -0.01	**<0.001**
Previous RTLog	0.05	0.05 – 0.06	**<0.001**
Participant			
zAge	0.02	-0.00 – 0.05	0.082
Male	-0.01	-0.06 – 0.04	0.749
Group			
2	0.07	-0.00 – 0.13	0.053
3	0.01	-0.06 – 0.07	0.813
4	0.08	0.01 – 0.15	**0.024**
VOICE (Diff.) × DIST. (1-0)	-0.07	-0.10 – -0.04	**<0.001**
VOICE (Diff.) × DIST. (2-1)	-0.03	-0.06 – -0.00	**0.024**
VOICE (Diff.) × DIST. (3-2)	0.02	-0.00 – 0.05	0.103
VOICE (Diff.) × DIST. (4-3)	-0.04	-0.07 – -0.01	**0.011**
VOICE (Diff.) × DIST. (5-4)	0.01	-0.02 – 0.04	0.433
VOICE (Diff.) × DIST. (7-5)	-0.00	-0.03 – 0.03	0.836
VOICE (Diff.) × DIST. (10-7)	0.02	-0.01 – 0.05	0.230
Random Effects			
σ^2	0.04		
τ_{00} Item	0.01		
τ_{00} Participant	0.01		
Observations	12343		
Marginal R^2 / Conditional R^2	0.281 / 0.517		

Table 28: Experiment 4 target RT model effect sizes

Condition	Percentage	Effect size (ms)
Diff-Voice	*1.9*	*10*
1-0	*12.9*	*72*
2-1	*6.2*	*35*
3-2	*0.7*	*4*
4-3	*1.3*	*7*
5-4	*0.4*	*2*
7-5	*1.0*	*5*
10	*0.2*	*1*

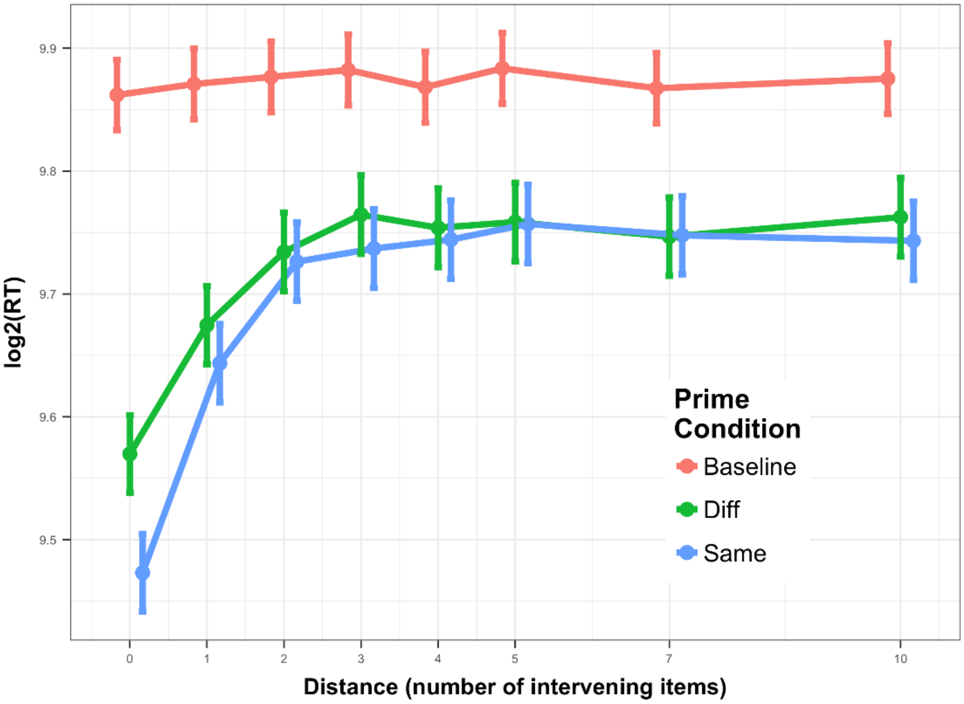

Figure 26: Experiment 4 target RT model predicted values

We now turn to additional models separated by the DISTANCE condition to interpret these interaction effects. These models were the same as the experimental model, except for the removal of the DISTANCE factor from the variables of experimental interest. The random effects structure was set at intercepts for by-subjects and by-items only. Only in the models of immediate distance ($\beta = 0.114$, $p < 0.001$, 8.4% / 47ms) and when one ($\beta =$

0.036, $p = 0.005$, 2.5% / 14ms) and three items intervened ($\beta = 0.031$, $p = 0.01$, 2.1% / 12ms) was there any effect of VOICE. In each of these cases, the different-voice targets were recognized slower than the same-voice targets. In none of the other models split by DISTANCE were main or interaction effects of VOICE significant. All models had an equivalent amount of data removed due to model criticism step (0: 40, 2.40%; 1: 23, 1.44%; 2: 34, 2.21%; 3: 30, 1.95%; 4: 30, 2.00%; 5: 22, 1.42%; 7: 39, 2.45%; 10: 28, 1.80%).

Again, we used these models to test for an additional impact of raw time over the interference due to intervening items. The chi-squared test compared models with and without an additional factor of z-scored, $\log_2$-transformed time between the end of the prime and the beginning of the target. In every model, the chi-squared test significantly reported a better model fit with the measure of raw time ($p < 0.001$ in each case). Summarizing the models with this predictor, all main effects of raw time were significant ($p < 0.001$), with effect sizes of $\beta = 0.062$, 0.070, 0.036, 0.042, 0.047, 0.055, 0.034, and 0.025. A plot of the trimmed data illustrating these findings is seen below.

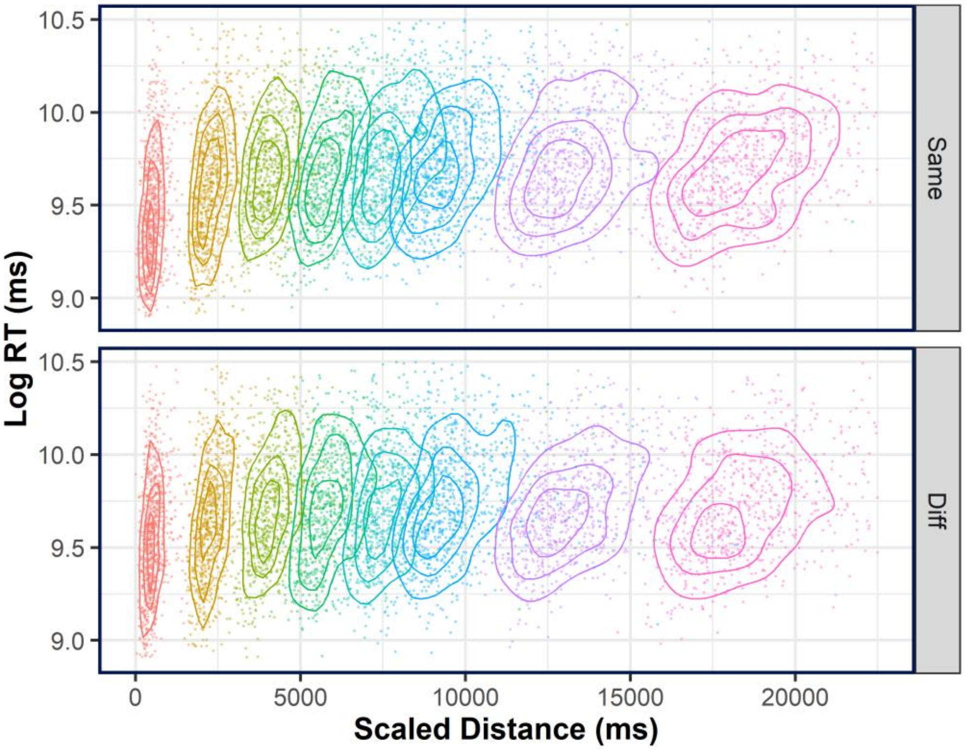

Figure 27: Experiment 4 effects of time vs. interveners

3.5.3 Discussion

The analysis of Experiment 4 indicates the following:

- **Abstract repetition priming**
    - *Accuracy:*
        - Numeric trends and some significant comparisons indicate targets recognized more accurately than primes, but no broad patterns
    - *Reaction time:*
        - All repeated targets in all distances recognized faster than primes
- **Talker-specific effects**
    - *Accuracy:*
        - No accuracy differences between same- and different-voice targets
    - *Reaction time:*

        - Separate models indicate different-voice targets recognized slower than same-voice ones at 0, 1, and 3 items intervened
        - Different short-term slopes between same- and different-voice targets
- Increasing raw time between prime and target consistently slows reaction time in all distance condition

These models indicate that voice-switch affects implicit priming as measured by the lexical decision task for up to around distances of three intervening items. After which, same- and different voice pairs are recognized equally as fast. Abstract, repetition priming persists strongly throughout the distances measured in this experiment though, indicating different priming decay rates for TSEs and abstract effects of words. Additionally, an increase of raw time on top of the slow down introduced by interveners slowed reaction time, as predicted by McKone (1998).

This pattern fits nicely with the findings in the literature. No TSEs were again found at a distance of ten intervening items. Instead, TSEs were found until around three items intervene. This concrete decay pattern of TSEs, along with the different slopes found between same- and different-voices in the target RT model suggests that TSEs affect the short-term source of repetition priming. Lastly, as discussed in Section 3.5.1.3 we mentioned the fact that this experiment had a repetition rate of 50%; which may have over-emphasized the presence of TSEs given the results of Hanique et al. (2013). If that is the case, then TSEs lasting until three intervening items is an upper limit, with the real decay pattern ending sooner.

3.6 Conclusion

In this chapter, we investigated the time-course of TSEs found in implicit tasks. Researchers who investigated a similar topic using the perceptual identification task have for the most part found long-lasting TSEs comparing same- to different-voice (most often by switching between male and female tokens). On the other hand, these long-lasting effects have not been found using the lexical decision task, leading to conflicting conclusions from the literature. Our experiments using the lexical decision task found that TSEs due to voice-switch (male to male tokens) persist until around three items intervene between prime and target. This finding of a discrete time-course fits nicely with the results of Orfanidou et al. (2006, 2011), who find no long-lasting TSEs due to voice.

One difference however comes from the results reported in Experiment 3. In that experiment, we similarly manipulated voice switch but instead of comparing same-voice to different-voice pairs, we compared different-voice, same-gender to different-voice, different-gender pairs. We unexpectedly found that different-voice, different-gender pairs were recognized more accurately and faster than different-voice, same-gender pairs at the longest distance tested; when ten items intervened between prime and target. This points to an asymmetry between perceptually similar and distinct stimuli. Previous results from Goldinger (1996) and Sheffert (1998) have found that increased similarity leads to increased accuracy in responses. Our results stand in contrast to theirs, as presumably the more perceptually similar male voice pairs were recognized slower and less accurately than pairs switching between male and female speakers.

In the next chapter, we turn to tasks tapping into explicit memory; namely the blocked and continuous word recognition tasks. Previous research has noted many differences between the tasks in whether TSEs should be found. Setting up a simple comparison between the two sheds light on the different findings presented in this chapter.

CHAPTER 4: Time-Course in Explicit Tasks

4.1 Introduction

In this chapter, we shift focus to look at the reported TSEs shown by tasks tapping into explicit memory. Recalling the conclusion of the previous chapter, TSEs persist up to distances of around three intervening items between prime and target while speaker-invariant abstract priming persists much longer in implicit lexical decision tasks. We now address why different results have been found in explicit memory tasks. These differences could stem from the possibility that representations of speech are separated into implicit and explicit types (Graf & Schacter, 1985), and therefore we are not comparing the same objects between the studies, or they could be due to other differences in the tasks, for example changing the dependent variable from reaction time to accuracy. This section begins with an overview of the literature advocating a separation between implicit and explicit memory systems before turning to a broad description of the types of tasks which have been used to investigate these issues. The conclusion of this section sets the stage for the presentation of two additional experiments to investigate the presence of TSEs in explicit tasks.

Following the introduction in Section 1.1.4, for a task to be labelled *explicit,* it must include instructions to access information about a specific, remembered experience. This is in contrast to *implicit* tasks, whose results do not depend on participants' awareness of the previous processing of an item (Graf & Schacter, 1985). There are multiple reasons to separate the memory systems accessed by explicit and implicit tasks. The fact that amnesiacs, who are unable to explicitly recall recent information, still exhibit the same implicit priming effects as typical controls indicates that explicit recall and implicit priming are separate. Investigating levels of processing, decay patterns, and modality switches also illuminates differences between the two. Generally, the assumption of memory researchers has been that explicit priming on recall accuracy is increased by a deeper level of processing of the prime items, whereas the level of processing in implicit tasks appears to be irrelevant. TSEs are also typically expected to be found longer in implicit than explicit tasks.

With these differences in mind, we now ask how relevant they are to the goal of this thesis; the investigation of the mental representations of speech. First and foremost, early work suggested that TSEs should only be found in implicit tasks, as reviewed in Section 4.2.1. Multiple influential studies were then conducted on the memory of written and spoken words using explicit tasks. Sections 4.2.2 and 4.2.3 review these results from two different paradigms, the blocked and continuous word recognition paradigms respectively. Overall, these studies showed that TSEs do arise in explicit tasks, contrary to earlier work. This finding led researchers to propose a unitary model of memory built around episodic details and exemplars of words. Without these results, support for a mental lexicon based on episodic memory traces would be quite weakened.

Additionally, an investigation of the differences between implicit and explicit memory is crucial to determining the relative effects of each type of memory. In implicit memory studies, like the ones presented in CHAPTER 3: Time-Course in Implicit Tasks using the lexical decision and shadowing tasks, it is difficult to operationally rule out the presence of explicit memory effects in the results. Perhaps the studies finding long-distance implicit priming in the perceptual identification studies were actually testing explicit recall for words after all. This point calls for studies using controlled sets of items and consistent analysis methods to investigate the nature of task effects in word recognition. We present two such studies at the end of this chapter.

As a final note, we emphasize the fact that this dissertation is not concerned solely with delineating implicit from explicit memory effects. Just as we do not necessarily care about the capacity of participants to judge whether a sound is a word or not or how fast people can mimic words, we similarly are not interested in whether participants have heard a word before. These tasks are useful inasmuch as they inform us about the representations of speech. The concepts of implicit and explicit memory have been well-discussed in the literature, and therefore we use them here to begin our investigation into speech representations.

4.2 Background

Many studies of word recognition have been conducted using both implicit and explicit memory tasks. Interestingly, effects of talker-specific details have been found with both. According to previous research however, there are differences between these tasks in what details are seen, how long they persist, and what influences their presence. The goal of this background review is to set up these differences in expectations. By doing so, we are better able to formulate the reasoning behind the models of speech recognition in the literature. This also provides us with the background to discuss the experimental results found in this thesis.

4.2.1 Motivating Implicit vs. Explicit Memory

We first discuss the reasons to make a distinction between implicit and explicit tests of memory. Perhaps the most convincing reason comes from studies of amnesiacs and other populations with neural lesions. As Graf & Schacter (1985) state, the inability to complete explicit recall tasks is a diagnostic property of anterograde amnesiacs. They conducted two visual lexical tasks, an implicit word completion task and an explicit cued recall task, comparing amnesiacs' performance with that of matched controls. While the group with amnesia predictably performed quite low on the explicit task, all groups performed similarly on the implicit word completion task. This surprising result among others (see Schacter (1987) for a review of other seminal findings) clearly points to a disassociation between the two systems of memory.

Additionally, it has been suggested that differences in the level of processing of words affect explicit tasks but not implicit ones. As we saw in the previous chapter, this understanding is far from conclusive even for implicit tasks, with McLennan & Luce (2005) and Goldinger (1996) for example disagreeing about whether increased processing time leads to the emergence of TSEs. Discussing possible sources of explicit memory effects, Jacoby & Dallas (1981) built on the idea that two factors contribute to the overall effects. The first is the general perceptual saliency of an item (the more salient, the more easily recognized) and the second is the amount of attention paid to the item. In a multi-experiment paper, they presented words either with a semantic encoding task (responding to the meaning of the word) or a perceptual encoding task (responding to the spelling of a word). They consistently find that this manipulation impacts explicit recognition tasks such that the semantic encoding condition produces higher recall than the perceptual encoding condition. This was not the case for an implicit task of word identification. Graf & Mandler (1984) conducted a similar set of studies comparing explicit cued recall and word recognition tasks against the implicit task of word completion. They found comparable results that explicit tasks and not implicit ones are affected by the level of processing of the primes.

Modality switches, where the presentation (either auditory or visual) of the studied word is different from the tested word, are also cited to show effects arising in explicit but not implicit studies. In one condition of their experiment, Jackson & Morton (1984) presented a block of visual words as primes and then had participants perform an auditory word identification task. No priming resulted in this implicit test. Similar results were obtained by Scarborough et al. (1979), who found that only within-modality presentation of primes resulted in implicit priming using the lexical decision task. Schacter & Graf (1989) compared both implicit and explicit tasks crossing auditory and visual primes with visual targets and found robust cross-modal priming only in the explicit cued-recall task. In the implicit word completion task, they found within-modality priming but severely reduced cross-modal priming.

Additionally, the relative importance of token-specific details is said to differ, with increased perceptual similarity only causing priming effects in implicit and not in explicit tasks. Using visual tests of implicit memory, Roediger & Blaxton (1987) and Jacoby & Hayman (1987) showed that keeping the presentation of items similar (i.e., both in lower-case) between study and test resulted in greater priming than when any changes between the two were present. No effect of gradation was found; that is, an increased difference between study and test did not result in worse priming effects. This is interesting for our purposes as this is the visual analogue between the perceptual similarity investigations of Goldinger (1996) and Sheffert (1998). Notably, these effects persisted in the Roediger & Blaxton (1987) study for distances of around a week between study and test blocks, as did the effects in Goldinger (1996).

Decay profile differences, the focus of this chapter, have also been found whereby TSEs in explicit tasks decay quicker than those in implicit tasks. Musen & Treisman (1990) tested visual pattern memory using both tasks and found stable implicit priming after one week while priming in the explicit memory task dropped off considerably in the same time frame. The studies presented in Goldinger (1996), which will be discussed at length in the next section, also provide evidence for a decay profile difference in that TSEs persist up to a week in implicit tasks but not after one day in explicit tasks. Lastly, in an interesting study comparing the interactions between these tasks, Wagner et al. (2000) tested the hypothesis that *previous implicit priming* of an item will decrease *subsequent explicit memory* for the same item. They tested words a day after the first presentation with an implicit classification task and two days after that with an explicit word recognition task. Both behavioral and neural evidence supported their hypothesis; increased implicit priming for an item decreased subsequent explicit recall memory of the same item.

Taken together, these findings clearly motivate a distinction between the memory systems tested in implicit and explicit tasks. To model this, Schacter (1987, 1990) proposes multiple separate but interacting memory systems. A pre-semantic perceptual memory system (discussed in CHAPTER 3: Time-Course in Implicit Tasks) models the findings from implicit memory tasks, while semantic and episodic memory systems account for more conceptually-driven results. This model of memory is interesting for our purposes, as it straightforwardly predicts that two separate representations of speech exist. Upon hearing a word, an implicit representation is created which is long-lasting (up to a week) and contains some sort of modality- and/or form-specific information. An explicit representation is also created depending on how much attention was paid to the word upon perception. This representation is available for conscious awareness but decays over a shorter time window, probably up to one day.

If true, an important question is then raised: how are we able to actually make a hard division between tasks testing implicit memory and those testing explicit memory? To put it another way, how are we able to prevent the explicit memory of a word from affecting recall in an implicit task? The fact that we are not testing explicit memory in these tasks, and therefore cannot measure its effects, indicates that explicit representations may be responsible for some part of the effects normally attributed to implicit tasks. As Schacter & Church (1990: 926) note, *"Because performance on nominally implicit tasks can often be contaminated by explicit retrieval, it is critical to provide evidence for implicit and explicit dissociation to make theoretical inferences about the nature of priming."* Masked priming, where a prime is presented for such a short window that conscious awareness of it does not occur, has been used to mitigate these concerns (for visual, see Forster & Davis, 1984; for auditory, Dupoux et al., 2003). The experiments presented in this thesis however are not masked and are therefore susceptible to this possibility. For this reason, it is important to look into the results of tasks tapping explicit memory. Comparing these results with those cited in the previous chapter using implicit tasks should help to distinguish between effects of potentially different implicit and explicit representations.

4.2.2 Blocked Word Recognition

As discussed in the previous section, the research motivating the distinction between implicit and explicit memory tasks concluded that implicit memory is where TSEs should be reliably found. The effects in these tasks were assumed to last for a long time and be insensitive to distinctions between levels of processing. Understanding these assumptions of the field from the past puts the current developments into a better perspective. Specifically, when this pattern of effects was also found in explicit studies (Palmeri et al., 1993; Goldinger, 1996; Sheffert, 1998; *inter alia*), a unitary memory hypothesis was proposed bringing implicit and explicit memory systems together under an episodic memory system (Goldinger, 1998). This episodic memory system is built from persistent traces of perceptual events stored with token-specific details. In this system, words are recognized by comparing incoming speech signals to this cloud of episodic traces, with the most similar trace being the most activated. Word recognition hinges on perceptual similarity comparisons, so the fact that token-specific detail affects the results in implicit and explicit tasks is straightforwardly predicted.

We now move on to the results which led to this proposal. This chapter, being concerned with explicit recognition tasks, presents studies that ask participants to respond as to whether they recall having heard a word before, thereby asking them to access their past memories. Additionally, as we are concerned with decay properties of speech information, we will make a distinction between *blocked word recognition* and *continuous word recognition* tasks. Different from continuous word recognition tasks, which present primes and targets with the same task (and are discussed in the next section), blocked word recognition tasks present two blocks of items: the study and the test block. The words in the study block are presented by asking participants to classify something about the word (e.g., abstract properties like animacy and concreteness or perceptual properties like enunciation and phonemes), which we term the *encoding* task. As much research in this field has been concerned with levels of processing manipulations, differences in encoding tasks tend to define these studies. Also, they typically involve long-distance manipulations due to delays between study and test blocks. This, along with randomly presenting items in each block, creates a variable amount of distance between each individual prime and target pair.

The most well-known blocked word recognition study is found in Goldinger (1996), which we will discuss at length here. Broadly, his results are cited as finding voice effects using both implicit (word identification task in noise) and explicit (blocked word recognition) tasks. In two ambitious experimental designs, he crossed experimental task, number of voices producing the stimuli (either 2, 6, or 10; 50% male each), time delay (5 minutes, 1 day, and 1 week), and the level of processing of the primes (gender, phonemic, and syntactic). In post-hoc tests, he also compared the relative effects of switching genders between prime and target as well as the perceptual similarity between any two given voices. To do so, he first conducted a paired word recognition experiment which contained all relevant prime and target voice combinations. By comparing the reaction time of recognition and hypothesizing that increased similarity would lead to increased reaction time, he calculated a two-dimensional similarity matrix of all possible voice pairings used in the experiment.

First, we discuss the results of the implicit tasks of the two experiments. The general conclusion was that TSEs were found for very long distances; up to at least one week intervening between prime and target. In the first implicit task, 180 participants were run, with 20 participants in each of nine conditions. These nine conditions were formed by a between-subjects crossing of the number of voices (two, six, or ten) and delay (5 minutes, 1 day, or 1 week). The task involved identifying two blocks of 300 words each, all presented in background noise. With such a complicated design, the large number of statistical tests in the experiment make the results difficult to interpret. However, the main results, as summarized by Goldinger, are that a same-voice advantage in response accuracy persisted up to even the 1 week delay condition and that no effect of number of voices was found. Interpreting the post-hoc correlational tests using the voice similarity matrix described above, he also reported that participants were sensitive to the perceptual similarity between voices, with similar voices being more accurate, at up to a day between study and test blocks. Similar results were found in the implicit version of the second experiment, which manipulated the level of processing of the primes. Increased processing of the primes led to improved responses to the targets, although only significantly so in the reaction time data.

