ACOUSTIC CORRELATES OF SYNTAX IN SENTENCE PRODUCTION AND COMPREHENSION

By

GWENDOLYN LOUISE REHRIG

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Karin Stromswold

And approved by

____

____

____

____

New Brunswick, New Jersey

October, 2017

ABSTRACT OF THE DISSERTATION

Acoustic Correlates of Syntax in Sentence Production and Comprehension

by GWENDOLYN LOUISE REHRIG

Dissertation Director:

Karin Stromswold

When saying active and passive sentences, speakers produce longer verb stems in passive sentences than actives due to phrase-final lengthening in passives and polysyllabic shortening in actives, and lengthening occurs on the verb stem vowel (Stromswold et al., 2002). Listeners may be able to use this acoustic cue to predict whether a sentence is active or passive prior to the verbal inflection (Stromswold et al., 2002; 2016). In three production experiments, we further investigated which acoustic cues speakers produced consistently, and to what extent polysyllabic shortening and phrase-final lengthening contributed to duration cues to syntax. In Experiment 1 we compared progressive active (*was punching*) and passive verb stems (*was punched*), and in Experiments 2 and 3 we added a comparison to perfective active verb stems (*has punched*). Experiments 1 and 2 showed that the most consistent cue was passive verb stem lengthening, while speakers also showed passive auxiliary lengthening to a lesser degree. Experiments 2 and 3 showed that both phrase-final lengthening and polysyllabic shortening were present, but polysyllabic shortening contributed most consistently to the progressive active-passive verb stem duration difference. Experiment 3 confirmed that the verb stem vowel was the

locus of passive verb stem lengthening. This suggests that listeners in Stromswold et al. (2002; 2016) may have used verb stem vowel duration to predict upcoming syntax.

In two comprehension experiments, we tested whether native English speakers used verb stem vowel duration as a cue to syntax, and whether L2 English speakers were able to both produce the same acoustic cues to syntax and use them in comprehension.

First, we manipulated the duration of the verb stem vowel to “swap” duration across active and passive sentences, inverting the vowel duration cue. In a visual-world paradigm task, listeners looked to the correct image prior to hearing the verbal inflection. The effect of the vowel duration manipulation on processing, however, was slight. This suggests that the earlier, but less consistent acoustic correlates may cue listeners to syntax in combination with the duration of the verb stem vowel, and that the parser is robust enough to ignore a single inconsistent cue.

Second, native Mandarin speakers who learned English as a second language completed the same visual-world paradigm task, but listened to only unaltered recordings. Mandarin speakers waited until the second noun phrase to identify the syntax of the sentence correctly. This may indicate an inability to use acoustic and morphosyntactic cues that are not present in the listener’s native language, or may reflect the difference in English proficiency.

Acknowledgements

I would like to thank my advisor, Karin Stromswold, for her invaluable support and guidance. I would also like to thank my committee members, Kristen Syrett, Dave Kleinschmidt, and Joseph Casillas. The dissertation is improved thanks to their input. I thank Paul de Lacy for his insights, which have vastly improved this work. I would also like to thank current and former members of the Language Acquisition and Language Processing lab at Rutgers University, without whom this work would not have been possible. Specifically, I thank Nicolaus Schrum, Eleonora Beier, Elizabeth Chalmers, Melinh Lai, Katie Aveni, Aldo Mayro, Megan Kenny, Sten Knutsen, Rachel Shu, and Karl Mulligan. I thank Eileen Kowler for her support and advice throughout my graduate career.

I would be remiss if I did not also thank the hardworking staff in the psychology department and the Center for Cognitive Science here are Rutgers for their help over the years. I thank Anne Sokolowski, Sue Cosentino, Jo’Ann Meli, Donna Tomaselli, Dianne McMillion, John Dixon, and Tamela Wilcox. Their hard work made it much easier for me to navigate a very difficult path.

My friends and family have been an extraordinary source of unconditional love and support. Their encouragement helped keep me going through my most difficult moments. I especially want to thank my father, Dr. David Johnson, who has always been there for me, and my sister, Elisabeth Johnson, who has gone out of her way on many occasions to support me. Lastly, I thank Dr. Andrew Dobson for his incredible support and companionship during the final leg of my graduate school journey.

Completion of this work was supported by the National Science Foundation for the Social, Behavioral, and Economic Sciences (BCS-0124095; BCS-0042561), the Integrative Graduate Education and Research Traineeship (IGERT DGE 0549115), and the Research Experience for Undergraduates program (REU CNS-1062735). Financial support was also provided by the Aresty Foundation and the Dorothy and David Cooper Scholarship.

Dedication

In loving memory of my mother, Dr. Paula Michal-Johnson, who was very proud of my research.

This dissertation is also dedicated to the memory of our research assistant, Maxwell Witkowski, whose work helped make this dissertation possible.

Table of Contents

ABSTRACT OF THE DISSERTATION ii

Acknowledgements iv

Dedication vi

Table of Contents vii

List of Tables xi

List of Figures xiv

1. Introduction 1

2. Linguistic Background 6

2.1. Incremental Speech Processing. 6

2.2. Theories of Sentence Processing 12

2.3. Second Language (L2) Processing 15

2.4. Summary 18

3. Sentence Production 19

3.0. Motivation 19

3.1. Experiment 1: Acoustic cues to syntax 21

3.1.1. Methods 22

3.1.2. Results 28

3.1.3. Discussion 40

3.2. Experiment 2: Phrase-final lengthening vs. polysyllabic shortening 43

3.2.1. Methods 45

3.2.2. Results 51

3.2.3. Discussion 71

3.3. Experiment 3: Segment duration 74

3.3.1. Methods 75

3.3.2. Results 81

3.3.3. Discussion 105

3.4. General Discussion 107

4. Sentence Comprehension by Native English Speakers 109

4.0. Motivation 109

4.1. Experiment 4a: Norming study 110

4.1.1. Methods 111

4.1.2. Results 117

4.1.3. Discussion 123

4.2. Experiment 4b: Manipulated vowel study 123

4.2.1. Methods 124

4.2.2. Results 130

4.2.3. Discussion 165

4.3. General Discussion 168

5. Sentence Comprehension and Production by L2 English Speakers 170

5.0. Motivation 170

5.1. Experiment 5a: L2 English sentence processing 172

5.1.1. Methods 173

5.1.2. Results 174

5.1.3. Discussion 197

5.2. Experiment 5b: L2 English production 198

5.2.1. Methods 199

5.2.2. Results 201

5.2.3. Discussion 215

5.3. General Discussion 216

6. Conclusions 219

6.1. Chapter 3 Summary: Sentence Production by Native English Speakers 219

6.2. Chapter 4 Summary: Sentence Comprehension by Native English Speakers 222

6.3. Chapter 5 Summary: Sentence Production and Comprehension by L2 English Speakers 225

6.4. Implications for Sentence Production 228

6.5. Implications for Sentence Processing 231

Appendix A 236

1. Experiment 1 Stimuli: Target Sentences 236

2. Experiment 1 Stimuli: Filler Sentences 238

3. Experiment 1 Stimuli: Verb Frequency 239

4. Experiment 1 Stimuli: Select Images Described by Sentences 240

Appendix B 242

1. Experiment 2 Stimuli: Target Sentences 242

2. Experiment 2 Stimuli: Filler Sentences 245

3. Experiment 2 Stimuli: Verb Frequency 249

Appendix C 251

1. Experiment 3 Stimuli: Target Sentences 251

2. Experiment 3 Stimuli: Filler Sentences 254

3. Experiment 3 Stimuli: Verb Frequency 255

Appendix D 257

1. Experiment 4b Code: [Python] Vowel Manipulation Script 257

2. Experiment 4b and 5a Code: [Python] Track-Loss Correcting Script 260

3. Experiment 4b and 5a Code: [R] Eye-Tracking Data Parsing Function 263

Appendix E 271

Bibliography 274

List of Tables

Table 3.1.1. 28

Table 3.1.2. 30

Table 3.1.3. 31

Table 3.1.4. 33

Table 3.1.5. 35

Table 3.1.6. 38

Table 3.1.7. 39

Table 3.2.1. 52

Table 3.2.2. 54

Table 3.2.3. 56

Table 3.2.4. 59

Table 3.2.5 62

Table 3.2.6. 64

Table 3.2.7. 66

Table 3.2.8. 68

Table 3.2.9. 70

Table 3.3.1. 82

Table 3.3.2. 85

Table 3.3.3. 87

Table 3.3.4. 91

Table 3.3.5. 95

Table 3.3.6. 98

Table 3.3.7. 101

Table 3.3.8. 103

Table 4.1.1. 119

Table 4.1.2. 121

Table 4.2.1. 131

Table 4.2.2. 134

Table 4.2.3. 137

Table 4.2.4. 150

Table 4.2.5. 163

Table 5.1.1. 175

Table 5.1.2. 177

Table 5.1.3. 179

Table 5.1.4. 181

Table 5.1.5. 184

Table 5.1.6. 193

Table 5.1.7. 196

Table 5.2.1. 201

Table 5.2.2. 203

Table 5.2.3. 205

Table 5.2.4. 207

Table 5.2.5. 209

Table 5.2.6. 210

Table 5.2.7. 212

Table 5.2.8. 214

List of Figures

Figure 1.1. 2

Figure 1.2 3

Figure 3.1.1. 26

Figure 3.1.2. 28

Figure 3.1.3. 30

Figure 3.1.4. 33

Figure 3.1.5. 35

Figure 3.1.6. 37

Figure 3.1.7. 39

Figure 3.1.8. 41

Figure 3.1.9. 42

Figure 3.2.1. 50

Figure 3.2.2. 54

Figure 3.2.3. 56

Figure 3.2.4. 59

Figure 3.2.5. 62

Figure 3.2.6. 64

Figure 3.2.7. 66

Figure 3.2.8. 70

Figure 3.3.1. 80

Figure 3.3.2. 82

Figure 3.3.3. 85

Figure 3.3.4. 87

Figure 3.3.5. 91

Figure 3.3.6. 95

Figure 3.3.7. 98

Figure 3.3.8. 101

Figure 4.1.1. 112

Figure 4.1.2. 113

Figure 4.1.3. 115

Figure 4.1.4. 119

Figure 4.1.5. 121

Figure 4.2.1. 126

Figure 4.2.2. 131

Figure 4.2.3. 134

Figure 4.2.4. 139

Figure 4.2.5. 140

Figure 4.2.6. 142

Figure 4.2.7. 143

Figure 4.2.8. 144

Figure 4.2.9. 145

Figure 4.2.10. 146

Figure 4.2.11. 147

Figure 4.2.12. 149

Figure 4.2.13. 153

Figure 4.2.14. 154

Figure 4.2.15. 156

Figure 4.2.16. 157

Figure 4.2.17. 158

Figure 4.2.18. 160

Figure 4.2.19. 163

Figure 5.1.1. 177

Figure 5.1.2. 181

Figure 5.1.3. 184

Figure 5.1.4. 186

Figure 5.1.5. 188

Figure 5.1.6. 189

Figure 5.1.7. 191

Figure 5.1.8. 191

Figure 5.1.9. 192

Figure 5.1.10. 195

Figure 5.2.1. 200

Figure 5.2.2 203

Figure 5.2.3 205

Figure 5.2.4 208

Figure 5.2.5 212

Figure 5.2.6 214

1. Introduction

To correctly understand a sentence, it is necessary to construct a representation of its underlying structure. Crucially, two different parses for the same surface structure can result in entirely different sentence meanings. For example, a prepositional phrase can embed within another syntactic phrase (noun, verb, prepositional phrase, etc.), and where a prepositional phrase occurs can affect the interpretation of the sentence. To illustrate the effect of attachment location on the interpretation of a sentence, consider the attachment of the prepositional phrase *in my pajamas* in the famous line (1):

1. *I shot an elephant in my pajamas.*

To obtain the plausible interpretation of the sentence, where the subject (*I*) wears pajamas during the event *shot an elephant*, the prepositional phrase must attach to the verb phrase (Figure 1A).[1] The implausible interpretation requires the prepositional phrase to attach to the noun phrase (Figure 1B), which is necessary to understand the subsequent joke, “How he got in my pajamas, I don’t know”.

[1] The tree diagrams shown in Figure 1.1 for the sentence given in example 1 do not reflect modern approaches to syntactic theory, and are simplified here for illustrative purposes. We realize it is not an accurate representation of the syntax as a whole, and use it solely to contrast between VP and NP attachment.

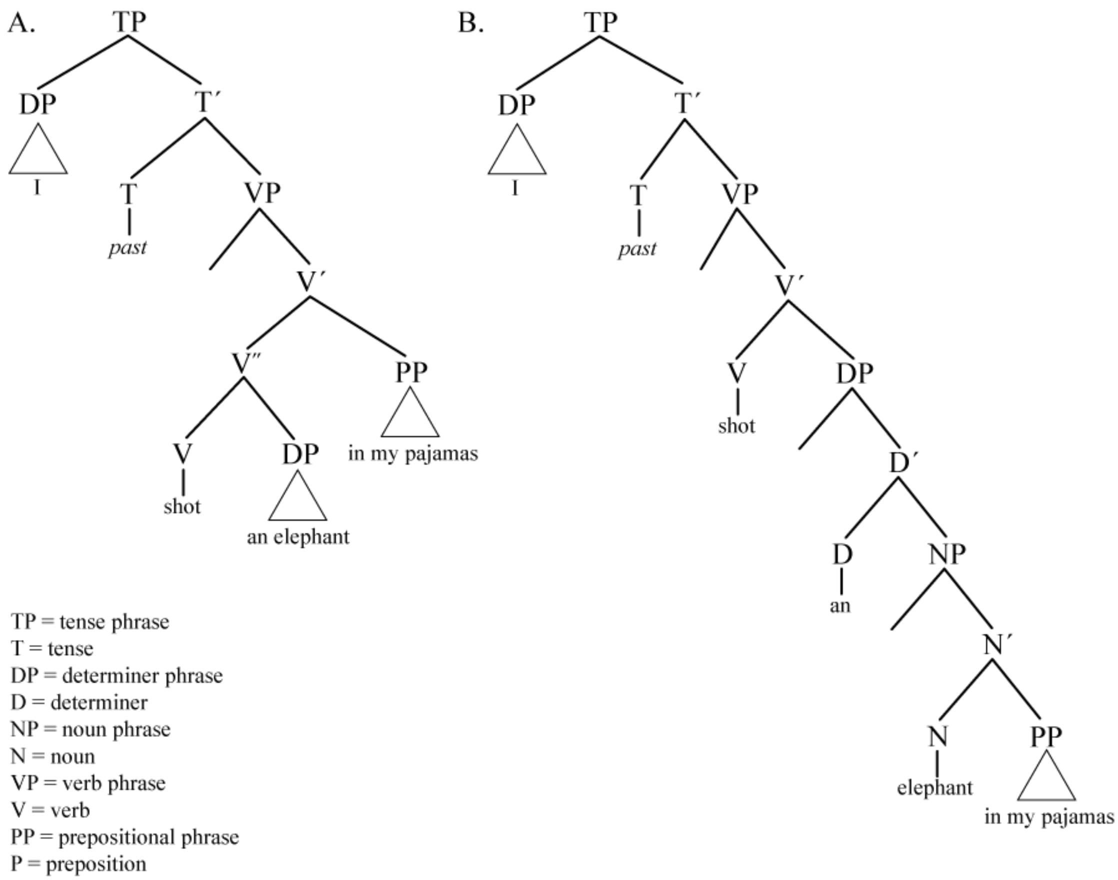


Figure 1.1. Two viable syntactic structures for the sentence *I shot an elephant in my pajamas*. A) shows the prepositional phrase modifying the verb *shot*, and B) shows the prepositional phrase modifying the noun *elephant*. The syntactic structures are simplified for convenience.

The above example is a case of permanent syntactic ambiguity: an ambiguity that is not resolved by the end of the sentence. In contrast, temporary syntactic ambiguity resolves as the sentence unfolds. Consider the progressive active sentence (2a) and the passive sentence (2b):

1. a) *The dog was punching the bear*.

   b) *The dog was punched by the bear*.

With respect to morphosyntax, the sentences in 2a and 2b are temporarily ambiguous through the verb stem (*punch*), after which the inflection on the verb disambiguates the two structures (shown in Figure 1.2). The dissertation investigates the role that acoustic correlates of syntax play in informing parsing decisions for temporarily ambiguous sentences, and what phonological processes give rise to those acoustic correlates.

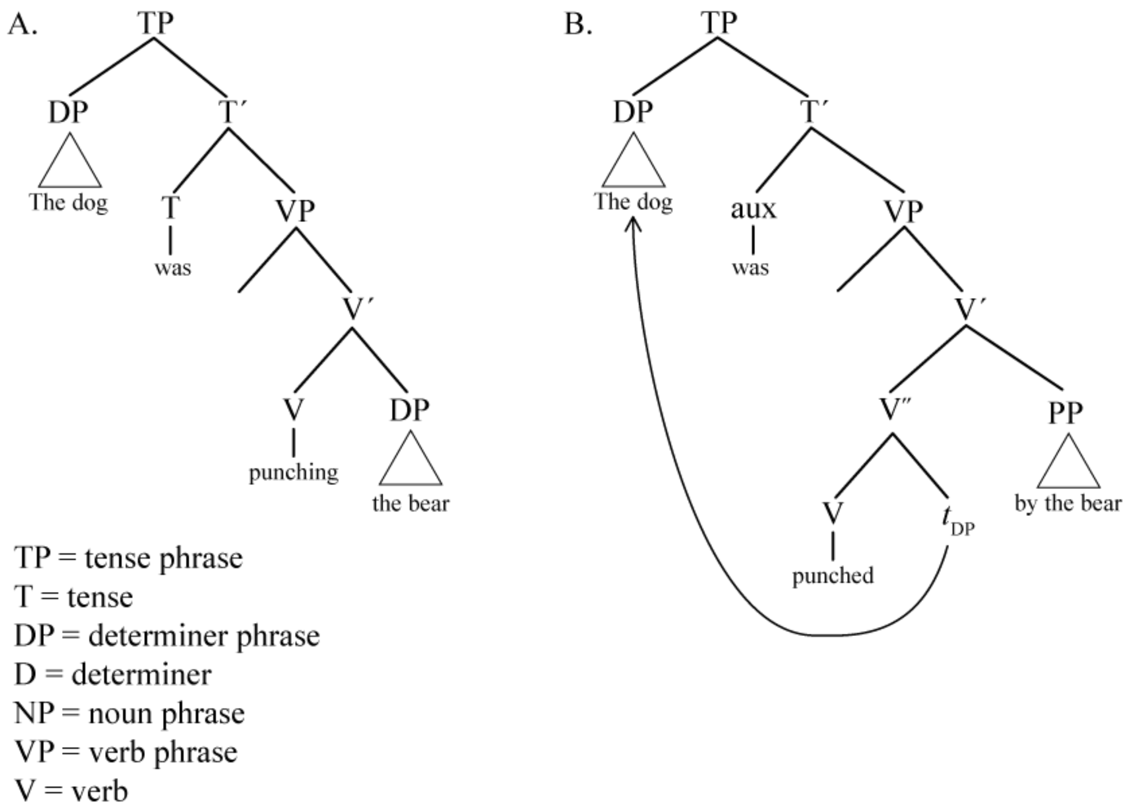


Figure 1.2. Simplified trees showing the syntactic structure for a progressive active sentence (A) and a passive sentence (B).

As discussed in Chapter 2, listeners incrementally construct a representation of an utterance in real-time, and can use acoustic information to form predictions about yet-to-be-heard sentence content. The identity of the next word can be predicted when coarticulatory information on a preceding vowel predicts the next segment (Salverda et al., 2014). Syllable duration can predict the total number of syllables in the word, and the location of the word within a prosodic phrase (Salverda et al., 2003; 2007). The presence of a pause can be used to determine whether a prepositional phrase modifies a noun or serves as an instrument of the action (Snedeker & Yuan, 2008). In sum, listeners are able to use various acoustic cues in order to predict upcoming content in an utterance. The dissertation addresses whether listeners can use acoustic cues to form predictions about sentences that are temporarily ambiguous (as in example 2).

Stromswold et al. (2002) found that when a single native English speaker said active and passive sentences, the verb stem vowel was longer in passive sentences (e.g., 2B) than in active sentences (e.g., 2A). They further reported results from an eye-tracking study and a gating study that indicate that native English-speaking adults can use acoustic cues to predict whether a sentence is active or passive prior to hearing the inflection on the verb (Stromswold et al., 2002; under review).

Chapter 2 reviews related work on several topics in psycholinguistic research. First, the relevant incremental processing literature that demonstrates the use of acoustic cues to predict upcoming sentence content is discussed. This includes the use of cues that can predict a word from part of a word, the next word from the end of the previous word, and syntactic structure. Second, theories of sentence processing are reviewed. Compatibility between these theories and the incremental sentence processing literature, and particularly the verb stem vowel duration cue, are discussed. Finally, the chapter discusses literature on sentence processing in a speaker's second language.

In Chapter 3, three experiments are presented, the goals of which are to further investigate under what circumstances native English speakers produce a verb stem duration difference between progressive active and passive sentences, and why the difference arises. Specifically, the first experiment demonstrates that naïve native English speakers reliably produce verb stems that are longer in passive sentences as compared to active sentences, and the second experiment identifies polysyllabic shortening and phrase-final lengthening as the phonological processes that give rise to the duration difference, with polysyllabic shortening playing a larger role. The third experiment confirms that the verb stem vowel is the only constituent that consistently undergoes lengthening, and drives the duration difference.

Chapter 4 presents two studies of sentence comprehension by native English speakers. In the first, vowel duration is manipulated to be longer in progressive active sentences and shorter in passive sentences. The experiment then demonstrates that the manipulation sounds natural to listeners overall, but sounds less natural in active sentences. In the second, altering verb stem vowel duration did not heavily interfere with processing in a visual world paradigm task, suggesting that listeners do not use verb stem vowel duration as the primary cue to active and passive syntax. Furthermore, a learning effect is observed in which listeners are able to learn that passive sentences occur more frequently in the experiment than in English as a whole, and adjust their processing strategies to reflect this learning.

Two experiments on production and comprehension by L2 English speakers are presented in Chapter 5. In the first, native Mandarin speakers who learned English later in life demonstrate considerable processing delays: Mandarin speakers wait until after the verbal inflection to form a prediction. In the second, native Mandarin speakers lengthen passive verb stems in English sentences.

The concluding chapter discusses the implications of the results of the experiments conducted in this dissertation for theories of sentence production and processing.

2. Linguistic Background

2.1. Incremental Speech Processing.

Research suggests that when listeners process sentences, they do so by incrementally building the underlying syntactic structure (e.g. Trueswell & Tanenhaus, 1994; see 2.3). This requires linguistic knowledge about which structures are common in the language, which words are likely to follow one another, etc. The beginning of a sentence is inherently uncertain; nevertheless, listeners actively predict upcoming sentence content in real time. One way that listeners predict yet-to-be-heard sentence content is by exploiting information in the acoustic signal. Of interest here, listeners may use the duration of a syllable, and even of an individual segment (e.g., a vowel), to predictively process sentences.

Some of the most persuasive evidence for predictive processing of speech comes from experiments that use the visual world paradigm. Since its inception, the visual world paradigm has been a boon for online language processing research (Tanenhaus, Spivey-Knowlton, Eberhard, Sedivy, 1996; Allopenna, Magnuson, Tanenhaus, 1998). In this paradigm, listeners hear an utterance and view an image array. While listening to the utterance, an eye tracker records the listener's gaze, which serves as an indirect measure of predictive processing. Using this paradigm, psycholinguists have shown that listeners use information in the acoustic signal to predict subsequent sound segments (Swingley, Pinto, & Fernald, 1999; Fernald, Swingley, & Pinto, 2001; Dahan, Tanenhaus, & Chambers, 2002; Salverda, Kleinschmidt, & Tanenhaus, 2014), syllables (Salverda, Dahan, & McQueen, 2003; Salverda, Dahan, Tanenhaus, Crosswhite, Masharov, &

McDonough, 2007), and syntactic structure (Snedeker & Yuan, 2008; Thothathiri & Snedeker, 2008; Choi & Trueswell, 2010).

When we speak, the way we articulate each segment is influenced by articulation of the preceding and following segments, a phenomenon known as coarticulation. Coarticulation has phonetic consequences, which may be perceptible to listeners, and may cue them to upcoming segments. Swingley et al. (1999) found that both children and adults processed spoken words incrementally (segment-by-segment) in a word recognition task. The authors showed pairs of images of familiar objects to 24-month-olds. In some of the object pairs, the names were phonological competitors: names that sound the same prior to a disambiguating sound segment. Specifically, names for the objects in each pair shared onsets (*doggie-doll*), rhymed (*duck-truck*), or sounded entirely dissimilar (doggie-truck). Unsurprisingly, children and adults both looked to the correct image when the name pairs sounded dissimilar or did not share onset sounds. Of greater interest, when the names did share onsets, children and adults alike still looked at the correct image prior to hearing the object's name in its entirety. Their findings were confirmed by Fernald et al. (2001), who showed similar evidence for incremental lexical processing in children as young as 18-months. These studies indicate that listeners use segment-by-segment processing to successfully predict the identity of a word.

For listeners to process sentences incrementally, they must be able to predict upcoming segments across word boundaries as well as within word boundaries. Salverda et al. (2014) tested the role of coarticulation effects on the vowel ə (in the word *the*) in predicting the onset of the next word. Adult listeners viewed an image array of four pictures, the names of which all began with different consonants (e.g., *ladder, fish,* *grapes*, *pineapple*), and listened to sentences (*The ladder is the target*). The sentences consisted of either natural, connected speech, or recordings that were spliced to remove coarticulatory effects on the first determiner (*the*). Listeners looked to the target image earlier when coarticulatory information was present, indicating that listeners can use this information to predict the next segment across a word boundary.

Listeners are able to use discourse context (pragmatic information) to predict the ending of a word. Dahan et al. (2002) investigated the influence of pitch accent, where a portion of an utterance is emphasized, on word recognition. For example, in the sentence *Now put the CANDLE above the square*, the word *candle* is accented. Accented information is usually new information, which does not refer to an entity that has already been introduced in the discourse (Cutler, Dahan, & van Donselaar, 1997). In other words, de-accented words can serve as anaphora, which refer back to an already introduced, or given, item. Adult listeners completed an interactive task: a display showed four images, two of which were phonological competitors, and the participants were asked to move the pictures using a mouse. In each trial, participants listened to an instruction sentence (*Put the candle below the triangle*), followed by a sentence containing a target noun that was either accented (*Now put the CANDLE above the square*) or de-accented (*Now put the candle ABOVE THE SQUARE*). When the target noun was accented, listeners looked to the object that was not mentioned in the instruction sentence (e.g., *candy* if *candle* was already mentioned). When it was de-accented, listeners looked to the same object named in the instruction (*candle* if *candle* was already mentioned). Their results indicate that listeners can use pitch accent to predict the ending of a word prior to hearing the disambiguating (word-final) syllable.

Listeners can use the duration of a syllable both to recognize a word and predict its position in a sentence. Salverda et al. (2003) spliced monosyllabic words (e.g., *ham*) and attached them to polysyllabic words where the first syllable of the polysyllabic word was the same as the monosyllabic word (e.g., *hamster*), replacing the first syllable of the polysyllabic word and the words that preceded it. The monosyllabic word was longer in duration than the first syllable of its polysyllabic counterpart. Participants viewed an array of four images, two of which were phonological competitors. When listeners heard the spliced monosyllable, they looked to the image described by the monosyllabic competitor first (e.g., *ham*), then looked to the target after hearing the next syllable (e.g., *hamster*). Their findings indicate that listeners use syllable duration to predict the identity of a word. Note, however, that the nature of this manipulation does not control for any acoustic cues that may be present in the portion of the sentence that preceded the target syllable, in which case listeners may be using more than just the duration of the target syllable to form predictions.

In a follow-up study, Salverda et al. (2007) manipulated the position of both monosyllabic (e.g., *cap*) and polysyllabic words (e.g., *captain*) within an utterance so that the target word occurred either phrase-medially (*Put the <u>cap</u> next to the square*) or phrase-finally (*Now click on the <u>cap</u>*). Listeners preferred to interpret the polysyllabic word as phrase-medial and the monosyllabic word as phrase-final. When the monosyllabic word was phrase-medial, participants initially mistook it for a polysyllabic word, and vice versa. Their results further demonstrate that listeners are sensitive to the duration of a syllable during word recognition, and indicate that listeners use duration to infer the position of a syllable within a phrase. Note, however, that the structure and content of the sentences Salverda et al. (2007) tested differ considerably across the key manipulation, which could conceivably have impacted their findings.

Evidence of predictive, incremental processing is not limited to the word recognition literature. Snedeker and Yuan (2008) investigated whether children can use prosody—in this case, the presence or absence of a pause—to predict the syntax of an utterance. 4- and 5-year-olds completed an interactive task. The displays contained two objects (*feather*, *candle*) and two stuffed animals (*frog*, *tiger*) holding miniature versions of the former objects (e.g., the stuffed frog held a miniature feather). Children heard sentences that were produced by a female actor, who modeled the prosody of the utterances after productions obtained in separate experiments that had either instrument or modifier prosody. Specifically, sentences either contained a pause after the verb (*You can feel…the frog with the feather*) or did not (*You can feel the frog with the feather*). The prepositional phrase *with the feather* could either attach to the object noun phrase (modifying *the frog*) or as an adjunct to the verb (an instrument of the action *feel*). A pause after the verb was compatible with the instrument interpretation. Children looked to the instrument (*the feather*) when a pause was present, and looked to the stuffed animal (*the frog with the feather*) when it was not, indicating that children used prosody to disambiguate the syntax.

Listeners can use the structure of sentences they have recently heard in order to predict the syntax of an utterance. Thothathiri and Snedeker (2008) tested syntactic priming in 3- and 4-year old children using an interactive task. Children viewed four toys, two of which were phonological competitors, and listened to instruction sentences. The instruction sentences were either double-object datives (*Show the horse the book*) or prepositional-object datives (*Show the horn to the dog*). During the ambiguous region of the sentence (*Show the hor*—), children who were primed on double-object datives looked to the recipient (*the horse*), and children who were primed on prepositional-object datives looked to the theme (*the horn*). Their findings indicate that children can use recent experience (the primed construction) to predict syntactic structure.

In a classic study, Trueswell, Sekerina, Hill, and Logrip (1999) showed that English-speaking children fail to correct the initial misinterpretation of garden-path sentences. For example, *Put the frog on the napkin in the box* is temporarily ambiguous—before *in the box*, one could easily misinterpret *on the napkin* (a modifier) as the target destination for the action *put*. Adults can backtrack in order to override their initial (incorrect) parse of these sentences, but children fail to do so. Choi and Trueswell (2010) built on the earlier study by investigating 4- and 5-year-old children's sentence processing in a verb-final language (Korean). Korean uses an ambiguous case marker (-*ey*) on nouns that could mark the noun as a locative (e.g., the destination for an action, such as put) or (less frequently) as a modifier. Because Korean is a verb-final language, listeners hear the subject, bearing the ambiguous case marker, and the object prior to hearing the verb. Korean-speaking children listened to sentences that were ambiguous prior to the verb (gloss: *napkin-on frog pick up, napkin-on frog put*). Children's looking behavior reflected the initial (locative) parse, even after hearing the disambiguating verb. The two studies indicate that children use early information to predict the syntax of an utterance, and do not deviate from the predicted parse in the face of (late) disambiguating evidence.

The work reviewed in this section presents compelling evidence that listeners process spoken sentences incrementally, and use a variety of acoustic cues and distributional cues to do so. Most of the incremental processing literature focuses on predictive word recognition, or on using acoustic information to incrementally disambiguate syntactically ambiguous sentences. Next, we discuss theories of sentence processing, and whether existing theories of sentence processing can account for incremental processing.

2.2. Theories of Sentence Processing

Sentence processing models generally fall into one of two categories (Traxler, 2014): those that rely solely on syntactic information unless forced to reanalyze (garden-path model: Frazier & Rayner, 1982; Ferreira & Clifton, 1986; good-enough processing: Ferreira, Bailey, & Ferraro, 2002) and those that allow extrasyntactic factors to facilitate processing (constraint-satisfaction model: Trueswell & Tanenhaus, 1994; noisy-channel model: Gibson, Bergen, & Piantadosi, 2013).

According to the garden-path (Frazier & Rayner, 1982; Ferreira & Clifton, 1986, among others) and good-enough processing models (Ferreira, Bailey, & Ferraro, 2002), syntactic constraints such as Minimal Attachment (construct the simplest syntactic tree possible) are prioritized in parsing decisions, while semantic, pragmatic, lexical, and acoustic information have little to no influence. The parser builds one structure at a time using the phrase structure rules of the language. If a parse based on syntactic information alone fails, the parser will reanalyze the sentence and allow extrasyntactic information to bear on parsing decisions (Frazier, 1995).

The primary difference between the two models is that the garden-path model expects the final parse to be faithful to the input, whereas good-enough processing allows for unfaithful—but more plausible—output by using world knowledge to bias parsing decisions. For example, when confronted with the implausible passive sentence *The dog was bitten by the man*, listeners do not always process the sentence accurately. Instead, they may process the sentence as a plausible active: *The dog bit the man* (Ferreira et al., 2002). This allows the parser to fix errors a speaker may make due to performance limits, such as failing to plan a long utterance correctly. Recently, good-enough processing models have incorporated a distinction between given and new information, where processing resources are devoted to new information, and given information, which usually occurs earlier in the sentences, is processed in a good-enough manner (Ferreira & Lowder, 2016). This strategy increases parsing efficiency.

In a similar two-stage processing model, listeners use simple heuristics and biases to inform parsing decisions early in the sentence, when the listener has little input to work with (Townsend & Bever, 2001; Crocker, 2002). For example, in an SVO language like English, listeners will exploit the extreme frequency of active sentences in the language overall by assuming that the first noun they encounter in the sentence is the agent of the action (hereby referred to as the NP1 = agent strategy). Because passive sentences are very infrequent (Stromswold et al., under review), this strategy will often work for the listener.

In all of the models summarized above, listeners make an initial—often, shallow—parse using the syntactic information available to them, and will revise the initial interpretation in light of conflicting evidence.

In contrast, constraint-satisfaction models (e.g., Trueswell & Tanenhaus, 1994) and probabilistic models (e.g., Gibson, Bergen, & Piantadosi, 2013, among others) allow all potentially relevant information, linguistic and non-linguistic (e.g., visual context) to influence parsing decisions. Constraint-satisfaction models allow multiple candidate parses to be considered in parallel, and "activated" based on their plausibility, the frequency of the construction in the language, predicate argument structure (verb transitivity, thematic constraints on arguments), and other constraints (Trueswell & Tanenhaus, 1994). Each candidate parse is activated in proportion to how well it satisfies the constraints, and the field of candidate parses narrows as the sentence unfolds. This process occurs incrementally on a word-by-word basis, until it reaches a uniqueness point in the sentence that determines the correct candidate.

In the noisy-channel framework, listeners must make probabilistic inferences to parse sentences, due to uncertainty introduced by noise (Gibson et al., 2013). Rather than processing a noisy utterance directly, listeners infer what the speaker most likely intended to say. Like good-enough processing, the model allows world knowledge to bias the final sentence interpretation toward a plausible sentence. However, like constraint-satisfaction models, it also allows other linguistic knowledge (frequency, semantics, etc.) to bias the final sentence interpretation.

Both constraint-satisfaction and noisy-channel modes allow extrasyntactic information to bias parsing decisions. Because these models do not strictly rely on syntactic information, they are better able to account for predictive behavior during on-line sentence processing tasks (Traxler, 2014).

2.3. Second Language (L2) Processing

In sections 2.1 and 2.2, a body of literature about speech processing in a speaker's native tongue (L1) was reviewed. The work reviewed in 2.1 indicated that native speakers are rapidly able to make predictions about upcoming elements in the speech signal using a variety of acoustic cues. The theories discussed in 2.2 rely strongly on the native speaker's knowledgebase to inform parsing decisions. We now turn to the unique challenges associated with processing sentences in a second language.

Whether L2 speakers can process speech incrementally in the way that native speakers do is not entirely clear. There is some evidence that L2 speakers can form predictions to facilitate speech processing in their L2 (Marull, 2017), at least in the face of a predictive cue that is related to, if not the same as, a cue in their native language. However, even proficient L2 speakers fail to form predictions from cues that are entirely absent in their native language, even when those cues are highly reliable (Lew-Williams & Fernald, 2007; Lew-Williams & Fernald, 2010; Grüter, Lew-Williams, & Fernald, 2012).

Marull (2017) compared native Spanish speakers to native English speakers who learned Spanish as a second language to determine whether non-native Spanish speakers can use the plurality of a definite article to facilitate processing. Like Spanish, English distinguishes between singular and plural forms of demonstratives (e.g., *that*, *those*) but unlike Spanish, English lacks distinct definite articles for singular and plural nouns (e.g., *the man*, *the men*). L2 speakers were divided into speakers with either intermediate or high Spanish proficiency. In a picture matching task, either demonstratives (*eso*$_{sg}$, *esos*$_{pl}$; English: *that*$_{sg}$, *those*$_{pl}$) or definite articles (*el*$_{sg}$, *los*$_{pl}$, English: *the*) were presented in recorded sentences while two pictures were displayed. When the target picture differed from the distractor only in the quantity of potential referents shown (e.g., a single magician in one image, two clowns in the other), listeners were able to use the plurality of the determiner to facilitate processing. For example, the plural demonstrative *esos* predicted the plural target *payasos* when the distractor image showed only one member of the circus. This was true across language groups, which suggests that L2 Spanish speakers were able to successfully use plural morphology to predict upcoming sentence content, even when the specific plural morphology used was not present in their L1. Similarity between English and Spanish with regard to plural morphology likely allowed L2 speakers to generalize to a plural definite article, despite the absence of plural morphology on definite articles in English.

L2 sentence processing is less successful when the predictive cue is entirely absent in the native language. A series of studies revealed that L2 Spanish speakers whose L1 lacks grammatical gender were unable to use the gender of a determiner to predict a gender-matched noun (Lew-Williams & Fernald, 2007; Lew-Williams & Fernald, 2010; Grüter et al., 2012). Native speakers of Spanish can use the gender of an article (e.g., el_{masc}, la_{fem}) to predict an upcoming noun of the same gender (e.g., $zapato_{masc}$, $pelota_{fem}$). In an eye-tracking task, native adult and 3-year-old Spanish speakers fixated to a target of the same gender as the article faster when the gender of the article was informative—when only one object in the display matched in gender (Lew-Williams & Fernald, 2007; Lew-Williams & Fernald, 2010). In contrast, native English speakers who learned Spanish as a second language were not able to use the gender of an article to facilitate processing of the upcoming gender-agreeing noun. This was true even when all L2 Spanish speakers were highly proficient (Grüter et al., 2012). Interestingly, native and L2 Spanish speakers alike were able to use the gender of a determiner to predict the gender of a nonce noun after training, but L2 Spanish speakers could only do so when the same determiner type (definite article, e.g., *el durino*) was used in both the training and test sentences (Lew-Williams & Fernald, 2010; Grüter, et al., 2012). Native Spanish speakers, on the other hand, could generalize beyond the articles used in the training phase: processing by native speakers was faster for gender informative articles, even when definite articles were used during training and indefinite articles were used in the test phase (e.g., *un durino*), but this was not the case for L2 speakers.

The findings of Lew-Williams & Fernald (2010) suggest that the predictive processing abilities demonstrated by native speakers occur because of the way a first language is necessarily acquired: infants recognize statistical regularities in continuous speech and use those regularities to segment the speech signal. Spanish speaking infants do not hear the noun *pelota* without a gender-matched determiner. In contrast, L2 speakers are given word boundaries in written form, and also learn words presented in relative isolation in a classroom setting. Because the acquisition process differs, L1 speakers have a rich base of knowledge about the statistical regularities of their language that they can use to make predictions in real time, while L2 speakers do not. However, L2 speakers may be able to generalize from the predictive relationships in their L1 when a predictive cue is shared, or is very similar (e.g., determiner plurality as a cue to noun plurality; Marull, 2017).

2.4. Summary

In this chapter, we have reviewed several bodies of work that motivate the dissertation. First, in 2.1 we reviewed the literature on incremental processing in a speaker's native language, which predominantly focuses on lexical processing (e.g., incremental word recognition). Next, in 2.2 we discussed theories of sentence processing and contrasted between those that prioritize syntactic information and those that allow all available information to bear on processing decisions. Finally, in 2.3 we reviewed a portion of the second language acquisition literature that speaks to the debate over whether L2 speakers can predict upcoming sentence content in native-like ways. Next, we present several production experiments in which native English speakers produce acoustic cues to progressive active and passive syntax.

3. Sentence Production

3.0. Motivation

In Chapter 2, we presented evidence that listeners predict subsequent acoustic information using subtle acoustic cues, and that they can use highly salient acoustic cues to correctly predict syntactic parsing decisions.

A series of studies on the processing of progressive active and passive English sentences revealed that duration cues through the verb stem may disambiguate the two syntactic structures (Stromswold, Eisenband, Norland, & Ratzan, 2002; Stromswold, Kharkwal, Sorkin, & Zola, under review; Rehrig, Beier, Chalmers, Schrum, & Stromswold, 2015; Stromswold, Lai, Rehrig, & de Lacy, 2016). The processing of progressive active and passive sentences is interesting because much of the surface form of the sentence can be identical; however, the syntactic structure of a passive sentence is quite different from that of a progressive active sentence. In passive sentences, the subject of a transitive verb is omitted, and the object of the verb moves to become the syntactic subject. The subject of the verb can optionally be expressed in a *by*-phrase. Consider, for example, the progressive active sentence (3a) and the passive sentence (3b):

3. a) *The girl was kicking the boy*

 b) *The girl was kicked by the boy*

In the progressive active sentence, *the girl* is the subject of the verb *kick*, and is also the agent of the verb *kick*: she is the one performing the action. In the passive sentence, *the girl* is the grammatical subject of the verb *kick*, and is the patient of the verb: she is on the receiving end of the action. See Figure 1.2. for simplified syntactic trees comparing the syntax of similar progressive active and passive sentences (example

2). From the first noun phrase to the verb stem, the two sentences are identical on the surface, but they describe very different events.

Stromswold et al. (2002) and Stromswold et al. (under review) recorded progressive active and passive sentences spoken by an adult native English speaker with linguistic training. The recordings served as stimuli for a sentence comprehension task that employed the visual world paradigm. To their surprise, listeners were able to correctly predict the syntactic structure of the sentence during the portion of the sentence that was supposed to be temporarily ambiguous (prior to the inflection on the verb). The surface form prior to the inflection of the verb (*the girl was kick*—) is ambiguous on paper, but acoustic information in the spoken sentences could potentially disambiguate the two structures. Analysis of the stimuli sentences revealed that monosyllabic verb stems in passive sentences (3b) were longer than the equivalent verb stems in the progressive active sentences (3a). Further analyses revealed that the vowel in the verb stem lengthened, and that the passive auxiliary was lengthened in one list of sentences, but no other duration differences were present between morphosyntactically ambiguous regions of the stimuli (from the first determiner through the verb stem). This suggests that listeners may have used the duration of the verb stem vowel to predict the inflection (*-ing* or *-ed*) on the verb, and by extension the syntactic structure of the sentence. However, the sentences were produced by a single speaker with linguistic training, and could have reflected that speaker's unique speech.

The experiments presented in this chapter expand on the work of Stromswold et al. (2002) by investigating the reliability of the verb stem duration difference they discovered, and by identifying the production processes that give rise to the duration difference. Experiment 1 (section 3.1) investigates whether the duration difference found by Stromswold et al. (2002) generalizes beyond a single speaker with linguistic training. Experiment 2 (section 3.2) determines which phonological processes yield the duration difference between progressive active and passive verb stems. In Experiment 3 (section 3.3), we identify which segment(s) in the verb stem undergo a duration change.

3.1. Experiment 1: Acoustic cues to syntax

Stromswold et al. (2002) found an acoustic correlate of syntax for a single native speaker with linguistic training. The purpose of Experiment 1 is to determine whether the findings of Stromswold et al. (2002) generalize beyond a single speaker with linguistic training, and whether the findings generalize to a broader set of verbs. Following Stromswold et al. (2002), we carried out a production study in which naïve native English speakers produced progressive active and passive sentences of the form of 2a and 2b (section 1). Their productions were then segmented into morphemes, and the acoustic properties of the morphemes within the morphosyntactically ambiguous region were analyzed.

If passive verb stem lengthening occurs as the result of general properties of the English language, then all of the naïve speakers should lengthen passive verb stems. It would then follow that listeners in Stromswold et al. (2002) may have used the duration of the verb stem to predict syntactic structure. However, it may be the case that listeners used other cues during the morphosyntactically ambiguous region. We assume that predictive acoustic cues must be consistent and reliable across speakers in order for listeners use them in incremental processing. If other cues are available to listeners in progressive active and passive sentences, we therefore expect naïve speakers in the current experiment to produce those same acoustic cues.

3.1.1. Methods

Participants. Nine adult monolingual native American English speakers participated. Native speaker status was determined via responses to the following yes/no questions[2]: 1) are you a native speaker of English?, 2) is English the only language you are fluent in?, and 3) was English the only language spoken in your home when you were growing up? Participants were naïve to the purpose of the experiment, did not have linguistic training, and had no history of language impairments. Speakers who did not have a dialectal accent typical of central New Jersey, or who hyperarticulated, were excluded from analysis (n = 2). Of the remaining subjects, 6 were female and 1 was male. All participants were paid volunteers. The study was approved by the Rutgers University Institutional Review Board, and was carried out in accordance with the Declaration of Helsinki.

Materials. *Verbs.* Sixteen regular verbs (*chase*, *comb*, *kick*, *kiss*, *lick*, *pat*, *pinch*, *poke*, *punch*, *push*, *scrub*, *shove*, *tickle*, *touch*, *trap*, *wash*) were selected using the MRC Psycholinguistic Database (Coltheart, 1981) and the SUBTLEXus database (Brysbaert & New, 2009). Verbs were selected to match a series of pictures depicting animals performing actions (see Appendix A.4 for representative images). All verbs were highly frequent and had similar frequency (Appendix A.3), were felicitous as passives (Levin, 1993), and took the -*ed* passive inflection. Preference was given to monosyllabic verbs with stop consonants in onset position, verbs without consonant clusters, and verbs for

[2] These questions were used to evaluate native English speaker status throughout the dissertation; specifically, in Experiments 1, 2, 3, 4a, and 4b.

which the passive participle was not syllabic; however, priority was necessarily given to verbs that best described the series of drawings, even if they violated the aforementioned phonological constraints.

Sentences. All passive sentences included an agentive *by*-phrase. Active and passive sentences were constructed using the 16 verbs selected and two of 19 animal names as nouns in the sentence, which were chosen to match the series of pictures (e.g., *The rabbit was washing the duck, The rabbit was washed by the duck*). Each noun occurred as both the first noun and as the second noun in separate sentences. This resulted in 64 total target sentences (16 verbs x 2 syntactic frames x 2 noun positions; Appendix A.1). Twenty-eight active filler sentences were constructed (Appendix A.2), resulting in 92 sentences total.

Design. Sentence order was pseudorandomized so that no sentence was followed by another sentence containing the same verb, or the same noun in subject position. No more than 2 passive sentences occurred consecutively, no more than 6 progressive active target sentences occurred consecutively, and there were at most 8 consecutive target sentences with no fillers in between. Nine simple yes/no comprehension questions were constructed (e.g., *Did the rabbit wash the duck?*) to ensure participants processed the sentences that they read aloud. Each comprehension question pertained to the preceding sentence: For example, the question *Did the rabbit wash the duck?* was asked after the participant had said the sentence *The rabbit was washing the duck*. One comprehension question followed a target passive sentence, 2 comprehension questions followed active target sentences, and the remaining 6 comprehension questions occurred after active filler sentences. Comprehension questions were interspersed at irregular intervals (every 5-14 trials). The first list was pseudorandomized, then that list was reversed to create a second order. List presentation order was counterbalanced across participants.

Apparatus. Participants sat in a sound-attenuated booth while wearing an AKG C 420 head-worn condenser microphone and Audio-Technica ATH-M50 headphones. Audio was recorded on a single channel at a rate of 44.1 kHz using an ART Digital MPA Microphone Preamplifier and GoldWave audio editing software (GoldWave Version 5.70; GoldWave Inc, 2013). The experiment was presented on a laptop equipped with Windows 7. Stimulus presentation and response measures were controlled by E-Prime 2.0 Professional experiment software. A keyboard mask was used to occlude all keys except *Q*, *P*, and the spacebar.

Procedure. Prior to the experimental trials, participants first said each word in the experiment in isolation. Participants were instructed to first read the sentence silently, then aloud. This was followed by 6 practice trials, including a practice comprehension question. A trial proceeded as follows. A fixation screen appeared for 2000 ms, followed by a blank screen for 250 ms. This was followed by a screen displaying a sentence. The sentence persisted until the participants read the sentence aloud, then pressed spacebar to initiate the beginning of the next trial. After an inter-trial-interval of 250 ms, the next trial began. On catch trials, the yes/no comprehension question was displayed after the inter-trial-interval and persisted until the participant responded via keypress (*Q* or *P*). Once the participant responded to the question, a fixation was displayed for 5000 ms, followed by the next sentence presentation. This procedure repeated for 92 trials. At the half-way point, participants took a 2-minute break. The experiment took approximately 30 minutes. After completing the experiment, participants completed the New York Times dialect survey (Katz, Andrews, & Buth, 2013), followed by a separate production study not reported here.

Segmentation. Morpheme onsets were marked by hand for each audio file using Praat (Boersma, 2001). Segmentation was carried out by 3 research assistants using the following criteria. The first syllable of all morphemes tested began with a consonant or consonant cluster in onset position. For each morpheme, a boundary was placed at the start of the consonant in onset position of the first syllable. Stop onset boundaries were placed immediately after a gap in the spectrogram, reflecting the closure of the consonant, and before the burst. Voiced onsets were marked at the beginning of the voicing bar in the spectrogram. Voiceless fricative onsets were marked at the first zero crossing of a noisy region in the waveform with no periodicity. For cases where a morpheme began with a sonorant, a boundary was placed at the zero-crossing preceding the first period that marked a qualitative change in the waveform. When segment transitions were ambiguous, as in the /r/ - /w/ transition in *The bear was punched by the dog*, the boundary was placed at the midpoint of the ambiguous region. See figure 3.1.1 for an example segmented sentence.

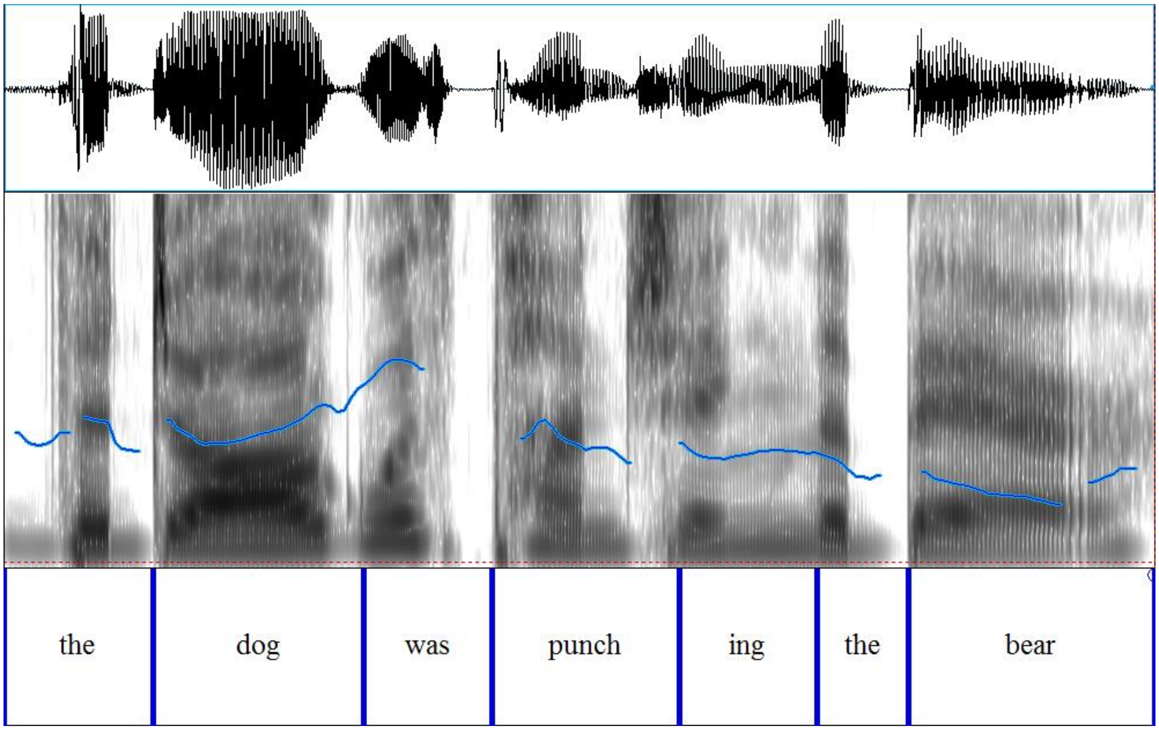


Figure 3.1.1. Experiment 1: Example of morpheme boundary placements for the sentence *The dog was punching the bear.* Boundaries (shown in blue) mark the onset of each morpheme per the segmentation criteria outlined in section 3.1.1.

Inter-rater reliability for morpheme duration on 40 target sentences from a single speaker was high (coders 1 and 2: $r(298) = 0.97$, $p < .001$; coders 1 and 3: $r(298) = 0.97$, $p < .001$; coders 2 and 3: $r(298) = 0.99$, $p < .001$). This indicates that coders consistently used the same segmentation criteria throughout.

Analysis. For each interval, defined from the start of one morpheme to the start of the next, the interval duration and mean intensity were calculated. Duration and mean intensity of auxiliary and verb stems were then analyzed using Bayesian linear mixed-effects models using the rstanarm package in R[3]. See Nicenboim & Vasishth (2016) for a practical guide on this analysis method. The rstanarm package was used to approximate a posterior distribution using Markov Chain Monte Carlo sampling. Weakly informative

[3] ANOVAs revealed no differences in duration or intensity prior to the auxiliary.

priors were used for all models, and were initialized using the default rstanarm parameters. To control for variation due to individual differences or due to the items tested, random slopes and random intercepts were included for the syntactic frame (active or passive) by subjects and by item (verb). For stan objects, we used the median as a measure of central tendency, and standard deviation to report variance (per rstanarm's default output). The standard deviation of the median absolute difference (MAD_{SD} hereafter) was reported for all Bayesian models.

The overall significance of random effects (e.g., individual differences by subject) were assessed via model comparison using the loo package, which allowed for comparison between two models and returned a value for the difference in *expected log predictive density* ($ELPD_{diff}$) between them. This value indicated which of the models predicted the data better: a negative value favored the first model, and a positive value favored the second. Throughout the dissertation, the more complex model was always used as the first argument, therefore a negative $ELPD_{diff}$ indicated that the more complex model predicted the data better, and a positive $ELPD_{diff}$ indicated that the simpler model predicted the data better.

The lsmeans package in R was used to obtain *p*-values, and to determine whether levels of the fixed effects differed from one another. Degrees of freedom were estimated using the Satterthwaite method (Satterthwaite, 1946), and *p*-values were adjusted for multiple comparisons using the Tukey method where applicable. Comparisons that were not significant according to statistical tests were further evaluated by determining the proportion of samples that fell within 1 just-noticeable difference (JND) of zero, and/or based on the size of the β coefficient. For duration, the JND range was set to 20% of the average duration (Klatt, 1976) and for intensity the JND range was 1 dB (Harris, 1963). The mean and standard error of the mean were reported as measures of central tendency and variance, respectively, obtained from the lsmeans package.

3.1.2. Results

Duration. *Auxiliary.* The duration of the auxiliary *was* was 10.41 ms longer in passive sentences (*M* = 219.02 ms, *SE* = 2.62 ms) than in active sentences (*M* = 208.61 ms, *SE* = 2.61 ms; see Figure 3.1.2). When the impact of syntax on auxiliary duration was analyzed using a Bayesian linear mixed-effects model, passive auxiliary lengthening was marginally significant (*z* = -1.76, *p* = .08; see Table 3.1.1.).

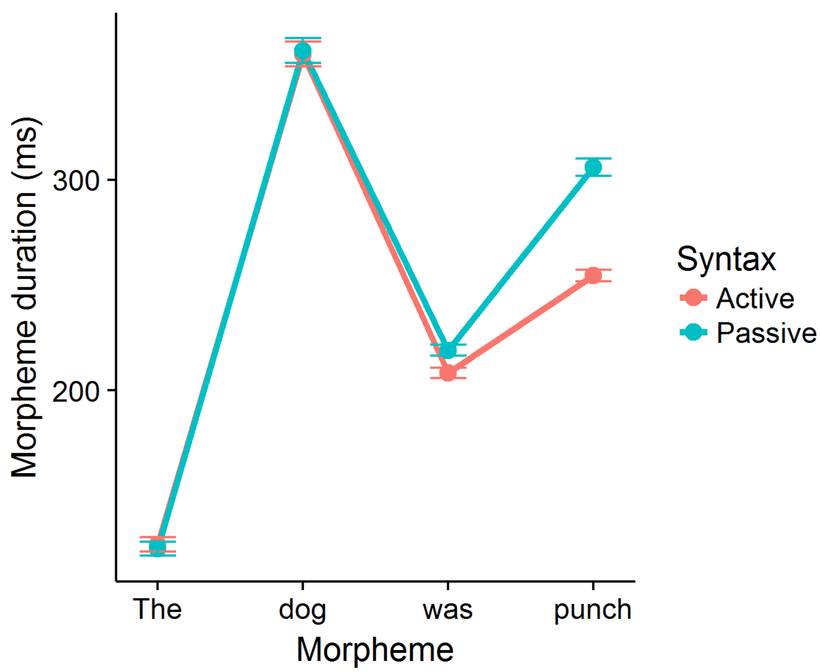


Figure 3.1.2. Experiment 1: Mean morpheme duration for active and passive sentences. Error bars indicate standard error of the mean.

Table 3.1.1.

Experiment 1: Bayesian Linear Mixed-Effects Model Summary for Syntax and Auxiliary Duration

					Median	MAD_{SD}
Mean Posterior Predictive Distribution					213.80 ms	1.9 ms
Random Effects	*SD*	*r*				
Subject (n = 7)						
Intercept	14.90	–				
Passive Syntax	11.80	-0.39				
Verb (n = 16)						
Intercept	26.80	–				
Passive Syntax	8.50	-0.20				
Residual	28.00	–				
Fixed Effects	*Median β*	*β* MAD_{SD}				
Intercept	209.10	9.1				
Error *SD*	27.90	1.10				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Passive Syntax	-9.99	5.66	-1.76	.08	0.82	19.38

The extent of auxiliary lengthening across active and passive sentences varied for different speakers (Figure 3.1.3). For 6 of the 7 speakers, the auxiliary was longer in passive sentences, and a sign test indicates that this was marginally significant (cumulative binomial: $p = .06$). Surprisingly, for one speaker (subject 6) the auxiliary was *shorter* in passive sentences than in active sentences.

A model that included subject as a random effect predicted the data better than a model that did not ($ELPD_{diff}$ = -25.5, *SE* = 7.3), which suggests individual differences in passive auxiliary lengthening. Three speakers lengthened the passive auxiliary to a significant extent (*ps* < .05), 3 lengthened the passive auxiliary, but not to a reliable degree, and 1 speaker lengthened the *active* auxiliary (see Table 3.1.2). For all speakers, over 98% of samples fell within 1 JND of zero, which suggests that the difference in auxiliary duration may not be perceptible to listeners. However, if passive auxiliary lengthening is perceptible, longer passive auxiliaries may serve as a very early cue to syntax (see 3.1.3. for discussion on this point).

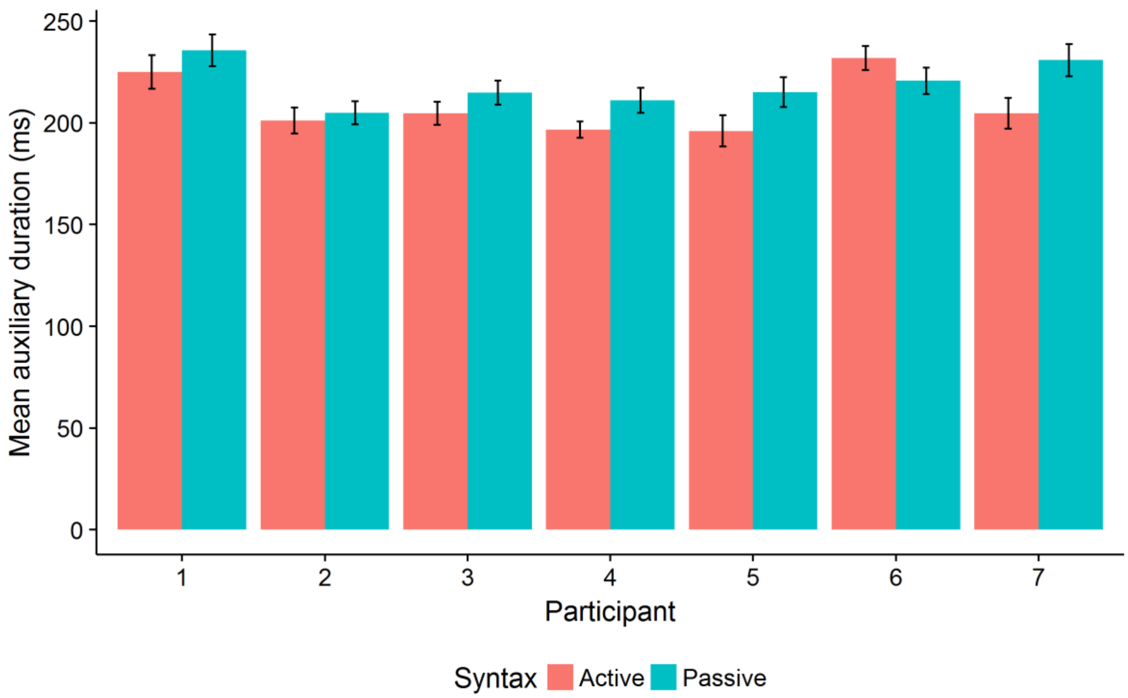


Figure 3.1.3. Experiment 1: Mean auxiliary duration in active and passive sentences, for individual participants. Error bars indicate standard error of the mean.

Table 3.1.2.

