Copyright 2017 Peiyun Zhou

AUDITORY PERCEPTUAL SIMULATION DURING SILENT READING: EFFECTS ON LANGUAGE PROCESSING AND COMPREHENSION

BY

PEIYUN ZHOU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
with a concentration in Secondary Language Acquisition and Teacher Education
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Kiel Christianson, Chair
Professor Emerita Susan Garnsey
Associate Professor Tania Ionin
Professor Elizabeth Stine-Morrow
Associate Professor Duane Watson

ABSTRACT

Readers who have seen the Harry Potter movies before reading the novels may "hear" actors' voices in their heads during subsequent reading. This phenomenon of mentally simulating the voice or characteristics of the depicted speakers has been referred as *auditory perceptual simulation* (APS). In three eye-tracking studies and one Event-related Potentials (ERP) study, APS effects on language processing and comprehension were investigated as people read utterances attributed to a native English speaker, a non-native English speaker, or no speaker at all. APS effects were measured via online eye movements, ERPs, and offline comprehension probes. Study 1, 2, 3 employed a new eye-tracking paradigm that triggered participants' auditory perceptual simulation of a native American-English accented and a non-native English speakers' voices during silent reading. Study 4 investigated how much is this mental simulation of voices like listening to voices by adapting the APS paradigm to the Event-related Potentials (ERPs) experiment.

Results from Studies 1 and 2 demonstrated that inducing APS during silent reading resulted in observable differences in reading speed when readers simulated the speech of faster compared to slower speakers, and compared to silent reading without APS among native English speakers. Social attitude survey results indicated that readers' attitudes towards native and non-native speech did not consistently influence APS-related effects. APS of both native speech and non-native speech increased native English speakers' reading speed, facilitated deeper, less good-enough, sentence processing, and improved comprehension compared to normal silent reading. Study 3 revealed that APS of native and non-native speech affects L2 English speakers' reading speed similarly to the APS effects among native English speakers in Studies 1 and 2; however, APS of non-native speech decreased L2 English speakers' comprehension accuracy of

implausible sentences. Study 4 revealed differences in N400 and P600 waveforms when imagining non-native speakers' voices compared to natives'. Importantly, when imagining non-native speakers committing subject-verb agreement errors, P600 amplitudes were no different from error-free items. The results demonstrated that APS is similar to listening to non-native speech, where readers tend to forgive the minor ungrammatical errors (e.g. subject-verb disagreement) when they simulate the non-native speech.

Keywords: Auditory Perceptual Simulation, language processing, reading comprehension, eye-tracking, ERPs, relative clauses, grammaticality

To my parents

Acknowledgements

In Chinese martial novels, the main characters usually learn a special powerful kind of kong fu through receiving the guidance of a master, getting support from mentors and peers, and enduring many tests of strength, intelligence and perseverance. On my path to learn the "kong fu" in psycholinguistics and educational research, I have received enormous support from a great number of individuals.

First and formost, I would like to thank my advisor, Dr. Kiel Christianson, who led me into the psycholinguistic field, for his patience and encouragement when guiding me through all the challenges in graduate study, equipping me with all skills, and helping me to reflect and learn from my mistakes. Kiel taught me how identify and articulate interesting and important research questions and how to address them thoroughly and he always provide constructive advice while encouraging me to explore and carry out these ideas. Kiel's optimism, sense of humor, and suggestions helped me when tackling a difficult programming problem, receiving a harsh review, getting a rejection letter, and facing other challenges. Kiel is a great mentor who truly cares about his students, and he not only guided and helped me for academic development, but also in life.

Dr. Susan Garnsey who led me into Event-related Potential (ERPs) research also played a significant role in my academic development. Susan's office is always open. I truly enjoyed and benefited greatly from the discussions with Susan for inspiring new research ideas, research strategies and data presentation skills. I also like to thank my committee members, Dr. Tania Ionin, Dr. Elizabeth Stine-Morrow, and Dr. Duane Watson for their thoughtful suggestions and challenging questions from forming research ideas to designing and completing my studies. I also appreciate their support in my future career path.

I am lucky to work with and learn from my excellent lab mates from the EdPsych Psycholinguistics Lab and Language and Brain Lab, Yun Yao, Anastasia Stoops, Xuefei Gao, Cassie Palmer, Nayoung Kim, Joseph Toscano, Aaron Hanford, and Heather Mauch. I am also very lucky to get support from friends at University of Illinois, Matthew Benedict, Shuyi Ma, Kejia Chen, Juan Guan, and Yujie Xia, etc. The discussion and collaboration with Yun is truly stimulating. I draw great pleasure from the time we spent programming experiments and being lunch buddy. Anastasia, Xuefei, and Cassie shared lots of valuable research experiences with me. Shuyi and Matt provided lots of valuable suggestions for my research applications from an outsider perspective. I also own a great deal of gratitude to the undergraduate assistants in these two labs for their help in data collection.

Last but not the least, I would like to thank my family members—my mom, Meiling Liu, my dad, Yongjian Zhou, and my husband, Yuliang Wang. Yuliang's patience, support, optimism, and sense of humor help me through the ups and downs of my Ph.D. My parents always value the importance of education and do their best to provide me all the learning opportunities. One of their gifts to me is cultivating my interests in reading since I was a little girl. Their unconditional love, support, and confidence in me propel me to pursuit a life that I truly enjoy.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: I "HEARD" WHAT YOU'RE SAYING: AUDITORY PERCEPTUAL SIMULATION 13

CHAPTER 3: AUDITORY PERCEPTUAL SIMULATION: SIMULATING SPEECH RATES OR ACCENTS? 54

CHAPTER 4: L2 ENGLISH SPEAKERS' AUDITORY PERCEPTUAL SIMULATION DURING SILENT READING 74

CHAPTER 5: IS IMAGINING A VOICE LIKE LISTENING TO IT?: EVIDENCE FROM ERPS 86

CHAPTER 6: GENERAL DISCUSSION AND FUTURE DIRECTIONS 123

REFERENCES 128

APPENDIX A: EXPERIMENTAL STIMULI USED IN EYE-TRACKING STUDIES 146

APPENDIX B: SOCIAL ATTRACTIVENESS SURVEY 149

APPENDIX C: EXPERIMENTAL STIMULI USED IN ERP STUDY 151

CHAPTER 1

INTRODUCTION

When readers read, their eyes quickly "scan" the orthographic information on the text at an average rate of 300 words per minute while their brains rapidly activate the meaning of the words, but this alone cannot let readers understand the whole text. Readers often first interpret a bundle of words while considering various factors, such as syntactic rules, tense, context, and even background knowledge on the topic or personal life experiences before they eventually reach their individual understanding of what they read. This process of "scanning" words, "sending" information to the brain, activating lexical information, integrating all the words following syntactic rules, and connecting them to the context to make sense is not easy. Readers sometimes would activate the sounds of the words to help their memory and comprehension of the text when they read silently.