In the explicit tasks of the two experiments, the results showed similar effects up to one day. In the one week delay condition, no TSEs were found, as expected from the literature in the previous section. The first experiment had another 180 participants, with the same nine conditions as the implicit task. A similar study block was used, with participants identifying words in noise, however only 150 items were presented in this block. The test block consisted of 300 items (50% from the study block) and participants had to respond whether they recalled hearing them before. Goldinger summarized the findings such that TSEs were present in the five minute and one day delay conditions. In the second examination, the same three levels of processing were implemented in presenting the study block. At deeper levels of processing (i.e., the syntactic classification), more abstract repetition priming resulted whereas the TSEs were reduced. This pattern held primarily in the explicit task, however we note that in the lowest level of processing (the gender classification), the hit rates of both same- and different-voice hovered around the chance level of 50%. Broad correlations of perceptual differences again significantly indicated that the more distinct voices were, the less priming resulted in the different-voice conditions. The fact that a deeper processing of study items increases explicit recall of test items replicated the findings from the explicit study in Schacter & Church (1992), who also contrasted implicit and explicit tasks with two different level of processing manipulations: category and pitch categorization. In the explicit task only, the category categorization increased recall of test items compared to the pitch categorization.

Two different hypotheses of the memory systems underlying these results have been put forward: an episodic memory system and a separation of perceptual and semantic systems. Starting with the former, Goldinger (1998) built on the previous results above to put forward an episodic model. He tested an episodic model of speech recognition called

MINERVA2 (Hintzman, 1988) against responses in a word shadowing task. The conclusion was that the model accurately predicted the response patterns of participants. As Goldinger (1998: 254) states, *"If episodic traces of words persist in memory and affect later perception, might they constitute the mental lexicon?"* The interpretation of Schacter & Church (1992) however centered on disassociating abstract priming effects from TSEs. Citing support from lesion studies, they advocated separate representations of each, located in different hemispheres in the brain.

The results of Karayianni & Gardiner (2003), a more recent study elaborating on level of processing effects, added additional data. Specifically, they examined the relationship between conscious *remembering* and subconscious *knowing*. By that, they asked participants in an explicit old/new recognition task to indicate how they knew a stimulus was previously presented. The additional response of *'remember'* indicated a conscious, episodic recollection of having heard the word before whereas the response *'know'* indicated a subconscious feeling that the word was previously presented. In doing so, they attempted to separate the implicit and explicit types of word activation highlighted in the previous section. They found that TSEs *reported as conscious recollections* decreased as the encoding task increased in difficulty, indicating that implicit memory may be underlying some of the results.

Goh (2005) adds another important contribution to this discussion. In explicit studies, the repeated presentation of words from one voice should strengthen an abstract representation of that speaker, apart from the words. This is hinted at by the voice learning studies in Nygaard & Pisoni (1998), who showed faster reaction time to well-studied voices in lexical decision experiments (compared with novel voices). To truly compare same- and different-voice conditions in recognition tasks, a new-voice condition should be added where words are presented in a voice completely new to the participant. The comparison between different-voice and new-voice conditions should indicate the effect of voice learning throughout these studies. To that end, Goh (2005) conducted an explicit recognition memory task using ten different male speakers (without the confounding influence of switching gender). A signal detection theory analysis only supported conclusions that the same-voice condition was recognized faster than the other conditions, although an analysis of response biases and hit rates showed that the different-voice condition was recognized more accurately than the new-voice condition. Lastly, to highlight the contradictory findings in this literature, an analysis of perceptual similarity between voices found no evidence that similar sounding voices were recognized better.

The study reported in Papesh et al. (2012) extends these results by presenting a similar paradigm while measuring the pupil dilation of the participants. Without getting too far afield, pupil dilation has been shown to index memory activity, with increased dilation indicating greater effort spent in encoding memories (Võ et al., 2008). Papesh et al. (2012) conduct an explicit old/new recognition task with same-, different-, and new-voice repetition conditions, similar to Goh (2005). Behaviorally, same-voice repetition conditions predictably were recognized with greater accuracy. Unlike Goh (2005) however, they found no statistically significant results of familiar vs. unfamiliar voices (although the numerical trend was in the expected pattern). The pupillometry results indicated that pupil dilation size for items in the study block correlated with explicit memory recall, where the more dilated the pupil was at study, the better memory for that item existed at test. Only considering the test block, the pupil diameter additionally correlated with recognition accuracy, with a greater diameter signifying a correct recognition response. Their results are one of the first to link behavioral and physiological indices together but overall support similar conclusions as to the nature of TSEs in word recognition.

The last study discussed here is found in Brown (2011) and Brown & Gaskell (2014), who tested whether a combined episodic memory system or separate abstract and episodic systems better explain talker-specific and lexical competition effects. The main contribution of this work for our purposes is that it attempts to test these effects using the same materials, where other studies have used vastly different designs. To do so, they introduced an artificial lexicon that they then taught to participants. Focusing on the recognition studies, they tested the recall rates of same- and different-voice primes on the same targets, an important design element which we implement in our studies in this chapter. TSEs were seen both the next day after study and surprisingly one week later as well, longer than the explicit results from Goldinger (1996). Follow-up studies showed that this effect persisted even if the novel non-words were produced with multiple voices, although there was more decay in this than the single-talker training version.

Overall, the studies presented in this section all seem to point to the same fact: TSEs can persist quite long in explicit tasks. As mentioned at the beginning, this result is the primary reason that researchers have proposed a single, combined representational account of word recognition built from episodic traces of speech. One thing to continue to keep in mind throughout this background section is the differences between these and the results found in the implicit tasks presented in the previous chapter. Long-distance TSEs were thought to be the hallmark of implicit tasks, and yet studies using indirect measures like reaction time (including our own experiments) do not show long-distance effects of voice. Only the earlier studies examining accuracy in perceptual identification tasks and the studies presented here using the word recognition task appear to show TSEs. These differences need to be resolved before a unified model of speech perception can be built, especially with the observation in the previous section that both implicit and explicit memory may be operating within the same task.

4.2.3 Continuous Word Recognition

In this section, we describe the other main set of recognition studies used to examine TSEs. These studies use what we term the continuous word recognition task, which presents primes and targets together in the same block. The distance between the two is manipulated to create varying lags which are then used to investigate the decay profile of TSES along with abstract repetition priming.

We begin with the classic study of Craik & Kirsner (1974), who found consistent talker-specific effects of switching genders. The tested lags of 0, 1, 3, 7, 15, and 31 intervening items between prime and target and found significant TSEs using both accuracy and reaction time measures, along with a general decline in performance over increasing distances. As the interaction did not come out significant, they concluded that the observed same-voice advantage did not decay over the distances tested. In their final auditory experiment, they added an additional manipulation asking participants to indicate which speaker (male or female) had previously spoken words which were marked as *'old'*. They found that participants could significantly do so, although their accuracy did decline to around 65% by the time one to three items intervened. This result hints at the fact that participants' episodic memories may not have been driving the responses. As Karayianni & Gardiner (2003) discussed, increasing distances resulted in participants not relying as much on explicit recall but more on general familiarity, as reported in the previous section.

Building on these results, Palmeri et al. (1993) went further in testing lags of 0, 1, 3, 7, 15, 31, and 63 intervening items. They also varied the number of speakers in the experiments to measure the joint effect of an increasing variability in the speech signal. Surprisingly, they found no effects of increasing the number of talkers from 2 to 20, but did find TSEs in both accuracy and response time at all distances, replicating Craik & Kirsner's (1974) results. Interaction effects were found in this study however, as increasing the distance between prime and target did result in proportionally smaller TSEs. Different-voice repetitions were also statistically significant when comparing between same-gender and different-gender switches, counter to the results of Geiselman & Bellezza (1976, 1977) who found only gender-switch effects in sentence recognition tasks. In their second experiment, Palmeri et al. (1993) added an additional voice discrimination task, which showed participants were able to recognize whether the voice speaking the word was the same or different. This effect was most pronounced early on but decreased (while still remaining significant) after a few intervening items. The results for the other comparisons replicated their first experiment, except for the fact that different-gender words were actually recognized *better* than same-gender, different-voice words; a finding which contradicts the hypothesis that increased perceptual similarity improves recognition accuracy.

The studies reported in Bradlow et al. (1999) provide further examples of a controlled investigation of the time-course of TSEs. They used the continuous word recognition task along with the perceptual identification modification that the two previously reported studies used, resulting in *old-different*, *old-same*, and *new* judgments. The results showed a linearly decreasing function of categorization accuracy between 2, 8, 16, and 32 intervening words. Between subjects, the words were presented in one of three different conditions testing different types of token-specific information; gender-switch, speech rate changes (between slow and fast productions of words), and amplitude changes (between 35 and 60 dB SPL). Only the voice and speech rate episodic information proved important for the categorization accuracy of participants, where switches resulted in significantly worse accuracy. The null results looking at the amplitude of words indicates that this information is not used in word recognition, which is an interesting albeit unsurprising finding. It does pose problems for accounts of word recognition based on pure perceptual similarity however. Recall the results of Sheffert (1998) in Section 3.2.2 who only found TSEs when primes and targets were similarly presented with background noise. Presumably, background noise is also not something we would expect to matter for word recognition. The contrast between the null results for amplitude in Bradlow et al. (1999) and the interpretation that background noise mattered for similarity from Sheffert (1998) questions the general importance of perceptual similarity in word recognition.

Next, the studies presented in Nygaard et al. (2000) investigate a similar hypothesis, namely that the typicality of a token affects whether TSEs are seen. They conducted multiple continuous word recognition studies using lags of 1, 7, 15, and 31 intervening items. The tested speech rate (slow, normal, and fast), vocal effort (soft and loud), and various measures of amplitude (including normalization and rescaling). Their interpretation was based on the finding that, for example, while TSEs were seen when slow words were preceded by fast words, the reverse was not the case. Results like these were found across their conditions, leading to the conclusion that token-specific information impacts speech perception, but only for productions of words that do not straightforwardly match the prototypical production. For these types of words, only abstract information is represented; which explains the lack of TSEs.

The last study we consider in this section is found in Campeanu et al. (2014), who looked at the effects of gender and accent. Their continuous word recognition study used lags of 1, 7, and 15 intervening items and contrasted same-voice pairs with three types of different-voice pairs: a different-gender speaker, a speaker with a different accent, or both. At all lags, they found that the same-voice condition was recognized more accurately and faster than the different-voice conditions. No comparisons between the various different-voice conditions were significant however, indicating no additional effects of gender or accent on top of those seen with voice-switches. They additionally presented ERP results which echoed the behavioral results. We interpret these results as negative support for a broad perceptual similarity account of word recognition. If perceptual similarity drove speech recognition, then the comparisons between the different-voice conditions would have reflected so.

The general conclusion of the studies presented in this section is again that same-voice pairs are recognized more accurately than different-voice pairs. Normally, this effect is found to decrease over increasing time intervals, but not always. Given the results from the blocked word recognition studies, we would expect a same-voice advantage to be present at all distances, but since these studies test much shorter lags, there is the possibility that a different priming pattern exists at early distances. This would mirror the dual-route source of priming effects hypothesized by McKone (1995), discussed in Section 3.2.4. Finally, the results from this section provide some overall skepticism for the account of perceptual similarity underlying word recognition. Multiple attempts at finding results supporting the importance of perceptual similarity have instead found the opposite.

4.2.4 Summary

In conclusion, this section has broadly discussed the effect of explicit memory tasks on word recognition studies. We first motivated the distinction between implicit and explicit tasks, which serves to be a very important distinction. Since the two tap into different memory systems, we noted that it is possible that two different types of representations are created upon perception of a word (similar to the *separate representations* discussion of Section 1.1.3). One of these would exist in explicit memory and the other in implicit memory, each with their own decay properties and sensitivity to task effects. This possibility raises the problem that implicit and explicit representations are operationally difficult to disentangle. Experiments need to be clear about which representation they are in fact testing, as explicit information may be contaminating the results of studies described as implicit tasks.

Next, we discussed research in the field using both blocked and continuous word recognition tasks. The general consensus from the former is that a same-voice advantage exists, lasting for quite a while in explicit recall. Whether this is due to actual memory recall or implicit priming effects is unknown. The decay patterns from the continuous word recognition paradigm indicate a general same-voice advantage as well. Unlike the results from the lexical decision and shadowing tasks presented in the last chapter, same-voice repetitions induce similar talker-specific effects at distances even beyond ten intervening words. General conclusions about perceptual similarity seem not to hold however; as multiple studies fail to find that increased similarity leads to better processing.

To investigate these effects, the next section presents two studies using the continuous and the blocked word recognition tasks. With similar items and analysis methods, we are able to check the decay patterns of TSEs in tasks designed to tap into explicit memory. Given the summary of the literature, we would expect to find TSEs at all tested delays, since we tested participants with delays less than one day between prime and target. Broadly speaking, these patterns should be different than the ones found in Experiment 2, Experiment 3, and Experiment 4, presented in CHAPTER 3: Time-Course in Implicit Tasks.

4.3 Experiment 5

The first study described in this chapter is an adaptation of the lexical decision task of Experiment 4 into a continuous word recognition task. Given the results presented in this chapter, we expect to find significant evidence of TSEs in the accuracy data from this experiment. This experiment tests the same discrete distances (0, 1, 2, 3, 4, 5, 7, and 10) as Experiment 4, measured in number of intervening items. Some of the studies find interaction effects such that TSEs diminish as distance increases, while others find consistently strong effects of voice at all the distances they test.

4.3.1 Method

4.3.1.1 Participants

A total of 156 (age range 18-77, mean 33.0; 86 male) participants were run at the end of the fall semester of 2017. They were recruited from the experimental platform Prolific and voluntarily completed the study online using a custom Ibex implementation of a continuous word recognition task.

4.3.1.2 Stimuli

This experiment consisted of the exact same set of 288 total stimuli from Experiment 4 (split halfway between words and non-words). Two tokens of each stimulus were recorded and used in the experiment by speakers MA1 and MA3. For a complete list of the properties of the stimuli in this experiment, see APPENDIX I: Experimental stimuli.

4.3.1.3 Design

The design of this experiment exactly mirrored that of Experiment 4. The only difference was in the response participants made to the stimuli. Instead of lexical decision responses, participants indicated whether a word had been heard before or not.

4.3.1.4 Procedure

The experiment, as presented on Prolific, was titled *Remembering Words* and had the description *“In this study, you will listen to three groups of sounds and indicate with the keyboard whether the sound you hear is new (first time in the experiment) or old (already heard in the experiment). You will also respond to a very brief demographic questionnaire. Completion time depends on your internet connection speed, as downloading the stimuli may take a while.”* Participants responded using the *‘F’* key if the item they heard was *‘old’* (i.e., previously heard in the experiment) and with the *‘J’* key if it was *‘new’*. Twelve practice responses were included before the experiment. These consisted of the following items presented per subject with an 800ms ISI (with subscript representing different speakers and bold face indicating an ‘old’ response was appropriate): *fluff*$_1$, *flame*$_2$, *nowk*$_1$, ***fluff***$_2$, *gaech*$_2$, *smell*$_2$, ***nowk***$_1$, *skrown*$_2$, ***skrown***$_2$, ***smell***$_1$, ***flame***$_1$, and *kwaanch*$_1$. These practice items were included to emphasize that repetitions could be between speakers and could be repetitions of both words and non-words. Feedback was given for these practice items, as the online nature of the experiment prevented participants from asking for clarification before beginning the experiment. Following the practice session, three blocks of 192 items each were presented (576 total recognition responses) using a custom Ibex continuous word recognition task implementation. Participants were encouraged to take breaks only at the end of a block. All experiments included in the analysis were completed on average within 24 minutes (range: 15-70 minutes). As mentioned, each participant had their own specific trial order due to the distance manipulation.

4.3.1.5 Analysis

Of the original 156 participants in the experiment, 39 participants were removed due to overall poor performance. Of these, 35 were removed due to an overall accuracy score of less than 60% correct. As this was a more difficult task than lexical decision, the overall accuracy score was lowered from 70% to 60%. An additional two subjects each were removed for having Hodges-Lehmann estimated RT distribution outliers and over 20 near-immediate responses. After this global participant removal, the overall results per each item were examined. With this experiment, no items had accuracy scores of less than 50% correct.

With a different task which is not focused on reaction time, it is not straightforward which data-trimming steps should be taken. Upon visualizing the reaction time data, the overall density closely matched that from Experiment 4 which involved the same items. For that reason, this analysis persists in globally trimming responses which took less than 300ms or greater than 3000ms. By-participant and by-item trimming of reaction time were not performed however, as the analysis of recognition task data centers on accuracy.

Table 29: Experiment 5 removal summary

	Observations	**Percentage**
Inaccurate trials	*4117*	*24.4*
RT trimming (300 > RT < 3000)	*790*	*4.7*
Total removed	*4907*	*29.1*
Total remaining	*11941*	

4.3.2 Results

The following table reports the distribution of the data per the factors VOICE, DISTANCE, and SPEAKER. These data are reported after only the minimal global trimming procedure of unreasonable reaction times was applied. Reaction time reports are central tendency measures from the non-parametric Hodges-Lehmann estimates and accuracy scores indicate the amount of correct responses out of the total, after all global participant and item removal was conducted.

Table 30: Experiment 5 data summary

DISTANCE			VOICE			SPEAKER			Total
0 Interveners	822	91.4	**Same**	808	91.7	**MA1**	828	91.4	**490**
						MA3	790	91.9	**540**
			Diff	838	91.2	**MA1**	840	92.5	**535**
						MA3	834	89.8	**488**
1 Interveners	854	93.2	**Same**	840	95.0	**MA1**	852	95.1	**510**
						MA3	828	94.9	**508**
			Diff	869	91.4	**MA1**	884	93.3	**479**
						MA3	854	89.7	**523**
2 Interveners	892	90.2	**Same**	886	91.1	**MA1**	904	90.5	**486**
						MA3	866	91.6	**513**
			Diff	898	89.3	**MA1**	908	89.3	**514**
						MA3	888	89.4	**490**
3 Interveners	922	88.7	**Same**	914	88.2	**MA1**	918	87.4	**509**
						MA3	909	89	**490**
			Diff	931	89.2	**MA1**	946	89.1	**485**
						MA3	916	89.2	**520**
4 Interveners	944	85.0	**Same**	936	85.8	**MA1**	964	85.2	**480**
						MA3	910	86.3	**510**
			Diff	954	84.2	**MA1**	966	88.1	**477**
						MA3	941	80.5	**497**
5 Interveners	983	85.5	**Same**	976	86.2	**MA1**	999	86.7	**467**
						MA3	954	85.7	**526**
			Diff	990	84.8	**MA1**	1000	86.8	**508**
						MA3	980	82.8	**494**
7 Interveners	1005	82.5	**Same**	1000	83.6	**MA1**	1012	83.7	**527**
						MA3	986	83.5	**474**
			Diff	1010	81.4	**MA1**	1038	82.5	**491**
						MA3	985	80.3	**503**
10 Interveners	1001	79.2	**Same**	986	80.2	**MA1**	991	78.9	**525**
						MA3	982	81.7	**471**
			Diff	1016	78.2	**MA1**	1047	78.7	**516**
						MA3	984	77.7	**485**
Primes	1039	87.7				**MA1**	1062	86.9	**8014**
						MA3	1016	88.4	**8021**

4.3.2.1 *Accuracy*

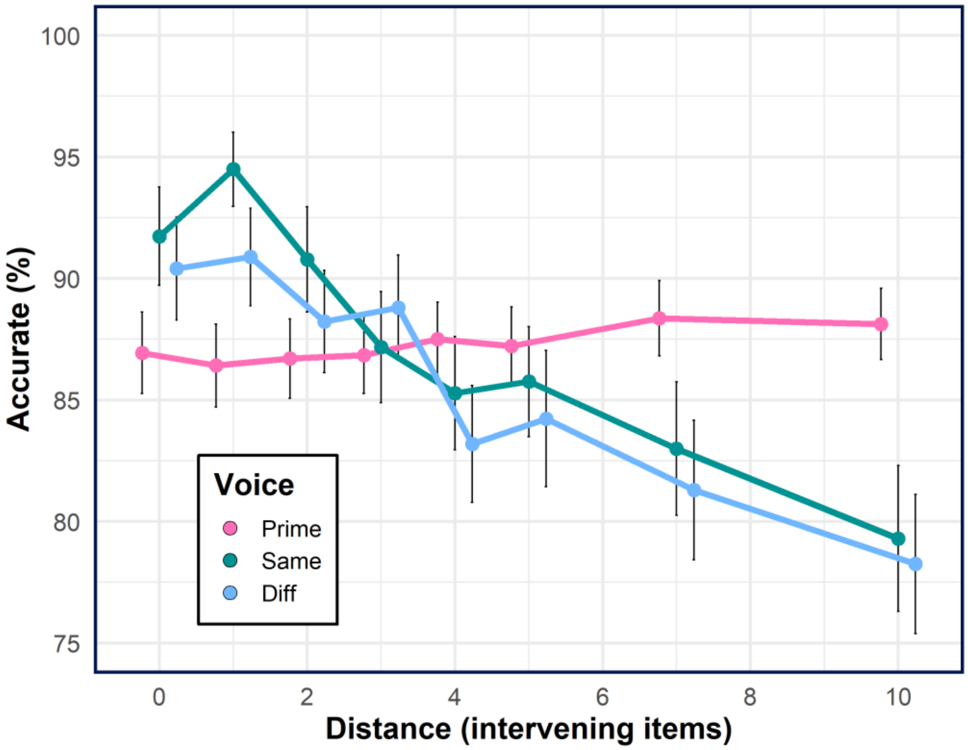

Figure 28: Experiment 5 accuracy data

A generalized linear mixed-effects model was first fit to the hit rates from the accuracy data. This model sets the primes as the reference level for the condition factor, with targets in each combination of VOICE, DISTANCE, and SPEAKER dummy-coded. Random effects were set as intercepts for participants and items. The outcome of this model is seen below.

Table 31: Experiment 5 combined accuracy model

	Accuracy (combined)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	8.11	7.07 – 9.31	**<0.001**
VOICE × DISTANCE			
Same at 0	1.58	1.14 – 2.20	**0.006**
Diff. at 0	1.93	1.38 – 2.70	**<0.001**
Same at 1	3.02	2.00 – 4.56	**<0.001**
Diff. at 1	2.11	1.46 – 3.06	**<0.001**
Same at 2	1.45	1.06 – 2.00	**0.022**
Diff. at 2	1.28	0.95 – 1.72	0.102
Same at 3	1.03	0.78 – 1.37	0.812
Diff. at 3	1.24	0.92 – 1.68	0.165
Same at 4	0.85	0.65 – 1.12	0.243
Diff. at 4	1.08	0.80 – 1.45	0.612
Same at 5	0.90	0.68 – 1.20	0.477
Diff. at 5	1.01	0.77 – 1.33	0.944
Same at 7	0.74	0.58 – 0.96	**0.022**
Diff. at 7	0.65	0.51 – 0.84	**0.001**
Same at 10	0.53	0.42 – 0.66	**<0.001**
Diff. at 10	0.52	0.41 – 0.66	**<0.001**
SPEAKER			
MA3	1.17	1.03 – 1.32	**0.014**
Item (z-scored)			
Frequency	0.99	0.94 – 1.04	0.592
PNH	1.06	1.00 – 1.11	**0.037**
AoA	0.98	0.93 – 1.03	0.460
Duration	1.03	0.97 – 1.10	0.305
Trial	0.92	0.89 – 0.95	**<0.001**
Participant			
zAge	1.00	0.89 – 1.12	0.988
Male	0.73	0.58 – 0.92	**0.009**
Group			
2	0.81	0.59 – 1.10	0.178
3	0.70	0.50 – 0.98	**0.035**
4	0.99	0.71 – 1.38	0.935
VOICE × DIST. (Same-0) × SPKR (MA3)	0.99	0.62 – 1.57	0.963
VOICE × DIST. (Diff.-0) × SPKR (MA3)	0.62	0.39 – 0.98	**0.042**
VOICE × DIST. (Same-1) × SPKR (MA3)	0.87	0.49 – 1.55	0.628
VOICE × DIST. (Diff.-1) × SPKR (MA3)	0.59	0.36 – 0.94	**0.028**
VOICE × DIST. (Same-2) × SPKR (MA3)	1.01	0.64 – 1.59	0.982
VOICE × DIST. (Diff.-2) × SPKR (MA3)	0.90	0.59 – 1.37	0.617
VOICE × DIST. (Same-3) × SPKR (MA3)	1.07	0.71 – 1.62	0.736
VOICE × DIST. (Diff.-3) × SPKR (MA3)	0.90	0.59 – 1.38	0.635
VOICE × DIST. (Same-4) × SPKR (MA3)	0.97	0.66 – 1.42	0.859
VOICE × DIST. (Diff.-4) × SPKR (MA3)	0.50	0.34 – 0.73	**<0.001**
VOICE × DIST. (Same-5) × SPKR (MA3)	0.87	0.59 – 1.28	0.473
VOICE × DIST. (Diff.-5) × SPKR (MA3)	0.60	0.41 – 0.87	**0.008**
VOICE × DIST. (Same-7) × SPKR (MA3)	0.86	0.60 – 1.24	0.416
VOICE × DIST. (Diff.-7) × SPKR (MA3)	0.81	0.57 – 1.16	0.253
VOICE × DIST. (Same-10) × SPKR (MA3)	1.09	0.77 – 1.55	0.614
VOICE × DIST. (Diff.-10) × SPKR (MA3)	0.83	0.59 – 1.15	0.262
Observations	32066		
Marginal R^2 / Conditional R^2	0.042 / 0.173		

This model shows whether any priming was seen when items were repeated by comparing each repetition condition with the accuracy of the primes. In this large model, we see that up until distances of two intervening items with the same speaker, targets were recognized significantly *more* accurately (0: $p = 0.006$, 1: $p < 0.001$, 2: $p = 0.022$). When the voice was switched between prime and target however, this effect was significant up until distances of one intervener (0: $p < 0.001$, 1: $p < 0.001$, 2: $p = 0.102$). At distances of seven and ten intervening items however, targets were recognized significantly *less* accurately than the primes (same-voice at 7: $p = 0.022$, different-voice at 7: $p = 0.001$, same-voice/different-voice at 10: $p < 0.001$). Overall, the main effect of speaker indicated that words spoken by speaker MA3 were recognized more accurately ($p = 0.014$). A few interaction effects between condition and speaker came out significant, which complicates the interpretation of the main effect of SPEAKER however.