Experiment 1: Progressive Active and Passive Auxiliary Duration by Subject

	Syntax

Descriptive Statistics	Sex	Active		Passive	
		M	*SE*	*M*	*SE*
Subject 1	Female	224.97	8.29	235.72	7.73
Subject 2	Female	201.14	6.38	204.90	5.65
Subject 3	Female	204.71	5.74	214.84	5.92
Subject 4	Female	196.67	4.12	211.05	6.21
Subject 5	Female	196.03	7.62	215.08	7.30
Subject 6	Male	231.89	5.80	220.69	6.46
Subject 7	Female	204.64	7.58	230.84	7.97
Overall		208.61	2.61	219.02	2.62
Statistical Comparisons		*Mean β*	*β SE*	*z*	*p*
Subject 1: Progressive vs. Passive		-10.97	7.32	-1.52	.13
Subject 2: Progressive vs. Passive		-4.08	7.29	-0.56	.58
Subject 3: Progressive vs. Passive		-10.38	7.25	-1.43	.15
Subject 4: Progressive vs. Passive		-14.50	7.11	-2.04	.04
Subject 5: Progressive vs. Passive		-16.35	7.52	-2.17	.03
Subject 6: Progressive vs. Passive		10.82	7.31	1.48	.14
Subject 7: Progressive vs. Passive		-26.47	7.65	-3.46	< .01

Verb stem. Consistent with Stromswold et al. (2002), the average verb stem duration in passive sentences (*M* = 305.87 ms, *SE* = 4.11 ms) was 51.51 ms longer than that of active verb stems (*M* = 254.36 ms, *SE* = 2.82 ms; see Figure 3.1.2.), and this difference was significant ($z = -4.64$, $p < .01$; see Table 3.1.3.).

Table 3.1.3.

Experiment 1: Bayesian Linear Mixed-Effects Model Summary for Active and Passive Verb Stem Duration

			Median			MAD_{SD}
Mean Posterior Predictive Distribution			280.10 ms			1.9 ms
Random Effects	*SD*	*r*				
Subject (n = 7)						
Intercept	21.00	–				
Passive Syntax	19.00	-0.07				
Verb (n = 16)						
Intercept	34.00	–				
Passive Syntax	31.00	0.29				
Residual	28.00	–				
Fixed Effects	*Median β*	*β* MAD_{SD}				
Intercept	254.30	11.50				
Error *SD*	28.20	1.0				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Passive Syntax	-51.01	11.00	-4.64	< .01	33.78	69.33

A model that included a random effect of subject predicted the data better than an otherwise equivalent model ($ELPD_{diff}$ = -71.2, *SE* = 11.1), indicating individual differences in passive verb stem lengthening. Despite individual differences overall, all subjects significantly lengthened passive verb stems as compared to active verb stems (all *ps* < .01, see Figure 3.1.4 and Table 3.1.4). This was unlikely to occur by chance alone according to a sign test (cumulative binomial: $p < .02$).

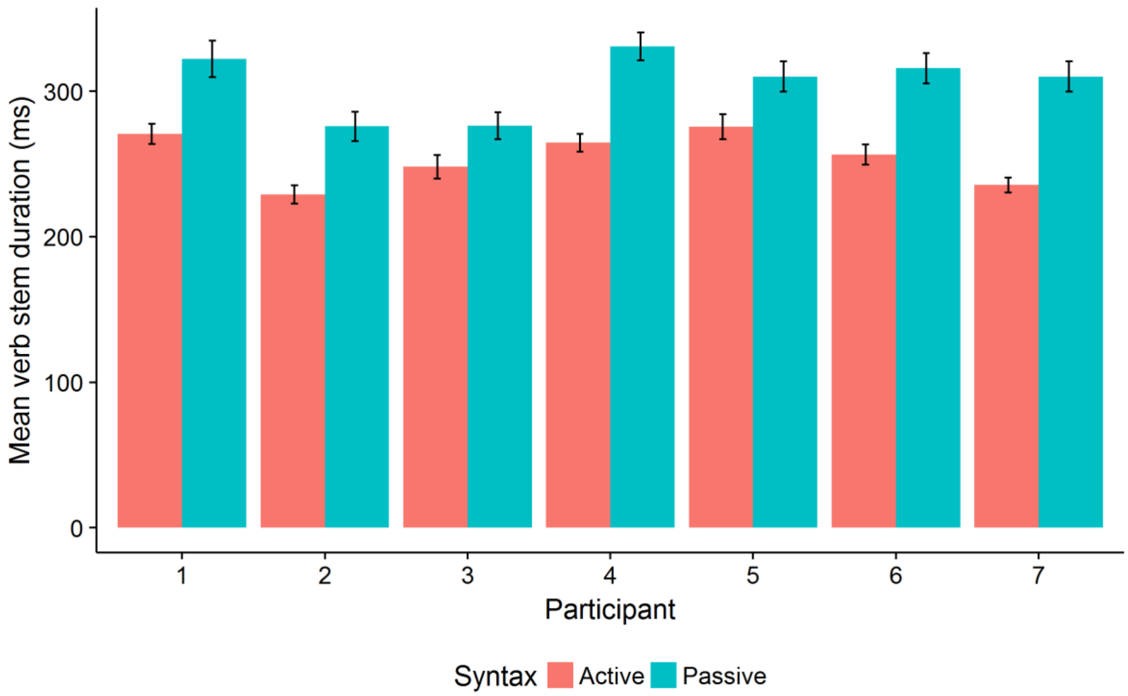


Figure 3.1.4. Experiment 1: Mean verb stem duration in active and passive sentences, for each participant. Error bars indicate standard error of the mean.

Table 3.1.4.

Experiment 1: Progressive Active and Passive Verb Stem Duration by Subject

		Syntax			
		Active		Passive	
Subject	Sex	*M*	*SE*	*M*	*SE*
Subject 1	Female	270.84	6.95	322.32	12.51
Subject 2	Female	229.14	6.31	275.90	10.10
Subject 3	Female	248.12	8.05	276.37	9.15
Subject 4	Female	264.64	6.01	330.84	9.60

Subject 5	Female	275.93	8.55	310.17	10.37
Subject 6	Male	256.56	7.04	315.84	10.43
Subject 7	Female	235.62	5.15	310.17	10.43
Overall		254.36	2.82	305.87	4.11

Progressive vs. Passive Syntax	*Mean* β	β *SE*	*z*	*p*
Subject 1: Progressive vs. Passive	-50.82	10.37	-4.90	< .001
Subject 2: Progressive vs. Passive	-47.20	10.49	-4.50	< .001
Subject 3: Progressive vs. Passive	-27.54	10.50	-2.62	.009
Subject 4: Progressive vs. Passive	-66.07	10.43	-6.33	< .001
Subject 5: Progressive vs. Passive	-35.43	10.70	-3.31	.001
Subject 6: Progressive vs. Passive	-59.29	10.49	-5.65	< .001
Subject 7: Progressive vs. Passive	-74.35	10.78	-6.90	< .001

Verb stems were longer in passive sentences for all verbs, which is unlikely to have occurred by chance alone (cumulative binomial: $p < .001$). However, the duration of the verb stem and the magnitude of the progressive active-passive verb stem duration difference varied depending on the verb (see Figure 3.1.5). When a model that accounted for differences due to different verbs was compared to one that did not, the model that accounted for variation due to verb was a better predictor of the data ($ELPD_{diff} = -234.80$, $SE = 20.50$), indicating an effect of items (verbs). Passive verb stem lengthening was significant for 13 of the 16 verbs (*ps* < .05, see Table 3.1.5). Note that for 3 verbs, the duration difference was not significant: namely, *pat*, *tickle*, and *trap*. For all 3 of these verbs, the proportion of posterior samples that fell within one JND of zero was at least

97%, suggesting that the size of the effect is too small to have practical significance. We will discuss why this is the case in section 3.1.3.

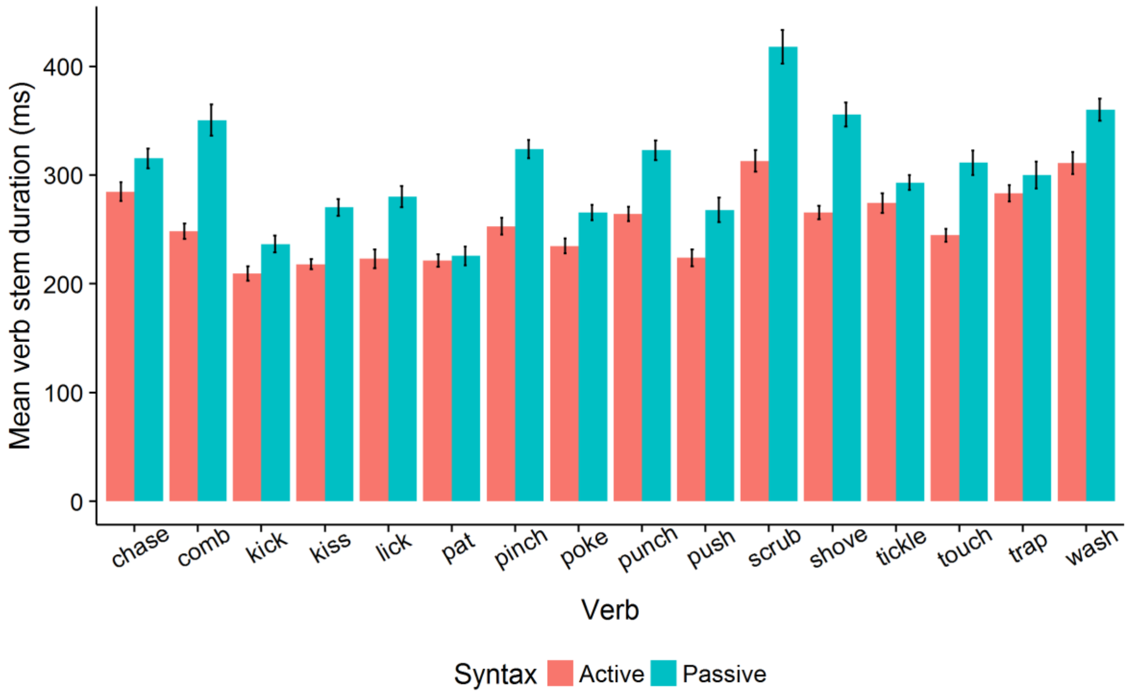


Figure 3.1.5. Experiment 1: Mean verb stem duration in active and passive sentences, for each target verb. Error bars indicate standard error of the mean.

Table 3.1.5.

Experiment 1: Progressive Active and Passive Verb Stem Duration by Verb

Verb	*Mean β*	*β SE*	*z*	*p*
chase: Progressive vs. Passive	-32.94	12.44	-2.65	.01
comb: Progressive vs. Passive	-101.73	12.98	-7.84	< .001
kick: Progressive vs. Passive	-26.93	13.03	-2.07	.03
kiss: Progressive vs. Passive	-51.54	13.11	-3.93	< .001
lick: Progressive vs. Passive	-57.31	12.83	-4.47	< .001
pat: Progressive vs. Passive	-4.40	12.90	-0.34	.73

pinch: Progressive vs. Passive	-69.21	12.99	-5.33	< .001
poke: Progressive vs. Passive	-30.40	12.61	-2.41	.02
punch: Progressive vs. Passive	-58.13	12.65	-4.59	< .001
push: Progressive vs. Passive	-44.09	13.24	-3.33	< .001
scrub: Progressive vs. Passive	-104.69	13.11	-7.99	< .001
shove: Progressive vs. Passive	-89.89	13.30	-6.76	< .001
tickle: Progressive vs. Passive	-18.65	13.31	-1.40	.16
touch: Progressive vs. Passive	-67.87	13.56	-5.01	< .001
trap: Progressive vs. Passive	-16.76	13.17	-1.27	.20
wash: Progressive vs. Passive	-48.53	12.87	-3.77	< .001

To recap, passive auxiliaries were longer than active auxiliaries for most participants, which suggests that auxiliary duration may be an early, but potentially less reliable, cue to syntax. Consistent with Stromswold et al. (2002), verb stems were longer in passive sentences as compared to progressive active sentences. This was true for all speakers, which shows that even linguistically naïve speakers demonstrate passive verb stem lengthening. However, the magnitude of the duration difference varied by verb, and in some cases, was negligible. Overall, verb stem duration appears to be a robust cue to passive syntax for most verbs.

Intensity. *Auxiliary.* Average intensity was similar between active (M = 62.59 dB, SE = 0.42 dB) and passive (M = 62.26 dB, SE = 0.38 dB) auxiliaries (z = 0.33, p = .74). The proportion of samples within 1 JND of zero was 88.03%, which suggests that the effect is either so small as to not be of practical significance, or that the effect size estimate is nearly zero (median β = -0.20, MAD_{SD} = 0.50).

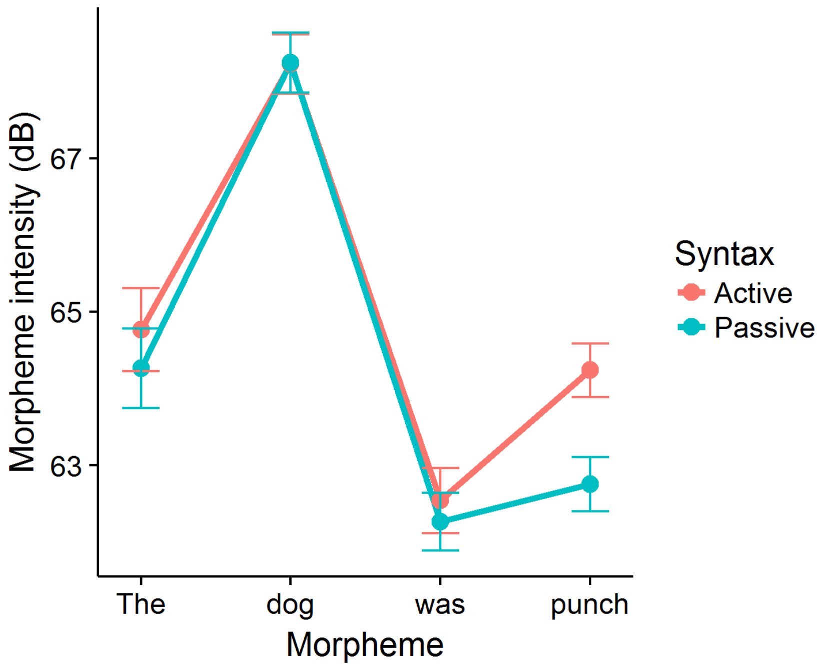


Figure 3.1.6. Experiment 1: Mean morpheme intensity in active and passive sentences. Error bars indicate standard error of the mean.

Verb stem. Average intensity was 1.46 dB higher for active verb stems (*M* = 64.21 dB, *SE* = 0.35 dB) than for passive verb stems (*M* = 62.75 dB, *SE* = 0.35 dB; see Figure 3.1.6), and the magnitude of the difference was reliable ($z = 2.65$, $p = .008$; see Table 3.1.7).

While all speakers produced verb stems with greater intensity in active sentences as compared to passive sentences, the magnitude of the difference varied by speaker (Figure 3.1.7). A model that included random slopes and intercepts by subjects predicted the data better than an equivalent model that did not account for variation from different subjects ($ELPD_{diff} = -234.80$, $SE = 20.50$), which indicates individual differences in average verb stem intensity across constructions. Active verb stem intensity was higher for all participants, which was unlikely to occur by chance (cumulative binomial: $p = .02$), but average intensity was only significantly higher for 1 of 7 speakers (see Table

3.1.7). However, it is possible that the effect is present, but too small to be of practical significance: for 3 of the 7 speakers, fewer than 24% of samples fell within 1 JND of zero, and for the remaining subjects, fewer than 64% of samples fell within 1 JND of zero. Though there is considerable uncertainty on the effect size, it is possible that intensity differences may have been perceptible even if they were not large.

Table 3.1.6.

Experiment 1: Bayesian Linear Mixed-Effects Model Fixed Effects Summary for Active and Passive Verb Stem Intensity

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			63.5 dB	0.20 dB
Random Effects	*SD*	*r*		
Subject (n = 7)				
Intercept	4.30	–		
Passive Syntax	1.22	-0.19		
Verb (n = 16)				
Intercept	2.79	–		
Passive Syntax	0.84	0.19		
Residual	2.56	–		
Fixed Effects	*Median* β	β MAD_{SE}		
Intercept	64.1	1.60		
Error *SD*	2.60	0.10		
			Posterior Interval	

	Mean β	*β SE*	*z*	*p*	5%	95%
Active vs. Passive	1.43	0.54	2.65	.008	-2.31	-0.55

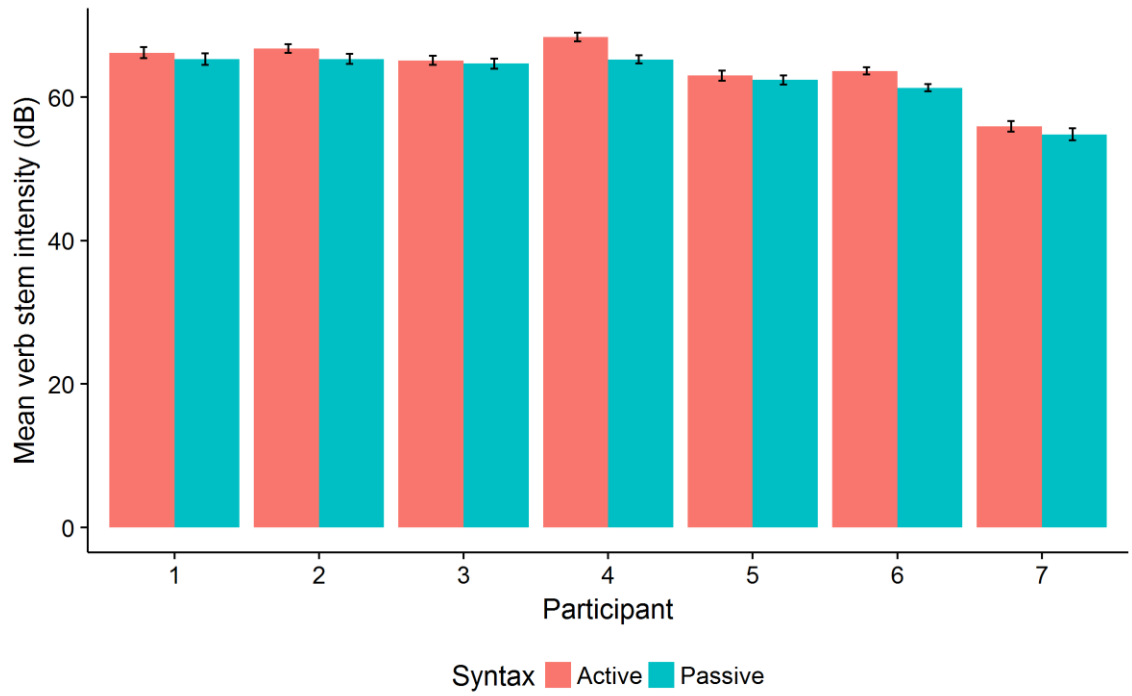


Figure 3.1.7. Experiment 1: Mean verb stem intensity in active and passive sentences, for each individual participant. Error bars indicate standard error of the mean.

Table 3.1.7.

Experiment 1: Progressive Active and Passive Verb Stem Intensity by Subject

		Syntax			
		Active		Passive	
Subject	Sex	*M*	*SE*	*M*	*SE*
Subject 1	Female	66.19	0.79	65.29	0.80
Subject 2	Female	66.77	0.58	65.31	0.69
Subject 3	Female	65.11	0.63	64.66	0.68

Subject 4	Female	68.37	0.60	65.25	0.56
Subject 5	Female	62.99	0.68	62.40	0.65
Subject 6	Male	63.64	0.49	61.30	0.53
Subject 7	Female	55.92	0.75	54.79	0.84
Overall		64.21	0.35	62.75	0.35

Subject	*Mean β*	*β SE*	*z*	*p*
Subject 1: Progressive vs. Passive	-0.77	0.78	-0.98	.33
Subject 2: Progressive vs. Passive	-0.23	0.77	-0.30	.77
Subject 3: Progressive vs. Passive	0.98	0.77	1.28	.20
Subject 4: Progressive vs. Passive	2.00	0.78	2.59	.01
Subject 5: Progressive vs. Passive	-0.77	0.81	-0.96	.34
Subject 6: Progressive vs. Passive	1.05	0.77	1.36	.18
Subject 7: Progressive vs. Passive	-0.97	0.79	0.22	.22

In summary, the acoustic analyses described above show that passive verb stem lengthening was the only reliable acoustic cue to syntax across all 7 speakers. Passive auxiliaries were lengthened reliably by some, but not all speakers. Average verb stem intensity was higher for active verb stems across speakers, though the magnitude of the difference varied. These findings are consistent with those of Stromswold et al. (2002), and lend credence to the possibility that listeners can use verb stem duration to predict active or passive syntax.

3.1.3. Discussion

The results of Experiment 1 indicate that the progressive active-passive verb stem duration difference is indeed robust. All speakers in the study lengthened passive verb stems for monosyllabic verbs. Verb stems in a polysyllabic passive context (e.g., *patted*) did not undergo lengthening. The same phonological processes that explain the verb stem duration difference for monosyllabic passive verbs also explains why polysyllabic passive verb stems did not differ reliably from progressive active verb stems in duration.

3.1.3.1. Why does lengthening take place?

Stromswold et al. (2002) speculated that the most likely explanation for the phenomenon they discovered, and that was replicated here, is that it arises as the result of prosodic structure: the abstract structure underlying speech that gives rise to the acoustics of an utterance (Wagner & Watson, 2010; Shattuck-Hufnagel & Turk, 1996; Beckman, 1996). Prosodic structure organizes linguistic units into larger constituents according to the prosodic hierarchy, where syllables (Figure 3.1.8) are the lowest constituents in the hierarchy, and the entire utterance occupies the top of the hierarchy (Selkirk, 1986; Shattuck-Hufnagel & Turk, 1996; Beckman, 1996).

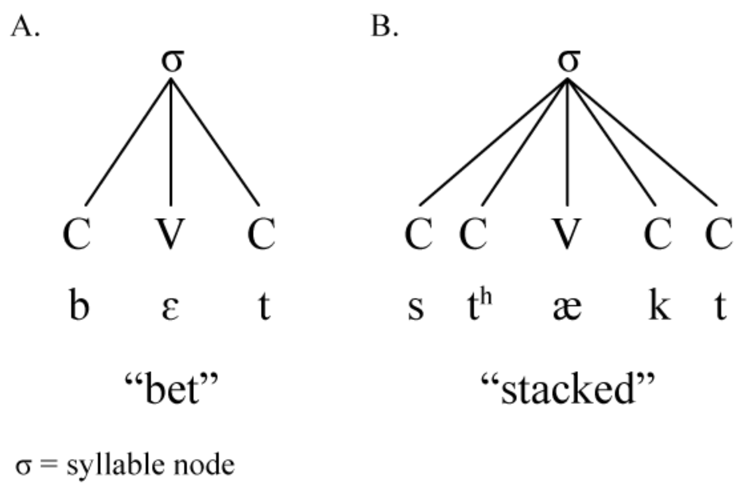


Figure 3.1.8. Experiment 1: Simplified syllable structure trees for syllables containing (A) simple onsets and codas only and (B) consonant clusters in onset and coda position (adapted from Zec, 2007).

Though there is disagreement over the exact nature of the prosodic hierarchy (e.g., Selkirk, 1986 vs. Beckman, 1996), it is generally accepted that the intonational phrase level sits below the utterance (highest) level of the hierarchy, and houses the intonational contour (see Figure 3.1.9).

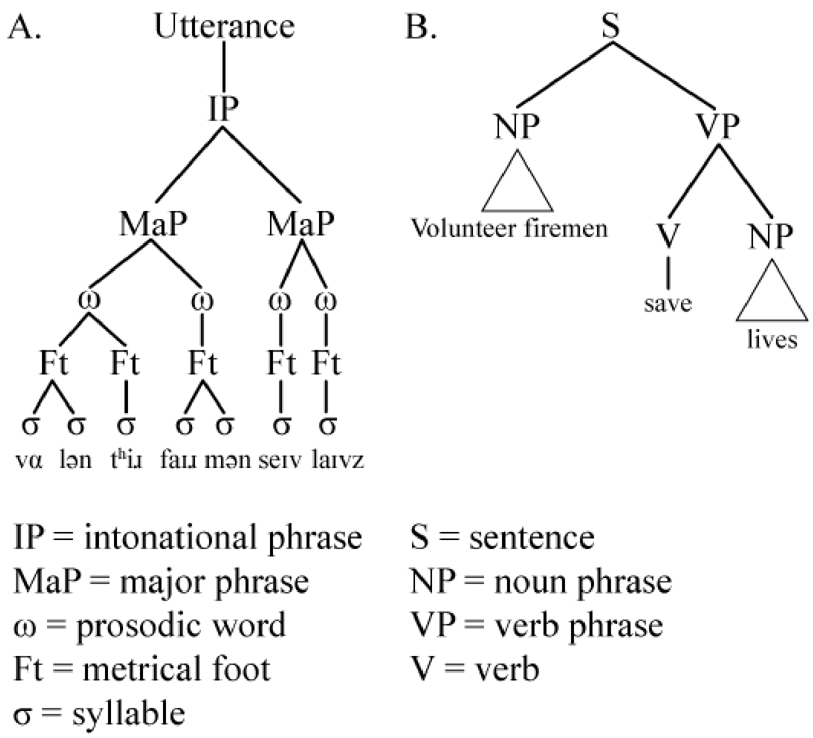


Figure 3.1.9. Experiment 1: Example of possible prosodic (A) and syntactic (B) structure for the sentence *Volunteer firemen save lives.* Adapted from Selkirk (1995).

One consequence of prosodic structure on the acoustics of an utterance is that the duration of a syllable varies depending on its location within a prosodic phrase. When a syllable immediately precedes a prosodic phrase boundary, that syllable undergoes lengthening (Lehiste, 1973; Klatt & Cooper, 1975; Klatt, 1976; Beckman & Edwards, 1990; van Santen, 1992). This phenomenon is called phrase-final lengthening. As suggested by Stromswold et al. (2002), the presence of phrase-final lengthening in passive verb stems may explain their results and the findings of Experiment 1, because a monosyllabic passive verb (*punched*) is often phrase-final, and usually undergoes lengthening. In contrast, the progressive active verb is phrase-medial, and would not undergo lengthening.

Another factor that may contribute to the duration difference is polysyllabic shortening. Polysyllabic shortening refers to the observation that a syllable is shortest when it occurs as the first syllable of a polysyllabic word (Lehiste, 1972; Port, 1981). As Stromswold et al. (2002) suggest, polysyllabic shortening likely contributes to their findings and ours: the progressive active verb takes a syllabic suffix (*-ing*), which results in a polysyllabic word. This further differentiates the progressive active verb stem from its passive counterpart, which would undergo no shortening. Polysyllabic passive verbs (e.g., *patted*), however, do undergo shortening, which explains why polysyllabic passive verb stems did not differ in duration from their progressive active counterparts.

Polysyllabic shortening and phrase-final lengthening could drive the progressive active-passive verb stem duration difference reported in Stromswold et al. (2002) and in Experiment 1. However, to determine whether these two processes underlie the duration difference, and the contribution of each process to said difference, it is not sufficient to compare passive and progressive active verb stems alone. Experiment 2 (section 3.2) was designed to investigate the contribution of both phrase-final lengthening and polysyllabic shortening to the verb stem duration difference.

3.2. Experiment 2: Phrase-final lengthening vs. polysyllabic shortening

While it is likely that both polysyllabic shortening and phrase-final lengthening contribute to the verb stem duration difference found by Stromswold et al. (2002) and in Experiment 1, drawing a comparison between progressive active and passive sentences does not allow us to distinguish between the contributions of each. Because passive verbs are usually monosyllabic and prosodically phrase-final, and progressive verbs are polysyllabic and *not* phrase final, either polysyllabic shortening or phrase-final lengthening, or a combination of the two, could account for passive verb stem lengthening. To assess the relative contributions of these phonological processes, Experiment 2 compares verb stem duration in progressive active (*was bribing*, 4a), perfective active (*has bribed*, 4b), and passive (*was bribed*, 4c) sentences.

4. a) *The governor was bribing the mayor with a new car.*

b) *The governor has bribed the mayor with a new car.*

c) *The governor was bribed by the mayor with a new car.*

Because the progressive active verb stem (4a) is susceptible to polysyllabic shortening, and the perfective active verb stem (4b) is not, the progressive active verb stem should be shorter in duration than the perfective active verb stem if polysyllabic shortening contributes to the progressive active-passive verb stem duration difference. Similarly, because the passive verb stem (4c) is prosodically phrase-final, it is susceptible to phrase-final lengthening, while the perfective active verb stem (4b) is not, therefore if phrase-final lengthening contributes to the progressive active-passive verb stem duration difference, then the passive verb stem should be longer than the perfective active verb stem.

Furthermore, Experiment 2 compares an equal number of verbs that are 1) monosyllabic as passives and perfectives and take a voiceless inflection [t], 2) monosyllabic as passives and perfectives and take a voiced inflection [d], and 3) bisyllabic as passives and perfectives due to a syllabic inflection [ɪd]. Due to polysyllabic shortening, verb stems in verbs that take a syllabic *-ed* inflection should be shorter, and should be comparable in duration to their progressive active counterparts. If phrase-final lengthening is the primary phonological force driving the duration difference observed by Stromswold et al. (2002) and in Experiment 1, verb stems in progressive actives and perfective actives should be comparable in duration, but passives, which are subject to lengthening, should be longer than both ({perfective actives, progressive actives} < passives). In contrast, if polysyllabic shortening drives the duration difference, then progressive active verb stems should be shorter in duration than both perfective actives and passives, which should be comparable to one another (progressive actives < {perfective actives, passives}). If both processes contribute equally, then verb stems in progressive actives should be shortest due to polysyllabic shortening, followed by perfective actives, which would *not* undergo phrase-final lengthening or polysyllabic shortening, while passives should be longest due to the lack of polysyllabic shortening and the presence of phrase-final lengthening (progressive actives < perfective actives < passives).

3.2.1. Methods

Participants. Eight monolingual adult native American English speakers participated. Data from 1 subject were not analyzed due to low accuracy on catch trials. Of the remaining subjects, 5 were female and 2 were male. All participants were paid volunteers. The study was approved by the Rutgers University Institutional Review Board, and was carried out in accordance with the Declaration of Helsinki.

Materials. *Verbs.* Fifteen verbs were selected using the MRC Psycholinguistic Database (Coltheart, 1981) and the SUBTLEXus database (Brysbaert & New, 2009). Similar criteria were used to those in Experiment 1 (see section 4.1.1) with respect to frequency and semantic constraints. All verbs were felicitous in progressive active, perfective active, and passive sentences.

Voiced -*ed* inflections occur on verbs with voiced post-vocalic coda consonants (e.g., [k^{h}ɔɫ] → [k^{h}ɔɫd]), and, similarly, voiceless inflections accompany voiceless codas (e.g., [k^{h}ɪk] → [k^{h}ɪkt]). An equal number of verbs (n = 5) were chosen for each realization of the -*ed* inflection: a voiceless stop [t] as in *kicked* (*chase*, *kick*, *kiss*, *mock*, *poke*), a voiced stop [d] as in *bribed* (*bribe*, *call*, *fool*, *praise*, *tease*), and a syllabic inflection [ɪd] as in *guided* (*guard*, *guide*, *hunt*, *quote*, *treat*), resulting in 15 verbs total (see Appendix B.3). Syllabic -*ed* inflections follow alveolar stops ([t], [d]) in verb stem codas (e.g., [hʌnt] → [hʌn.tɪd]). To facilitate segmentation, preference was given to verbs that began with stop consonants, and that lacked complex consonant clusters in either onset or coda position.

Target sentences. Each verb occurred twice in a progressive active sentence (*The governor was bribing the mayor with a new car*; 4A), a perfective active sentence (*The governor has bribed the mayor with a new car*; 4B), and a passive sentence (*The governor was bribed by the mayor with a new car*; 4B). Passive sentences contained an agentive *by*-phrase. Thirty high frequency nouns were selected for their plausibility with the 15 verbs. Nouns were animate entities, either people or animals, that could serve either as the agent or the patient of an action to ensure that all sentences were semantically reversible. As in Experiment 1, each noun occurred once in subject position and object position for each construction. To prevent the second noun from occurring in utterance-final position, and to maintain a roughly equal number of syllables across all constructions[4], syntactically optional phrases (e.g., a prepositional phrase, an adverbial phrase) were included. On average, passive sentences were 12.39 syllables long, progressive actives were 12.07 syllables long, and perfective active sentences were 11.35 syllables long. Ninety total target sentences were tested (15 verbs x 3 syntactic frames x 2 noun positions; Appendix B.1).

Filler sentences. An additional 120 controlled active filler sentences were created: 90 of the filler sentences contained the same verbs and nouns as the target sentences (e.g., *The governor will bribe the mayor with a new car*, *The reporter suspected the governor bribed the mayor*; *The reporter suspected the governor bribing the mayor*), and the remaining 30 filler sentences used the same nouns as the target sentences, but with *has* as a possessive verb rather than a perfective auxiliary (e.g., *The governor has brilliant ideas for new laws*; Appendix B.2).

A total of 210 unique sentences (90 target sentences + 120 filler sentences) were tested.

Design. As in Experiment 1, sentence order was pseudorandomized so that consecutive sentences did not contain the same verb or the same noun in subject position. Additionally, no more than two sentences in a row were of the same construction type (progressive active, perfective active, or passive). 30 comprehension questions (2 per verb) were interspersed at irregular intervals (every 5-9 trials) to ensure participants processed each sentence prior to reading it aloud. Each comprehension question pertained

[4] This experiment was originally designed to test 6 syntactic constructions, including future tense sentences that were like the experimental sentences presented here, and two construction types that included a matrix verb and either a relative clause or a small clause. Because these sentence types do not speak to the contribution of phrase-final lengthening and polysyllabic shortening to the progressive active-passive verb stem duration difference, we have chosen to treat them as controlled fillers here.

to the preceding sentence: For example, the target sentence *The mayor has bribed the governor with a new car* was followed by the comprehension question *Did the governor bribe the mayor?*. Comprehension questions were followed by a filler trial. This was done to ensure that any surprise caused by the comprehension question did not influence the production of a target sentence. Six comprehension questions occurred after passive target sentences, another 6 occurred after progressive active targets, 5 occurred after perfective active targets, and the remaining 13 comprehension questions followed controlled filler sentences. After one list was pseudorandomized, that list was then reversed to create a second order. List presentation order was counterbalanced across participants.

Apparatus. The same apparatus as Experiment 1 was used (see 4.1.1).

Procedure. Prior to the experimental trials, participants first said each word in the experiment in isolation. Participants were instructed to first read the sentence silently, then aloud. This instruction was followed by 6 practice trials, including a practice comprehension question. A trial proceeded as follows. A fixation screen appeared for 1500 ms, followed by a blank screen for 250 ms. This was followed by a screen displaying a sentence, which persisted until the participant said the sentence, then pressed the spacebar. After an inter-trial-interval of 250 ms, the next trial began. The same trial procedure was used for comprehension questions on catch trials, except that participants responded “yes” or “no” (via the *Q* or *P* keys) to a question displayed on the screen rather than reading a sentence aloud. Halfway through the experimental trials, a 10-minute break occurred during which the participant completed the New York Times dialect survey (Katz, Andrews, & Buth, 2013). After the break, the trial procedure was repeated for the remainder of the experimental trials. There were 210 experimental trials in total, and the experiment took approximately 1 hour to complete.

Segmentation. To precisely capture morpheme duration, both morpheme onsets and offsets were marked by hand for each audio file using Praat (Boersma, 2001). Criteria for identifying the onset of each morpheme were the same as Experiment 1. Morpheme offsets were determined similarly, with some exceptions. For instance, when a period of silence reflecting closure was encountered, it was attributed to the preceding morpheme. This was done to be consistent with the segmentation strategies used in Experiment 1. The syntactically optional phrases (which were added to the end of experimental sentences to make them similar in length to filler sentences) were not segmented. Segmentation was carried out by 1 research assistant. See Figure 3.2.1. for an example segmented sentence.

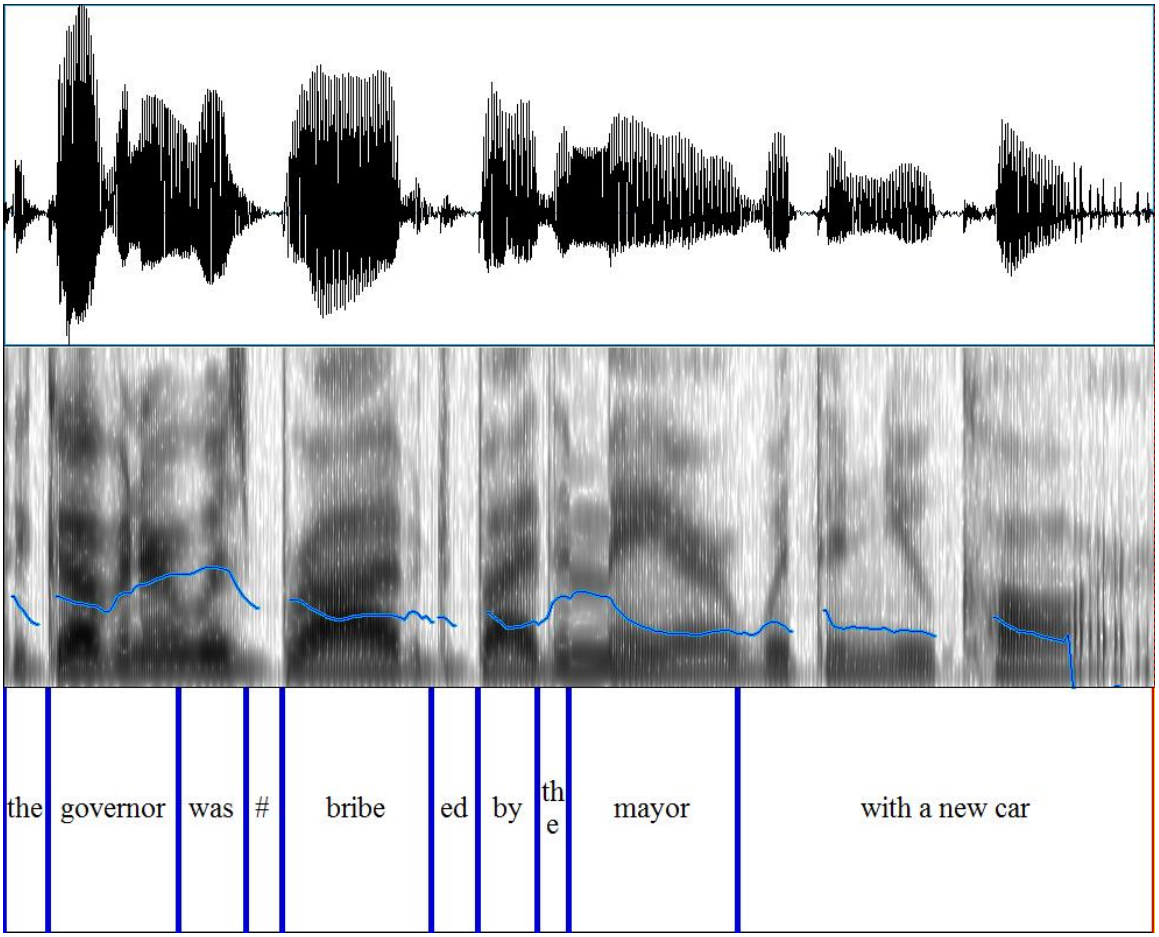


Figure 3.2.1. Experiment 2: Example of morpheme boundary placements for the sentence *The governor was bribed by the mayor with a new car*. Boundaries (shown in blue) mark the onset and offset of each morpheme per the segmentation criteria outlined in section 3.2.1.

Analysis. Productions from seven participants were analyzed. For each interval, defined from the start of a morpheme to the end of the morpheme, the interval duration was calculated. Durations of auxiliaries and verb stems were then analyzed using Bayesian linear mixed-effects models using the rstanarm package in R. The rstanarm package was used to approximate a posterior distribution using Markov Chain Monte Carlo sampling. Weakly informative priors were used for all models, and were initialized using the default rstanarm parameters. For stan objects, we used the median as a measure of central tendency, and standard deviation to report variance (per rstanarm’s default output). To control for variation due to individual differences or due to the items tested, random slopes and random intercepts were included for the syntactic frame (active or passive) by subjects and by item (verb). The standard deviation of the median absolute difference (MAD_{SD} hereafter) was reported for all Bayesian models.

The overall significance of random effects (e.g., individual differences by subject) were assessed via model comparison using the loo package, which allowed for comparison between two models and returned a value for the difference in *expected log predictive density* ($ELPD_{diff}$) between them. This value indicated which of the models predicts the data better: a negative value favored the first model, and a positive value favored the second.

The lsmeans package in R was used to obtain *p*-values, and to determine whether levels of the fixed effects differed from one another. Degrees of freedom were estimated using the Satterthwaite method (Satterthwaite, 1946), and *p*-values were adjusted for multiple comparisons using the Tukey method where applicable. Comparisons that were not significant according to statistical tests were further evaluated by determining the proportion of samples that fell within 1 just-noticeable difference (JND) of zero, and/or based on the size of the β coefficient. For duration, the JND range was set to 20% of the average duration (Klatt, 1976). The mean and standard error of the mean were reported as measures of central tendency and variance, respectively, obtained from the lsmeans package.

3.2.2. Results

Auxiliary. To determine whether passive auxiliary lengthening occurred in the current experiment, auxiliary duration was compared between progressive active and passive sentences only. Perfective active sentences were excluded because the difference in phonological form of the auxiliary *has* likely results in a difference in auxiliary duration, for reasons that are not of interest here.

Consistent with Stromswold et al. (2002) and Experiment 1, passive auxiliaries were 8.87 ms longer (*M* = 162.48 ms, *SE* = 2.29 ms) than progressive active auxiliaries (*M* = 153.60 ms, *SE* = 2.13 ms). The magnitude of passive auxiliary lengthening was significant ($\chi^2 = 8.79$, $p = 0.003$; see Table 3.2.1). Though the size of the difference is small (8.87 ms), it appears that passive auxiliary lengthening is a relatively robust phenomenon.

Table 3.2.1.

Experiment 2: Bayesian Linear Mixed-Effects Model Summary for Progressive Active and Passive Auxiliary Duration

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			158.00	1.80
Random Effects	*SD*	*r*		
Subject (n = 7)				
Intercept	14.70	–		
Progressive Active	6.50	-0.04		
Verb (n = 10)				
Intercept	15.10	–		
Progressive Active	5.00	-0.24		
Residual	26.10	–		

Fixed Effects	*Median β*	*β MAD$_{SE}$*				
Intercept	162.40	6.50				
Error *SD*	26.10	0.90				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Progressive Active	-9.0	3.40	-2.43	.02	-14.78	-3.06

Note: Four divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Consistent with Experiment 1, the data were predicted better by a model that accounted for variation due to individual speakers ($ELPD_{diff}$ = -36.60, *SE* = 8.3), which suggests that there were individual differences in passive auxiliary lengthening (see Figure 3.2.2). All 7 speakers lengthened auxiliaries in passive sentences, which is unlikely to occur by chance, according to a sign test (cumulative binomial: $p = .008$). To determine which speakers produced significant passive auxiliary lengthening, a model that included syntax and participant as fixed effects was tested. Passive auxiliary lengthening was marginal for participants 4 and 6 ($ps < .10$), was significant for participant 7 ($z = 2.59$, $p = .01$), and was not reliable for the remaining 4 participants (see Table 3.2.2). For all participants, at least 98% of samples fell within 1 JND of zero, which suggests that the effect size is either so small as to have no practical significance, or may be near zero (Median $\beta = -9.0$).

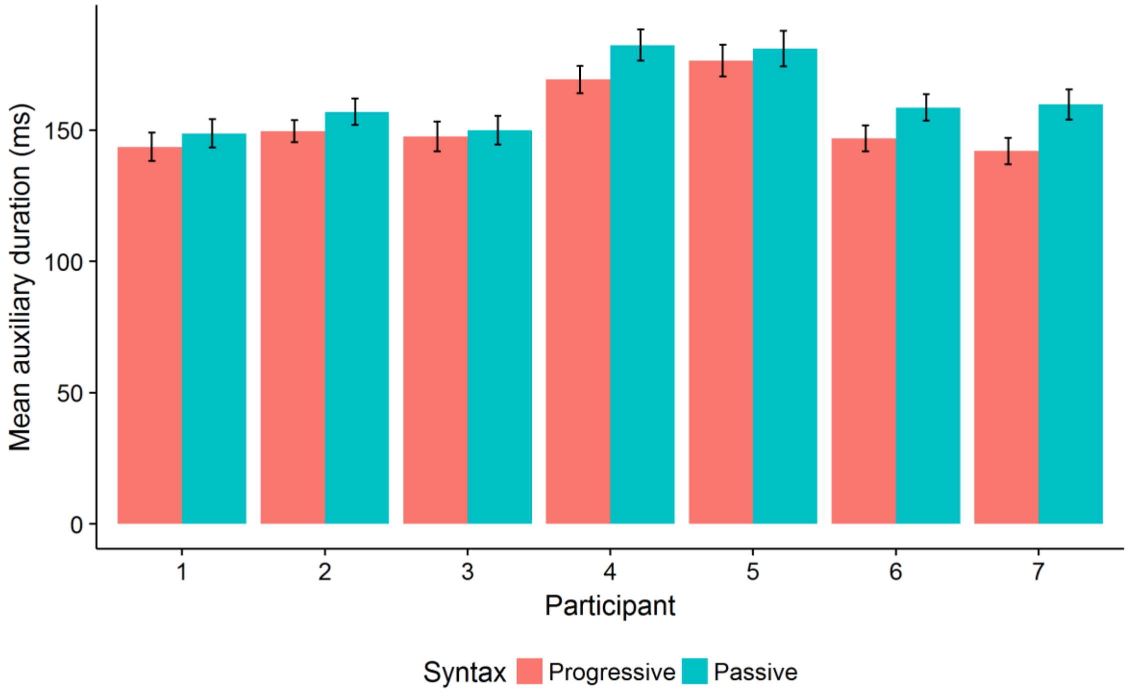


Figure 3.2.2. Experiment 2: Mean auxiliary duration in active and passive sentences, for individual participants. Error bars indicate standard error of the mean.

Table 3.2.2.

Experiment 2: Progressive Active and Passive Auxiliary Duration by Subject

Subject	Sex	*Mean β*	*β SE*	*z*	*p*
Subject 1: Passive vs. Progressive	Female	5.78	6.67	0.87	.39
Subject 2: Passive vs. Progressive	Female	7.64	6.99	1.09	.27
Subject 3: Passive vs. Progressive	Female	2.52	6.74	0.37	.71
Subject 4: Passive vs. Progressive	Female	13.00	6.80	1.91	.06
Subject 5: Passive vs. Progressive	Male	4.56	6.86	0.67	.51
Subject 6: Passive vs. Progressive	Male	11.78	6.80	1.73	.08
Subject 7: Passive vs. Progressive	Female	17.87	6.90	2.59	.01

Note: One divergent transition occurred during sampling, which indicates a possibility that the sampler was biased.

Verb stem. In order to determine the contribution of phrase-final lengthening and polysyllabic shortening to the progressive active-passive verb stem duration difference, we performed an analysis where verbs that took a syllabic *-ed* inflection were excluded. This was done so that bisyllabic passive and perfective active verbs, which would be subject to polysyllabic shortening, do not confound the analysis.[5]

Overall, verb stems were 15.87 ms longer in passive sentences (*M* = 282.75 ms, *SE* = 5.96) than in perfective active sentences (*M* = 266.88 ms, *SE* = 5.83 ms), and perfective active verb stems were 26.18 ms longer than progressive active verb stems (*M* = 240.70 ms, *SE* = 3.77 ms; see Figure 3.2.3). A model that accounted for variation in verb stem duration across syntactic constructions predicted the data better than a model that did not ($ELPD_{diff}$ = -1.40, *SE* = 1.3), and the progressive active-passive duration difference appeared to drive this difference (*z* = 3.83, *p* = .01; see Table 3.2.3). Duration differences across the other syntactic comparisons were smaller such that passive verb stems were not reliably longer than perfective actives, and perfective active verb stems were not reliably longer than progressive active verb stems (*ps* > .10). At least 99.8% of samples fell within 1 JND of zero for perfective active vs. passive sentences, which suggests that any effect of phrase-final lengthening was too small to be of practical significance.

[5] An analysis containing all verbs had similar results, except that passive verb stems were marginally longer than perfective active verb stems (*t*(8.35) = 2.64, *p* = .07).

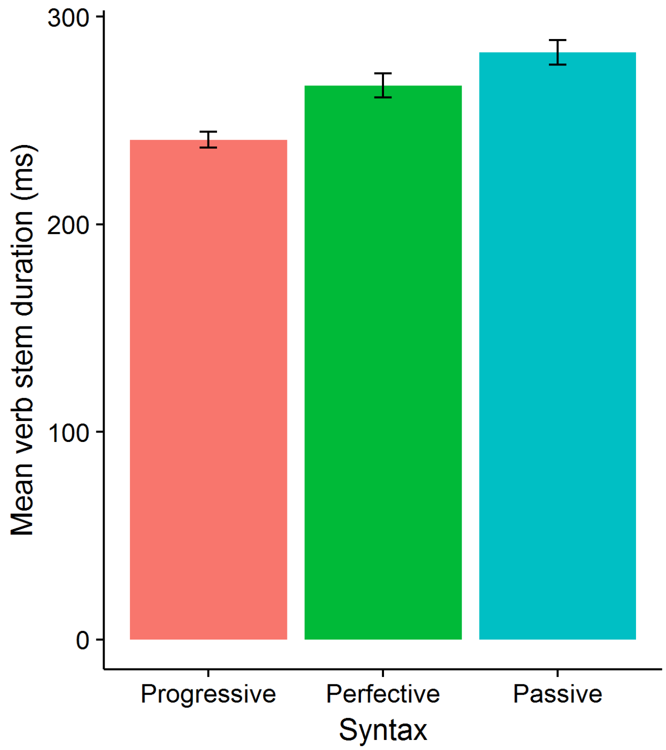


Figure 3.2.3. Experiment 2: Mean verb stem duration in progressive active, perfective active, and passive sentences. Error bars indicate standard error of the mean.

Table 3.2.3.

Experiment 2: Bayesian Linear Mixed-Effects Model Summary for Syntax and Verb Stem Duration

				Median	*MAD_{SD}*
Mean Posterior Predictive Distribution				263.20	2.6
Random Effects	*SD*	*r*			
Subject (n = 7)					
Intercept	39	–	–		
Perfective Active	22	0.00	–		
Progressive Active	21	-0.70	-0.03		
Verb (n = 10)					

Intercept	39	–	–			
Perfective Active	16	0.15	–			
Progressive Active	21	-0.30	0.20			
Residual	38	–	–			
Fixed Effects	*Median β*	*β MAD_{SE}*				
Intercept	284.00	18.60				
Error *SD*	37.80	1.40				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Passive vs. Perfective	16.35	10.59	1.55	.27	-33.19	1.49
Progressive vs. Passive	43.30	11.32	3.83	< .01	-61.00	-23.99
Perfective vs. Progressive	26.95	14.39	1.87	.15	–	–

Note: Eleven divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Individual differences: Verb stem duration. Consistent with Experiment 1 (section 3.1), all speakers produced longer passive verb stems than progressive active verb stems.[6] This is unlikely to be due to chance alone per a sign test (cumulative binomial: p = .008). However, the duration of the perfective active verb stem, as compared to progressive active and passive verb stems, was not consistent across speakers (see Figure 3.2.4).

For some speakers (e.g., participants 2 and 6), the perfective active verb stem was, on average, as short or shorter in duration than the progressive active verb stem, though these differences were not reliable (*ps* > .10). For other speakers (n = 5), perfective active

[6] Patterns in verb stem duration for each subject were essentially the same when all 15 verbs were included in the analysis.

verb stem durations were similar to those of passive verb stems (*ps* > .10). To determine which verb stem duration differences across syntactic constructions were reliable for each subject, pairwise comparisons were performed on a model that included syntax and participant as fixed factors. For 4 participants, the progressive active-passive verb stem duration difference was significant (*ps* < .05), but for 3 of the 7 participants the difference was too small to be of practical significance (Table 3.2.4). The difference between passive and perfective verb stem durations was significant for two participants (*ps* < .05), indicating the presence of phrase-final lengthening. The difference between perfective and progressive active verb stem durations was significant for 4 participants (*ps* < .01), which reflects the contribution of polysyllabic shortening.

Participant 1 appeared to demonstrate both polysyllabic shortening and phrase-final lengthening, reflected by differences in verb stem duration across each sentence type comparison. Participants 4, 5, and 7, on the other hand, primarily demonstrate polysyllabic shortening, as evidenced by duration differences between passive and progressive active verb stems, and progressive and perfective active verb stems, but no reliable difference between passive and perfective active verb stems. Note that one participant (6) produced a reliable duration difference between perfective active and passive sentences, but not between other sentence types. Participant 6 is also the only speaker who produced shorter perfective active verb stems than progressive active verb stems. However, this likely was not due to phrase-final lengthening, because phrase-final lengthening would also be expected in the comparison between progressive actives and passives, which were not reliably different in duration for participant 6. Note, however, that fewer than 21.9% of samples in the comparison between progressive active and passive sentences fell within 1 JND of zero, which may suggest that a very small effect was present, but was not large enough to be of practical significance. No participant produced a pattern consistent with *only* phrase-final lengthening contributing to differences in verb stem duration without also demonstrating polysyllabic shortening. These findings indicate individual differences in the magnitude of polysyllabic shortening and phrase-final lengthening. Furthermore, they suggest that polysyllabic shortening contributes to the progressive active-passive verb stem duration difference more than phrase-final lengthening does.

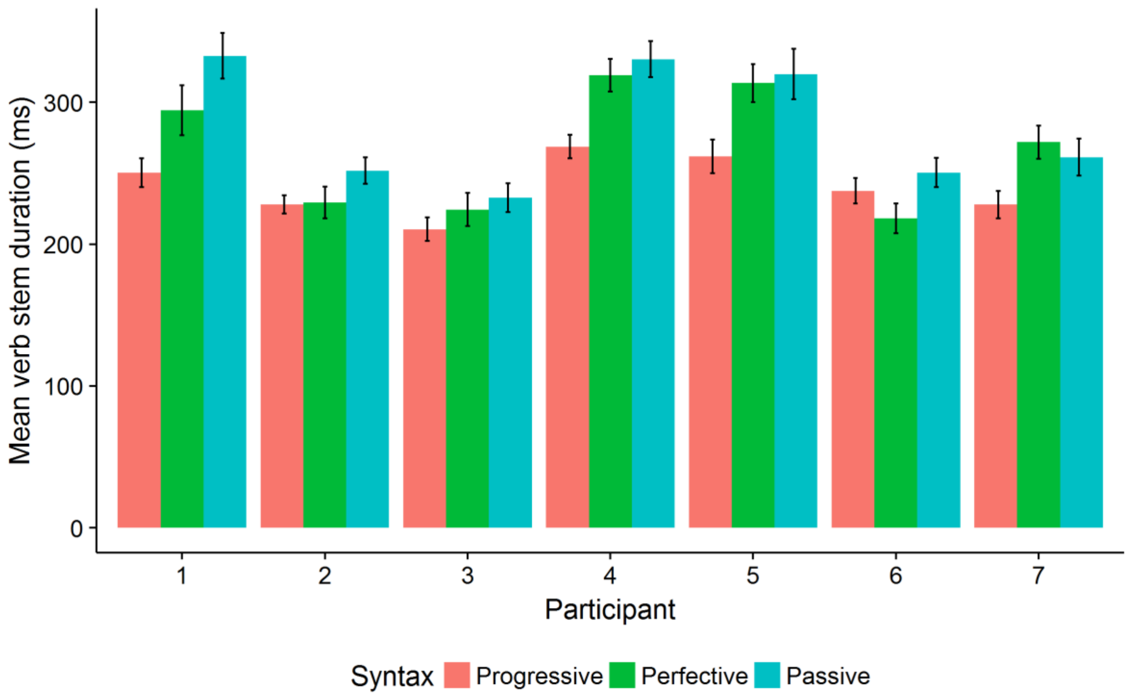


Figure 3.2.4. Experiment 2: Mean verb stem duration in progressive active, perfective active, and passive sentences, for each individual participant. Error bars indicate standard error of the mean.

Table 3.2.4.

Experiment 2: Pairwise Comparisons for Syntax and Verb Stem Duration by Subject

Subject	Sex	*Mean β*	*β SE*	*z*	*p*
Subject 1	Female				
Passive vs. Perfective		37.51	13.00	2.89	.01
Progressive vs. Passive		80.63	13.65	5.91	< .01
Perfective vs. Progressive		43.12	13.98	3.08	.01
Subject 2	Female				
Passive vs. Perfective		22.95	12.89	1.78	.18
Progressive vs. Passive		24.46	13.74	1.78	.18
Perfective vs. Progressive		1.51	13.81	0.11	.99
Subject 3	Female				
Passive vs. Perfective		10.20	12.85	0.79	.71
Progressive vs. Passive		24.27	13.63	1.78	.18
Perfective vs. Progressive		14.08	14.20	0.99	.58
Subject 4	Female				
Passive vs. Perfective		11.26	13.17	0.86	.67
Progressive vs. Passive		63.11	13.86	4.55	< .01
Perfective vs. Progressive		51.86	14.38	3.61	< .01
Subject 5	Male				
Passive vs. Perfective		8.02	13.36	0.60	.82
Progressive vs. Passive		59.42	14.13	4.21	< .01
Perfective vs. Progressive		51.40	14.60	3.52	< .01
Subject 6	Male				

Passive vs. Perfective		34.37	13.26	2.59	.03
Progressive vs. Passive		14.81	13.70	1.08	.53
Perfective vs. Progressive		-19.56	13.89	-1.41	.34
Subject 7	Female				
Passive vs. Perfective		-11.74	13.38	-0.88	.65
Progressive vs. Passive		35.19	14.00	2.51	.03
Perfective vs. Progressive		46.92	14.15	3.32	< .01

Note: Seven divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Next, we test further for the presence of phrase-final lengthening by investigating the duration of syllabic -*ed* inflections.

Duration of the syllabic -ed inflection. To test for only the effect of phrase-final lengthening on syllable duration, we compared the duration of the syllabic -*ed* inflection [ɪd] across passive and perfective active sentences. In passive sentences, the syllabic -*ed* inflection is the final syllable in the prosodic phrase, and therefore is subject to phrase-final lengthening. In perfective actives, the syllabic -*ed* inflection is *not* phrase-final, and therefore is not subject to lengthening.

Syllabic -*ed* inflections were 23.96 ms longer on average for passive verbs (M = 131.04 ms, SE = 3.79 ms) than for perfective active verbs (M = 107.08 ms, SE = 3.59 ms; Figure 3.2.5). The duration of the inflection was analyzed using a linear mixed-effects model where syntax (perfective active or progressive) was specified as a fixed effect, and random slopes and intercepts were specified for syntax by subject and by item (verb). The magnitude of the duration difference was reliable (z = 2.49, p = .01), which indicates that phrase-final lengthening took place on syllabic -*ed* inflections in passive sentences

(Table 3.2.5). Note, however, that all samples fell within 1 JND of 0, which suggests that this is a small—but reliable—effect.

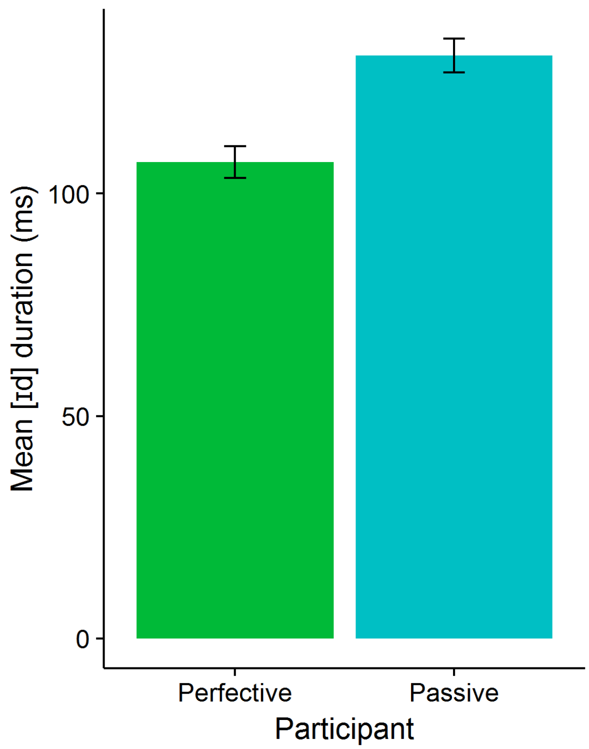


Figure 3.2.5. Experiment 2: Mean syllabic *-ed* inflection duration for perfective active and passive verbs. Error bars indicate standard error of the mean.

Table 3.2.5

Experiment 2: Bayesian Linear Mixed-Effects Model Summary for Syntax and Duration of the Syllabic -ed Inflection

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			155.80	3.9
Random Effects	*SD*	*r*		
Subject (*n* = 7)				
Intercept	31.30	–		

Perfective Active	32.30	-0.54				
Verb (*n* = 5)						
Intercept	6.50	–				
Perfective Active	6.90	-0.25				
Residual	32.10	–				
Fixed Effects	*Median* *β*	*β* MAD_{SD}				
Intercept	172.70	12.10				
Error *SD*	31.90	2.0				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Perfective vs. Passive	-33.50	12.70	2.49	.01	-55.89	-11.96

Six of the 7 participants produced longer syllabic -*ed* inflections in passive sentences (see Figure 3.2.6), which is marginally unlikely to have occurred by chance per a sign test (cumulative binomial: $p = .06$). Given that verb stem duration revealed individual differences in phrase-final lengthening, we tested a model that included an interaction between syntax and subject, with random slopes specified for subject and syntax by item (verb). For 4 of the 7 participants, the magnitude of phrase-final lengthening was reliable (*ps* < .05, Table 3.2.6). This provides further evidence that the magnitude of phrase-final lengthening can vary for different speakers.