The voice that readers activate in head plays a critical role in reading (Rayner, Pollatsek, Ashby, & Clifton, 2012). It has been argued that this voice, also refers as inner speech, helps readers access the meaning of the words, aids higher-order comprehension by bolstering short-term memory, and creates acoustic code that may "be beneficial for creating a short-memory representation in which words are held while other tasks are done, such as working out the syntactic structure of the sentence or holding an easily accessible representation of nouns to search for the referent of a pronoun" (Rayner et al., 2012, p. 189). When inner speech is suppressed during silent reading (e.g. by asking readers to count from one to six while doing the reading task), readers often fail to detect errors of word order and anomalous words, because inner speech creates a supplementary

articulatory code which is produced and utilized in parallel with other aspects of reading (Baddeley, Eldridge, & Lewis 1981).

Researchers further found that the inner speech that readers activated can contain rich information of the depicted characteristics of a character or a voice. If you are a fan of the Harry Potter books and films, when you read Harry Potter's line in the book, i.e. "Give it here, Malfoy, or I'll kick you off your broom" silently, you may "see" the scenario from the movie in your mind while "hearing" actor Daniel Radcliff 's voice "saying" this line. This phenomenon when readers mentally simulate characteristics of someone else's (or their own) voice or the voice of a character in the texts is called Auditory perceptual simulation refers to (Kurby, Magliano, Rapp, 2009; Yao and Scheepers 2011; Stites, Luke, and Christianson, 2013; see Hubbard, 2010, for a review).

My dissertation consists of four studies investigating Auditory Perceptual Simulation (APS) during silent reading, affects native English speakers' and L2 English leaners' language processing and comprehension. The first three studies (Studies 1, 2, and 3) employed eye-tracking to explore how APS of native and non-native speech affects sentence processing and comprehension during silent reading among native English speakers and L2 English speakers. Study 4 used event-relative potentials (ERPs) to examine readers' electrophysiological responses to grammatical errors when they engage in APS of native versus non-native speech in sentence processing. I will first introduce the rationale for each study. Then, I will review the relevant literature. Because I thoroughly discussed the background of auditory perceptual simulation in Studies 1, 2 and 4, I will not repeated it in the introduction. I will review the literature on English subject/object relative clauses sentences processing because these sentences were used as

the target materials in eye-tracking studies (Study 1-3). I will also review ERPs research on ungrammatical sentence processing because it's highly relevant to Study 4.

Rationale for the present studies

In Chapter 2, Study 1 investigated whether APS of a native and a non-native speech would differentially affect the sentence processing speed, pattern, and comprehension through three eye-tracking experiments. Experiment 1 examines readers' sentence processing in normal silent reading and it serves as a baseline for the comparison of APS and normal silent reading. Experiment 2 used a new eye-tracking paradigm with photo cues of the different speakers to trigger the APS effects. Experiment 3 used a different cue—recordings of the speakers' voices—in the same paradigm, because one cue might be more effective in triggering the APS effects than the other.

In Chapter3, Study 2 explored whether the APS effects were triggered by speakers' faster/slower speech rates or readers' biases towards native and non-native speakers. In Study 1, there was a confounding effect of accents and speech rates, because sociolinguistics literature (see Study 2 for more details) reveal that listeners often perceive the non-native speech more negatively and require extra efforts to comprehension, it is hard to tell whether the slower online reading speed during APS of the non-native speech was derived from the difficulty of simulating the non-native speech or simulating the slower non-native speaker's speech rate. Therefore, the same paradigm from Study 1 with manipulation of a faster non-native speech and a slower native speech was used to address this question. Because study 1 and study 2 are published papers, there is some repeated information in the literature review section on auditory perceptual simulation.

In Chapter 4, Study 3 extended the investigation of APS to L2 English speakers' language processing and comprehension using the same paradigm and experimental materials from Study 1, because no one had explored the APS effects among L2 English speakers and L2 English speakers usually need more recourses than native speakers in processing in L2. So it is possible that L2 English speakers may not be able to perform APS in their L2 English. But if they can, the benefits of APS on reading speed and comprehension have the potential to benefit L2 English speakers as well. Thus, study 3 examined whether L2 English speakers can perform auditory perceptual simulation and how APS of native and non-native speech affect their reading speed, reading patterns, and comprehension.

In Chapter 5, Study 4 employed two ERP experiments to explore how APS of a native speech differs from normal silent reading and whether APS of a non-native speech affects readers' electrophysiological responses of grammatical errors differently than the APS of a native speech. Experiment 1 examined reader' brain responses in normal silent reading and its results served as baseline for the comparison with APS conditions. Experiment 2 adapted the paradigm from Study 2 and investigated readers' brain responses to grammatical errors in the APS of native and non-native speech.

Relative clause processing in English

A large body of psycholinguistic research deals with differences in processing difficulty between two types of English relative clauses. They differ in what role the noun that is modified by the relative clause (called its head noun) plays in it. In subject-relative clauses (SRCs), as in (1a), the head noun (*employees*) plays the role of agent/subject of both the relative clause verb (*noticed*) and the whole sentence. In object-relative clauses

(ORCs), as in (1b), the head noun (*fireman*) plays the role of patient/object of the relative clause verb (*noticed*) but agent/subject of the whole sentence. The surface difference between the two types of relatives is a difference in word order. In SRCs, the head and its relative are in the default SVO order, while in ORCs, they are instead in OSV order. Consistent evidence from a variety of paradigms has demonstrated that ORCs are harder to process and comprehend than SRCs in English (Ferreira, 2003; Gibson, 1998; Traxler, Williams, Blozis, & Morris, 2005). In English, readers and listeners take longer to process ORCs and make more errors on them than SRCs in comprehension tasks or lexical decision tasks (c.f. Ford, 1983; Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; Van Dyke & McElree, 2006; Wells, Christiansen, Race, Acheson, & MacDonald, 2009).

(1) a. The employees that noticed the fireman hurried across the open field.

b. The fireman that the employees noticed hurried across the open field.

Several competing theories have attempted to explain this processing asymmetry, such as subject-verb-object structure is more frequent (Roland, Dick, & Elman, 2007), so that structure mapping is easier when reading the SRCs than ORCs. One influential line of research argues that working memory accounts for the processing differences between relative clauses (Ford, 1983; Frazier & Fodor, 1978; Gibson, 1998; Grodner & Gibson, 2005). According to this view, the difference in word order in the two kinds of relatives has the consequence that linear distance between the head noun and the relative clause verb is longer in ORCs than in SRCs, so the latter impose more demands on working

memory retrieval. Results from self-paced reading studies (Grodner & Gibson, 2005) have shown that readers take longer to read the embedded verbs in ORCs than in SRCs, but display no concomitant difficulty at the subject in the ORCs, indicating that resource constraints, rather than structural mapping (S-V-O), affected readers' relative clause processing.

Another line of research focuses more on world knowledge and experience-based expectations to explain the processing differences between ORCs and SRCs (Hale, 2001; Levy, 2008). Because SVO is the most common word order in English (Townsend & Bever, 2001; Roland, Dick, & Elman, 2007), readers usually expect the first noun in the relative clause to be its subject. When they encounter the true subject and the relative clause verb in ORCs, they are surprised that the head noun was not the subject and slow down to integrate the information and reorganize the structure. Another factor is that animate head nouns tend to be modified by SRCs (because animate entities are more likely to be subjects), so an ORC following an animate head noun is especially unexpected, leading to processing difficulty.