The model examining the interaction of the experimental predictors of VOICE, DISTANCE, and SPEAKER only on target accuracy is presented in the table below. The factors VOICE and SPEAKER dummy-coded as in the full model. The contrast coding for DISTANCE however was backwards-difference coded; similar to Experiment 4, which compares each level of the DISTANCE condition with the prior level.

Table 32: Experiment 5 target accuracy model

	Accuracy (targets)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	10.06	8.26 – 12.24	**<0.001**
Voice			
Diff.	1.02	0.88 – 1.18	0.842
Distance			
1-0 interveners	2.04	1.21 – 3.44	**0.008**
2-1 interveners	0.46	0.28 – 0.78	**0.004**
3-2 interveners	0.70	0.46 – 1.06	0.095
4-3 interveners	0.84	0.57 – 1.23	0.370
5-4 interveners	1.02	0.69 – 1.50	0.922
7-5 interveners	0.82	0.57 – 1.19	0.289
10-7 interveners	0.72	0.51 – 1.00	**0.048**
Speaker			
MA3	1.10	0.90 – 1.35	0.348
Item (z-scored)			
Frequency	0.92	0.85 – 1.00	0.056
PNH	0.99	0.91 – 1.07	0.742
AoA	0.95	0.87 – 1.03	0.231
Duration	0.99	0.90 – 1.08	0.817
Trial	1.08	1.03 – 1.14	**0.002**
Participant			
zAge	0.96	0.83 – 1.12	0.599
Male	0.83	0.62 – 1.12	0.229
Group			
2	0.69	0.46 – 1.06	0.088
3	0.72	0.46 – 1.13	0.153
4	1.01	0.65 – 1.55	0.976
Voice (Diff.) × Dist. (1-0)	0.53	0.26 – 1.08	0.082
Voice (Diff.) × Dist. (2-1)	1.35	0.67 – 2.70	0.403
Voice (Diff.) × Dist. (3-2)	1.40	0.78 – 2.51	0.265
Voice (Diff.) × Dist. (4-3)	1.01	0.57 – 1.77	0.982
Voice (Diff.) × Dist. (5-4)	0.91	0.53 – 1.58	0.747
Voice (Diff.) × Dist. (7-5)	0.82	0.49 – 1.38	0.453
Voice (Diff.) × Dist. (10-7)	1.09	0.68 – 1.75	0.710
Voice (Diff.) × Speaker (MA3)	0.72	0.59 – 0.88	**0.002**
Dist. (1-0) × Speaker (MA3)	0.85	0.41 – 1.78	0.674
Dist. (2-1) × Speaker (MA3)	1.15	0.55 – 2.39	0.713
Dist. (3-2) × Speaker (MA3)	1.07	0.58 – 1.96	0.827
Dist. (4-3) × Speaker (MA3)	0.89	0.52 – 1.55	0.691
Dist. (5-4) × Speaker (MA3)	0.88	0.52 – 1.51	0.649
Dist. (7-5) × Speaker (MA3)	1.10	0.66 – 1.86	0.712
Dist. (10-7) × Speaker (MA3)	1.17	0.72 – 1.91	0.526
V. (Diff.) × D. (1-0) × S. (MA3)	1.18	0.44 – 3.14	0.741
V. (Diff.) × D. (2-1) × S. (MA3)	1.31	0.50 – 3.45	0.582
V. (Diff.) × D. (3-2) × S. (MA3)	0.92	0.40 – 2.14	0.848
V. (Diff.) × D. (4-3) × S. (MA3)	0.63	0.29 – 1.39	0.255
V. (Diff.) × D. (5-4) × S. (MA3)	1.31	0.62 – 2.77	0.478
V. (Diff.) × D. (7-5) × S. (MA3)	1.20	0.58 – 2.47	0.627
V. (Diff.) × D. (10-7) × S. (MA3)	0.85	0.43 – 1.68	0.638
Observations	16031		
Marginal R^2 / Conditional R^2	0.067 / 0.257		

In this model, we find a more discernable pattern. TSEs are not found, as indicated by the non-significant main effect of VOICE (p = 0.842). The factor DISTANCE did influence the accuracy hit rates however, with distances of 1 intervener being recognized more accurately than 0 (p = 0.008), 2 interveners being less accurate than 1 (p = 0.004) and 10 interveners being less accurate than 7 (p = 0.048). No main effects of SPEAKER are found here (p = 0.348), except for one significant interaction term. Overall, this indicates that the increase in the number of interveners between prime and target did cause discernable effects on accuracy. The VOICE manipulation did not appear to have any substantial effects.

4.3.2.2 Reaction Time

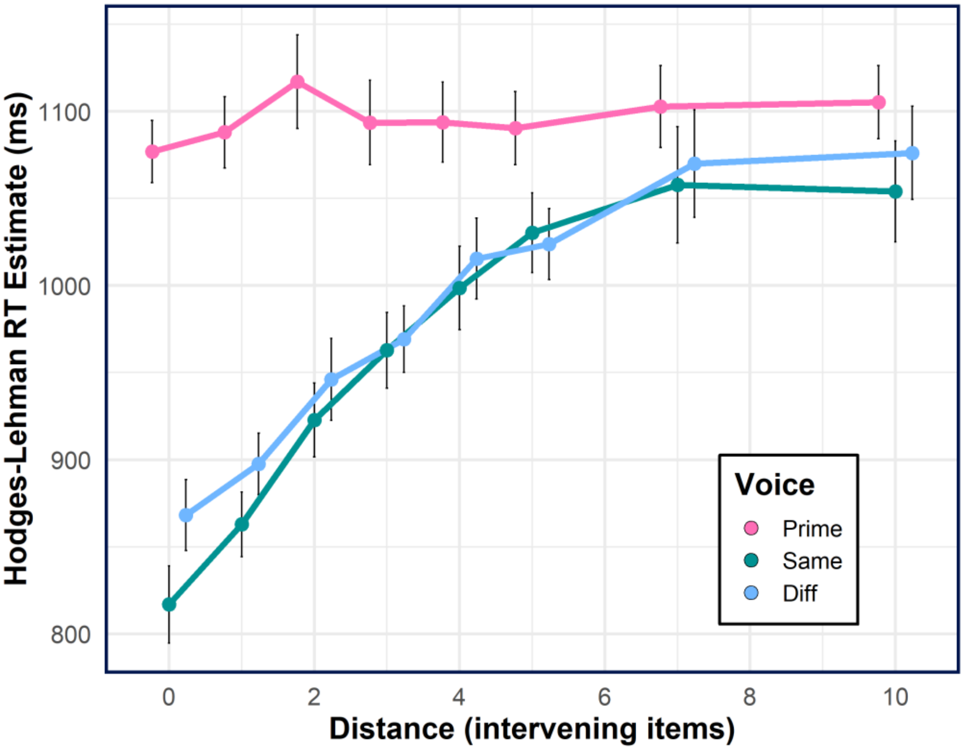

Figure 29: Experiment 5 reaction time data

The first analysis of reaction time comes from the large model investigating whether repetition priming existed in all experimental conditions. The responses to primes and targets were combined into one dataset and a linear mixed-effect model was run examining the log$_2$-transformed response time. Random effects included by-subject and by-item intercepts. Fixed effects of interest were a dummy-coded variable with the baseline being responses to the primes and factor levels indicating each of the four VOICE by DISTANCE conditions, which interacted with a dummy-coded variable indicating the SPEAKER$_{[\text{MA1 vs. MA3}]}$ of the target. Model criticism resulted in 623 additional observations being removed, a total of 2.61% of the remaining data.

For the comparisons of interest, all factor levels came out significant (all p < 0.05) except for the 10-intervening distance, different voice condition (p = 0.157). For the most part, this indicates that responding *‘old’* to an item was generally faster than responding *‘new’* to an item.

Table 33: Experiment 5 combined RT model

	Log$_2$-transformed RT (combined)		
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.99	9.91 – 10.07	**<0.001**
VOICE × DISTANCE			
Same at 0	-0.33	-0.36 – -0.30	**<0.001**
Diff. at 0	-0.30	-0.33 – -0.27	**<0.001**
Same at 1	-0.32	-0.35 – -0.29	**<0.001**
Diff. at 1	-0.23	-0.26 – -0.20	**<0.001**
Same at 2	-0.22	-0.25 – -0.19	**<0.001**
Diff. at 2	-0.20	-0.23 – -0.17	**<0.001**
Same at 3	-0.19	-0.22 – -0.16	**<0.001**
Diff. at 3	-0.15	-0.18 – -0.12	**<0.001**
Same at 4	-0.15	-0.18 – -0.11	**<0.001**
Diff. at 4	-0.13	-0.17 – -0.10	**<0.001**
Same at 5	-0.05	-0.09 – -0.02	**0.002**
Diff. at 5	-0.08	-0.11 – -0.05	**<0.001**
Same at 7	-0.05	-0.08 – -0.02	**0.001**
Diff. at 7	-0.04	-0.07 – -0.01	**0.013**
Same at 10	-0.08	-0.12 – -0.05	**<0.001**

Diff. at 10	-0.02	-0.06 – 0.01	0.157
Speaker			
MA3	-0.00	-0.02 – 0.01	0.655
Item (z-scored)			
Frequency	0.00	-0.01 – 0.01	0.733
PNH	0.00	-0.01 – 0.01	0.768
AoA	0.01	0.00 – 0.02	**0.018**
Duration	0.07	0.07 – 0.08	**<0.001**
Trial	-0.04	-0.04 – -0.03	**<0.001**
Previous RTLog			
Participant			
zAge	0.03	-0.00 – 0.06	0.091
Male	0.04	-0.20 – 0.28	0.730
Group			
2	-0.02	-0.10 – 0.07	0.653
3	0.06	-0.03 – 0.15	0.224
4	-0.01	-0.10 – 0.08	0.839
Voi × Dist. (Same-0) × Spkr (MA3)	-0.02	-0.07 – 0.02	0.282
Voi × Dist. (Diff.-0) × Spkr (MA3)	0.01	-0.04 – 0.05	0.721
Voi × Dist. (Same-1) × Spkr (MA3)	0.04	-0.01 – 0.08	0.090
Voi × Dist. (Diff.-1) × Spkr (MA3)	-0.03	-0.07 – 0.02	0.208
Voi × Dist. (Same-2) × Spkr (MA3)	-0.01	-0.05 – 0.03	0.677
Voi × Dist. (Diff.-2) × Spkr (MA3)	0.01	-0.03 – 0.06	0.611
Voi × Dist. (Same-3) × Spkr (MA3)	0.03	-0.02 – 0.07	0.261
Voi × Dist. (Diff.-3) × Spkr (MA3)	-0.02	-0.06 – 0.03	0.464
Voi × Dist. (Same-4) × Spkr (MA3)	-0.02	-0.07 – 0.02	0.350
Voi × Dist. (Diff.-4) × Spkr (MA3)	0.02	-0.03 – 0.06	0.489
Voi × Dist. (Same-5) × Spkr (MA3)	-0.08	-0.12 – -0.03	**0.001**
Voi × Dist. (Diff.-5) × Spkr (MA3)	-0.01	-0.06 – 0.03	0.568
Voi × Dist. (Same-7) × Spkr (MA3)	-0.03	-0.08 – 0.01	0.163
Voi × Dist. (Diff.-7) × Spkr (MA3)	0.00	-0.04 – 0.05	0.940
Voi × Dist. (Same-10) × Spkr (MA3)	0.02	-0.02 – 0.07	0.319
Voi × Dist. (Diff.-10) × Spkr (MA3)	-0.03	-0.07 – 0.02	0.291
Random Effects			
σ^2	0.09		
τ_{00} Item	0.00		
τ_{00} Participant	0.03		
Observations	23259		
Marginal R^2 / Conditional R^2	0.151 / 0.363		

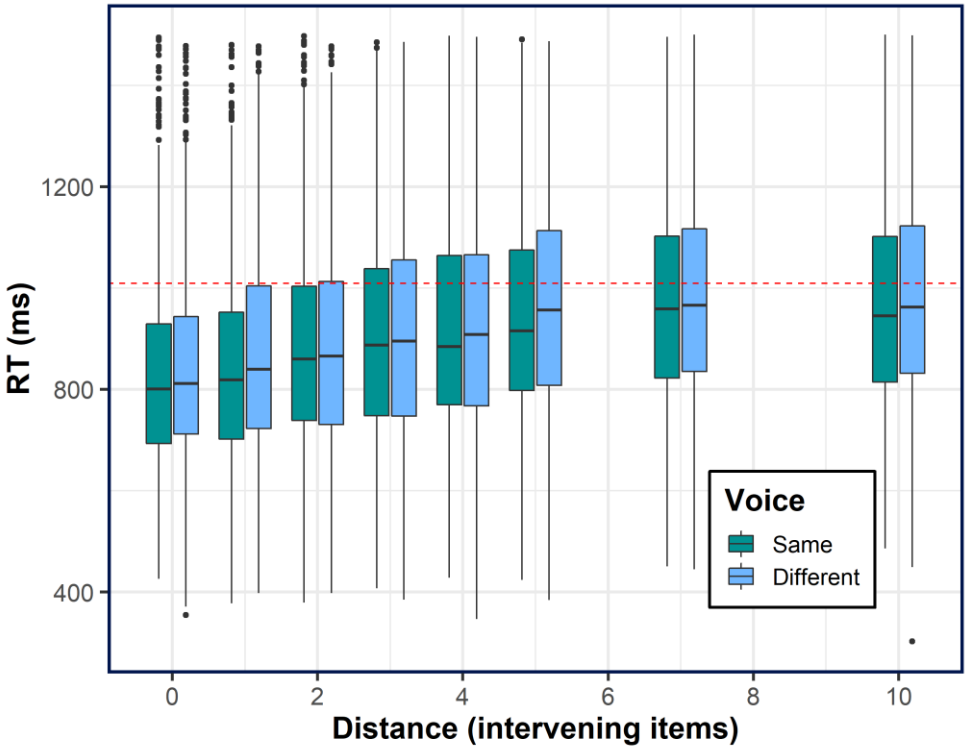

Figure 30: Experiment 5 trimmed RT data

Next, we examine the model testing the differences between same- and different-voice conditions at immediate and long distances. The random effects for this model consisted of by-subject and by-item intercepts only. The fixed effects of interest were formed by the interactions between two terms: the dummy-coded factors of VOICE (baseline = same-voice) and the backwards-difference coded factor of DISTANCE. Again, backwards-difference coding compares each level of a factor to the level previous to it. The dummy-coded factor of SPEAKER (baseline = speaker MA1) was treated only as fixed effect without participating in any interactions. After fitting the model, 327 additional values (2.74% of the remaining data) with residuals > 2.5 SDs from the mean were removed.

Table 34: Experiment 5 target RT model

	Log$_2$-transformed RT (targets)		
Predictors	*Estimates*	*CI*	*p-values*
Intercept	9.82	9.79 – 9.85	**<0.001**
VOICE			
Diff.	0.03	0.02 – 0.04	**<0.001**
DISTANCE			
1-0 interveners	0.04	0.02 – 0.07	**0.002**
2-1 interveners	0.22	0.19 – 0.25	**<0.001**
3-2 interveners	0.04	0.01 – 0.07	**0.012**
4-3 interveners	0.03	-0.00 – 0.06	0.077
5-4 interveners	0.07	0.04 – 0.10	**<0.001**
7-5 interveners	0.02	-0.01 – 0.05	0.229
10-7 interveners	-0.15	-0.18 – -0.12	**<0.001**
SPEAKER			
MA3	-0.01	-0.03 – 0.01	0.273
Item (z-scored)			
Frequency	0.01	-0.00 – 0.02	0.065
PNH	0.00	-0.01 – 0.01	0.945
AoA	0.01	0.00 – 0.02	**0.020**
Duration	0.07	0.06 – 0.08	**<0.001**
Trial	-0.05	-0.05 – -0.04	**<0.001**
Previous RTLog	0.06	0.05 – 0.07	**<0.001**
Participant			
zAge	0.02	-0.00 – 0.05	0.087
Male	-0.05	-0.11 – 0.00	0.062
Group			
2	0.02	-0.06 – 0.09	0.653
3	0.05	-0.03 – 0.13	0.219
4	-0.00	-0.08 – 0.07	0.910
VOICE (Diff.) × DIST. (1-0)	0.01	-0.03 – 0.05	0.577
VOICE (Diff.) × DIST. (2-1)	-0.01	-0.05 – 0.03	0.580
VOICE (Diff.) × DIST. (3-2)	-0.00	-0.05 – 0.04	0.841
VOICE (Diff.) × DIST. (4-3)	-0.01	-0.05 – 0.03	0.611
VOICE (Diff.) × DIST. (5-4)	0.01	-0.03 – 0.05	0.608
VOICE (Diff.) × DIST. (7-5)	-0.02	-0.06 – 0.02	0.302
VOICE (Diff.) × DIST. (10-7)	0.02	-0.02 – 0.06	0.339
Random Effects			
σ^2	0.08		
τ_{00} Item	0.00		
τ_{00} Participant	0.02		
Observations	11614		
Marginal R^2 / Conditional R^2	0.201 / 0.389		

In this model, we see a main effect of VOICE introducing a slow-down from same to different voice pairs ($p < 0.001$). The backwards-difference coded DISTANCE factor showed multiple main effects ($p < 0.05$), indicating slower response times for gradually increasing distances, except for the comparisons between 4 to 3 (p = 0.077) and 6 to 5 interveners (p = 0.229). No interactions between DISTANCE and VOICE were found to be significant however. Overall, these models show a difference in reaction time between same-voice and different-voice pairs. Additionally, response time to *'old'* targets increased as the distance between prime and target increased. Without any prior notions about the response time patterns for recognition responses though, we hesitate to interpret these patterns in meaningful ways.

4.3.3 Discussion

The analysis of Experiment 5 indicates the following:

- **Abstract repetition priming**
  - *Accuracy:*
    - General pattern for earlier distances to be recognized more accurately and later distances less accurately
  - *Reaction time:*
    - All repeated targets in all distances recognized faster than primes (except in the different-voice condition at the 10 intervener distance)
- **Talker-specific effects**
  - *Accuracy:*
    - Potential indications that accuracy improved in the same-voice condition up to 2 interveners but only 1 in the different-voice condition
  - *Reaction time:*
    - Significant main effect of voice-switch with no significant interaction terms, indicating overall slow-down

Contrary to the expectations from the literature, we did not find evidence for TSEs in this implementation of a continuous word recognition task. Increasing the number of intervening items from 0 up to 10 did cause effects on recall accuracy which did not interact with the voice-switch manipulation. One potential reason is that our study looked at switching voices within-gender (i.e., both speakers were male), while most of the other continuous word recognition studies tested male to female voice switches. However, the studies presented in Palmeri et al. (1993) and Campeanu et al. (2014) both found significant TSEs when comparing between same and different voices within the same gender. Another possibility for the different results in our study is our inclusion of non-words. One half of the experiment tested participants' ability to remember having heard a non-word before, which is not strictly necessary in a recognition task. We kept this manipulation however to keep Experiment 4 and Experiment 5 maximally similar, as the point of the study was to investigate the potentially different nature of TSEs in implicit and explicit tasks. The next experiment tests whether the same results hold in a blocked word recognition study.

4.4 Experiment 6

Following up on the results of Experiment 5, which did not find strong evidence for TSEs in a continuous word recognition task, Experiment 6 is designed to find TSEs in a blocked word recognition task. This experiment can be thought of as an attempt to replicate the explicit task of the second experiment found in Goldinger (1996). In that experiment (discussed at length in Section 4.2.2), words in the study block were presented (between-subjects) with three different encoding tasks: gender of the speaker, first phoneme, and syntactic part-of-speech. Six voices (three male, three female) presented these words to 35 participants in each encoding condition. The strongest abstract priming effects were found in the deepest level of processing (the syntactic judgment). Goldinger cites however that this level showed the lowest amount of TSEs compared to the gender-classification encoding task. As noted above however, the gender-classification task resulted in hit rates hovering around 50%, which is the chance level for a two-choice response task. Interpreting the low processing level task as not reliably finding any effects, we choose a higher level, semantic encoding task which we expect to find more realistic effects of both TSEs and abstract repetition priming. Lastly, to keep a similar voice manipulation as Goldinger (1996), we choose to present two speakers here; one male and one female.

The expectation given the literature is that a strong effect of voice switch will be seen between the items in the test block. The semantic encoding task may reduce the effect size of TSEs, given the heightened focus on abstract attributes. However, the fact that the voice switch between male and female voices is quite salient should highlight any effects of voice. Lastly, the relatively small distances between study and test block, with only the time it takes to read the instructions in between, is much shorter than the experiments in the literature finding long-distance TSEs.

4.4.1 Method

4.4.1.1 Participants

A total of 106 (age range 8-71, mean 29.2; 66 male) participants were run at the end of the fall semester of 2017. They were recruited from the experimental platform Prolific and voluntarily completed the study online using a custom Ibex implementation of a blocked word recognition task.

4.4.1.2 Stimuli

This experiment consisted of a new set of 432 total word stimuli. No non-words were included in this experiment. Two tokens of each stimulus were recorded and used in the experiment by speakers MA1 and FM1. These words were chosen to encompass a wide range of frequency values, according to the *SUBTLEX-US* database (Brysbaert & New, 2009) to allow for potential effects of word frequency on recall memory to arise. For a complete list of the properties of the stimuli in this experiment, see APPENDIX I: Experimental stimuli.

A unique semantic associate and a unique semantically unrelated word were chosen for each of the 432 auditory stimuli. These pairs of words were presented visually in the study block to create a semantic classification task. This was accomplished by choosing a word from the *University of South Florida Free Association Norms* (Nelson, 1998). Care was taken to choose a highly associated word from these by preferring normed words with high forward and backward reliability. When this was not possible, the list of related words generated from *Latent Semantic Analysis* (LSA: Dennis, 2007) was consulted. In rare cases, neither of these methods resulted in a usable word, in which case the author determined a suitable candidate. Unlike the monosyllabic auditory stimuli exclusively used in this thesis, both semantically related and un-related words were not restricted to be monosyllabic. The LSA relationship between an auditory word and its related and unrelated words was consulted and roughly indicated a divide between the two types of words. Without a better measure of semantic relatedness, it is impossible to definitely clarify the semantic relatedness between two given items. However, we note that the purpose of these words was to give participants a forced choice response between only two options. This classification was only included to have the participants engage with the meaning of the auditory word; the relationship between that word and the semantic associate will not be discussed further here.

The figure below illustrates the relationship between frequency (mean = 2.19, standard deviation = 0.75, range = [0.30, 3.89]), age of acquisition (mean = 7.39, standard deviation = 2.77, range = [2.50, 15.27]), and phonological neighborhood density (mean = 13.18, standard deviation = 9.14, range = [0, 42]) for the words in this experiment. Unlike the other experiments in this thesis, this set of stimuli was constructed to vary greatly in terms of frequency, as mentioned above. Additionally, age of acquisition and phonological neighborhood density appear highly correlated with frequency, which makes it problematic to investigate the joint contributions of each of these properties.