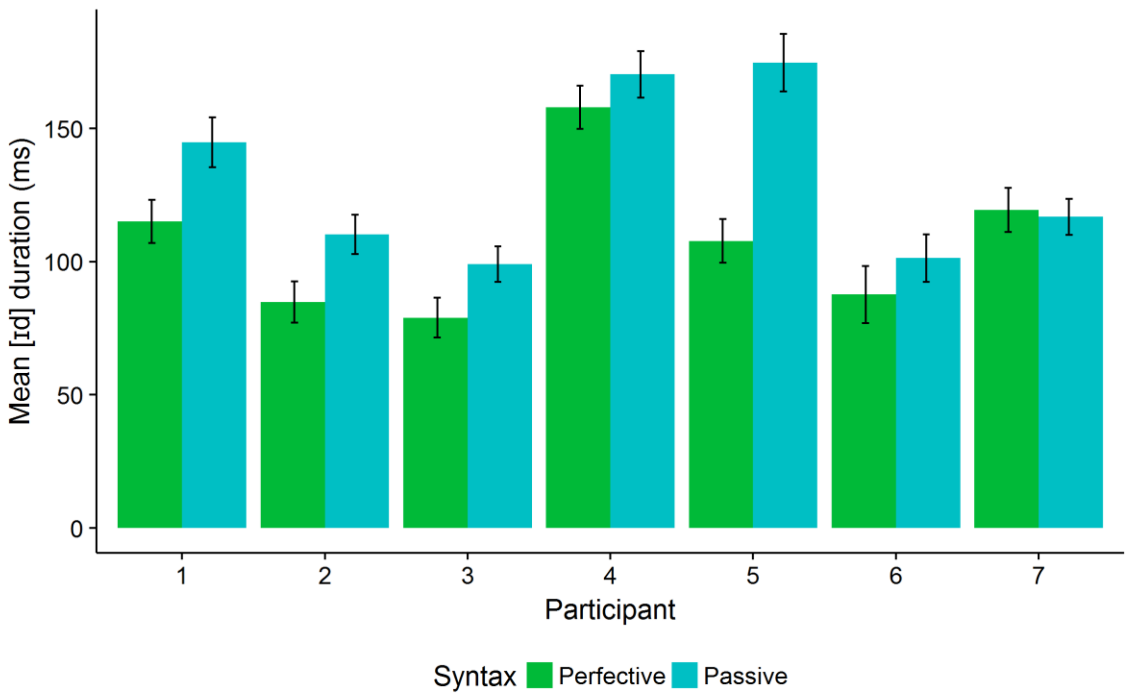


Figure 3.2.6. Experiment 2: Mean duration of syllabic *-ed* inflection [ɪd] in perfective active and passive verbs, for each participant. Error bars indicate standard error of the mean.

Table 3.2.6.

Experiment 2: Perfective Active and Passive Syntax and Syllabic -ed Inflection Duration by Subject

Subject	Sex	*Mean β*	*SE β*	*z*	*p*
Subject 1: Passive vs. Perfective	Female	50.23	13.89	3.62	< .01
Subject 2: Passive vs. Perfective	Female	28.70	14.40	1.99	< .05
Subject 3: Passive vs. Perfective	Female	22.96	14.48	1.59	.11
Subject 4: Passive vs. Perfective	Female	9.34	14.52	0.64	.52
Subject 5: Passive vs. Perfective	Male	94.95	14.51	6.54	< .01
Subject 6: Passive vs. Perfective	Male	-2.64	14.51	-0.18	.86

Subject 7: Passive vs. Perfective	Female	30.64	14.32	2.14	.03

Next, we evaluate the duration of verb stems in bisyllabic verbal contexts across syntactic constructions.

Duration of verb stems in bisyllabic verbs. To confirm that polysyllabic shortening and phrase-final lengthening are the phonological processes that impact verb stem duration in our stimuli, we compared verb stem duration for verbs that were bisyllabic in all constructions (e.g., *guide*, *hunt*, etc.). Because both the *-ed* and *-ing* inflections add a syllable for these verbs, they undergo polysyllabic shortening across the sentence types tested here. Furthermore, because the verb stem would not be the phrase-final syllable for passive sentences containing these verbs, phrase-final lengthening would not affect verb stem duration.

Passive verb stems (M = 221.20 ms, SE = 4.20 ms) were 4.41 ms longer than progressive active verb stems (M = 216.29 ms, SE = 4.36 ms), which were 2.77 ms longer than perfective active verb stems (M = 214.02 ms, SE = 4.20 ms; see Figure 3.2.7). To confirm that these differences were not reliable, the duration of verb stems in bisyllabic contexts were analyzed using a linear mixed-effects model with syntax as a fixed factor, and random slopes and intercepts specified by subjects and by items (Table 3.2.7). As predicted, verb stem duration was similar across syntactic constructions when they appeared in a bisyllabic context ($ps > .10$). All samples fell within 1 JND of zero, which suggests that effect estimate may be near 0 (*Median* $\beta_{Passive\text{-}Progressive} = -7.2$, *Median* $\beta_{Passive\text{-}Perfective} = -4.4$), or that it is too small to be of practical consequence. This supports the hypothesis that phrase-final lengthening and polysyllabic shortening affected verb stem duration in Experiment 1, and in verbs that do not take a syllabic *-ed* inflection in the current experiment.

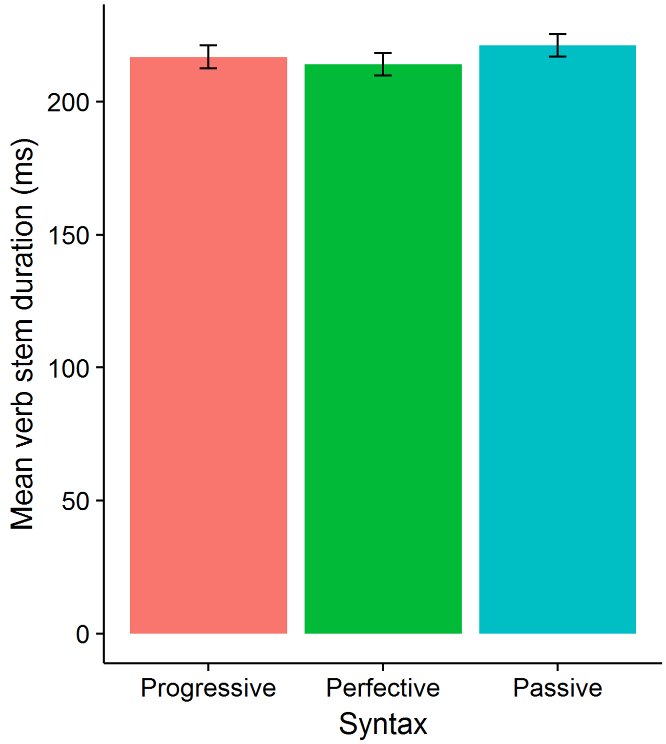


Figure 3.2.7. Experiment 2: Mean verb stem duration for bisyllabic verbs across syntactic constructions. Error bars indicate standard error of the mean.

Table 3.2.7.

Experiment 2: Bayesian Linear Mixed-Effects Model Summary for Syntax and Verb Stem Duration for Bisyllabic Verbs

				Median	*MAD*$_{SD}$
Mean Posterior Predictive Distribution				217.30	2.20
Random Effects	*SD*	*r*			
Subject (*n* = 7)					
Intercept	22.30	–	–		

Perfective Active	6.3	0.00	—			
Progressive Active	8.3	-0.02	0.05			
Verb (*n* = 5)						
Intercept	18.10	—	—			
Perfective Active	10.00	0.05	—			
Progressive Active	10.90	0.17	0.46			
Residuals	24.00	—	—			
Fixed Effects	*Median β*	*β* MAD_{SD}				
Intercept	221.20	11.20				
Error *SD*	23.90	1.30				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Passive vs. Perfective	7.18	6.36	1.13	.50	-17.33	3.35
Progressive vs. Passive	4.43	7.05	0.63	.80	-15.56	7.09
Perfective vs. Progressive	-2.75	7.20	-0.38	.92	—	—

Note: Eight divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Verb stem: Deletion of the -ed inflection. English speakers show a tendency to reduce consonant clusters in coda position (e.g., [k^{h}ɪkt] → [k^{h}ɪk]), particularly for coda final [t] and [d] (Labov, 1989). Because verbs in two of the tested constructions formed complex codas with the *-ed* inflection, we looked at the rate at which the *-ed* inflection was deleted in passive sentences and perfective active sentences. For completeness, we also included verbs that took a syllabic *-ed* inflection, though only one syllabic *-ed* inflection was deleted (in a perfective active sentence). Speakers deleted the *-ed* inflection more frequently in perfective active sentences (30.00% deletion rate) than in passive sentences (14.30% deletion rate). A Bayesian logistic mixed-effects regression analysis was performed to determine whether the difference in deletion rates was significant. For this analysis, the dependent variable was whether the inflection was present (1) or absent (0). Progressive actives were excluded from the analysis because they always bear a syllabic inflection, though note that one subject *did* delete a syllabic -*ed* inflection in a perfective active sentence. The difference in -*ed* inflection deletions was not reliable ($\beta = 1.03$, $z = 1.63$, $p = .10$; see Table 3.2.8). However, the proportion of samples that fell below 0 was 94.7%, which may suggest a tendency for speakers to drop inflections more often in perfective active sentences.

Table 3.2.8.

Experiment 2: Bayesian Logistic Mixed-Effects Model Summary Deletion of the -ed Inflection by Syntax

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			0.8	0.0
Random Effects	*SD*	*r*		
Subject (*n* = 7)				
Intercept	1.40	–		
Perfective	1.00	-0.11		
Active				
Verb (*n* = 15)				
Intercept	1.90	–		

Perfective Active	1.10	0.21				

Fixed Effects	*Median β*	*β MAD*$_{SD}$				
Intercept	2.80	0.80	–	–	–	–
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Perfective Active	1.03	0.63	1.63	.10	-2.03	0.028

However, deletion rates varied by subject, indicating that some speakers were more likely to delete the *-ed* inflection than others, both overall and in perfective active sentences relative to passives (see Figure 3.2.8). For all but 1 of the 7 speakers, perfective active verb stem codas were deleted more frequently (cumulative binomial: $p = .06$). A model that accounted for variation due to subjects predicted the data better than a model that did not ($ELPD_{diff} = -23.80$, $SE = 6.9$), indicating that individual differences were present. Rates of *-ed* inflection deletion were significantly different across perfective active and passive sentences for 3 participants ($ps < .05$), but were similar for the remaining 4 participants (Table 3.2.9).

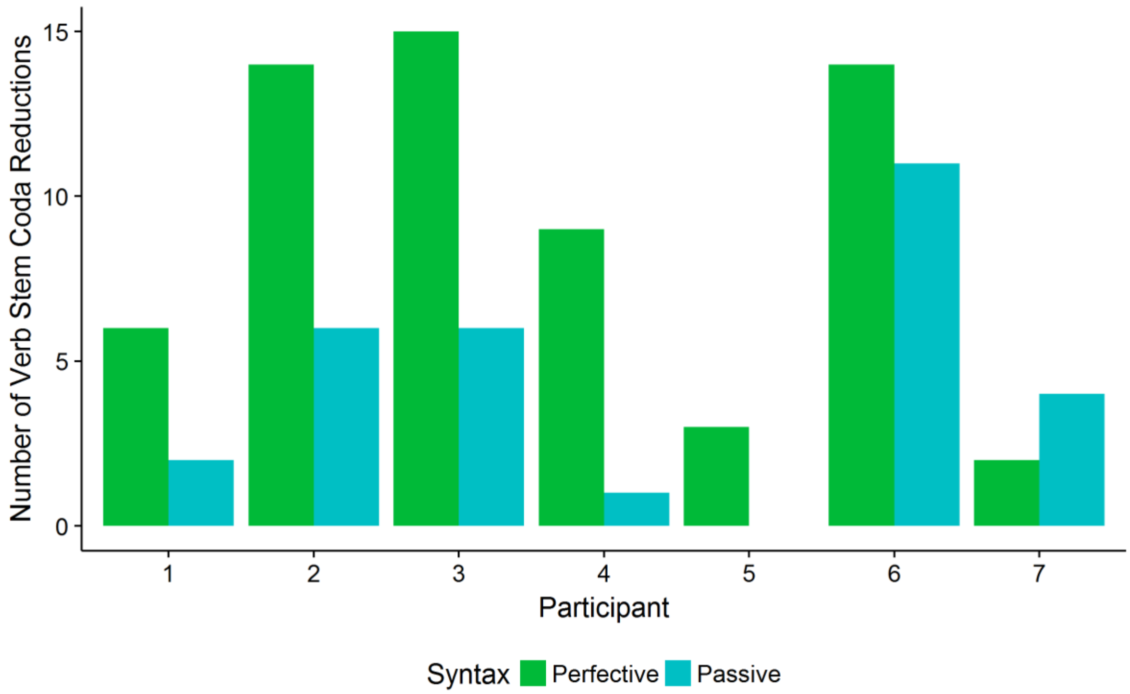


Figure 3.2.8. Experiment 2: Number of instances of *-ed* deletion for each subject, and each syntactic construction containing *-ed* inflected verbs.

Table 3.2.9.

Experiment 2: Perfective Active and Passive -ed Inflection Deletion by Subject

Subject	Sex	*Mean β*	*SE β*	*z*	*p*
Subject 1: Passive vs. Perfective	Female	0.99	0.77	1.29	.20
Subject 2: Passive vs. Perfective	Female	1.59	0.81	1.96	< .05
Subject 3: Passive vs. Perfective	Female	1.76	0.83	2.13	.03
Subject 4: Passive vs. Perfective	Female	2.52	1.09	2.32	.02
Subject 5: Passive vs. Perfective	Male	1.88	1.43	1.32	.19
Subject 6: Passive vs. Perfective	Male	0.48	0.77	0.63	.53
Subject 7: Passive vs. Perfective	Female	-1.02	1.07	-0.95	.34

In sum, verb stem duration was affected by syntactic construction and by individual differences across speakers. As in Experiment 1, passive verb stems that did not take a syllabic inflection were consistently longer than active verb stems. Our findings suggest that, while phrase-final lengthening does occur, it is less robust than polysyllabic shortening. This suggests that both phrase-final lengthening and polysyllabic shortening contribute to the progressive active-passive verb stem duration difference observed in Experiment 1, with polysyllabic shortening playing a more substantial role. The relative duration of the verb stem in perfective active sentences as compared to passive and progressive active sentences varied across speakers. This suggests that the contribution of both of these processes, and phrase-final lengthening in particular, varies dependent on the speaker.

3.2.3. Discussion

The results of Experiment 2 make clear that both phrase-final lengthening and polysyllabic shortening contribute to progressive active-passive verb stem duration difference. Verb stem duration was consistently longer in passive sentences as compared to progressive active sentences for verbs that form monosyllabic passive verbs. The combined role of phrase-final lengthening and polysyllabic shortening in the duration difference is supported by the lack of a duration difference between progressive active verb stems and passive verb stems that took a syllabic inflection [ɪd]. Because the *-ed* inflection is syllabic in this case, the passive verb stem is subject to polysyllabic shortening. Similarly, because the verb stem is not the phrase-final syllable, phrase-final lengthening did not increase the duration of the passive verb stem.

However, variation in perfective active verb stem duration suggests that, while the sum of the effects of these processes is consistent, the relative contribution of each of these processes is not consistent across speakers. While the duration of perfective active verb stems fell between the duration of progressive active and passive verb stems, overall there was individual variation in the extent to which perfective active verb stems were longer than progressive active verb stems, or shorter than passive verb stems. As mentioned previously, one possibility is that there are individual differences in the magnitude of phrase-final lengthening and/or polysyllabic shortening. Given these findings, it is unlikely that either process is sufficient by itself to yield a reliable cue to syntax.

For verbs where the *-ed* inflection was not syllabic, and therefore formed a complex coda, deletion of the inflection occurred more frequently in perfective active sentences than in passive sentences for 3 speakers. This may have occurred due to differences in the phonological environment, or because the inflection is redundant in perfective active sentences from an information perspective: the auxiliary *has* must be followed by a verb bearing an *-ed* inflection in order to form a grammatical English sentence. Speakers may reduce or delete informationally redundant portions of the speech signal in order to make communication more efficient (Jaeger, 2010; Turk, 2010; see section 3.3.3 for discussion on this point). However, differences in the phonological environment between perfective active and passive sentences may also have driven the observed difference in *-ed* inflection deletion. In our passive sentences, the verbal inflection was immediately followed by the preposition *by*, which begins with a bilabial stop [b], while the verbal inflection was followed by the determiner *the*, beginning with the voiced dental fricative [ð], in perfective active sentences. Coda consonant cluster deletion is more likely when the deleted consonant ([t], or [d]) is followed by a consonant with a similar place of articulation (Fasold, 1972). Thus, speakers may have dropped the -*ed* inflection in perfective active sentences because its alveolar place of articulation is similar to the dental place of articulation in the subsequent determiner. In order to confirm that the surrounding context yielded the difference in -*ed* deletion, the subsequent phonological context would need to be held constant across syntactic constructions.

Another factor that was not controlled in this study, but may have affected verb stem duration, is the phonology of the verbs themselves. In this study, there were 8 verbs with voiced post-vocalic consonants in verb stems, and 7 verbs with voiceless post-vocalic consonants in coda position. Verb stems containing voiced post-vocalic consonants may lengthen because vowels that precede voiced consonants usually undergo lengthening (Klatt, 1976). However, verbs with voiced codas in this study also differed from one another in other ways: 3 of the 5 verbs that took the [t] -*ed* inflection had stop consonants in coda position (*kick*, *mock*, *poke*), whereas only 1 of the 5 verbs that took the [d] -*ed* inflection had a stop consonant coda (*bribe*). For verbs that took the syllabic [ɪd] -*ed* inflection, it is unclear how the bisyllabic verbs would syllabify, and whether or not the post-vocalic consonant would be in the coda of the verb stem (e.g., [t^hri.tɪd]). Because the semantics of the verbs were prioritized over controlling for phonological properties in this study, it is not possible to determine whether manner of articulation (stop, liquid, fricative) or voicing of the coda consonant drive the difference in duration here. For the same reason, we could not address the role of vowel duration in Experiment

2 because the verbs we selected for semantic and syntactic reasons were not controlled with respect to the verb stem vowel (e.g., diphthongs were present in some verbs).

Verb stem duration was likely affected by the surrounding phonological environment, both with respect to verb stem codas and to differences in post-verbal portions of the sentences across syntactic constructions. Because the materials in the current experiment were not primarily designed to minimize phonological differences across sentences, there is no way to determine which part of the verb stem undergoes a duration change across progressive active and passive sentences. Given that Stromswold et al. (2002) identified a duration difference between progressive active and passive verb stem vowels, it is likely that a change in duration of the verb stem vowel drives the duration difference. To determine whether this is the case, Experiment 3 was conducted to measure each of the verb stem syllable segments and identify which segment undergoes lengthening consistently.

3.3. Experiment 3: Segment duration

Stromswold et al. (2002) found evidence that the vowel undergoes lengthening in passive sentences. To further investigate which segment undergoes lengthening, speakers produced progressive active (*was picking*, 5a), perfective active (*has picked*, 5b), and passive sentences (*was picked*, 5c).

5. a) *The little girl was picking a pumpkin to carve.*

b) *The new husband has picked a wedding ring.*

c) *The red apple was picked a couple of hours ago.*

As in Experiment 2, speakers produced progressive actives, perfective actives, and passives. Unlike Experiment 2, speakers said truncated passive to ensure that the passive verb stem occurred in a prosodically phrase-final position. Another difference between the current experiment and Experiment 2 is that only verbs that took [t] or [d] -*ed* inflections were tested. Furthermore, because the goal of this study was to analyze acoustic information at a fine-grained level, phonological restrictions were prioritized over semantic constraints. This means that the sentences in the current experiment were not as semantically sound as those in Experiments 1 and 2, and were not balanced as well due to a lack of semantic reversibility.

The verb stem in each production was segmented into onset, vowel, and coda, and the duration of each segment was then compared across sentence types to determine which verb stem segment undergoes a duration change.

3.3.1. Methods

Participants. Eight monolingual adult native American English speakers participated. Of these, 5 were females and 3 were males. All participants were paid volunteers. The study was approved by the Rutgers University Institutional Review Board, and was carried out in accordance with the Declaration of Helsinki.

Materials. *Verbs.* Eighteen regular verbs were selected using the MRC Psycholinguistic Database (Coltheart, 1981). All verbs were frequent, regular verbs of the form CVC (see Appendix C3), but sentences were not semantically reversible (e.g., a verb could take an inanimate patient). All verbs were felicitous as passives (Levin, 1993), and took the -*ed* passive inflection. To facilitate segmentation, onsets were restricted to non-sibilant consonants, and the codas [v] and [ð] were avoided (Turk et al., 2006). Preference was given to verbs with stop consonants in onset position. In order to control for inherent vowel duration, all verbs contained only the phonologically short vowels [æ] as in *cash*ed (9 verbs: *ban*, *can*, *cash*, *fan*, *pack*, *pass*, *ram*, *tag*, *wrap*), [ɪ] as in *picked* (4 verbs: *kiss*, *miss*, *pick*, *pin*), or [ʌ] as in *hugged* (5 verbs: *hug*, *hum*, *hush*, *rush*, *tug*). Though [æ] is phonetically long (Peterson & Lehiste, 1960), it is treated as a phonologically short vowel in American English, as evidenced by the fact that syllables with nuclei containing only [æ] (e.g., [bæ]) do not meet the minimal word restriction in English (Mayro et al., 2016). Furthermore, including a phonetically long vowel in half of the verbs tested (9 of 18) allowed for a comparison between the duration of vowels that are both phonologically and phonetically short to that of a vowel that is phonologically short, but phonetically long. For each vowel type (phonetically long or short), half of the verbs contained stop consonant codas, and the other half had non-stop codas (nasals, sibilants). In addition, for each vowel type, half of the verbs had voiced coda consonants and the other half had voiceless coda consonants, because coda voicing can affect the duration of the preceding vowel (Klatt, 1976).

Target sentences: Syntax. Each verb occurred in three syntactic frames: progressive active (*The little girl was picking a pumpkin to carve*; 5a), perfective active (*The new husband has picked a wedding ring*; 5b), and passive (*The red apple was picked a couple of hours ago*; 5c). Verbs were the same across syntactic constructions, but different nouns were used across sentences.

Sentences: Phonological environment. Considerable care was taken to ensure that the phonological environment surrounding the verb was the same in all sentence types. With respect to the number of syllables, there were no between-sentence statistical differences in the number of syllables either before the verb or in the entire sentence. This was done to ensure the phonological environment surrounding the verb was as similar as possible in all sentence types. In all 3 syntactic constructions, verbs were followed by the vowel [ə] in either the determiner *a* or in a preposition (e.g., *around*) in order to facilitate segmentation, and to investigate whether the phonological environment affected the *-ed* deletion rate in Experiment 2. Verbs were preceded by the consonant [z] in the end of an auxiliary (e.g., *was*), except for sentences that did not contain an auxiliary (e.g., filler sentences), where verbs were preceded by a plural noun with a voiced coda (e.g., *players*) to keep the preverbal phonological environment consistent.

Filler sentences. An additional sentence type, simple past active (e.g., *The careful players picked a new teammate*), was also included in the original design. Because the semantics of these sentences differ, and due to the lack of an auxiliary, for this study the past active sentences were treated as controlled filler sentences. These 18 filler sentences used the same verbs that appeared in the target sentences.

Sentence distribution. Because controlled fillers were included in the experiment, and also bear the *-ed* inflection, we included 3 times as many progressive active sentences (54) as either perfective active (18) or passive sentences (18) to balance the number of sentences containing the 2 inflectional morphemes *-ing* and *-ed* in the experiment overall. This resulted in 90 total target sentences (Appendix C.1). An additional 18 filler sentences were constructed that used similar syntactic frames, with different verbs (Appendix C.2), yielding 126 unique sentences in total (90 target sentences + 18 controlled fillers + 18 additional fillers).

Design. Sentence order was pseudorandomized so that no sentence was followed by another sentence containing the same verb, or the same noun in subject position. No more than three passive sentences occurred consecutively, no more than four progressive active target sentences occurred consecutively, and no more than two perfective active sentences occurred consecutively. There were at most seven consecutive target sentences with no fillers in between. Thirty-six comprehension questions (2 per verb) were interspersed at irregular intervals (every 4-7 trials) to ensure participants processed each sentence prior to saying it. Each comprehension question pertained to the preceding sentence: For example, the target sentence *The new baby was missing a clean blanket* was followed by the comprehension question *Was the new baby missing a clean blanket?*. Comprehension questions were followed by a filler trial. This was done to ensure that any surprise caused by the comprehension question did not influence the production of a target sentence. Six comprehension questions occurred after passive target sentences, another 18 occurred after progressive active targets, 6 occurred after perfective active targets, and the remaining 6 comprehension questions followed controlled filler sentences. Each sentence appeared twice during the experiment, once in the first half of the experimental and once in the second half. The halves were pseudorandomized separately. Two lists were constructed by combining each pseudorandomized half of the trials to create the first list, then reversing that list to create a second order. List presentation order was counterbalanced across participants.

Apparatus. The same apparatus as Experiments 1 and 2 was used (see 3.1.1).

Procedure. The trial procedure was identical to that of Experiment 2. After completing the experimental trials once, a 10-minute break occurred during which the participant completed the New York Times dialect survey (Katz, Andrews, & Buth, 2013). After the break, the experimental trials were repeated using a different pseudorandomized list order. There were 252 trials in total (126 total sentences x 2 presentations). The experiment took approximately 1 hour to complete.

Segmentation. Because the purpose of the experiment was to identify the verb stem phone(s) implicated in the progressive active-passive duration difference, the onset, vowel, and coda of each verb stem were segmented by hand using Praat (Boersma, 2001). Consonants were segmented using similar criteria to that used in Experiments 1 and 2, except that consonant closures were treated differently: the closure was attributed to the consonant it preceded. Vowel onset boundaries were placed at the first zero-crossing before a smooth peak. Criteria for vowel offset boundary placement varied depending on the type of segment following the vowel. When the coda consonant was a fricative, the vowel offset boundary was placed at the first zero-crossing before the lowest smooth peak. For cases where the coda consonant was a sonorant, the vowel offset boundary was placed at the zero-crossing preceding the first period that marked a qualitative change in the waveform. When the post-vocalic consonant was a stop, the vowel offset boundary was placed at the last zero-crossing before the closure of the stop. Segmentation was carried out by 2 research assistants. Inter-rater reliability for vowel duration in 214 sentences was high ($r(212) = 0.93$, $p < .0001$). See Figure 3.3.1. for an example segmented verb.

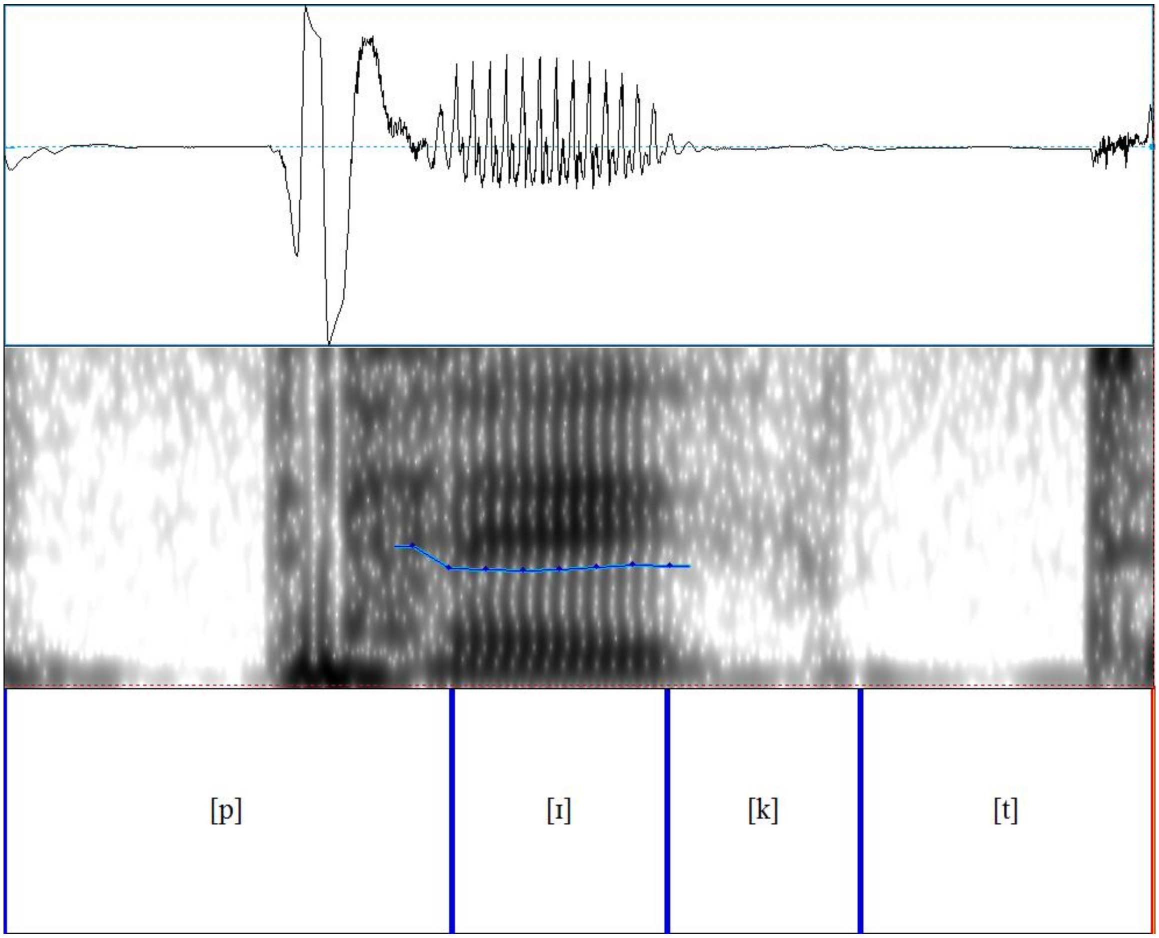


Figure 3.3.1. Experiment 3: Example of segment boundary placements for the sentence *The red apple was picked a couple of hours ago.* Boundaries (shown in blue) mark the onset and offset of each segment per the segmentation criteria outlined in section 3.3.1.

Analysis. For each interval, defined from the start of a segment to the end of the segment, the interval duration was calculated. Verb stem duration was calculated separately by summing the duration of each verb stem segment. Durations were then analyzed using Bayesian linear mixed-effects models using the rstanarm package in R. The rstanarm package was used to approximate a posterior distribution using Markov Chain Monte Carlo sampling. Weakly informative priors were used for all models, and were initialized using the default rstanarm parameters.

To control for variation due to individual differences or due to the items tested, random slopes and random intercepts were included for the syntactic frame (active or passive) by subjects and by item (verb). The overall significance of random effects (e.g., individual differences by subject) were assessed via model comparison using the loo package, which allowed for comparison between two models and returned a value for the difference in *expected log predictive density* ($ELPD_{diff}$) between them. This value indicated which of the models predicts the data better: a negative value favored the first model, and a positive value favored the second.

The lsmeans package in R was used to obtain *p*-values, and to determine whether levels of the fixed effects differed from one another. Degrees of freedom were estimated using the Satterthwaite method (Satterthwaite, 1946), and *p*-values were adjusted for multiple comparisons using the Tukey method where applicable. Comparisons that were not significant according to statistical tests were further evaluated by determining the proportion of samples that fell within 1 just-noticeable difference (JND) of zero, and/or based on the size of the β coefficient. For duration, the JND range was set to 20% of the average duration (Klatt, 1976).

3.3.2. Results

Verb stem duration. Consistent with Experiments 1 and 2, passive verb stems (M = 295.19 ms, SE = 3.90 ms) were 10.82 ms longer than perfective active verb stems (M = 284.37 ms, SE = 3.79 ms), which were 17.35 ms longer than progressive active verb stems (M = 267.02 ms, SE = 1.74 ms; see Figure 3.3.2). However, model comparison revealed that a model without a fixed effect of syntax was a slightly better predictor of the data than one that included a fixed effect of syntax ($ELPD_{diff}$ = 0.2, SE = 1.2).

Consistent with Experiment 2, the progressive active-passive verb stem duration difference was reliable, where passive verb stems were significantly longer than progressive active verb stems ($z = 3.42$, $p < .01$; see Table 3.3.1.)[7].

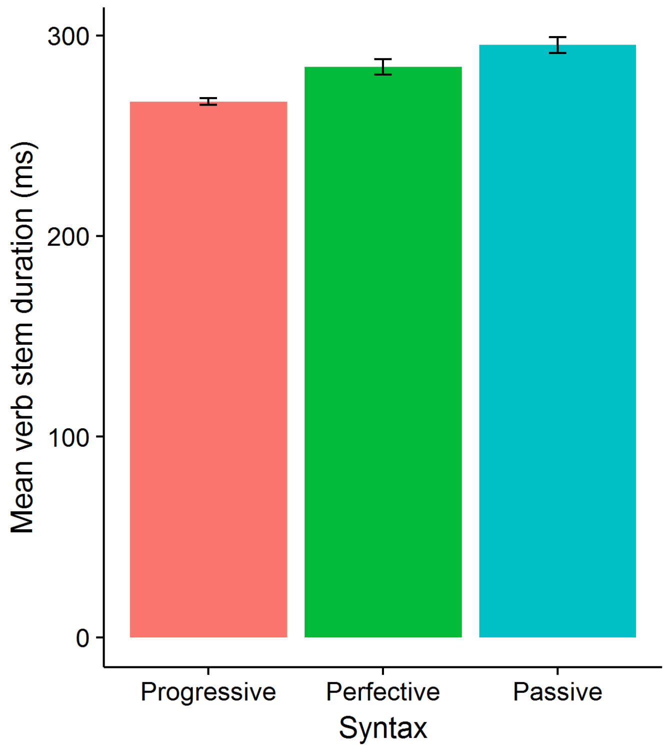


Figure 3.3.2. Experiment 3: Mean verb stem duration for progressive active, perfective active, and passive sentences. Error bars indicate standard error of the mean.

Table 3.3.1.

Experiment 3: Bayesian Linear Mixed-Effects Model Summary for Verb Stem Duration by Syntax

	Median	MAD_{SD}
Mean Posterior Predictive Distribution	276.20	1.30

[7] Note that the verb stem duration effect was found despite different segmentation strategies between the current study and Experiments 1 and 2, which indicates that the progressive active-passive was not the result of attributing the consonant closure for the -*ed* inflection to the verb stem in Experiments 1 and 2.

Random Effects	*SD*	*r*	
Subject (n = 8)			
Intercept	31	–	–
Perfective Active	14	-0.04	–
Progressive Active	16	-0.53	0.25
Verb (n = 18)			
Intercept	42	–	–
Perfective Active	18	-0.35	–
Progressive Active	26	-0.29	0.01
Residual	36	–	–

Fixed Effects	*Median β*	*β MAD$_{SD}$*
Intercept	297.00	14.20
Error *SD*	35.60	0.70

					Posterior Interval	
	β	*SE*	*z*	*p*	5%	95%
Passive vs. Progressive	11.29	7.28	1.55	.27	-23.36	0.38
Passive vs. Progressive	29.14	8.53	3.42	< .01	-43.07	-15.10
Perfective vs. Progressive	17.85	10.30	1.73	.19	–	–

Note: Nine divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Consistent with the verb stem duration results of Experiments 1 and 2, there were individual differences in the pattern of verb stem durations across syntactic constructions (see Figure 3.3.3). For 7 of the 8 participants, passive verb stems were longer than progressive active verb stems, which was unlikely to occur by chance (cumulative binomial: $p = .04$). This difference in duration was reliable for 6 of the 8 participants (*ps* < .01; Table 3.3.2). Perfective active verb stems were significantly longer than progressive active verb stems for 3 participants (*ps* < .01), and marginally longer for 2 participants. Passive verb stems were significantly longer than perfective verb stems for 1 participant (participant 2, $p = .006$) and marginally longer for another (participant 8, $p = .08$). These 2 participants both demonstrated a pattern that was not observed in Experiment 2: they demonstrated no verb stem duration difference between progressive active and perfective actives, but *did* show differences between the other sentence types. This suggests that phrase-final lengthening had a greater impact than polysyllabic shortening in their verb stem durations. Three participants showed the opposite pattern, where polysyllabic shortening appeared to have a greater impact than phrase-final lengthening, as evidenced by differences between progressive active and passive verb stems, and between perfective and progressive active verb stems.

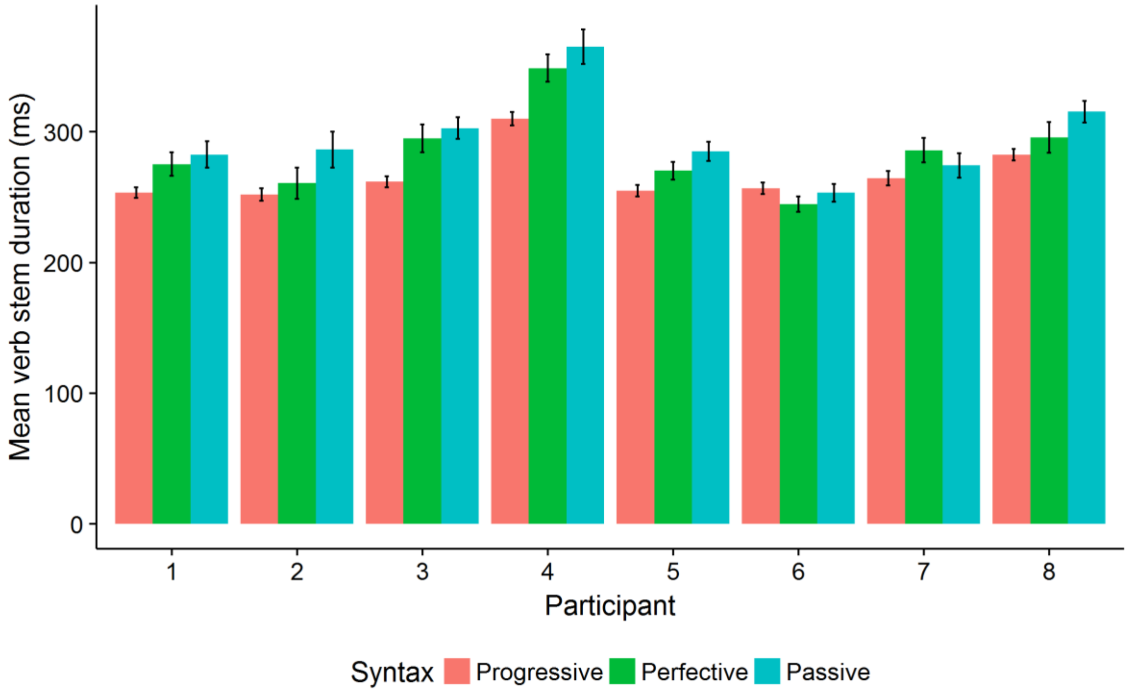


Figure 3.3.3. Experiment 3: Mean verb stem vowel duration for each syntactic construction, by participant. Error bars indicate standard error of the mean.

Table 3.3.2.

Experiment 3: Pairwise Comparisons for Verb Stem Vowel Duration across Syntactic Constructions by Participant

Subject	Sex	*β*	*SE*	*z*	*p*
Subject 1	Female				
Passive vs. Perfective		7.66	9.42	0.81	.70
Progressive vs. Passive		29.28	9.41	3.11	< .01
Perfective vs. Progressive		21.62	10.39	2.08	.09
Subject 2	Female				
Passive vs. Perfective		29.41	9.61	3.06	< .01
Progressive vs. Passive		37.44	9.38	3.99	< .01
Perfective vs. Progressive		8.04	10.47	0.77	.72

Subject 3	Female				
Passive vs. Perfective		7.65	9.49	0.81	.70
Progressive vs. Passive		41.39	9.13	4.53	< .01
Perfective vs. Progressive		33.74	10.24	3.29	< .01
Subject 4	Male				
Passive vs. Perfective		15.88	9.61	1.65	.22
Progressive vs. Passive		54.50	9.34	5.84	< .01
Perfective vs. Progressive		38.63	10.24	3.77	< .01
Subject 5	Male				
Passive vs. Perfective		14.69	9.66	1.52	.28
Progressive vs. Passive		30.33	9.37	3.24	< .01
Perfective vs. Progressive		15.64	10.33	1.52	.28
Subject 6	Male				
Passive vs. Perfective		8.90	9.42	0.95	.61
Progressive vs. Passive		-3.20	9.27	-0.35	.94
Perfective vs. Progressive		-12.11	10.28	-1.18	.47
Subject 7	Female				
Passive vs. Perfective		-11.54	9.40	-1.23	.44
Progressive vs. Passive		9.94	9.34	1.06	.54
Perfective vs. Progressive		-12.48	10.20	2.11	.09
Subject 8	Female				
Passive vs. Perfective		19.87	9.45	2.10	.09
Progressive vs. Passive		32.72	9.33	3.51	< .01

Perfective vs. Progressive	12.85	10.30	1.25	.43

Note: Five divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Overall, the duration of verb stem segments was affected by syntax (Figure 3.3.4.). A model with an interaction between syntax and verb stem segment (onset, vowel, coda) was better able to predict the data than an otherwise equivalent model without the interaction (*ELPDdiff* = -90.7, *SE* = 13.8). Pairwise comparisons indicate that the interaction was driven primarily by changes in the duration of the vowel, which were significantly different between perfective active and progressive active sentences, and between passive and progressive active sentences (*ps* < .001), but was only slightly different between perfective active and passive sentences (z = 1.84, p = .16; see Table 3.3.3).

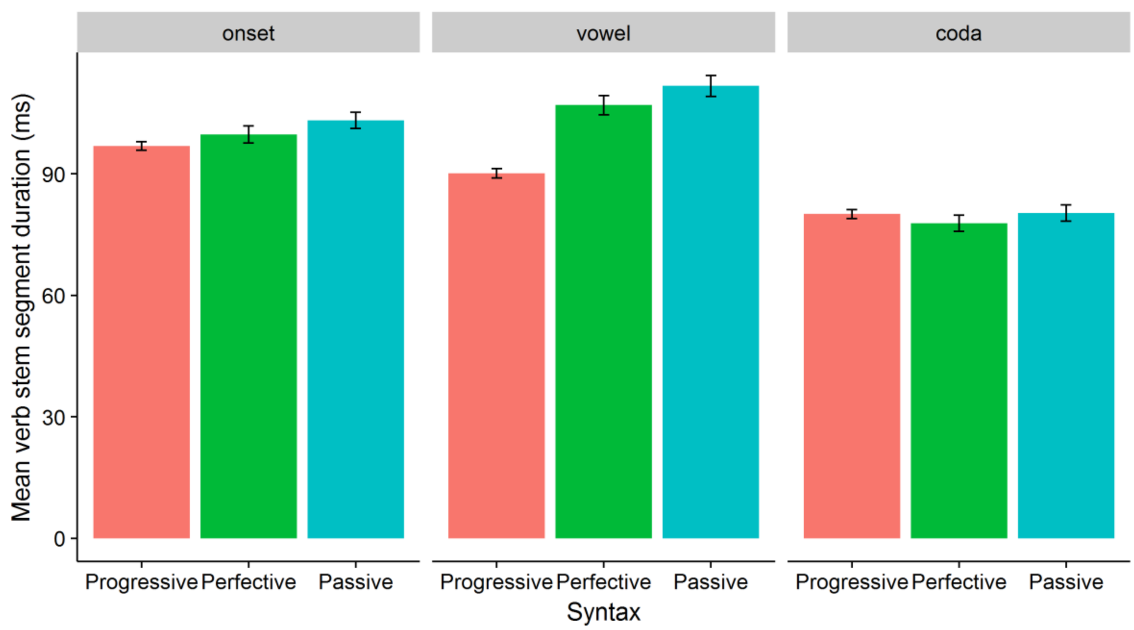


Figure 3.3.4. Experiment 3: Mean duration of progressive active, perfective active, and passive verb stem segments. Error bars indicate standard error of the mean.

Table 3.3.3.

Experiment 3: Bayesian Linear Mixed-Effects Model Summary for Verb Segment Duration by Syntax

					Median	*MAD$_{SD}$*
Mean Posterior Predictive Distribution					92.0	0.4
Random Effects	*SD*	*r*				
Subject (n = 8)						
Intercept	11.0	–	–	–	–	
Perfective Active	5.1	0.00	–	–	–	
Progressive Active	5.9	-0.38	0.24	–	–	
Onset	10.4	-0.09	-0.09	0.05	–	
Vowel	13.3	0.05	0.03	-0.26	-0.12	
Verb (n = 18)						
Intercept	24.8	–	–	–	–	
Perfective Active	6.2	-0.08	–	–	–	
Progressive Active	12.1	-0.02	0.06	–	–	
Onset	33.4	-0.38	-0.02	-0.14	–	
Vowel	43.0	-0.55	-0.27	-0.26	0.32	
Residual		–	–	–	–	
Fixed Effects	*Median β*	*β MAD$_{SD}$*				
Intercept	80.9	7.0				
Error *SD*	19.8	0.2				
					Posterior Interval	
	β	*SE*	*z*	*p*	5%	95%

Syntax						
Passive vs. Perfective	3.80	2.48	1.53	.28	-7.25	1.99
Prog. vs. Passive	9.71	3.66	2.66	.02	-6.88	5.59
Perfective vs. Prog.	5.91	4.07	1.45	.31	—	—
Onset						
Passive vs. Perfective	3.57	2.87	1.25	.43	-4.80	3.09
Prog. vs. Passive	6.54	3.81	1.72	.20	-9.09	-2.77
Perfective vs. Prog.	2.97	4.21	0.71	.76	—	—
Vowel						
Passive vs. Perfective	5.22	2.84	1.84	.16	-6.45	1.35
Prog. vs. Passive	21.95	3.82	5.74	< .01	-24.69	-18.14
Perfective vs. Prog.	16.73	4.23	3.96	< .01	—	—
Coda						
Passive vs. Perfective	2.59	2.82	0.92	.63	—	—
Prog. vs. Passive	0.64	3.85	0.17	.99	—	—
Perfective vs. Prog.	-1.95	4.22	-0.46	.89	—	—

In the next paragraphs, the impact of syntactic construction on each verb stem segment will be considered in turn.

Onset duration. Onsets were 3.53 ms longer in passive verb stems (*M* = 103.22 ms, *SE* = 1.99 ms) than perfective active onsets (*M* = 99.69 ms, *SE* = 2.13 ms), which were in turn 2.82 ms longer than progressive active onsets (*M* = 96.87 ms, *SE* = 1.02 ms; see Figure 3.3.3.). However, none of the above differences were reliable (see Table 3.3.2), and all samples fell within 1 JND of zero, which suggests that the effects were so small that they are of no practical significance, or that the effect estimate is near zero ($Median\ \beta_{Perfective\text{-}Passive}$ = -3.4, $Median\ \beta_{Progressive\text{-}Passive}$ = -6.3).

Differences in verb stem onset duration varied for different speakers (Figure 3.3.5.). For all speakers, passive verb stem onsets were longer than progressive active verb stem onsets, but the magnitude of the difference varied considerably across speakers, and for 5 speakers the effect was likely small enough to have on practical significance. Perfective active verb onset duration was inconsistent across speakers: for 3 subjects, perfective active verb stem onsets were similar in duration to either the duration of passive or progressive active onsets; for 2 subjects, perfective active verb onsets were shorter than progressive active verb onsets, and for another 2 subjects, perfective active verb stem onsets were longer than passive verb stem onsets. A model that included a random effect of subject was a better predictor of onset duration than one that did not ($ELPD_{diff}$ = -100.1, *SE* = 16.2), indicating individual differences in onset duration. Pairwise comparisons revealed that passive verb stem onsets were marginally longer than progressive active verb stem onsets for 3 of the 8 participants (*ps* = .05; Table 3.3.4). For comparisons across other construction types, the effect was so small as to have no practical significance.

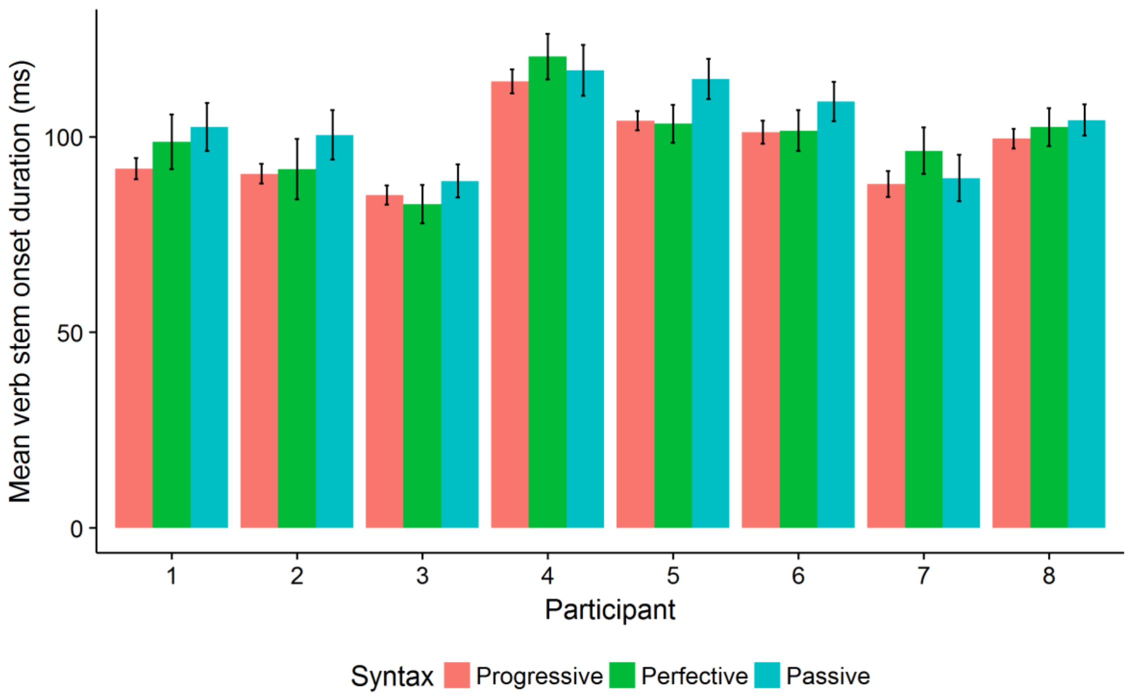


Figure 3.3.5. Experiment 3: Mean verb stem onset duration in progressive active, perfective active, and passive sentences, for each individual participant. Error bars indicate standard error of the mean.

Table 3.3.4.

Experiment 3: Pairwise Comparisons for Verb Stem Onset Duration across Syntactic Constructions by Subject

Subject	Sex	*β*	*SE*	*z*	*p*
Subject 1	Female				
Passive vs. Perfective		3.49	5.35	0.65	.79
Progressive vs. Passive		10.02	4.48	2.24	.07
Perfective vs. Progressive		6.53	4.87	1.34	.37
Subject 2	Female				
Passive vs. Perfective		8.80	5.59	1.57	.26
Progressive vs. Passive		10.61	4.77	2.23	.07

Perfective vs. Progressive		1.81	4.86	0.37	.93
Subject 3	Female				
Passive vs. Perfective		5.78	5.59	1.03	.56
Progressive vs. Passive		3.73	4.68	0.80	.71
Perfective vs. Progressive		-2.05	4.93	-0.42	.91
Subject 4	Male				
Passive vs. Perfective		-4.46	5.60	-0.80	.71
Progressive vs. Passive		2.37	4.79	0.49	.87
Perfective vs. Progressive		6.83	4.84	1.41	.34
Subject 5	Male				
Passive vs. Perfective		11.40	5.59	2.04	.10
Progressive vs. Passive		10.31	4.66	2.21	.07
Perfective vs. Progressive		-1.09	4.89	-0.22	.97
Subject 6	Male				
Passive vs. Perfective		7.43	5.42	1.37	.36
Progressive vs. Passive		7.75	4.68	1.66	.22
Perfective vs. Progressive		0.32	4.70	0.07	.99
Subject 7	Female				
Passive vs. Perfective		-6.75	5.44	-1.24	.43
Progressive vs. Passive		1.43	4.67	0.31	.95
Perfective vs. Progressive		8.18	4.85	1.69	.21
Subject 8	Female				
Passive vs. Perfective		1.82	5.49	0.33	.94

Progressive vs. Passive	4.45	4.74	0.94	.62
Perfective vs. Progressive	2.63	4.82	0.55	.85

Note: Three divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Vowel duration. As predicted, verb stem vowels were 4.76 ms longer in passive sentences (*M* = 111.69 ms, *SE* = 2.58 ms) than in perfective active sentences (*M* = 106.93 ms, *SE* = 2.34 ms), and were 16.85 ms longer in perfective active sentences than in progressive active sentences (*M* = 90.08 ms, *SE* = 1.16 ms; see Figure 3.3.3). The difference between perfective active and passive verb stem vowel durations was marginally significant (*p* = .09). This is consistent with phrase-final lengthening taking place on the vowel of the phrase-final syllable (Klatt, 1976). However, differences in vowel duration between progressive active and passive vowels, and between perfective and progressive active vowels, were both reliable (*ps* <.001; see comparisons in Table 3.3.2). These findings are consistent with those of Experiment 2, and further suggest that polysyllabic shortening plays a greater role than phrase-final lengthening in the progressive active-passive verb stem duration difference. Additionally, 99.95% of samples fell within 1 JND of zero for the comparison between perfective active vs. passives sentences, whereas only 30.6% of samples for the comparison between progressive active and passive verb stem vowel durations fell within 1 JND of zero. This provides further evidence that polysyllabic shortening had a greater influence on verb stem vowel duration than did phrase-final lengthening.

Consistent with the verb stem duration results here and in Experiments 1 and 2, there were significant differences among individuals in the pattern of verb stem vowel durations across syntactic constructions (see Figure 3.3.6.), as indicated by the greater predictive value of a model with a term for the random effect of subject vs. one without ($ELPD_{diff}$ = -344.40, *SE* = 24.0). Passive verb stem vowels were always longer on average than progressive verb stem vowels, which a sign test indicates is unlikely to occur by chance (cumulative binomial: $p = .004$). However, the magnitude of the difference varied across speakers: pairwise comparisons revealed that 6 out of 8 participants lengthened passive verb stems vowels relative to progressive active verb stem vowels to a significant extent (*ps* < .001; see Table 3.3.5). Perfective active verb stem vowels were longer than progressive active verb stem vowels for 6 out of 8 speakers (*ps* < .05). For 6 speakers, the effect was so small as to be of no practical significance, while 2 speakers reliably lengthened passive verb stem vowels relative to perfective active verb stem vowels (*ps* < .05). The same 6 speakers who lengthened passive verb stem vowels relative to progressive active verb stem vowels also produced reliably longer perfective active verb stem vowels than progressive active verb stem vowels. This is consistent with the findings collapsed across subjects that, while phrase-final lengthening was present for at least some speakers, polysyllabic shortening appeared to have greater influence on verb stem vowel duration, and verb stem duration by extension.

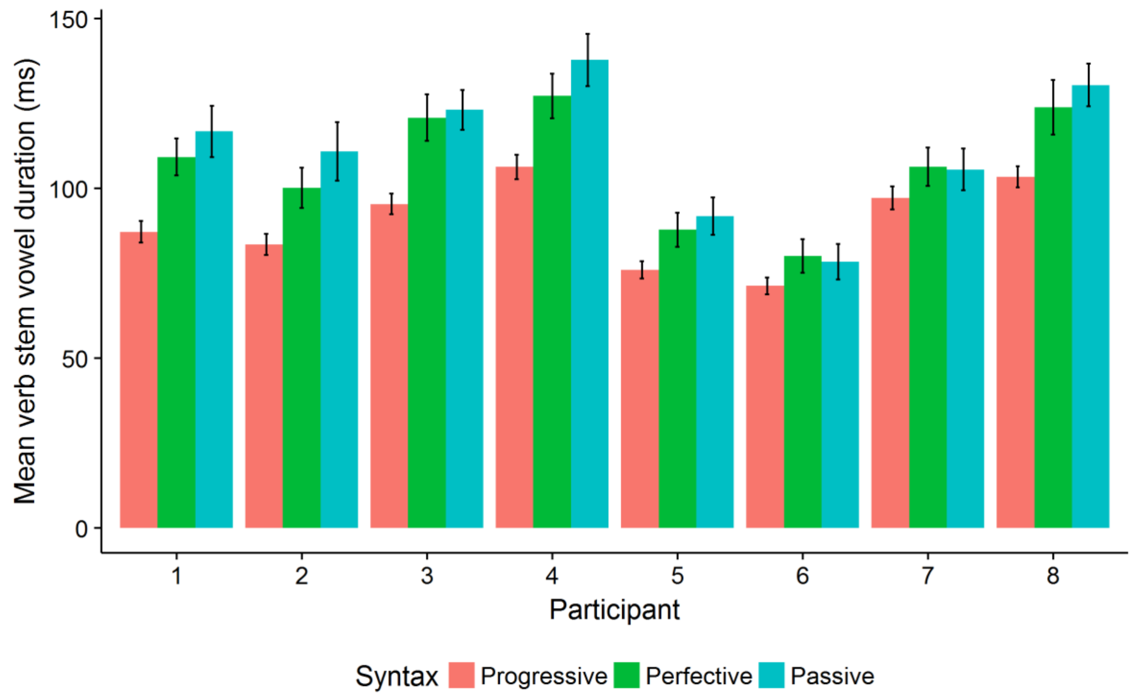


Figure 3.3.6. Experiment 3: Mean verb stem vowel duration in progressive active, perfective active, and passive sentences, for each individual participant. Error bars indicate standard error of the mean.

Table 3.3.5.

Experiment 3: Pairwise Comparisons for Verb Stem Vowel Duration across Syntactic Constructions by Subject

Subject	*Mean β*	*β SE*	*z*	*p*
Subject 1				
Passive vs. Perfective	7.51	4.90	1.53	.28
Progressive vs. Passive	29.39	4.05	7.26	< .001
Perfective vs. Progressive	21.87	4.21	5.20	< .001
Subject 2				
Passive vs. Perfective	12.76	4.95	2.58	.03
Progressive vs. Passive	28.17	4.08	6.91	< .001

Perfective vs. Progressive	15.41	4.24	3.64	.001
Subject 3				
Passive vs. Perfective	2.90	4.90	0.47	.89
Progressive vs. Passive	28.20	4.05	6.96	< .001
Perfective vs. Progressive	25.91	4.21	6.16	< .001
Subject 4				
Passive vs. Perfective	11.92	4.98	2.40	.05
Progressive vs. Passive	32.67	4.11	7.95	< .001
Perfective vs. Progressive	20.75	4.23	4.90	< .001
Subject 5				
Passive vs. Perfective	3.96	4.90	0.81	.70
Progressive vs. Passive	15.61	4.05	3.86	< .001
Perfective vs. Progressive	11.64	4.21	2.77	.02
Subject 6				
Passive vs. Perfective	-1.69	4.90	-0.35	.94
Progressive vs. Passive	6.92	4.05	1.71	.21
Perfective vs. Progressive	8.61	4.21	2.05	.11
Subject 7				
Passive vs. Perfective	-0.80	4.90	-0.16	.99
Progressive vs. Passive	8.18	4.05	2.02	.12
Perfective vs. Progressive	8.97	4.21	2.13	.09
Subject 8				
Passive vs. Perfective	6.53	4.90	1.33	.38

Progressive vs. Passive	26.84	4.05	6.63	< .001
Perfective vs. Progressive	20.31	4.21	4.83	< .001

Note: Three divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Vowel duration: Phonetic vowel length. As expected, the phonetically long vowel [æ] was 53.23 ms longer (M = 124.35 ms, SE = 1.16 ms) than the phonetically short vowels [ʌ] (M = 80.08 ms, SE = 1.19 ms) and [ɪ] (M = 60.55 ms, SE = 1.12 ms; see Figure 3.3.7; z = 7.37, p < .001). A model that included an interaction between phonetic vowel length and syntax predicted the data slightly better than one without an interaction between these factors ($ELPD_{diff}$ = -1.7, SE = 1.7). For both long and short vowels, passive verb stem vowels were longer than progressive vowels, and perfective verb stem vowels were longer than progressive verb stem vowels (see Table 3.3.6). The difference between passive and perfective vowels was significant for long vowels (z = 2.50, p = .03), and this difference appeared to drive the interaction. This suggests that different vowels undergo different degrees of phrase-final lengthening, and that this is true even across vowels that are phonologically short.

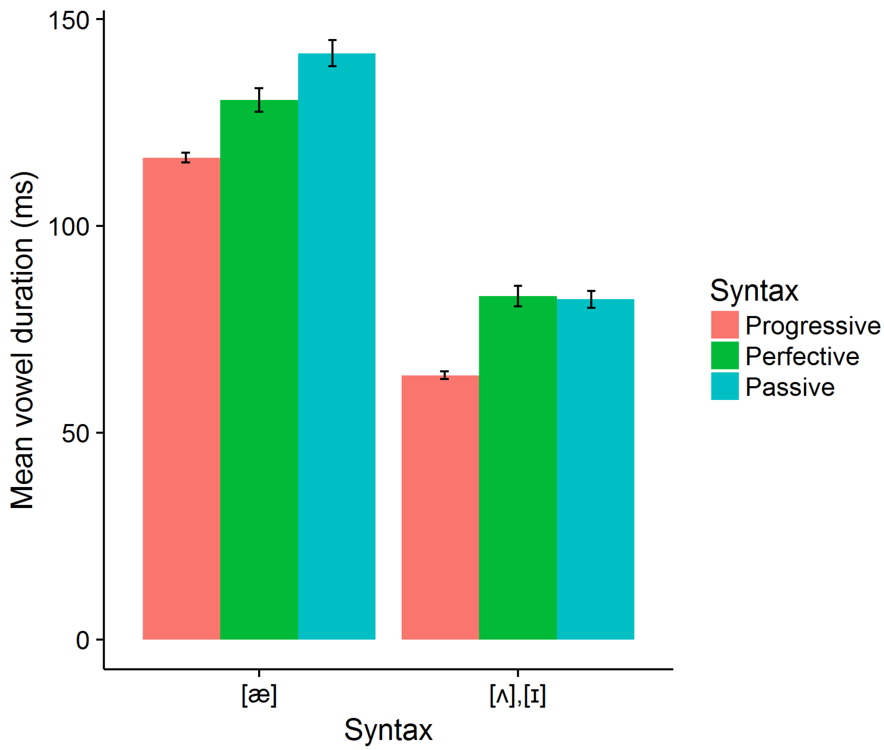


Figure 3.3.7. Experiment 3: Mean verb stem vowel duration by phonetic vowel type and syntactic construction. Error bars indicate standard error of the mean.

Table 3.3.6.