Staub (2010) tested these two theories using two eye-tracking experiments. He found that readers regressed more often from the subject noun phrases of ORCs than from the same noun phrases in SRCs. Relative clause type also affects reading times on the embedded verb, including first fixation duration, gaze durations, and total reading times, in the direction that embedded verb with shorter linear distance to the head noun (SRCs) were processed faster than the embedded verb in ORCs, supporting the memory-based theory. Therefore, the authors concluded that both memory retrieval and readers' experience-based expectations contribute to the difficulty of processing object relatives.

Another proposed account argues that ORCs are hard when they modify the subject of a sentence because the different roles played by the head noun in the relative clause (object) and the whole sentence (subject), requires a shift in perspective (MacWhinney, 1982; MacWhinney & Pleh, 1988). In contrast, in SRCs the head noun plays the same role in both the relative and the sentence, thus requiring no shift in perspective and making processing easier.

Others accounts emphasize the potential confusability of the nouns in the relative clause. A series of self-paced reading experiments (Fedorenko, Gibson, & Rohde, 2006; Gordon, Hendrick, & Johnson, 2001) has shown that similarity of the nouns that play the subject and object roles in the relative clause increased processing difficulty. When the subjects and objects were from different categories (e.g. Occupation vs. Name), readers' processing speed of the ORCs and SRCs were not reliably different (Fedorenko et al., 2006).

Another important factor is the animacy of the nouns playing the subject and object roles in relative clauses. As mentioned earlier, animate head nouns are more likely to be modified by SRCs, because animate nouns are more likely to be subjects. And inanimate head nouns are more likely to be modified by ORCs because they are more likely to be objects. Eye-tracking studies have revealed that difficulty in processing object-relative clauses was greatly reduced when the head noun of the relative clause was inanimate (Traxler, Morris, & Seely, 2012). ERP experiments manipulating animacy of the head noun in SRCs and ORCs have found that animate and inanimate head nouns elicit different N400 amplitudes, indicating that readers were surprised to encounter an animate noun at the beginning of the sentence. Animacy effects were also found later in

the sentence. Both the relative clause verbs and the main clause verbs in ORCs elicited larger P600 amplitude than the same words in SRCs, indicating difficulty in integrating them with the syntactic structure during sentence processing (Weckerly & Kutas, 1999).

Another series of ERP studies have investigated relative clause processing in Dutch by manipulating semantic fit in SRCs and ORCs (Kolk, Chwilla, van Herten, & Oor, 2002; Kolk & Chwilla, 2007), as illustrated below in (2).

(2) a. The cat that fled from the mice ran through the room. (anomalous ORC)

b. The mice that fled from the cat ran through the room.

They were surprised to find a "semantic P600" effect instead of an N400 effect in response to semantic reversal anomalies in SRCs even though the sentences were syntactically correct, unambiguous, and had the exactly same words as their counterpart sentences. The "semantic P600" has been observed in this and other studies when the two nouns in a sentence could play plausible roles for the verb but their position in the sentence indicates that they instead play the opposite roles from what would make sense (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kim & Osterhout, 2005). Structural complexity modulated the semantic P600 effects. In the more complex ORCs, there was no semantic P600, while there was in the structurally simpler SRCs. The authors argued that readers try to reach a plausible interpretation if possible, and for the role-reversed SRCs, that requires revising the structure, generating a P600 effect. The absence of a similar effect in ORCs is presumably due to greater difficulty in figuring out how to revise the structure to make things make sense.

There are two current theories of sentence processing that could help explain the semantic P600 effects in relative clauses. Non-canonical syntactic structures are usually more difficult to process and thus more likely to be misinterpreted, especially when the correct interpretation is implausible and there is another highly plausible option, though it is inconsistent with the apparent structure. The Good-Enough Processing theory (Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002) was proposed to explain the processing of sentences containing conflicting syntactic and semantic information. It argues that when the syntactic structure is difficult and semantic information is confusing, readers are more like to rely on the superficial semantic information to reach a quick and dirty interpretation of the sentence, overlooking the deeper syntactic structure (Ferreira et al., 2002). Noisy channel theory (Gibson et al., 2013; Levy, 2008) proposes a similar explanation. On this view, syntactic and semantic routes both contribute to sentence processing, and when one is noisier and/or harder to analyze, readers abandon that route and rely on the easier other route to come up with the meaning of the sentence.

Because the pattern of subject/object relative clauses with manipulation of plausibility are consistent across different studies (e.g. ORCs are harder than SRCs and implausible sentences are hard than plausible ones; Kim & Christianson, 2013; Lim & Christianson, 2013a,b), and because Good-Enough Processing theory provides a clear theoretical explanation for these patterns, I will use relative clauses sentences with manipulations of plausibility in three eye-tracking studies.

ERPs research on ungrammatical sentence processing

Event-related Potential (ERP) method has been used to investigate native speakers (see reviews in Hagoort, Brown, & Osterhout, 1999; Kutas, Van Petten, & Kluender,

2006) and second language learners' (e.g. Gabriele, Fiorentino, & Bañón, 2013; Tokowicz & MacWhinney, 2005) sentence processing because it is a very sensitive measure of the brain responses to language-related events. Three ERP components, N400, Left-Anterior Negativity (LAN), and P600, are associated with different aspects of sentence processing (see Kaan, 2007 for a review). The N400 is an increased electrophysiological response in negativity peaking at around 400 milliseconds (ms) after the onset of offending words containing lexical/semantic difficulties or anomalies (Friederici, Steinhauer, & Frisch, 1999; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). LAN is a left-anterior negativity that peaks around 300-500ms after the onset of words with a syntactic violation (Friederici, Pfeifer, & Hahne, 1993; Kutas & Hillyard 1983; Neville, Nicol, Barss, Forster, & Garrett, 1991). Some scholars posit that LAN is associated with syntactic processing (Friederici, 2002; Hagoort, Wassenaar, & Brown, 2003) while others claim it is associated with working memory load during sentence processing (Kluender and Kutas, 1993a,b; Coulson, King, & Kutas, 1998). Yet another group of scholars suggest that LAN does not exit (Tanner, 2015). The P600 is an increased electrophysiological response in positivity peaking at around 600 milliseconds (ms) after the onset of offending words containing morphosyntactic errors (Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995; Osterhout & Nicol, 1999).

Different types of violations modulated the type, latency, and the size of the ERP components (Molinaro, Barber, & Carreiras, 2011; Steinhauer & Connolly, 2008). Coulson et al (1998) found that subject verb disagreement, i.e. *They <u>suns</u> themselves every Sunday*, led to greater P600 effects than pronoun case mismatch, i.e. *Ray fell down and skinned <u>he</u> knee*. They proposed that this difference derived from the proportion

differences of ungrammatical sentences presented each block. Another possibility is the salience and frequency of grammatical errors. Because it is rare to encounter pronoun case mismatch in written texts or in daily conversation, pronoun case mismatch was more salient. Barber and Carreiras (2005) observed that the size of the P600 effects were greater in the gender violation than the number violation in Spanish, especially in the later phrase of the P600. They proposed that differences in the reanalysis processes caused the different P600 effects— when participants encountered a gender disagreement, they checked both the syntactic integration processes and lexical access; while facing the number violation, they only examined the syntactic integration processes.