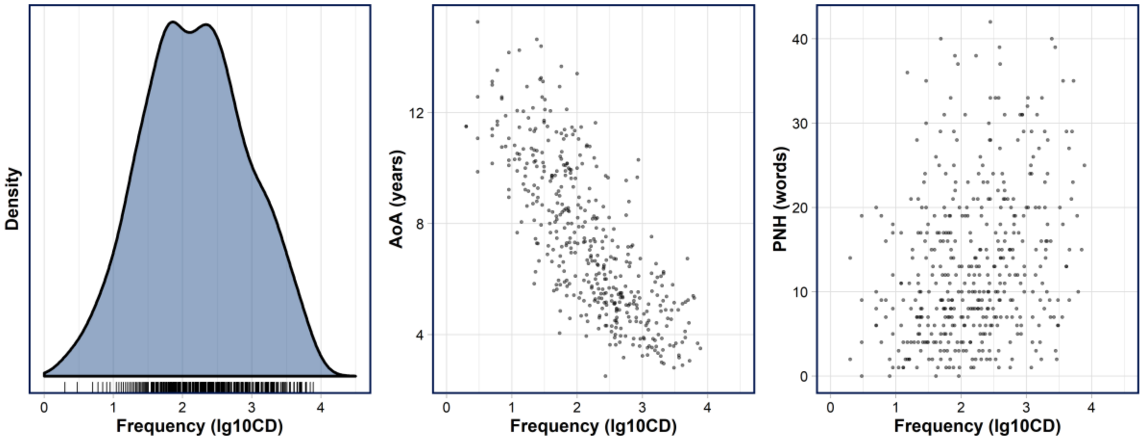

Figure 31: Properties of Experiment 6 word stimuli

4.4.1.3 Design

The design of this experiment, being quite different from the others presented in this thesis, deserves some explanation. To begin, the experiment was conducted in two blocks: the study and the test block. The study block was presented as a semantic classification task in which 288 words were auditorily presented (50% from each of the speakers MA1 and FM1). The participant had to choose which of two visually presented words on the screen was most semantically related (described as a *'meaning association'* in the instructions) to the auditory word. These two words were presented in a random order per trial. Additionally, the stimuli triplets (auditory cue word and related/unrelated visual words) were randomly presented per participant, mitigating any potential list effects. The two possible responses were presented first for 500ms and then the auditory word began playing (matching the design in Goldinger, 1996). After the response, a random ISI per trial of 250 to 500ms was inserted between the response and the beginning of the next item. The test block then consisted of an *'old'/'new'* recognition task, in similar fashion to Experiment 5. Again, 288 words were auditorily presented and participants indicated whether they recalled hearing the item before in the study block. These words, similar to the study block, were randomly presented by participant, with a random ISI per trial between 750 and 1000ms.

Participants were recruited into one of twelve lists, which are now described. The 432 items were divided into 12 groups of 32 words each, which roughly spanned the entire frequency range. These 12 groups of words were then assigned in rotating Latin Squares fashion following the table below, with each row being assigned one of the 12 groups of words:

Table 35: Experiment 6 design

Word Group	Study Block SPEAKER	Test Block SPEAKER	CONDITION
1		MA1	***New***
2		MA1	
3		FM1	
4		FM1	
5	MA1		***Study***
6	MA1		
7	FM1		
8	FM1		
9	FM1	FM1	***Old-Same***
10	MA1	MA1	
11	FM1	MA1	***Old-Diff.***
12	MA1	FM1	

These were chosen for the following reasons:

- This design equates the length of the study and test blocks, which was not done in a number of the previous studies
- Each of the two voices is represented 50% in each block
- Not all of the words heard in the study block were later heard in the test block, unlike in previous studies
- Any frequency effects are roughly equated across conditions

The CONDITION factor allows us to test for TSEs in the comparison between the *Old-Same* and *Old-Different* conditions. The SPEAKER factor is also counter-balanced across lists. This design lead to a repetition rate of 1/3 in the test block (1/6 overall). As both blocks were randomized, the amount of intervening items between prime and target was vastly different. Lastly, the same target items were used across all the statistical comparisons; that is, the same sound-file was presented between-subjects for the *New*, *Old-Same*, and *Old-Different* conditions.

4.4.1.4 Procedure

The experiment, as presented on Prolific, was titled *Word Association Study* and had the description *“In this study, you will listen to words and indicate with the keyboard one of two responses, printed on the screen, that best fits the word. You will also respond to a very brief demographic questionnaire. Completion time depends on your internet connection speed, as downloading the stimuli may take a while.”* The study block was introduced with the following practice triplets (italics indicating the auditory cue and bold-face indicating the semantic associate): *mile*$_1$ – **kilometer**/visa; *pin*$_2$ – **tack**/glee; *jam*$_2$ – **berry**/chapel; *blue*$_1$ – **purple**/sonnet; *nice*$_2$ – **kind**/knuckle; and *straw*$_1$ – **fodder**/content. The instructions for the study block were *“Press* 'F' *for the word on the left or* 'J' *for the word on the right."* The visual response possibilities were displayed on the screen for 500ms and then the auditory cue began, which was the same procedure in the study block. A by-trial random ISI of 250 to 500ms intervened between each practice item and each item in the study block. These practice items were presented with feedback about the correct response, as the online nature of the experiment prevented participants from asking for clarification before beginning the experiment.

Between the study and test blocks, the recognition task was explained such that participants had to indicate with *‘F’* if the word they heard was *‘old’* and *‘J’* if it was *‘new’*. No practice trials were included between study and test blocks. All experiments included in the analysis were completed on average within 28 minutes (range: 16-82 minutes). As mentioned, each participant had a unique trial order, as both study and test block items were randomly presented by participant.

4.4.1.5 Analysis

Of the original 106 participants in the experiment, 23 were removed for poor performance. Eight of these were removed due to accuracy less than 70% on the study block alone. Another eight were removed due to having over twenty near-immediate responses, and seven more due to having Hodges-Lehmann estimated RT distribution outliers. After this global participant removal, the overall results per each item were examined. With this experiment, no items had accuracy scores of less than 50% correct.

With a different task not focused on reaction time, it is again not straightforward which data-trimming steps should be taken. Upon visualizing the reaction time data, the overall density roughly matched that from Experiment 4 and Experiment 5 even though different items were involved. The range was extended slightly for this experiment, leading to a global trimming of responses which took less than 500ms or greater than 3500ms. By-participant and by-item trimming of reaction time was not performed however, as the analysis of recognition task data centers on accuracy.

Table 36: Experiment 6 removal summary

	Observations	**Percentage**
Inaccurate trials	*4149*	*34.7*
RT trimming (300 > RT < 3000)	*499*	*4.2*
Total removed	*4648*	*38.9*
Total remaining	*7304*	

4.4.2 Results

The following table reports the distribution of the data per the factors BLOCK, CONDITION, and SPEAKER. These data are reported after only the minimal global trimming procedure of unreasonable reaction times was applied. Reaction time reports are central tendency measures from the non-parametric Hodges-Lehmann estimates and accuracy scores indicate the amount of correct responses out of the total, after all global participant and item removal was conducted.

Table 37: Experiment 6 data summary

BLOCK			**CONDITION**			**SPEAKER**			**Total**
Study	*1181*	*95.1*	**Same**	*1182*	*95.1*	**MA1**	*1198*	*94.5*	***2764***
						FM1	*1167*	*95.7*	***2739***
			Diff	*1180*	*95.2*	**MA1**	*1194*	*94.8*	***2753***
						FM1	*1165*	*95.5*	***2776***
Test	*1165*	*70.9*	**Same**	*1158*	*66.4*	**MA1**	*1176*	*65.4*	***2842***
						FM1	*1138*	*67.3*	***2838***
			Diff	*1150*	*65*	**MA1**	*1166*	*65*	***2860***
						FM1	*1134*	*65*	***2853***
			New	*1176*	*76.2*	**MA1**	*1204*	*76.7*	***5685***
						FM1	*1146*	*75.8*	***5697***

4.4.2.1 Accuracy

For this experiment, only one generalized linear mixed-effects model of accuracy was fit to the data. This model, reported below, tests the difference between the same-voice and different-voice CONDITIONS in the test BLOCK. This model sets the same-voice repetitions as the reference level for the CONDITION factor, with the factor SPEAKER dummy-coded. The distance between prime and target (a continuous range between 7 and 574 intervening items) was *z*-scored and allowed to interact with the CONDITION and SPEAKER factors. Random effects were set as intercepts for participants and items. The outcome of this model is seen below.

Table 38: Experiment 6 target accuracy model

	Accuracy (targets)		
Predictors	*Odds Ratios*	*CI*	*p-values*
Intercept	2.19	1.87 – 2.56	**<0.001**
VOICE			
Old-Diff.	0.97	0.87 – 1.10	0.671
DISTANCE			
zDistance	0.89	0.81 – 0.98	**0.020**
SPEAKER			
FM1	1.10	0.97 – 1.26	0.139
Item (z-scored)			
Frequency	0.84	0.78 – 0.91	**<0.001**
PNH	0.92	0.87 – 0.97	**0.001**
AoA	1.20	1.11 – 1.30	**<0.001**
Duration	1.02	0.97 – 1.08	0.377
Trial	0.93	0.88 – 0.99	**0.019**
Participant			
zAge	0.94	0.82 – 1.07	0.324
Male	0.67	0.51 – 0.89	**0.006**
Group			
2	1.56	0.84 – 2.89	0.159
3	1.95	1.08 – 3.53	**0.027**
4	1.67	0.88 – 3.17	0.119
5	1.06	0.56 – 2.00	0.858
6	0.84	0.45 – 1.57	0.589
7	0.66	0.35 – 1.25	0.206
8	0.68	0.38 – 1.24	0.207
9	1.41	0.74 – 2.68	0.295
10	0.81	0.42 – 1.55	0.522
11	1.23	0.69 – 2.19	0.473
12	0.70	0.37 – 1.32	0.267
VOICE (Old-Diff.) × DIST. (zDist)	1.02	0.91 – 1.15	0.742
VOICE (Old-Diff.) × SPKR. (FM1)	0.92	0.78 – 1.08	0.312
DIST. (zDist) × SPKR. (FM1)	0.98	0.87 – 1.10	0.713
V. (Old-Diff.) × D. (zDist) × S. (FM1)	1.01	0.85 – 1.19	0.929
Observations	11366		
Marginal R^2 / Conditional R^2	0.079 / 0.190		

This model shows whether any priming was seen when items were repeated by different voices. The non-significant term VOICE$_{\text{Old-Diff.}}$ ($p = 0.671$) indicates that the numerical trend from 65% to 66.4% hit rate accuracy is not a significant difference. The *z*-scored predictor for distance was significant (p = 0.02), with increased distance leading to decreased accuracy. The main effect of speaker was not significant (p = 0.139) and neither were any interaction effects of interest. Overall, this model indicates that the data revealed only effects of increasing distance on accuracy priming. The plot below illustrates the effect of intervening items on the accuracy hit rates in this experiment.

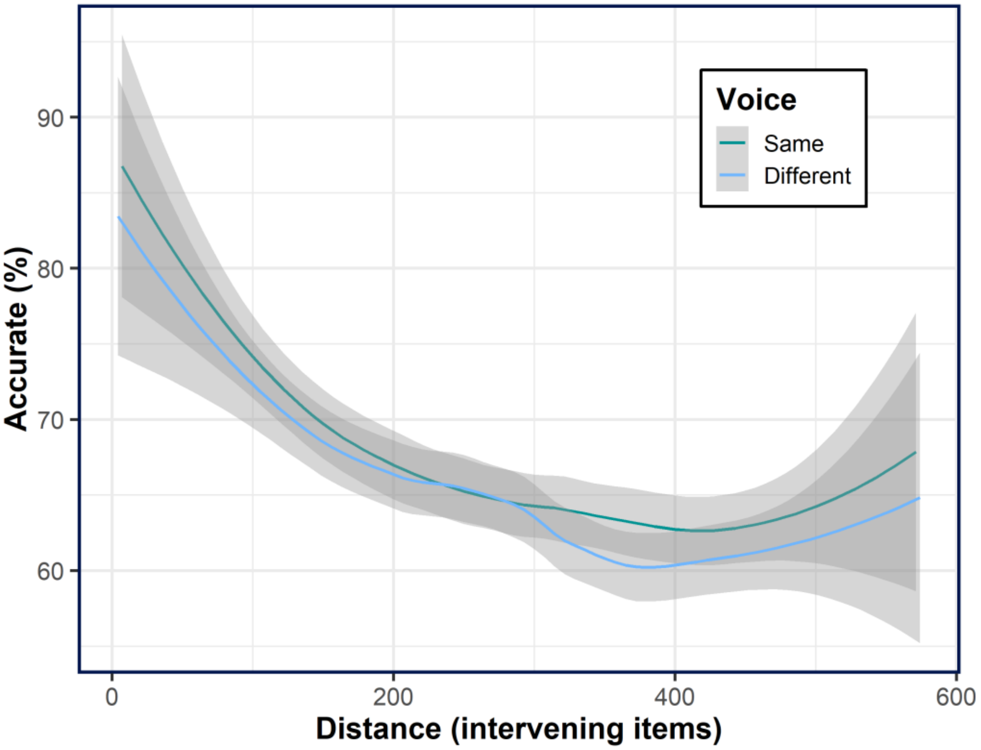

Figure 32: Experiment 6 accuracy over distance

4.4.2.2 Reaction Time

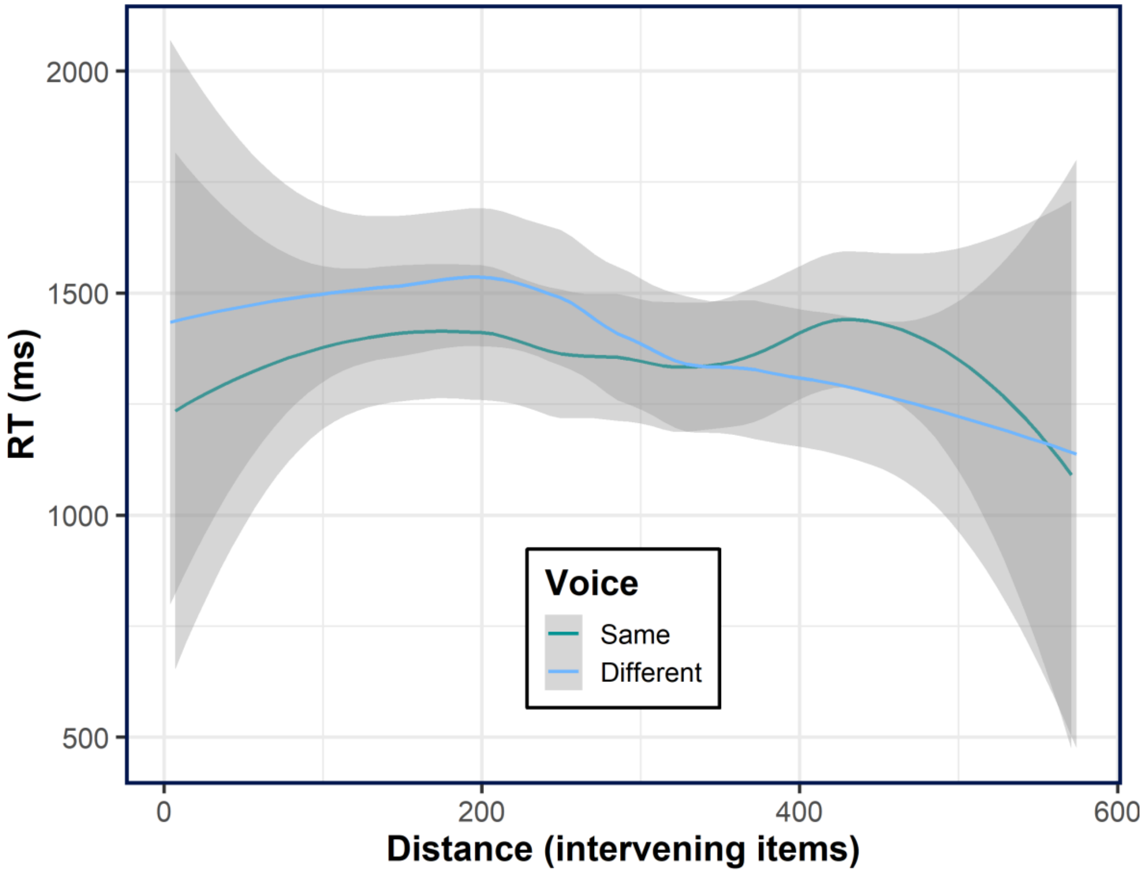

Figure 33: Experiment 6 RT over distance

The graph above illustrate the effect of the number of intervening items on the reaction time to the targets in the test block. No analyses were conducted on reaction time data in this experiment, as we have no predictions about how the reaction time should change from study to test block.

4.4.3 Discussion

The analysis of Experiment 6 indicates the following:

- **Talker-specific effects**

  - *Accuracy:*
    - No main effects found between targets in same- and different-voice pairs
  - *Reaction time:*
    - Not tested

Overall, the results from this experiment are easily summarized. In a controlled study testing the effect of switching speaker (between gender) from study to test blocks, no TSEs are found in an analysis of accuracy hit rates. This study used an encoding task asking participants to pick which of two visually presented words was the most semantically associated to the word they heard. It is possible, as the argument goes in the literature, that the focus of the encoding task on abstract content of the word caused participants to only generate abstract representations throughout the experiment. However, the voice switches occurred between male and female tokens; presumably causing a largely reduced amount of perceptual similarity between the two. This finding dramatically contrasts with the findings of Goldinger (1996) *inter alia* presented in the review of blocked word recognition studies.

4.5 Conclusion

In this chapter, we investigated the time-course of TSEs found in explicit tasks. First, the difference between implicit and explicit tasks was motivated. Much evidence supports a distinction between the memory systems that each engages with. That being said, the conclusion from the literature is that moderately long-lasting effects (up to around a day) of voice switches have been found using explicit tasks. These come from both blocked word recognition studies, with primes and targets presented separately in study and test blocks and from continuous word recognition studies, with primes and targets mixed together.

Adapting the stimuli and design from Experiment 4, we conducted a continuous word recognition task. The only main difference between Experiment 4 and Experiment 5 was in the response that participants were asked to give upon hearing an item. Contrary to what was found in the literature, no robust TSEs emerged during the entire experiment. Moderate conclusions can be made that a same-voice advantage persists up to two intervening items, but nothing resembling the long-distance effects of Palmeri et al. (1993) was found. Then, we conducted a blocked word recognition experiment designed to replicate the results found by Goldinger (1996) as straightforwardly as possible. The results of this study yielded only a 1.5% numerical same-voice advantage; a non-significant result.

These results confirm the conclusions of the past chapter that TSEs are only found in short distances between prime and target. They stand in stark contrast to the results cited from the literature that TSEs exist for up to a day in explicit tasks and up to a week in implicit tasks. In the next chapter, we conclude this thesis by discussing possible differences in our studies that may have contributed to the contrasting effects.

CHAPTER 5: General Discussion

5.1 Experimental Summary

The experiments throughout this thesis were designed to determine the impact of talker-specific information on speech perception. This investigation centered on comparing conditions in which prime and target were spoken either by the same speaker or by different speakers. Using the lexical decision task to examine implicit memory, the findings point to a relatively short effect of talker-specific information. After distances of around three intervening items, same- and different-voice pairs cease being statistically different. Robust abstract priming effects are found at all tested distances however. This finding is expected given the literature using the lexical decision task but unexpected from the numerous studies testing recall accuracy in the perceptual identification task.

Turning to explicit memory, two additional studies were conducted using the continuous and blocked word recognition tasks. The expectation from nearly all studies in the literature is that robust talker-specific effects on recall accuracy should be found at long-distances, at least up until one day intervenes between prime and target. However, in these two experiments, this expectation was not borne out. In the continuous word recognition task, effects of voice on recall accuracy were only seen up until two intervening items. In the blocked word recognition task, only a small, statistically insignificant trend for talker-specific effects was found. If supported by future attempts at replication, these results call into question the primary focus on talker-specific information, and by extension token-specific information, in building models of speech perception. The following sections discuss potential reasons the findings of these experiments differ from the predictions in the literature.

5.1.1 Frequency Effects

One potential issue surrounding studies of this nature are the abstract properties of the stimuli themselves. It is certainly possible that the use of different words and non-words creates different expectations for participants, which may lead to conflicting results. Additionally, token-specific effects may be dependent on properties like word frequency or neighborhood density for instance. This is far from a novel idea, but a full discussion of this potential issue is outside the current field of investigation.

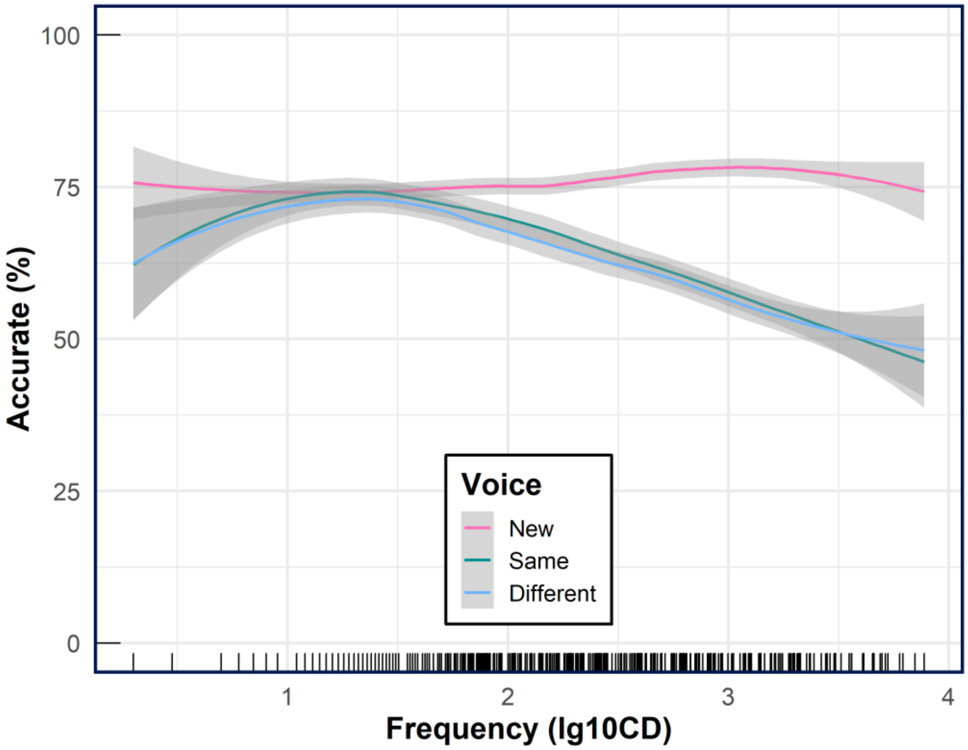

Figure 34: Experiment 6 accuracy by frequency

As a small demonstration however, Figure 34 plots the relationship between recall accuracy and word frequency in the test block of Experiment 6. In this experiment, word frequency was explicitly manipulated in order to provide a window onto differential effects for future research. As can be seen, increasing word frequency decreases the ability of participants to recall having heard words before. Interestingly, no frequency effects are seen when words are presented for the first time. This pattern is apparently only seen in explicit memory tasks, as Figure 35 below shows.