Experiment 3: Bayesian Linear Mixed-Effects Model Summary for Verb Stem Vowel Duration by Syntax and Vowel Length Category

					Median	MAD_{SD}
Mean Posterior Predictive Distribution					97.8	0.7
Random Effects	*SD*		*r*			
Subject (n = 8)						
Intercept	19.3	—	—	—		
Perfective	7.4	-0.13	—	—		
Active						

Progressive Active	7.9	-0.44	0.37	—
Short Vowel	7.3	-0.64	0.37	0.31
Verb (n = 18)				
Intercept	15.4	—	—	—
Perfective Active	9.5	0.06	—	—
Progressive Active	9.3	-0.60	0.25	—
Residual	17.6	—	—	—

Fixed Effects	*Median β*	*β MAD$_{SD}$*
Intercept	142.3	8.2
Error *SD*	17.6	0.3

					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Long vs. Short	53.39	7.24	7.37	< .01	-72.35	-46.99
Long [æ]						
Passive vs. Perf.	11.56	4.62	2.50	.03	—	—
Prog. vs. Passive	25.53	4.43	5.76	< .01	—	—

Perf. vs. Prog.	13.96	5.24	2.67	.02	—	—
Short [ʌ],[ɪ]						
Passive vs. Perf.	-0.83	4.60	-0.18	.98	3.93	21.15
Prog. vs. Passive	18.20	4.39	4.15	< .01	-0.93	15.03
Perf. vs. Prog.	19.02	5.18	3.67	< .01	—	—

Note: Twelve divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Vowel duration: Coda voicing. Because the voicing of the post-vocalic consonant can affect vowel duration, we examined whether verb stem coda voicing had an impact on vowel duration in the current study. On average, vowels with voiced coda consonants (M = 105.61 ms, SE = 1.56 ms) were 15.92 ms longer than vowels with voiceless codas (M = 89.99 ms, SE = 1.23 ms). Across syntactic constructions, vowels with voiced codas were 20.83 ms longer than vowels with voiceless codas in passive verb stems, 19.70 ms longer in perfective active verb stems, and 12.40 ms longer in progressive active verb stems (see Figure 3.3.8). A Bayesian linear mixed-effects model including an interaction between syntax and verb stem coda voicing was used to determine whether these differences were reliable. Contrary to Klatt (1976), the overall difference in vowel duration between vowels that were followed by voiced or voiceless codas was not reliable (z = -1.18, p = .24). This was also true for comparisons within syntactic constructions (all *ps* > .10), indicating no interaction between syntax and coda voicing (see Table 3.3.7). However, only 47.4% of all samples fell within 1 JND of zero, which suggests that the effect may be present, but is so small as to not have practical significance.

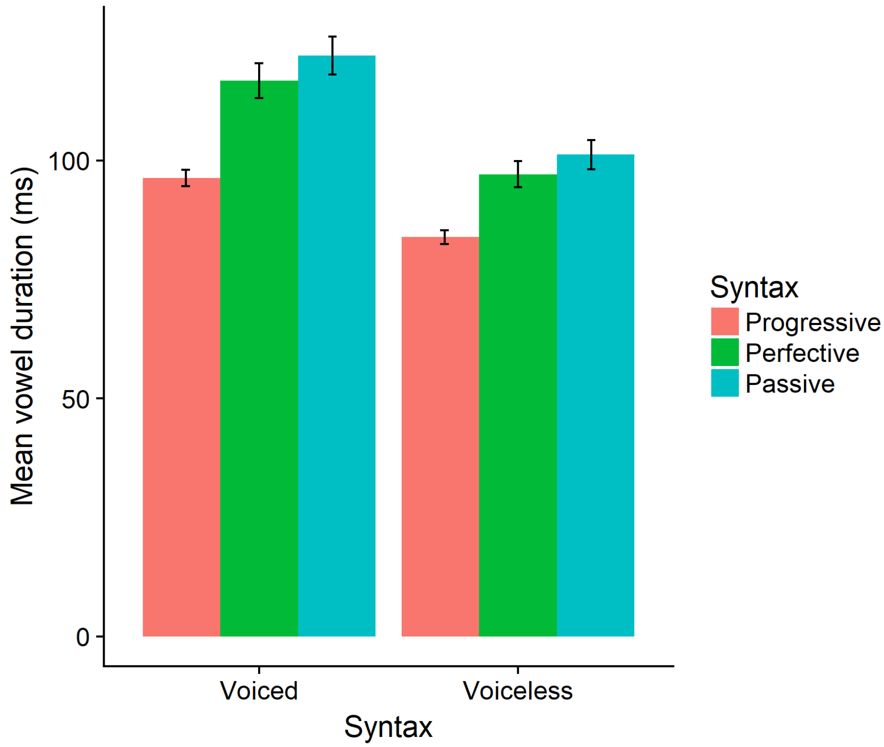


Figure 3.3.8. Experiment 3: Mean verb stem vowel duration for vowels with voiced and voiceless post-vocalic consonants, by syntax. Error bars indicate standard error of the mean.

Table 3.3.7.

Experiment 3: Bayesian Linear Mixed-Effects Model Summary for Verb Stem Vowel Duration by Syntax and Coda Voicing

					Median	*MAD*$_{SD}$
Mean Posterior Predictive Distribution					97.8	0.6
Random Effects	*SD*		*r*			
Subject (n = 8)						
Intercept	17.3	–	–	–		

Voiced Consonant		9.4	0.13	—	—	
Perf. Active		8.6	-	-	—	
			0.08	0.34		
Prog. Active		8.6	-	-	0.41	
			0.42	0.60		
Verb (n = 18)						
Intercept		29.8	—	—	—	
Voiced Consonant		16.8	0.09	—	—	
Perf. Active		11.8	-	-	—	
			0.30	0.34		
Prog. Active		10.1	-	-	0.34	
			0.33	0.29		
Residual		17.2	—	—	—	
Fixed Effects	*Median*	*β*				
	β	*MAD*$_{SD}$				
Intercept	102.1	11.5				
Error *SD*	17.2	0.3				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Voiceless vs. Voiced	-17.41	14.74	-	.24	-5.66	45.73
			1.18			
Passive: Voiceless vs.	-20.19	15.79	-	.20	—	—
Voiced			1.28			

Perf.: Voiceless vs. Voiced	-19.72	14.99	-1.32	.19	-10.45	9.78
Prog.: Voiceless vs. Voiced	-12.33	14.65	-0.84	.40	-16.95	1.03

Note: Six divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Next, we investigate whether the duration of the verb stem coda itself changed across syntactic constructions.

Coda duration. Verb stem codas were comparable in duration for passive verbs (M = 80.28 ms, SE = 1.98), perfective active verbs (M = 77.74 ms, SE = 1.99), and progressive active verbs (M = 80.07 ms, SE = 1.12 ms; see Figure 3.3.3.). Nearly all (99%) of samples fell within 1 JND of zero, indicating that differences in coda duration across syntactic constructions were so small as to be of no practical significance, (see summary in Table 3.3.8).

Table 3.3.8.

Experiment 3: Linear Mixed-Effects Model Summary for Verb Stem Coda Duration by Syntax

				Median	MAD_{SD}
Mean Posterior Predictive Distribution				79.6	0.6
Random Effects	*SD*	*r*			
Subject (n = 8)					
Intercept	12.9	—	—		
Perfective Active	9.5	-0.23	—		

Progressive Active		9.6	-0.72	0.00		
Verb (n = 18)						
Intercept		26.1	–	–		
Perfective Active		5.5	-0.11	–		
Progressive Active		19.7	-0.01	-0.02		
Residual		15.8	–	–		
Fixed Effects	*Median β*	*β MAD$_{SD}$*				
Intercept	80.5	7.7				
Error *SD*	15.8	0.3				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Passive vs. Perfective	2.80	3.76	0.75	.74	-8.77	3.50
Progressive vs. Passive	0.48	5.85	0.08	.99	-10.27	9.21
Perfective vs. Progressive	-2.32	6.88	-0.34	.94	–	–

Note: Six divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Deletion of the -ed inflection. Deletion of the *-ed* inflection was less frequent in this experiment than in Experiment 2: perfective active *-ed* inflections were deleted 10.84% of the time, and passive *-ed* inflections were deleted 9.47% of the time. Note that these deletion rates compare to that of passives in Experiment 2. To determine whether deletion of the past-tense verbal inflection was affected by syntax, a Bayesian logistic mixed-effects model analysis was performed (see 3.2.2 for a more detailed explanation). Unlike Experiment 2, speakers did not delete *-ed* inflections reliably in perfective active sentences as compared to passive sentences, and the effect estimate is nearly zero (β =

0.02, $z = 0.03$, $p = .98$). This is consistent with speakers in Experiment 2 deleting -*ed* inflections due to the phonological environment: the adjacent consonantal onset [ð] in *the* was similar in place of articulation to that of an alveolar stop ([t] or [d]).

In sum, the findings for verb stem duration were consistent with the findings of Experiments 1 and 2: passive verb stems were longer than progressive active verb stems, and all speakers produced this difference. Consistent with Experiment 2, there was greater variability in perfective active verb stem duration across speakers. Verb stem onsets were also longer in passive sentences than progressive active sentences, but not for all speakers. Vowels were consistently lengthened by all speakers in passive sentences as compared to progressive active sentences. For both onset and vowel segments, perfective active segment duration was less consistent: the duration was usually shorter than passive verb stem segments and longer than progressive active verb stem segments, but this was not consistent across speakers. The verb stem coda consonant did not vary in duration for different syntactic constructions.

3.3.3. Discussion

Experiment 3 revealed that two segments of the verb stem undergo lengthening in passive sentences: the onset and the vowel. Vowels were consistently longer in passive sentences as compared to progressive active sentences across speakers, but the difference was more variable for onset duration. These findings suggest that the vowel primarily undergoes lengthening in passive sentences and contributes to passive verb stem lengthening, while the onset contributes to a lesser extent. The consistency of passive verb stem vowel lengthening leaves open the possibility that listeners can use vowel duration as a cue to syntax.

The trends in verb stem duration and verb stem vowel duration across syntactic environments support the results of Experiment 2, and generate further evidence that the progressive active-passive verb stem duration difference occurs as a consequence of both phrase-final lengthening and polysyllabic shortening.

Unlike in Experiment 2, speakers in the current experiment did not delete *-ed* inflections at different rates between perfective active and passive sentences. As explained in section 3.2.3, the difference likely occurred in Experiment 2 because the place of articulation was similar between the verbal inflection and the onset of the subsequent determiner. The results of the current study suggest that the similar place of articulation indeed led speakers to reduce complex codas more frequently in perfective active sentences. The subsequent phonological context was controlled in this experiment across passive and perfective active sentences, and we observe deletion of the *-ed* inflection equally often in these constructions as a result. This also indicates that the differences in deletion rates in Experiment 2 were not the result the speaker modulating the information content of the acoustic signal.

Overall, we can conclude that the verb stem vowel is the most likely candidate acoustic cue that listeners may use to predict syntax. While there may be other acoustic cues present (the onset of the verb stem, the duration of the auxiliary, the intensity of the verb stem), none of them are as consistent or as robust as the progressive active-passive vowel duration difference across the 3 studies presented in this chapter and Stromswold et al. (2002).

3.4. General Discussion

Following Stromswold et al. (2002), in this chapter we investigated whether progressive active and passive sentences produced by native English speakers differ acoustically during the morphosyntactically ambiguous region of the sentence.

Experiments 1-3 show that duration differences are the most reliable differences across progressive active and passive sentences. The most consistent duration cue was the difference between progressive active and passive verb stems, where passive verb stems are consistently lengthened relative to progressive actives. Auxiliaries also lengthened in passive constructions, but this difference was less reliable in Experiments 1 and 2 than verb stem duration. Experiment 1 found that speakers consistently lengthen monosyllabic passive verb stems relative to progressive active verb stems, and Experiments 2 and 3 replicated this finding. Experiment 1 also found that verb stem intensity was higher for actives than for passives, but the difference was less robust than it was for the duration cues.

As discussed in section 3.1.3, differences in prosodic structure across these constructions give rise to passive verb stem lengthening. The results reported in Experiments 2 and 3 indicate that both phrase-final lengthening and polysyllabic shortening contribute to the verb stem duration difference. Experiment 2 provided evidence that polysyllabic shortening contributes more to the verb stem duration difference than phrase-final lengthening does. While Experiment 3 produced similar findings, there was more involvement of phrase-final lengthening, likely due to the careful control of the phonological environment in Experiment 3. Finally, Experiment 3 confirmed that the vowel is the verb stem segment that is primarily responsible for the progressive active-passive verb stem duration difference. We conclude from this that both processes do contribute to the progressive active-passive verb stem duration difference, though the contribution by polysyllabic shortening outweighs that of phrase-final lengthening, and that duration changes occur on the verb stem vowel.

The experiments reported in this chapter show that passive verb stem lengthening is a robust phenomenon, and is realized primarily on the verb stem vowel. These findings support the possibility that listeners use verb stem vowel duration to facilitate processing of active and passive sentences. However, the experiments reported in this chapter and Stromswold et al. (2002) identified several other acoustic cues to syntax in the morphosyntactically ambiguous region that were less consistent: passive auxiliary lengthening, passive verb stem onset lengthening, and higher verb stem intensity in active sentences. It is possible that listeners could use these acoustic correlates to parse the sentence as it unfolds. The comprehension study reported in Chapter 4 was designed to assess whether listeners use verb stem vowel duration to facilitate the auditory processing of progressive active and passive sentences.

4. Sentence Comprehension by Native English Speakers

4.0. Motivation

In Chapter 2, we reviewed literature demonstrating that listeners can recruit acoustic cues in order to form predictions about yet-to-be-heard speech in an utterance. We also reviewed theories of sentence processing that either do or do not allow extrasyntactic information, such as information in the acoustic signal, to bear on processing decisions. The experiments presented in the current chapter speak to the both of these topics by testing whether verb stem vowel duration can be used to predictively process active and passive sentences.

Recall that listeners tend to prioritize early information in processing sentences even when it is noisy, and may be misleading (e.g., Trueswell et al., 1999; Choi & Trueswell, 2010). The production experiments presented in Chapter 3 show that the progressive active-passive verb stem duration difference is robust for monosyllabic verbs, and that the duration difference is driven by phonological mechanisms that affect vowel duration. Because the duration difference is the most robust cue, it may serve as a relatively reliable indicator of progressive active or passive syntax: in these sentence structures, short verb stem vowels cue progressive active syntax, while long verb stem vowels cue passive syntax. Given that the verb stem vowel precedes a definite morphosyntactic cue (inflectional morphology on the verb), and that across studies and participants in Chapter 3, no other acoustic differences *consistently* cue syntax, it is possible that listeners used vowel duration to predict syntax in both the eye-tracking study and gating study reported by Stromswold et al. (2002). In other words, it is possible that listeners form hypotheses about the syntactic structure of an utterance conditioned on the duration of a syllable—or perhaps even of a vowel—when its duration is affected by the next morpheme and the location of the syllable in a prosodic phrase. This would be consistent with evidence that listeners can use syllable duration to infer the prosodic context of a syllable and predict subsequent syllable(s) (Salverda et al., 2003; Salverda et al., 2007).

Two experiments were carried out to test the claim that listeners can use vowel duration to facilitate processing of progressive active and passive sentences. In Experiment 4a, recordings from Experiment 1 were manipulated so that vowels were lengthened in progressive active sentences and were shortened in passive sentences. A control study was then carried out to ensure that the recordings containing manipulated vowels sounded natural. In Experiment 4b, the manipulated recordings from Experiment 4a were used as stimuli in an eye-tracking study. If listeners are sensitive to vowel duration as a cue to syntax, the manipulated vowels should interfere with online processing.

4.1. Experiment 4a: Norming study

Because verb stem duration is the only acoustic correlate of syntax that was consistent across speakers (Chapter 3), and because the verb stem vowel is the locus of the duration change, listeners may use verb stem vowel duration to disambiguate progressive active or passive syntax. To investigate the suspected role of verb stem vowel duration in the online processing of active and passive sentences, the duration of the verb stem vowel was altered so that it differed from what listeners expect, and was then used in a comprehension study (4b). In this experiment, the verb stem vowel was lengthened in active sentences to match verb stem vowel duration in equivalent passive sentences, and was similarly shortened in passive sentences.

To verify that the vowel duration manipulation sounded natural, a norming study was conducted. This involved asking adult native English speakers to judge the naturalness of the audio for all sentences tested in Experiment 4b. If the manipulation did not sound natural, then the audio containing manipulated verb stem vowels should be rated as less natural (lower Likert rating) than the unmanipulated audio.

4.1.1. Methods

Participants. 9 monolingual adult native American English speakers participated in exchange for course credit. The participants did not participate in any of the other experiments (1-3, or 4b). The study was approved by the Rutgers University Institutional Review Board, and was carried out in accordance with the Declaration of Helsinki.

Materials and Design. *Acoustics*. The sentences used in the current study were a subset of those recorded from a single speaker in Experiment 1.[8] For the target sentences recorded in Experiment 1, the speaker (participant 4) lengthened passive auxiliaries by 20.56 ms ($t(48.01) = 2.65$, $p = .01$), and lengthened passive verb stems by 74.20 ms ($t(46.81) = 6.07$, $p < .001$; Figure 4.1.1).

[8] Note that this information corresponds to 56 of the 64 target sentences produced by a single speaker (Experiment 1, Participant 4). Eight sentences containing the verbs *tickle* and *pat* were omitted. Because the current summary reflects only a subset of the original productions, the values here differ from those reported in Experiment 1 (Ch. 3) for this speaker.

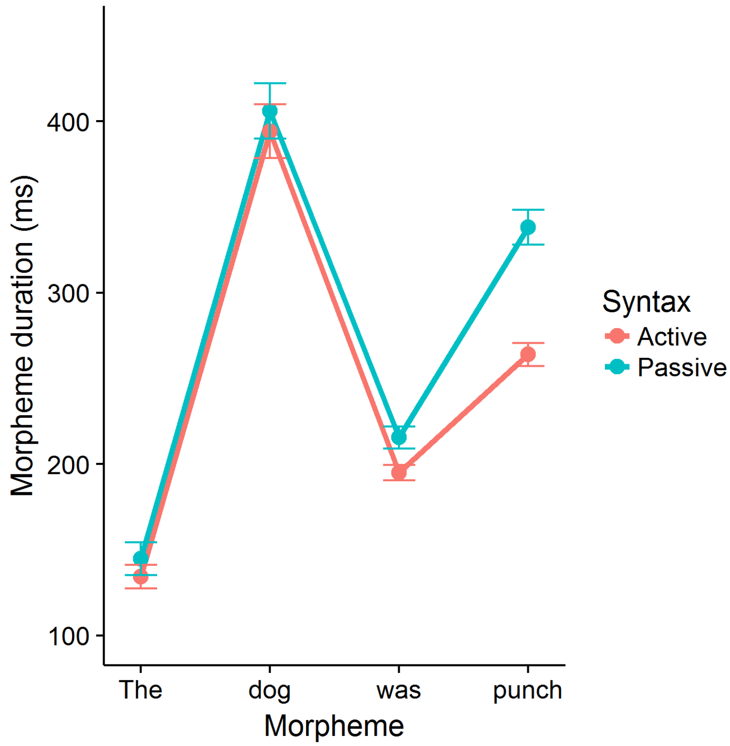


Figure 4.1.1. Experiment 4a: Average morpheme duration for each morpheme during the morphosyntactically ambiguous region (first determiner through the verb stem) of active and passive sentences, for the unmanipulated sentences produced by a speaker in Experiment 1. Error bars indicate standard error of the mean.

The speaker produced auxiliaries with marginally higher intensity (2.40 dB) in active sentences ($t(48.72) = 1.71$, $p = .09$), and produced verb stems that were 3.46 dB higher on average in active sentences ($t(53.17) = 3.88$, $p < .001$; Figure 4.1.2).

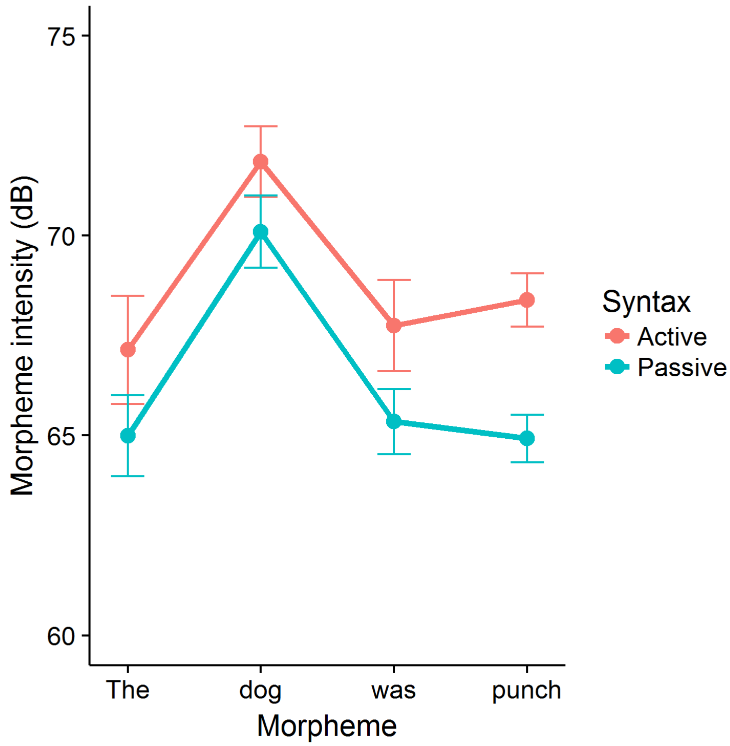


Figure 4.1.2. Experiment 4a: Average morpheme intensity for each morpheme during the morphosyntactically ambiguous region (first determiner through the verb stem) of active and passive sentences, for the unmanipulated sentences produced by a speaker in Experiment 1. Error bars indicate standard error of the mean.

Vowel duration manipulation. Verb stem duration was manipulated by altering the duration of the verb stem vowel. The verb stem vowel was chosen because it was the segment that consistently demonstrated duration changes in Experiment 3, and because it can be manipulated without causing distortion. Vowel duration was “swapped” across active and passive sentences. For each unique sentence, the duration of the verb stem vowel was manipulated to match its duration in the other construction type: for example, the verb stem vowel in the sentence *The dog was punching the bear* was lengthened to match the duration of the vowel [ʌ] in *The dog was punched by the bear*.

To manipulate the duration of the verb stem vowel, a Praat manipulation function that operates on duration was used. The algorithm either increased or decreased vowel duration by identifying a full vowel period near the midpoint of the vowel, then either copied the vowel period to lengthen the vowel, or removed it to shorten the vowel (see Figure 4.1.3.).

The duration altering process was carried out using a custom Python script that interfaced with Praat (see Appendix D.1). The Python script calculated the percent change in vowel duration required for the original vowel (e.g., in a passive sentence) to match the duration of the vowel in the opposite syntactic construction (e.g., the complementary active sentence), then passed the calculated onset, midpoint, and offset to Praat. Because the Praat algorithm operates on the level of vowel periods, the final vowel duration must be a multiple of the period duration. As a result, it was possible for the final vowel duration to differ from the target vowel duration. To avoid this, the Python script removed or added the duration of a vowel period to the calculated time points: if the resulting vowel duration would fall short of the target lengthening for actives, a vowel period was added, and vice versa for the target shortening in passives. See Figure 4.1.3 for a comparison between an unmanipulated vowel and a manipulated vowel. For filler sentences, because there was no passive sentence to swap duration with, verb stem vowels were instead lengthened by 20% of the vowel's duration.

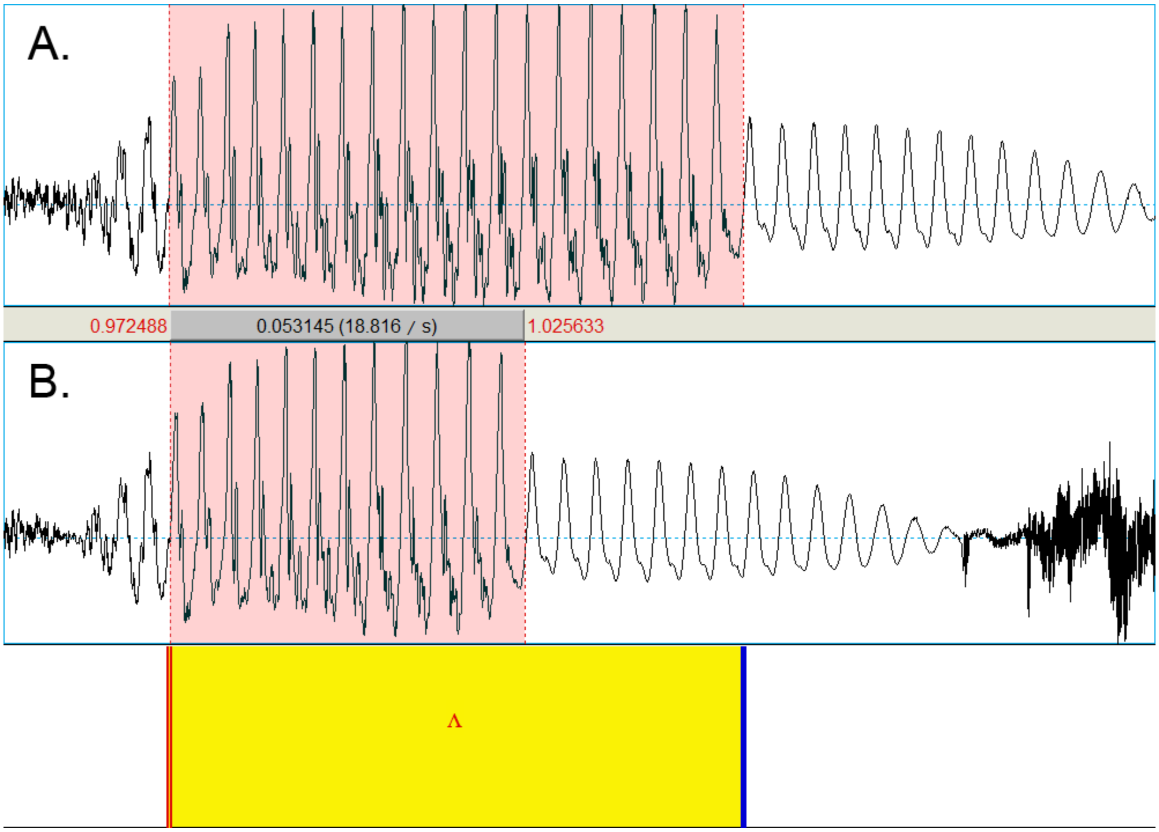


Figure 4.1.3. Experiment 4a: Acoustic wave form showing the duration of the verb stem vowel [ʌ] in the sentence *The dog was punched by the bear* in an unmanipulated sentence (panel A, [ʌ] duration = 85.91 ms), and in a sentence where the verb stem vowel has been shortened (panel B, [ʌ] duration = 53.15 ms).

Target sentences. A subset of the sentences in Experiment 1 served as the stimuli (see Appendix A). Sentences from Experiment 1 containing the verbs *tickle* and *pat* were excluded, resulting in 56 unique target sentences. In the manipulated audio condition, verb stem vowel durations were swapped across active and passive sentences for all 56 target sentences.

Filler sentences. Twenty-eight unique active filler sentences were used. Because 8 filler sentences were each repeated once, there were 36 total filler sentences in each audio manipulation condition. Twenty of the filler sentences contained main verbs. For these sentences, verb stem vowels were lengthened by 20% in the manipulated audio condition.

The remaining 16 filler sentences were identical to those in the unmanipulated control condition, and were included for completeness.

The stimuli consisted of 92 unmanipulated control sentences and 92 sentences in the manipulated audio condition (56 manipulated target sentences + 20 manipulated fillers + 16 neutral fillers), resulting in a total of 184 trials. Each subject heard all 184 sentences during a single experimental session. Sentences were presented in randomized order.

Apparatus. The experiment was carried out using a Lenovo ThinkPad Yoga 12 20DL running Windows 8.1, with a resolution of 1920 x 1080. Viewing distance was whatever felt comfortable to the participant, approximately 20 inches from the screen. Audio was presented using a pair of Sennheiser HD 202 headphones. Stimulus presentation and response logging were performed using E-Prime Professional 2.0 software.

Procedure. Prior to the experimental trials, participants were given the following instruction: “In this task, you will hear audio recordings of spoken sentences. For each, listen to the recording and indicate how natural the speech sounds to you”. After receiving this instruction, participants pressed the spacebar to proceed to the experimental trials. A trial proceeded as follows. First, a blank screen was displayed for 250 ms. After the blank screen, a recording played, followed immediately by a prompt to rate the naturalness of the speech in the recording using the number keys. A Likert scale from 1-5 was used to solicit judgments, where a response of 1 indicated that the audio sounded “not at all natural” and 5 indicated that the audio sounded “very natural”. Ratings were recorded via keypress, at which point response time was also recorded. After responding, the next trial began. This procedure was repeated for a total of 184 trials.

Analysis. Responses for the 112 target sentences (56 with manipulated vowels, 56 with unmanipulated vowels) were analyzed. Trials for which response times fell below 250 ms or over 5000 ms were excluded (31.85% of trials).

Likert ratings. To determine whether listeners judged the manipulated verb stem vowels to be unnatural, Likert ratings were analyzed by Bayesian linear mixed-effects models. Syntax and audio manipulation were included as fixed effects. Item (verb) and subject were included as random effects, with random intercepts and random slopes (for syntax and audio manipulation) included in the random effect structure. The lsmeans package in R was used to obtain *p*-values, and to determine whether levels of the fixed effects differed from one another. Comparisons that were not significant according to statistical tests were further evaluated based on the size of the β coefficient, and/or based on the proportion of samples that fell below or above zero. Degrees of freedom were estimated using the Satterthwaite method (Satterthwaite, 1946), and *p*-values were adjusted for multiple comparisons using the Tukey method.

Reaction time. Reaction times were adjusted to the onset of the first determiner in each sentence, then analyzed using Bayesian linear mixed effects model comparison as specified above (see *Likert ratings*).

4.1.2. Results

4.1.2.1. Likert ratings

Overall, participants indicated that the recordings sounded relatively natural (see Figure 4.1.4). On average, manipulated audio (M = 3.19, SE = 0.05) was rated to sound less natural by 0.11 on the Likert scale than unmanipulated audio (*M* = 3.30, *SE* = 0.05). Listeners rated the naturalness of the audio for active sentences (*M* = 2.95, *SE* = 0.065) to be lower by 0.16 on the Likert scale than for passives sentences (*M* = 3.11, *SE* = 0.064).

Because the difference in Likert ratings appeared to be different for active sentences on the dimension on audio manipulation, we tested a model containing an interaction between syntax and audio manipulation. Overall, the effect of the audio manipulation on naturalness ratings was so small as to be of no practical significance, and the effect estimate may have been zero (*Median* β = -0.3; *p* = .17). The proportion of samples that fell within 0.5 of zero on the Likert scale was 95.25%, which further suggests that the effect of the audio manipulation was very small. Across active and passive sentences, however, the difference in Likert ratings was marginally significant (*z* = -1.74, *p* = 0.08; see Table 4.1.1). Manipulated active sentences (*M* = 2.82, *SE* = 0.09) were rated to be less natural than manipulated passive sentences (*M* = 3.10, *SE* = 0.09) by 0.28 points on the Likert scale. A model with an interaction between syntax and audio manipulation was better able to predict the data than the model without, but the difference was slight ($ELPD_{diff}$ = -0.40, *SE* = 1.3).

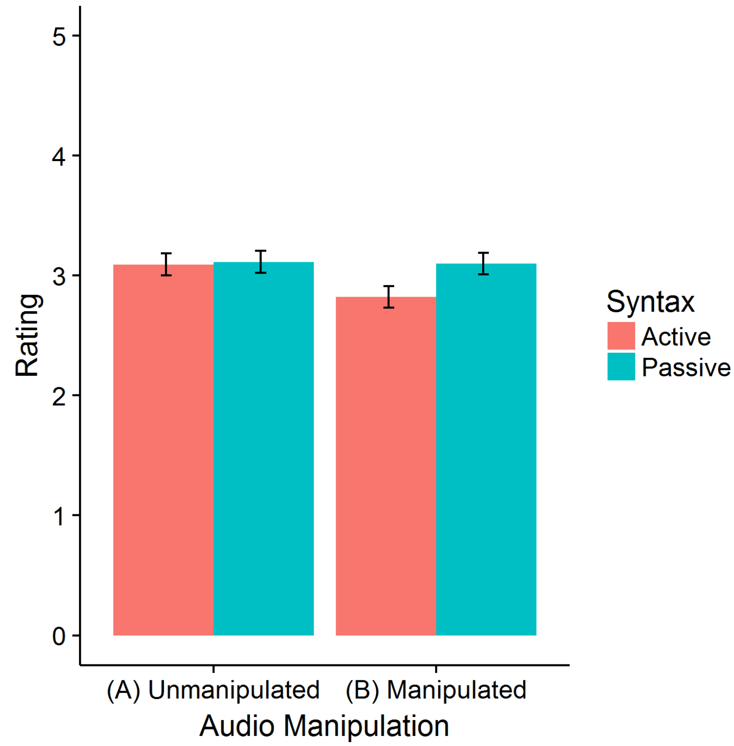


Figure 4.1.4. Experiment 4a: Mean Likert ratings for active and passive sentences when audio was either unmanipulated or manipulated. Error bars indicate standard error of the mean.

Table 4.1.1.

Experiment 4a: Bayesian Linear Mixed-Effects Model Summary for Audio Manipulation, Syntax and Naturalness Rating

				Median	*MAD$_{SD}$*
Mean Posterior Predictive Distribution				3.0	0.1
Random Effects	*SD*	*r*			
Subject (*n* = 9)					
Intercept	0.57	–	–		
Passive Syntax	0.19	0.29	–		
Manipulated	0.25	-0.08	-0.18		
Verb (*n* = 14)					

Intercept	0.20	–	–			
Passive Syntax	0.10	-0.21	–			
Manipulated	0.12	0.04	-0.13			
Residual		–	–			
Fixed Effects	*Median β*	*β MAD_{SD}*				
Intercept	3.1	0.2				
Error *SD*	1.0	0.0				
					Posterior Interval	
	β	*SE*	*z*	*p*	5%	95%
Unmanipulated vs. Manipulated	0.16	0.12	1.36	.17	-0.50	-0.03
Active vs. Passive	-0.18	0.10	-1.74	.08	-0.13	0.29
Unmanipulated:						
Active vs. Passive	-0.08	0.13	-0.59	.56	–	–
Manipulated:						
Active vs. Passive	-0.28	0.13	-2.18	.03	-0.06	0.47

Note: One divergent transition occurred during sampling, which indicates a possibility that the sampler was biased.

It is unclear why listeners rated active sentences to sound less natural than passive sentences, but did not consider the vowel duration manipulation to sound unnatural overall. One possibility is that differences in duration caused by vowel lengthening are more salient than those caused by vowel shortening, which may have cued listeners to the presence of the manipulation. However, it should be noted that these results are based on data from only 9 participants, and so the analysis should be interpreted with caution.

4.1.2.2. Reaction time

There effect of audio manipulation or syntax on reaction time was so small as to be of no significance (see Figure 4.1.5). Analysis using a Bayesian linear mixed-effects model with an interaction between syntax and audio manipulation suggested that any effect of either syntax (*Median* β = -26.1) or audio manipulation condition (*Median* β = -40.8) on response times was too small to be of consequence (Table 4.1.2). A model containing an interaction between audio manipulation and syntax did not predict the data as well as an otherwise equivalent model without this interaction, though the difference was slight ($ELPD_{diff}$ = 0.6, *SE* = 1.0).

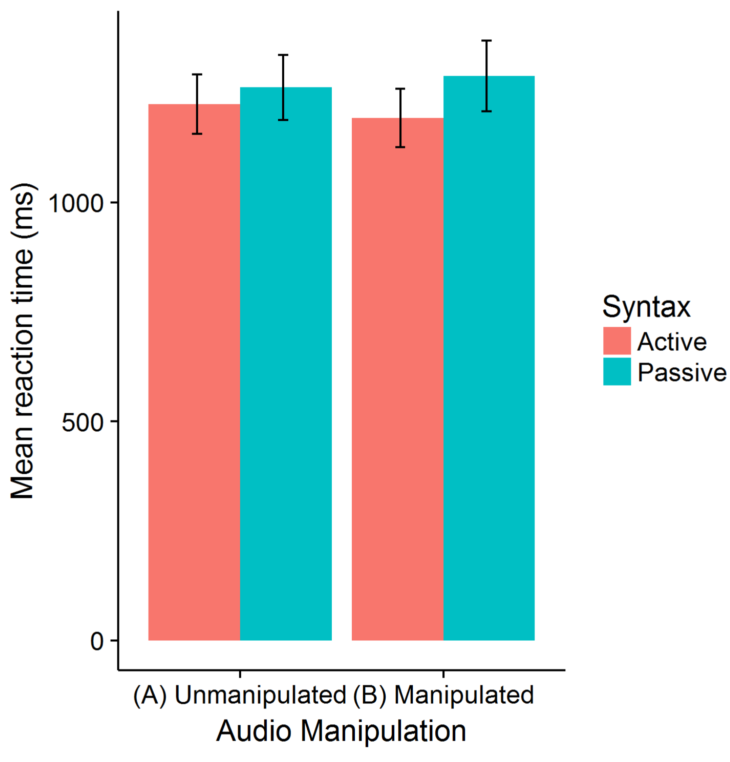


Figure 4.1.5. Experiment 4a: Mean reaction time for active and passive sentences when audio was either unmanipulated or manipulated. Error bars indicate standard error of the mean.

Table 4.1.2.

Experiment 4a: Bayesian Linear Mixed-Effects Model Summary for Naturalness Rating by Audio Manipulation and Syntax

						Median	MAD_{SD}
Mean Posterior Predictive Distribution						1243.8	46.1
Random Effects	*SD*	*r*					
Subject (*n* = 9)							
Intercept	399	–	–				
Passive Syntax	209	0.34	–				
Manipulated	185	0.03	0.09				
Verb (*n* = 14)							
Intercept	59	–	–				
Passive Syntax	77	-0.14	–				
Manipulated	72	-0.10	- 0.19				
Residual	850	–	–				
Fixed Effects	*Median* *β*	*β* MAD_{SD}					
Intercept	1184.4	145.5					
Error *SD*	849.9	24.9					
						Posterior Interval	
	β	*SE*	*z*	*p*		5%	95%

Unmanipulated vs. Manipulated	-25.77	97.60	-0.26	.79	-237.55	156.48
Active vs. Passive	-40.46	101.13	-0.4	.69	-222.51	168.32
Unmanipulated: Active vs. Passive	26.09	121.23	.22	.83	–	–
Manipulated: Active vs. Passive	-107.02	119.35	-0.90	.37	-82.29	344.10

Note: Five divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

4.1.3. Discussion

In this study, we switched the duration of active and passive verb stem vowels by either adding or removing vowel periods at the midpoint of the vowel. The results indicate that the manipulation sounded natural to listeners, which suggests that any differences across audio manipulation conditions in the comprehension study would not be due to the naturalness of the vowel duration manipulation. Furthermore, for reasons that are currently unclear, recordings of active sentences were rated as less natural sounding than passive sentences. A caveat that attaches to both of these analyses is that the low number subjects (n = 9) may have affected the outcome of the analyses.

We turn now to a comprehension study using the manipulated audio as stimuli for one group of participants, and the unmanipulated audio for another.

4.2. Experiment 4b: Manipulated vowel study

There is evidence that verb stem lengthening in passive sentences is robust (Chapter 3). In the current study, we investigated whether listeners use verb stem vowel duration to disambiguate active and passive sentences prior to hearing the verbal inflection. To assess this, listeners heard the recordings that were used in Experiment 4a and selected 1 of 2 images that matched the sentence. Each participant heard either recordings containing verb stem vowels that were manipulated, or unmanipulated recordings.

Response accuracy, response time, and eye-gaze were recorded and analyzed. If being a native English speaker entails knowing (unconsciously) that the verb stem should be longer when the verb is monosyllabic and/or phrase-final, and if listeners do not use earlier disambiguating cues, then we expected processing to be impaired when the cue differs from listeners' expectations. If listeners use vowel duration to facilitate processing, response time was expected to be higher for sentences containing manipulated vowels than for unmanipulated sentences. This would reflect a processing delay when the cue was altered. Accuracy was expected to be lower for sentences containing manipulated vowels, as we expected processing difficulty may also result in poor comprehension. Furthermore, listeners were expected to look to the correct picture later when verb stem vowel duration is different from expected: specifically, we expected listeners to look to the correct image after hearing the verbal inflection.

If verb stem vowel duration is not the primary acoustic cue that listeners use to facilitate processing, there should be no difference in response time, accuracy, or eye-gaze between sentences containing manipulated or unmanipulated vowel durations, save for individual differences between participants.

4.2.1. Methods

Participants. Forty-four monolingual adult native English speakers participated, and received either credit toward a course requirement or monetary compensation. Three participants failed to calibrate, and instead completed Experiment 4a (see section 4.1). One participant in the unmanipulated audio condition was excluded due to substantial track loss during 97.65% of target trials, and for response times in excess of 5000 ms on 24.11% of the target trials. Data collection for two participants in the manipulated audio condition was incomplete, and so these subjects were not analyzed. Data from the remaining 38 participants were analyzed. All participants had normal or corrected-to-normal vision. The study was approved by the Rutgers University Institutional Review Board, and was carried out in accordance with the Declaration of Helsinki.

Materials. *Sentences.* The same sentences tested in Experiment 4a were used, but each participant only heard either recordings from the unmanipulated audio condition or the manipulated audio condition (see Experiment 4a, section 4.1.1). Audio manipulation was carried out as a between-subjects design to prevent the presence or absence of the vowel duration manipulation from trial-to-trial from affecting processing[9], and to allow for the possibility that listeners may learn over the course of the experiment. Two coders marked boundaries by hand using Praat (Boersma, 2001) for the onset and offset of each morpheme, and of the verb stem vowel. Segmentation was carried out by two coders using the same criteria used to segment morphemes in Experiment 2, and to segment vowels in Experiment 3. The two coders showed high agreement in boundary placement ($r(440) = 0.93$, $p < .001$).

Images. Visual stimuli consisted of a series of 87 unique 1200 x 800 images depicting two events involving animals performing actions against a white background (see Figure 4.2.1). A black divider line positioned at the midpoint separated the event on

[9] This was done to maintain as much consistency in the acoustics of the recordings as possible, both within and across the audio manipulation condition.

the left side of the image from the event on the right. The images were stretched by E-Prime to occupy the entire width of the screen (1920 pixels).

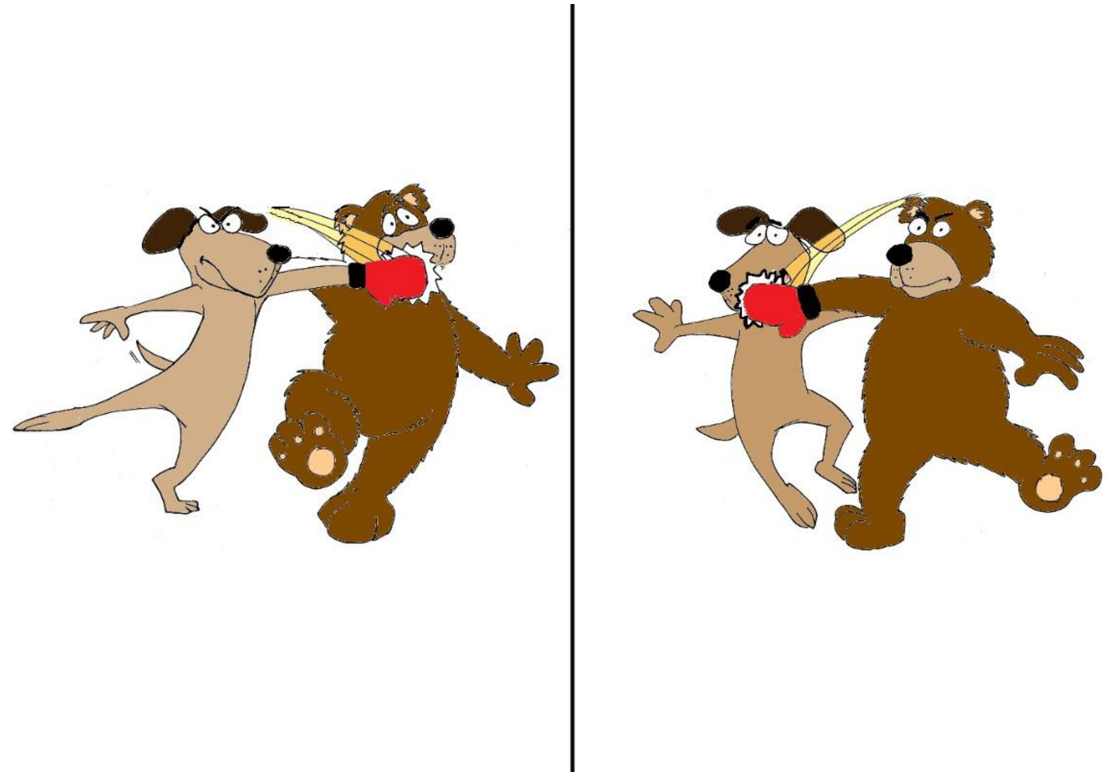


Figure 4.2.1. Experiment 4b: Example of visual stimulus in the eye-tracking study, where the left image matches the sentence *The dog was punching the bear*, and the right image matches *The dog was punched by the bear.*

Design. As in the production experiments (Chapter 3), two lists were generated by creating a single pseudorandomized list, then reversing that list to create a second list. Sentences containing the same main verb or the same noun in subject position did not occur consecutively. No more than 3 passive or progressive active target sentences occurred consecutively, and no more than 4 target sentences occurred between filler sentences. Target images were spatially balanced so that the left and right sides of the screen showed the target image equally often. Images were also balanced so that the animal that was the agent of the action appeared equally often on both the left and right side of each image. Participants were randomly assigned to the manipulated (n = 20) or unmanipulated (n = 18) audio condition.

Apparatus. The experiment was carried out using a Lenovo ThinkPad Yoga 12 20DL running Windows 8.1, with a resolution of 1920 x 1080. Viewing distance was whatever felt comfortable to the participant, which was generally about 20 inches from the screen. A keyboard mask was used to occlude all keys except *A*, *L*, and the spacebar. Audio was presented using a pair of Sennheiser HD 202 headphones.

Experiment software. Stimulus presentation and logging of response data were performed using E-Prime Professional 2.0 software. A time stamp with millisecond precision was included to mark the start of each trial.

Eye-tracker. Eye-gaze data were collected using an Eye Tribe ET1000 eye tracker, with a sampling rate of 30 Hz and accuracy of 0.5° - 1° visual angle. Custom logging software written in Java was used to record the x- and y- coordinates for each eye, and a time stamp with millisecond precision. Gaze coordinates recorded for the right eye were used in subsequent analyses, except when right eye experienced track loss but the left eye did not. In the latter case, gaze coordinates for the left eye were used instead.

Procedure. Participants completed 4 practice trials prior to beginning the experimental trials. The procedure for the practice trials was identical to that of the experimental trials, using sentences that were not included in the experimental trials. The participant was instructed to indicate which picture was best described by the sentence. This instruction was followed by the experimental trials.

A trial proceeded as follows. First, a blank screen appeared for 100 ms, followed by a shrinking fixation cross. This shrinking fixation was presented for 1000 ms, during which time it changed in font size from 900 pt to 60 pt, decreasing in size by 120 pt increments every 125 ms. Following the shrinking fixation, the stimulus image was presented for 3 seconds, after which the image persisted while a recording of a sentence played. Listeners indicated which image matched the sentence they heard via keypress (*A* for the left side, *L* for right side), at which point response time and accuracy were recorded. If the response was made prior to the end of the sentence, the image and audio persisted until the end of the recording. Following either the response or the end of the recording (if the response was made while the recording was playing), the next trial began. This procedure repeated for 92 trials.

After completing the first experimental block, the New York Times dialect survey (Katz, Andrews, & Buth, 2013) was completed as a distractor task. The distractor task took approximately 5-10 minutes. After the distractor task, the comprehension experiment was repeated using the reverse of the order in the first block. The experiment lasted approximately 1 hour.

Analysis. Analyses were carried out using Bayesian mixed-effects models using the rstanarm package in R (see section 3.1.1 for a detailed description). The lsmeans package in R was used to obtain *p*-values, and to determine whether levels of the fixed effects differed from one another. Degrees of freedom were estimated using the Satterthwaite method (Satterthwaite, 1946), and *p*-values were adjusted for multiple comparisons using the Tukey method. Comparisons that were not significant according to statistical tests were further evaluated by determining the proportion of β samples that fell above or below zero, and/or based on the size of the overall β coefficient (mean or median). This was done to differentiate between the presence of an effect that was too small to result in statistical significance, and an effect size that was too small to be of any real consequence (effectively zero).

Accuracy. To determine whether the altered verb stem vowels affected comprehension, response accuracy was analyzed. Accuracy data were analyzed using a Bayesian logistic mixed effects model. To control for variation due to individual differences or due to the items tested, random slopes and random intercepts were included for the syntactic frame (active or passive) and the audio condition (manipulated or unmanipulated), by subjects and by item (verb).

Reaction time. To determine whether unexpected verb stem vowel durations impaired processing, reaction time was analyzed. Reaction time was corrected to the onset of the first determiner, and was then analyzed with Bayesian linear mixed-effects models using the rstanarm package in R. Response times for incorrect responses were excluded from analysis.[10] To control for variation due to individual differences or due to the items tested, random slopes and random intercepts were included for the syntactic frame (active or passive) and the audio condition (manipulated or unmanipulated), both by subjects and by item (verb).

Eye-gaze data: Processing. The eye-gaze data were parsed into individual trials by identifying the timestamped data collected after the onset of a trial and before the start of the subsequent trial. Entire trials were discarded if no valid data could be obtained (e.g., total track loss). Within each trial, gaze x-coordinates were discarded if their values were either negative, exceeded the pixel width of the screen (1920), equaled the midpoint of the screen (960), or were zeroes (track loss). The eye-gaze data for each trial were then

[10] When response times for incorrect trials were included, the results of the reaction time analyses were not affected.

time-locked 50 ms prior to the onset of the first determiner, to ensure no frames were skipped. For each frame, the constituent heard by the subject and the accuracy of the subject's gaze were recorded. Frames that were recorded after the end of the second noun were discarded.

Eye-gaze data: Analysis. In order to analyze changes in gaze behavior, we analyzed saccades between the target and distractor image within a region of interest. In the first analysis, we examined saccades in response to the verb stem during a region defined by the onset of the verb stem and the onset of the inflection. In the second analysis, we assessed whether listeners used the NP1 = agent strategy by examining saccades in response to the first noun, during a region defined by the onset of the first noun and the onset of the auxiliary. See 4.2.2.3. for a detailed description of these analyses. Except where indicated otherwise, trials where listeners responded incorrectly were excluded from these analyses.

4.2.2. Results

4.2.2.1. Accuracy

Listeners were more accurate in response to actives (96.18%) than passives (95.28%; see Figure 4.2.2), but the impact of syntax on accuracy was not reliable (*Mean* β = -0.18, z = 0.65, p = .51; see Table 4.2.1), but 66.38% of samples fell below zero, which may reflect the presence of a small effect of syntax on accuracy. Similarly, though listeners were more accurate for audio containing manipulated verb stem vowels (96.46%) as compared to unmanipulated audio (94.91%), the difference was not reliable (z = -0.77, p = .44); however, 79.1% of samples in this comparison fell above zero, which suggests a small effect of accuracy across the audio manipulation condition. A model with an interaction term for syntax and audio manipulation was slightly less able to predict the data as compared to an otherwise equivalent model with no interaction ($ELPD_{diff}$ = 0.7, *SE* = 0.3).

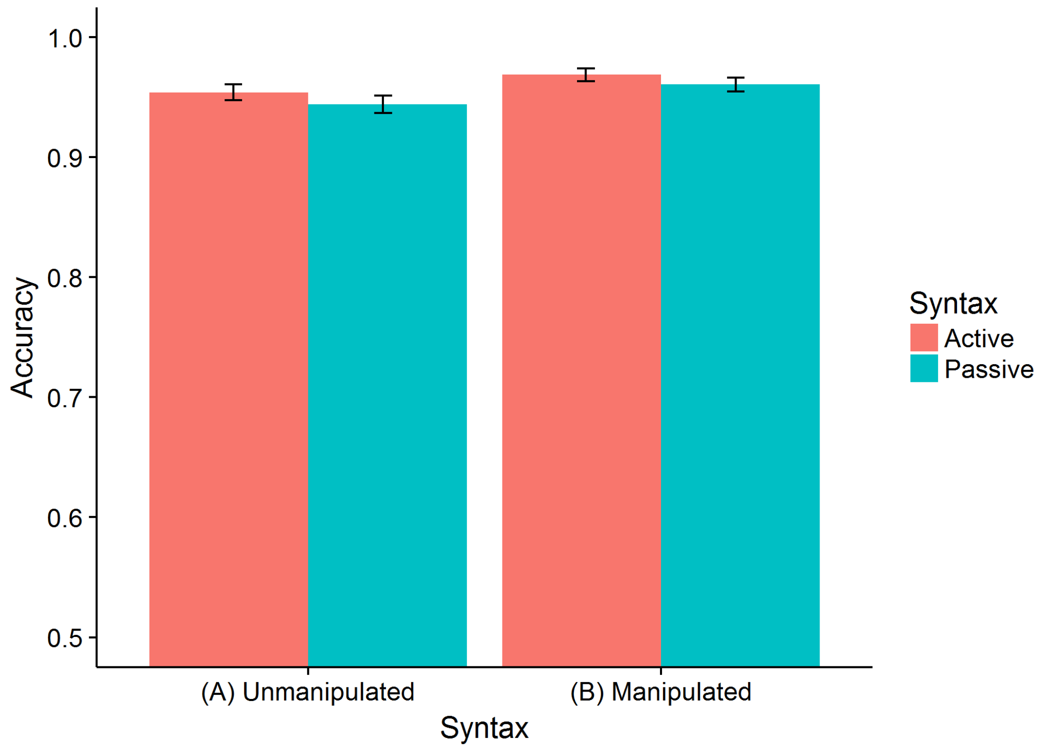


Figure 4.2.2. Experiment 4b: Accuracy for active and passive sentences for both (A) unmanipulated audio and (B) audio containing manipulated verb stem vowels.

Table 4.2.1.

Experiment 4b: Bayesian Logistic Mixed-Effects Model Summary for Syntax and Audio Manipulation on Accuracy

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			1.0	0.0
Random Effects	*SD*	*r*		
Subject (*n* = 38)				

Intercept	1.09	–	
Passive Syntax	0.83	-0.20	
Verb (*n* = 14)			
Intercept	0.23	–	–
Manipulated Audio	0.16	-0.09	–
Passive Syntax	0.17	-0.22	-0.21

Fixed Effects	*Median* *β*	*β* MAD_{SD}
Intercept	3.5	0.3

					Posterior Interval	
	β	*SE*	*z*	*p*	5%	95%
Audio: Unmanipulated vs. Manipulated	-0.31	0.41	-0.77	.44	-0.35	1.07
Active vs. Passive	0.18	0.27	0.65	.51	-0.69	0.43

In summary, comprehension accuracy was high, and did not appear to differ due to syntax or the audio manipulation. Next, we assess processing of the sentences by analyzing response time.

4.2.2.2. Reaction time

Syntax. Reaction time was 170.27 ms higher for passive sentences (M = 2222.66 ms, SE = 13.12 ms) than for active sentences (M = 2052.39, SE = 12.89; see Figure 4.2.3), consistent with the classic finding that, overall, passive sentences are more difficult to comprehend than active sentences (for a review, see Stromswold et al., under review). The difference in response time across constructions was significant (z = -6.62, p < .01; see Table 4.2.2).

Audio manipulation. Response times were 141.18 ms slower for listeners who heard manipulated audio (M = 2203.35 ms, SE = 12.38 ms) vs. unmanipulated audio (M = 2062.17, SE = 13.80). The difference was not significant according to traditional statistical tests (z = -1.37, p - .17); however, 93.03% of samples exceed 0, which may indicate that the effect estimate is not zero, but may be too small to detect in this analysis. Recall that the audio manipulation condition was a between-subjects manipulation. Despite the fact that participants were randomly assigned to audio manipulation conditions, it is possible that the apparent differences in response time for manipulated and unmanipulated audio files actually reflect differences between individuals in the two audio manipulation conditions—but this seems unlikely. Note that because the audio manipulation was realized between-subjects, random slopes by audio manipulation condition could not be included in the by subject random effect structure. A model that contained an interaction between syntax and audio manipulation did not predict the data better than an otherwise equivalent model without the interaction term ($ELPD_{diff}$ = 0.4, SE = 0.8).

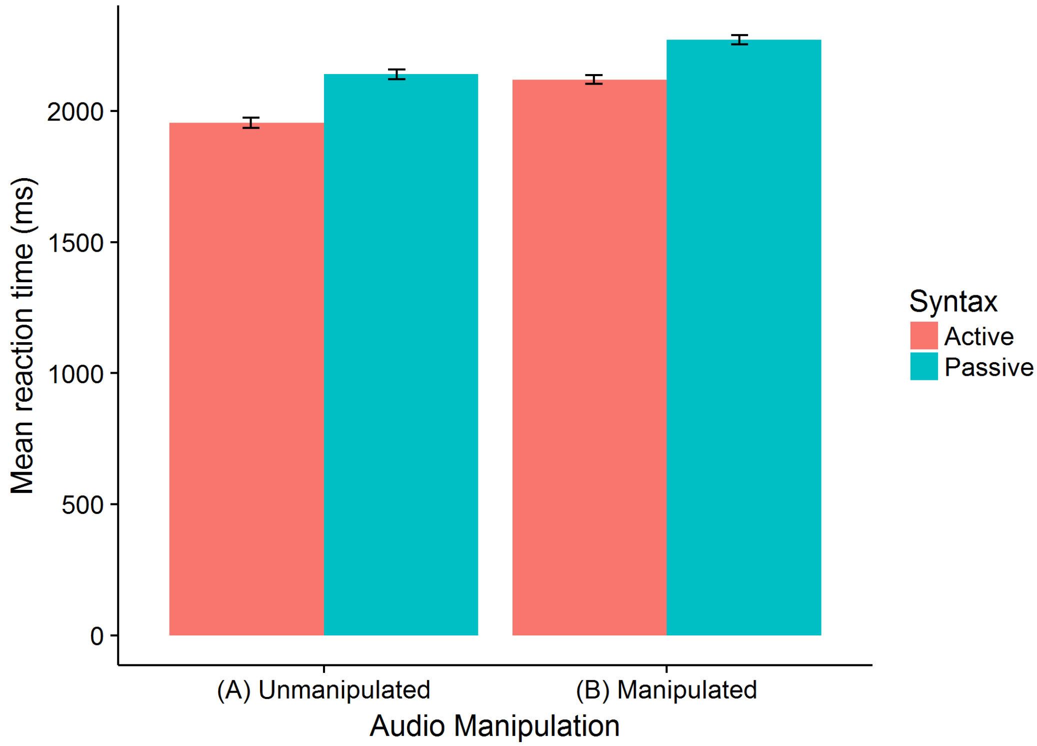


Figure 4.2.3. Experiment 4b: Mean reaction time for active and passive sentences for both (A) unmanipulated audio and (B) audio containing manipulated verb stem vowels. Error bars indicate standard error of the mean.

Table 4.2.2.

Experiment 4b: Bayesian Linear Mixed-Effects Model Summary for Syntax and Audio Manipulation on Reaction Time

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			2125.60	10.0
Random Effects	*SD*	*r*		
Subject (*n* = 38)				
Intercept	363	–		
Passive Syntax	62	-0.03		
Verb (*n* = 14)				

Intercept	95	–	–			
Manipulated Audio	44	-0.06	–			
Passive Syntax	70	-0.26	0.12			
Residual	461	–	–			
Fixed Effects	*Median β*	*β* *MAD_{SD}*				
Intercept	1950.4	85.5				
Error *SD*	460.6	5.0				
					Posterior Interval	
	β	*SE*	*z*	*p*	5%	95%
Syntax: Active vs. Passive	-168.05	25.37	-6.62	< .01	132.35	234.09
Audio: Unmanipulated vs. Manipulated	-159.18	116.28	-1.37	.17	-22.46	360.51

Note: Four divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

Speed-accuracy tradeoff. In order to determine whether listeners were slower to process sentences in favor of comprehension accuracy, a speed-accuracy tradeoff analysis was performed. This was done by testing a Bayesian linear mixed-effects model where reaction time was the dependent variable, and response accuracy (0 or 1) was included as a fixed factor along with syntax (active or passive). Random slopes for response accuracy were included in both the by subject and by item random effects terms.

Response time was 271.87 ms higher overall for trials where responses were accurate (M = 2137.13 ms, SE = 9.29 ms, n = 4057) than when they were inaccurate (M =

1865.26 ms, *SE* = 40.82 ms, *n* = 181; *z* = -2.87, *p* = .004), which indicates the presence of a speed-accuracy tradeoff. Even when accuracy was included as a factor, response times were longer for passives than for actives (*z* = -6.54, *p* < .001; see Table 4.2.3).

Listeners responded 277.44 ms slower for active sentences when their response was accurate (*M* = 2052.39 ms, *SE* = 12.89 ms, *n* = 2038) than when it was inaccurate (*M* = 1774.95 ms, *SE* = 57.60 ms, *n* = 81; see Figure 4.2.4). The speed-accuracy tradeoff for active sentences was significant (*z* = -2.32, *p* = .02). This also occurred for passive sentences, and the difference was reliable (*z* = 2.49, *p* = .01): listeners were 284.24 ms slower for passive sentences when responses were accurate (*M* = 2222.66 ms, *SE* = 13.12 ms, *n* = 2019) than when inaccurate (*M* = 1938.42 ms, *SE* = 56.50 ms, *n* = 100). This suggests that listeners incur a processing cost in order to maximize comprehension. Additionally, processing time was 163.47 ms higher for passive sentences than active sentences, even when listeners responded incorrectly (*z* = -2.12, *p* = .03). Overall, speed-accuracy tradeoffs were present for both passive and active sentences.

Listeners who heard unmanipulated audio were 289.57 ms faster to respond incorrectly (*M* =1772.60 ms, *SE* = 59.15 ms) than correctly (*M* = 2062.17 ms, *SE* = 13.80 ms; *z* = -3.01, *p* = .003; see Figure 4.2.5). The difference was not reliable for participants who heard sentences containing manipulated verb stem vowels (*z* = -1.22, *p* = .22), but 96.4% of samples for this comparison fell below zero, which may suggest the presence of a small effect that was not otherwise captured by our analysis. The difference between unmanipulated and manipulated audio was marginally significant in this analysis (*z* = -1.84, *p* = .07; see Table 4.2.3), which further suggests that there may be a small—but not reliable—effect of the audio manipulation on processing.

For incorrect trials, participants in the unmanipulated audio condition responded 141.18 ms faster than participants who heard manipulated audio, and this difference was reliable (z = -2.19, p = .03). This suggests that listeners who heard unmanipulated audio responded incorrectly when they were too fast, but the same explanation does not apply to listeners in the manipulated audio condition. When listeners who heard manipulated audio responded incorrectly, they may have done so for a different reason. Another possibility is that listeners in the manipulated audio condition were slower to respond due to individual differences, and were slow enough that they could maintain high accuracy.

Model comparison was used to test for 2- and 3-way interactions between the fixed effects (syntax, accuracy, and audio manipulation). For all comparisons, the model with fewer (or zero) interactions terms predicted the data better than the models without interaction terms (all $ELPD_{diff} > 0$).

Table 4.2.3.