Researchers further compared brain responses to ungrammatical sentences using different paradigms, reading vs. listening (Hagoort & Brown, 2000). In the reading paradigm, researchers presented the sentences at a rate that is similar to listening to speech (250ms/word). Results showed that participants' electrophysiological responses to the ungrammatical errors are similar in the reading and listening paradigms.

Interestingly, recent studies have found native speakers' sensitivity to the grammatical error can be modulated by listeners' perception of the speakers (Grey & Van Hell, 2017; Hanulíková, van Alphen, van Goch, & Weber, 2012). A study by Hanulíková et al (2012) revealed that native Dutch speakers respond differently to ungrammatical sentences read by native and non-native Dutch speakers. When a native speech contains gender disagreement errors, listeners generated a P600 effect. However, when the same sentences were read by a non-native Dutch speaker, this P600 effect was gone, indicating that listeners were more forgiving towards the non-native speech. Grey and Van Hell (2017) also observed significantly reduced P600 effects in listeners' electrophysiological

responses to the non-native speech with ungrammatical errors, but not to the non-native speech with semantic anomalies. They also observed listeners' individual differences in responses to native and non-native speech—whether listeners succeeded in identifying the non-native accent as Asian accent influenced their brain responses to the grammatical errors. In Study 4, I will further investigate how native and non-native speech as well as individual differences affect sentence processing when readers perform Auditory Perceptual Simulation of the written input.

CHAPTER 2

I "HEARD" WHAT YOU'RE SAYING: AUDITORY PERCEPTUAL SIMULATION

Chapter 2 contains my previous published work "I "heard" what you're saying: Auditory perceptual simulation"[1].

Syntactic structure is fragile, susceptible to both decay in memory (Sachs, 1967) and interference from competing lexical-semantic information (Christianson, Luke, & Ferreira, 2010; Ferreira, 2003; Lim & Christianson, 2013a, b). Consequently, misinterpretations often arise when non-canonical syntactic structure is used to convey semantically implausible information. For example, readers regularly misinterpret both passive sentences (1) and object-relative clauses (2) below to mean "the bird ate the worm". Ferreira (2003) proposes that such misinterpretations derive from a word-order heuristic in English which interprets, based on frequency and/or syntactic canonicity, Noun-Verb-Noun strings as Subject-Verb-Object structures with Agent-Verb-Patient thematic alignments (Townsend & Bever, 2001). The interpretation derived from the application of this heuristic is then supported by general world knowledge, which, in the case of (1-2) below, tells us that birds usually eat worms, rather than the other way around. Ferreira does not argue that the syntactic structure is not computed. Rather, the syntax is fragile, and under certain circumstances, is overridden by competing plausibility and heuristic-based processing. Substantial related research on visual language processing (i.e., reading) shows that readers often misinterpret sentences like (1-2) as well as other difficult structures (such as so-called garden-path sentences, (3)), despite

[1] Zhou, P. & Christianson, K. (2016a). I "hear" what you're "saying". Auditory perceptual simulation, reading speed, and comprehension. *Quarterly Journal of Experimental Psychology, 69, 972-995.*

signs of both re-reading and co-existing partially correct interpretations (Christianson, 2008; Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Christianson & Luke, 2011; Christianson, Williams, Zacks, & Ferreira, 2006; Lim & Christianson, 2013a,b; Patson, Darowski, Moon, & Ferreira, 2009; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013; Swets, Desmet, Clifton, & Ferreira, 2008; van Gompel, Pickering, Pearson, & Jacob, 2006).

(1) The bird was eaten by the worm.

(2) The bird that the worm ate was slow.

(3) While Anna dressed the baby that was cute played in the crib.

This line of research, all of which falls within the "Good Enough Processing" framework (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007), has so far focused on documenting misinterpretations and accounting for how they arise. Importantly, the evidence suggests that these types of misinterpretations are not haphazard and do not stem from simple inattention. Instead, misinterpretations are predictable and systematic, and derive from competition from information sources – various processing heuristics – that overwhelm the syntactic structure that, if built and maintained sufficiently, will lead unfailingly to the correct interpretation. The aim of the present study is to investigate and document one factor that may strengthen the syntactic representation: the prosodic representation. We hypothesized that if readers could be cued to construct a more perceptually salient prosodic representation during silent reading, the

corresponding syntactic representation would be "buttressed" against intrusions from heuristics.

Considerable research has demonstrated that syntactic structure is intimately intertwined with prosodic structure (e.g., Bader, 1998; Breen, 2014; Carlson, Clifton, & Frazier, 2001; Cutler, Dahan, & van Donselaar, 1997; Fodor, 2002; Slowiaczak & Clifton, 1980; Speer, Kjelgaard, & Dobroth, 1996; Watson & Gibson, 2005). In particular, Fodor's (2002) Implicit Prosody Hypothesis proposes that during normal silent reading, a default prosodic contour is projected onto the stimulus, which reflects, all things being equal, the most natural prosody for the structure. With respect to reanalysis of difficult syntactic structures, Bader (1998) proposed the Prosodic Constraint on Reanalysis, which states that syntactic reanalysis is more difficult if the prosody must be reanalyzed simultaneously. As discussed at length in Breen (2014), these results suggest that suprasegmental or prosodic phrasing provides information about the input's syntactic phrasing, reinforcing the grouping of words into phrasal units (cf. Slowiaczak & Clifton, 1980). The proposal here is that a more conscious instantiation of this suprasegmental phrasing, in the form of APS, results in a more salient, resilient prosodic structure, and by association, syntactic structure.

It has also been long recognized that during silent reading, skilled readers automatically generate an "inner voice" (e.g., Breen, 2014; Edfeldt, 1960; Huey, 1908/1968; Rayner, Pollatsek, Ashby, & Clifton, 2012). Consistent with this observation, eye movement data (Ashby & Clifton, 2005) and event-related potential (ERP) data (Ashby, Sanders, & Kingston, 2009) provide evidence that rough phonological representations are included in inner speech during silent reading, as well as evidence that

readers create an implicit metrical structure during silent reading (Breen & Clifton, 2011). Under certain conditions, this basic or default prosodic representation can be elaborated, such that it contains richer, more veridical information, i.e., a virtual auditory representation of the text approximating the representation that would result from hearing, rather than reading the text. Evidence for this phenomenon, called *auditory perceptual simulation* (APS), is both behavioral and neurological. As far back as 1977, Kosslyn and Matt reported that the prose of a purported "slow speaker" was likely to be read more slowly than the prose of a purported "fast speaker," a result that has since been replicated in several ways (e.g., Alexander & Nygaard, 2008; Drumm & Klin, 2011). Also, specific speakers' voices are simulated under certain conditions during silent reading (Filik & Barber, 2011; Kurby, Magliano, & Rapp, 2009).