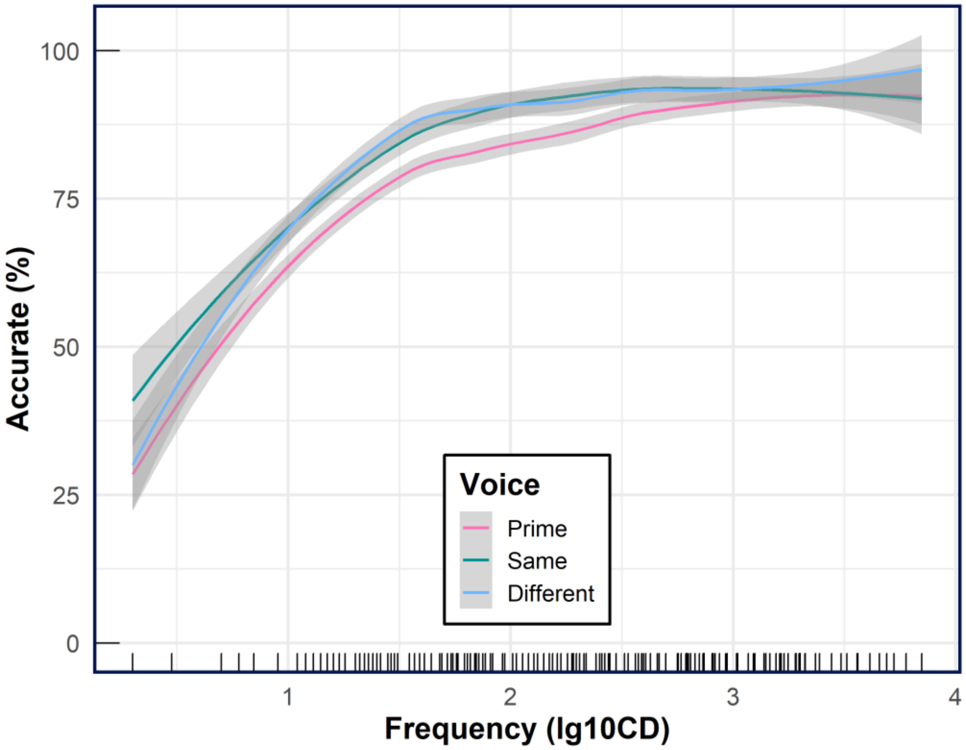

Figure 35: Accuracy by frequency in a lexical decision task

This figure comes from an additional study not reported here which used a set of stimuli similar to Experiment 6. The purpose of this study was to examine the effects of accuracy in a lexical decision task, similar in design with Experiment 4. As can be seen, frequency effects are apparent starting around a *lg10CD* frequency measure of 1.5, before which the majority of participants do not recognize the low-frequency words. An abstract priming effect is seen, as less frequent words exhibit greater accuracy priming effects.

It is certainly possible that other studies in the literature use words with a different frequency range than the studies in this thesis. These differences may be a potential source of the conflicting results found here. However, we note in passing that, while frequency does seem to stronger impact abstract repetition priming, it does not appear to differentially affect talker-specific effects. To conclusively determine this, further work is warranted.

5.1.2 Talker-specific representations

Another possible locus of difference between the experiments in this thesis and those in the literature concerns the nature of talker-specific representations. As discussed previously, one of the main reasons to investigate token-specific detail is to model the phenomenon of speaker recognition. Nygaard & Pisoni (1998), among others, have demonstrated that a participant's perception is improved when presented with stimuli from someone known to that participant. It is therefore possible that the creation of an abstract representation of a speaker over the course of an experiment can influence the presence of token-specific effects. Perhaps, given the long-distance manipulations used in previous studies which often span across days, talker-specific representations are highlighted in ways that were not present in this thesis' experiments. Further investigation of this possibilty is required to determine the impact of talker-specific representations; which promises to provide useful data in generalizing results from these types of experiments to actual speech perception outside of the lab.

5.1.3 Lexicality Effects

Lastly, another angle which was not addressed in this thesis concerns the data from non-word stimuli. As mentioned, results from investigations of this type have been cited to support hybrid models of speech perception. Models solely built on abstract representations have no straightforward way to account for the presence of non-word priming effects without recourse to sub-lexical, abstract representations.

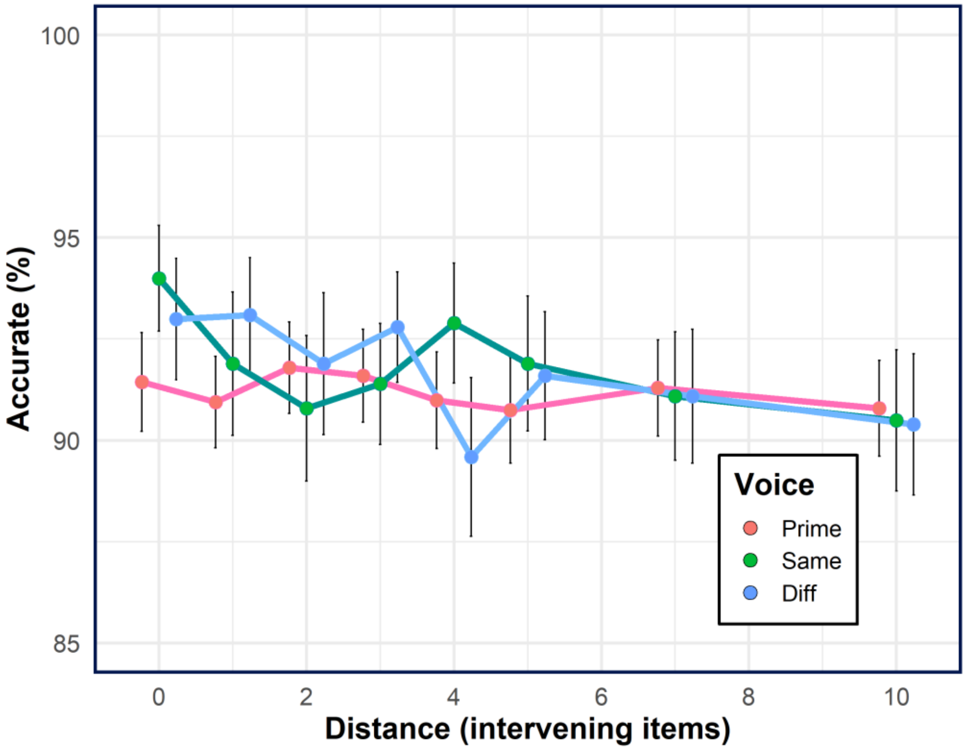

Figure 36: Experiment 4 non-word accuracy

Without delving into a deeper investigation here, we present graphs of the non-word data from Experiment 4 and Experiment 5. These two experiments, using the lexical decision and continuous word recognition tasks, were identical except for the response task performed by the participants. Figure 36 presents the non-word accuracy data and Figure 37 presents the reaction time data from the lexical decision task in Experiment 4.

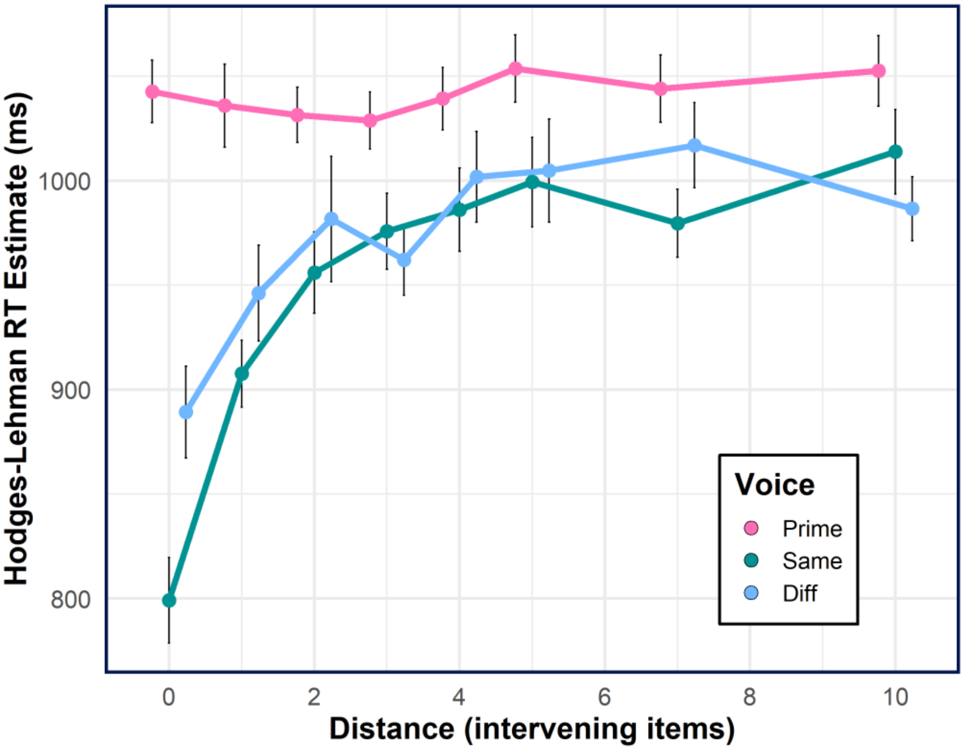

Figure 37: Experiment 4 non-word RT

The pattern seen in the figures is remarkably similar to the results from words presented in Section 3.5.2. No visible effect is seen in accuracy while short-term effects, again up until around three words intervene, are seen in the reaction time data. The following two figures plot the same type of information for the continuous word recognition task of Experiment 5. Figure 38 plots the non-word accuracy data while Figure 39 plots the non-word reaction time data.

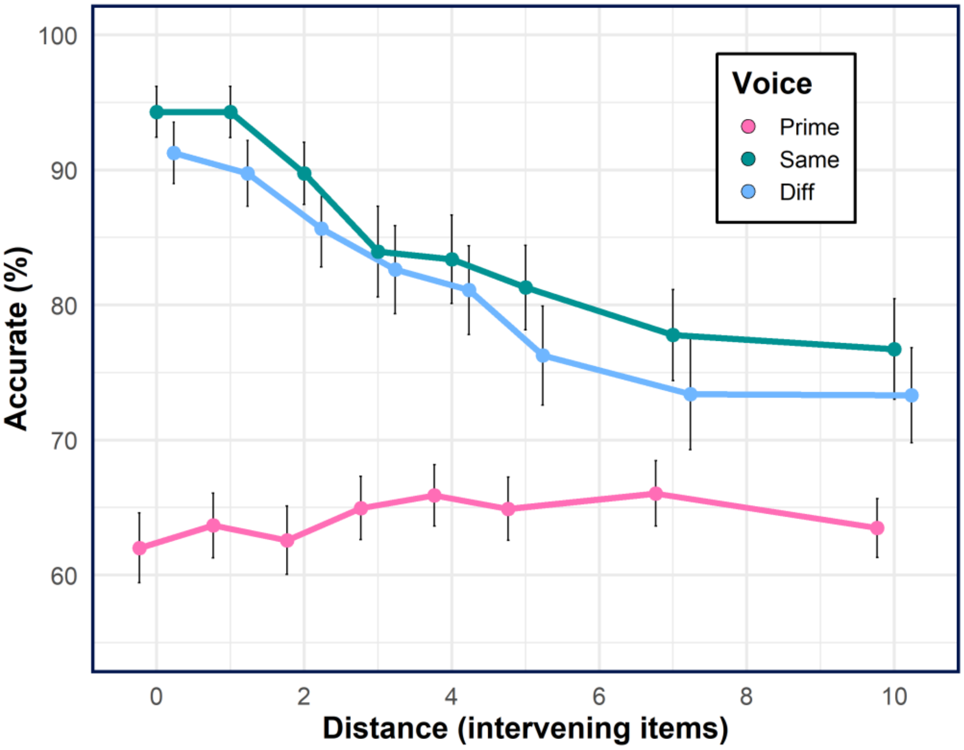

Figure 38: Experiment 5 non-word accuracy

Except for a much-reduced ability to respond to non-word primes, participants' ability to recall word stimuli mirrors that of the word data. The same numerical trend is seen whereby same-voice pairs are recalled more accurately than different-voice pairs. In the reaction time data, again a similar pattern to the word data is seen.

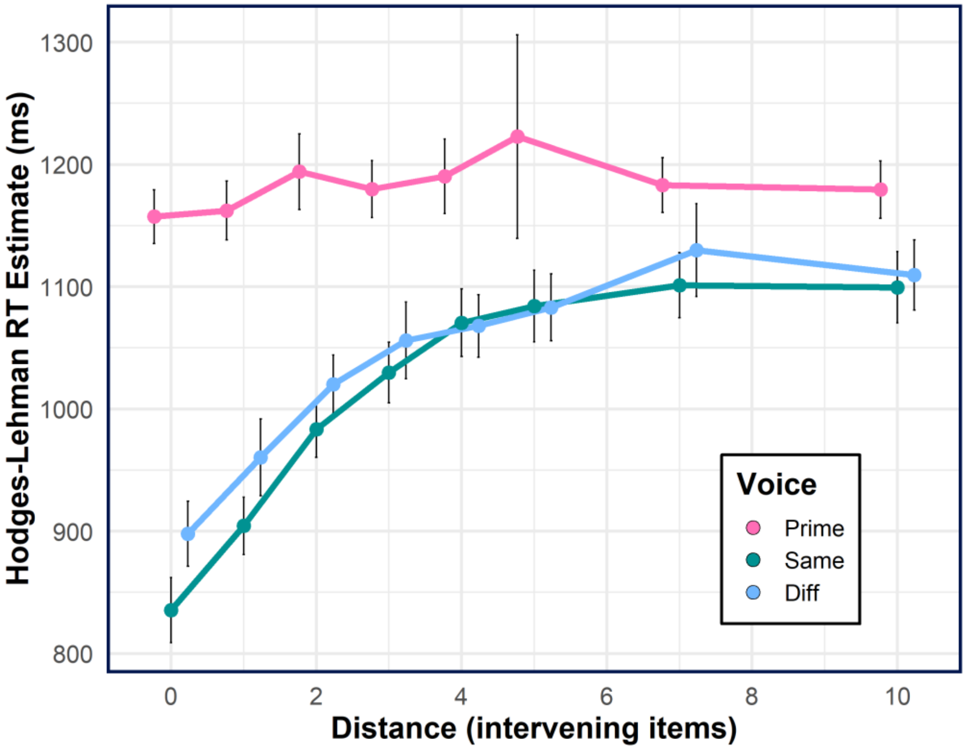

Figure 39: Experiment 5 non-word RT

If supported by future in-depth analyses, these graphs support the conclusions of Orfanidou et al. (2011), among others, who cite non-word priming effects as demonstrating the necessity of hybrid models of speech perception.

5.2 Conclusion

In the introduction, we set up the distinction between two types of hybrid models of speech perception. The distinction hinges on how the hybrid model packages the joint information of abstract content and token-specific detail. One way of doing so, which we term *combined* representations or single-route representational accounts, consists of late stage abstractions constructed from detail-rich representations. This type of hybrid model emphasizes detail-rich information as the primary basis upon which speech perception occurs. The other type of hybrid model posits a separation between abstract and token-specific types of speech information. In these hybrid models, which we term *separate* or dual-route representational accounts, abstract representations are not based on a prior detail-rich representation. Rather, both abstract and token-specific information have different representations, account for different patterns of priming, and are located in different structures in the brain.

The goal of this thesis is to distinguish between these two types of hybrid accounts of speech perception. This thesis focuses on the decay patterns of talker-specific effects, as discerning how long these effects last compared to abstract repetition priming points to their relative importance in word recognition. If talker-specific effects are robustly found at long-distances, then it is quite likely that detail-rich representations underlie word recognition. If they are only found at early stages while abstract repetition priming is found to be strong and long-lasting, then the focus of modelling word recognition should be on abstract speech information.

In investigating the presence of talker-specific effects, we first set up the distinction between implicit and explicit priming effects. This turns out to be quite an important distinction, as much research shows that these are indeed separate phenomena. Therefore, different tasks used to determine the presence of priming effects may be selectively highlighting or prioritizing the creation of certain types of representations. The experiments presented here are broken down into two types: implicit lexical decision and explicit word recognition experiments.

The literature investigating the presence of talker-specific effects using implicit tasks yields contradictory results. From studies testing recognition accuracy on perceptual identification tasks, long-lasting effects of voice have been found. Using indirect tasks which test differences in reaction time, like shadowing or lexical decision, only very short-term voice effects have been found. Our three experiments using the lexical decision task mirror the findings in the literature. We find short-term effects of talker-specific information, contrasting same-voice with different-voice (same-gender) pairs, which last until three items intervene between prime and target. After which, only abstract repetition priming effects are seen. In a further experiment comparing same-gender to different-gender pairs (both different voices), small effects of voice were found at distances of ten intervening items. These results however indicated a speed up in reaction time when the prime and target mismatched in gender. This result contradicts the hypothesis that increased perceptual similarity should lead to increased priming effects.

Turning to explicit tasks, talker-specific effects have been found lasting up to, and even past, one day intervening between prime and target. To investigate this, we first implemented a direct comparison between an implicit lexical decision task and an explicit continuous word recognition task. These two experiments used the same stimuli, same design, and had similar analysis methods. The only difference involved what response task the participants were performing upon hearing the items. Contrary to the literature, we again found marginal effects of voice only at short distances.

Finally using a blocked word recognition task, we attempted a replication of some of the more established findings in the literature. Words were presented in two blocks – a study and test block. In the study block, participants had to indicate which of two visually presented words was more semantically associated with an auditory cue. In the test block, participants indicated whether they had previously heard a word before in the study block. The overwhelming prediction from the literature was that robust talker-specific effects should be found, even when the two voices involved were male and female voices. However, again we found no statistically significant effects of voice in the recall accuracy hit rates.

Overall, these findings motivate a return to looking at speech perception models which emphasize the importance of abstract information. Instead of being causally linked to word recognition and underlying abstract representations, the pattern of talker-specific effects in the studies in this thesis indicates that token-specific information is stored in representations separate from abstract, invariant information. Within early time-windows only, token-specific information impacts word recognition. After which, the word recognition process proceeds by abstracting away from token-specific detail. These findings straightforwardly point to a dual-route representational hybrid model of speech perception. In the field of speech perception, these findings motivate future work in investigating the nature and interaction of these two separate representations.

APPENDIX I: Experimental stimuli

AI.1 Words

The following table displays the list of words used in the experiments in this thesis. The column names ***1-10*** are described below:

1. Word, with all statistics calculated from the given orthography in the case of homophones
2. Number of phonemes in the word
3. Frequency of the word, the *lg10CD* measure from the SUBTLEX-US (Brysbaert & New, 2009)
4. Age of acquisition measure, the average rating from Kuperman et al. (2012)
5. Phonological neighborhood density (excluding homophones) measure, the *PNH* measure from the *English Lexicon Project* (Balota et al., 2007)
6. Concreteness measure, from Brysbaert et al. (2014)
7. Frequency of the onset of the word, rounded to three decimal places (described in Section 2.2.2)
8. As above, but for the vowel of the word
9. As above, but for the coda of the word
10. List of experiment numbers the word was used in (most often 1 – 3, 1 – 5, or 1 – 6)

1	2	3	4	5	6	7	8	9	10
arm	3	3.29	3.26	4	4.96	0.123	0.042	0	1–3
art	3	3.2	6.21	4	4.17	0.123	0.042	0.001	1–3
babe	3	3	6.61	15	3.67	0.038	0.053	0.005	1–3
bag	3	3.4	4.28	27	4.9	0.038	0.065	0.006	1–3
bait	3	2.59	7.11	41	4.45	0.038	0.053	0.023	1–3
ball	3	3.32	2.9	33	5	0.038	0.022	0.035	1–3
bar	3	3.35	6.9	26	4.67	0.038	0.042	0.004	1–3
bear	3	3.18	3.58	37	4.88	0.038	0.073	0.004	1–3
beat	3	3.5	6.15	40	3.97	0.038	0.091	0.023	1–3
bill	3	3.27	6.42	34	4.68	0.038	0.177	0.035	1–3
bin	3	2.23	4.68	33	4.72	0.038	0.177	0.089	1–3
board	4	3.23	6.37	29	4.57	0.038	0.022	0.002	1–3
buck	3	2.84	7.68	27	4.67	0.038	0.25	0.03	1–3
cab	3	2.98	6.94	21	4.88	0.054	0.065	0.005	1–3
can	3	3.92	4.32	30	4.55	0.054	0.065	0.089	1–3
car	3	3.71	3.37	21	4.89	0.054	0.042	0.004	1–3
chick	3	2.87	5.53	24	4.93	0.011	0.177	0.03	1–3
choice	3	3.49	5.17	4	1.9	0.011	0.004	0.03	1–3
club	4	3.33	5.89	3	3.78	0.006	0.25	0.005	1–3
cook	3	3.14	4.22	13	4.32	0.054	0.007	0.03	1–3
cop	3	3.17	4.94	23	4.3	0.054	0.042	0.009	1–3
crane	4	2.5	6.78	13	4.68	0.008	0.053	0.089	1–3
crawl	4	2.67	3.89	4	4.27	0.008	0.022	0.035	1–3
crew	3	3.02	7.56	13	4.36	0.008	0.027	0.514	1–3
cry	3	3.28	2.78	7	4	0.008	0.039	0.514	1–3
dice	3	2.47	6.37	17	4.86	0.052	0.039	0.03	1–3
dig	3	3.1	4.19	19	4.33	0.052	0.177	0.006	1–3
duck	3	2.85	3.5	25	4.86	0.052	0.25	0.03	1–3
elk	3	1.79	7.05	4	4.93	0.123	0.073	0	1–3
fan	3	3.04	5.63	26	4.71	0.031	0.065	0.089	1–3
fat	3	3.31	5.15	25	4.52	0.031	0.065	0.023	1–3
film	4	3.01	6.95	6	4.71	0.031	0.177	0	1–3
fish	3	3.23	4.05	12	5	0.031	0.177	0.004	1–3
fort	4	2.61	6.48	17	4.72	0.031	0.022	0.001	1–3
fox	4	2.61	5.02	12	4.97	0.031	0.042	0.005	1–3
gas	3	3.21	5.32	16	4.29	0.015	0.065	0.03	1–3
grave	4	2.95	7.06	16	4.56	0.007	0.053	0.006	1–3
green	4	3.28	3.79	15	4.07	0.007	0.091	0.089	1–3
lamb	3	2.51	4.15	29	4.97	0.066	0.065	0.021	1–3
laugh	3	3.29	3.79	16	4.21	0.066	0.065	0.003	1–3
limb	3	2.29	7.16	21	4.64	0.066	0.177	0.021	1–3
limp	4	2.18	7.16	7	4.15	0.066	0.177	0.001	1–3
lock	3	3.27	5.74	29	4.65	0.066	0.042	0.03	1–3
lung	3	2.38	7.16	12	4.82	0.066	0.25	0.034	1–3
mark	4	3.2	6.48	13	4.21	0.053	0.042	0.001	1–3
moan	3	1.92	7.72	28	3.72	0.053	0.036	0.089	1–3
nail	3	2.85	5.42	34	4.93	0.05	0.053	0.035	1–3
net	3	2.73	7	24	4.53	0.05	0.073	0.023	1–3
nut	3	2.74	5.21	27	4.52	0.05	0.25	0.023	1–3
pail	3	1.61	6.16	47	4.93	0.04	0.053	0.035	1–3
pant	4	1.6	5.94	16	4.38	0.04	0.065	0.01	1–3
park	4	3.27	4.47	11	4.74	0.04	0.042	0.001	1–3
pass	3	3.5	5.39	21	2.71	0.04	0.065	0.03	1–3
pin	3	2.74	4.53	33	4.92	0.04	0.177	0.089	1–3
plow	3	1.87	7.11	9	4.46	0.006	0.011	0.514	1–3
pump	4	2.66	6.06	11	4.31	0.04	0.25	0.001	1–3
race	3	3.06	6	26	3.59	0.048	0.053	0.03	1–3
rain	3	3.16	3.6	46	4.97	0.048	0.053	0.089	1–3
rinse	4	2.01	4.95	7	4.1	0.048	0.177	0.004	1–3
roll	3	3.28	4.47	39	4.16	0.048	0.036	0.035	1–3
rug	3	2.55	4.61	21	4.79	0.048	0.25	0.006	1–3
sail	3	2.54	6.47	43	4.59	0.061	0.053	0.035	1–3
screw	4	3.13	6.65	3	4.81	0.001	0.027	0.514	1–3
shop	3	3.21	5.78	15	4.31	0.027	0.042	0.009	1–3
size	3	3.23	4.84	34	3.13	0.061	0.039	0.033	1–3
smell	4	3.38	4.22	4	3.7	0.001	0.073	0.035	1–3
smoke	4	3.28	4	6	4.96	0.001	0.036	0.03	1–3
soul	3	3.31	6.17	36	1.86	0.061	0.036	0.035	1–3
speech	4	3.04	6.22	2	3.37	0.007	0.091	0.002	1–3
store	4	3.33	4.76	10	4.5	0.013	0.022	0.004	1–3
straw	4	2.39	4.22	2	4.77	0.005	0.022	0.514	1–3
sun	3	3.29	3.4	34	4.83	0.061	0.25	0.089	1–3
swing	4	2.93	4.16	16	4.54	0.002	0.177	0.034	1–3
tag	3	2.68	5	17	4.25	0.09	0.065	0.006	1–3
tape	3	3.17	4.42	13	4.9	0.09	0.053	0.009	1–3
tick	3	2.31	6.05	30	4.57	0.09	0.177	0.03	1–3