Experiment 4b: Bayesian Linear Mixed-Effects Model Summary for Syntax and Audio Manipulation on Speed-Accuracy Tradeoff

				Median	MAD_{SD}
Mean Posterior Predictive Distribution				2125.8	10.3
Random Effects	*SD*	*r*			
Subject (n = 38)					
Intercept	285	–	–		
Passive Syntax	58	0.19	–		
Accuracy = 1	220	0.01	-0.05		
Verb (n = 14)					

Intercept	77	–	–	–		
Passive Syntax	67	-0.41	–	–		
Manipulated Audio	50	-0.05	0.19	–		
Accuracy = 1	62	0.02	-0.09	-0.10		
Residual	458	–	–	–		
Fixed Effects	*Median β*	*β MAD_{SD}*				
Intercept	1677.7	99.1				
Error *SD*	458.2	4.8				
					Posterior Interval	
	β	*SE*	*z*	*p*	5%	95%
Accuracy (1 vs. 0)	-167.73	58.55	-2.87	< .01	137.75	451.09
Syntax (Passive vs. Active)	-164.76	42.96	-3.84	< .01	126.41	454.48
Audio (Unmanipulated vs. Manipulated)	-198.69	108.06	-1.84	.07	161.98	638.55
Accuracy x Syntax						
Active (0 vs. 1)	-165.77	71.33	-2.32	.02	–	–
Passive (0 vs. 1)	-169.69	68.11	-2.49	.01	-273.12	53.14
Accuracy = 0 (Syntax)	-162.80	76.89	-2.12	.03	–	–
Accuracy = 1 (Syntax)	-166.72	25.51	-6.54	< .01	–	–
Accuracy x Audio						
Unmanipulated (0 vs. 1)	-235.94	78.32	-3.01	< .01	–	–
Manipulated (0 vs. 1)	-99.52	81.59	-1.22	.22	-471.01	-31.88

Accuracy = 0 (Audio)	-266.90	121.93	-2.19	.03	–	–
Accuracy = 1 (Audio)	-130.47	120.09	-1.09	.28	–	–

Note: Five divergent transitions occurred during sampling, which indicates a possibility that the sampler was biased.

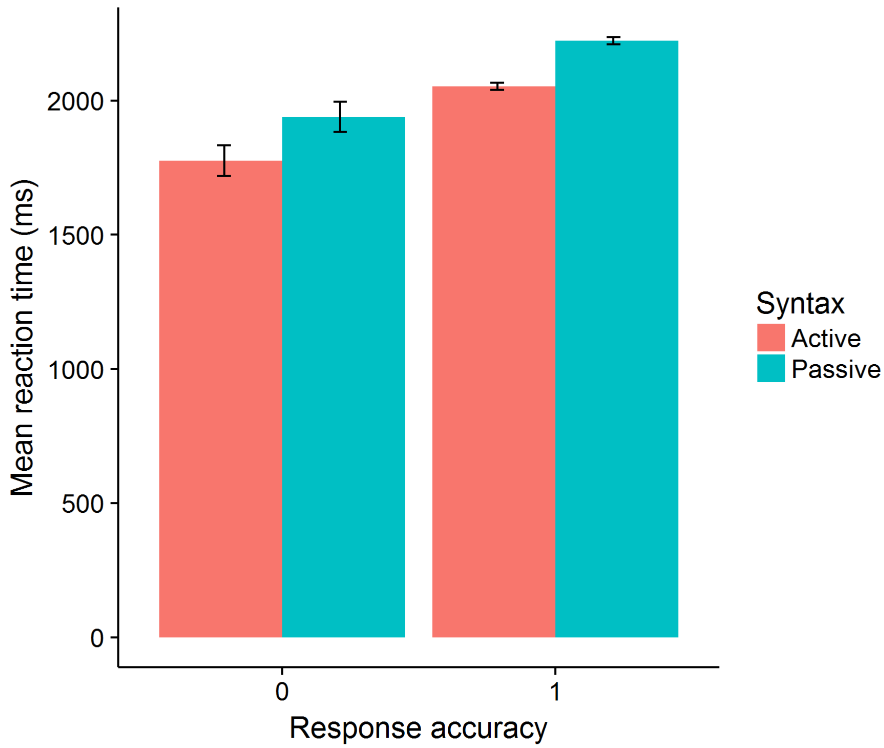


Figure 4.2.4. Experiment 4b: Mean reaction time across sentence syntax (active and passive) and response accuracy. Error bars indicate standard error of the mean.

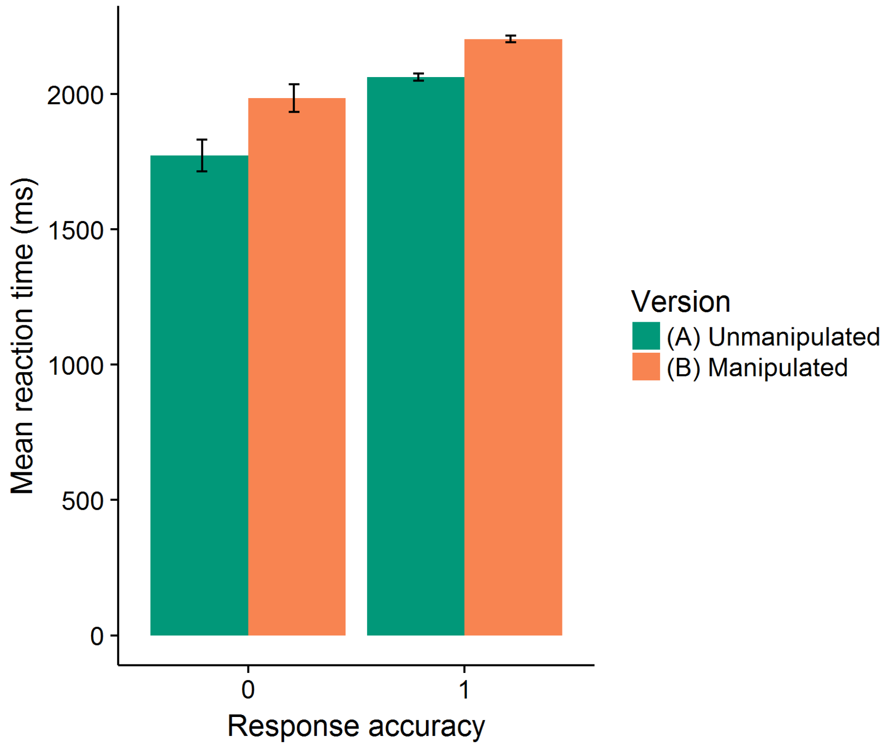


Figure 4.2.5. Experiment 4b: Mean response time in across the audio condition (manipulated or unmanipulated) and accuracy of responses. Error bars indicate standard error of the mean.

In sum, processing was affected by the syntax of the sentence. Passives were more difficult to process than actives, and this was reflected in higher response times for passive sentences. In contrast, response times for listeners who heard manipulated verb stem vowels were only slightly higher than for listeners who heard unmanipulated audio. We found evidence for speed-accuracy tradeoffs for both syntax conditions (active and passive), and within the unmanipulated audio condition. Next, we turn to gaze behavior to assess processing at a more fine-grained level.

4.2.2.3. Gaze behavior

Qualitative description. First, we present a qualitative description of eye-gaze behavior. In order to visualize gaze traces, we calculated the proportion of looks to the target image for each frame in a sentence. In all visualizations, gaze traces were aligned to the onset of the verb stem. Average morpheme onset markers are included in each visualization, and have been shifted approximately 200 ms to the right in order to account for the time required to plan and execute a saccade in response to speech (6 frames * 33.33 ms = 199.98 ms). All visualizations include a line marking 0.50 on the y-axis, reflecting chance level performance.

For active sentences in the unmanipulated audio condition, listeners appeared to look to the correct image as early as the auxiliary, with a monotonic increase in proportion looks to the target during the end of the verb stem. For passives, listeners look to the incorrect image (reflect an active bias) until late in the verb stem, after which a shift in looking behavior towards the correct image takes place. This pattern is reflected both in the proportion looks to the target image (Figure 4.2.6 A) and in the number of saccades made to the target image (Figure 4.2.6. B).

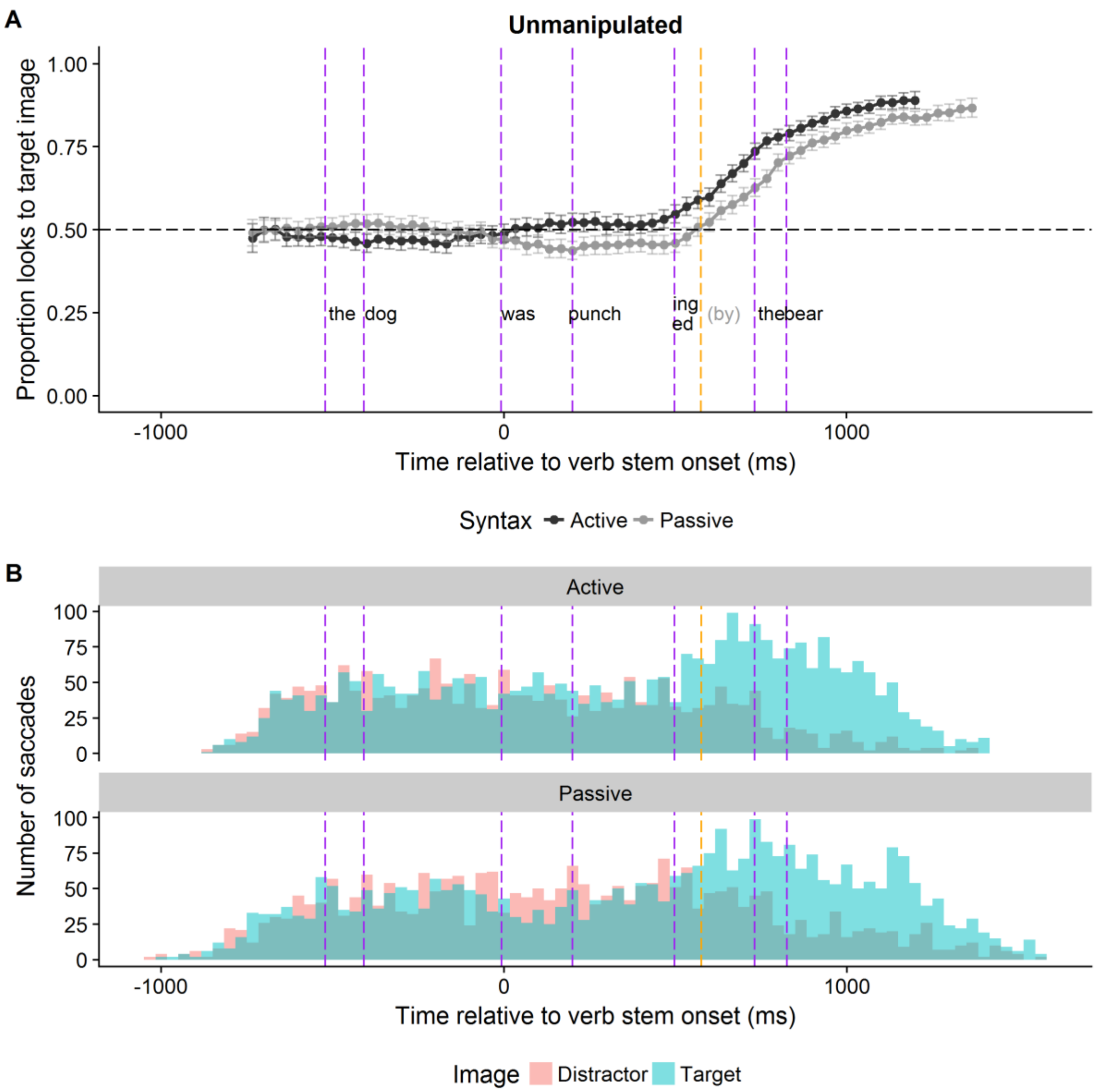


Figure 4.2.6. Experiment 4b: A) Proportion of looks to the target image for active and passive sentences in the unmanipulated audio condition. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. B) Number of saccades to the target or the distractor image relative to the onset of the verb stem in the unmanipulated audio condition, shown separately for active and passive sentences. Dark blue reflects overlap between looks to the target and the distractor. For both A and B, average morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

A similar trend occurs in the manipulated audio condition, except that for both active and passive sentences, the monotonic increase in looks to the target image occurs after the verb stem. This pattern is reflected both in the proportion looks to the target image (Figure 4.2.7 A) and in the number of saccades made to the target image (Figure 4.2.7. B).

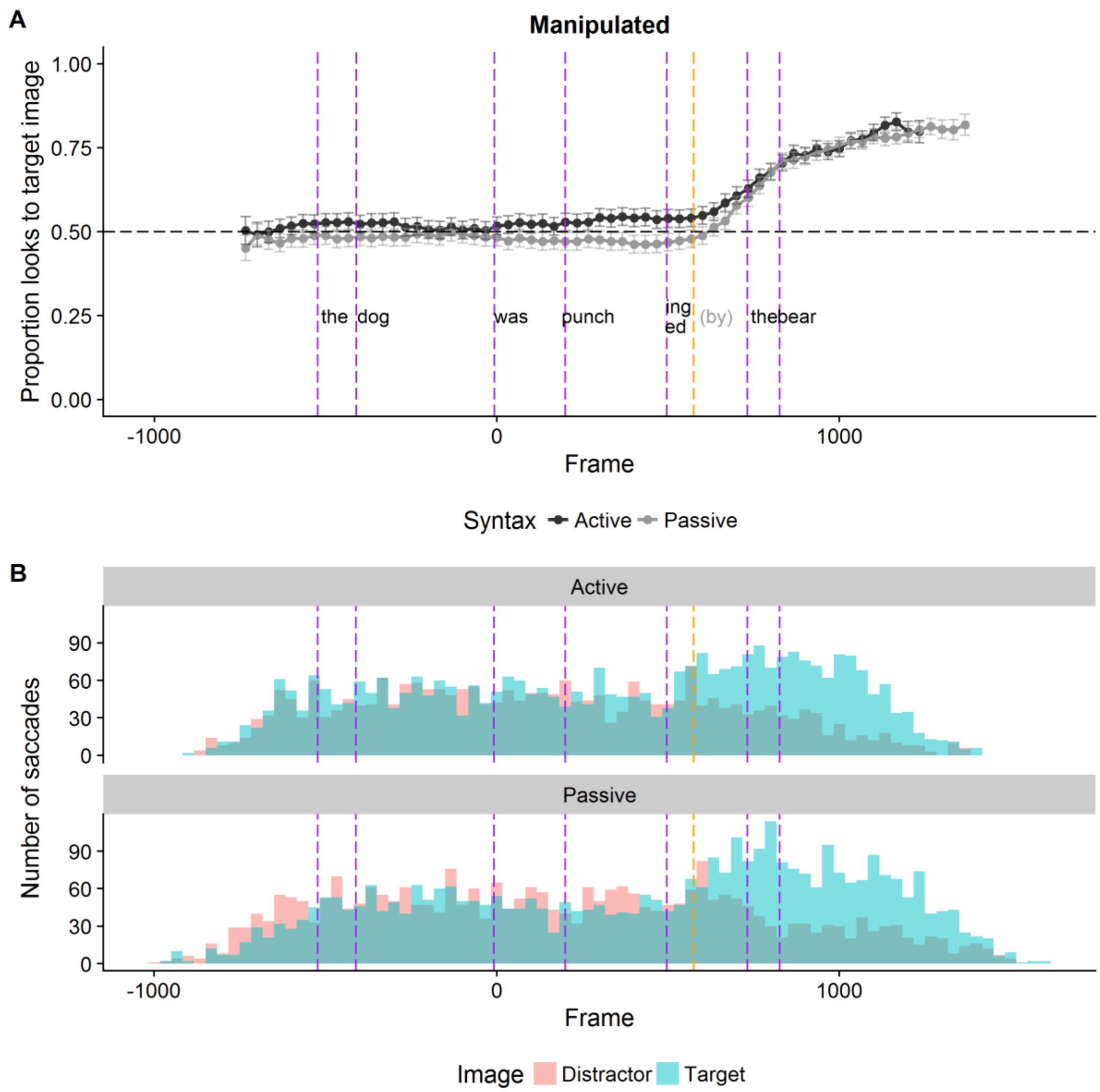


Figure 4.2.7. Experiment 4b: A) Proportion of looks to the target image for active and passive sentences in the manipulated audio condition. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. B) Number of saccades to the target or the distractor image relative to the onset of the verb stem in the unmanipulated audio condition, shown separately for active and passive sentences. Dark blue reflects overlap between looks to the target and the distractor. For both A and B, average morpheme onset markers are

shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

Gaze behavior changed over blocks in the unmanipulated audio condition (Figure 4.2.8). In block 1, the presence of an active bias was more apparent, as demonstrated by the separation between gaze traces across active and passive sentences. By bock 2, however, the difference decreased.

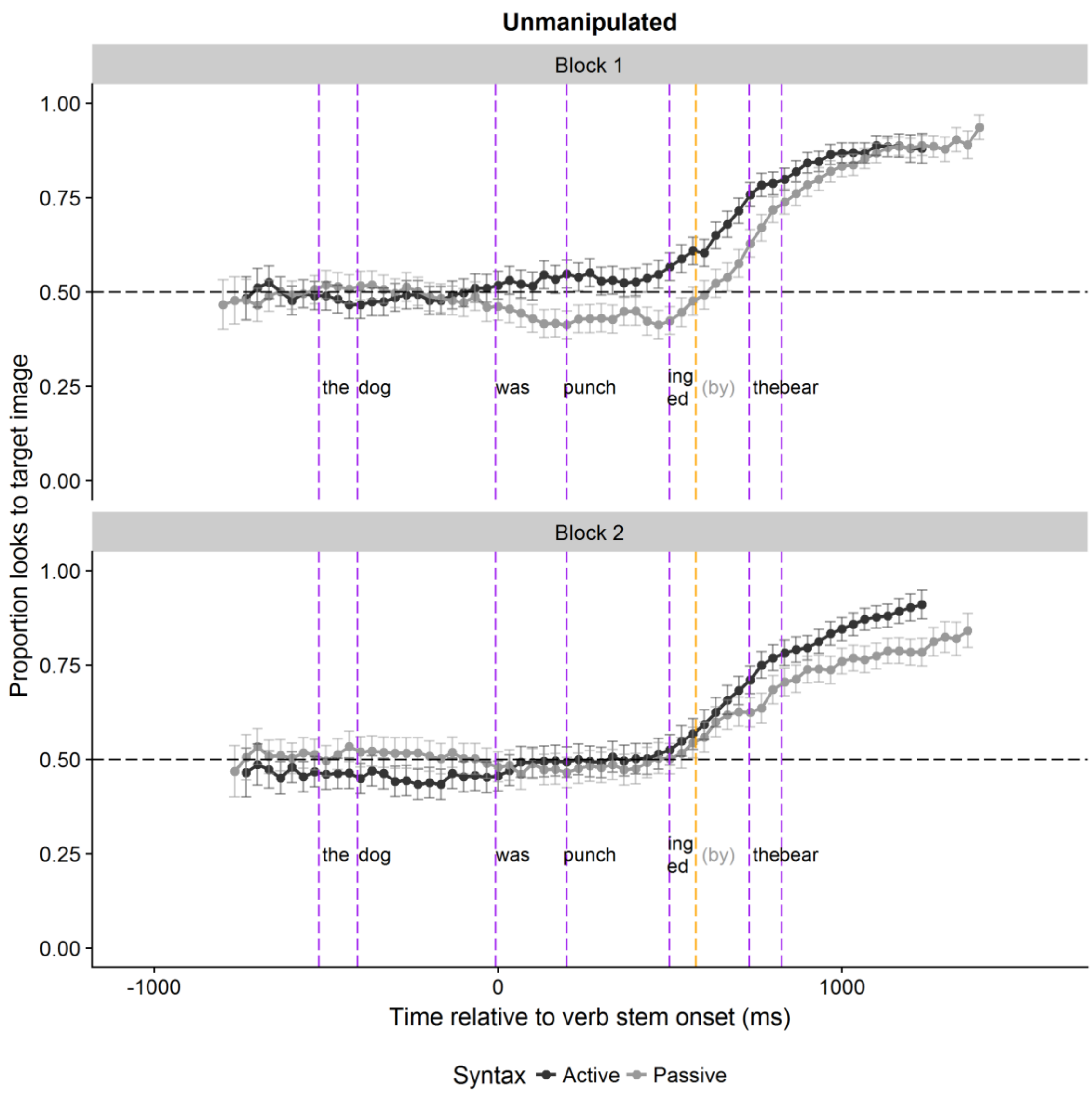


Figure 4.2.8. Experiment 4b: Proportion of looks to the target image for active and passive sentences in the unmanipulated audio condition, separated by experimental block. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance

performance (.50), and vertical lines indicate average morpheme onsets. Morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

In the manipulate audio condition, however, gaze behavior did not appear to change across blocks: in both blocks, there is a slight active bias (Figure 4.2.9). See section 4.2.2.3.1 for further analysis and discussion on learning behavior.

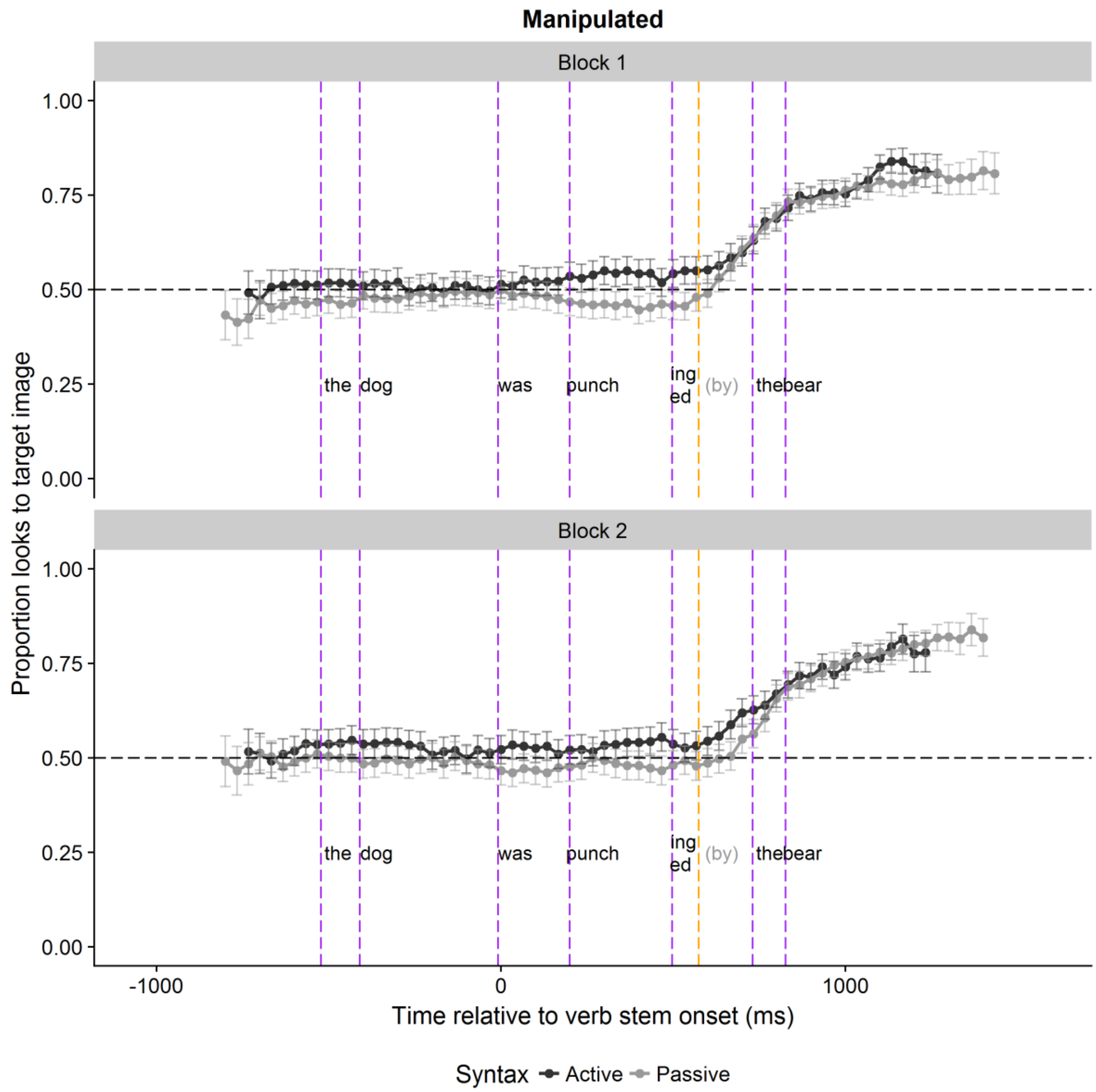


Figure 4.2.9. Experiment 4b: Proportion of looks to the target image for active and passive sentences in the manipulated audio condition, separated by experimental block. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate average morpheme onsets. Morpheme onset

markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

For active sentences (Figure 4.2.10), listeners appeared to look more frequently to the target image during the verbal inflection. Additionally, listeners in the manipulated verb stem vowel condition trend toward looking at the correct image slightly later than the listeners who heard unmanipulated audio files. Note that for the unmanipulated audio condition only, listeners look more frequently to the distractor image early on. This may be the result of learning over the course of the experiment, and will be discussed in more detail in the next section (4.2.2.3.1.).

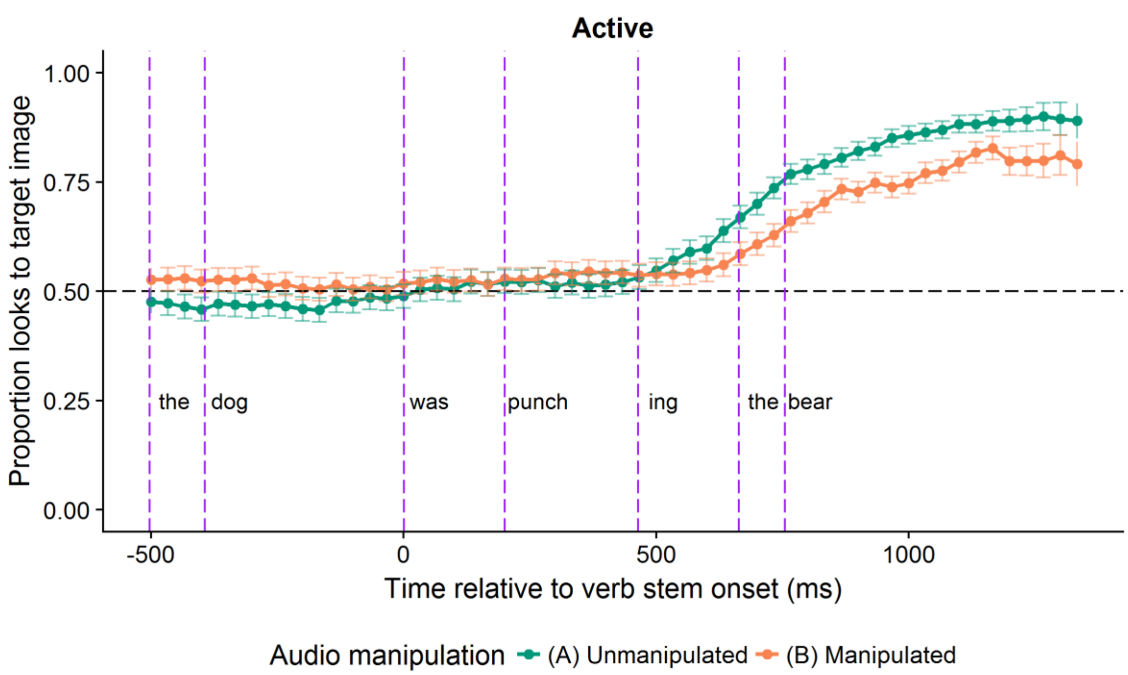


Figure 4.2.10. Experiment 4b: Proportion of looks to the target image for active sentences. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. Morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

For passive sentences (Figure 4.2.11.), listeners in both groups looked to the distractor image more prior to the verb stem. An inflection point appeared to occur late in the verb stem, where listeners shifted to look to the target image more frequently. This shift occurred slightly sooner for listeners who heard only unmanipulated verb stem vowels.

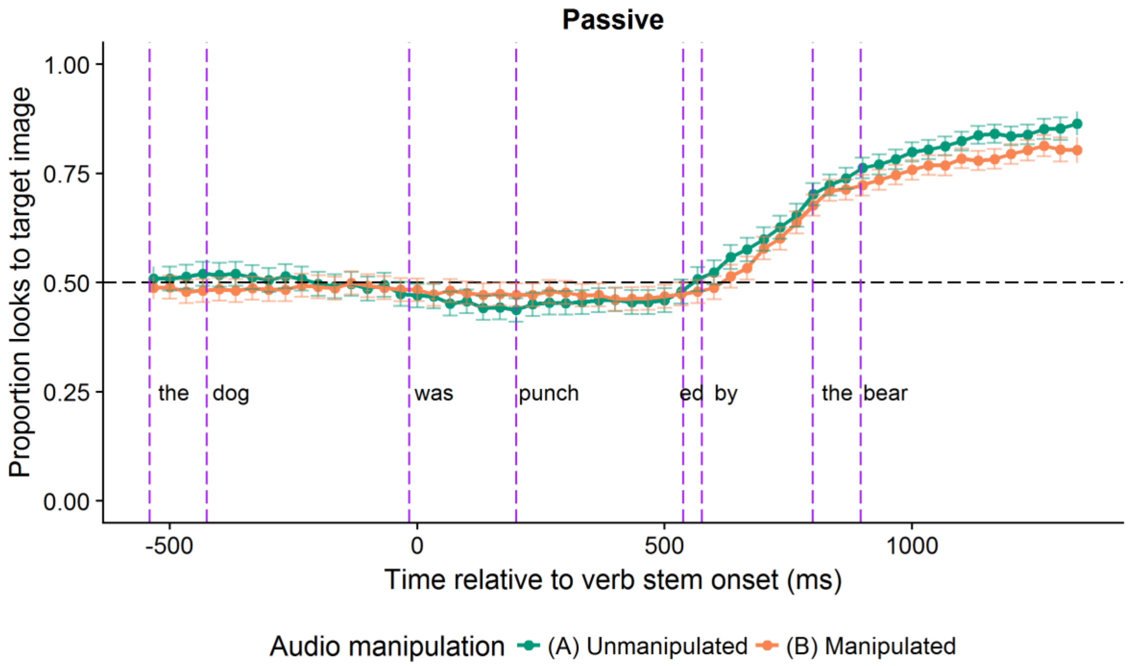


Figure 4.2.11. Experiment 4b: Proportion of looks to the target image for passive sentences, separated by whether verb stem vowel duration was manipulated or unmanipulated. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. Morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

In summary, listeners appeared to look to the correct image during the verb stem for passive sentences, and during the inflection for active sentences. Some separation in gaze traces occurred across the audio manipulation conditions, but it was slight. Next, we turn to quantitative analyses of saccade behavior to determine what factors influenced listeners' saccades after they heard the verb stem vowel.

Verb stem analysis. To determine whether listeners looked to the target or the distractor image in response to acoustic information in the verb stem, an analysis of saccade behavior was performed. If acoustic cues within the verb stem facilitate processing, saccades should be made within this region. The analysis of saccade behavior was carried out as follows. For each frame in the parsed eye-gaze data, whether a saccade occurred, the region where the saccade landed, and when the saccade occurred were recorded. The occurrence of a saccade was determined by a change greater than 110 pixels (approximately 1° visual angle using our apparatus) in the x-coordinate. The y-coordinate was not used for this purpose due to high noise and data loss.

If a saccade was made to the target, this was assigned a value of 1, and saccades to the distractor were assigned the value 0. Saccades that occurred after the onset of the verb stem but prior to the onset of the verbal inflection were analyzed. The analysis window was adjusted to account for the time required to plan and execute a saccade in response to speech, approximately 200 ms (Salverda et al., 2014). Saccades within this analysis window were then analyzed using a Bayesian logistic mixed-effects model. To control for variation due to individual differences or due to the items tested, random slopes and random intercepts were included for the syntactic frame (active or passive), the audio condition (manipulated or unmanipulated), and experimental block (1 or 2), by subjects and by item (verb). See Figure 4.2.12 for a visualization of the number of saccades made to either the target image or the distractor during the region of interest.

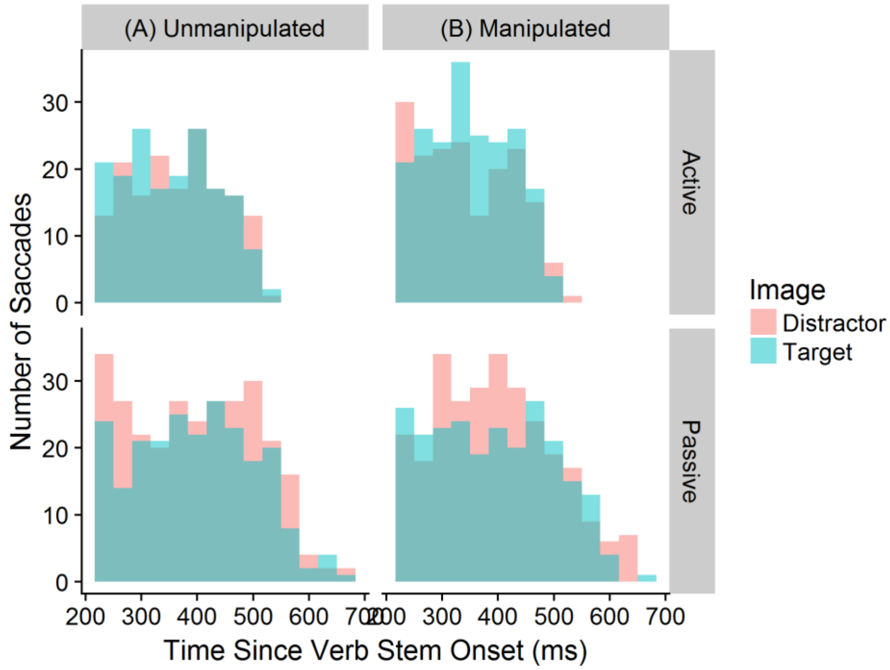


Figure 4.2.12. Experiment 4b: Histograms of saccades to the target (blue) and distractor (red) in response to the verb stem. Rows correspond to syntax (active or passive) and columns indicate the audio manipulation (unmanipulated or manipulated). Bin size corresponds to the sampling frequency of the eye tracker (1 Hz, or 1 frame every 33.33 ms).

Listeners were more likely to make a saccade to the target image during this region in active sentences than in passive sentences (z = 2.62, p < .01; see Table 4.2.4 for model summary). This is in line with the processing advantage found for actives in response time, and suggests that listeners form an expectation about the syntax of the sentence prior to hearing the verbal inflection. Saccade behavior in this region was not reliably predicted by whether the listener heard manipulated audio compared to unmanipulated audio (*Mean* β = -0.08, z = -0.77, p = .44), and the magnitude of the mean β suggests that the effect size estimate is near zero. The findings for the effect of block on saccade behavior were similar, except that listeners were marginally less likely to look at the target image in block 1 as compared to block 2 for passive sentences only (z = -1.90,

p = .06), which suggests that listeners learn to expect to hear passives by the second half of the experiment.

Model comparison was used to test whether 2- and 3-way interactions between fixed effects improved the ability of the model to predict the data. In all cases, models with interactions did not predict the data better than models without those interactions (all $ELPD_{diff} > 1.5$).

Table 4.2.4.

Experiment 4b: Bayesian Logistic Mixed-Effects Model Summary for Saccade Behavior

					Median	MAD_{SD}
Mean Posterior Predictive Distribution					0.5	0.0
Random Effects	*SD*		*r*			
Subject (n = 38)						
Intercept	0.062	—	—			
Passive Syntax	0.078	-0.24	—			
Block 2	0.082	-0.08	-0.19			
Verb (n = 14)						
Intercept	0.060	—	—	—		
Passive Syntax	0.075	-0.21	—	—		
Manipulated Audio	0.069	-0.01	-0.29	—		
Block 2	0.066	-0.13	-0.02	-0.13		
Fixed Effects	*Median β*	*β* MAD_{SD}				
Intercept	0.10	0.10				

					Posterior Interval	
	Mean β	*SE β*	*z*	*p*	5%	95%
Active vs. Passive	0.26	0.10	2.62	< .01	-0.78	-0.13
Manipulated Audio	-0.08	0.10	-0.77	.44	-0.41	0.25
Block 2	-0.15	0.10	-1.45	.15	-0.50	0.24
Block 1 vs. Block 2						
Active Syntax	-0.043	0.15	-0.28	.78	—	—
Passive Syntax	-0.25	0.13	-1.90	.06	-0.04	0.89
Audio Manipulation						
Active Syntax	-0.097	0.15	-0.63	.53	—	—
Passive Syntax	-0.058	0.13	-0.44	.66	-0.25	0.62

Overall, listeners looked to the target image more for active sentences than passives prior to hearing the verbal inflection, which suggests some degree of predictive processing. Saccade behavior was also influenced by learning, where listeners looked to the target image more in the second block of the experiment for passive sentences. The audio manipulation, however, did not appear to affect saccade behavior, or the size of the effect may be too small to detect in this analysis. This suggests that predictive processing for these sentences is not informed solely by the duration of the verb stem vowel. Next, we turn to the role experience may play in saccade behavior earlier in the sentence.

4.2.2.3.1. Use of heuristic parsing strategies.

Qualitative description. As in the previous discussion of gaze behavior, we will first consider a qualitative presentation of the data, starting with active sentences (Figure 4.2.13). To assess changes in parsing strategies over time—namely, use of the NP1 =

agent strategy—we examined changes in gaze traces across blocks. A separation occurred relatively early in the sentence for the proportion of looks to the target in block 1 as opposed to block 2, but only for audio where verb stem vowels were unmanipulated. In block 1, there was a tendency for listeners to look more frequently to the target image, which most likely reflects a bias toward active sentences (the first NP = agent strategy). By block 2, however, listeners are more likely to look to the distractor image. This suggests that listeners learn passives are more frequent in our experiment than in English overall, and learn to expect them. If this effect is reliable, it supports the findings of Thothathiri & Snedeker (2008), and syntactic priming more generally.

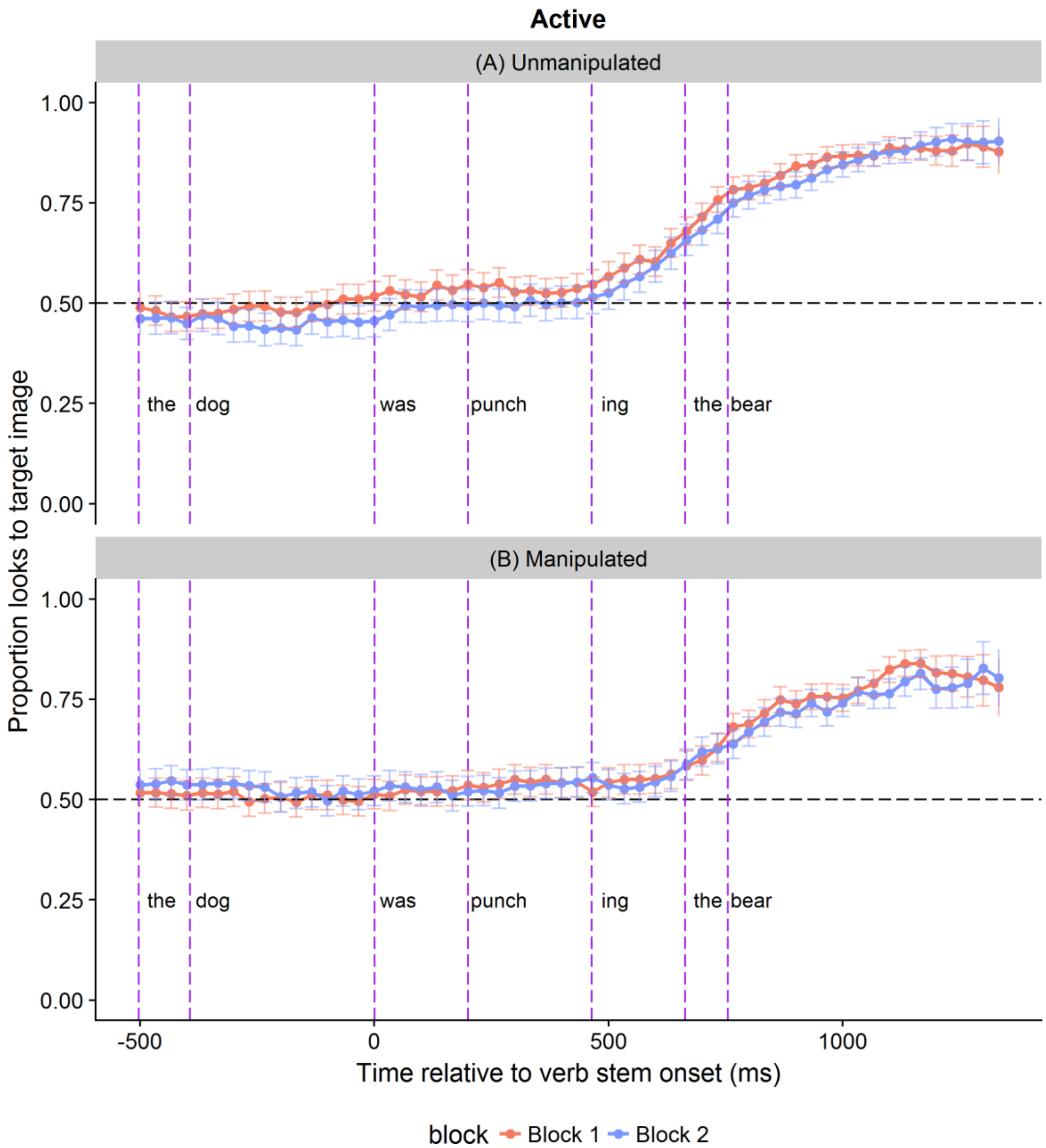


Figure 4.2.13. Experiment 4b: Proportion of looks to the target image for active sentences, separated by audio manipulation and block. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. Morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

This separation is even more pronounced when looks to the target are compared between the first quarter of the experiment and the last quarter of the experiment (Figure 4.2.14). Note that because of the way the sentences were balanced, the first quarter of the experiment and the last quarter of the experiment contain the same sentences. This means the differences did not arise due to differences between sentences.

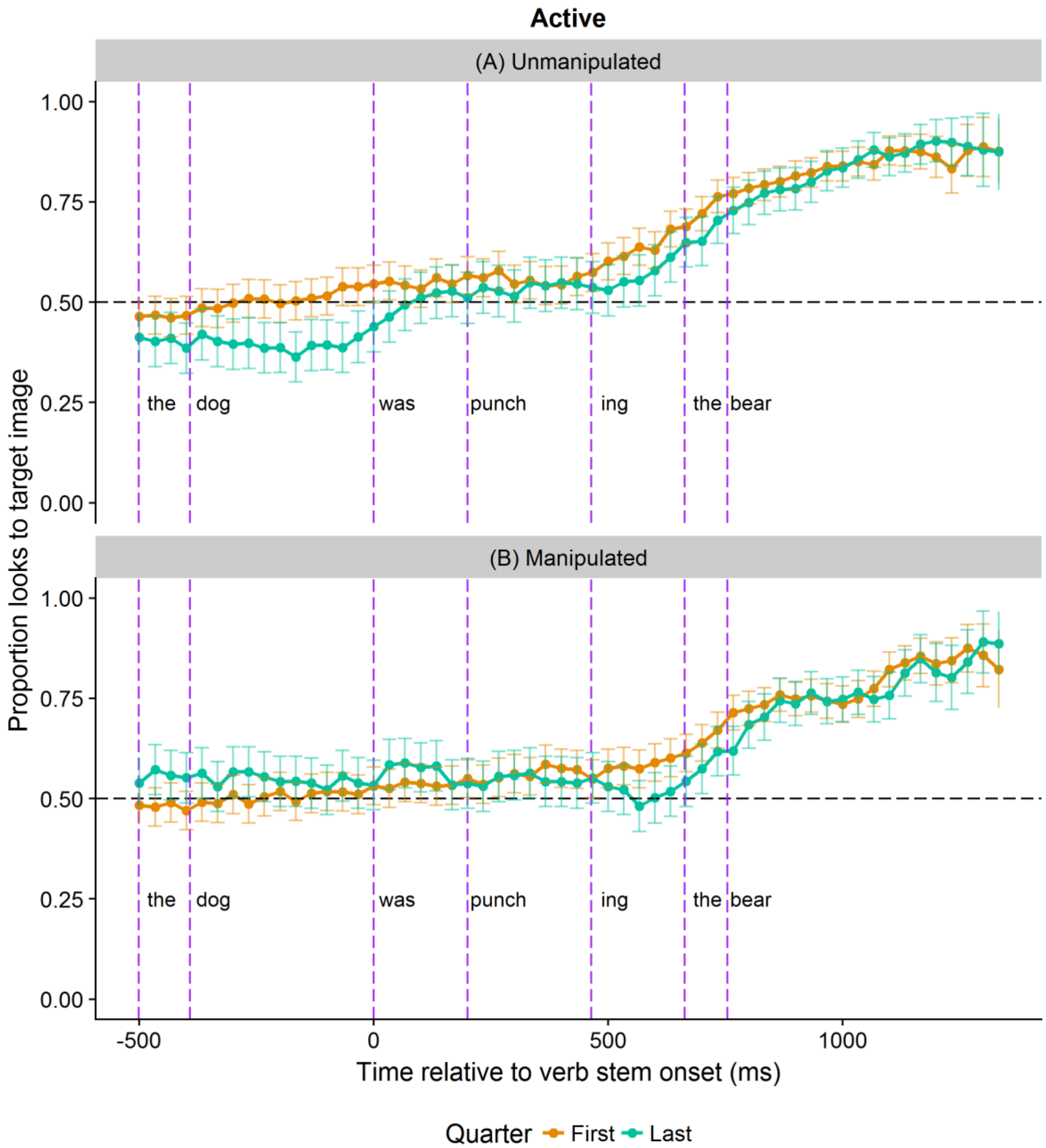


Figure 4.2.14. Experiment 4b: Proportion of looks to the target image for active sentences, separated by the audio manipulation, then by first quarter (first half of bock 1) and last quarter (second half of block 2) of the experimental trials. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. Morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

A similar trend can be observed for passive sentences (see Figure 4.2.15). In the first block, listeners looked more frequently to the distractor image (at least for sentences with unmanipulated audio) which corresponds to the active completion of the sentence. Again, this most likely reflects a bias towards active sentences, due to the relative frequency of active sentences compared to passives in the wild. By the second block, looks to the target approached chance level, indicating a decrease in the active bias.

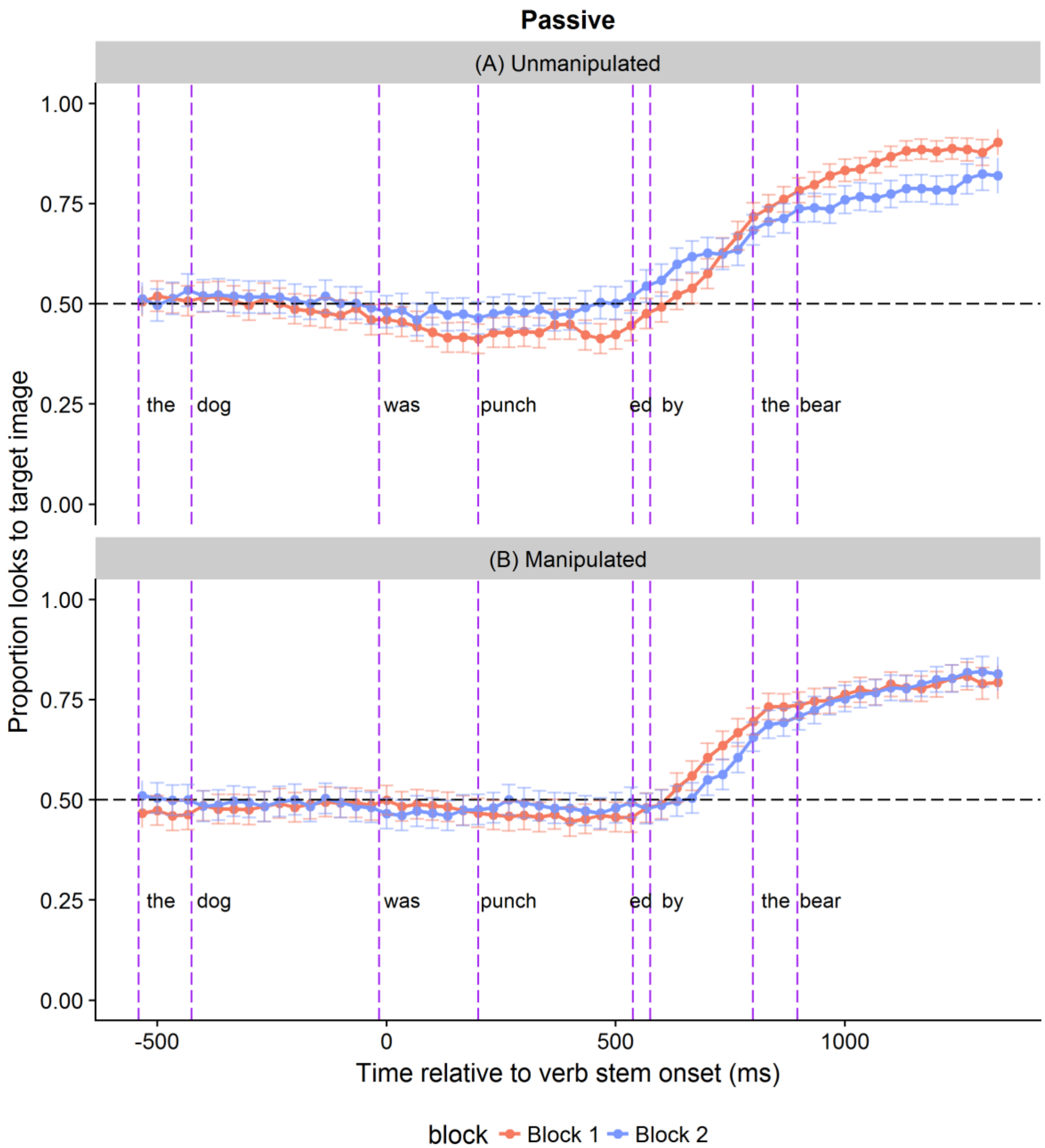


Figure 4.2.15. Experiment 4b: Proportion of looks to the target image for passive sentences, separated by audio manipulation and block. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. Morpheme onset markers are shifted to the right by 6 frames to account for the time required to plan and execute an eye movement.

The separation is once again made more apparent by separating the data into the first and last quarter of the experiment (Figure 4.2.16).

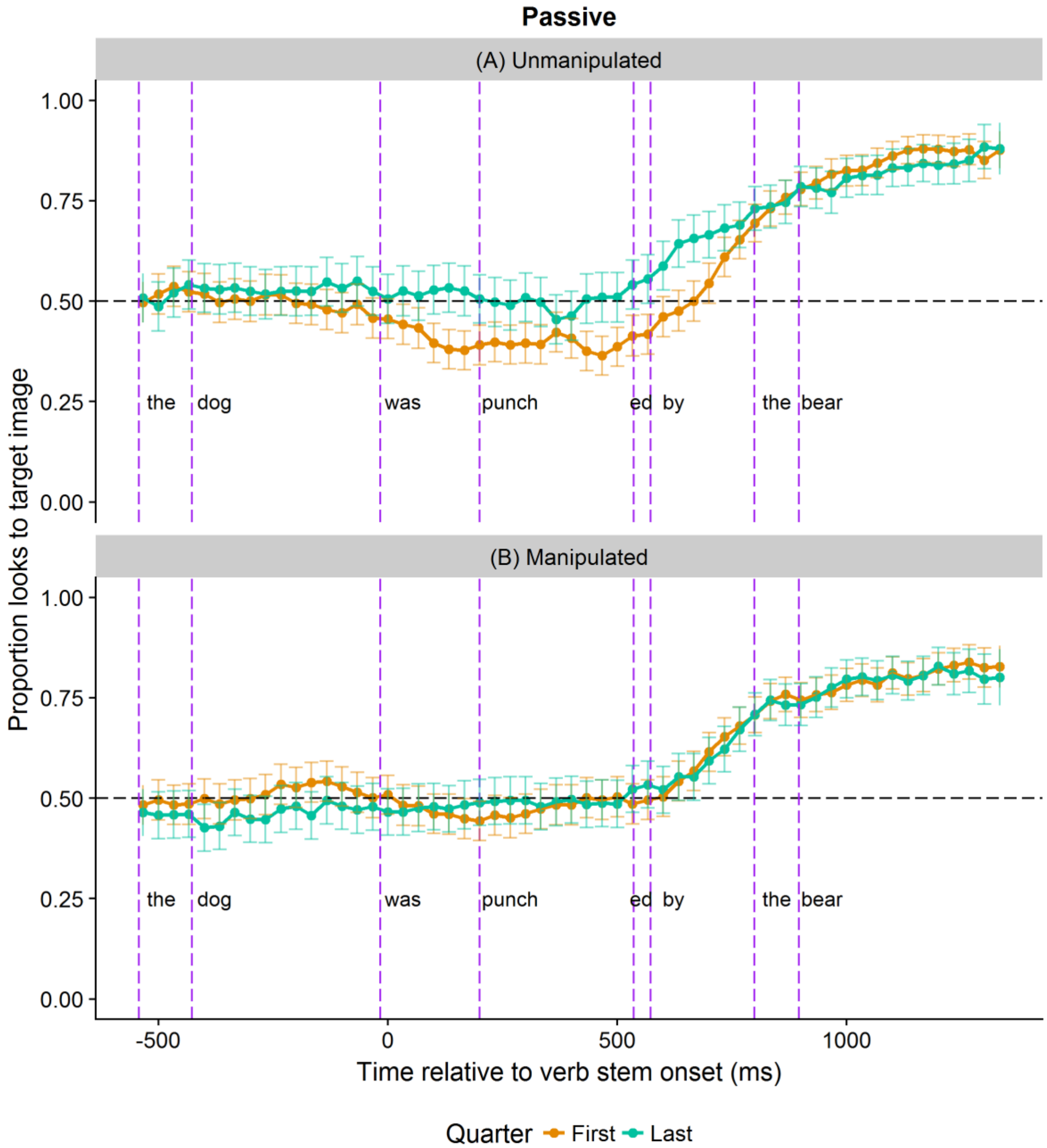


Figure 4.2.16. Experiment 4b: Proportion of looks to the target image for passive sentences, separated by audio manipulation, then by the first quarter (first half of bock 1) and last quarter (second half of block 2) of the experimental trials. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. Morpheme onset markers are shifted to the right by 6 frames to account for the time required to plan and execute an eye movement.

Recall from Chapter 2 that some models of sentence processing (e.g., Townsend & Bever, 2001; Crocker, 2002; Ferreira et al., 2002) propose two-stage processing models in which the initial parse is shallow and relies on heuristics. The second stage then checks the result of the initial shallow stage and performs a deeper analysis of the sentence only if necessary. This type of processing model is compatible with an active bias, where the first noun is assumed to be the agent of the sentence (Bever, 1970). To assess qualitatively whether listeners may be using such a strategy, we visualized gaze traces when listeners responded correctly vs. when their responses were incorrect.

When listeners responded incorrectly to active sentences in the unmanipulated audio condition (Figure 4.2.17), they initially demonstrated an active bias, where they looked to the active image more frequently during the first noun. During the auxiliary, however, they began to look at the incorrect image, and did not recover until well beyond the end of the sentence. In contrast, for active sentences in the manipulated audio condition it appeared that listeners looked to the image that is compatible with a passive sentence more frequently during the first noun, recovered during the auxiliary and verb stem, but then returned to the passive image prior to the verbal inflection. In the manipulated audio condition, we did not see an active bias when listeners responded incorrectly, which is somewhat surprising given that the overall data appeared to show an active bias. However, this may be due to the limited number of trials where listeners responded incorrectly.

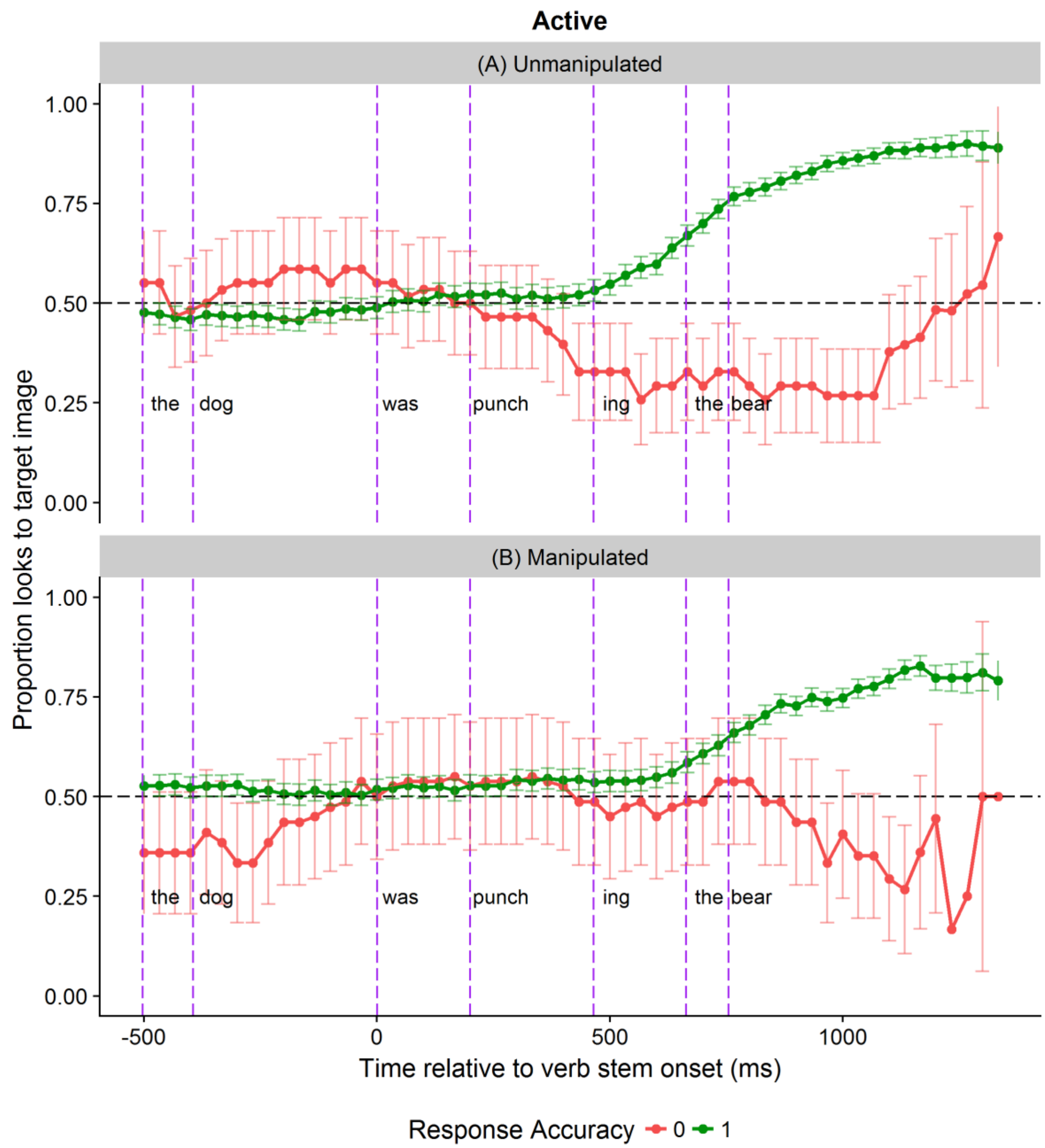


Figure 4.2.17. Experiment 4b: Proportion of looks to the target image for active sentences, separated by whether the response was correct (1) or incorrect (0). Rows correspond to the audio manipulation. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate average morpheme onsets. Morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

Interestingly, when listeners responded incorrectly to passive sentences, their gaze traces appeared similar regardless of whether they heard unmanipulated or manipulated audio (Figure 4.2.18). In these cases, listeners appeared to show a slight *passive* bias where they looked to the correct (passive) image more frequently during the first noun, switched to the active image during the auxiliary, returned to the passive image during the verb, and then looked instead to the active image for the remainder of the sentence. This may indicate some confusion on the part of the listener, perhaps in an attempt to resolve a conflict between a shallow parse and a more in-depth analysis of the sentence.

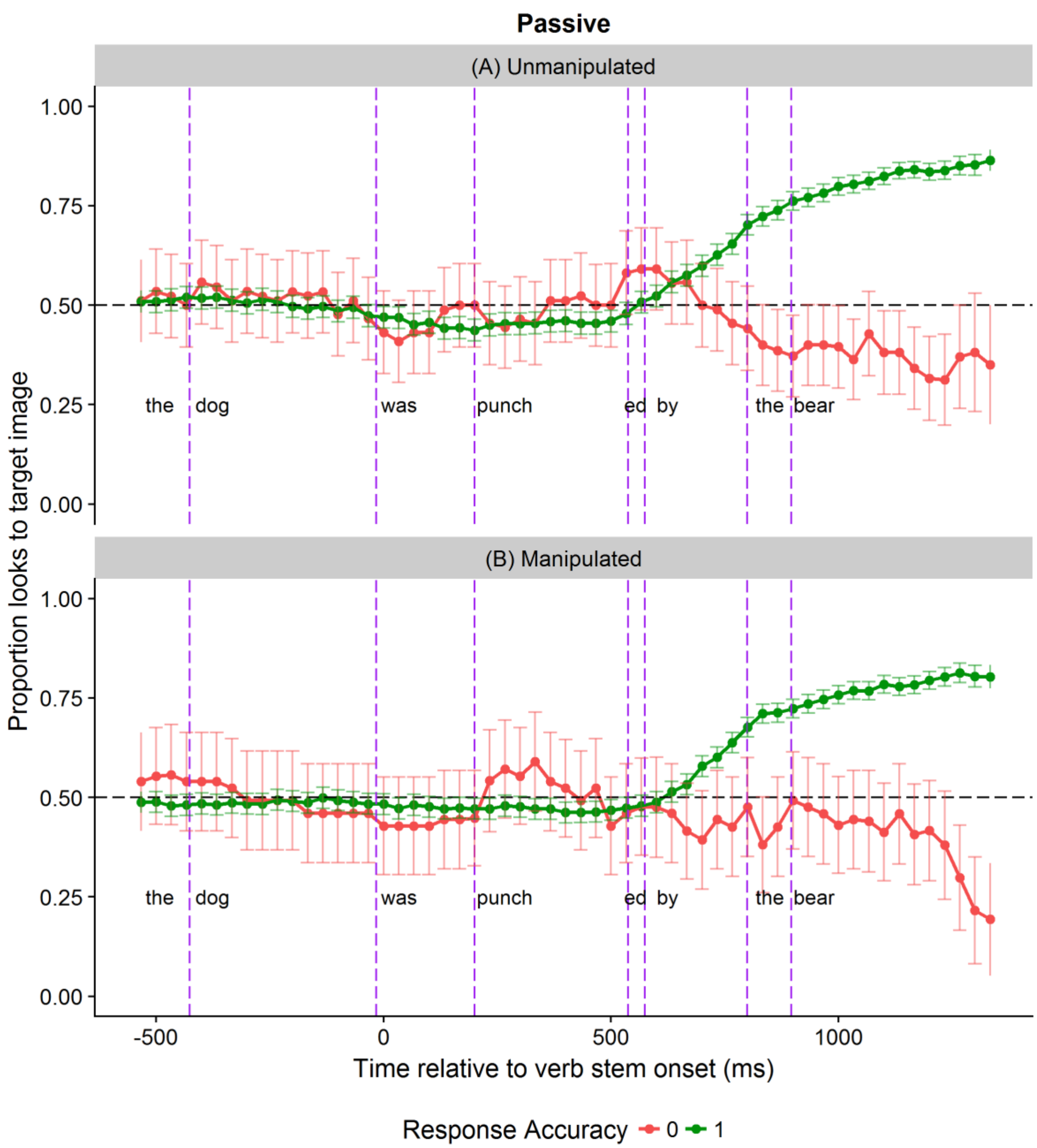


Figure 4.2.18. Experiment 4b: Proportion of looks to the target image for passive sentences, separated by whether the response was correct (1) or incorrect (0). Rows correspond to the audio manipulation. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate average morpheme onsets. Morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

In summary, for active and passives we observed some qualitative evidence of a two-stage processing model in the comparison between gaze traces for accurate and inaccurate responses. Next, we perform a saccade analysis to further investigate use of the NP1 = agent heuristic.