These findings have been interpreted within embodied cognition theory, which proposes that semantic meaning is derived from the activation of perceptual features that are associated with physical experiences and sensations (Barsalou, 1999; Barsalou, & Hale, 1993; Glenberg, & Robertson, 2000; Zwaan, 2004, 2008; Zwaan, & Radvansky, 1998). Studies have demonstrated that language processing elicits perceptual simulation: readers perceptually simulate the motion, action, odor, and other information described in text during reading (Bergen & Wheeler, 2010; Glenberg & Kaschak, 2002; Hubbard, 2010; Louwerse & Connell, 2011; Louwerse & Jeuniaux, 2010; Pecher, Zeelenberg, & Barsalou, 2003, 2004; Pecher & Zwaan, 2005; Solomon & Barsalou, 2004). Thus, during language processing, readers "see" the depicted objects and events, "smell" the described odor, "act" the described motion, and, importantly for the present study, "hear" the depicted voice in their minds.

More recently, eye-tracking experiments have provided more fine-grained results confirming that not only global reading rates, but also discrete eye fixation measures reflect properties of depicted speech in texts. Both Yao and Scheepers (2011) and Stites, Luke, and Christianson (2013) showed that direct quotes, but not indirect quotes, were read more quickly if they were described as having been said by a fast talker (Yao & Scheepers) or simply as having been said quickly, as in (4), from Stites et al.

(4) David walked into the room and said quickly (slowly), "I finally found my car keys!"

Yao et al. (2011) used fMRI to show a significant boost in brain activity in auditory cortices as people read direct quotes. Assuming that the vast literature on "inner voice" cited above is not wrong, and that basic, default phonological and prosodic representations are indeed generated naturally during silent reading, the additional effects observed in these eye tracking and fMRI studies can be taken as evidence of an elaborated, prosodically richer representation–APS–which can be generated under certain circumstances.

The study reported here is the first to explore APS as a potential strategy to bolster reading comprehension. In particular, we examined whether cuing readers to undertake APS during reading of difficult sentences such as in (2) above would reduce the likelihood that readers would end up with "good enough" interpretations of such sentences. As discussed above, earlier evidence supports the connection between "inner voice" and reading comprehension, showing that suppressing the inner voice hampers

comprehension (Slowiaczek & Clifton, 1980) and interferes with readers' ability to notice anomalous words or syntax (Baddeley, Eldridge, & Lewis, 1981). On the other hand, some other work suggests that suppression or lack of inner voice affects only verbatim memory for text (Levy, 1975). No one to our knowledge has examined APS as a more elaborated "inner voice" with respect to its effect on reading comprehension.

In order to investigate the effectiveness of inducing APS to help readers avoid misinterpretation, we developed a novel paradigm. In it, participants were familiarized with the voices of two speakers, one a native American English speaker, and one a non-native American English speaker. These speakers differed in accent and, crucially, speech rate, with the non-native speaker speaking more slowly. Then, prior to silently reading each stimulus sentence, participants were cued with pictures or names of these speakers and instructed to imagine that the given speaker was saying the given sentence. Participants' eye movements were tracked as they read sentences that were manipulated with respect to their structure (subject-relative clause vs. object-relative clause) and plausibility of thematic roles (plausible, implausible). After each sentence had been read, participants' comprehension was probed using a paraphrase verification task (Kim & Christianson, 2013). An example item in all four conditions, and the paraphrase verification probe (the correct answer to which was always TRUE for the test items), are provided in (5).

(5) a. The bird that ate the worm was small. (subject-relative, plausible)

b. The worm that ate the bird was small. (subject-relative, implausible)

c. The worm that the bird ate was small. (object-relative, plausible)

d. The bird that the worm ate was small. (object-relative, implausible)

e. The bird/worm ate the worm/bird. The bird/worm was small. (T/F)

Three experiments were conducted. In Experiment 1, participants read sentences such as in (5) while their eyes were being tracked without any cue to perform APS. This experiment established baseline normal silent reading rates, patterns, and comprehension for items of this sort. In Experiment 2, participants were cued prior to each sentence to perform APS with pictures of "speakers," and in Experiment 3, participants were cued with recordings of the "speakers" saying their names. A between-subjects design was used to compare APS manipulations across experiments to non-APS normal silent reading. The rationale for this design was based on logistical concerns. If APS vs. non-APS reading conditions were included in the same experiment, participants might be likely to maintain whichever reading mode was used first throughout the experiment, and differences between APS reading and non-APS could be washed out. Counterbalancing the order of the reading modes across participants would not have eliminated this concern.

We predicted that the speech rates of the "speakers" would affect the reading speeds of participants who performed APS: simulating the native speaker would result in faster reading times than simulating the non-native speaker. We also predicted that inducing readers to perform APS would improve comprehension compared to the non-APS baseline, especially in the condition in which the most "good enough"-driven misinterpretation was predicted to occur, i.e., in implausible object-relative sentences. We were not certain how APS would affect reading speed compared to normal silent reading without APS, however.

Experiment 1

In Experiment 1, we examined how readers read and process sentences such as (5) silently, without being cued to initiate APS. The purpose of this experiment was to establish baselines for reading speed, eye movement patterns, and comprehension.

Method

Participants

Fifty-three native English speakers with normal or corrected-to-normal vision were recruited through the Educational Psychology subject pool at University of Illinois at Urbana-Champaign. These participants did not participate in either of the other two experiments. A survey was used to collect participants' language and education backgrounds, ensuring the relative homogeneity of the sample in these aspects. Participants received either one research credit or $7 for their participation. Thirteen participants' data were excluded due to experimental errors. The data from the remaining 40 participants were analyzed.

Materials

Forty-eight target sentences were used in Experiment 1 (and the same sentences were used in Experiments 2-3, see Appendix A for the target sentences). One hundred thirty filler sentences were included, consisting of a mix of complex and simple sentences, all of which were grammatical. As described above, the target sentences were manipulated with respect to structure and plausibility, resulting in a 2(subject-relative, object-relative) x 2(plausible, implausible) fully-crossed, within-participants and within-items design. Materials were distributed pseudo-randomly for each participant across four lists according to a Latin square, so that every participant saw each item only once.

One-half of the items were converted from the passives used in Ferreira (2003) and Christianson et al. (2010) into relative clauses by Lim and Christianson (2013b). The other half were created by Lim and Christianson. Plausibility of thematic role assignment, for example whether it is more likely for a bird to eat a worm or a worm to eat a bird, was normed by Ferreira (2003) for half the items, and Lim and Christianson (2013b) for the other half of the items.

There is a large psycholinguistic literature detailing the increased processing difficulty of object-relative clauses compared to subject-relative clauses and several competing theories attempting to account for this processing asymmetry, none of which are relevant to the current investigation (e.g., De Villiers, Tager, Flusberg, Hakuta, & Cohen, 1979; Gennari & MacDonald, 2008; Gibson, 1998; Gordon, Hendrick, & Johnson, 2004; Gordon, Hendrick, & Levine, 2002; Gordon, Hendrick, Johnson, & Lee, 2006; Mecklinger, Schriefers, Steinhauder, & Friederici, 1995; Traxler, Morris, & Seely, 2002; Van Dyke, & McElree, 2006; Weckerly & Kutas, 1999; Wells, Christiansen, Race, Acheson & MacDonald, 2009). What is relevant, however, is the novel finding by Ferreira (2003) that syntactic structures that are more difficult to process –described as having noncanonical word order – are likely to be misinterpreted, especially when those structures communicate meanings that are implausible. For example, Ferreira showed that native English speakers are more likely to misinterpret implausible passive sentences like *The bird was eaten by the worm* than plausible passive sentences like *The worm was eaten by the bird* or implausible active sentences like *The worm ate the bird*. We therefore predicted that reading speeds, eye movement patterns, and comprehension accuracy should all be most impacted in the implausible object-relative condition.