guess	3	3.83	5.44	12	2.28	0.015	0.073	0.03	1–3
gun	3	3.47	5.58	24	4.83	0.015	0.25	0.089	1–3
gym	3	2.73	6	18	4.83	0.017	0.177	0.021	1–3
head	3	3.8	3.42	28	4.75	0.019	0.073	0.03	1–3
hog	3	2.29	5.7	9	4.66	0.019	0.042	0.006	1–3
hump	4	2.21	6.11	11	4.36	0.019	0.25	0.001	1–3
inch	3	2.7	5.11	2	4.37	0.123	0.177	0.001	1–3
jam	3	2.66	6.56	22	4.71	0.017	0.065	0.021	1–3
joke	3	3.33	5.2	15	2.9	0.017	0.036	0.03	1–3
knife	3	3.1	4.15	9	4.9	0.05	0.039	0.003	1–3
knock	3	3.32	4.63	25	4.24	0.05	0.042	0.03	1–3
knot	3	2.18	6.05	32	4.87	0.05	0.042	0.023	1–3
lad	3	2.53	7.35	31	4.28	0.066	0.065	0.03	1–3
bath	3	3	3.23	20	4.85	0.038	0.065	0.002	1–3,6
bead	3	1.69	5.63	40	4.9	0.038	0.091	0.03	1–3,6
beak	3	1.95	5.42	37	4.96	0.038	0.091	0.03	1–3,6
bed	3	3.61	2.89	29	5	0.038	0.073	0.03	1–3,6
blood	4	3.55	4.89	7	4.86	0.005	0.25	0.03	1–3,6
cash	3	3.28	4.84	23	4.48	0.054	0.065	0.004	1–3,6
cheek	3	2.45	5.06	28	4.83	0.011	0.091	0.03	1–3,6
claw	3	2.14	4.7	4	4.83	0.006	0.022	0.514	1–3,6
cloak	4	1.97	6.95	13	4.71	0.006	0.036	0.03	1–3,6
fluff	4	1.74	4.85	5	3.8	0.005	0.25	0.003	1–3,6
foot	3	3.31	3.44	10	4.9	0.031	0.007	0.023	1–3,6
grain	4	2.26	7.44	21	4.8	0.007	0.053	0.089	1–3,6
hat	3	3.24	3.33	33	4.88	0.019	0.065	0.023	1–3,6
hoop	3	1.97	6.11	20	4.74	0.019	0.027	0.009	1–3,6
ink	3	2.44	5.16	2	4.56	0.123	0.177	0.002	1–3,6
kite	3	1.89	4.58	25	5	0.054	0.039	0.023	1–3,6
loaf	3	2.21	6.84	14	4.79	0.066	0.036	0.003	1–3,6
lump	4	2.18	5.89	10	4.56	0.066	0.25	0.001	1–3,6
mitt	3	1.74	6.83	31	4.76	0.053	0.177	0.023	1–3,6
mouse	3	2.6	4.94	14	4.83	0.053	0.011	0.03	1–3,6
mug	3	2.42	5.15	21	4.8	0.053	0.25	0.006	1–3,6
neck	3	3.29	3	14	5	0.05	0.073	0.03	1–3,6
noun	3	1.4	7.28	8	3.15	0.05	0.011	0.089	1–3,6
rag	3	2.27	5.22	24	4.67	0.048	0.065	0.006	1–3,6
ramp	4	2.01	7.28	10	4.69	0.048	0.065	0.001	1–3,6
skill	4	2.52	6.8	13	2.17	0.005	0.177	0.035	1–3,6

ton	3	2.51	7.42	30	4.17	0.09	0.25	0.089	1–3
tongue	3	3.05	4.47	15	4.93	0.09	0.25	0.034	1–3
trap	4	2.91	6.44	8	4.3	0.011	0.065	0.009	1–3
trip	4	3.37	4.22	11	3.71	0.011	0.177	0.009	1–3
type	3	3.28	6.58	17	3.3	0.09	0.039	0.009	1–3
van	3	2.96	5.2	15	4.72	0.025	0.065	0.089	1–3
view	3	3.15	5.63	10	3.21	0	0.027	0.514	1–3
voice	3	3.38	4.83	8	4.13	0.025	0.004	0.03	1–3
war	3	3.41	7.67	34	3.63	0.015	0.022	0.004	1–3
wave	3	2.83	4.26	28	4.55	0.015	0.053	0.006	1–3
wine	3	3.18	7.9	25	4.79	0.015	0.039	0.089	1–3
worth	3	3.52	7.08	11	1.89	0.015	0.082	0.002	1–3
wreck	3	2.72	7.59	26	4.07	0.048	0.073	0.03	1–3
soap	3	2.71	3.17	21	4.93	0.061	0.036	0.009	1–3,6
sound	4	3.61	3.72	13	3.7	0.061	0.011	0.007	1–3,6
spoon	4	2.43	2.5	11	4.96	0.007	0.027	0.089	1–3,6
tide	3	2.45	6.68	33	4.1	0.09	0.039	0.03	1–3,6
west	4	3.2	5.89	20	3.44	0.015	0.073	0.009	1–3,6
wheat	4	2.31	6.53	8	4.89	0.015	0.091	0.023	1–3,6
wolf	4	2.58	4.5	1	4.79	0.015	0.007	0	1–3,6
boot	3	2.57	3.89	33	4.96	0.038	0.027	0.023	1–3,6
broom	4	2.26	5.5	8	4.89	0.006	0.027	0.021	1–3,6
bulb	4	2.17	6.56	2	4.93	0.038	0.25	0	1–3,6
chief	3	3.14	7.53	12	4.26	0.011	0.091	0.003	1–3,6
dog	3	3.48	2.8	9	4.85	0.052	0.022	0.006	1–3,6
door	3	3.72	3.05	35	4.81	0.052	0.022	0.004	1–3,6
field	4	3.28	6.1	20	4.26	0.031	0.091	0.004	1–3,6
flake	4	1.84	5.95	8	4.36	0.005	0.053	0.03	1–3,6
flame	4	2.51	6.25	7	4.67	0.005	0.053	0.021	1–3,6
food	3	3.56	3.25	18	4.8	0.031	0.027	0.03	1–3,6
groom	4	2.4	7.78	8	4.54	0.007	0.027	0.021	1–3,6
hair	3	3.56	3.17	35	4.97	0.019	0.073	0.004	1–3,6
north	4	3.21	6.55	4	4.14	0.05	0.022	0	1–3,6
plum	4	2.05	5.5	11	4.85	0.006	0.25	0.021	1–3,6
rock	3	3.28	3.22	29	4.91	0.048	0.042	0.03	1–3,6
sash	3	1.59	7.67	19	4.67	0.061	0.065	0.004	1–3,6
seed	3	2.45	4.72	42	4.71	0.061	0.091	0.03	1–3,6
shack	3	2.29	6.15	26	4.93	0.027	0.065	0.03	1–3,6
sheet	3	2.67	5.33	31	4.93	0.027	0.091	0.023	1–3,6

snail	4	1.74	5.79	3	4.93	0.001	0.053	0.035	1–3,6
band	4	3.05	6.16	21	4.68	0.038	0.065	0.007	1–5
bank	4	3.2	6.44	13	4.78	0.038	0.065	0.002	1–5
base	3	3.01	6.42	24	3.86	0.038	0.053	0.03	1–5
beach	3	3.11	4.8	24	4.79	0.038	0.091	0.002	1–5
block	4	3.12	4.79	8	4.48	0.005	0.042	0.03	1–5
boss	3	3.37	6.16	13	3.83	0.038	0.042	0.03	1–5
bowl	3	2.83	4.26	38	4.87	0.038	0.036	0.035	1–5
bridge	4	3.02	5.58	6	4.97	0.006	0.177	0.003	1–5
bus	3	3.19	3.85	20	4.9	0.038	0.25	0.03	1–5
camp	4	3.04	5.78	8	4.35	0.054	0.065	0.001	1–5
cat	3	3.14	3.68	32	4.86	0.054	0.065	0.023	1–5
chest	4	3.07	5.05	17	4.93	0.011	0.073	0.009	1–5
church	3	3.13	5.15	6	4.9	0.011	0.082	0.002	1–5
class	4	3.39	4.95	11	3.85	0.006	0.065	0.03	1–5
crack	4	3.08	6.33	16	4.53	0.008	0.065	0.03	1–5
crime	4	3.25	7.67	11	3.03	0.008	0.039	0.021	1–5
dirt	3	2.91	3.83	14	4.86	0.052	0.082	0.023	1–5
doll	3	2.81	3.68	13	5	0.052	0.042	0.035	1–5
dress	4	3.34	4.05	4	4.93	0.004	0.073	0.03	1–5
face	3	3.75	3.75	20	4.87	0.031	0.053	0.03	1–5
faith	3	3.09	7.62	11	1.63	0.031	0.053	0.002	1–5
fear	3	3.28	4.79	40	2.57	0.031	0.177	0.004	1–5
floor	4	3.44	4.44	6	4.8	0.005	0.022	0.004	1–5
game	3	3.58	4.26	19	4.5	0.015	0.053	0.021	1–5
girl	3	3.79	4	15	4.85	0.015	0.082	0.035	1–5
glass	4	3.27	4.47	5	4.82	0.002	0.065	0.03	1–5
group	4	3.32	5.94	10	4.12	0.007	0.027	0.009	1–5
guard	4	3.2	6.25	15	4.04	0.015	0.042	0.002	1–5
guilt	4	2.76	7.05	16	1.93	0.015	0.177	0.001	1–5
hall	3	3.2	5.35	36	4.67	0.019	0.022	0.035	1–5
horn	4	2.77	4.84	18	5	0.019	0.022	0.001	1–5
badge	3	2.67	6.11	11	4.93	0.038	0.065	0.003	1–6
beast	4	2.74	5.74	15	4.63	0.038	0.091	0.009	1–6
bird	3	3.08	3.52	29	5	0.038	0.082	0.03	1–6
blade	4	2.59	6.72	9	4.93	0.005	0.053	0.03	1–6
bone	3	2.93	5.53	31	4.9	0.038	0.036	0.089	1–6
box	4	3.37	4.3	16	4.9	0.038	0.042	0.005	1–6
bread	4	2.96	3.58	19	4.92	0.006	0.073	0.03	1–6
test	4	3.32	6.26	18	3.93	0.09	0.073	0.009	1–3,6
horse	4	3.2	4.15	17	5	0.019	0.022	0	1–5
jail	3	3.23	5.74	31	4.83	0.017	0.053	0.035	1–5
jeep	3	2.38	6.32	14	4.8	0.017	0.091	0.009	1–5
juice	3	2.92	4.4	13	4.89	0.017	0.027	0.03	1–5
light	3	3.59	4.05	34	4.21	0.066	0.039	0.023	1–5
line	3	3.66	4.85	34	4.5	0.066	0.039	0.089	1–5
luck	3	3.59	6.53	23	1.33	0.066	0.25	0.03	1–5
mess	3	3.42	4.28	21	3.9	0.053	0.073	0.03	1–5
mud	3	2.7	4.05	21	4.86	0.053	0.25	0.03	1–5
night	3	3.87	3.61	33	4.52	0.05	0.039	0.023	1–5
page	3	3.06	5.16	16	4.9	0.04	0.053	0.003	1–5
ring	3	3.32	4.53	26	4.81	0.048	0.177	0.034	1–5
rose	3	2.92	6.11	38	4.9	0.048	0.036	0.033	1–5
scout	4	2.61	6.94	13	4	0.005	0.011	0.023	1–5
ship	3	3.07	5.33	20	4.87	0.027	0.177	0.009	1–5
snow	3	2.89	4.11	3	4.85	0.001	0.036	0.514	1–5
song	3	3.29	4.26	14	4.46	0.061	0.022	0.034	1–5
space	4	3.22	5.67	5	3.54	0.007	0.053	0.03	1–5
spot	4	3.31	5.39	11	4.21	0.007	0.042	0.023	1–5
star	4	3.26	3.89	8	4.69	0.013	0.042	0.004	1–5
steak	4	2.72	6.63	21	4.96	0.013	0.053	0.03	1–5
steam	4	2.61	6.26	7	4.5	0.013	0.091	0.021	1–5
style	4	3.04	7.58	9	2.67	0.013	0.039	0.035	1–5
suit	3	3.29	6.67	27	4.97	0.061	0.027	0.023	1–5
team	3	3.42	6	25	3.79	0.09	0.091	0.021	1–5
toast	4	3.04	4.67	11	4.93	0.09	0.036	0.009	1–5
tune	3	2.74	7.32	24	3.5	0.09	0.027	0.089	1–5
wall	3	3.32	3.79	25	4.86	0.015	0.022	0.035	1–5
wheel	4	2.93	4.4	6	4.86	0.015	0.091	0.035	1–5
wood	3	2.88	4.58	19	4.85	0.015	0.007	0.03	1–5
sheep	3	2.59	4.25	27	4.9	0.027	0.091	0.009	1–6
soup	3	2.88	5.37	17	4.72	0.061	0.027	0.009	1–6
sword	4	2.67	5.45	22	4.93	0.061	0.022	0.002	1–6
tank	4	2.83	7.17	12	4.8	0.09	0.065	0.002	1–6
tent	4	2.65	5.16	21	4.96	0.09	0.073	0.01	1–6
throat	4	3.09	5.09	7	4.97	0.001	0.036	0.023	1–6
thumb	3	2.61	4.42	14	4.96	0.005	0.25	0.021	1–6
trash	4	2.91	4.47	7	4.7	0.011	0.065	0.004	1–6

bride	4	2.81	5.1	17	4.63	0.006	0.039	0.03	1–6
cave	3	2.56	6.74	22	4.96	0.054	0.053	0.006	1–6
cheese	3	3.05	4.33	24	4.7	0.011	0.091	0.033	1–6
chin	3	2.67	4.22	21	4.89	0.011	0.177	0.089	1–6
clay	3	2.33	5.32	7	4.93	0.006	0.053	0.514	1–6
cloud	4	2.59	3.63	12	4.54	0.006	0.011	0.03	1–6
coat	3	3.1	3.58	26	4.97	0.054	0.036	0.023	1–6
corn	4	2.61	4.61	20	4.96	0.054	0.022	0.001	1–6
desk	4	3.15	5.56	3	4.87	0.052	0.073	0	1–6
dish	3	2.65	4.89	14	4.9	0.052	0.177	0.004	1–6
dream	4	3.45	4.88	3	2.6	0.004	0.091	0.021	1–6
frog	4	2.43	4.32	3	5	0.004	0.042	0.006	1–6
fruit	4	2.85	3.63	6	4.81	0.004	0.027	0.023	1–6
gate	3	2.96	5.32	29	4.96	0.015	0.053	0.023	1–6
gift	4	3.27	5.05	8	4.56	0.015	0.177	0.001	1–6
golf	4	2.78	7.16	2	4.52	0.015	0.042	0	1–6
goose	3	2.58	5.15	15	4.81	0.015	0.027	0.03	1–6
grass	4	2.74	3.94	12	4.93	0.007	0.065	0.03	1–6
guest	4	3.14	6.21	20	3.83	0.015	0.073	0.009	1–6
heart	4	3.66	5.17	15	4.52	0.019	0.042	0.001	1–6
hood	3	2.68	5.5	17	4.88	0.019	0.007	0.03	1–6
king	3	3.23	5.42	16	4.1	0.054	0.177	0.034	1–6
lamp	4	2.59	4	9	4.97	0.066	0.065	0.001	1–6
land	4	3.32	5.22	16	4.57	0.066	0.065	0.007	1–6
lunch	4	3.42	3.61	7	4.31	0.066	0.25	0.001	1–6
month	4	3.46	5.79	2	4.2	0.053	0.25	0	1–6
mouth	3	3.48	3.58	6	4.74	0.053	0.011	0.002	1–6
path	3	2.94	6.11	16	4.41	0.04	0.065	0.002	1–6
phone	3	3.65	4.11	27	4.86	0.031	0.036	0.089	1–6
pig	3	3.02	3.84	20	5	0.04	0.177	0.006	1–6
plate	4	2.97	3.84	13	4.77	0.006	0.053	0.023	1–6
prize	4	2.87	5.11	12	4.45	0.013	0.039	0.033	1–6
queen	4	3.08	4.42	1	4.45	0.004	0.091	0.089	1–6
rice	3	2.65	3.72	23	4.86	0.048	0.039	0.03	1–6
bleach	4	1.86	8	7	4.74	0.005	0.091	0.002	6
bliss	4	2.08	10.16	3	1.37	0.005	0.177	0.03	6
blouse	4	2.27	6.65	3	4.96	0.005	0.011	0.03	6
cask	4	0.7	13.12	6	0	0.054	0.065	0	6
chimp	4	1.76	7.17	5	4.96	0.011	0.177	0.001	6
tree	3	3.21	3.57	9	5	0.011	0.091	0.514	1–6
world	4	3.79	5.32	10	4.36	0.015	0.082	0.004	1–6
beef	3	2.81	6.58	19	4.74	0.038	0.091	0.003	1–6
bell	3	3.02	3.89	32	4.96	0.038	0.073	0.035	1–6
boat	3	3.16	3.84	31	4.93	0.038	0.036	0.023	1–6
book	3	3.51	3.68	20	4.9	0.038	0.007	0.03	1–6
brain	4	3.3	5.76	16	4.69	0.006	0.053	0.089	1–6
bush	3	2.61	4.9	7	4.9	0.038	0.007	0.004	1–6
cage	3	2.75	5.06	15	5	0.054	0.053	0.003	1–6
card	4	3.32	6.2	24	4.9	0.054	0.042	0.002	1–6
case	3	3.69	6.74	21	3.93	0.054	0.053	0.03	1–6
chain	3	2.87	5.22	28	4.55	0.011	0.053	0.089	1–6
clock	4	3.21	4.42	13	5	0.006	0.042	0.03	1–6
crown	4	2.56	7.8	10	4.81	0.008	0.011	0.089	1–6
deck	3	2.8	6.45	21	4.77	0.052	0.073	0.03	1–6
dust	4	2.91	5.06	11	4.4	0.052	0.25	0.009	1–6
flag	4	2.7	5.33	7	4.79	0.005	0.065	0.006	1–6
lawn	3	2.63	5.45	24	4.93	0.066	0.022	0.089	1–6
meal	3	3.02	4.74	33	4.62	0.053	0.091	0.035	1–6
moon	3	3.07	4.83	25	4.9	0.053	0.027	0.089	1–6
name	3	3.85	3.68	15	3.5	0.05	0.053	0.021	1–6
noise	3	3.1	4.5	10	3.52	0.05	0.004	0.033	1–6
pole	3	2.6	5.63	37	4.66	0.04	0.036	0.035	1–6
purse	3	2.83	5.53	19	4.9	0.04	0.082	0.03	1–6
road	3	3.44	4.55	39	4.75	0.048	0.036	0.03	1–6
sand	4	2.79	4.63	15	5	0.061	0.065	0.007	1–6
school	4	3.66	3.89	9	4.79	0.005	0.027	0.035	1–6
seat	3	3.39	4.58	40	4.78	0.061	0.091	0.023	1–6
shirt	3	3.15	3.53	15	4.94	0.027	0.082	0.023	1–6
south	3	3.23	6.06	2	3.84	0.061	0.011	0.002	1–6
stone	4	2.97	4.44	12	4.72	0.013	0.036	0.089	1–6
town	3	3.61	5.11	13	4.64	0.09	0.011	0.089	1–6
train	4	3.25	4	11	4.79	0.011	0.053	0.089	1–6
truck	4	3.19	3.79	6	4.84	0.011	0.25	0.03	1–6
germ	3	1.72	5.95	9	3.89	0.017	0.082	0.021	6
gist	4	1.68	11.3	13	1.81	0.017	0.177	0.009	6
glaze	4	1.62	8.42	8	4	0.002	0.053	0.033	6
globe	4	2.32	6.5	4	4.59	0.002	0.036	0.005	6
gloom	4	1.73	9	8	1.86	0.002	0.027	0.021	6

chore	3	1.68	5.79	29	3.78	0.011	0.022	0.004	6
chrome	4	1.74	8.32	8	4.5	0.008	0.036	0.021	6
clot	4	1.79	11.05	10	4.2	0.006	0.042	0.023	6
conch	4	1.11	10.5	4	4.52	0.054	0.042	0.001	6
cord	4	2.42	6	30	4.63	0.054	0.022	0.002	6
cove	3	1.91	9.63	16	4.57	0.054	0.036	0.006	6
cowl	3	0.78	13.53	19	3.96	0.054	0.011	0.035	6
creed	4	1.72	9.17	12	2.1	0.008	0.091	0.03	6
crumb	4	1.81	5.89	9	4.8	0.008	0.25	0.021	6
crust	5	2.1	5.95	7	4.59	0.008	0.25	0.009	6
cube	4	1.96	7.32	3	4.62	0	0.027	0.005	6
dirge	3	0.95	14.17	8	0	0.052	0.082	0.003	6
dome	3	1.88	8.44	21	4.74	0.052	0.036	0.021	6
dune	3	1.46	9.65	30	4.46	0.052	0.027	0.089	6
earth	2	3.37	5.37	7	4.8	0.123	0.082	0.002	6
fang	3	1.57	8.47	13	4.26	0.031	0.065	0.034	6
farm	4	2.86	3.85	5	4.59	0.031	0.042	0	6
fate	3	2.94	10.3	31	1.53	0.031	0.053	0.023	6
fleece	4	1.48	10.2	4	4.75	0.005	0.091	0.03	6
foam	3	2.15	6.15	16	4.85	0.031	0.036	0.021	6
foe	2	1.89	9.95	29	2.96	0.031	0.036	0.514	6
frond	5	0.3	11.5	2	0	0.004	0.042	0.007	6
glade	4	1.3	10.71	9	3.96	0.002	0.053	0.03	6
glove	4	2.53	4.3	2	4.97	0.002	0.25	0.006	6
gnome	3	1.4	8.94	20	4.59	0.05	0.036	0.021	6
gourd	4	1.34	10.65	21	4.86	0.015	0.022	0.002	6
grid	4	2.28	11.63	12	4.55	0.007	0.177	0.03	6
grime	4	1.32	8.72	7	3.85	0.007	0.039	0.021	6
grove	4	2.01	8.21	13	4.76	0.007	0.036	0.006	6
gust	4	1.34	8.12	11	3.85	0.015	0.25	0.009	6
haunch	4	0.48	12.57	4	3.23	0.019	0.022	0.001	6
hemp	4	1.23	12.94	4	4.85	0.019	0.073	0.001	6
hilt	4	1.45	14.4	16	0	0.019	0.177	0.001	6
hive	3	1.54	6.89	15	4.83	0.019	0.039	0.006	6
hoax	4	1.92	11.47	12	2.67	0.019	0.036	0.005	6
hound	4	2.23	5.74	14	4.48	0.019	0.011	0.007	6
husk	4	1.23	8.63	8	4.86	0.019	0.25	0	6
ice	2	3.3	3.86	15	4.89	0.123	0.039	0.03	6
jade	3	1.85	10.17	15	4.38	0.017	0.053	0.03	6
gloss	4	1.82	9.9	1	3.8	0.002	0.022	0.03	6
glue	3	2.34	4.67	10	4.65	0.002	0.027	0.514	6
goat	3	2.53	5.21	17	5	0.015	0.036	0.023	6
gown	3	2.42	6.16	10	4.61	0.015	0.011	0.089	6
grape	4	2.14	3.94	17	5	0.007	0.053	0.009	6
grasp	5	2.31	8.2	2	3.63	0.007	0.065	0	6
grouch	4	1.38	7.29	3	2.73	0.007	0.011	0.002	6
ground	5	3.36	4.89	9	4.77	0.007	0.011	0.007	6
grub	4	2.03	8.95	4	3.86	0.007	0.25	0.005	6
guild	4	1.73	12.94	17	2.96	0.015	0.177	0.004	6
guile	3	1.45	13.25	24	1.88	0.015	0.039	0.035	6
ham	3	2.53	4.1	33	4.9	0.019	0.065	0.021	6
hearth	4	1.49	10.05	6	4.38	0.019	0.042	0	6
hinge	4	1.45	7.95	4	4.57	0.019	0.177	0.001	6
hub	3	1.79	10.11	16	3.59	0.019	0.25	0.005	6
imp	3	1.11	11.47	1	0	0.123	0.177	0.001	6
itch	2	2.19	5.05	11	3.15	0.123	0.177	0.002	6
jug	3	1.88	5.83	14	4.96	0.017	0.25	0.006	6
junk	4	2.76	6.62	9	3.88	0.017	0.25	0.002	6
keg	3	2	13.41	7	4.75	0.054	0.073	0.006	6
laud	3	0.7	13	20	1.88	0.066	0.022	0.03	6
liege	3	1.38	14.64	17	0	0.066	0.091	0.003	6
lilt	4	0.85	12.5	16	0	0.066	0.177	0.001	6
lint	4	1.69	7.47	9	4.96	0.066	0.177	0.01	6
lobe	3	1.85	10.17	14	4.44	0.066	0.036	0.005	6
lodge	3	2.26	8.26	9	4	0.066	0.042	0.003	6
loft	4	2.02	11.05	9	4.32	0.066	0.022	0.001	6
mast	4	1.76	10.59	24	4.92	0.053	0.065	0.009	6
maze	3	1.9	7.11	38	4.45	0.053	0.053	0.033	6
mirth	3	1.08	12.07	11	2.83	0.053	0.082	0.002	6
moose	3	2.1	5.22	23	4.97	0.053	0.027	0.03	6
moth	3	1.85	5.74	8	4.69	0.053	0.022	0.002	6
mound	4	1.68	7.71	16	4.63	0.053	0.011	0.007	6
muck	3	1.91	8	23	3.77	0.053	0.25	0.03	6
mulch	4	1.2	9.22	2	4.59	0.053	0.25	0	6
mule	4	2.33	5.65	3	5	0.001	0.027	0.035	6
myth	3	2.4	8.61	14	2.17	0.053	0.177	0.002	6
nape	3	1.08	10.53	11	4.44	0.05	0.053	0.009	6
nerve	3	2.92	9.67	6	3.89	0.05	0.082	0.006	6