NP1 =agent saccade analysis. To determine whether listeners looked to the target or the distractor image in response to the first noun, an analysis of saccade behavior was performed. When listeners have heard only the beginning of the sentence and have very little information to inform parsing decisions, they tend to rely on simple processing biases, such as assuming that the first noun phrase encountered is the agent of the verb (Bever, 1970; Dowty, 1991). Because active sentences are more frequent, an active bias of this form (NP1 = agent strategy) can often be effective.

To assess the presence of a NP1 = agent processing strategy across syntactic constructions, the procedure used in the vowel duration analysis was repeated for a different analysis window to assess saccade behavior in response to the first noun. The analysis window began 200 ms after the onset of the first noun and ended 200 ms after the onset of the auxiliary *was*.

We expected to see more saccades to the target image for active sentences than for passives if listeners use the NP1 = agent strategy. Because, when visualized, the trends in looking behavior appear to differ depending on the audio manipulation condition, audio manipulation and syntax were included as fixed effects in the analysis. The Bayesian logistic mixed-effects model included syntax (active or passive), audio condition (manipulated or unmanipulated), and experimental block (1 or 2) as fixed effects. See Figure 4.2.19 for a visualization of the number of saccades made to either the target image or the distractor during the region of interest.

The overall effects of block, syntax, and audio manipulation were very small, and the effect estimates may have been near zero (all *Mean* β <= .05; see Table 4.2.5.). However, 80.28% of samples fell below 0 across the audio manipulation condition, suggesting that listeners who heard manipulated audio may have had a tendency to look at the distractor. In the comparison across blocks, 94.6% of samples fell below zero, which may suggest that listeners looked to the distractor more in block 2, which would be consistent with a change in strategy over blocks. Nevertheless, these effects were very small. Saccade behavior was marginally affected by the audio manipulation in block 1 only ($z = 1.72$, $p = .09$). Model comparisons between a model with interactions between all three fixed effects and models without these interactions revealed that interactions did not improve the ability of the model to predict the data (all $ELPD_{diff} > 0$). The results suggest that the effect of learning was very small during this region of the sentence. Based on this analysis, we cannot determine with confidence whether the verb stem vowel duration manipulation interfered with listeners' ability to learn the distribution of sentences in the experiment, and change their processing strategies accordingly.

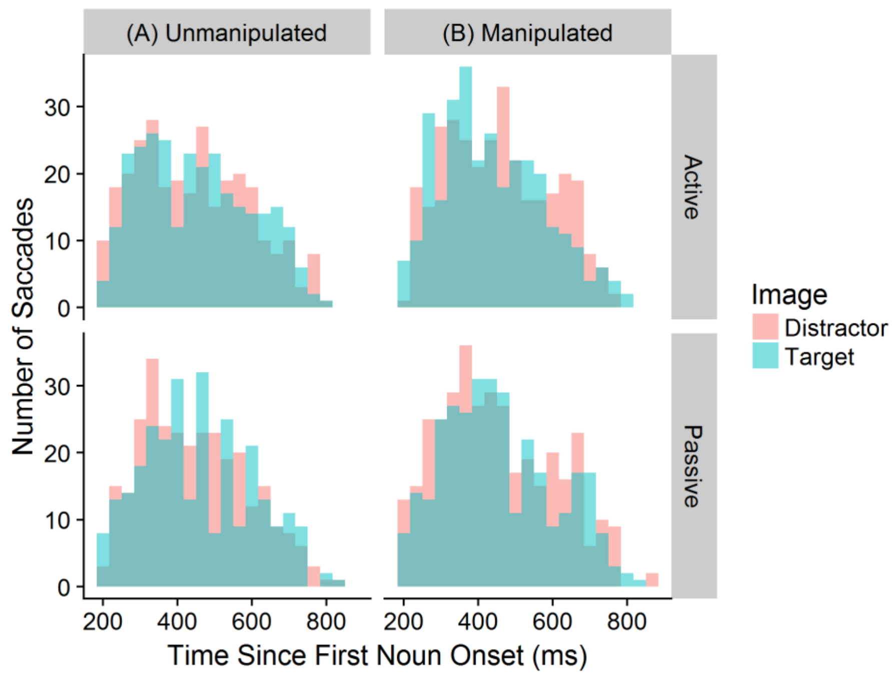


Figure 4.2.19. Experiment 4b: Histograms of saccades to the target (blue) and distractor (red) in response to the first noun. Darker blue regions indicate overlap between looks to the target and looks to the distractor. Rows correspond to syntax (active and passive) and columns indicate the audio manipulation (unmanipulated or manipulated). Bin size corresponds to the sampling frequency of the eye tracker (1 Hz, or 1 frame every 33.33 ms).

Table 4.2.5.

Experiment 4b: Bayesian Logistic Mixed-Effects Model Summary for Saccade Behavior by Audio Manipulation and Syntax

				Median	*MAD$_{SD}$*
Mean Posterior Predictive Distribution				0.5	0.1
Random Effects	*SD*	*r*			
Subject (*n* = 38)					
Intercept	0.071	—	—		
Passive Syntax	0.085	-0.20	—		
Block 2	0.090	-0.12	-0.13		

Verb (n = 14)						
Intercept	0.072	–	–	–		
Block 2	0.068	-0.10	–	–		
Passive Syntax	0.067	-0.29	-0.15	–		
Manipulated	0.073	0.00	-0.02	-0.21		
Audio						
Fixed Effects	*Median β*	*β MAD$_{SD}$*				
Intercept	0.10	0.10	–	–		
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Active vs. Passive	0.05	0.08	0.65	.52	-0.36	0.18
Audio Manipulation	0.03	0.09	0.34	.74	-0.40	0.12
Block 1 vs. 2	0.05	0.08	0.65	.52	-0.55	0.005
Block = 1						
Audio	0.19	0.11	1.72	.09	–	–
Block = 2						
Audio	-0.13	0.13	-1.03	.30	-0.11	0.65

Note: One divergent transition occurred during sampling, which indicates a possibility that the sampler was biased.

To summarize, gaze behavior revealed that incremental processing during the verb stem was not significantly impacted by the duration of the verb stem vowel, but was impacted by the syntax of the utterance. Additionally, we did not find reliable evidence from saccade behavior that listeners used the NP1 = agent strategy to process sentences.

4.2.3. Discussion

Syntax. Listeners responded more quickly to active sentences than passive sentences in both the manipulated and unmanipulated audio conditions. Qualitatively, gaze traces suggest that listeners looked to the target image more often during the verb stem for passives, but waited until the verbal inflection for actives. An analysis of saccade behavior during the verb stem revealed a reliable difference in gaze behavior across syntactic constructions. These results suggest that listeners predict whether the sentence will end as a passive or an active prior to hearing the verbal inflection.

Verb stem vowel duration: Incremental processing. Response time and gaze data revealed only small differences in processing that were not reliable as the result of the verb stem vowel duration manipulation. This suggests that verb stem vowel duration is not the primary cue listeners use to incrementally process active and passive sentences, and is not solely responsible for the behavior observed by Stromswold et al. (2002) and Stromswold et al. (2016). There are several possible explanations for this finding.

First, verb stem vowel duration may have simply been the acoustic tipping point in previous studies. Recall from Chapter 3 that the duration of the auxiliary was often longer in passives, and the intensity of the verb stem was higher in actives. In addition to verb stem vowel duration, listeners may have used these other acoustic cues, or other yet undiscovered acoustic cues, that are correlated with the syntax of the utterance. In this case, vowel duration was not the only acoustic cue to syntax in the morphosyntactically ambiguous region. Across speakers, there is some evidence of passive auxiliary lengthening (Experiments 1 and 2), greater average intensity on the verb stem in actives (Experiment 1), and reliable passive auxiliary lengthening was present in the stimuli used in this study (see 4.1.1). Changing any one of the acoustic cues present in this region may not be enough to lead the parser astray. This also raises the concern that, despite our best efforts, cue conflict may have been present within the sentences containing manipulated verb stem vowels. By the time the verb stem vowel was heard, listeners may have had enough evidence to ignore the conflicting cue and successfully continue to parse the sentence. It is important to note that incremental syntactic processing differs from incremental auditory processing more generally because listeners are simultaneously predicting upcoming segments *and* building the underlying representation of the sentence. Global cues to sentence structure may be prioritized over local acoustic cues, particularly those as subtle as the cue tested in this study (see Chapter 6 for more discussion on this point). If this is the case, and auxiliary duration serves as an additional cue, then future work could manipulate auxiliary vowel duration as well to determine whether listeners use a combination of auxiliary and verb stem vowel duration as cues to syntax.

There are several other potential explanations for our findings that pertain to our methods. It may be the case that the effect was present, but was simply too subtle to detect in our experiment, as suggested by the fact that differences in behavior across audio conditions was marginally significant in some comparisons, and that the direction—but not the magnitude—of the effect was consistent with manipulated vowel durations affecting processing. Use of a between-subjects design to test the audio manipulation was a limiting factor in this study. We chose to use a between-subjects design in order to prevent cue conflict across the audio manipulation condition. An alternative method would have been to run one condition, then ask the same subjects to return to the lab to complete the other after some period of time. However, this is problematic for two reasons: first, because the attrition rate would be high, and second, we have no way of knowing how long listeners would retain information from the first session, or what effect the retained information may have on the other condition.

In the current study, we "swapped" verb stem vowel duration across active and passive sentences to determine whether listeners use verb stem vowel duration to facilitate processing. Future work could apply the same vowel duration manipulation method in several ways to further test the role of duration cues in the processing of active and passive sentences. For example, the verb stem vowel could be manipulated so that it is longer than usual in passives, and shorter than usual in actives, in order to see whether accentuating the verb stem vowel duration difference facilitates processing. This method does face a limitation, however, in that shortening the active verb stem vowel may cause vowel devoicing, or may otherwise cause noticeable distortion of the vowel. Another possibility, mentioned previously, is that vowel duration in auxiliaries could also be swapped to test whether listeners use a combination of verb stem vowel duration and auxiliary vowel duration during processing. Finally, the acoustic duration cues could be neutralized in auxiliaries and verb stems so that vowels in these morphemes have equal duration across sentence types. This would clarify whether listeners experience any processing benefit from duration cues during the morphosyntactically ambiguous region.

Given the current evidence, we conclude that verb stem vowel duration does not strongly influence parsing decisions for these sentences. Future work should further investigate the acoustic cues that listeners may recruit to inform parsing decisions.

4.3. General Discussion

Unlike in the incremental lexical processing studies by Salverda et al. (2003; 2007), we did not find that listeners were able to use the duration of a syllable (the verb stem) to infer its prosodic context, and therefore the syntax of the utterance. However, like Stromswold et al. (2002; 2016), we did find evidence for incremental processing in predicting active or passive syntax, though prediction was not as early or as robust as expected.

Listeners generally use early, uncertain information when it is available in order to form predictions about upcoming sentence content (see Chapter 2). Even though the verb stem vowel occurred earlier than disambiguating morphosyntactic cues, listeners did not appear to rely heavily on its duration to form predictions about upcoming sentence content. There may be several reasons why listeners do not use verb stem vowel duration to form predictions. One possibility is that it simply does not occur early enough in the sentence relative to a definite morphosyntactic cue: the inflection on the verb. If prediction comes at a cost of any kind, it may not be worth the cost to use an uncertain cue that immediately precedes a definite cue.

Another possibility is that verb stem vowel duration is simply not reliable enough to cue progressive active or passive syntax consistently. All of our native English speakers lengthened verb stems in passive sentences when the passive verb was monosyllabic, and the progressive active verb was necessarily bisyllabic through the addition of the syllabic progressive active inflection. However, some monosyllabic verb roots take syllabic *-ed* inflections (e.g., *guided, patted*) and do not undergo a duration change. Similarly, many of the most frequent verbs in English are irregular verbs for which vowel duration would not reliably cue syntax (e.g., *eaten*). Furthermore, not all English verb stems are monosyllabic, and it is currently not known whether polysyllabic passive verb stems lengthen appreciably. In sum, the verb stem vowel duration cue may not to be sufficiently reliable to pass a threshold necessary to use in predictive processing. The implications of this possibility for theories of sentence processing are discussed in the concluding chapter (Chapter 6).

In this chapter, we showed that manipulating the duration of a verb stem vowel by adding or removing vowel periods near the midpoint of the vowel did not change how natural the vowel sounded overall. More importantly, we showed that listeners to not rely heavily on verb stem vowel duration to guide sentence processing, though it may be used in conjunction with other cues. Next, we turn to the processing and production of these same sentences by non-native speakers of English.

5. Sentence Comprehension and Production by L2 English Speakers

5.0. Motivation

In Chapter 4 we determined that native English speakers may use a combination of duration cues during processing. The findings of Salverda et al. (2003; 2007) and Stromswold et al. (2002) suggest that native English speakers may accumulate cues from duration to make inferences about where a word is located within a prosodic phrase, and how many syllables are in the word. If listeners successfully make these inferences, they must have knowledge of the way phonological processes (e.g., polysyllabic shortening, phrase-final lengthening) operate in their native language. It is possible that native English speakers have knowledge of the way phrase-final lengthening and polysyllabic shortening operate in English as a component of their linguistic competence, and that they may employ that knowledge when processing sentences on-line. If so, then there should be cross-linguistic differences in the production and use of syllable duration as a cue to syntax.

In the current chapter, we revisit the processing challenges involved in processing a second language. We specifically consider the challenges faced by native speakers of Mandarin who learned English as a second language.

Recall from section 2.3 that non-native speakers can form predictions about upcoming morphosyntax in their second language when they can generalize from a related cue in their native language (e.g., plural morphology on determiners as a syntactic category) even when they lack the specific cue tested (e.g., plural morphology on definite articles; Marull, 2017). However, if a predictive cue in the second language is entirely absent in a listener's native language, even proficient speakers are unable to use the cue to predict upcoming morphosyntax (Lew-Williams & Fernald, 2007; Lew-Williams & Fernald, 2010). This discrepancy may be explained by differences in the way listeners acquire a second language as opposed to a first language: when acquiring a first language, listeners gain distributional evidence from statistical regularities in the input, while second language acquisition occurs only through the lens of an existing native language (Lew-Williams & Fernald, 2010).

The distributional evidence explanation of second language acquisition and processing has implications for the current study. If native English speakers indeed use durational cues to syntax in order to facilitate processing of active and passive sentences, they would have learned the statistical regularities in the input that predict syntax when acquiring their first language. This suggests that L2 English speakers who lack some portion of the predictive chain in their native language (e.g., no durational cues to syntax, no inflectional morphology on verbs) might be unable to predict active or passive syntax based on acoustic evidence accumulated prior to the inflection on the verb, and might rely less on inflectional morphology than native English speakers do.

To test this hypothesis, native English speakers were compared with native speakers of another language who have acquired English as a second language. Mandarin Chinese differs from English in two key ways that are relevant to our hypothesis: 1) while Mandarin does have phrase-final lengthening, there is no evidence of polysyllabic shortening in Mandarin, and Mandarin may even have polysyllabic lengthening (Lai et al., 2010), and 2) Mandarin does not have inflectional morphology on verbs. Mandarin speakers were expected to produce no duration difference between progressive active and passive verb stems, and were not expected to use the duration difference produced by a native speaker to facilitate processing, due to lack of sensitivity to verb stem duration as an acoustic cue to syntax, and because they may be less sensitive to inflectional morphology on the verb. Processing by Mandarin speakers was then compared to processing by monolingual native English speakers.

5.1. Experiment 5a: L2 English sentence processing

The purpose of the current experiment was to determine whether L2 English speakers can make use of acoustic cues to syntax when processing active and passive sentences. Specifically, Mandarin speakers who learned English as a second language completed the same eye-tracking task as the manipulated vowel study, but only heard unmanipulated audio. Mandarin speakers also completed a language history questionnaire to assess their proficiency in English. Recall from Experiments 2 and 3 that polysyllabic shortening appeared to drive the progressive active-passive verb stem duration difference. If L2 speakers' linguistic knowledge of phrase-final lengthening and polysyllabic shortening must be like that of native English speakers to facilitate processing, then we expect that Mandarin speakers would incur a considerable processing delay because they lack polysyllabic shortening in their native language. Similarly, if linguistic knowledge of inflectional morphology facilitates processing, Mandarin speakers should require more processing time than English speakers.

As in Experiment 4b, response accuracy, response time, and eye-gaze were recorded and analyzed. Given the findings reported in the literature, we predicted that native Mandarin speakers would have higher response times than native English speakers. We expected lower accuracy for native Mandarin than for native English speakers, who should have less difficulty with comprehension. With respect to eye-gaze data, Mandarin speakers were expected to wait until after the verbal inflection to look at the correct image.

5.1.1. Methods

Participants. Sixteen adult self-reported native Mandarin speakers who learned English as a second language participated in the study. Data from 2 Mandarin speakers who moved to an English-speaking country before puberty (specifically: at 1 year of age and 7 years of age) were excluded from analysis. Mandarin speakers were compared to 14 of the native English speakers from the unmanipulated audio condition in Experiment 4a. Because Mandarin speakers only completed one experimental block, only data from block 1 of Experiment 4b were used. Native English speakers were matched against Mandarin speakers based on the same sentence order they received in block 1. Additionally, priority was given to native English speakers who showed very little track loss in the eye gaze data. For Mandarin speakers, native speaker status and English proficiency level were determined via a language history questionnaire (see Appendix E). In exchange for completing the study, participants received either course credit or monetary compensation. The study was approved by the Rutgers University Institutional Review Board, and was carried out in accordance with the Declaration of Helsinki.

Materials and Design. The materials and design of the comprehension study were identical to those in the control condition (unmanipulated audio) in Experiment 4a (see section 4.1.1.).

Apparatus. The apparatus was the same as that used in Experiment 4a (see section 4.1.1.).

Procedure. The procedure for this experiment was the same as the procedure used in Experiment 4a (see section 4.1.1), with a few exceptions. Listeners in this experiment only completed one block of the comprehension study. Additionally, L2 listeners went on to complete the language history questionnaire (see Appendix E) and a production study (Experiment 5b). The comprehension experiment required half an hour to complete.

Analysis. Response time and accuracy data were analyzed in the same manner as Experiment 4b (see section 4.2.2). Eye-gaze data also received the same treatment as in Experiment 4b (see section 4.2.2.3).

5.1.2. Results

Participant language background. Language background for 14 native Mandarin speakers whose data were analyzed were as follows. On average, native Mandarin speakers in the current study moved to an English-speaking country for the first time after puberty (*M* = 20.07 years old, *SE* = 0.93 years, range = 16-29 years). While, on average, participants did not attend a school where instruction was given in English until after puberty (*M* = 16.29 years of age, *SE* = 1.80 years, range = 3-29 years), 2 of the participants first attended such a school prior to the age of 10. Overall, Mandarin speakers first took an English class in prior to puberty (*M* = 8.07 years, *SE* = 0.80 years, range = 3-14 years). In addition, 12 of the 14 participants reported scores on the reading, writing, listening, and speaking sections of the Test of English as a Foreign Language™ (TOEFL®), where higher scores indicate higher English proficiency. Thirteen of the 14 participants rated their proficiency in reading, writing, listening, and speaking English on a 7-point scale (where 7 = fluent; see Table 5.1.1).

Table 5.1.1.

Experiment 5a: Descriptive Statistics for Mandarin Native Speaker Language History Questionnaire Responses

	Response	
Prompt	*M*	*SE*
Age when first moved to an English-speaking country	20.07	0.93
Age when first attended English-speaking school	16.29	1.80
Age when first took an English class	8.07	0.80
TOEFL (range = 0-30, *n* = 12)		
Reading	24.33	1.35
Writing	23.58	1.29
Speaking	23.08	0.66
Listening	25.75	0.68
Self-assessment (range = 1-7, *n* = 13)		
Reading	5.46	0.23
Writing	4.77	0.29
Speaking	5.38	0.34
Listening	5.31	0.28

Next, we assess comprehension by native and non-native English speakers by analyzing response accuracy.

5.1.2.1. Accuracy

We predicted that Mandarin speakers would be less accurate than native English speakers, due to the difference in English proficiency. However, this was not the case.

Mandarin speakers were 1.76% *more* accurate than English speakers (98.06% vs. 96.30%, respectively; see Figure 5.1.1). Accuracy was very high for both active (96.96%) and passive sentences (96.63%). A Bayesian logistic mixed-effects regression analysis was used to test whether native language (English or Mandarin) or syntax (active or passive) affected accuracy. Because native English speakers appeared to be less accurate on passive sentences, this model also tested whether there was an interaction between native language and syntax. The difference in accuracy across native language was too small to be reliable (*Mean* $\beta = -0.95$, $z = -1.23$, $p = .22$; Table 5.1.2), but note that 72.4% of samples for this comparison fell above zero, which may indicate a small effect of accuracy. In contrast, the effect of syntax on accuracy was so small as to be of no practical consequence, and the effect estimate may have been near zero (*Mean* $\beta = -0.19$, $z = -0.29$, $p = .77$). A model that included an interaction between native language and accuracy did not predict the data better than a model that did not include this interaction, but was otherwise equivalent, and this difference was slight ($ELPD_{diff} = 0.2$, $SE = 0.9$).

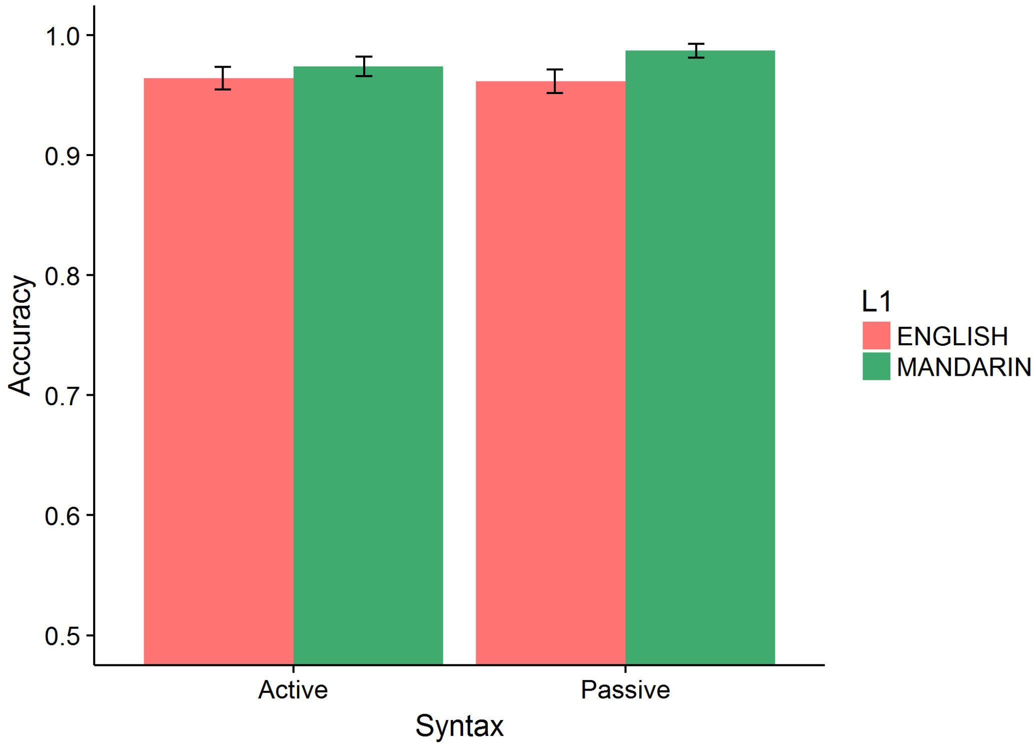


Figure 5.1.1. Experiment 5a: Accuracy for active or passive sentences, separated by native language. Error bars indicate standard error of the mean.

Table 5.1.2.

Experiment 5a: Bayesian Logistic Mixed-Effects Model Summary for Native Language and Accuracy

				Median	*MAD*$_{SD}$
Mean Posterior Predictive Distribution				1.0	0.0
Random Effects	*SD*	*r*			
Subject (*n* = 28)					
Intercept	1.71	–			
Passive Syntax	1.27	-0.43			
Verb (*n* = 14)					
Intercept	0.39	–	–		

Passive Syntax	0.73	-0.34	–			
L1: Mandarin	0.71	0.02	-0.52			
Fixed Effects	*Median β*	*β MAD_{SD}*				
Intercept	4.4	0.7				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
L1	-0.95	0.77	-1.23	.22	-0.94	1.93
Passive Syntax	-0.19	0.67	-0.29	.77	-1.44	0.92
L1: English						
Active vs. Passive	0.24	0.72	0.34	.74	–	–
L1: Mandarin						
Active vs. Passive	-0.63	0.90	-0.70	.49	-0.60	2.41

Note: One divergent transition occurred during sampling, which indicates a possibility that the sampler was biased.

5.1.2.2. Reaction time

Mandarin speakers: English proficiency. A linear regression was performed to determine which English proficiency measures best predicted response time for native Mandarin speakers. Specifically, we included response times for Mandarin speakers as the dependent variable, and tested the age when the speaker first moved to an English-speaking country, the age when the speaker first took an English class, and speakers' self-assessments of their ability to produce and understand spoken English.[11] This model accounted for 8% of the variance in response time (R^2 = 0.08, F(4,716) = 17.20, p <

[11] These measures were chosen because they were all highly correlated with response time. Because these measures were likely also intercorrelated, multiple regression analysis was used to determine which, if any, of these factors were independent predictors of response time.

.001). Only the age when the speaker first moved to an English-speaking country was an independent predictor of response time (β = 63.31, t = 7.39, p < .001), where older age predicted higher response times (Table 5.1.3).

Table 5.1.3.

Experiment 5a: Multiple Regression Analyses for Language Background Measures and Response Time

Fixed Effects	β	*SE*	*t*	*p*
Intercept	1122.59	249.86	4.49	< .001
Age when Moved to English-Speaking Country	63.31	8.57	7.39	< .001
Age when first took an English class	-12.52	9.99	-1.25	.21
Ability to speak English (self-assessment, scale: 0-5)	-35.47	40.50	-0.88	.38
Ability to understand spoken English (self-assessment, scale: 0-5)	57.75	51.35	1.13	.26

However, there was no difference between a linear mixed-effects model that *did* include the age when the participant first moved to an English-speaking country and one that *did not* include this variable (reported in Table 5.1.4; χ^2 = 3.26, p = .78), so it was not included as a factor in subsequent reaction time analyses.

As in Experiment 4b, response times were 125.18 ms higher on average for passive sentences (M = 2320.57 ms, SE = 22.65 ms) than for active sentences (M =

2195.39 ms, *SE* = 25.15 ms; see Figure 5.1.2.). As predicted, native English speakers processed sentences 430.53 ms more quickly (*M* = 2044.41 ms, *SE* = 19.68 ms) than native Mandarin speakers (*M* = 2474.84 ms, *SE* = 25.54; see Figure 5.1.2). Native Mandarin speakers were slower to respond to sentences, but were also more accurate in their responses, which could indicate a speed-accuracy tradeoff (see *Speed-accuracy tradeoff*). The difference in processing speed between passive sentences and active sentences was less pronounced for native Mandarin speakers than for native English speakers (Figure 5.1.2). On average, native English speakers responded 173.87 ms later when the sentence was passive, while Mandarin speakers responded 73.00 ms later on passive trials.

Given that native English speakers appeared to be slower to respond in Figure 5.1.2 to passives than actives, and Mandarin speakers did not, the magnitude of response time differences was assessed using a Bayesian linear mixed-effects model that included an interaction between native language and syntax. As in Experiment 4b, the processing delay for passive sentences was significant ($z = -3.42$, $p < .01$; see Table 5.1.4). There was also a main effect of native language ($z = -3.22$, $p < .01$). A model containing an interaction between native language and syntax predicted the data slightly better than an otherwise equivalent model with no interaction ($ELPD_{diff} = -0.7$, $SE = 1.7$). Further analysis revealed that the difference in response time across active and passive sentences was significant for native English speakers ($z = -3.88$, $p < .01$), but the difference was not reliable for native Mandarin speakers ($z = -1.45$, $p = .15$); however, 96.68% of samples fell below zero in the comparison between active and passive sentences for Mandarin speakers. This, combined with the size of the mean β estimate (-67.67) suggests that the effect may have been present, but was very small. This suggests that native Mandarin speakers incurred a reduced processing penalty for passive sentences in comparison to native English speakers.

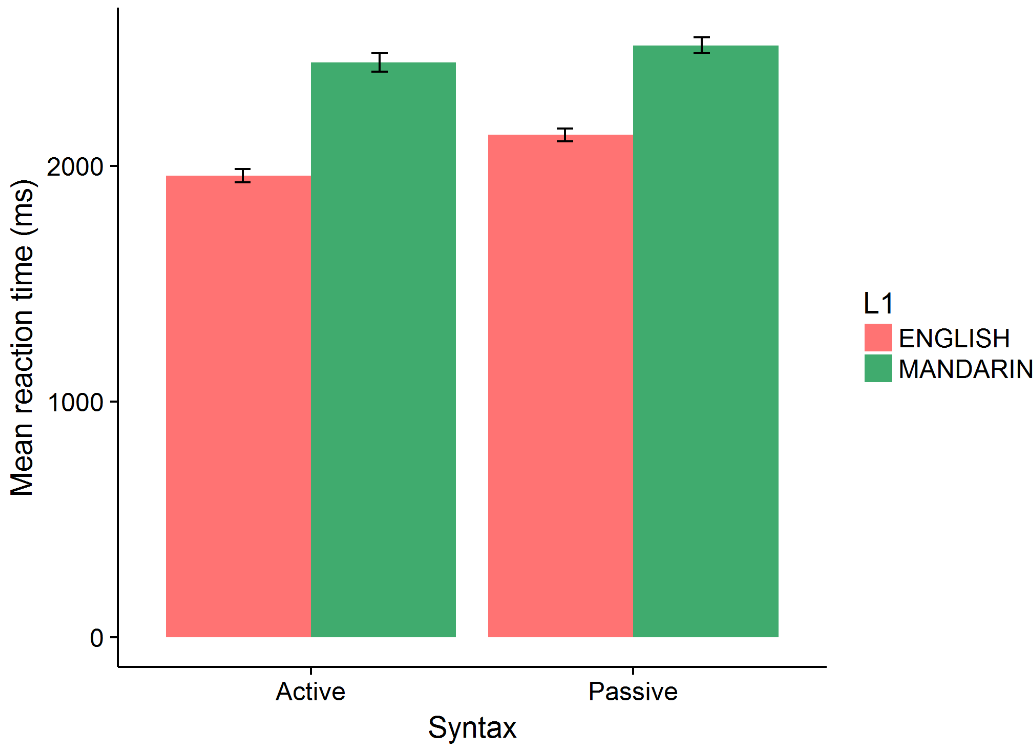


Figure 5.1.2. Experiment 5a: Mean reaction time for active and passive sentences, separated by native language. Error bars indicate standard error of the mean.

Table 5.1.4.

Experiment 5a: Bayesian Linear Mixed-Effects Model Summary for Syntax and Native Language on Response Time

			Median	*MAD*$_{SD}$
Mean Posterior Predictive Distribution			2258.2	18.9
Random Effects	*SD*	*r*		
Subject (*n* = 28)				

Intercept	379	–	–				
Passive Syntax	66	-0.42	–				
Verb (*n* = 14)							
Intercept	74	–	–				
L1: Mandarin	89	0.17	–				
Passive Syntax	73	-0.50	-0.02				
Residual	520	–	–				
Fixed Effects	*Median β*	*β MAD$_{SD}$*					
Intercept	1958.5	98.2					
Error *SD*	519.9	9.3					
					Posterior Interval		
	Mean β	*β SE*	*z*	*p*	5%	95%	
L1	-435.34	135.41	-3.22	< .01	251.45	720.00	
Syntax	-121.48	35.57	-3.42	< .01	100.54	248.10	
L1: English							
Active vs. Passive	-175.29	45.24	-3.88	< .01	–	–	
L1: Mandarin							
Active vs. Passive	-67.67	46.55	-1.45	.15	-202.66	-12.48	

Note: One divergent transition occurred during sampling, which indicates a possibility that the sampler was biased.

Speed-accuracy tradeoff. Listeners were 455.04 ms faster to respond when they were inaccurate (*M* = 1815.94 ms, *SE* = 98.80 ms, *n* = 44) than when they were accurate (*M* = 2270.98 ms, *SE* = 17.14 ms, *n* = 1511; see Figure 5.1.3). Similarly, listeners were 478.19 ms faster to respond incorrectly for active sentences (*M* = 1731.99 ms, *SE* =

137.98 ms) than when they responded correctly (*M* = 2210.18 ms, *SE* = 25.31 ms). For passives, listeners were 414.55 ms faster to respond incorrectly (*M* = 1916.67 ms, *SE* = 141.31 ms) than correctly (*M* = 2331.21 ms, *SE* = 22.83 ms). Native English speakers were 199.08 ms faster to respond incorrectly (*M* = 1852.70, *SE* = 114.03, *n* = 29) than correctly (*M* = 2051.78 ms, *SE* = 19.93 ms, *n* = 754). For Mandarin speakers, the size of this difference was more pronounced: Mandarin speakers were 744.45 ms faster to respond incorrectly (*M* = 1744.85 ms, *SE* = 192.66, *n* = 15) than when they responded correctly (*M* = 2489.30, *SE* = 25.50, *n* = 757).

To determine whether the above differences were reliable, a speed-accuracy tradeoff analysis was performed. This was done by testing a Bayesian linear mixed-effects model where reaction time was the dependent variable, and an interaction between response accuracy (0 or 1), syntax (active or passive), and native language (English or Mandarin) was included based on the apparent differences in Figures 5.1.3. and 5.1.4. Random slopes for response accuracy were included in both the by subject and by item random effects terms. The difference in response time between accurate and inaccurate responses was significant (*z* = -2.21, *p* = .03; see Table 5.1.5). The tradeoff was marginally significant for active sentences (*z* = -1.83, *p* = .07) but the effect was smaller for passive sentences (*z* = 1.50, *p* = .13); however, 85.25% of samples fell below zero for the comparison between accurate and inaccurate passive trials, which suggests that the effect may have been present, but was too small to be reliable. For native Mandarin speakers, the difference in response time across accurate and inaccurate trials was significant (*z* = -2.51, *p* = .01), but the effect was too small to be of practical significance for native English speakers (*Mean β* = -34.07, *z* = -0.30, *p* = .77). Model comparison was used to test for interactions between fixed effects. A model containing a 3-way interaction did not predict the data better than models containing 2-way interactions (all $ELPD_{diff} > 0$), but this model did predict the data slightly better than a model with no interaction terms ($ELPD_{diff} = -0.3$, $SE = 2.8$) suggesting that a small effect of the 3-way interaction may have been present.

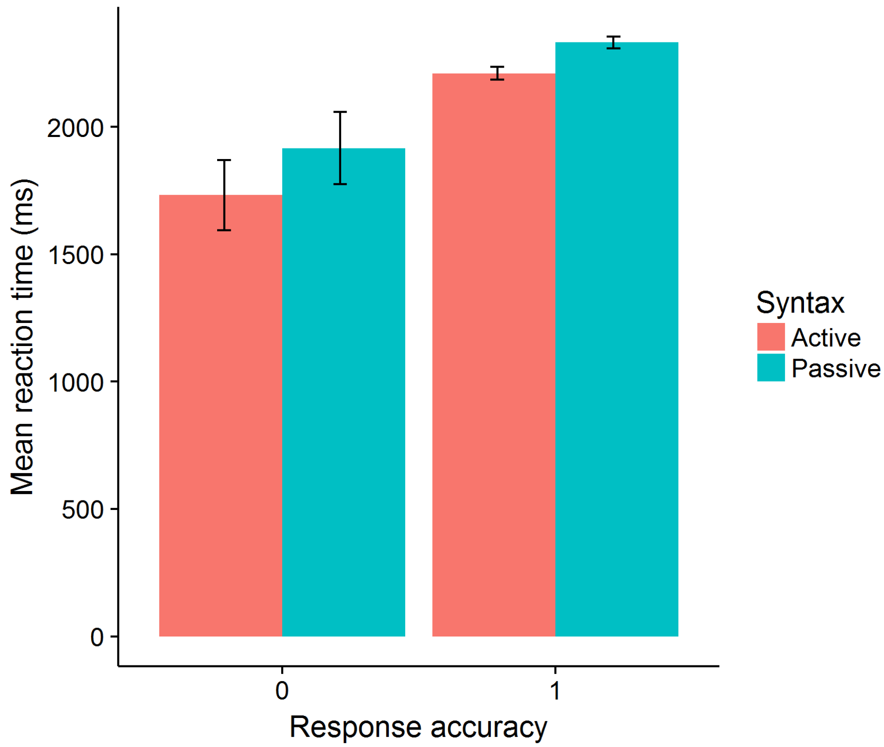


Figure 5.1.3. Experiment 5a: Mean response time by response accuracy for active and passive syntax. Error bars indicate standard error of the mean.

Table 5.1.5.

Experiment 5a: Bayesian Linear Mixed-Effects Model Summary for Speed-Accuracy Tradeoff Analysis by Syntax

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			2257.5	18.7
Random Effects	*SD*	*r*		

Subject (n = 28)				
Intercept	348	—	—	
Passive Syntax	64	-0.32	—	
Acc. = 1	141	-0.02	-0.01	
Verb (n = 14)				
Intercept	67	—	—	—
L1 = Mandarin	85	0.14	—	—
Passive Syntax	67	-0.30	0.02	—
Acc. = 1	60	-0.21	-0.05	-0.34
Residual	520	—	—	—

Fixed Effects	*Median β*	*β MAD*
Intercept	1822.1	174.8
Error *SD*	519.2	9.8

					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Accuracy (1 vs. 0)	-228.39	103.48	-2.21	.03	-119.09	396.81
Actives (1 vs. 0)	-230.89	126.36	-1.83	.07	—	—
Passives (1 vs. 0)	-225.89	150.73	-1.50	.13	-549.78	129.92
L1	-249.56	163.32	-1.53	.13	-129.66	742.16

L1: English (1 vs. 0)	-34.07	113.98	-0.30	.77	—	—
L1: Mandarin (1 vs. 0)	-422.71	168.34	-2.51	.01	-202.50	577.66

Note: One divergent transition occurred during sampling, which indicates a possibility that the sampler was biased.

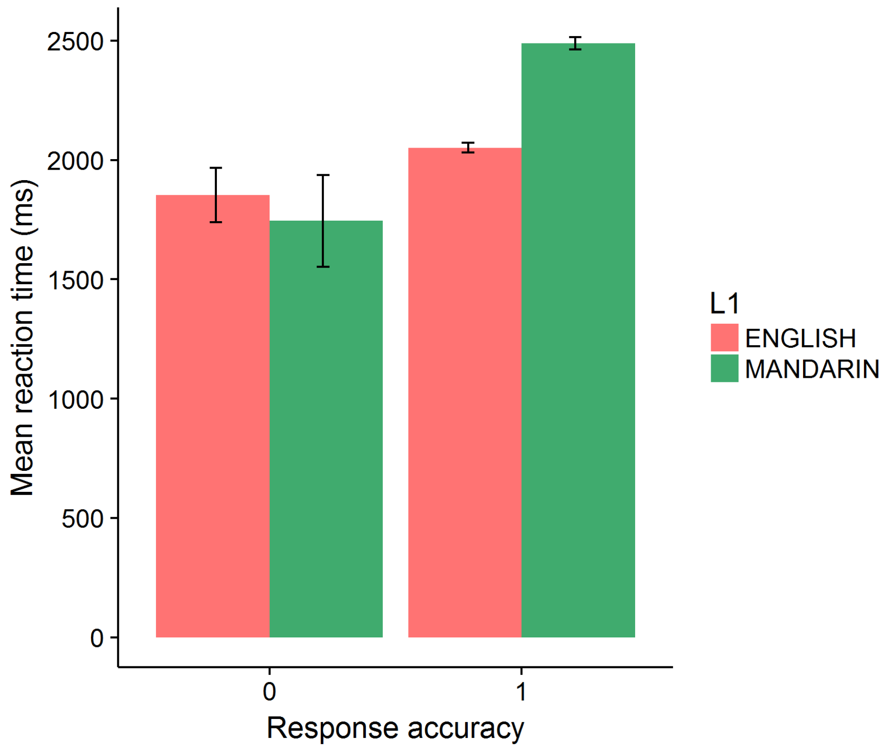


Figure 5.1.4. Experiment 5a: Mean response time by accuracy for native English and Mandarin speakers. Error bars indicate standard error of the mean.

Unlike in Experiment 4b, we did not find evidence of a speed-accuracy tradeoff for native English speakers in this experiment. This is surprising given that native English speakers who heard unmanipulated audio in Experiment 4b—the control for this study—showed a speed-accuracy tradeoff. The most likely explanation for this surprising finding is that in Experiment 4b, the native English speakers used as a comparison in this study were 2.33% more accurate in block 1 (96.30% accuracy) than in block 2 (93.97% accuracy), which means there were fewer inaccurate responses to work with in this analysis, where only the first block was used. Another possibility is that a select few of the native English-speaking subjects who were not included in this analysis were responsible for the difference in response times observed in Experiment 4b.

In sum, response times were higher for passive sentences, which reflects greater difficulty in the processing of passive sentences. Processing was slower for native Mandarin speakers, which likely reflects processing difficulty due to English proficiency, but may also arise due to more fundamental linguistic differences (see 5.1.3). A speed-accuracy tradeoff was present, and the tradeoff was larger for active sentences than passives, and for native Mandarin speakers more so than native English speakers. Next, we turn to the eye gaze data to evaluate processing at a fine-grained level.

5.1.2.3. Gaze behavior

Qualitative description. As in section 4.2.2 (experiment 4b), we present a qualitative description of eye-gaze behavior. In order to visualize gaze traces, we calculated the proportion of looks to the target image for each frame in a sentence. This was done separately for active sentences and passive sentences. Average morpheme onset markers are included in each visualization, and have been shifted 6 frames to the right in order to account for the time required to plan and execute a saccade in response to speech (6 frames * 33.33 ms = 199.98 ms). All visualizations also include a line marking 0.50 on the y-axis, reflecting chance level performance.

As in Experiment 4b, native English speakers began to look to the target image consistently toward the end of the verb stem (Figure 5.1.5 A), and looks to the correct target increased further after the verbal inflection, particularly for active sentences (Figure 5.1.5 B).

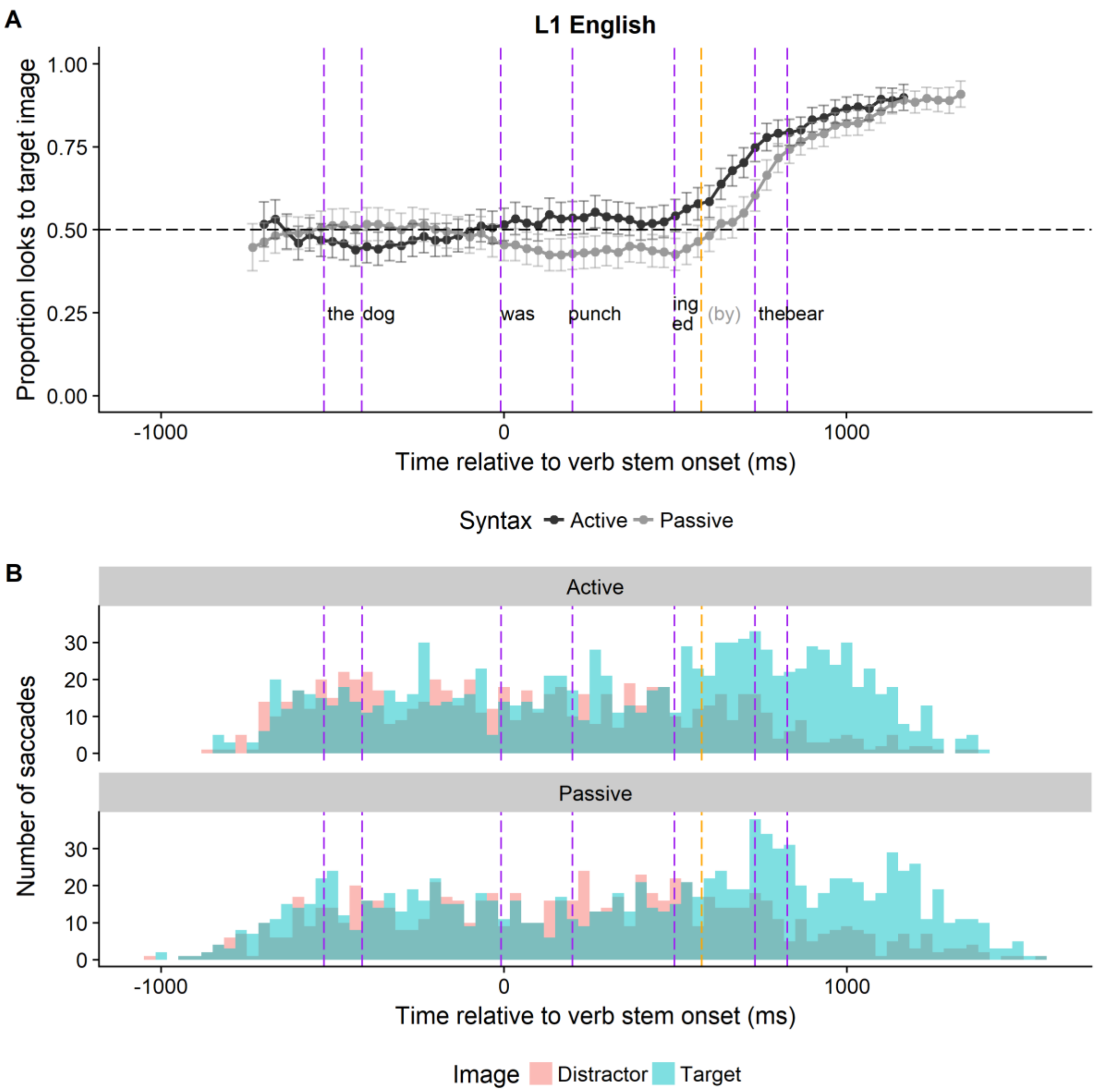


Figure 5.1.5. Experiment 5a: A) Proportion of looks to the target image for native English speakers hearing active and passive sentences. Error bars indicate 95% confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. B) Number of saccades to the target or the distractor image relative to the onset of the verb stem, shown separately for active and passive sentences. Dark blue reflects overlap between looks to the target and the distractor. For both A and B, average morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

Unlike native English speakers, native Mandarin speakers did not begin to look at the target image consistently until after the inflection in active sentences, and after the preposition in passive sentences (Figure 5.1.6 A). This delay is also reflected in saccades, where looks to the target image increased during the verbal inflection for actives, and during the preposition for passives (Figure 5.1.6 B).

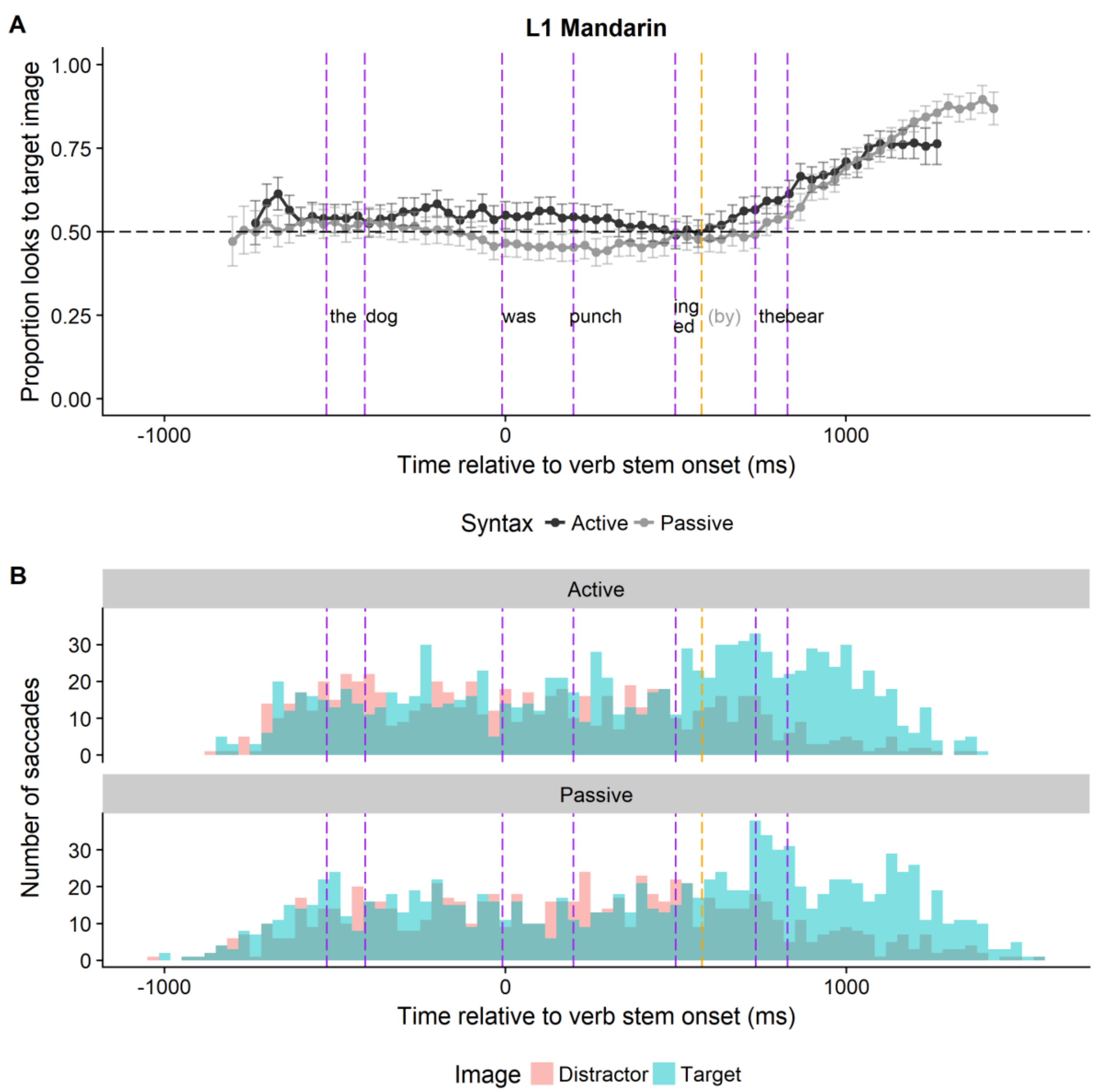


Figure 5.1.6. Experiment 5a: A) Proportion of looks to the target image for native Mandarin speakers hearing active and passive sentences. Error bars indicate 95%

confidence intervals. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. B) Number of saccades to the target or the distractor image relative to the onset of the verb stem, shown separately for active and passive sentences. Dark blue reflects overlap between looks to the target and the distractor. For both A and B, average morpheme onset markers are shifted to the right by 200 ms to account for the time required to plan and execute an eye movement.

The differences were more apparent when gaze traces for native Mandarin and native English speakers were compared directly. As predicted, Mandarin speakers waited until after the inflection on the verb before looking consistently at the target image. For active sentences, native English speakers appeared to look to the target image more frequently within the verbal inflection, while Mandarin speakers appeared to wait until the end of the verbal inflection to look to the correct image (Figure 5.1.7). Furthermore, Mandarin speakers appeared to rely more on an active bias, or the first NP = agent processing strategy, as indicated by looking to the active image more frequently during the first noun.

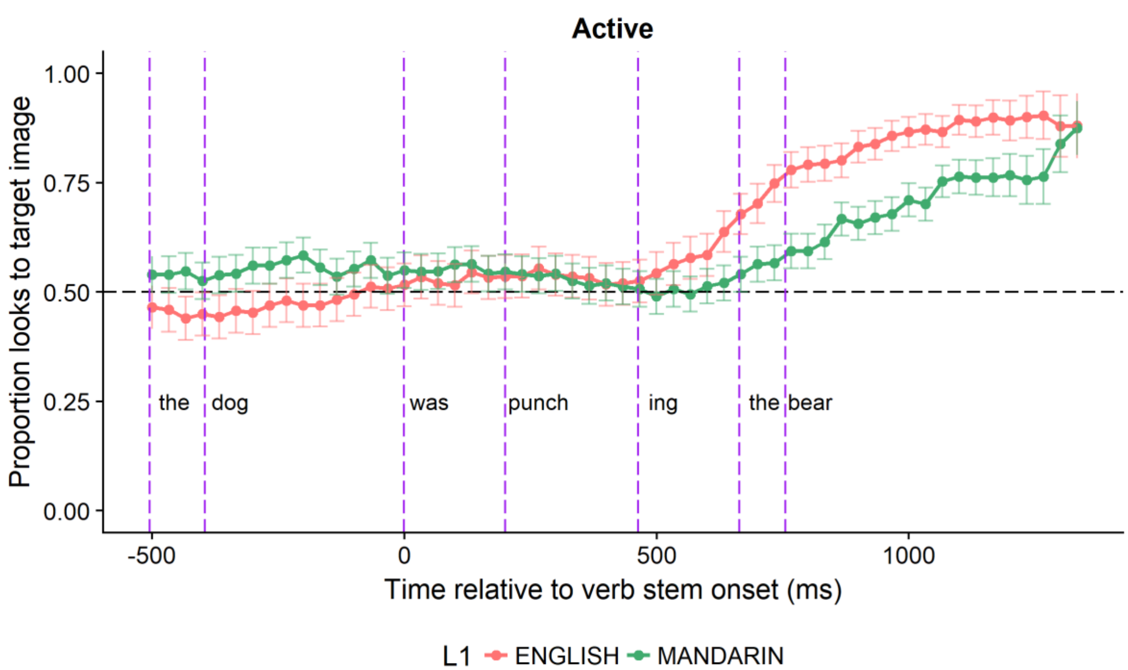


Figure 5.1.7. Experiment 5a: Proportion of looks to the target image for active sentences, separated by native language. The black horizontal line indicates chance performance (.50), and vertical lines indicate average morpheme onsets. Morpheme onset markers are shifted to the right by 6 frames to account for the time required to plan and execute an eye movement.

For passive sentences, an inflection point in the gaze trace occurred within the verb stem for native English speakers, after which they looked more frequently at the target image; Mandarin speakers did not look more frequently at the target image until the end of the preposition *by* (Figure 5.1.8).

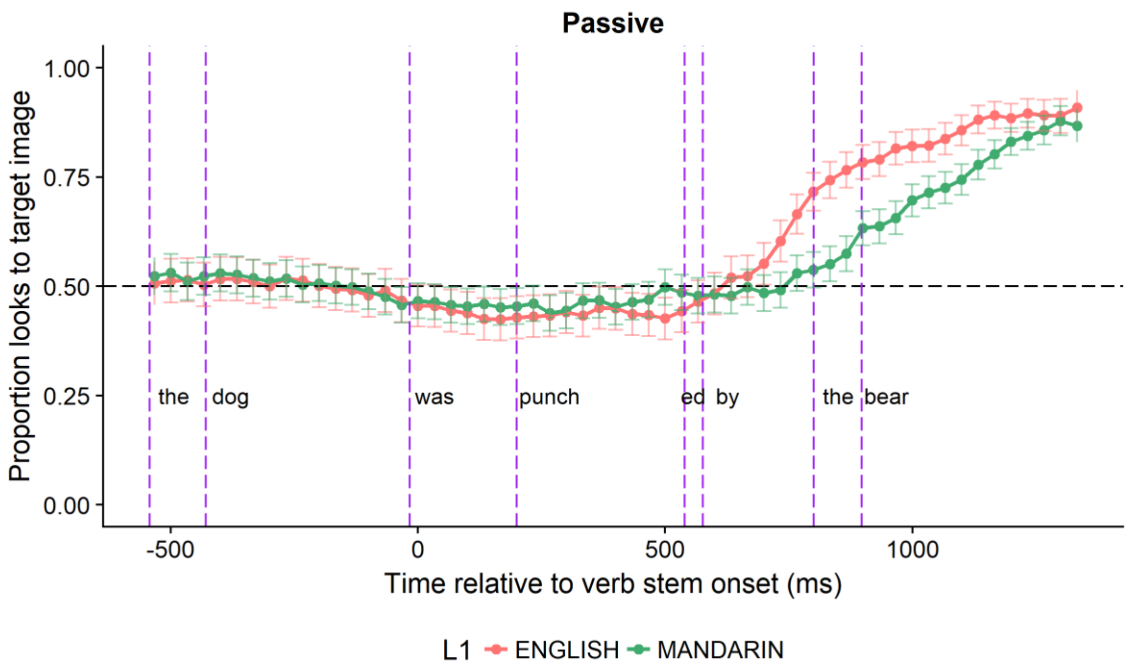


Figure 5.1.8. Experiment 5a: Proportion of looks to the target image for passive sentences, separated by native language. The black horizontal line indicates chance performance (.50), and vertical lines indicate morpheme onsets. Morpheme onset markers are shifted to the right by 6 frames to account for the time required to plan and execute an eye movement.

Analysis. To determine whether listeners looked to the target or the distractor image in response to acoustic cues through the verb stem vowel, an analysis of saccade behavior was performed (see section 4.2.2 for a full description of the analysis preparation). Saccades that occurred after the offset of the verb stem vowel but prior to the onset of the verbal inflection were analyzed. The analysis window was adjusted to account for the time required to plan and execute a saccade in response to speech, approximately 200 ms (Salverda et al., 2014). Saccades within the analysis window were then analyzed using a Bayesian logistic mixed-effects model. To control for variation due to individual differences or due to the items tested, random slopes and random intercepts were included for the syntactic frame (active or passive) and native language (English or Mandarin), by subjects and by item (verb), except where otherwise noted. An interaction term for syntax and native language was included to test for apparent differences across syntactic construction in gaze traces (figures 5.1.7 and 5.1.8). See Figure 5.1.9 for a visualization of the number of saccades made to either the target image or the distractor during the region of interest.

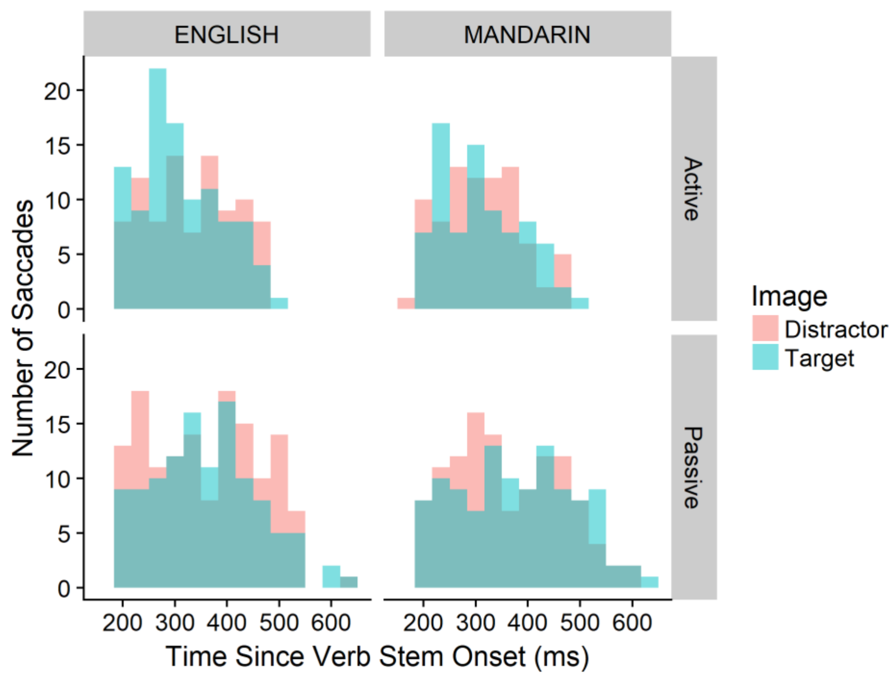


Figure 5.1.9. Experiment 5a: Histograms of saccades to the target (blue) and distractor (red) in response to the verb stem. Darker blue regions indicate overlap between looks to the target and looks to the distractor. Rows correspond syntax (active and passive) and columns indicate the audio manipulation (unmanipulated or manipulated). Bin size

corresponds to the sampling frequency of the eye tracker (1 Hz, or 1 frame every 33.33 ms).

Unlike in Experiment 4b, the overall influence of syntax on saccade behavior was too small to be reliable (*Mean* β = 0.18, z = 1.27, p = .21; see Table 5.1.6); however, 95.1% of samples fell below zero, which suggests that the effect may have been present, but very small. The overall effect of native language on saccade behavior was so small as to have no practical significance (*Mean* β = 0.02, z = 0.13, p = 0.89). Like in Experiment 4b, syntax had a marginal effect on saccade behavior for native English speakers (*Mean* β = 0.33, z = 1.69, p = .09), but for native Mandarin speakers the effect of syntax was too small to be of practical consequence (*Mean* β = 0.03, z = 0.15, p = .88). A model that included an interaction between native language and syntax did not predict the data better than a model without the interaction ($ELPD_{diff}$ = 0.4, *SE* = 1.0), though this difference was slight.

Table 5.1.6.

Experiment 5a: Bayesian Logistic Mixed-Effects Model Summary for Syntax and Native Language on Saccade Behavior

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			0.5	0.0
Random Effects	*SD*	*r*		
Subject (*n* = 28)				
Intercept	0.10	–		
Passive Syntax	0.14	-0.29		
Verb (*n* = 14)				

Intercept	0.08	—	—			
Passive Syntax	0.09	-0.18	—			
Mandarin	0.10	-0.10	-0.09			
Fixed Effects	*Median β*	*β MAD$_{SD}$*				
Intercept	0.1	0.1				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Syntax	0.18	0.14	1.27	.21	-0.66	-0.001
L1	0.02	0.15	0.13	.89	-0.53	0.20
L1 English Syntax	0.33	0.20	1.69	.09	—	—
L1 Mandarin Syntax	0.03	0.21	0.15	.88	-0.19	0.78

To summarize, saccade behavior during the verb stem was marginally impacted by the syntax of the utterance for native English speakers, but the effect was much smaller—and may have been near zero—for native Mandarin speakers. Next, we evaluated early sentence processing biases by analyzing saccades within the region of the sentence occupied by the first noun.