Structure and plausibility should have independently observable effects in plausible object-relatives and implausible subject-relatives. The plausible subject-relative condition should be the easiest by all measures, and served as the baseline for comparison.

Apparatus

Eye movement data and offline comprehension data were collected using an SR Research EyeLink 1000 remote desktop eye tracker running SR Research Experiment Builder software. A chin rest and forehead rest were used to stabilize participants' heads. Sentences were presented in black Courier New monotype font (14pt) on a white background on the computer monitor, which was approximately 70cm away from the participants. At this distance, 1° of visual angle subtended approximately 3 characters. All viewing was binocular, but data were recorded from only the right eye.

Procedure

The entire experimental session lasted on average 50 minutes. Participants provided informed consent and were then calibrated on the eye tracker using a nine-point calibration procedure. After a six-item practice session, participants were asked to read each sentence silently before completing a paraphrase verification task. After the eye-tracking portion of the experiment, the social attractiveness survey was administered on a different computer in the same room.

Experiment 1 Results

Fixations and sentence reading time data were trimmed before data analysis. Fixation times shorter than 80ms and longer than 1200ms were excluded. Sentence reading times were log transformed and scaled to reduce collinearity. Remaining data points were examined and trimmed within each condition: reading times greater or

smaller than two standard deviations from the mean of that particular condition were deleted (approximately 3.6% of the data in total). Reading time data for the entire sentence and four specific regions of interest were analyzed using linear mixed-effects models (Baayen, 2008); logit mixed models were applied to examine binomial response accuracy data (Jaeger, 2008). All analyses were performed using maximal random effects structures (Barr et al., 2013), i.e., with random slopes and intercepts for participants and items. These maximal random effects models were compared to models with only random intercepts; in the interest of space, we report below (for all experiments) the full maximal random effects models only in cases in which the addition of the random slopes changed the results.

Sentence reading time

A stepwise selection procedure was applied to determine the best-fitted linear mixed-effects model, and only factors that were significant or nearly so, or which participated in interactions, were included in the final, best-fitted model. Model output tables for all analyses in all experiments can be found in the supplemental online materials. The best-fitted model revealed that syntactic structure (t = -3.26, $p < 0.001$), plausibility (t = -4.54, $p < 0.001$), and trial presentation order (t = -6.14, $p < 0.001$) had significant effects on sentence reading time without any interactions; subjects and items were significant random intercepts and slopes in the model. Trial presentation order was included to account for changes in reading speed as participants progressed through the experiment (i.e., practice effects). Participants read plausible sentences faster than implausible sentences, and read SRCs faster than ORCs, as expected. Across the experiment, participants read sentences that were presented later faster than those

presented earlier. Descriptive statistics for sentence reading time are presented in Table 2.1.

Table 2.1

Means and Standard Deviations of Sentence Reading Times (msec.) and Response Accuracy in Three Experiments

Experiment	"Speaker"	Structure	Sentence reading time		Accuracy	
			Plausible	Implausible	Plausible	Implausible
1	No APS	Subject-RC	4071.77 (2124.4)	4679.54 (2525.83)	0.94 (0.24)	0.85 (0.36)
		Object-RC	4376.42 (2290.88)	4950.72 (2679.64)	0.86 (0.35)	0.58 (0.49)
2	Native	Subject-RC	3576.54 (1924.85)	3864.53 (2194.14)	0.96 (0.19)	0.95 (0.22)
		Object-RC	4196.10 (2244.21)	4603.20 (2758.85)	0.88 (0.32)	0.70 (0.46)
	Non-native	Subject-RC	3933.12 (2647.61)	4010.09 (2060.25)	0.95 (0.21)	0.92 (0.28)
		Object-RC	4546.79 (2528.96)	4800.14 (2850.15)	0.87 (0.34)	0.69 (0.46)
3	Native	Subject-RC	3884.62 (1934.45)	4100.96 (2322.18)	0.94 (0.23)	0.96 (0.20)
		Object-RC	4226.56 (2168.60)	4529.98 (2478.67)	0.90 (0.30)	0.73 (0.45)
	Non-native	Subject-RC	3970.37 (1968.32)	4124.54 (1882.06)	0.95 (0.22)	0.94 (0.23)
		Object-RC	4482.30 (2311.70)	4642.13 (2348.27)	0.88 (0.33)	0.74 (0.44)

Interest area analyses

The four specific interest areas were: the first noun (FN) region, which was the subject noun of the main clause; the relative clause verb region (RVC); the second noun (SN) region, which was the noun within the relative clause; and the main clause (MC) region, consisting of the remainder of the sentence. Reading times for the words within the MC region were summed to create a single dependent variable. An example of the specific regions is presented in (6).

(6) The | bird | that | ate | the | worm | was small.
FN RCV SN MC

Seven eye movement measures were included in the interest area analyses. *First*

fixation duration is the duration of the first fixation made in the target region. *Gaze duration* is the summed duration of all fixations in the target region from when it is first fixated until the eyes move to another region in either direction. When there is only one fixation on a region, first fixation duration equals gaze duration. *Go-past time*, also called regression path duration, is the total time of all fixations in a region from when it is first entered until it is exited to the right (or until the button is pressed to move on to the comprehension probe, in the case of the MC region), including fixations on previous regions that are refixated after regressive saccades. *Total reading time* is the summed duration of all fixations in the target region during the trial. *Regressions-in* reflects whether a regression was made into the target region from a later part of the sentence. *Regressions-out* reflects whether a regression was made out of the target region to an earlier part of the sentence. *Skipping* shows whether the region was fixated or not. The four duration variables are recorded and reported descriptively in milliseconds, while the other three are binomial (1 if a regression in/out was made or a region was skipped, 0 if otherwise). In the First Noun region, we only examined gaze duration, total reading time, and regressions-in. In the other three regions, we examined all seven measures. See the descriptive data in Table 2.2.

Table 2.2

Means and Standard Deviations of First Fixation Duration (FFD), Gaze Duration (GD), Go-past time (GP), Total Times (TT), Regressions-in (RI), Regressions-out (RO), and Skipping (SK) on Four Interest Areas in Three Experiments

Expt	Region	Speaker	Structure	Plausibility	FFD	GD	GP	TT	RI	RO	SK
E1	FN	N/A	SRC	plaus	---	302.87(224.91)	---	501.32 (436.61)	0.58 (0.49)	---	---
				implaus	---	303.37(197.17)	---	611.6 (523.88)	0.70 (0.46)	---	---

Table 2.2 (cont.)