jinx	5	2.03	7.89	7	2.03	0.017	0.177	0	6
job	3	3.78	5.39	19	3.19	0.017	0.042	0.005	6
joust	4	1.38	10.63	7	4.03	0.017	0.011	0.009	6
joy	2	2.97	6.74	10	2.37	0.017	0.004	0.514	6
kelp	4	1.34	10.22	4	4.9	0.054	0.073	0	6
kiln	4	1.18	10.69	4	4.82	0.054	0.177	0	6
lake	3	2.9	4.61	31	4.88	0.066	0.053	0.03	6
lard	4	1.82	9.74	19	4.88	0.066	0.042	0.002	6
leaf	3	2.34	4.6	23	5	0.066	0.091	0.003	6
lore	3	1.46	12.56	35	2.35	0.066	0.022	0.004	6
moat	3	1.61	9.65	26	4.69	0.053	0.036	0.023	6
mold	4	2.19	9.26	21	4.85	0.053	0.036	0.004	6
peach	3	2.33	4.21	25	4.9	0.04	0.091	0.002	6
pill	3	2.59	6.06	39	4.72	0.04	0.177	0.035	6
prow	3	0.85	12.56	7	0	0.013	0.011	0.514	6
quake	4	1.69	8.15	5	3.64	0.004	0.053	0.03	6
raft	4	2.12	7.35	13	5	0.048	0.065	0.001	6
rind	4	0.95	8.95	18	4.48	0.048	0.039	0.007	6
salt	4	2.79	5.05	9	4.89	0.061	0.022	0.001	6
scab	4	1.42	6.65	4	4.71	0.005	0.065	0.005	6
scheme	4	2.47	9.65	6	2.41	0.005	0.091	0.021	6
scribe	5	0.95	10.1	1	4.04	0.001	0.039	0.005	6
scruff	5	1.11	10.32	1	3.15	0.001	0.25	0.003	6
shard	4	1.08	9.9	17	4.21	0.027	0.042	0.002	6
shelf	4	2.44	5.5	5	4.96	0.027	0.073	0	6
shield	4	2.41	6.5	10	4.66	0.027	0.091	0.004	6
shrine	4	1.97	9.63	1	4.47	0	0.039	0.089	6
shrub	4	1.15	8.06	2	4.92	0	0.25	0.005	6
skit	4	1.26	9.17	15	3.86	0.005	0.177	0.023	6
sky	3	3.09	4.17	4	4.45	0.005	0.039	0.514	6
sleet	4	1.49	11.72	18	4.78	0.004	0.091	0.023	6
sleeve	4	2.38	4.94	7	4.84	0.004	0.091	0.006	6
sleuth	4	1.04	10.82	3	3.07	0.004	0.027	0.002	6
smock	4	1.36	6.26	5	4.78	0.001	0.042	0.03	6
spawn	4	1.76	11.25	8	3.9	0.007	0.042	0.089	6
spice	4	2.21	6.78	8	4.54	0.007	0.039	0.03	6
spire	4	0.78	11.69	9	4	0.007	0.039	0.004	6
spore	4	0.85	10.76	9	4.14	0.007	0.022	0.004	6
storm	5	2.92	4.94	5	4.7	0.013	0.022	0	6
nest	4	2.57	5.11	16	4.86	0.05	0.073	0.009	6
node	3	1.32	12.17	28	4	0.05	0.036	0.03	6
nook	3	1.61	9.74	11	4.37	0.05	0.007	0.03	6
nose	3	3.33	2.95	28	4.89	0.05	0.036	0.033	6
notch	3	2.12	8.47	10	4.23	0.05	0.042	0.002	6
oak	2	2.27	7.35	13	4.82	0.123	0.036	0.03	6
oat	2	1.34	5.83	17	5	0.123	0.036	0.023	6
ooze	2	1.36	7.45	7	4	0.123	0.027	0.033	6
orb	3	1.83	11.17	2	4.12	0.123	0.022	0	6
plaid	4	1.86	8.56	9	4.23	0.006	0.065	0.03	6
plank	5	1.81	7.84	4	5	0.006	0.065	0.002	6
pond	4	2.32	5.16	9	4.9	0.04	0.042	0.007	6
porch	4	2.53	5.4	10	4.92	0.04	0.022	0	6
prong	4	0.9	10.78	2	4.73	0.013	0.022	0.034	6
prop	4	2.15	10.83	8	4.46	0.013	0.042	0.009	6
prose	4	1.59	12.06	15	0	0.013	0.036	0.033	6
pun	3	1.93	11.21	28	2.3	0.04	0.25	0.089	6
quail	4	1.61	9.72	4	4.65	0.004	0.053	0.035	6
quartz	6	1.4	9.28	7	4.72	0.004	0.022	0	6
quirk	4	1.28	11.39	4	2.25	0.004	0.082	0.03	6
quiz	4	2.19	7.05	6	4.43	0.004	0.177	0.033	6
realm	4	2.19	10.78	1	2.96	0.048	0.073	0	6
ridge	3	2.29	8.84	15	4.48	0.048	0.177	0.003	6
rink	4	1.77	8.69	17	4.56	0.048	0.177	0.002	6
rogue	3	2.06	11.48	19	2.32	0.048	0.036	0.006	6
sage	3	1.81	11.39	17	4.54	0.061	0.053	0.003	6
sauce	3	2.69	5.37	13	4.75	0.061	0.022	0.03	6
scalp	5	2.15	6.68	1	4.82	0.005	0.065	0	6
scar	4	2.49	5.68	8	4.74	0.005	0.042	0.004	6
scarf	5	2.22	5.68	2	4.97	0.005	0.042	0	6
scone	4	1.36	10.26	6	4.85	0.005	0.036	0.089	6
scroll	5	1.69	9.89	2	4.11	0.001	0.036	0.035	6
shade	3	2.4	6.37	23	3.38	0.027	0.053	0.03	6
shark	4	2.41	5.47	8	4.93	0.027	0.042	0.001	6
shawl	3	1.66	9.74	25	5	0.027	0.022	0.035	6
sheaf	3	0.48	11.07	19	0	0.027	0.091	0.003	6
shrimp	5	2.39	7.11	2	4.8	0	0.177	0.001	6
shroud	4	1.49	11.65	4	4.34	0	0.011	0.03	6
silk	4	2.54	7.39	7	4.7	0.061	0.177	0	6

stripe	5	1.73	4.05	6	4.72	0.005	0.039	0.009	6
thief	3	2.85	7.22	8	4.37	0.005	0.091	0.003	6
trunk	5	2.75	8.3	1	4.71	0.011	0.25	0.002	6
tube	3	2.66	5.5	10	4.82	0.09	0.027	0.005	6
tuft	4	0.48	9.87	9	3.85	0.09	0.25	0.001	6
vest	4	2.31	5.83	19	4.52	0.025	0.073	0.009	6
wasp	4	1.64	5.58	1	4.96	0.015	0.042	0	6
wing	3	2.76	4.79	20	4.86	0.015	0.177	0.034	6
wisp	4	1.2	9.71	2	2.87	0.015	0.177	0	6
barb	4	1.49	9.88	7	4.52	0.038	0.042	0	6
barn	4	2.6	4.5	11	4.79	0.038	0.042	0.001	6
beard	4	2.6	4.84	18	4.96	0.038	0.177	0.002	6
blaze	4	1.89	7.47	8	4.28	0.005	0.053	0.033	6
blimp	5	1.61	6.63	0	4.76	0.005	0.177	0.001	6
blurb	4	0.78	11.56	4	3.04	0.005	0.082	0.005	6
booth	3	2.62	7.16	15	4.42	0.038	0.027	0.002	6
branch	5	2.57	5.11	3	4.9	0.006	0.065	0.001	6
brawn	4	1.38	10.06	10	2.79	0.006	0.022	0.089	6
brick	4	2.41	6.43	12	4.83	0.006	0.177	0.03	6
brine	4	1.11	14.25	9	4.24	0.006	0.039	0.089	6
brink	5	2.05	10.47	6	2.48	0.006	0.177	0.002	6
bronze	5	2	10	4	4.47	0.006	0.042	0.01	6
brow	3	1.9	9.82	6	4.39	0.006	0.011	0.514	6
cairn	0	0.48	15.27	0	0	0.054	0.073	0.001	6
cake	3	3.08	3.26	24	4.81	0.054	0.053	0.03	6
carp	4	1.36	11.42	10	4.62	0.054	0.042	0	6
chair	3	3.19	3.43	27	4.58	0.011	0.073	0.004	6
chalk	3	2.16	4.47	13	4.9	0.011	0.042	0.03	6
champ	4	2.44	7.53	9	3.29	0.011	0.065	0.001	6
child	4	3.49	5.15	6	4.78	0.011	0.039	0.004	6
clam	4	2.2	7.37	12	4.89	0.006	0.065	0.021	6
cloth	4	2.37	5.3	6	4.9	0.006	0.022	0.002	6
coal	3	2.23	6.65	38	4.66	0.054	0.036	0.035	6
cog	3	1.34	11.33	12	4.46	0.054	0.022	0.006	6
cone	3	2.06	4.67	31	4.86	0.054	0.036	0.089	6
crab	4	2.28	5.28	9	4.9	0.008	0.065	0.005	6
cub	3	1.84	5.4	17	4.67	0.054	0.25	0.005	6
curd	3	1.23	10.17	21	4.19	0.054	0.082	0.03	6
cyst	4	1.26	11.81	14	4.23	0.061	0.177	0.009	6
sink	4	2.81	4.47	18	4.74	0.061	0.177	0.002	6
skunk	5	2.03	5.32	3	4.88	0.005	0.25	0.002	6
slab	4	1.83	8.53	11	4.29	0.004	0.065	0.005	6
slice	4	2.55	5.69	6	3.85	0.004	0.039	0.03	6
sling	4	2.02	8.67	19	4.52	0.004	0.177	0.034	6
sloop	4	0.95	11.07	12	0	0.004	0.027	0.009	6
slot	4	2.31	6.85	20	4.45	0.004	0.042	0.023	6
sludge	4	1.38	8.8	8	4.23	0.004	0.25	0.003	6
slug	4	2.28	6	10	4.64	0.004	0.25	0.006	6
snob	4	1.94	9.32	4	2.7	0.001	0.042	0.005	6
soot	3	1.56	7.2	12	4.61	0.061	0.007	0.023	6
speck	4	1.8	8.48	12	4.46	0.007	0.073	0.03	6
sphere	4	1.68	8.26	4	4.44	0	0.177	0.004	6
spine	4	2.38	7.35	15	4.88	0.007	0.039	0.089	6
spleen	5	1.96	11.35	0	4.7	0	0.091	0.089	6
spouse	4	1.91	9.94	3	3.85	0.007	0.011	0.03	6
sprig	5	1.04	10.31	1	4.29	0	0.177	0.006	6
spud	4	1.32	9.05	7	4.76	0.007	0.25	0.03	6
squad	5	2.75	9.55	4	3.65	0.001	0.042	0.03	6
squid	5	1.88	7.32	1	4.71	0.001	0.177	0.03	6
squire	5	1.56	11	2	4.16	0.001	0.039	0.004	6
stance	5	1.87	10	6	4.04	0.013	0.065	0.004	6
stove	4	2.45	4.32	6	4.96	0.013	0.036	0.006	6
tact	4	1.74	10.42	20	1.76	0.09	0.065	0.003	6
teal	3	1.18	9.42	36	4.07	0.09	0.091	0.035	6
thrift	5	1.32	10.47	1	2.36	0.001	0.177	0.001	6
tinge	4	0.95	10.28	5	2.65	0.09	0.177	0.001	6
tint	4	1.2	10	11	4	0.09	0.177	0.01	6
tithe	3	0.3	11.5	14	0	0.09	0.039	0	6
tome	3	1.43	13.2	23	0	0.09	0.036	0.021	6
torch	4	2.29	7.84	5	4.76	0.09	0.022	0	6
tract	5	1.76	13.67	12	3.46	0.011	0.065	0.003	6
tribe	4	2.29	8.17	6	4.14	0.011	0.039	0.005	6
trough	4	1.77	8.41	1	4.17	0.011	0.022	0.003	6
trout	4	2.05	8.56	6	4.72	0.011	0.011	0.023	6
truce	4	2.18	8.8	8	2.47	0.011	0.027	0.03	6
tub	3	2.66	3.95	16	4.64	0.09	0.25	0.005	6
tusk	4	1.08	7.67	6	4.76	0.09	0.25	0	6
twang	4	0.9	11.89	0	2.96	0.001	0.042	0.034	6

daub	3	0.7	10.17	6	0	0.052	0.022	0.005	6
day	2	3.89	3.5	25	3.92	0.052	0.053	0.514	6
deed	3	2.44	9.72	28	3.86	0.052	0.091	0.03	6
dent	4	2.1	7.33	19	4.63	0.052	0.073	0.01	6
disk	4	2.11	7.47	8	4.8	0.052	0.177	0	6
dock	3	2.49	8.22	24	4.64	0.052	0.042	0.03	6
drawl	4	0.7	11.17	8	3.08	0.004	0.022	0.035	6
dusk	4	1.7	8.74	9	4.24	0.052	0.25	0	6
eel	2	1.72	6.47	16	4.69	0.123	0.091	0.035	6
elm	3	1.68	9.06	3	4.69	0.123	0.073	0	6
farce	4	1.84	12.12	5	2.32	0.031	0.042	0	6
fern	3	1.46	8.67	20	5	0.031	0.082	0.089	6
feud	4	1.76	10.33	7	3.18	0	0.027	0.03	6
fig	3	1.66	8.06	17	4.97	0.031	0.177	0.006	6
fin	3	1.84	7.3	33	4.76	0.031	0.177	0.089	6
fleck	4	0.7	11.41	10	3.75	0.005	0.073	0.03	6
flock	4	2.21	7.18	10	4.67	0.005	0.042	0.03	6
fluke	4	1.82	10.1	7	2.34	0.005	0.027	0.03	6
flute	4	1.86	8.47	9	5	0.005	0.027	0.023	6
fog	3	2.33	6.21	5	4.66	0.031	0.042	0.006	6
font	4	1.54	11.5	5	3.93	0.031	0.042	0.01	6
fork	4	2.5	3.63	13	4.9	0.031	0.022	0.001	6
garb	4	1.18	12.05	2	4.19	0.015	0.042	0	6
gauze	3	1.76	9.32	19	4.62	0.015	0.022	0.033	6
gel	3	1.83	7.21	19	4.72	0.017	0.073	0.035	6
gem	3	1.96	7.68	10	4.88	0.017	0.073	0.021	6
tweed	4	1.51	12.11	2	4.81	0.001	0.091	0.03	6
twig	4	1.8	6.28	7	4.75	0.001	0.177	0.006	6
twine	4	1.34	9.38	5	4.03	0.001	0.039	0.089	6
valve	4	2.14	10.78	4	4.83	0.025	0.065	0	6
vase	3	2.09	7.89	20	5	0.025	0.053	0.03	6
verb	3	1.58	8	4	2.85	0.025	0.082	0.005	6
verse	3	2.23	8.17	13	3.19	0.025	0.082	0.03	6
vine	3	1.91	6.95	22	4.86	0.025	0.039	0.089	6
waltz	5	2.26	10.37	8	4.52	0.015	0.022	0	6
wax	4	2.44	6	21	4.97	0.015	0.065	0.005	6
welt	4	0.95	9.22	17	4.43	0.015	0.073	0.001	6
whale	4	2.37	5.47	5	4.96	0.015	0.053	0.035	6
wig	3	2.39	5.63	18	4.72	0.015	0.177	0.006	6
wool	3	2.12	8.06	12	4.86	0.015	0.007	0.035	6
word	3	3.71	4.42	23	3.56	0.015	0.082	0.03	6
wreath	3	1.63	7.06	17	4.93	0.048	0.091	0.002	6
yacht	3	2.3	10.06	15	4.97	0.012	0.042	0.023	6
yarn	4	1.79	6.61	3	4.93	0.012	0.042	0.001	6
year	3	3.7	5.24	29	3.25	0.012	0.177	0.004	6
yeast	4	1.59	9.53	6	4.72	0.012	0.091	0.009	6
youth	3	2.78	6.89	7	3.28	0.012	0.027	0.002	6
zeal	3	1.48	13.12	17	2.33	0.015	0.091	0.035	6
zest	4	1.46	10.56	14	2.27	0.015	0.073	0.009	6
zinc	4	1.51	12.47	10	4.4	0.015	0.177	0.002	6
zone	3	2.79	8.79	14	3.07	0.015	0.036	0.089	6

AI.2 Non-Words

The following table displays the list of non-words used in the experiments in this thesis. The column names ***1-6*** are described below:

1. Non-word transcribed using the ARPABET phonetic alphabet used by the CMUDict Pronunciation Dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict)
2. Number of phonemes in the non-word

3. **3.** Frequency of the onset of the non-word, rounded to three decimal places (described in Section 2.2.2)

4. **4.** As above, but for the vowel of the non-word

5. **5.** As above, but for the coda of the non-word

6. **6.** List of experiment numbers the non-word was used in (either 1 – 3, 1 – 5, or 4 – 5)

1	2	3	4	5	6
ahft	3	0.123	0.25	0.001	1–3
baepth	4	0.038	0.065	0	1–3
bays	3	0.038	0.039	0.03	1–3
bihm	3	0.038	0.177	0.021	1–3
bihp	3	0.038	0.177	0.009	1–3
blay	3	0.005	0.039	0.514	1–3
blayl	4	0.005	0.039	0.035	1–3
bleyt	4	0.005	0.053	0.023	1–3
bluhk	4	0.005	0.007	0.03	1–3
bluhn	4	0.005	0.007	0.089	1–3
chaab	3	0.011	0.042	0.005	1–3
chaengk	4	0.011	0.065	0.002	1–3
chaorn	4	0.011	0.022	0.001	1–3
chawn	3	0.011	0.011	0.089	1–3
chert	3	0.011	0.082	0.023	1–3
cheyd	3	0.011	0.053	0.03	1–3
cheyl	3	0.011	0.053	0.035	1–3
daapt	4	0.052	0.042	0.002	1–3
daarf	4	0.052	0.042	0	1–3
daask	4	0.052	0.042	0	1–3
daeg	3	0.052	0.065	0.006	1–3
daek	3	0.052	0.065	0.03	1–3
daeks	4	0.052	0.065	0.005	1–3
dahft	4	0.052	0.25	0.001	1–3
dayt	3	0.052	0.039	0.023	1–3
deyk	3	0.052	0.053	0.03	1–3
deymp	4	0.052	0.053	0.001	1–3
deyth	3	0.052	0.053	0.002	1–3
diht	3	0.052	0.177	0.023	1–3
diyst	4	0.052	0.091	0.009	1–3
driyz	4	0.004	0.091	0.033	1–3
duwt	3	0.052	0.027	0.023	1–3
dwiyr	4	0	0.091	0.004	1–3
faajh	3	0.031	0.042	0.003	1–3
faesh	3	0.031	0.065	0.004	1–3
faht	3	0.031	0.25	0.023	1–3
fawt	3	0.031	0.011	0.023	1–3
feht	3	0.031	0.073	0.023	1–3
luhb	3	0.066	0.007	0.005	1–3
luhd	3	0.066	0.007	0.03	1–3
luhlb	4	0.066	0.007	0	1–3
luhlf	4	0.066	0.007	0	1–3
maat	3	0.053	0.042	0.023	1–3
mawt	3	0.053	0.011	0.023	1–3
mayd	3	0.053	0.039	0.03	1–3
meyrk	4	0.053	0.053	0.001	1–3
mihp	3	0.053	0.177	0.009	1–3
muhgth	4	0.053	0.007	0	1–3
muwm	3	0.053	0.027	0.021	1–3
muwp	3	0.053	0.027	0.009	1–3
naap	3	0.05	0.042	0.009	1–3
naef	3	0.05	0.065	0.003	1–3
naeng	3	0.05	0.065	0.034	1–3
naes	3	0.05	0.065	0.03	1–3
nahk	3	0.05	0.25	0.03	1–3
naol	3	0.05	0.022	0.035	1–3
nayd	3	0.05	0.039	0.03	1–3
nerth	3	0.05	0.082	0.002	1–3
neyjh	3	0.05	0.053	0.003	1–3
neyn	3	0.05	0.053	0.089	1–3
neynt	4	0.05	0.053	0.01	1–3
neyz	3	0.05	0.053	0.033	1–3
niym	3	0.05	0.091	0.021	1–3
niyn	3	0.05	0.091	0.089	1–3
niynz	4	0.05	0.091	0.01	1–3
nowk	3	0.05	0.036	0.03	1–3
nuhd	3	0.05	0.007	0.03	1–3
nuwp	3	0.05	0.027	0.009	1–3
paeg	3	0.04	0.065	0.006	1–3
paengk	4	0.04	0.065	0.002	1–3
pays	3	0.04	0.039	0.03	1–3
pehch	3	0.04	0.073	0.002	1–3
peyb	3	0.04	0.053	0.005	1–3
peykth	4	0.04	0.053	0	1–3
peym	3	0.04	0.053	0.021	1–3
plawl	4	0.006	0.011	0.035	1–3