To determine whether Mandarin speakers relied more strongly on an active bias early in the sentence, as suggested by gaze traces (shown in Figures 5.1.7 and 5.1.8), saccades to the target and distractor during the first noun were compared. The region of interest began 200 ms after the onset of the first noun, and ended 200 ms after the start of the auxiliary. Saccades between the target and distractor within this window should only reflect responses to the first noun (see Figure 5.1.10 for histograms of saccades in this region).

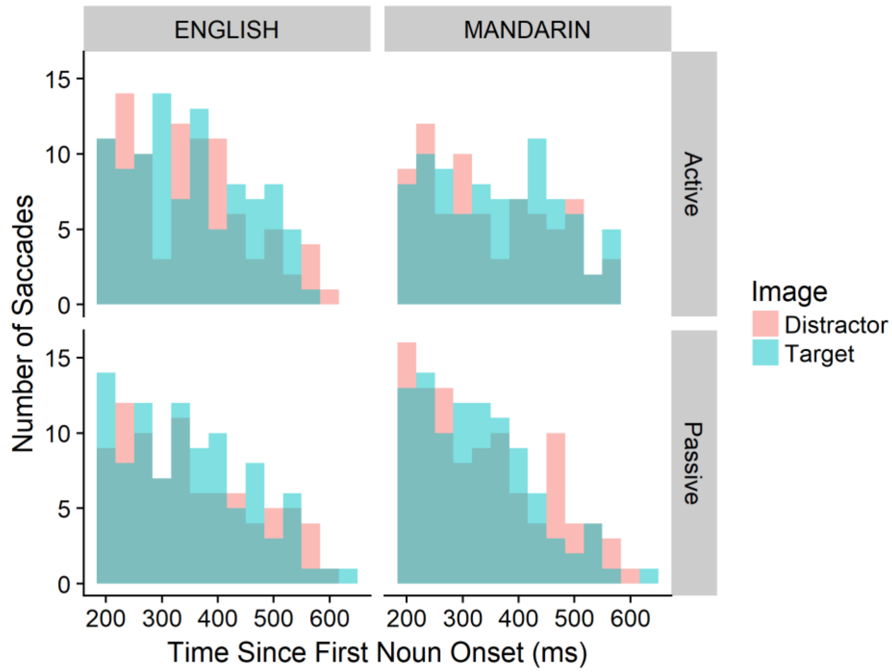


Figure 5.1.10. Experiment 5a: Histograms of saccades to the target (blue) and distractor (red) in response to the first noun. Darker blue regions indicate overlap between looks to the target and looks to the distractor. Rows correspond to syntax (active and passive) and columns indicate the audio manipulation (unmanipulated or manipulated). Bin size corresponds to the sampling frequency of the eye tracker (1 Hz, or 1 frame every 33.33 ms).

During this region, the effect of native language on saccade behavior was so small as to be of no practical consequence (*Mean* β = -.03, z = -0.24, p = .81). The effect of the syntax of the sentence was similarly small (*Mean* β = -.03, z = -0.27, p = .79), contrary to the predictions of an NP1 = agent processing strategy, where an active bias would predict more looks to the distractor image in passive sentences, and more looks to the target image in active sentences (see Table 5.1.7). A model with an interaction between native language and syntax did not predict the data better than a model that was otherwise the same, but did not include an interaction ($ELPD_{diff}$ = 1.0, SE = 0.6).

Table 5.1.7.

Experiment 5a: Logistic Mixed-Effects Model Summary for Syntax and Native Language and Saccade Behavior

					Median	*MAD*$_{SD}$
Mean Posterior Predictive Distribution					0.5	0.0
Random Effects	*SD*	*r*				
Subject (*n* = 28)						
Intercept	0.11	–				
Passive Syntax	0.17	-0.27				
Verb (*n* = 14)						
Intercept	0.09	–	–			
Passive Syntax	0.10	-0.16	–			
Mandarin	0.11	-0.16	-0.06			
Fixed Effects	*Median* *β*	*β* *MAD*$_{SD}$				
Intercept	0.0	0.1				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
L1: English vs. Mandarin	-0.03	0.13	-0.24	.81	-0.17	0.40
Syntax: Active vs. Passive	-0.03	0.13	-0.27	.79	-0.18	0.40

Note: One divergent transition occurred during sampling, which indicates a possibility that the sampler was biased.

This suggests that native Mandarin speakers did not have a stronger active bias than native English speakers early in the sentence, perhaps because native English speakers also showed some active bias. Next, we interpret these findings in greater detail.

5.1.3. Discussion

Consistent with Experiment 4b, passive sentences were more difficult to process than active sentences, reflected in higher processing time for passives. Although native English speakers' response times suggested that they had less processing difficulty overall, the syntax of the utterance impacted native English speakers' processing speed to a greater extent than it did for native Mandarin speakers, as evidenced by a small interaction between syntax and native language for response times. The syntax of the sentence affected saccade behavior (marginally) for native English speakers, but the effect was too small to have practical significance for native Mandarin speakers. This suggests that native Mandarin speakers did not form expectations about the syntax of the utterance prior to hearing the verbal inflection.

Native Mandarin speakers who learned English after puberty were more accurate than native English speakers overall, but also required more time to process sentences. The presence of a speed-accuracy tradeoff for native Mandarin speakers only suggests that Mandarin speakers slowed down to achieve greater comprehension accuracy. However, native English speakers did not show a similar tradeoff, despite the presence of a speed-accuracy tradeoff in Experiment 4b. Another possible explanation is that Mandarin speakers were more uncertain about processing English sentences, due either to their English proficiency or due to differences between Mandarin and English that prevent predictive processing, and waited for deterministic cues in order to respond. This would result in greater accuracy, because unambiguous morphosyntactic information is available to them, and would also result in slower response times.

Next, we turn to a production study that was conducted in order to test whether differences between the way phonological processes operate in Mandarin and in English may be implicated in the processing delay found in this study.

5.2. Experiment 5b: L2 English production

The previous study revealed that native Mandarin speakers experienced a processing delay relative to native English speakers when processing simple active and passive sentences. Experiment 5a was unable to disentangle whether the processing delay was due to an inability to use acoustic cues to syntax or to limited English proficiency. As a first step to determining whether proficiency alone is implicated in the processing delay, the current experiment investigates whether native Mandarin speakers produce the duration cues to syntax that native English speakers produced in Chapter 3. Mandarin has phrase-final lengthening, but may not have polysyllabic shortening (Lai et al., 2010), and lacks inflectional morphology for verbs.

In the current experiment, L2 English speakers who speak Mandarin as a native language said active (2a) and passive (2b) sentences from Experiment 1 (Appendix A), which were then segmented into morphemes. The verb stem vowel was also segmented. The duration of the auxiliary, verb stem, and the verb stem vowel in active and passive sentences were compared. We predicted that Mandarin speakers may not produce longer verb stems in passive sentences, because Mandarin does not demonstrate evidence of polysyllabic shortening, and may have polysyllabic lengthening. It is possible that Mandarin speakers may lengthen verb stems in progressive active contexts, in which case we may have found a duration difference in the opposite direction. Mandarin speakers may have dropped inflections more often than English speakers, because inflectional morphology is lacking in their native language. Furthermore, Mandarin speakers were predicted to drop -*ed* inflections frequently because 1) they are low in salience, and 2) the -*ed* inflection results in a complex coda, which does not occur in Mandarin (Hansen, 2001).

5.2.1. Methods

Participants. The same 14 native Mandarin speakers who completed Experiment 5a participated in this study. Productions from 5 of the 14 native Mandarin speakers were analyzed. Of these participants, 3 were female and 2 were male. The study was approved by the Rutgers University Institutional Review Board, and was carried out in accordance with the Declaration of Helsinki.

Materials and Design. Materials and design for this experiment were identical to those in Experiment 1. Note that the same participants heard a subset of these sentences (56 of the 64 sentences, excluding verbs *pat* and *tickle*) during the comprehension study.

Apparatus. Participants sat in a quiet room while wearing a Plantronics Audio995H-02 USB headset. The experiment was carried out using a Lenovo ThinkPad Yoga 12 20DL running Windows 8.1, with a resolution of 1920 x 1080. A keyboard mask was used to occlude all keys except *Q*, *P*, and the spacebar. Stimulus presentation and response measures were controlled by E-Prime 2.0 Professional experiment software. Productions were recorded in stereo using Audacity at a sampling rate of 48 kHz.

Procedure. Participants completed the experiment after doing the language history questionnaire and comprehension study (Experiment 5a). The experimental procedure was otherwise identical to that of Experiment 1 (section 3.1.1).

Segmentation. Boundaries marking the onset and offset of each morpheme in the sentence were placed by hand using Praat (Boersma, 2001). In addition, the onset, vowel, and coda of each verb stem were also segmented in Praat. Segmentation was carried out by two coders per the same criteria used to segment morphemes in Experiment 2, and to segment vowels in Experiment 3. See Figure 5.2.1 for an example segmented sentence.

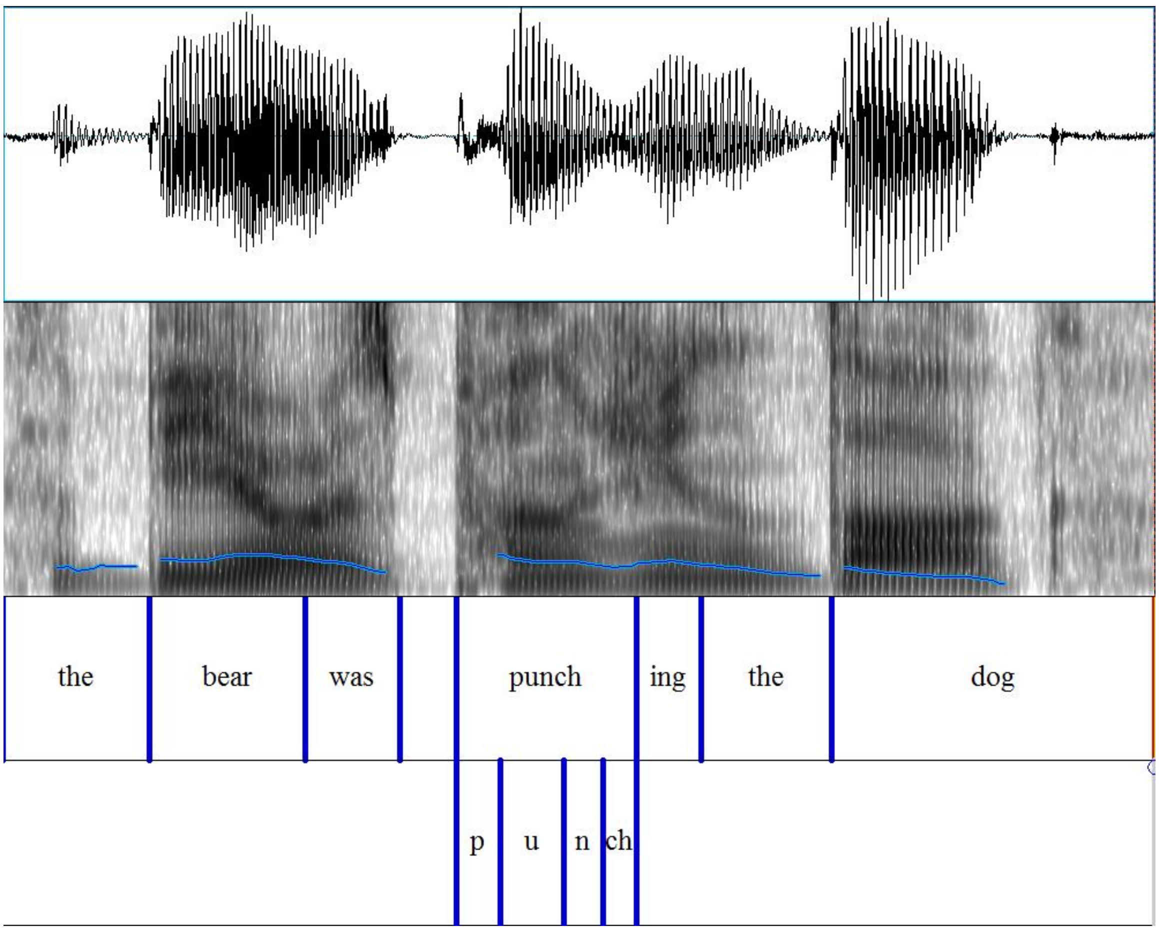


Figure 5.2.1. Experiment 5b: Example of segment boundary placements for the sentence *The bear was punching the dog.* Boundaries (shown in blue) mark the onset and offset of each segment per the segmentation criteria outlined in section 5.2.1.

Analysis. Eight sentences containing the verbs *pat* or *tickle* were excluded from analysis, and the remaining 56 target sentences were analyzed. Filler sentences were not analyzed. As a caveat, due to the prevalence of disfluencies in the data, sentences containing disfluencies were analyzed. For each interval, spanning either a morpheme or a segment, the duration, mean intensity, and mean pitch were calculated. The effect of syntactic frame on verb stem duration and on vowel duration was then assessed via Bayesian linear mixed-effects models. To control for variation due to individual differences or due to the items tested, random slopes and random intercepts were included for the syntactic frame (progressive active or passive) by subjects and by item (verb), unless otherwise noted. See section 3.1.1 for a more detailed explanation of the type of analyses used here.

5.2.2. Results

Errors. Due to limited English proficiency, speech errors were common in our data (Table 5.2.1).

Table 5.2.1.

Experiment 5b: Measures of English Proficiency and Error Rates by Error Type for Native Mandarin Speakers

	Participant				
	1	2	3	4	5
Sex	Female	Male	Female	Male	Female
Language History	1	2	3	4	5
Age when first moved to an English-speaking country (years)	18	18	22	29	16

Age when first attended a school where the language of instruction was English (years)	18	10	22	29	8
Age when first took and English class (years)	6	10	10	10	8
Error Type	1	2	3	4	5
Missing Verbal Inflection					
Active	0.00%	3.85%	0.00%	0.00%	0.00%
Passive	44.44%	40.00%	0.00%	0.00%	71.43%
Focused Verb					
Active	0.00%	0.00%	0.00%	0.00%	0.00%
Passive	0.00%	4.00%	0.00%	0.00%	0.00%
Epenthesis in Verb					
Active	0.00%	0.00%	0.00%	0.00%	0.00%
Passive	0.00%	20.83%	0.00%	0.00%	0.00%
Other					
Active	18.52%	69.23%	16.00%	0.00%	0.00%
Passive	11.11%	70.83%	15.38%	0.00%	0.00%

Unsurprisingly, Mandarin speakers frequently dropped inflectional morphology in passives (31.73% drop rate), where the inflection formed a complex coda. Quite surprisingly, one speaker even dropped the inflection in 1 progressive active sentence; however, this speaker (participant 2) had a very high error rate relative to other speakers. The same speaker epenthesized syllables in passive verbs: for example, [k^hɪkt] became [k^hɪk.ɪd]. Other errors, such as reducing consonant clusters outside of the verbal inflection (e.g., complex onsets, consonant clusters in nouns), epenthesizing vowels outside of verbs to avoid complex codas (e.g., [kʰæt] → [kʰæ.tə]), and errors in pronunciation (incorrect vowel, e.g., [ʃʌv] → [ʃov]) were common.

Duration. *Auxiliary.* Although Mandarin speakers produced 7.27 ms longer auxiliaries in passives (*M* = 224.05 ms, *SE* = 4.56 ms) relative to the actives (*M* = 216.58 ms, *SE* = 4.94 ms; see Figure 5.2.2), the effect was too small to be of practical consequence, and 99.83% of samples fell within 1 JND of zero (*Mean* β = -7.11, *z* = -0.77, *p* = .44; Table 5.2.2).

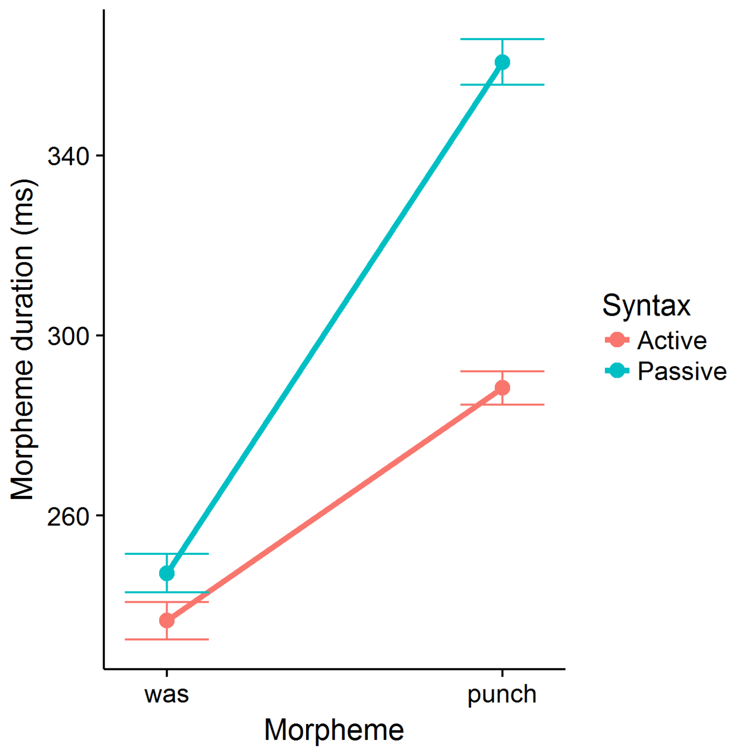


Figure 5.2.2. Experiment 5b: Mean duration for auxiliaries and verb stems produced by native Mandarin speakers in active and passive sentences. Error bars indicate standard error of the mean.

Table 5.2.2.

Experiment 5b: Bayesian Linear Mixed-Effects Model Summary for Syntax and Auxiliary Duration

				Median		*MAD_{SD}*
Mean Posterior Predictive Distribution				218.8		3.8
Random Effects	*SD*	*r*				
Subject (n = 5)						
Intercept	38.1	–				
Passive Syntax	15.8	-0.16				
Verb (n = 14)						
Intercept	20.1	–				
Passive Syntax	9.4	0.06				
Residual	43.9	–				
Fixed Effects	*Median β*	*β MAD_{SD}*				
Intercept	214.9	15.7				
Error *SD*	43.7	2.0				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Passive vs. Active Syntax	-7.11	9.24	-0.77	.44	-7.20	21.72

This finding is not consistent with the reliable auxiliary lengthening produced by native speakers (Chapter 3). Indeed, none of the native Mandarin speakers produced reliable auxiliary lengthening, and one speaker (participant 4) produced longer auxiliaries in *active* sentences (see Figure 5.2.3).

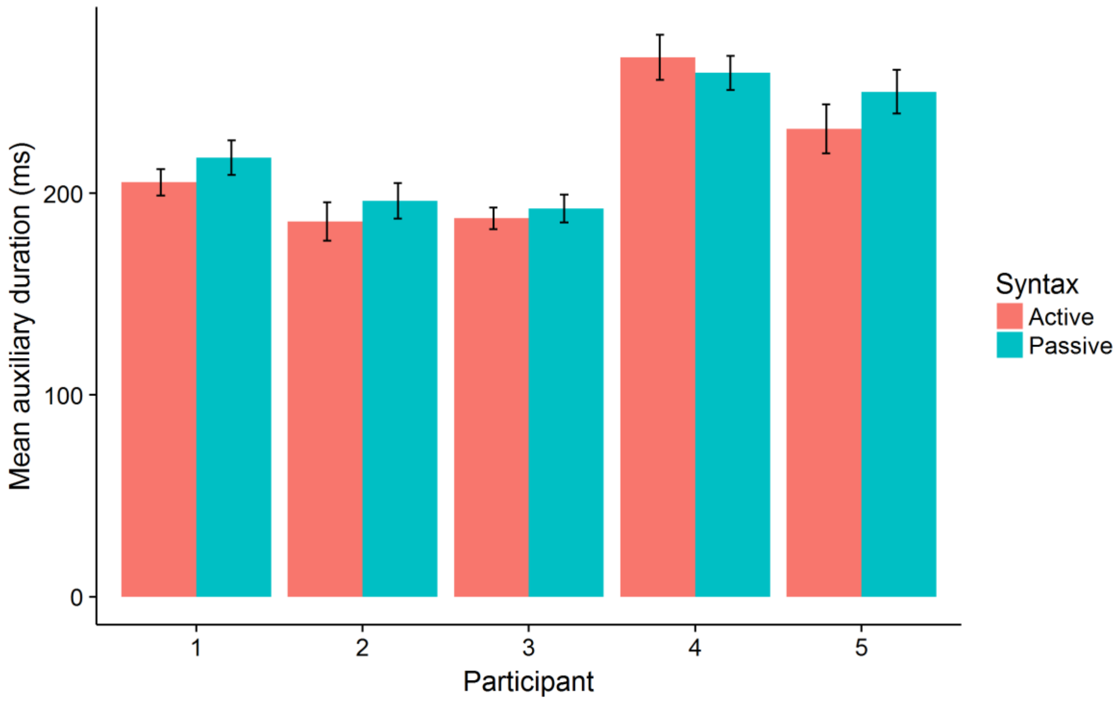


Figure 5.2.3. Experiment 5b: Mean duration for auxiliaries produced by native Mandarin speakers, shown for each speaker. Error bars indicate standard error of the mean.

To analyze performance by individuals in greater detail, another model was constructed where subject was included as a fixed effect rather than a random effect (Table 5.2.3). Because we expected differences in duration across constructions, this model included an interaction between subject and syntax.

Table 5.2.3.

Experiment 5b: Bayesian Linear Mixed-Effects Model Summary for Syntax and Subject on Auxiliary Duration

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			218.7	3.9
Random Effects	*SD*	*r*		
Verb (n = 14)				

Intercept	19.6	–				
Passive Syntax	9.3	0.06				
Residual	44.1	–				
Fixed Effects	*Median β*	*β MAD$_{SD}$*				
Intercept	205.6	9.8				
Error *SD*	44.0	2.0				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Active vs. Passive Syntax	-6.90	6.01	-1.15	.25	-8.00	31.23
Subject 1: Active vs. Passive	-11.54	11.96	-0.96	.33	–	–
Subject 2: Active vs. Passive	-9.42	12.71	-0.74	.46	-29.72	26.62
Subject 3: Active vs. Passive	-5.38	12.63	-0.43	.67	-33.21	21.67
Subject 4: Active vs. Passive	8.55	12.70	0.67	.51	-47.23	7.88
Subject 5: Active vs. Passive	-16.72	12.22	-1.37	.17	-21.39	33.02

For all participants, the difference in auxiliary duration was too small to be of practical consequence, and 91.88% of samples fell within1 JND of zero.

To determine whether native Mandarin speakers who learned English as a second language produce the passive verb stem lengthening observed in Chapter 3, we analyzed the duration of the verb stem and the verb stem vowel.

Verb stem. On average, Mandarin speakers did demonstrate passive verb stem lengthening. Passive verb stems were 60.29 ms longer on average (M = 348.63 ms, SE = 6.84 ms) than active verb stems (M = 288.34 ms, SE = 5.19 ms; see Figure 5.2.2). Consistent with the results presented in Chapter 3, passive verb stem lengthening was significant (*Mean* β = -62.39, z = -4.59, p < .01; Table 5.2.4).

Table 5.2.4.

Experiment 5b: Bayesian Linear Mixed-Effects Model Summary for Syntax and Verb Stem Duration

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			318.8	4.4
Random Effects	*SD*	*r*		
Subject (n = 5)				
Intercept	30	–		
Passive Syntax	24	0.25		
Verb (n = 14)				
Intercept	45	–		
Passive Syntax	22	-0.02		
Residual	50	–		
Fixed Effects	*Median* β	β MAD_{SD}		
Intercept	288.2	17.5		

	Mean β	β SE	z	p	Posterior Interval 5%	95%
Error *SD*	50.2	2.4				
Active vs. Passive Syntax	-62.39	13.60	-4.59	< .01	41.11	83.72

Passive verb stem lengthening was produced by all speakers (Figure 5.2.4).

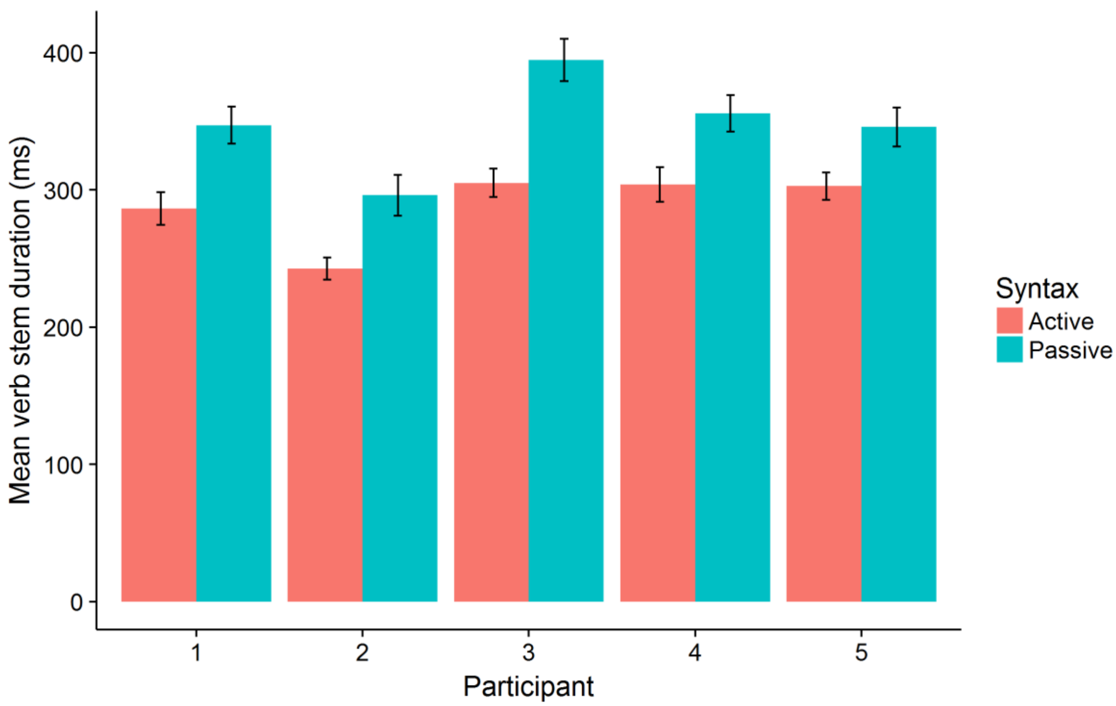


Figure 5.2.4. Experiment 5b: Mean duration for verb stems produced by native Mandarin speakers, shown for each speaker. Error bars indicate standard error of the mean.

A model that included a random effect structure by subject predicted the data better than an otherwise equivalent model without this term ($ELPD_{diff}$ = -26.2, *SE* = 8.3), which indicates the presence of individual differences. To test for passive verb stem lengthening for each speaker, we once again tested a model that included subject and syntax as an interaction, and did not contain a random effect by subject. Consistent with the results reported in Chapter 3 for native English speakers, all native Mandarin speakers produced passive verb stem lengthening (*ps* < .01; see Table 5.2.5).

Table 5.2.5.

Experiment 5b: Bayesian Linear Mixed-Effects Model Summary for Syntax and Subject on Verb Stem Duration

				Median	*MAD_{SD}*	
Mean Posterior Predictive Distribution				318.9	4.4	
Random Effects	*SD*	*r*				
Verb (n = 14)						
Intercept	44	–				
Passive Syntax	22	0.00				
Residual	50	–				
Fixed Effects	*Median β*	*β MAD_{SD}*				
Intercept	284.9	15.0				
Error *SD*	50.1	2.5				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Active vs. Passive Syntax	-63.15	8.65	-7.3	< .01	37.79	85.79
Subject 1: Active vs. Passive	-61.77	14.70	-4.20	< .01	–	–

Subject 2: Active vs. Passive	-51.41	15.50	-3.32	< .01	-42.87	21.03
Subject 3: Active vs. Passive	-97.04	15.24	-6.37	< .01	2.49	66.09
Subject 4: Active vs. Passive	-58.61	15.80	-3.71	< .01	-34.85	29.11
Subject 5: Active vs. Passive	-46.91	14.86	-3.16	< .01	-46.87	16.87

This shows that passive verb stem lengthening is robust, even for non-native English speakers.

Verb stem duration: Missing -ed inflection. Given that our speakers frequently dropped the *-ed* inflection, we performed an analysis to determine whether verb stems were longer when the inflection was omitted. Because this only applies to passive sentences, active sentences were excluded from the analysis. Passive verb stems were 3.95 ms longer when the inflection was present (*M* = 349.87 ms, *SE* = 8.36 ms) than when it was absent (*M* = 345.92 ms, *SE* = 12.00), and this difference was so small as to have no practical significance (*Mean β* = -3.60, *z* = -0.12, *p* = .90; Table 5.2.6).

Table 5.2.6.

Experiment 5b: Bayesian Linear Mixed-Effects Model Summary for Dropped -ed Inflection on Verb Stem Duration

	Median	MAD_{SD}
Mean Posterior Predictive Distribution	349.8	7.5

Random Effects	*SD*	*r*				
Subject (n = 5)						
Intercept	47	–				
-ed Dropped	45	-0.23				
Verb (n = 14)						
Intercept	45	–				
-ed Dropped	35	0.03				
Residual	59	–				
Fixed Effects	*Median β*	*β MAD_{SD}*				
Intercept	346.6	22.6				
Error *SD*	59.0	4.2				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
-ed Dropped vs. Present	-3.60	29.03	-0.12	.90	-41.21	48.00

This indicates that effect of deleting the *-ed* inflection on verb stem duration was too small to have practical consequence.

Vowel duration. Mandarin speakers, on average, produced 24.71 ms longer vowels in passive verb stems (*M* = 111.94 ms, *SE* = 3.50 ms) than active verbs stems (*M* = 87.23 ms *SE* = 3.50 ms; Figure 5.2.5).

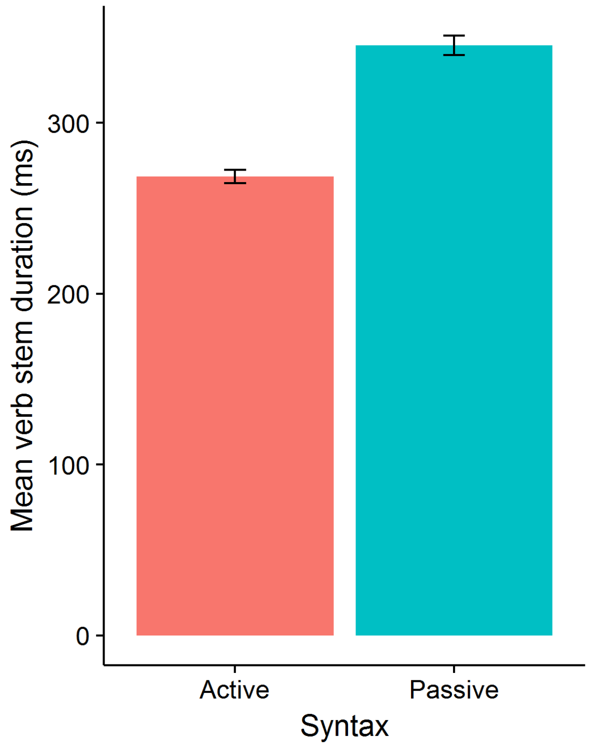


Figure 5.2.5. Experiment 5b: Mean duration for verb stem vowels produced by native Mandarin speakers. Error bars indicate standard error of the mean.

The difference in vowel duration between passive and active verb stem vowels was significant ($z = 2.84$, $p < .01$; Table 5.2.7).

Table 5.2.7.

Experiment 5b: Bayesian Linear Mixed-Effects Model Summary for Syntax and Verb Stem Vowel Duration

			Median	MAD_{SD}
Mean Posterior Predictive Distribution			100.3	2.3
Random Effects	*SD*	*r*		
Subject (n = 5)				
Intercept	11.6	–		
Passive Syntax	17.1	0.38		

Verb (n = 14)						
Intercept	22.4	–				
Passive Syntax	8.3	0.15				
Residual	25.4	–				
Fixed Effects	*Median β*	*β MAD$_{SD}$*				
Intercept	87.7	7.5				
Error *SD*	25.4	1.2				
					Posterior Interval	
	Mean β	*β SE*	*z*	*p*	5%	95%
Active vs. Passive Syntax	-24.57	8.66	-2.84	< .01	10.69	38.45

Once again, verb stem vowel duration across syntactic constructions varied for different speakers (Figure 5.2.6).

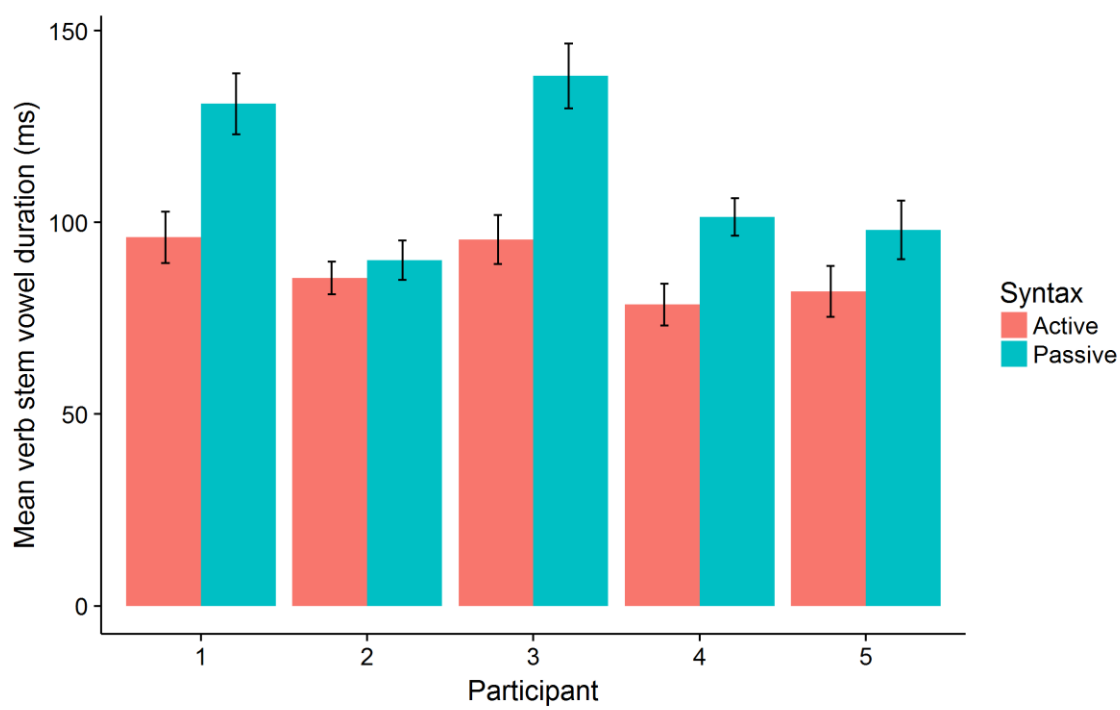


Figure 5.2.6. Experiment 5b: Mean duration for verb stem vowels produced by native Mandarin speakers, shown for each speaker. Error bars indicate standard error of the mean.

To determine whether a change in verb stem vowel duration drove passive verb stem lengthening for all speakers, we once again performed an analysis to compare duration differences across syntactic constructions for each individual. (Table 5.2.8)

Table 5.2.8.

Experiment 5b: Bayesian Linear Mixed-Effects Model Summary for Syntax and Subject on Verb Stem Vowel Duration

				Median	MAD_{SD}	
Mean Posterior Predictive Distribution				100.3	2.2	
Random Effects	*SD*	*r*				
Verb (n = 14)						
Intercept	22.6	–				
Passive Syntax	8.3	0.12				
Residual	25.5	–				
Fixed Effects	*Median* β	β MAD_{SD}				
Intercept	95.5	7.7				
Error *SD*	25.5	1.2				
					Posterior Interval	
	Mean β	β *SE*	*z*	*p*	5%	95%
Active vs. Passive Syntax	-24.37	3.80	-6.42	< .01	22.03	46.50

Subject 1: Active vs. Passive	-34.57	7.38	-4.68	< .01	–	–
Subject 2: Active vs. Passive	-5.22	7.46	-0.70	.48	-45.75	-12.69
Subject 3: Active vs. Passive	-43.65	7.32	-5.96	< .01	-7.36	25.31
Subject 4: Active vs. Passive	-22.56	7.43	-3.04	< .01	-28.22	4.07
Subject 5: Active vs. Passive	-15.86	7.17	-2.21	.03	-34.41	-2.79

For 4 of the speakers, vowels were reliably longer in passive verb stems (*ps* < .05), while for one speaker the effect of syntax on verb stem vowel duration was too small to be reliable (*Mean β* = -5.22, *z* = -0.70, *p* = .48); however, 99.83% of samples fell below 0 in the comparison between active and passive verb stem vowels for this speaker, which suggests that the effect may have been present, but was very small. Note, however, that this speaker is the same speaker who produced the most errors overall.

In sum, native Mandarin speakers who learned English as a second language lengthened passive verb stems and passive verb stem vowels relative to their active counterparts, but did not lengthen passive auxiliaries. This demonstrates that non-native English speakers who have a very different L1 phonology are able to produce some, but not all, of the acoustic cues to syntax that native English speakers produce. The implications of this will be discussed in the next section.

5.2.3. Discussion

The results of this study indicate that L2 English speakers produce some of the same duration cues to passive syntax as native English speakers do. All native Mandarin speakers in this study lengthened passive verb stems, and all but one speaker lengthened passive verb stem vowels to a reliable degree. What remains unclear is whether native Mandarin speakers achieve the progressive active-passive verb stem duration difference through the process of phrase-final lengthening alone, or whether polysyllabic shortening also plays a role, even though it is not present in their native language. It is also unclear why passive auxiliary lengthening did not occur. Future work along this vein should carry out a production study like Experiments 2 and 3 in order to determine which processes are implicated in duration for non-native speakers.

It should also be noted that we reported results for only 5 native Mandarin speakers, when we have data from a total of 13 speakers to work with.[12] It is possible that the trends we found for the 5 speakers reported here differ from those of the whole group. Further analysis is required to determine whether native Mandarin speakers consistently produce verb stem lengthening, and *only* verb stem lengthening, in passive English sentences.

Next, we discuss the implications of the production findings, combined with the comprehension findings from Experiment 5a, for sentence processing in a second language.

5.3. General Discussion

In experiment 5a, native Mandarin speakers were slower to process sentences overall than native English speakers. Experiment 5b revealed that 4 of the 5 native

---

[12] This figure excludes the 2 subjects who moved to an English-speaking country before puberty, and 1 subject with incomplete data for the production experiment.

Mandarin speakers produced 2 of the 3 duration cues to passive syntax that native English speakers produced in Chapter 3 (lengthened auxiliaries, lengthened verb stems, and lengthened verb stem vowels), and 1 only produced reliable passive verb stem lengthening. It is currently unclear whether the processing delay experienced by Mandarin speakers is due to non-native English proficiency, or to fundamental differences between the phonology and morphosyntax of Mandarin and English that prevented native Mandarin speakers from exploiting cues to syntactic structure in processing.

If it is not the case that non-native English proficiency alone is responsible for processing differences across language groups, then one possible explanation for Mandarin speakers' processing difficulty is that Mandarin speakers did not acquire the predictive relationship between acoustic cues that accumulate as the sentences unfolds, and the morphosyntax predicted by those cues. Mandarin does have phrase-final lengthening in the language, and native speakers of Mandarin seem to be able to produce some duration cues to English syntax in their utterances. Why, then, do native Mandarin speakers seem to be unable to use the acoustic correlates of active and passive syntax to predict its syntax? If the lack of passive auxiliary lengthening in production is mirrored by an insensitivity to auxiliary lengthening in comprehension, then this may prevent native Mandarin speakers from accumulating sufficient duration cues to facilitate processing for English sentences.

It is also possible that Mandarin has duration cues to syntax in some constructions, and that Mandarin speakers would be able to exploit those cues to facilitate processing of Mandarin sentences. However, Mandarin lacks inflectional morphology on verbs to indicate tense. Recall that Lew-Williams & Fernald (2010) found evidence that listeners can only use cues to predict upcoming morphosyntax if they have learned the statistical regularities between those cues and the sentence content they predict. For English speakers, it is possible that longer auxiliaries and longer verb stems predict the passive participle, and passive syntax by extension. Given that Mandarin lacks inflectional tense morphology on verbs, native Mandarin speakers necessarily lack a chain in this predictive sequence. Native Mandarin speakers would not have learned the statistical relationship between morpheme duration and syntax in order to make use of morpheme duration as a cue to syntax during processing, which would prevent them from using the cue despite their ability to produce it.

In order to investigate this possibility further, processing by native Mandarin speakers must be compared to that of L2 English speakers who speak a different native language, such as Spanish. Spanish has inflectional morphology for verbs and phrase-final lengthening (Vaissière, 1983), but has less polysyllabic shortening than English does (Nam et al., 2008). If native Spanish speakers do not incur the same processing delay shown by Mandarin speakers, then we can conclude that differences in native language drive the processing difficulties faced by Mandarin speakers. Data collection is underway to examine this possibility.

6. Conclusions

The dissertation makes three main contributions to the sentence processing literature. First, the dissertation demonstrates that acoustic cues may be used in a cumulative fashion in order to form predictions about upcoming sentence content, but that disrupting one such cue does not dramatically hinder processing.

Second, the dissertation generates some evidence that listeners update their processing strategies over time in light of information gained from recent experience. Specifically, listeners can learn that a distribution of sentences differs from their initial expectations based on the distribution of sentences in the language as a whole.

Third, the dissertation makes a potential contribution to a growing body of evidence which shows that predictive processing in a second language fails when prediction is conditioned on relationships that are absent in the speaker's native language.

The first goal of the dissertation was to determine what acoustic cues to syntax native English speakers produce during morphosyntactically ambiguous regions of temporarily ambiguous sentences, and to identify the phonological processes implicated in the aforementioned acoustic cues. The second goal was to determine which, if any, of these acoustic cues listeners use to facilitate the processing of temporarily ambiguous sentences. The third goal was to investigate whether L2 English speakers could produce the same cues that native English speakers produce, and whether they could use those cues to help them process English sentences.

6.1. Chapter 3 Summary: Sentence Production by Native English Speakers

Stromswold et al. (2002) recorded progressive active and passive sentences said by a single native English speaker as part of the stimuli for a comprehension study, and found that the speaker produced longer verb stems in passive sentences. The verb stem segment that changed in duration was the verb stem vowel. The authors proposed that phrase-final lengthening and polysyllabic shortening may have contributed to the duration cue (see Chapter 3). In Chapter 3, we presented 3 experiments that were designed to confirm the acoustic cues to progressive active and passive syntax that native English speakers consistently produce during the morphosyntactically ambiguous regions of temporarily ambiguous sentences.

Consistent with Stromswold et al. (2002), across all 3 experiments, speakers consistently produced longer verb stems in passive sentences in comparison to progressive active sentences. In Experiments 1 and 2, participants also lengthened passive auxiliaries, though with less consistency, and Experiment 1 showed evidence of greater intensity in progressive active verb stems. These findings indicate that the duration cues to syntax that Stromswold et al. (2002) found are robust.

Experiments 2 and 3 confirmed that phrase-final lengthening and polysyllabic shortening give rise to the progressive active-passive verb stem duration difference (see Chapter 3 for more background on this point). This was achieved by comparison of verb stem duration in progressive active sentences, perfective active sentences, and passive sentences. Perfective active sentences served as an appropriate comparison despite the difference in the auxiliary (*has*) because progressive active verbs are not subject to phrase-final lengthening, and are also not subject to polysyllabic shortening. Because the difference in verb stem duration across perfective and progressive actives was more consistent than the duration difference between perfective actives and passives, we concluded that, while the effects of both processes contribute, polysyllabic shortening contributes to the progressive active-passive duration difference more than phrase-final lengthening does.

In addition to assessing the contribution of phrase-final lengthening and polysyllabic shortening to the verb stem duration cue to syntax, Experiment 3 examined which verb stem segments undergo lengthening and give rise to said cue. The findings of Experiment 3 confirmed that the verb stem vowel is the segment that undergoes lengthening consistently, which supported the findings of Stromswold et al. (2002), and suggested that listeners could potentially use verb stem vowel duration to predict syntax.

To examine acoustic cues to syntax in temporarily ambiguous sentences, we focused on progressive active and passive sentences. We chose this comparison for three reasons. First, because it demonstrated the most robust acoustic differences found by Stromswold et al. (2002). Second, because they have similar surface structure, which allows for temporary syntactic ambiguity. Finally, because they can be used to communicate about similar events, but they have radically different meanings (see Chapter 1). In principle, however, duration cues of this form could occur in any syntactic alternation that also differs in prosodic structure: for example, because polysyllabic shortening and phrase-final lengthening drive the difference, a different alternation where two structures differ in the addition of a syllable or the position of a syllable within a prosodic phrase may yield similar acoustic cues. Future production studies should seek out other interesting comparisons of this sort and determine whether speakers provide duration cues to syntax in those instances as well.

In summary, native English speakers do produce acoustic cues to passive syntax, though some cues are more consistent than others. Next, we summarize sentence processing behavior when the most consistent duration cue was altered.

6.2. Chapter 4 Summary: Sentence Comprehension by Native English Speakers

Stromswold et al. (2002; 2016) found evidence that listeners can use acoustic cues during the morphosyntactically ambiguous region of a sentence to predict whether it is a progressive active or a passive sentence. Across the 3 studies presented in Chapter 3 and the study done by Stromswold et al. (2002), the most consistent acoustic difference in this region was lengthening of the passive verb stem, which was realized on the verb stem vowel. Taken as a whole, these findings suggested that listeners may have used the duration of the verb stem vowel to help them form predictions about the syntax of the sentence.

To test whether listeners rely on verb stem vowel duration in order to form predictions about the syntax of these temporarily ambiguous sentences, we manipulated the duration of the verb stem vowel to be longer in progressive active verbs and shorter in passive verbs. In a norming study, we found that listeners judged the manipulation to sound natural overall, except that they rated active sentences with manipulated verb stem vowels as *less* natural. We then used the manipulated and unmanipulated recordings as auditory stimuli in a visual world paradigm comprehension task. If native English speakers expect verb stem vowels to be longer in passives because they do not undergo polysyllabic shortening, and vice versa for progressive active verb stem vowels, then listeners should have incurred a processing delay of some kind when the duration of the verb stem vowel disagreed with their expectations. However, the processing delay we found was very small, which suggests that listeners do not rely heavily on verb stem vowel duration to infer structure.

Processing was delayed for passive sentences as evidenced by longer response times on trials with passive sentences as compared to active sentences. A much smaller difference was found for sentences containing manipulated as opposed to unmanipulated verb stem vowels, but the difference was not reliable. Gaze traces suggested that listeners looked to the correct picture sooner when they heard unmanipulated audio, particularly in active sentences. This may reflect that the verb stem vowel manipulation sounded less natural to listeners in active sentences (see Experiment 4a). Gaze traces also suggested that listeners learned the distribution of sentences in the experiment only when they heard unmanipulated audio, but this learning effect only appeared to affect saccades during the verb stem. Additionally, while we found evidence of predictive processing, listeners did not look to the target image as soon as we anticipated based on the findings of Stromswold et al. (2002).

Salverda et al. (2003; 2007) found that listeners formed predictions about the location of a syllable within a prosodic phrase (phrase-final or phrase-medial), and about the total number of syllables in the word. If listeners can make inferences of this sort, then in our study they should have been able to predict the syntax of the utterance using verb stem vowel duration, given that vowel duration affects syllable duration, and would inform the listener about the presence of phrase-final lengthening and/or polysyllabic shortening. Listeners were able to form predictions, but did not appear to use verb stem vowel duration—and syllable duration, by extension—to do so.

There are important methodological differences between our study and their work that may explain the difference in listeners' behavior. Salverda et al. (2003) spliced sentences together so that they shared a target syllable (e.g., *ham*), leaving any acoustic cues that may have preceded the target syllable intact. This means that subtle cues in the region of the sentence before the target syllable may have helped listeners form the correct inference: determining the context of the syllable based not only on its duration, but also on acoustic information that preceded the duration cue. In our study, we manipulated only the duration of the verb stem vowel, so acoustic cues preceding the verb stem vowel were consistent with the original recording. In other words, listeners in our study had more consistent acoustic cues to work with than those in Salverda et al. (2003). Future work along this line should minimize the presence of extraneous acoustic cues in order to identify whether one particular acoustic cue is informative to listeners.

Similarly, Salverda et al. (2007) used very different sentences in order to manipulate the location of the target syllable within a prosodic phrase. Because the sentence content and phonological environment were not held constant preceding the target syllable, the differences across their test sentences may have affected listeners' behavior. In our study, preverbal sentence content was the same across syntactic constructions. Future work along the line of Salverda et al. (2007) should take care to minimize differences in test sentences preceding the target syllable.

Lastly, Salverda et al. (2003; 2007) were interested in incremental lexical processing using durational cues. In our study, we examined sentence processing using similar durational cues. These are not mutually exclusive processes: indeed, incremental sentence processing requires incremental lexical processing. In our study, listeners needed to identify the entire verb (the stem and its inflection) in order to infer syntax. However, in addition to identifying individual words in the utterance, listeners in our study were also constructing a syntactic tree. A deeper explanation for the difference between our findings and those of Salverda et al. (2003; 2007) is that sentence processing requires the resources necessary for lexical processing *and* additional resources to infer the underlying structure. The parser may not have enough resources available to devote to predictive processing in this context, or the payoff may not be worthwhile (see 6.4 for more discussion on this point). Furthermore, lexical processing may be more susceptible to subtle cues in the input than sentence processing as a whole because it is less costly to reanalyze the identity of a word than to re-parse an entire sentence, and because lexical processing may be less constrained by long distance relationships in the sentence.

In summary, listeners in our study did not appear to use acoustic cues to predictively process progressive active and passive sentences, and were not strongly affected by the verb stem vowel manipulation. Next, we summarize production and comprehension of these sentences by L2 English speakers.

6.3. Chapter 5 Summary: Sentence Production and Comprehension by L2 English Speakers

If native English speakers learn acoustic cues to syntax from distributional evidence by identifying statistical regularities in the input during the process of first language acquisition (Lew-Williams & Fernald, 2007; 2010), then it may be the case that L2 English speakers are unable to use those same cues to inform parsing decisions, because they acquired English through the use of their native language rather than using distributional evidence.

Furthermore, there is conflicting evidence over whether L2 speakers can form predictions using cues that are not present in their native language (Marull, 2017; Lew-Williams & Fernald, 2007; 2010). There is evidence that L2 speakers cannot form predictions if their native language lacks the cue entirely (e.g., English has no grammatical gender; Lew-Williams & Fernald, 2007; 2010), but *can* form predictions if their native language lacks a specific cue (e.g., plural morphology on definite articles) but can generalize from similar cues that are present in the native language (e.g., plural morphology on demonstratives; Marull, 2017; see Chapter 2 for a more detailed review).

In Chapter 5, we tested whether native Mandarin speakers could use acoustic cues to facilitate the processing of progressive active and passive English sentences. We expected that native Mandarin speakers would wait until after the verbal inflection to look at the correct image, which would reflect an inability to use durational cues to English syntax at all. Furthermore, we expected that native Mandarin speakers may have waited until the second noun phrase if they were unable to use the verbal inflection at all.

Our findings supported both of these predictions. Mandarin speakers appeared to wait until the second noun phrase in order to look to the correct image. This may support Lew-Williams & Fernald (2007; 2010) if they waited because they were unable to use a cue that was entirely absent in their native tongue—inflectional morphology on the verb—in order to form predictions. However, another possibility that we cannot currently refute is that Mandarin speakers simply waited as long as they did because their English proficiency was too low, and sentence processing is just more difficult in a listener's non-native language.

To determine which is responsible for our findings, native Mandarin speakers must be compared to another group of non-native English speakers with similar English proficiency, but whose native language has inflectional morphology on verbs (e.g., Spanish speakers). Data collection for this comparison is still underway.

In Chapter 5, we also investigated whether native Mandarin speakers were able to produce similar duration cues when they say the sentences themselves. There is reason to believe this would not be the case. Mandarin does have phrase-final lengthening, but does not show polysyllabic shortening, and might even have polysyllabic lengthening (Lai et al., 2010). If the same phonological processes that drive the duration difference for native English speakers must be present in Mandarin for native Mandarin speakers to produce the same cues, then we expected no passive verb stem lengthening. Similarly, because Mandarin lacks inflectional morphology on verbs, we expected that speakers would fail to produce verbal inflections in their utterances.

We found that native Mandarin speakers *did* lengthen verb stems in passive sentences, though they did not reliably lengthen passive auxiliaries. Future research should carry out a production study like Experiment 2 or Experiment 3 in order to determine whether native Mandarin speakers achieved passive verb stem lengthening using phrase-final lengthening alone, or whether they learned to employ polysyllabic shortening in their English utterances. One speaker who produced longer verb stems in passives did not lengthen the verb stem vowel, which suggests that lengthening was realized elsewhere on the verb stem for a single speaker. In the event that this pattern is not limited to the single speaker in our study, then a follow-up study should be carefully designed to determine which verb stem segment undergoes lengthening when native

Mandarin speakers produce progressive active and passive English sentences. This entails testing only verbs containing monophthongs with no consonant clusters in onset or coda position, in addition to other phonological constraints (see Experiment 3).

To recap, native Mandarin speakers showed an overall processing delay, which may have been due either to low English proficiency or to the inability to use the acoustic and morphosyntactic cues English speakers produce that are not present in Mandarin. Mandarin speakers, however, were able to produce passive verb stem lengthening when they said English sentences, though lengthening was not always realized on the verb stem vowel, and they did not produce any auxiliary lengthening.

6.4. Implications for Sentence Production

Speakers leave a trail of bread crumbs in the form of very subtle acoustic cues during the morphosyntactically ambiguous region of temporarily ambiguous sentences, at least for progressive active and passive sentences. Ordered from least reliable to most reliable, speakers lengthened passive auxiliaries, produced active verb stems with higher intensity, and lengthened passive verb stems by lengthening passive verb stem vowels (see Chapter 3).[13]

Prosodic structure differs between sentences of these types as a consequence of English morphosyntax, and the difference in prosodic structure yields acoustic differences that are correlated with syntax. Why might prosodic structure operate in this way? It stands to reason that a lengthened vowel is more salient than its shortened counterpart, because the lengthened vowel occupies a greater portion of the speech signal.

[13] Note that we did not analyze verb stem segments in Experiments 1 and 2, and did not analyze the auxiliary in Experiment 3. If we had, the order of cues with respect to reliability may be different.

According to some theories of language production (Jaeger, 2010; Aylett & Turk, 2004; Turk, 2010), speakers "boost" informative portions of the speech signal (e.g., via lengthening) and reduce the parts that are less informative. Crucially, speakers do not make a *conscious* effort to modulate information in this way (Jaeger, 2010). The uniform information density (UID) hypothesis and the smooth signal redundancy hypothesis, advocated by Jaeger (2010) and Turk (2010), respectively, explain that speakers keep the information content of their utterances as uniform as possible. Information content in this context refers to the predictability of the word given its lexical frequency (how common the word is in the language), its meaning (how plausible is the word given the current context), syntactic constraints (does this word make the sentence grammatical), and other linguistic knowledge. Speakers may use prosody to modulate the salience of the speech signal based on its information value (Turk, 2010). This is supported by the observation that words that are highly likely are shorter (Tanenhaus, Kurumada, & Brown, 2015; Jaeger, 2010; Turk, 2010) and are articulated less carefully than unlikely words: for example, they may undergo vowel reduction by shortening in duration, or by changing the vowel to schwa (ə). Function words (e.g., determiners: the, a, prepositions: to, for) have low information value (Jaeger, 2010; Turk, 2010; citations therein) and may even be deleted. This would not be detrimental to the listener, because high information words provide sufficient context to fill in the gaps if low information words are not clearly perceived (Umeda, 1976). In contrast, words that are less likely in a sentence are longer in duration, and less frequently undergo vowel reduction, or truncation.

Perhaps speakers lengthen passive verb stem vowels because passive sentences are infrequent in English, and therefore less likely in a sentence. Speakers may boost the speech signal during the region that disambiguates the passive structure (the verb), which is highly informative, and use prosody to achieve signal boosting. When the passive verb is monosyllabic, signal boosting would be realized on the verb stem vowel as lengthening.

However, it is also possible that prosody is not an engine that speakers use to modulate the information content of their utterances. Furthermore, it seems unlikely that speakers would use prosody to help listeners process temporarily ambiguous sentences, given that speakers only use prosody to help listeners resolve permanently ambiguous sentences if they realize there is an ambiguity (Snedeker & Trueswell, 2003; Wagner & Watson, 2010). For temporarily ambiguous sentences, the speaker is likely unaware that the sentence is ambiguous early on, and would be unlikely to signal any ambiguity to the listener given that morphosyntactic disambiguation would occur down the road. For listeners, on the other hand, all sentences are temporarily ambiguous early on. It stands to reason that listeners would use any useful information present in the acoustic signal to inform processing decisions, whether it was produced for that purpose or not.

This also raises the question of whether passive verb stem lengthening takes place to help the listener. In all of our studies, speakers read sentences aloud in a sound-attenuated booth. Because speakers were not producing sentences to communicate with a listener, it is unlikely that they signaled passive syntax in order to help a hypothetical listener process their sentences. We did not find evidence that speakers emphasized the passive verb stem to make it more salient: a focus analysis (not reported in the dissertation) did not reveal a reliable difference in stem duration when the verb stem was focused, and verb stem intensity was higher for *active* verb stems, not for passives.

However, it is possible that the language production machinery does optimize for communication, and does so whether a listener is present or not. This could be tested experimentally using a task where the speaker must produce progressive active and passive sentences with the explicit goal of communicating with a listener. For example, the speaker could produce a sentence that corresponds to 1 of 2 images for a listener, who must then select which image is best described by the sentence. The speaker would then produce the same sentences in a task without the goal of communication, and the acoustics of the utterances would be compared across conditions.

6.5. Implications for Sentence Processing

If speakers consistently produce an acoustic cue to syntax, and listeners are capable of using acoustic cues to incrementally process speech, why was predictive processing less robust than expected in Chapter 4? Furthermore, why was processing relatively unaffected by the verb stem vowel manipulation tested in Experiment 4b? There are several possible reasons for this, some of which have been laid out in 4.2.3.

One such possibility is that changing a single cue may not be enough to lead the parser astray. In many ways, this indicates that the parser is robust. Eventually, the parser must map noisy input to a single structural representation (Gibson et al., 2013). If the parser failed in the absence of a single acoustic cue, sentence processing would be slow, and highly error-prone. When the parser has received a certain amount of input containing cues that are consistent with one structure (e.g., longer auxiliary and lower verb stem intensity → passive syntax), a single inconsistent cue—particularly one that has a lot of uncertainty attached to it, like the duration of the verb stem vowel—would not immediately point to another syntactic representation. If the parser operated in this manner, listeners would garden-path on a regular basis. Instead, once the parser has some information, it may make asymmetric use of subsequent cues that either are consistent or inconsistent with that information. This would mean the parser requires more inconsistent evidence to reject the current parse than it needs consistent evidence to continue along the current path. Further investigation is required to shed light on this possibility.

Recall that listeners appeared to learn the distribution of sentences in the unmanipulated audio condition, but not in the manipulated audio condition, when they heard manipulated verb stem vowels. This may indicate how the parser handles noisy input. Listeners in the manipulated audio condition heard an acoustic cue (verb stem vowel duration) that was inconsistent with preceding acoustic cues. This may have lead listeners to rely more heavily on prior knowledge (e.g., the distribution of active and passive sentences in English as a whole) rather than integrating information from recent experience, which was acoustically inconsistent. This would support the idea that failures of prediction are opportunities to update prior beliefs (Ferreira & Lowder, 2016).

Another potential reason is that the cue did not occur early enough in the sentence to drive prediction. Recall from the review in Chapter 2 that listeners prioritize early information when forming an initial parse (Trueswell et al., 1999; Choi & Trueswell, 2010) The verb stem vowel occurs relatively early in the sentence, and verb stem vowel duration was the most reliable cue to syntax in Chapter 3. However, it also immediately preceded the point of morphosyntactic disambiguation in the sentence.

Given that the verb stem was preceded by a *be* auxiliary, inflectional morphology *must* be present on the verb in order for the sentence to be grammatical in English. If predictive processing is computationally expensive, it may not be worth the effort to form a prediction about upcoming sentence content if a deterministic cue is on the horizon. Even though earlier cues (e.g., auxiliary duration) are less reliable cues to syntax, it is possible that listeners forgive the low reliability of the cue in favor of information that occurs early enough to be useful. If this is the case, it may indicate a minimum distance between a probabilistic cue and the point of morphosyntactic disambiguation that the parser requires in order to invest processing effort in prediction. This could be tested using a similar paradigm, but manipulating an earlier cue (e.g., auxiliary duration).

The alternative, however, is also possible: the cue was sufficiently early, but was not sufficiently reliable, or salient. The cues listeners used in Trueswell et al. (1999) and Choi & Trueswell (2010) were morphosyntactic, and the acoustic cues used by Dahan et al. (2002) and Snedeker & Yuan (2008) were highly salient. Verb stem vowel duration consistently cued passive syntax more than the other candidate acoustic cues that speakers produced did, but in 2 of the experiments presented in Chapter 3 (specifically, Experiments 2 and 3), not all speakers produced passive verb stems that were longer than progressive active verb stems. Even if all speakers *had* lengthened passive verb stems consistently, the cue does not occur for all English verbs: many of the most common English verbs are irregular, and would not be temporarily ambiguous across progressive active and passive sentences (e.g., *was holding*, *was held*). This means that the cue would not always be present or informative in the syntactic alternations tested here, even if speakers would otherwise produce them. Additionally, the duration cues in question are much subtler than the cues that have been investigated previously (e.g., the salience of a pause in Snedeker & Yuan, 2008; see Chapter 2). In other words, the parser may set a higher threshold for evidence in order to form predictions, which the acoustic cues in our study failed to breach.

An alternative argument on the relative salience of acoustic cues in our study as compared to the studies summarized in section 2.1 is that listeners do not commit to a prediction based on acoustic evidence unless that acoustic evidence is *extremely* salient—arguably, to a degree not typically encountered when listening to spoken utterances—such that listener use the information more readily than they would otherwise. It is possible that predictive processing of this sort has been oversold in the literature. While prediction is certainly part of sentence processing, it must operate under reasonable constraints. Future research should investigate situations where listeners do and do not form prediction based on available evidence.