Expt	Region	Speaker	Structure	Plausibility	FFD	GD	GP	TT	RI	RO	SK
E1	FN	N/A	ORC	plaus	---	287.79 (171.6)	---	548.35 (477.2)	0.65 (0.48)	---	---
				implaus	---	299.91 (192.5)	---	612.07 (493.08	0.70 (0.46)	---	---
	RCV		SRC	plaus	237.01 (96.81)	284.17 (151.4)	377.83 (247.96)	517.73 (423.85)	0.33 (0.47)	0.21 (0.41)	0.14(0. 34)
				implaus	225.87 (95.37)	279.08(167. 00)	408.41 (376.10)	606.42 (480.76)	0.43 (0.50)	0.23 (0.42)	0.15(0. 35)
			ORC	plaus	251.01(103. 99)	325.36(197. 35)	454.13 (453.90)	599.29 (468.67)	0.32 (0.47)	0.15 (0.35)	0.10(0. 30)
				implaus	252.29 (98.24)	319.83(172. 87)	479.24 (616.44)	675.17 (493.39)	0.29 (0.45)	0.16 (0.37)	0.09(0. 29)
	SN		SRC	plaus	224.95 (79.43)	265.89(136. 31)	370.81 (348.67)	444.56 (400.91)	0.29 (0.45)	0.14 (0.35)	0.20(0. 40)
				implaus	232.98 (85.64)	290.95(156. 66)	430.89 (495.28)	543.93 (455.22)	0.29 (0.45)	0.19 (0.40)	0.15(0. 36)
			ORC	plaus	220.19 (79.75)	265.10(141. 93)	397.24 (360.70)	494.22 (414.18)	0.33 (0.47)	0.20 (0.40)	0.18(0. 38)
				implaus	220.21 (81.99)	257.69(135. 06)	372.27 (304.78)	584.15 (523.41)	0.38 (0.49)	0.19 (0.39)	0.14(0. 35)
	MC		SRC	plaus	268.03(134. 18)	796.73(480. 06)	2030.02(1468 .39)	1158.30(653 .8)	0	0.69(0. 46)	0.02(0. 13)
				implaus	277.36(147. 01)	727.23(412. 76)	2454.48(1793 .78)	1267.82(766 .49)	0	0.80(0. 40)	0.02(0. 14)
			ORC	plaus	269.02(136. 88)	747.4 (493.36)	2213.54(1648 .29)	1148.64 (738.72)	0	0.76 (0.43)	0.03(0. 16)
				implaus	280.23(163. 03)	719.00(440. 15)	2600.80(1980)	1268.25(825 .71)	0	0.77(0. 42)	0.01(0. 11)
E2	FN	Native	SRC	plaus	---	275.37(142. 06)	---	467.32 (390.32)	0.57 (0.50)	---	---
				implaus	---	275.00(158. 53)	---	499.34 (417.1)	0.65 (0.48)	---	---
			ORC	plaus	---	267.86(158. 08)	---	542.21 (440.69)	0.70 (0.46)	---	---
				implaus	---	264.40(148. 81)	---	560.92 (509.98)	0.69 (0.46)	---	---
		Non-native	SRC	plaus	---	284.69(166. 19)	---	508.31 (438.29)	0.61 (0.49)	---	---
				implaus	---	279.95(162. 47)	---	532.22 (420.52)	0.66 (0.48)	---	---
			ORC	plaus	---	294.42(206. 29)	---	644.34 (569.16)	0.73 (0.44)	---	---
				implaus	---	277.16(153. 15)	---	609.00 (488.71)	0.70 (0.46)	---	---
	RCV	Native	SRC	plaus	216.29 (79.99)	251.85(128. 33)	357.99 (269.61)	435.85 (378.59)	0.35 (0.48)	0.22 (0.41)	0.22(0. 41)
				implaus	217.66 (81.94)	259.37(127. 12)	341.78 (236.38)	494.64 (437.51)	0.39 (0.49)	0.18 (0.38)	0.19(0. 39)
			ORC	plaus	234.22 (89.28)	286.14(140. 44)	407.59 (378.75	514.84 (358.46)	0.27 (0.45)	0.16 (0.37)	0.11(0. 32)
				implaus	238.76(101. 27)	283.06(143. 06)	411.72 (512.74)	575.32 (459.01)	0.31 (0.46)	0.14 (0.35)	0.10(0. 29)
		Non-native	SRC	plaus	219.38 (89.60)	268.8 (142.22)	386.55 (292.45)	517.37 (421.58)	0.37 (0.48)	0.20 (0.40)	0.16 (0.37
				implaus	219.62 (88.89)	262.52(135. 13)	356.40 (304.97)	504.72 (392.3)	0.40 (0.49)	0.17 (0.37)	0.16(0. 37)
			ORC	plaus	246.19 (100.4)	300.15(160. 03)	413.97 (420.41)	580.87 (418.01)	0.32 (0.47)	0.14 (0.35)	0.10(0. 31)
				implaus	242.02 (89.75)	302.99 (179.7)	401.42 (342.87)	627.35 (486.14)	0.36 (0.48)	0.14 (0.35)	0.14(0. 34)
	SN	Native	SRC	plaus	219 (83.86)	254.77(123. 32)	254.77 (123.32)	406.35(334. 69)	0.25(0. 43)	0.25(0. 43)	0.14(0. 35)
				implaus	223.83 (88.48)	260.35(122. 74)	260.35 (122.74)	450.39(374. 41)	0.28(0. 45)	0.28(0. 45)	0.17(0. 37)
			ORC	plaus	210.87 (80.00)	239.4 (116.18)	239.4 (116.18)	450.81(396. 23)	0.32(0. 47)	0.32(0. 47)	0.20(0. 40)
				implaus	211.74 (83.16)	238.86(118. 63)	238.86 (118.63)	532.35(532. 07)	0.33(0. 47)	0.33(0. 47)	0.19(0. 39)

Table 2.2 (cont.)