fiyk	3	0.031	0.091	0.03	1–3
fiyn	3	0.031	0.091	0.089	1–3
fiyp	3	0.031	0.091	0.009	1–3
fluhk	4	0.005	0.007	0.03	1–3
fowd	3	0.031	0.036	0.03	1–3
fowst	4	0.031	0.036	0.009	1–3
foyd	3	0.031	0.004	0.03	1–3
fruw	3	0.004	0.027	0.514	1–3
fuhd	3	0.031	0.007	0.03	1–3
fuhlf	4	0.031	0.007	0	1–3
fuwn	3	0.031	0.027	0.089	1–3
gaam	3	0.015	0.042	0.021	1–3
gaech	3	0.015	0.065	0.002	1–3
gaen	3	0.015	0.065	0.089	1–3
gaeth	3	0.015	0.065	0.002	1–3
gahb	3	0.015	0.25	0.005	1–3
gayn	3	0.015	0.039	0.089	1–3
gehd	3	0.015	0.073	0.03	1–3
geyk	3	0.015	0.053	0.03	1–3
geys	3	0.015	0.053	0.03	1–3
gihng	3	0.015	0.177	0.034	1–3
gihngk	4	0.015	0.177	0.002	1–3
giyd	3	0.015	0.091	0.03	1–3
giyl	3	0.015	0.091	0.035	1–3
giym	3	0.015	0.091	0.021	1–3
giyn	3	0.015	0.091	0.089	1–3
giyp	3	0.015	0.091	0.009	1–3
glay	3	0.002	0.039	0.514	1–3
glehs	4	0.002	0.073	0.03	1–3
gler	3	0.002	0.082	0.514	1–3
gliyr	4	0.002	0.091	0.004	1–3
grawd	4	0.007	0.011	0.03	1–3
grayd	4	0.007	0.039	0.03	1–3
grihr	4	0.007	0.177	0.004	1–3
guhbth	4	0.015	0.007	0	1–3
guhk	3	0.015	0.007	0.03	1–3
guhkt	4	0.015	0.007	0.003	1–3
hhaep	3	0.019	0.065	0.009	1–3
hhehst	4	0.019	0.073	0.009	1–3
hheht	3	0.019	0.073	0.023	1–3
hheyd	3	0.019	0.053	0.03	1–3
hheyk	3	0.019	0.053	0.03	1–3
hheym	3	0.019	0.053	0.021	1–3
hheyv	3	0.019	0.053	0.006	1–3
hhihlm	4	0.019	0.177	0	1–3
hhihng	3	0.019	0.177	0.034	1–3
hhiypth	4	0.019	0.091	0	1–3
hhowk	3	0.019	0.036	0.03	1–3
hhoys	3	0.019	0.004	0.03	1–3
hhuws	3	0.019	0.027	0.03	1–3
iykth	3	0.123	0.091	0	1–3
jhaart	4	0.017	0.042	0.001	1–3
jhaelk	4	0.017	0.065	0	1–3
plehl	4	0.006	0.073	0.035	1–3
pliyr	4	0.006	0.091	0.004	1–3
plown	4	0.006	0.036	0.089	1–3
pluw	3	0.006	0.027	0.514	1–3
puhm	3	0.04	0.007	0.021	1–3
raart	4	0.048	0.042	0.001	1–3
raol	3	0.048	0.022	0.035	1–3
raorb	4	0.048	0.022	0	1–3
rehl	3	0.048	0.073	0.035	1–3
reht	3	0.048	0.073	0.023	1–3
reyg	3	0.048	0.053	0.006	1–3
saemp	4	0.061	0.065	0.001	1–3
saht	3	0.061	0.25	0.023	1–3
sawrt	4	0.061	0.011	0.001	1–3
sayl	3	0.061	0.039	0.035	1–3
sehlk	4	0.061	0.073	0	1–3
sert	3	0.061	0.082	0.023	1–3
seyr	3	0.061	0.053	0.004	1–3
shaag	3	0.027	0.042	0.006	1–3
shahst	4	0.027	0.25	0.009	1–3
shehk	3	0.027	0.073	0.03	1–3
shihns	4	0.027	0.177	0.004	1–3
shiych	3	0.027	0.091	0.002	1–3
skehs	4	0.005	0.073	0.03	1–3
skuw	3	0.005	0.027	0.514	1–3
slaak	4	0.004	0.042	0.03	1–3
slihg	4	0.004	0.177	0.006	1–3
sluhch	4	0.004	0.007	0.002	1–3
smawn	4	0.001	0.011	0.089	1–3
sney	3	0.001	0.053	0.514	1–3
sniyd	4	0.001	0.091	0.03	1–3
sniyl	4	0.001	0.091	0.035	1–3
spaeg	4	0.007	0.065	0.006	1–3
spowf	4	0.007	0.036	0.003	1–3
spuhd	4	0.007	0.007	0.03	1–3
staes	4	0.013	0.065	0.03	1–3
stiyf	4	0.013	0.091	0.003	1–3
suhf	3	0.061	0.007	0.003	1–3
suhls	4	0.061	0.007	0	1–3
suwm	3	0.061	0.027	0.021	1–3
taab	3	0.09	0.042	0.005	1–3
taark	4	0.09	0.042	0.001	1–3
thaorg	4	0.005	0.022	0	1–3
theyz	3	0.005	0.053	0.033	1–3
thriys	4	0.001	0.091	0.03	1–3
tihg	3	0.09	0.177	0.006	1–3
toys	3	0.09	0.004	0.03	1–3
traeth	4	0.011	0.065	0.002	1–3
traor	4	0.011	0.022	0.004	1–3
trawd	4	0.011	0.011	0.03	1–3
trow	3	0.011	0.036	0.514	1–3
vaag	3	0.025	0.042	0.006	1–3
vaak	3	0.025	0.042	0.03	1–3

jhaend	4	0.017	0.065	0.007	1–3
jhaesh	3	0.017	0.065	0.004	1–3
jhahng	3	0.017	0.25	0.034	1–3
jhayl	3	0.017	0.039	0.035	1–3
jhayt	3	0.017	0.039	0.023	1–3
jheyjh	3	0.017	0.053	0.003	1–3
jheyrp	4	0.017	0.053	0	1–3
jhihkt	4	0.017	0.177	0.003	1–3
jhowt	3	0.017	0.036	0.023	1–3
jhreyz	4	0	0.053	0.033	1–3
jhuwp	3	0.017	0.027	0.009	1–3
kaark	4	0.054	0.042	0.001	1–3
kehk	3	0.054	0.073	0.03	1–3
kehs	3	0.054	0.073	0.03	1–3
kehsk	4	0.054	0.073	0	1–3
kehst	4	0.054	0.073	0.009	1–3
kiych	3	0.054	0.091	0.002	1–3
kiyt	3	0.054	0.091	0.023	1–3
kiyv	3	0.054	0.091	0.006	1–3
klaos	4	0.006	0.022	0.03	1–3
kleys	4	0.006	0.053	0.03	1–3
kuhg	3	0.054	0.007	0.006	1–3
kuhmp	4	0.054	0.007	0.001	1–3
kuhngk	4	0.054	0.007	0.002	1–3
kuws	3	0.054	0.027	0.03	1–3
laar	3	0.066	0.042	0.004	1–3
laelp	4	0.066	0.065	0	1–3
lahn	3	0.066	0.25	0.089	1–3
laors	4	0.066	0.022	0	1–3
lawp	3	0.066	0.011	0.009	1–3
laych	3	0.066	0.039	0.002	1–3
leyjh	3	0.066	0.053	0.003	1–3
leyrn	4	0.066	0.053	0.001	1–3
bihngk	4	0.038	0.177	0.002	1–5
braem	4	0.006	0.065	0.021	1–5
briyl	4	0.006	0.091	0.035	1–5
daar	3	0.052	0.042	0.004	1–5
daard	4	0.052	0.042	0.002	1–5
draed	4	0.004	0.065	0.03	1–5
driyt	4	0.004	0.091	0.023	1–5
faek	3	0.031	0.065	0.03	1–5
fayz	3	0.031	0.039	0.033	1–5
frehl	4	0.004	0.073	0.035	1–5
frehs	4	0.004	0.073	0.03	1–5
freyk	4	0.004	0.053	0.03	1–5
frihsh	4	0.004	0.177	0.004	1–5
gaend	4	0.015	0.065	0.007	1–5
grehl	4	0.007	0.073	0.035	1–5
grihk	4	0.007	0.177	0.03	1–5
grihsh	4	0.007	0.177	0.004	1–5
jhehk	3	0.017	0.073	0.03	1–5
jhihng	3	0.017	0.177	0.034	1–5
kihnt	4	0.054	0.177	0.01	1–5
vaap	3	0.025	0.042	0.009	1–3
vaeb	3	0.025	0.065	0.005	1–3
vaebd	4	0.025	0.065	0	1–3
vaep	3	0.025	0.065	0.009	1–3
vayt	3	0.025	0.039	0.023	1–3
vehs	3	0.025	0.073	0.03	1–3
veyk	3	0.025	0.053	0.03	1–3
vihk	3	0.025	0.177	0.03	1–3
vowz	3	0.025	0.036	0.033	1–3
vuhmp	4	0.025	0.007	0.001	1–3
vuwn	3	0.025	0.027	0.089	1–3
waef	3	0.015	0.065	0.003	1–3
waelf	4	0.015	0.065	0	1–3
waelv	4	0.015	0.065	0	1–3
wahng	3	0.015	0.25	0.034	1–3
wahst	4	0.015	0.25	0.009	1–3
weylth	4	0.015	0.053	0	1–3
wown	3	0.015	0.036	0.089	1–3
yaaks	4	0.012	0.042	0.005	1–3
yaeg	3	0.012	0.065	0.006	1–3
yaend	4	0.012	0.065	0.007	1–3
yayn	3	0.012	0.039	0.089	1–3
yiht	3	0.012	0.177	0.023	1–3
yown	3	0.012	0.036	0.089	1–3
zaed	3	0.015	0.065	0.03	1–3
zaet	3	0.015	0.065	0.023	1–3
zahg	3	0.015	0.25	0.006	1–3
zehl	3	0.015	0.073	0.035	1–3
ziyr	3	0.015	0.091	0.004	1–3
nayz	3	0.05	0.039	0.033	1–5
nihng	3	0.05	0.177	0.034	1–5
nihngk	4	0.05	0.177	0.002	1–5
niyk	3	0.05	0.091	0.03	1–5
pehs	3	0.04	0.073	0.03	1–5
plihn	4	0.006	0.177	0.089	1–5
pliyn	4	0.006	0.091	0.089	1–5
praen	4	0.013	0.065	0.089	1–5
priyf	4	0.013	0.091	0.003	1–5
priyl	4	0.013	0.091	0.035	1–5
priym	4	0.013	0.091	0.021	1–5
priyz	4	0.013	0.091	0.033	1–5
prowd	4	0.013	0.036	0.03	1–5
pruw	3	0.013	0.027	0.514	1–5
raes	3	0.048	0.065	0.03	1–5
rihl	3	0.048	0.177	0.035	1–5
rihn	3	0.048	0.177	0.089	1–5
riyn	3	0.048	0.091	0.089	1–5
riyst	4	0.048	0.091	0.009	1–5
rowk	3	0.048	0.036	0.03	1–5
saard	4	0.061	0.042	0.002	1–5
shihng	3	0.027	0.177	0.034	1–5
skawn	4	0.005	0.011	0.089	1–5
sliyd	4	0.004	0.091	0.03	1–5

klay	3	0.006	0.039	0.514	1–5
kleyt	4	0.006	0.053	0.023	1–5
klow	3	0.006	0.036	0.514	1–5
krihng	4	0.008	0.177	0.034	1–5
kweyz	4	0.004	0.053	0.033	1–5
lehk	3	0.066	0.073	0.03	1–5
blayn	4	0.005	0.039	0.089	4–5
blihjh	4	0.005	0.177	0.003	4–5
blihn	4	0.005	0.177	0.089	4–5
blihv	4	0.005	0.177	0.006	4–5
braek	4	0.006	0.065	0.03	4–5
brehr	4	0.006	0.073	0.004	4–5
brihn	4	0.006	0.177	0.089	4–5
brihv	4	0.006	0.177	0.006	4–5
briyn	4	0.006	0.091	0.089	4–5
drehr	4	0.004	0.073	0.004	4–5
drihjh	4	0.004	0.177	0.003	4–5
drihk	4	0.004	0.177	0.03	4–5
drihng	4	0.004	0.177	0.034	4–5
flahm	4	0.005	0.25	0.021	4–5
flayn	4	0.005	0.039	0.089	4–5
fleyn	4	0.005	0.053	0.089	4–5
flihjh	4	0.005	0.177	0.003	4–5
flihsh	4	0.005	0.177	0.004	4–5
flihv	4	0.005	0.177	0.006	4–5
freyn	4	0.004	0.053	0.089	4–5
frihk	4	0.004	0.177	0.03	4–5
frihp	4	0.004	0.177	0.009	4–5
glaek	4	0.002	0.065	0.03	4–5
glihjh	4	0.002	0.177	0.003	4–5
glihk	4	0.002	0.177	0.03	4–5
glihp	4	0.002	0.177	0.009	4–5
glihsh	4	0.002	0.177	0.004	4–5
glihv	4	0.002	0.177	0.006	4–5
graek	4	0.007	0.065	0.03	4–5
grahn	4	0.007	0.25	0.089	4–5
grehr	4	0.007	0.073	0.004	4–5
grihjh	4	0.007	0.177	0.003	4–5
grihng	4	0.007	0.177	0.034	4–5
klahm	4	0.006	0.25	0.021	4–5
klihjh	4	0.006	0.177	0.003	4–5
klihv	4	0.006	0.177	0.006	4–5
krihjh	4	0.008	0.177	0.003	4–5
krihn	4	0.008	0.177	0.089	4–5
krihsh	4	0.008	0.177	0.004	4–5
krihv	4	0.008	0.177	0.006	4–5
kriyn	4	0.008	0.091	0.089	4–5
kwehr	4	0.004	0.073	0.004	4–5
kwihn	4	0.004	0.177	0.089	4–5
kwihsh	4	0.004	0.177	0.004	4–5
stihp	4	0.013	0.177	0.009	1–5
stiyz	4	0.013	0.091	0.033	1–5
traak	4	0.011	0.042	0.03	1–5
traet	4	0.011	0.065	0.023	1–5
trowz	4	0.011	0.036	0.033	1–5
vihng	3	0.025	0.177	0.034	1–5
lihjh	3	0.066	0.177	0.003	4–5
lihng	3	0.066	0.177	0.034	4–5
mihsh	3	0.053	0.177	0.004	4–5
plihng	4	0.006	0.177	0.034	4–5
plihp	4	0.006	0.177	0.009	4–5
plihsh	4	0.006	0.177	0.004	4–5
plihv	4	0.006	0.177	0.006	4–5
prihn	4	0.013	0.177	0.089	4–5
prihng	4	0.013	0.177	0.034	4–5
prihp	4	0.013	0.177	0.009	4–5
prihsh	4	0.013	0.177	0.004	4–5
rihsh	3	0.048	0.177	0.004	4–5
skihng	4	0.005	0.177	0.034	4–5
slahn	4	0.004	0.25	0.089	4–5
slayn	4	0.004	0.039	0.089	4–5
slihjh	4	0.004	0.177	0.003	4–5
slihv	4	0.004	0.177	0.006	4–5
sliyn	4	0.004	0.091	0.089	4–5
snahm	4	0.001	0.25	0.021	4–5
snihn	4	0.001	0.177	0.089	4–5
snihsh	4	0.001	0.177	0.004	4–5
spihng	4	0.007	0.177	0.034	4–5
stihn	4	0.013	0.177	0.089	4–5
stihv	4	0.013	0.177	0.006	4–5
stiyn	4	0.013	0.091	0.089	4–5
straek	5	0.005	0.065	0.03	4–5
strayn	5	0.005	0.039	0.089	4–5
strehr	5	0.005	0.073	0.004	4–5
strihjh	5	0.005	0.177	0.003	4–5
strihsh	5	0.005	0.177	0.004	4–5
strihv	5	0.005	0.177	0.006	4–5
striyn	5	0.005	0.091	0.089	4–5
strown	5	0.005	0.036	0.089	4–5
swihk	4	0.002	0.177	0.03	4–5
swihn	4	0.002	0.177	0.089	4–5
tihv	3	0.09	0.177	0.006	4–5
trahm	4	0.011	0.25	0.021	4–5
trehr	4	0.011	0.073	0.004	4–5
trihjh	4	0.011	0.177	0.003	4–5
trihn	4	0.011	0.177	0.089	4–5
trihng	4	0.011	0.177	0.034	4–5
trihv	4	0.011	0.177	0.006	4–5
trown	4	0.011	0.036	0.089	4–5
vihjh	3	0.025	0.177	0.003	4–5

BIBLIOGRAPHY

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.

Baddeley, A. (1992). Working memory: The interface between memory and cognition. *Journal of cognitive neuroscience, 4*(3), 281-288.

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior research methods, 39*(3), 445-459.

Bartoń, K., 2018. MuMIn: Multi-Model Inference. R package version 1.42.1. URL CRAN.R-project.org/package=MuMIn

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967.*

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 317.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.17. URL http://www.praat.org/

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2009). The neural basis of the speed–accuracy tradeoff. *Trends in neurosciences*, *33*(1), 10-16.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & psychophysics, 61*(2), 206-219.

Brennan, J., Lignos, C., Embick, D., & Roberts, T. P. (2014). Spectro-temporal correlates of lexical access during auditory lexical decision. *Brain and language, 133,* 39-46.

Brown, H. (2011). *Talker-specificity and lexical competition effects during word learning* (Doctoral dissertation, University of York).

Brown, H., & Gaskell, M. G. (2014). The time-course of talker-specificity and lexical competition effects during word learning. *Language, Cognition and Neuroscience, 29*(9), 1163-1179.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods, 41*(4), 977-990.

Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904-911.

Campeanu, S., Craik, F. I., Backer, K. C., & Alain, C. (2014). Voice reinstatement modulates neural indices of continuous word recognition. *Neuropsychologia*, *62*, 233-244.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 521.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior, 12*(4), 335-359.

Cousineau, D. (2005) Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*, 75-78.

Cowan, N. (1998). *Attention and memory: An integrated framework*. Oxford University Press.

Craik, F. I., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *The Quarterly Journal of Experimental Psychology*, *26*(2), 274-284.

Damer, E., and colleagues, 2018. Prolific. Crowdsourcing platform for research participant recruitment. URL prolific.ac

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*(1536), 3773-3800.

Dennis, Simon. 2007. How to use the LSA web site. Handbook of latent semantic analysis 57–70. URL http://lsa.colorado.edu/.

Drummond, A., 2017. Ibex: Internet Based Experiments. URL spellout.net/ibexfarm/

Dupoux, E., Kouider, S., & Mehler, J. (2003). Lexical access without attention? Explorations using dichotic priming. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 172.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950-1953.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of experimental psychology: Learning, Memory, and Cognition*, *10*(4), 680.

Geiselman, R. E., & Bellezza, F. S. (1976). Long-term memory for speaker's voice and source location. *Memory & Cognition, 4,* 483-489.

Geiselman, R. E., & Bellezza, F. S. (1977). Incidental retention of speaker's voice. *Memory & Cognition, 5,* 658-665.

Goh, W. D. (2005). Talker variability and recognition memory: instance-specific and voice-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 40.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, Memory, and Cognition*, *22*(5), 1166.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review, 105*(2), 251.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In J. Trouvain, & W. J. Barry (Ed.), *Proceedings of the 16th international congress of phonetic sciences,* (pp. 49-54).

Goodwin Davies, A. (2018). *Morphological representations in lexical processing* (Doctoral dissertation, University of Pennsylvania).

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Memory and Language*, *23*(5), 553.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 501.

Grossberg, S. D. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition* *by humans and machines, Vol. 1: Speech perception* (pp. 187–294). New York: Academic Press.

Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance, 23*(2), 481.

Gylfadóttir, G. (2018). *The Effective Borrowing of a Phonemic Contrast* (Doctoral dissertation, University of Pennsylvania).

Hanique, I., Aalders, E., & Ernestus, M. (2013). How robust are exemplar effects in word comprehension? *The mental lexicon, 8*(3), 269-294.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528-551.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition, 12*(6), 568-574.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*(3), 306.

Jacoby, L. L., & Hayman, C. A. (1987). Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 456.

Karayianni, I., & Gardiner, J. M. (2003). Transferring voice effects in recognition memory from remembering to knowing. *Memory & Cognition*, *31*(7), 1052-1059.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169-226). London: MIT Press.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology, 51*(2), 141-178.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978-990.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2015). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-30. *lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-30*. Retrieved from http://CRAN.R-project.org/package=lmerTest

Lee, C. Y., & Zhang, Y. (2018). Processing lexical and speaker information in repetition and semantic/associative priming. *Journal of psycholinguistic research*, *47*(1), 65-78.

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. Journal of Statistical Software, 69(1), 1-33.<doi:10.18637/jss.v069.i01>

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, *54*(5), 358.

Lüdecke D (2018). sjPlot: Data Visualization for Statistics in Social Science. doi: 10.5281/zenodo.1308157, R package version 2.6.1, https://CRAN.R-project.org/package=sjPlot.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71-102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1-86.

McKone, E. (1995). Short-term implicit memory for words and nonwords. *Journal of experimental psychology: Learning, Memory, and Cognition*, *21*(5), 1108.

McKone, E. (1998). The decay of short-term implicit memory: Unpacking lag. *Memory & cognition*, *26*(6), 1173-1186.

McKone, E., & Dennis, C. (2000). Short-term implicit memory: Visual, auditory, and cross-modality priming. *Psychonomic bulletin & review*, *7*(2), 341-346.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 539.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2005). Representation of lexical form: evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1308.

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

Mimura, M., Verfaellie, M., & Milberg, W. P. (1997). Repetition priming in an auditory lexical decision task: Effects of lexical status. *Memory & cognition*, *25*(6), 819-825.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason*, *4*(2), 61-64.

Morton, J. (1969). Interaction of information in word recognition. *Psychological review*, *76*(2), 165.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365-378.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, *102*(6), 963-968.

Musen, G., & Treisman, A. (1990). Implicit and explicit memory for visual patterns. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 127-137.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251-269.

Nairne, J. S. (1996). Short-term/working memory. In E. L. Bjork (Ed.), *Memory* (pp. 101-126). NewYork: Academic

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual review of psychology*, *53*(1), 53-81.

Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining r^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132-142.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, *47*(2), 204-238.

Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1228-1244.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & psychophysics*, *60*(3), 355-376.

H. Oja, S. Sirkia, and J. Eriksson (2006). Scatter matrices and independent component analysis. *Austrian Journal of Statistics, 35*, 175-189.

Orfanidou, E., Davis, M. H., Ford, M. A., & Marslen-Wilson, W. D. (2011). Perceptual and response components in repetition priming of spoken words and pseudowords. *The Quarterly Journal of Experimental Psychology*, *64*(1), 96-121.

Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *Journal of cognitive neuroscience*, *18*(8), 1237-1252.

Otgaar, H., Peters, M., & Howe, M. L. (2012). Dividing attention lowers children's but increases adults' false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 204-210.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309-328.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, *83*(1), 56-64.

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8-13.

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

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech communication, 13*(1), 109-125.

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing.* Vienna, Austria. Retrieved from http://www.R-project.org

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological review, 111*(1), 159-182.

Roediger, H. L., & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & cognition*, *15*(5), 379-388.

Satterthwaite, F. E., 1946. An approximate distribution of estimates of variance components. *Biometrics bulletin, 2*(6), 110–114.

Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory & Cognition*, *7*(1), 3-12.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of experimental psychology: Learning, Memory, and Cognition*, *13*(3), 501-518.

Schacter, D. L. (1990). Perceptual representation systems and implicit memory. *Annals of the New York Academy of Sciences*, *608*(1), 543-571.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 915-930.

Schacter, D. L., & Graf, P. (1989). Modality specificity of implicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 3-12.

Sheffert, S. M. (1998). Voice-specificity effects on auditory word priming. *Memory & Cognition*, *26*(3), 591-598.

Sigurðsson, E. F. (2017). *Deriving case, agreement and voice phenomena in syntax* (Doctoral dissertation, University of Pennsylvania).

Sneller, B. (2018). *Mechanisms of Phonological Change: Allophonic restructuring of /æ/ in Philadelphia* (Doctoral dissertation, University of Pennsylvania).

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. The Journal of the Acoustical Society of America, 111(4), 1872-1891.

Tremblay, A., & Ransijn, J. (2015). LMERConvenienceFunctions: Model Selection and Post-hoc Analysis for (G)LMER Models. R package version 2.10. *LMERConvenienceFunctions: Model Selection and Post-hoc Analysis for (G)LMER Models. R package version 2.10*. Retrieved from http://CRAN.R-project.org/package=LMERConvenienceFunctions

Tyler, D. E., Critchley, F., Dümbgen, L., & Oja, H. (2009). Invariant co-ordinate selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71*(3), 549-592.

Võ, M. L. H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. Psychophysiology, 45(1), 130-140.

Wagenmakers, E. J., Van Der Maas, H. L., & Grasman, R. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic bulletin & review*, *14*(1), 3-22.

Wagner, A. D., Maril, A., & Schacter, D. L. (2000). Interactions between forms of memory: when priming hinders new episodic learning. *Journal of Cognitive Neuroscience*, *12*(Supplement 2), 52-60.

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 387-401.

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software, 21*(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software, 40*(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

Wickham, H., & Francois, R. (2015). dplyr: A Grammar of Data Manipulation. R package version 0.4.3. *dplyr: A Grammar of Data Manipulation. R package version 0.4.3*. Retrieved from https://github.com/hadley/dplyr

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832