For reasons that are discussed in section 5.3, the results of the dissertation as it stands cannot speak to the debate over whether second language learners can learn to use cues that are not present in their native language. It may be the case that native Mandarin speakers incur a processing delay compared to native English speakers because they are unable to use cues that are present in English, but absent in Mandarin, to facilitate processing. However, it is likely the case that the processing delay we observed in Chapter 5 was due to differences in English proficiency between native English speaking adults, who are the gold standard for English proficiency, and native Mandarin speakers who learned English later in life. Future experiments, such as those proposed in this chapter and in Chapter 5, could provide more compelling evidence that speaks to this debate.

In sum, we cannot conclude whether listeners use subtle acoustic cues to syntax in order to predictively process sentences based on the findings of the dissertation. However, we do find some evidence that the parser is robust in the face of noisy input, and may rely more strongly on prior knowledge when subtle cues in the input are inconsistent. Future experiments of the sort proposed in this section must be conducted in order to clarify the role of acoustic cues to syntax in sentence processing.

Appendix A

1. Experiment 1 Stimuli: Target Sentences

Sentence	Syntax
The hippo was chasing the turtle.	Progressive active
The hippo was chased by the turtle.	Passive
The turtle was chasing the hippo.	Progressive active
The turtle was chased by the hippo.	Passive
The fox was combing the lion.	Progressive active
The fox was combed by the lion.	Passive
The lion was combing the fox.	Progressive active
The lion was combed by the fox.	Passive
The elephant was kicking the kangaroo.	Progressive active
The elephant was kicked by the kangaroo.	Passive
The kangaroo was kicking the elephant.	Progressive active
The kangaroo was kicked by the elephant.	Passive
The pig was kissing the sheep.	Progressive active
The pig was kissed by the sheep.	Passive
The sheep was kissing the pig.	Progressive active
The sheep was kissed by the pig.	Passive
The bear was licking the dog.	Progressive active
The bear was licked by the dog.	Passive
The dog was licking the bear.	Progressive active
The dog was licked by the bear.	Passive
The duck was patting the rabbit.	Progressive active
The duck was patted by the rabbit.	Passive
The rabbit was patting the duck.	Progressive active
The rabbit was patted by the duck.	Passive
The monkey was pinching the rabbit.	Progressive active

The monkey was pinched by the rabbit.	Passive
The rabbit was pinching the monkey.	Progressive active
The rabbit was pinched by the monkey.	Passive
The cow was poking the zebra.	Progressive active
The cow was poked by the zebra.	Passive
The zebra was poking the cow.	Progressive active
The zebra was poked by the cow.	Passive
The bear was punching the dog.	Progressive active
The bear was punched by the dog.	Passive
The dog was punching the bear.	Progressive active
The dog was punched by the bear.	Passive
The cat was pushing the mouse.	Progressive active
The cat was pushed by the mouse.	Passive
The mouse was pushing the cat.	Progressive active
The mouse was pushed by the cat.	Passive
The pig was scrubbing the sheep.	Progressive active
The pig was scrubbed by the sheep.	Passive
The sheep was scrubbing the pig.	Progressive active
The sheep was scrubbed by the pig.	Passive
The frog was shoving the monkey.	Progressive active
The frog was shoved by the monkey.	Passive
The monkey was shoving the frog.	Progressive active
The monkey was shoved by the frog.	Passive
The fox was tickling the lion.	Progressive active
The fox was tickled by the lion.	Passive
The lion was tickling the fox.	Progressive active
The lion was tickled by the fox.	Passive
The cat was touching the rhino.	Progressive active
The cat was touched by the rhino.	Passive

The rhino was touching the cat.	Progressive active
The rhino was touched by the cat.	Passive
The frog was trapping the monkey.	Progressive active
The frog was trapped by the monkey.	Passive
The monkey was trapping the frog.	Progressive active
The monkey was trapped by the frog.	Passive
The duck was washing the rabbit.	Progressive active
The duck was washed by the rabbit.	Passive
The rabbit was washing the duck.	Progressive active
The rabbit was washed by the duck.	Passive

2. Experiment 1 Stimuli: Filler Sentences

Sentence	Syntax
The bear was holding the book.	Progressive active
The bear was holding the cookie.	Progressive active
The cat was holding the ball.	Progressive active
The cat was holding the medal.	Progressive active
The cow was dirtier than the kangaroo.	Active
The cow was behind the hippo.	Active
The dog was holding the book.	Progressive active
The dog was holding the cookie.	Progressive active
The elephant was near the turtle.	Active
The elephant was cleaner than the rhino.	Active
The elephant was wearing a green shirt.	Progressive active
The fox was wearing a red shirt.	Progressive active
The hippo was behind the cow.	Active
The hippo was dirtier than the mouse.	Active
The hippo was wearing a blue shirt.	Progressive active

The kangaroo was dirtier than the cow.	Active
The kangaroo was behind the lion.	Active
The kangaroo was wearing a yellow shirt.	Progressive active
The lion was behind the kangaroo.	Active
The mouse was dirtier than the hippo.	Active
The mouse was holding the medal.	Progressive active
The mouse was holding the ball.	Progressive active
The rhino was cleaner than the elephant.	Active
The rhino was near the zebra.	Active
The turtle was cleaner than the zebra.	Active
The turtle was near the cow.	Active
The zebra was near the mouse.	Active
The zebra was cleaner than the turtle.	Active

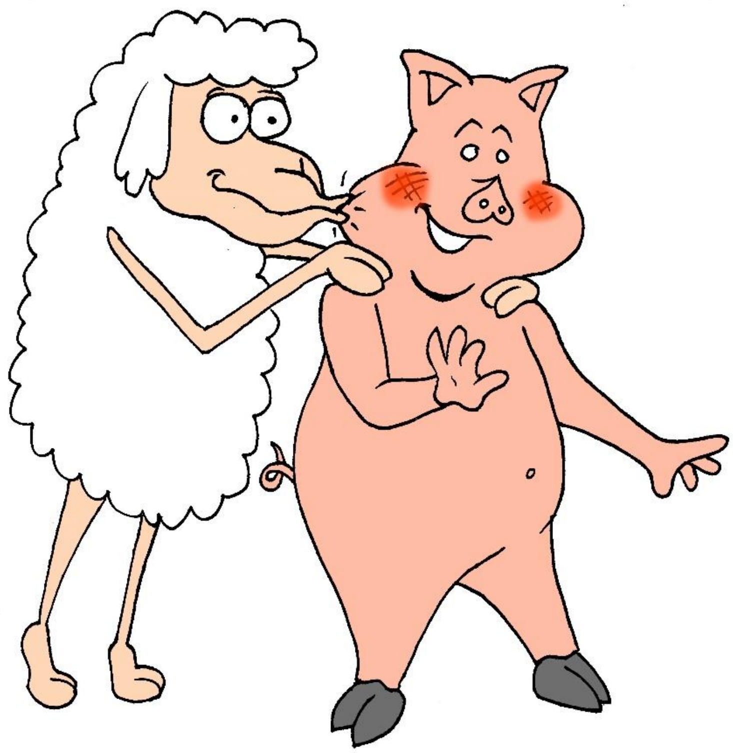
3. Experiment 1 Stimuli: Verb Frequency

	Frequency	
Verb	T-LWF	SUBTL
chase	142	32.80
comb	96	6.06
kick	248	73.41
kiss	1027	121.16
lick	128	10.96
pat	285	—
pinch	86	6.12
poke	60	5.84
punch	78	29.69
push	543	70.55
scrub	109	6.24
shove	110	13.22

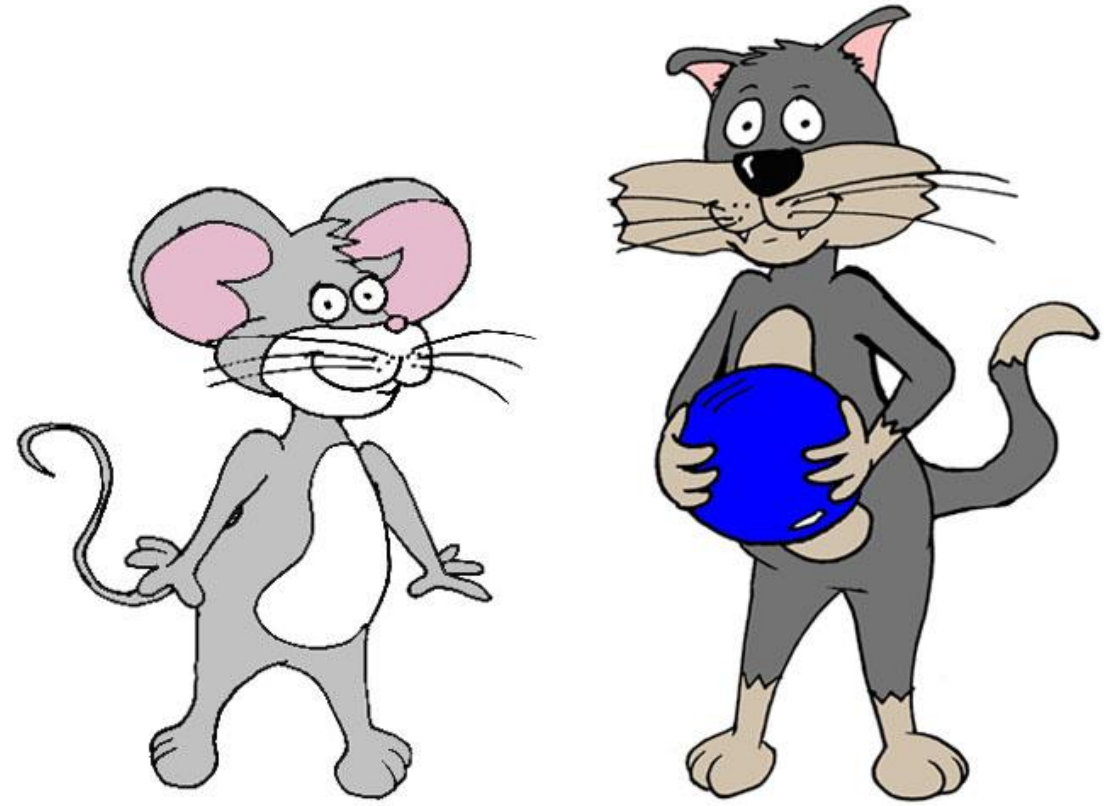
tickle	33	4.80
touch	1016	147.73
trap	199	23.84
wash	563	40.73

4. Experiment 1 Stimuli: Select Images Described by Sentences

Image and Compatible Sentence(s)



Target sentences: *The sheep was kissing the pig, The pig was kissed by the sheep.*



Filler sentence: *The cat was holding the ball.*

Appendix B

1. Experiment 2 Stimuli: Target Sentences

Sentence	Syntax
The actress has mocked the artist publicly.	Perfective active
The actress was mocked by the artist publicly.	Passive
The actress was mocking the artist publicly.	Progressive active
The agent has guarded the cop on the street.	Perfective active
The agent was guarded by the cop on the street.	Passive
The agent was guarding the cop on the street.	Progressive active
The artist has mocked the actress publicly.	Perfective active
The artist was mocked by the actress publicly.	Passive
The artist was mocking the actress publicly.	Progressive active
The author has quoted the journalist in her book.	Perfective active
The author was quoted by the journalist in her book.	Passive
The author was quoting the journalist in her book.	Progressive active
The boss has called the secretary on the phone.	Perfective active
The boss was called by the secretary on the phone.	Passive
The boss was calling the secretary on the phone.	Progressive active
The boy has teased the girl about her hair.	Perfective active
The boy was teased by the girl about his hair.	Passive
The boy was teasing the girl about her hair.	Progressive active
The bride has kissed the groom in the church.	Perfective active
The bride was kissed by the groom in the church.	Passive
The bride was kissing the groom in the church.	Progressive active
The brother has poked the sister a lot.	Perfective active
The brother was poked by the sister a lot.	Passive
The brother was poking the sister a lot.	Progressive active
The bully has kicked the child in the hallway.	Perfective active

The bully was kicked by the child in the hallway.	Passive
The bully was kicking the child in the hallway.	Progressive active
The child has kicked the bully in the hallway.	Perfective active
The child was kicked by the bully in the hallway.	Passive
The child was kicking the bully in the hallway.	Progressive active
The coach has guided the trainer onto the field.	Perfective active
The coach was guided by the trainer onto the field.	Passive
The coach was guiding the trainer onto the field.	Progressive active
The cop has guarded the agent on the street.	Perfective active
The cop was guarded by the agent on the street.	Passive
The cop was guarding the agent on the street.	Progressive active
The criminal has fooled the officer in the deal.	Perfective active
The criminal was fooled by the officer in the deal.	Passive
The criminal was fooling the officer in the deal.	Progressive active
The dentist has treated the medic very quickly.	Perfective active
The dentist was treated by the medic very quickly.	Passive
The dentist was treating the medic very quickly.	Progressive active
The dog has chased the fox in the yard.	Perfective active
The dog was chased by the fox in the yard.	Passive
The dog was chasing the fox in the yard.	Progressive active
The fox has chased the dog in the yard.	Perfective active
The fox was chased by the dog in the yard.	Passive
The fox was chasing the dog in the yard.	Progressive active
The girl has teased the boy about his hair.	Perfective active
The girl was teased by the boy about her hair.	Passive
The girl was teasing the boy about his hair.	Progressive active
The governor has bribed the mayor with a new car.	Perfective active
The governor was bribed by the mayor with a new car.	Passive
The governor was bribing the mayor with a new car.	Progressive active

The groom has kissed the bride in the church.	Perfective active
The groom was kissed by the bride in the church.	Passive
The groom was kissing the bride in the church.	Progressive active
The journalist has quoted the author in her book.	Perfective active
The journalist was quoted by the author in her book.	Passive
The journalist was quoting the author in her book.	Progressive active
The judge has praised the politician at the rally.	Perfective active
The judge was praised by the politician at the rally.	Passive
The judge was praising the politician at the rally.	Progressive active
The leopard has hunted the lion in the desert.	Perfective active
The leopard was hunted by the lion in the desert.	Passive
The leopard was hunting the lion in the desert.	Progressive active
The lion has hunted the leopard in the desert.	Perfective active
The lion was hunted by the leopard in the desert.	Passive
The lion was hunting the leopard in the desert.	Progressive active
The mayor has bribed the governor with a new car.	Perfective active
The mayor was bribed by the governor with a new car.	Passive
The mayor was bribing the governor with a new car.	Progressive active
The medic has treated the dentist very quickly.	Perfective active
The medic was treated by the dentist very quickly.	Passive
The medic was treating the dentist very quickly.	Progressive active
The officer has fooled the criminal in the deal.	Perfective active
The officer was fooled by the criminal in the deal.	Passive
The officer was fooling the criminal in the deal.	Progressive active
The politician has praised the judge at the rally.	Perfective active
The politician was praised by the judge at the rally.	Passive
The politician was praising the judge at the rally.	Progressive active
The secretary has called the boss on the phone.	Perfective active
The secretary was called by the boss on the phone.	Passive

The secretary was calling the boss on the phone.	Progressive active
The sister has poked the brother a lot.	Perfective active
The sister was poked by the brother a lot.	Passive
The sister was poking the brother a lot.	Progressive active
The trainer has guided the coach onto the field.	Perfective active
The trainer was guided by the coach onto the field.	Passive
The trainer was guiding the coach onto the field.	Progressive active

2. Experiment 2 Stimuli: Filler Sentences

Sentence	Syntax
The woman noticed the dentist treating the medic.	Embedded small clause
The reporter suspected the mayor bribing the governor.	Embedded small clause
The fan believed the journalist quoting the author.	Embedded small clause
The man feared the lion hunting the leopard.	Embedded small clause
The man feared the leopard hunting the lion.	Embedded small clause
The reporter suspected the governor bribing the mayor.	Embedded small clause
The woman noticed the medic treating the dentist.	Embedded small clause
The fan believed the author quoting the journalist.	Embedded small clause
The director heard the actress mocking the artist.	Embedded small clause
The uncle saw the brother poking the sister.	Embedded small clause
The resident heard the politician praising the judge.	Embedded small clause
The athlete knew the coach guiding the trainer.	Embedded small clause
The father heard the girl teasing the boy.	Embedded small clause
The lawyer knew the officer fooling the criminal.	Embedded small clause
The friend saw the groom kissing the bride.	Embedded small clause
The citizen saw the agent guarding the cop.	Embedded small clause
The teacher saw the bully kicking the child.	Embedded small clause
The rat saw the fox chasing the dog.	Embedded small clause

The employee heard the secretary calling the boss.	Embedded small clause
The father heard the boy teasing the girl.	Embedded small clause
The citizen saw the cop guarding the agent.	Embedded small clause
The director heard the artist mocking the actress.	Embedded small clause
The friend saw the bride kissing the groom.	Embedded small clause
The resident heard the judge praising the politician.	Embedded small clause
The employee heard the boss calling the secretary.	Embedded small clause
The athlete knew the trainer guiding the coach.	Embedded small clause
The lawyer knew the criminal fooling the officer.	Embedded small clause
The teacher saw the child kicking the bully.	Embedded small clause
The uncle saw the sister poking the brother.	Embedded small clause
The rat saw the dog chasing the fox.	Embedded small clause
The father heard the boy teased the girl.	Embedded past active
The athlete knew the trainer guided the coach.	Embedded past active
The reporter suspected the governor bribed the mayor.	Embedded past active
The teacher saw the child kicked the bully.	Embedded past active
The friend saw the groom kissed the bride.	Embedded past active
The resident heard the politician praised the judge.	Embedded past active
The uncle saw the sister poked the brother.	Embedded past active
The director heard the artist mocked the actress.	Embedded past active
The fan believed the author quoted the journalist.	Embedded past active
The lawyer knew the officer fooled the criminal.	Embedded past active
The employee heard the boss called the secretary.	Embedded past active
The man feared the lion hunted the leopard.	Embedded past active
The woman noticed the dentist treated the medic.	Embedded past active
The citizen saw the agent guarded the cop.	Embedded past active
The rat saw the dog chased the fox.	Embedded past active
The friend saw the bride kissed the groom.	Embedded past active
The fan believed the journalist quoted the author.	Embedded past active

The citizen saw the cop guarded the agent.	Embedded past active
The father heard the girl teased the boy.	Embedded past active
The director heard the actress mocked the artist.	Embedded past active
The lawyer knew the criminal fooled the officer.	Embedded past active
The woman noticed the medic treated the dentist.	Embedded past active
The teacher saw the bully kicked the child.	Embedded past active
The employee heard the secretary called the boss.	Embedded past active
The rat saw the fox chased the dog.	Embedded past active
The reporter suspected the mayor bribed the governor.	Embedded past active
The man feared the leopard hunted the lion.	Embedded past active
The athlete knew the coach guided the trainer.	Embedded past active
The resident heard the judge praised the politician.	Embedded past active
The uncle saw the brother poked the sister.	Embedded past active
The groom will kiss the bride in the church.	Simple future
The artist will mock the actress publicly.	Simple future
The lion will hunt the leopard in the desert.	Simple future
The fox will chase the dog in the yard.	Simple future
The cop will guard the agent on the street.	Simple future
The author will quote the journalist in her book.	Simple future
The boy will tease the girl about her hair.	Simple future
The criminal will fool the officer in the deal.	Simple future
The dentist will treat the medic very quickly.	Simple future
The bully will kick the child in the hallway.	Simple future
The boss will call the secretary on the phone.	Simple future
The coach will guide the trainer onto the field.	Simple future
The judge will praise the politician at the rally.	Simple future
The sister will poke the brother a lot.	Simple future
The mayor will bribe the governor with a new car.	Simple future
The agent will guard the cop on the street.	Simple future

The bride will kiss the groom in the church.	Simple future
The journalist will quote the author in her book.	Simple future
The girl will tease the boy about his hair.	Simple future
The governor will bribe the mayor with a new car.	Simple future
The brother will poke the sister a lot.	Simple future
The politician will praise the judge at the rally.	Simple future
The dog will chase the fox in the yard.	Simple future
The child will kick the bully in the hallway.	Simple future
The secretary will call the boss on the phone.	Simple future
The actress will mock the artist publicly.	Simple future
The leopard will hunt the lion in the desert.	Simple future
The medic will treat the dentist very quickly.	Simple future
The trainer will guide the coach onto the field.	Simple future
The officer will fool the criminal in the deal.	Simple future
The cop has garbage to throw away at the end of his shift.	Active
The leopard has hundreds of dark spots on its back.	Active
The actress has modest taste in clothes and shoes.	Active
The boss has coffee in her travel mug.	Active
The medic has trophies on her bookshelf.	Active
The child has kittens to take care of after school.	Active
The secretary has colorful dress shoes.	Active
The sister has poker chips in her backpack.	Active
The girl has teenage friends at her new high school.	Active
The mayor has brand new sneakers to wear.	Active
The groom has kitchen supplies in his apartment.	Active
The bride has keys to the new apartment.	Active
The officer has food in the refrigerator.	Active
The criminal has foolish plans to escape the prison.	Active
The fox has cherries it can eat today.	Active

The bully has candy in his lunchbox today.	Active
The dog has chains on its purple collar.	Active
The lion has hungry cubs to feed.	Active
The judge has proof the man was guilty of the crime.	Active
The coach has guidelines for his football players.	Active
The governor has brilliant ideas for new laws.	Active
The dentist has trees in his large backyard.	Active
The brother has Pokémon cards in his backpack.	Active
The trainer has guys waiting for him at the gym.	Active
The boy has t-shirts in the back of his closet.	Active
The journalist has questions to ask the witness.	Active
The artist has modern furniture in his house.	Active
The agent has gardening tools in the backyard.	Active
The politician has practical goals for the city.	Active
The author has quickly run out of ideas.	Active

3. Experiment 2 Stimuli: Verb Frequency

	Frequency		
Verb	T-LWF	SUBTL	*-ed* Inflection
chase	142	32.80	voiceless stop (t)
kick	248	73.41	voiceless stop (t)
kiss	1027	121.16	voiceless stop (t)
mock	102	5.37	voiceless stop (t)
poke	60	5.84	voiceless stop (t)
bribe	31	6.04	voiced stop (d)
call	3533	861.39	voiced stop (d)
fool	1140	89.33	voiced stop (d)
praise	116	9.45	voiced stop (d)
tease	105	5.69	voiced stop (d)

guard	423	58.20	syllabic (ɪd)
guide	287	17.84	syllabic (ɪd)
hunt	353	25.86	syllabic (ɪd)
quote	123	9.57	syllabic (ɪd)
treat	394	51.88	syllabic (ɪd)

Appendix C

1. Experiment 3 Stimuli: Target Sentences

Sentence	Syntax
The new novel was banned across the nation.	Truncated Passive
The mayor was banning a rally in the city limits.	Progressive Active
The governor was banning a movie in his state.	Progressive Active
The clerk was banning a customer from the store.	Progressive Active
The pilot has banned a passenger from smoking.	Perfective Active
The fruit was canned a day after harvest.	Truncated Passive
The maid was canning a quart of peaches.	Progressive Active
The wife was canning a large amount of yams.	Progressive Active
The chef was canning a few tomatoes.	Progressive Active
The cook has canned a pound of mushrooms.	Perfective Active
The check was cashed around eleven o'clock.	Truncated Passive
The landlord was cashing a rent check.	Progressive Active
The woman was cashing a big paycheck.	Progressive Active
The worker was cashing a welfare check.	Progressive Active
The teenager has cashed a birthday check.	Perfective Active
The rich man was fanned aboard his yacht.	Truncated Passive
The gambler was fanning a deck of cards.	Progressive Active
The butler was fanning a houseguest.	Progressive Active
The kind maid was fanning a princess.	Progressive Active
The servant has fanned a wealthy lady.	Perfective Active
The mother was hugged around the waist.	Truncated Passive
The nurse was hugging a sick patient.	Progressive Active
The racecar was hugging a tight corner.	Progressive Active
The baby was hugging a tiny puppy.	Progressive Active
The parent has hugged a newborn baby.	Perfective Active

The song was hummed around town.	Truncated Passive
The girl was humming a catchy tune.	Progressive Active
The actor was humming a theme song.	Progressive Active
The mother was humming a lullaby.	Progressive Active
The choir has hummed a church song.	Perfective Active
The student was hushed a dozen times today.	Truncated Passive
The nanny was hushing a noisy baby in the crib.	Progressive Active
The teacher was hushing a prankster in her class.	Progressive Active
The mother was hushing a disobedient toddler.	Progressive Active
The politician has hushed a rude heckler at the rally.	Perfective Active
The young girl was kissed under the starry sky.	Truncated Passive
The suitor was kissing a potential bride.	Progressive Active
The mother was kissing a newborn child.	Progressive Active
The president was kissing another baby.	Progressive Active
The daughter has kissed another old aunt.	Perfective Active
The deadline was missed about a month ago.	Truncated Passive
The first grader was missing a front tooth.	Progressive Active
The new baby was missing a clean blanket.	Progressive Active
The new student was missing a notebook.	Progressive Active
The old woman has missed a doctor's appointment.	Perfective Active
The picnic was packed a couple of days ago.	Truncated Passive
The player was packing up his equipment.	Progressive Active
The traveler was packing a large suitcase.	Progressive Active
The camper was packing a bag for his hike.	Progressive Active
The rock star has packed up his instruments.	Perfective Active
The note was passed around the classroom.	Truncated Passive
The truck was passing a black limousine.	Progressive Active
The guest was passing a bowl of salad.	Progressive Active
The driver was passing another red car.	Progressive Active

The taxi has passed another pedestrian.	Perfective Active
The red apple was picked a couple of hours ago.	Truncated Passive
The little girl was picking a pumpkin to carve.	Progressive Active
The farmer was picking a few ripe pears.	Progressive Active
The cowboy was picking a new horse to ride.	Progressive Active
The new husband has picked a honeymoon spot.	Perfective Active
The criminal was pinned against a brick wall.	Truncated Passive
The teenager was pinning a post to the webpage.	Progressive Active
The president was pinning a medal on the soldier.	Progressive Active
The new intern was pinning a note on the wall.	Progressive Active
The grandmother has pinned a skirt for the girl.	Perfective Active
The castle wall was rammed until it collapsed.	Truncated Passive
The young knight was ramming a castle gate.	Progressive Active
The mountain goat was ramming another goat.	Progressive Active
The angry bull was ramming a wood fence.	Progressive Active
The huge truck has rammed a taxi on the highway.	Perfective Active
The mayor was rushed around the busy town.	Truncated Passive
The train was rushing across the steel bridge.	Progressive Active
The taxicab was rushing along the city street.	Progressive Active
The soldier was rushing across the battlefield.	Progressive Active
The student has rushed around his campus.	Perfective Active
The girl was tagged about a dozen times.	Truncated Passive
The boy was tagging along with his sister.	Progressive Active
The grocer was tagging a bag of apples.	Progressive Active
The child was tagging a friend at recess.	Progressive Active
The mailman has tagged another package.	Perfective Active
The girl's hair was tugged a couple of times.	Truncated Passive
The toddler was tugging a nurse's arm.	Progressive Active
The puppy was tugging a red chew toy.	Progressive Active

The jogger was tugging a dog by the leash.	Progressive Active
The rescuer has tugged a boat to port.	Perfective Active
The package was wrapped a while ago.	Truncated Passive
The skier was wrapping a scarf around himself.	Progressive Active
The soldier was wrapping a bad injury.	Progressive Active
The scout was wrapping a rope around the tree.	Progressive Active
The big snake has wrapped around the tree branch.	Perfective Active

2. Experiment 3 Stimuli: Filler Sentences

Sentence	Syntax
The pie was eaten by the dinner guests.	Passive
The donkey was kicking a farmer.	Progressive Active
The guitarist has played that song before.	Perfective Active
The monkey was climbing a tall tree.	Progressive Active
The attic was filled with useless junk.	Truncated Passive
The garbage men grabbed the trash yesterday.	Past Active
The chicken was cooked at seven o'clock.	Truncated Passive
The eagle was flying above the mountains.	Progressive Active
The artist has painted a beautiful picture.	Perfective Active
The mobsters stole a bag of cash from the bank.	Past Active
The monsters scared the children after dark.	Past Active
The beach was crowded with teenagers.	Truncated Passive
The dictator has ruled the country for years.	Perfective Active
The patient was treated by the skilled doctor.	Passive
The trees shaded the children from the sun.	Past Active
The fish was swimming far below the surface.	Progressive Active
The cartoonist was drawing a new character.	Progressive Active
The old woman collected stamps in her free time.	Past Active

The strict teachers banned a controversial book.	Past active
The two bankers cashed a check for their new client.	Past active
The school girls hugged a favorite teacher.	Past active
The angry nuns hushed a student at morning prayer.	Past active
The suspended girls missed a day of school.	Past active
The super spies passed a secret message.	Past active
The scoutmasters pinned a badge on the boy.	Past active
The truck drivers rushed across the highway.	Past active
The two huskies tugged a large dog sled.	Past active
The farmers canned a bunch of potatoes.	Past active
The gusty winds fanned across the fields.	Past active
The singers hummed a beautiful melody.	Past active
The brave schoolgirls kissed a couple of boys.	Past active
The college seniors packed a snack for the trip.	Past active
The careful players picked a new teammate.	Past active
The brave warriors rammed a hostile fortress.	Past active
The workers tagged another tree to remove.	Past active
The caring sons wrapped a birthday gift for their mom.	Past active

3. Experiment 3 Stimuli: Verb Frequency

	Frequency		
Verb	T-LWF	SUBTL	Vowel
ban	38	3.14	æ
can	6733	5247.45	æ
cash	302	72.43	æ
fan	145	35.14	æ
pack	469	43.82	æ
pass	1815	108.12	æ
ram	24	6.43	æ

tag	52	13.88	æ
wrap	293	17.80	æ
hug	94	19.33	ʌ
hum	83	4.82	ʌ
hush	90	—	ʌ
rush	730	31.41	ʌ
tug	94	2.75	ʌ
kiss	1027	121.16	ɪ
miss	2408	—	ɪ
pick	986	198.39	ɪ
pin	201	16.37	ɪ

Appendix D

1. Experiment 4b Code: [Python] Vowel Manipulation Script

dur_swap.py

```
import sys
from os import listdir
from open_and_extract import extract_xmin_xmax
from open_and_extract import extract_period
from vst import vowel_strectching_time
from dur_swap_logger import create_log

##############################################
#
#   Extracts active/passive verb stem vowel
#   information from Praat TextGrid files and
#   creates new Duration Tiers which will
#   effectively "swap" the duration of the
#   active and passive verb stems.
#
#   written by Sten Knutsen
#
##############################################


input_filenames_raw = listdir("/Users/stenknutsen/Desktop/IO_folder")
print input_filenames_raw

input_filenames=[]
for f in input_filenames_raw:
    if f.startswith("."):
        continue
    else:
        input_filenames.append(f)


print(input_filenames)


active_filename = input_filenames[0]
passive_filename = input_filenames[1]


print(active_filename)
print(passive_filename)

active_xmin_xmax =
extract_xmin_xmax("/Users/stenknutsen/Desktop/IO_folder/"+
active_filename)
passive_xmin_xmax =
extract_xmin_xmax("/Users/stenknutsen/Desktop/IO_folder/"+
passive_filename)

print(active_xmin_xmax)
print(passive_xmin_xmax)
```

```
active_xmin = active_xmin_xmax["xmin"]
print(active_xmin)
active_xmax = active_xmin_xmax["xmax"]
print(active_xmax)
passive_xmin = passive_xmin_xmax["xmin"]
passive_xmax = passive_xmin_xmax["xmax"]
print(passive_xmin)
print(passive_xmax)

active_vowel_dur = active_xmax - active_xmin
passive_vowel_dur = passive_xmax - passive_xmin
print("active vowel dur:")
print(active_vowel_dur)
print("passive vowel dur:")
print(passive_vowel_dur)

active_period = extract_period("/Users/stenknutsen/Desktop/IO_folder/"
+ active_filename)
print("active period:")
print(active_period)
passive_period = extract_period("/Users/stenknutsen/Desktop/IO_folder/"
+ passive_filename)
print("passive period:")
print(passive_period)

#let the swapping begin!
#
new_active_vowel_dur = active_vowel_dur
new_passive_vowel_dur = passive_vowel_dur

while (new_active_vowel_dur<passive_vowel_dur):
    new_active_vowel_dur = new_active_vowel_dur + active_period


while (new_passive_vowel_dur>active_vowel_dur):
    new_passive_vowel_dur = new_passive_vowel_dur - passive_period

print("new active vowel dur:")
print(new_active_vowel_dur)
print("new passive vowel dur:")
print(new_passive_vowel_dur)


#Find percentage increase/decrease for active/passive
#
percent_lengthen_active = (new_active_vowel_dur/active_vowel_dur)-1
print("percent lengthen active:")
print(percent_lengthen_active)

percent_shorten_passive = (-1)*(1-
(new_passive_vowel_dur/passive_vowel_dur))
print("percent shorten passive:")
print(percent_shorten_passive)


#Create log for set of files
```

```
#
create_log(active_filename,passive_filename,active_vowel_dur,passive_vo
wel_dur,active_period,

passive_period,new_active_vowel_dur,new_passive_vowel_dur,percent_lengt
hen_active,percent_shorten_passive)

vowel_strectching_time(active_filename, active_xmin, active_xmax,
percent_lengthen_active)
vowel_strectching_time(passive_filename, passive_xmin, passive_xmax,
percent_shorten_passive)
```

vst.py

```
import sys

##############################################
#
#   vst.py
#   Calculates the "curve" for the Duration
#   Tier from data extracted by dur_swap.py
#   and creates a new Praat duration tier
#   file.
#
#   written by Sten Knutsen
#
#
##############################################


def vowel_strectching_time(file_name, begin_v, end_v,
percent_legnthen):
    print("Hi from VST!!!!")
    dif = end_v - begin_v
    perc = float(dif*0.40)
    begin_vowel = begin_v+perc
    end_vowel = end_v - perc
    print("begin vowel 40 percent is: ")+str(begin_vowel)
    print("end vowel 40 percent is: ")+str(end_vowel)
    total_area = end_v - begin_v
    midpoint = float((end_v-begin_v)/2.0)+begin_v
    new_area = float(percent_legnthen*total_area)
    x = float((end_vowel-begin_vowel)/2.0)
    y = float(new_area/x)
    print('percent lengthen is: ')+str(percent_legnthen)
    print('total area (before lengthening) is: ')+str(total_area)
    print('adding new area of: ')+str(new_area)
    print('for a total area of: ')+str(new_area+total_area)
    print('midpoint is: ')+str(midpoint)
    print('x is: ')+str(x)
    print('y is: ')+str(y)
    print('height of peak is: ')+str(y+1.0)
    tier_name = ("dur_"+ file_name.split(".")[0])
    print(tier_name)
    file_name =
("/Users/stenknutsen/Desktop/IO_folder/dur_"+file_name.split(".")[0]+".
praat")
```

```
    print(file_name)
    print('Creating new praat file')

    file = open(file_name,'a')
    file.write('Create DurationTier: \"'+tier_name+'\", '+
str(begin_vowel)+', '+str(end_vowel)+'\n')
    file.write('Add point: '+str(begin_vowel)+', 1\n')
    file.write('Add point: '+str(midpoint)+', '+str(y+1.0)+'\n')
    file.write('Add point: '+str(end_vowel)+', 1\n')

    file.close()
```

2. Experiment 4b and 5a Code: [Python] Track-Loss Correcting Script

```
###########
# this script identifies track loss in the gaze data (0 values). if the
#     track loss spans fewer than 6 frames (the time required to plan
#     and execute an eye movement) and gaze is directed to the same
#     side of the screen on both sides of the gap, then the gap is
#     filled.
#
# written by Sten Knutsen
# modified by Gwendolyn Rehrig
###########

import os

# folder locations
pre =
'D:/Documents/School/Graduate/Rutgers/Research/LALP/Dissertation/L2
study/gaze/'

infile = pre+'original/'
outfile = pre+'filled/'
details = pre+'notes/'

global notes
notes = []

def file_len(fname):
    with open(fname) as f:
        for i, l in enumerate(f):
            pass
    return i + 1

# function definitions
def file_to_list(fname):
    list = []
    with open(fname) as f:
        for line in f:
            if line == "\n":
                break
            line = line.rstrip()
            list.append(line)
    return list
```

```
def strip_leading_zeroes(lfile):
    list = []
    end_run = False;

    for line in lfile:
        if(line.split(",")[3]=='0')&(end_run==False):
            continue
        else:
            list.append(line)
            end_run = True
    return list


def strip_trailing_zeroes(lfile):
    list = []
    while(len(lfile)!= 0):
        list.append(lfile.pop())

    newlist = strip_leading_zeroes(list)

    ret_list = []

    while(len(newlist)!=0):
        ret_list.append(newlist.pop())


    return ret_list


def find_zeroes(lfile):
    list = []
    for line in lfile:
        if (line.split(",")[3]=='0'):
            list.append(line+"<")
        else:
            list.append(line)
    return list

def count_zeroes(lfile):
    count=0
    for line in lfile:
        if line[33]=="0":
            count = count + 1
    return count

def tag_start(lfile):
    list = []
    prevLine = "
"
    for line in lfile:
        if (line.endswith("<"))& (prevLine[33]!="0"):
            list.pop()
            list.append(prevLine+"<s>")
            list.append(line)
        else:
            list.append(line)
```

```
        prevLine = line

    return list

def count_starts(lfile):
    count=0
    for line in lfile:
        if line.endswith("<s>"):
            count = count+1
    return count

def start_index(lfile):
    i = 0
    j = 0
    k=[]
    for line in lfile:
        if line.endswith("<s>"):
            i = lfile.index(line)
            break
    for line in range(i+1, len(lfile)):
        if lfile[line].endswith("<"):
            j = j+1
        else:
            break
    k.append(i)
    k.append(j)
    return k

def replace_zeroes(start_index_and_num, lfile):
    start = start_index_and_num[0]
    num_zeroes = start_index_and_num[1]

    rep = lfile[int(start)]
    rep = rep[33:].split("<s>")[0]

    start_x = int(lfile[start].split(",")[2])
    end_x = int(lfile[start +num_zeroes+1].split(",")[2])
    if (start_x>960)&(end_x>960)&(num_zeroes<7):
        lfile[start] = lfile[start].split("<s>")[0]
        for i in range(start+1, start+num_zeroes+1):
            lfile[i] = lfile[i][:33].split("<")[0]
            lfile[i] = lfile[i]+rep
    elif (start_x<=960)&(end_x<=960)&(num_zeroes<7):
        lfile[start] = lfile[start].split("<s>")[0]
        for i in range(start+1, start+num_zeroes+1):
            lfile[i] = lfile[i][:33].split("<")[0]
            lfile[i] = lfile[i]+rep

    else:
        lfile[start] = lfile[start].split("<s>")[0]
        for i in range(start+1, start+num_zeroes+1):
            lfile[i] = lfile[i].split("<")[0]
    return lfile

# main
files = os.listdir(infile)
```

```
for f in files:
    # Transfer contents of file into list
    l = file_to_list(infile+f)

    # Remove zeroes from beginning and end of file
    l = strip_leading_zeroes(l)
    l = strip_trailing_zeroes(l)

    l = find_zeroes(l)
    l = tag_start(l)
    num = count_starts(l)
    zeroes_before = count_zeroes(l)

    for i in range(0,num):
        k = start_index(l)
        l = replace_zeroes(k,l)

    new_file_name = (outfile+f)

    total = len(l)

    file = open(new_file_name, 'a')
    for line in l:
        file.write(line +"\n")

    file.close()
    zeroes_after = count_zeroes(l)

    notes.append("File was "+ str(len(l))+" lines long")
    notes.append("Number of zeroes before: " + str(zeroes_before)+ ", "
      + str(float(zeroes_before)/total) + " % of the data")
    notes.append("Number of zeroes after: " + str(zeroes_after)+ ", " +
      str(float(zeroes_after)/total) + " % of the data")
    file = open(details+f,'a')
    for x in notes:
        file.write(str(x) +"\n")
    file.close()
    notes = []
```

3. Experiment 4b and 5a Code: [R] Eye-Tracking Data Parsing Function

```
###################
# this function takes a row from an eprime data file (one trial) as
#     input, along with a data structure containing gaze data, and
#     the following row of eprime data. it then parses the gaze data
#     into trials as defined in the eprime data, populates a data
#     structure with information that we need for analyses, and returns
#     that data structure.
# written by Gwendolyn Rehrig
###################

match_gaze <- function(vec,gaze,nextrow){
  # what are the arguments to this function?
  #   vec: the e-prime data for the current row (trial)
```

```
  #   gaze: the imported gaze data (unparsed, raw)
  #   nextrow: the e-prime data for the next row (trial)
  #     need this to get the time stamp for the next trial,
  #     defines endpoint of current trial

  # calculating start time: get the time the audio started playing,
  # add the ms into the file
  # when the first determiner starts, and subtract 50 ms to get the
  # frame before that point.
  # this should help correct for audio files with leading silences.

  # I use the timestamp from the start of the next trial and subtract
  # the time the stimulus presented from that to get the ms since
  # launch when the stimulus displayed.
  # I subtract 50ms from the start of the first determiner to get at
  # least one frame before, in case the eye tracking data doesn't
  # sample during the exact start

  t_start <- (nextrow$trial_ms-(nextrow$ms_since_launch.Block.-
      vec$Stimulus.OnsetTime))+(vec$det1_onset_ms)-50

  # whatever timestamp marks the start of the next trial
  t_stop <- nextrow$trial_ms

  # midpoint (x dimension) of the screen, to be used in determining
  # what side of the screen the participant looked at during that frame

  mid <- (vec$x_width)/2

  # get all the data that falls within the start and end of the trial
  this_trial <-
      which(gaze$time_ms>t_start&gaze$time_ms<t_stop&gaze$date==vec$Ses
      sionDate)

if((length(this_trial)>0)&(gaze$date[this_trial[1]]==vec$SessionDate)&(
vec$Stimulus.ACC==1)){
    print(paste0("Working on.... Trial: ",vec$Block," Subject:
      ",vec$Subject," Block: ",vec$block))

    # get a temporary data frame to work with
    current_trial <- gaze[this_trial,]

    # preallocating vectors
    look_active <-
      vector(mode="numeric",length=length(current_trial[,1]))
    look_passive <-
      vector(mode="numeric",length=length(current_trial[,1]))
    vowel_type <-
      vector(mode="character",length=length(current_trial[,1]))
    sentence <-
      vector(mode="character",length=length(current_trial[,1]))
    actpsv <- vector(mode="character",length=length(current_trial[,1]))
    subj <- vector(mode="numeric",length=length(current_trial[,1]))
    gaze_stamp <-
      vector(mode="numeric",length=length(current_trial[,1]))
    trial <- vector(mode="numeric",length=length(current_trial[,1]))
    verb <- vector(mode="character",length=length(current_trial[,1]))
```

```
block <- vector(mode="numeric",length=length(current_trial[,1]))
version <-
  vector(mode="character",length=length(current_trial[,1]))
frame <- vector(mode="numeric",length=length(current_trial[,1]))
np1 <- vector(mode="character",length=length(current_trial[,1]))
np2 <- vector(mode="character",length=length(current_trial[,1]))
time <- vector(mode="numeric",length=length(current_trial[,1]))
raw_x <- vector(mode="numeric",length=length(current_trial[,1]))
raw_y <- vector(mode="numeric",length=length(current_trial[,1]))
n1_onset <- vector(mode="numeric",length=length(current_trial[,1]))
det1_onset <-
  vector(mode="numeric",length=length(current_trial[,1]))
aux_onset <-
  vector(mode="numeric",length=length(current_trial[,1]))
prep_onset <-
  vector(mode="numeric",length=length(current_trial[,1]))
vstem_onset <-
  vector(mode="numeric",length=length(current_trial[,1]))
infl_onset <-
  vector(mode="numeric",length=length(current_trial[,1]))
det2_onset <-
  vector(mode="numeric",length=length(current_trial[,1]))
n2_onset <- vector(mode="numeric",length=length(current_trial[,1]))
constituent <-
  vector(mode="character",length=length(current_trial[,1]))
nmorph <- vector(mode="numeric",length=length(current_trial[,1]))
vowel_diff <-
  vector(mode="numeric",length=length(current_trial[,1]))
group <- vector(mode="numeric",length=length(current_trial[,1]))
sex <- vector(mode="character",length=length(current_trial[,1]))
hand <- vector(mode="character",length=length(current_trial[,1]))
look_acc <- vector(mode="numeric",length=length(current_trial[,1]))
looking <-
  vector(mode="character",length=length(current_trial[,1]))
JND <- vector(mode="character",length=length(current_trial[,1]))
vowel_offset <-
  vector(mode="numeric",length=length(current_trial[,1]))
is_vowel <-
  vector(mode="character",length=length(current_trial[,1]))
time_bin <- vector(mode="numeric",length=length(current_trial[,1]))
saccade <- vector(mode="numeric",length=length(current_trial[,1]))

# here's the only loop in this function
for(ts in 1:length(current_trial[,1])){
  # staggering the audio onset timestamps so that it makes sense
  # with the correction we made earlier (setting the start of the
  # audio/trial to 50ms before det1)
  # this will now tell us what time, wrt our modified start time,
  # the participant starts hearing each constituent

  if(current_trial$time_ms[ts]-t_start>0){
    n1_onset[ts] <- (vec$n1_onset_ms-vec$det1_onset_ms)+50
    det1_onset[ts] <- 50
    aux_onset[ts] <- (vec$aux_onset_ms-vec$det1_onset_ms)+50
    if(vec$actpsv=="passive"){
      prep_onset[ts] <- (vec$prep_onset_ms-vec$det1_onset_ms)+50
    } else {
```

```
      prep_onset[ts] <- NA
    }
    vstem_onset[ts] <- (vec$vstem_onset_ms-vec$det1_onset_ms)+50
    infl_onset[ts] <- (vec$infl_onset_ms-vec$det1_onset_ms)+50
    det2_onset[ts] <- (vec$det2_onset_ms-vec$det1_onset_ms)+50
    n2_onset[ts] <- (vec$n2_onset_ms-vec$det1_onset_ms)+50

    # assigning the x and y coordinates from the gaze data. since
    # sometimes the right eye is unreliable, we need to switch to
    # the left eye if needed. or, if both are bad, just set it to
    # NA so it doesn't affect our analysis
    if(current_trial$right_x[ts]<1&current_trial$left_x[ts]<0){
      look_x <- NA
      look_y <- NA
    } else
  if(current_trial$right_x[ts]>0&current_trial$right_x[ts]<=vec$x_w
  idth){
      look_x <- current_trial$right_x[ts]
      if(current_trial$right_y[ts]<=vec$y_height){
        look_y <- current_trial$right_y[ts]
      } else {
        look_y <- NA
      }
    } else
  if(current_trial$left_x[ts]>0&current_trial$left_x[ts]<=vec$x_wid
  th){
      look_x <- current_trial$left_x[ts]
      if(current_trial$left_y[ts]<=vec$y_height){
        look_y <- current_trial$left_y[ts]
      } else {
        look_y <- NA
      }
    } else {
      look_x <- current_trial$right_x[ts]
      look_y <- current_trial$right_y[ts]
    }

    vowel_type[ts] <- vec$vowel_type
    group[ts] <- vec$Group
    sex[ts] <- vec$Sex
    hand[ts] <- vec$Handedness
    sentence[ts] <- vec$Sentence
    actpsv[ts] <- vec$actpsv
    subj[ts] <- vec$Subject
    vowel_offset[ts] <- vec$vowel_offset

    gaze_stamp[ts] <- current_trial$time_ms[ts]
    trial[ts] <- vec$Block
    verb[ts] <- vec$verb
    np1[ts] <- vec$np1
    np2[ts] <- vec$np2
    time[ts] <- current_trial$time_ms[ts]-t_start
    time_bin[ts] <- (floor(((current_trial$time_ms[ts]-t_start)-
        vstem_onset[ts])/200))*200
    frame[ts] <- floor(time[ts]/33.33)
    raw_x[ts] <- look_x
    raw_y[ts] <- look_y
```

```
block[ts] <- vec$block
version[ts] <- vec$Version
vowel_diff[ts] <- vec$diff
JND[ts] <- vec$JND

# get the indices for the gaze data that fall within the
# beginning and end of the audio

# figure out which constituent they should be hearing. if there
# is missing data, skip the trial.
if(current_trial$time_ms[ts]>=t_start+50
   &current_trial$time_ms[ts]<=(t_start+(vec$n1_onset_ms-
    vec$det1_onset_ms)+50)){
  constituent[ts] <- 'det1'
  nmorph[ts] <- 1
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else if(current_trial$time_ms[ts]>=t_start+(vec$n1_onset_ms-
    vec$det1_onset_ms)+50
    &current_trial$time_ms[ts]<=t_start+(vec$aux_onset_ms-
    vec$det1_onset_ms)+50){
  constituent[ts] <- 'n1'
  nmorph[ts] <- 2
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else if(current_trial$time_ms[ts]>=t_start+(vec$aux_onset_ms-
    vec$det1_onset_ms)+50
    &current_trial$time_ms[ts]<=t_start+(vec$vstem_onset_ms-
    vec$det1_onset_ms)+50){
  constituent[ts] <- 'aux'
  nmorph[ts] <- 3
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else
    if(current_trial$time_ms[ts]>=t_start+(vec$vstem_onset_ms-
    vec$det1_onset_ms)+50
    &current_trial$time_ms[ts]<=t_start+(vec$infl_onset_ms-
    vec$det1_onset_ms)+50){
  constituent[ts] <- 'stem'
  nmorph[ts] <- 4
  if(current_trial$time_ms[ts]>=t_start+(vec$vowel_onset-
    vec$det1_onset_ms)+50
     &current_trial$time_ms[ts]<=t_start+(vec$vowel_offset-
    vec$det1_onset_ms)+50){
    is_vowel[ts] <- 'yes'
  } else {
    is_vowel[ts] <- 'no'
  }
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else
    if(vec$actpsv=="Passive"&current_trial$time_ms[ts]>=t_start
    +(vec$infl_onset_ms-vec$det1_onset_ms)+50
```

```
    &current_trial$time_ms[ts]<=t_start+(vec$prep_onset_ms-
    vec$det1_onset_ms)+50){
  constituent[ts] <- 'inflection'
  nmorph[ts] <- 5
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else
    if(vec$actpsv=="Active"&current_trial$time_ms[ts]>=t_start+
    (vec$infl_onset_ms-vec$det1_onset_ms)+50
    &current_trial$time_ms[ts]<=t_start+(vec$det2_onset_ms-
    vec$det1_onset_ms)+50){
  constituent[ts] <- 'inflection'
  nmorph[ts] <- 5
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else
    if(vec$actpsv=="Passive"&current_trial$time_ms[ts]>=t_start
    +(vec$prep_onset_ms-
    vec$det1_onset_ms)+50&current_trial$time_ms[ts]<=t_start+(v
    ec$det2_onset_ms-vec$det1_onset_ms)+50){
  constituent[ts] <- 'preposition'
  nmorph[ts] <- 6
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else
    if(current_trial$time_ms[ts]>=t_start+(vec$det2_onset_ms-
    vec$det1_onset_ms)+50
    &current_trial$time_ms[ts]<=t_start+(vec$n2_onset_ms-
    vec$det1_onset_ms)+50){
  constituent[ts] <- 'det2'
  nmorph[ts] <- 7
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else if(current_trial$time_ms[ts]>=t_start+(vec$n2_onset_ms-
    vec$det1_onset_ms)+50&current_trial$time_ms[ts]<=t_start+(v
    ec$sound_duration_ms-vec$det1_onset_ms)+50){
  constituent[ts] <- 'n2'
  nmorph[ts] <- 8
  if(!is.na(look_x)&(look_x<1|look_x>vec$x_width)){
    return(0)
  }
} else {
  constituent[ts] <- ""
  nmorph[ts] <- NA
}

if(!is.na(look_x)){
  if((look_x>mid)&(look_x<((mid*2)+1))){
    # if you looked at the right-hand side
    looking[ts] <- "right"
    if(vec$correct_side=="right"){
      # and you were right:
      look_acc[ts] <- 1
```

```
          } else if(vec$correct_side=="left"){
            # and you were wrong:
            look_acc[ts] <- 0
          }
        } else if(look_x<mid&look_x>0){
          # if you looked at the left-hand side
          looking[ts] <- "left"
          if(vec$correct_side=="left"){
            # and you were right:
            look_acc[ts] <- 1
          } else if(vec$correct_side=="right"){
            look_acc[ts] <- 0
          }

        } else {
          # if you looked right at the midpoint?

          look_acc[ts] <- NA
        }
      } else {
        look_acc[ts] <- NA
      }

      if(!is.na(look_acc[ts])&ts>1){
        if(!is.na(look_acc[ts-1])){
          # let's get more saccades! to be conservative, setting this
to 110 pixels to reflect the error of the eye tribe
          if(((raw_x[ts]&raw_x[ts-1])>0)&(abs(raw_x[ts]-raw_x[ts-
1])>110)){
            saccade[ts] <- look_acc[ts]

          if(!is.na(saccade[ts])&!is.na(saccade[ts-1])){
            if(saccade[ts]==saccade[ts-1]){
              saccade[ts-1] <- NA
            }
          }
        }
      }
    }
  }


  print("     Wrapping up this trial...")
  if(length(look_active)==length(subj)){
    temp <- data.frame(subj, trial, verb, vowel_type, np1, np2,
          actpsv, sentence, raw_x, raw_y, time, nmorph, constituent,
          frame, det1_onset, n1_onset, aux_onset, vstem_onset,
          vowel_offset, infl_onset, prep_onset, det2_onset, n2_onset,
          is_vowel, vowel_diff, version, block, group, sex, hand,
          look_acc, looking, JND, time_bin,saccade)
    # this screens out all the data before and after the sentence
    # audio
    temp <- temp[which(!is.na(temp$nmorph)),]
    return(temp)
  } else {
    print(paste0("     Missing data in participant ", vec$Subject,"
          trial ",vec$Block))
```

```
    }
rm(temp,subj,trial,verb,vowel_type,np1,np2,actpsv,sentence,raw_x,raw_y,
      frame,time,look_active, look_passive, nmorph, constituent,
      saccade, det1_onset, n1_onset, aux_onset, vstem_onset,
      vowel_offset, infl_onset, prep_onset, det2_onset, n2_onset,
      vowel_diff, version, block, group, sex, hand, look_acc, looking,
      JND, is_vowel, time_bin)
  }
}
```

Appendix E

L2 English Speakers: Language History Questionnaire

Question	Response Options
What language did you learn first?	English Other (type your Native Language)
Did you learn English before kindergarten (age 5)?	Yes No Unsure
If you were not born in an English-speaking country, how old were you when you moved to an English-speaking country?	I was born in an English-speaking country Age (years)
How old were you when you first went to a school where the language of instruction was English?	(text field)
How old were you when you first took an English class?	(text field)
If you took the TOEFL (Test of English as a Foreign Language), what were your scores for each section? If you did not take the TOEFL, leave the sliders at 0.	Slider ranging from 0:30 for the following sections: Reading Listening Speaking Writing
List any adults who lived in the same household as you when you were under 6 years of age, and give their native language. For example, Mother (Spanish), Grandmother (Spanish), Aunt (English)…	(text field)
If you took the SAT, what were your scores for each section? If you did not take the SAT, leave the sliders at 200.	Slider ranging from 200:800 for the following sections:

	Critical Reading Writing Math
Which language do you prefer to read in?	English Other (type your Native Language)
Which language do you prefer to write in?	English Other (type your Native Language)
Which language do you prefer to listen in?	English Other (type your Native Language)
Which language do you prefer to speak in?	English Other (type your Native Language)
How would you rate your reading, writing, speaking, and listening in English? (1 = cannot do, 7 = fluent)	Slider ranging from 1:7 for the following sections: Reading Writing Speaking Listening
How would you rate your reading, writing, speaking, and listening in your native language? (1 = cannot do, 7 = fluent)	Slider ranging from 1:7 for the following sections: Reading Writing Speaking Listening
Have you ever been diagnosed with a spoken or written language disorder? If yes, please indicate the type of disorder.	Yes (text box) No
Have you ever been diagnosed with a non- language learning disorder? If yes, please indicate the type of disorder.	Yes (text box) No

Do any immediate family members (parents, siblings, children) have a spoken or written language disorder? If yes, please indicate who and the type of disorder.	Yes (text box) No
Do any immediate family members have a non-language learning disorder? If yes, please indicate who and the type of disorder.	Yes (text box) No
Are any members of your immediate family left handed or ambidextrous? If yes, please indicate who.	Yes (text box) No

Bibliography

Allopenna, P. D., Magnuson, J. S., Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419-439.

Aveni, K., de Lacy, P., Rehrig, G., Stromswold, K. (2016). Acoustic correlates of syntactic structure. Poster presented March 4th 2016, *Eastern Psychological Association Conference*. New York Marriott Marquis, New York City, NY.

Aylett, M. P. & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech, 47*(1), 31-56.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes, 11*(1/2), 17-67.

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. C. Kingston & M. E. Beckman (Eds.), *Laboratory Phonology: Volume 1, Between the Grammar and Physics of Speech* (pp. 152-178). Cambridge University Press.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279-362). New York: Wiley.

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5:9/10, 341-345.

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.

Choi, Y., Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology, 106,* 41-61.

Crocker, M. W. (2002). Sentence comprehension: The integration of habits and rules. *Computational Linguistics, 28*(2), 238-241.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech, 40,* 141-201.

Dahan, D., Tanenhaus, M. K., Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47,* 292-314.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*(3), 547-619.

Fasold, R. W. (1972). *Tense marking in Black English: A linguistic and social analysis* (Vol. 8). Harcourt College Pub.

Fernald, A., Swingley, D., Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72*(4), 1003-1015.

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation, Vol. 65,* 217-247.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11-15.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348-368.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research, 24*(6), 437-468.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051-8056.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Rsearch, 28*(2), 191-215.

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, *22*(3), 338-365.

Harris, J. D. (1963). Loudness discrimination. *Journal of Speech & Hearing Disorders. Monograph Supplement*.

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

Katz, J., Andrews, W., & Buth, E. (2013, December 21). *How y'all, youse and you guys talk.* Retrieved from http://www.nytimes.com/interactive/2013/12/20/sunday-review/dialect-quiz-map.html.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59(5)*, 1208-1221.

Klatt, D. H., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In A. Cohen, & S. G. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 69-89). Springer.

Labov, W. (1989). The child as linguistic historian. *Language variation and change, 1*(1), 85-97.

Lai, C., Sui, Y., & Yuan, J. (2010). A corpus study of the prosody of polysyllabic words in Mandarin Chinese. In *Proceedings of Speech Prosody, Vol. 2010*.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America, 51*(6), 2018-2024.

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *The Journal of the Acoustical Society of America, 54*(5), 1228-1234.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Lew-Williams, C., & Fernald, A. (2007). Young children learnig Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18,* 193-198.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63*, 447-464.

Mayro, A., de Lacy, P., Rehrig, G., & Stromswold, K. (2016). Locus and cause of passive verb stem lengthening. (unpublished honors thesis). Rutgers University—New Brunswick, Piscataway.

Marull, C. (2017). Second language processing efficiency: Experience and cognitive effects on L2 morphosyntactic integration and anticipation. In M. LaMendola and J. Scott (Eds.), *Proceedings of the 41st annual Boston University Conference on Language Development*, 466-480. Somerville, MA: Cascadilla Press.

Nam, H., Saltzman, E., Krivokapić, J., & Goldstein, L. (2008). Modeling the durational difference of stressed vs. unstressed syllables. In *Proceedings of the 8th Phonetic Conference of China*.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—Part II. *Lang. Linguist. Compass., 10,* 591-613.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32,* 693-703.

Port, R. (1981). Linguistic timing factors in combination. *The Journal of the Acoustical Society of America, 69*(1)*, 262-274.*

Rehrig, G., Beier, E., Chalmers, E., Schrum, N., & Stromswold, K. (2015). Robust acoustic cues indicate upcoming structure in active and passive sentences. *The 28th Annual CUNY Conference on Human Sentence Processing.* Midtown Los Angeles Radisson Hotel, Los Angeles, CA.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51-89.

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation in lexical competition. *Cognition, 105,* 466-476.

Salverda, A. P., Kleinschmidt, D., Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71,* 145-163.

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

Selkirk, E. O. (1986). *Phonology and syntax: The relationship between sound and structure.* MIT Press.

Selkirk, E. O. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 550-569). Blackwell.

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

Shih, S. (2016). Prosodic structure and final vowel lengthening in Taiwan Mandarin. (unpublished qualifying paper). Rutgers University—New Brunswick, New Brunswick.

Snedeker J, Trueswell J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48,* 103–130.

Snedeker, J., Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language, 58*(2), 574-608.

Stromswold, K., Eisenband, J., Norland, E., & Ratzan, J. (2002). Tracking the acquisition and processing of English passives: Using acoustic cues to disambiguate actives and passives. *The 16th Annual CUNY Conference on Human Sentence Processing.* Cambridge, MA.

Stromswold, K., Kharkwal, G., Sorkin, J. E., & Zola, S. (under review). Tracking the elusive passive: The processing of spoken passives.

Stromswold, K., Lai, M., Rehrig, G., & de Lacy, P. (2016). Passive sentences can be predicted by adults. *The 29th Annual CUNY Conference on Human Sentence Processing.* Gainesville, FL.

Swingley, D., Pinto, J. P., Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71,* 73-108.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In: Innui, T., McClelland, J. L. (Eds.), *Attention and Performance 16: Information Integration in Perception and Communication* (pp. 457-478). MIT Press, Cambridge, MA.

Tanenhaus, M., Kurumada, C., & Brown, M. (2015). Prosody and intention recognition. In L. Frazier, & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing* (pp. 99-118). Springer International Publishing.

Thothathiri, M., Snedeker, J. (2008). Syntactic priming during language comprehension in three- and four-year-old children. *Journal of Memory and Language, 58,* 188-213.

Townsend, D. J. & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules.* Cambridge, MA: MIT Press.

Traxler, M. J. (2014). Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences, 18*(11), 605-611.

Trueswell, J. C., Sekerina, I., Hill, N.M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*, 89-134.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist framework for constraint-based syntactic ambiguity resolution. In L. Frazier & K. Rayner (Eds). *Perspectives on Sentence Processing* (pp. 155-179). Lawrence Erlbaum Assoc.

Turk, A. (2010). Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis. *Laboratory Phonology, 1*(2), 227-262.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in Empirical Prosody Research, 3,* 1-28.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28*, 397-440.

Umeda, N. (1976). Consonant duration in American English. *Journal of the Acoustical Society of America, 61(3),* 846-858.

Vaissière, J. (1983). Language-independent prosodic features. *Prosody: Models and Measurements* (53-65), Springer Verlag.

van Santen, J. P. H. (1992). Contextual effects on vowel duration. *Speech Communication, 11*, 513-546.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*(7-9), 905-945.

Zec, D. (2007). The Syllable. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 161-194). Cambridge University Press.