Expt	Region	Speaker	Structure	Plausibility	FFD	GD	GP	TT	RI	RO	SK
E2	SN	Non-native	SRC	plaus	226.56 (88.2)	260.48(128.42)	260.48 (128.42)	420.30(351.6)	0.25(0.43)	0.25(0.43)	0.18(0.39)
				implaus	226.36 (81.78)	261.04(128.73)	261.04 (128.73)	437.85(343.63)	0.23(0.42)	0.23(0.42)	0.16(0.37)
			ORC	plaus	213.91(84.93)	239.31(116.29)	239.31 (116.29)	484.09((449.11)	0.35(0.48)	0.35(0.48)	0.21(0.41)
				implaus	215.02 (94.53)	244.76(141.67)	244.76 (141.67)	519.26(440.69)	0.37(0.48)	0.37(0.48)	0.20(0.40)
	MC	Native	SRC	plaus	238.19(104.79)	779.34(479.12)	1878.15(1374.34)	1088.18(647.85)	0	0.73(0.44)	0.00(0.06)
				implaus	233.91(86.60)	741.79(484.83	2155.13(1632.20)	1154.59(704.83)	0	0.81(0.39)	0.02(0.14)
			ORC	plaus	244.69(116.11)	732.19(507.92)	2230.53(1632.53)	1159.61(746.96)	0	0.82(0.38)	0.01(0.10)
				implaus	243.75(118.44)	706.33(455.97)	2489.99(2011.95)	1179.43(720.92)	0	0.83(0.37)	0.01(0.08)
		Non-native	SRC	plaus	236.37(94.93)	790.30(518.91)	2096.33(1549.82)	1180.38(801.48)	0	0.79(0.41)	0.01(0.12)
				implaus	238.03(98.73)	785.94(546.26)	2129.51(1434.96)	1168.66(762.38)	0	0.80(0.40)	0.01(0.12)
			ORC	plaus	248.38(110.87)	767.77(518.53)	2490.54(1769.75)	1223.15(799.41)	0	0.85(0.36)	0.01(0.12)
				implaus	246.53(116.71)	718.64(470.74)	2703.80(2025.52)	1268.80(856.68)	0	0.85(0.35)	0.01(0.10)
E3	FN	Native	SRC	plaus	---	334.17(293.23)	---	477.25 (504.10)	0.30 (0.46)	---	---
				implaus	---	329.48(294.37)	---	489.83 (493.40)	0.31 (0.46)	---	---
			ORC	plaus	---	357.52(307.73)	---	594.07(536.38)	0.33 (0.47)	---	---
				implaus	---	357.54(305.67)	---	634.11 (583.07)	0.34(0.47)	---	---
		Non-native	SRC	plaus	---	342.07(319.82)	---	482.89 (512.52)	0.29 (0.46)	---	---
				implaus	---	335.93(302.98)	---	501.71 (494.34)	0.32 (0.47)	---	---
			ORC	plaus	---	365.97(322.77)	---	630.26 (572.35)	0.33 (0.47)	---	---
				implaus	---	366.22(324.07)	---	653.24 (584.74)	0.35 (0.48)	---	---
	RCV	Native	SRC	plaus	216.43(83.44)	334.17(293.23)	327.18 (212.13)	477.25 (504.10)	0.36 (0.48)	0.17 (0.37)	0.20(0.40)
				implaus	216.30(84.00)	329.48(294.37)	332.54 (234.26)	489.83 (493.40)	0.45 (0.50)	0.17 (0.38)	0.20(0.40)
			ORC	plaus	219.09(89.68)	357.52(307.73)	374.18 (268.17)	594.07 (536.38)	0.28 (0.45)	0.13 (0.34)	0.11(0.32)
				implaus	222.16(91.11)	357.54(305.67)	373.11 (314.00)	634.11 (583.07)	0.28 (0.45)	0.11 (0.31)	0.11(0.31)
		Non-native	SRC	plaus	216.08(82.68)	342.07(319.82)	359.37 (263.87)	482.89 (512.52)	0.36 (0.48)	0.19 (0.39)	0.19(0.39)
				implaus	219.59(87.05)	335.93(302.98)	370.09 (288.42)	501.71 (494.34)	0.45 (0.50)	0.20 (0.40)	0.19(0.39)
			ORC	plaus	222.45(88.41)	365.97(322.77)	381.20 (308.34)	630.26 (572.35)	0.31(0.46)	0.13 (0.34)	0.12(0.32)
				implaus	224.64(96.23)	366.22(324.07)	431.33 (416.00)	653.24 (584.74)	0.33 (0.47)	0.16 (0.37)	0.10(0.30)
	SN	Native	SRC	plaus	215.90 (84.72)	246.68(133.30)	306.39 (238.65)	371.46 (317.48)	0.26 (0.44)	0.13 (0.33)	0.20(0.40)
				implaus	224.18 (87.92)	260.53(126.47)	340.42 (273.89)	414.71 (330.45)	0.28 (0.45)	0.13 (0.34)	0.19(0.39)
			ORC	plaus	207.68 (68.73)	235.6 (116.43)	343.00 (278.33)	428.02 (377.67)	0.36 (0.48)	0.20 (0.40)	0.21(0.41)
				implaus	202.76 (70.47)	228.89(102.80)	332.20 (266.70)	453.22 (417.31)	0.36 (0.48)	0.17 (0.38)	0.24(0.43)

Table 2.2 (cont.)

Expt	Region	Speaker	Structure	Plausibility	FFD	GD	GP	TT	RI	RO	SK
E3	SN	Non-native	SRC	plaus	211.35 (79.10)	245.40(122.89)	320.69 (286.86)	390.46 (324.50)	0.26 (0.44)	0.12 (0.33)	0.19(0.39)
				implaus	226.65 (93.73)	265.57(143.65)	359.03 (310.79)	416.77 (306.91)	0.29 (0.45)	0.17 (0.37)	0.19(0.39)
			ORC	plaus	217.34 (82.91)	247.81(121.74)	357.79 (287.73)	457.38 (420.83)	0.37 (0.48)	0.19 (0.39)	0.26(0.44)
				implaus	211.52(104.34)	246.63(154.31)	355.24 (315.84)	486.46 (466.98)	0.38 (0.49)	0.20 (0.40)	0.20(0.40)
	MC	Native	SRC	plaus	240.24(101.58)	784.76(446.47)	2020.84(1439.80)	1184.08(763.30)	0	0.79(0.41)	0.01(0.08)
				implaus	247.03(107.08)	765.25(489.10)	2098.86(1459.36)	1193.75(708.62)	0	0.81(0.40)	0.00(0.06)
			ORC	plaus	247.94(122.15)	763.77(475.85)	2224.33(1571.36)	1190.45(698.24)	0	0.83(0.38)	0.01(0.09)
				implaus	253.26(122.15)	744.92(483.50)	2406.14(1890.92)	1180.19(714.78)	0	0.83(0.38)	0.00(0.06)
		Non-native	SRC	plaus	242.92(102.36)	812.41(527.39)	2033.68(1467.43)	1193.44(799.43)	0	0.80(0.40)	0.01(0.10)
				implaus	245.73(111.85)	788.91(505.09)	2141.97(1398.12)	1212.98(705.02)	0	0.83(0.38)	0.01(0.10)
			ORC	plaus	245.05(103.15)	770.12(527.61)	2365.65(1724.80)	1234.42(813.72)	0	0.83(0.38)	0.01(0.12)
				implaus	257.92(126.44)	764.53(531.91)	2425.81(1784.77)	1249.99(743.32)	0	0.85(0.36)	0.00(0.05)

In the first noun region, plausibility was a significant predicator of total reading time (t = -4.81, $p < 0.001$). Implausible sentences led to longer total reading times than plausible ones on the first noun. No factors were significant in either gaze duration or regressions-in.

In the relative clause verb region, structure significantly influenced first fixation durations and gaze durations, and marginally affected go-past times. Readers had longer first fixation durations (t = -7.17, $p < 0.001$), gaze durations (t = -4.54, $p < 0.001$), and go-past times (t = -1.82, $p < 0.1$) in ORCs than in SRCs. Total reading times were significantly influenced by plausibility (t = -2.52, $p < 0.01$) and structure (t = -4.01, $p < 0.001$) without any interaction (ORC > SRC, implausible > plausible). Both conditions marginally affected regressions-in (plausibility: z = -1.8, $p < 0.1$; structure: z = -1.77, $p < 0.1$) and regressions-out in (plausibility: z = -1.8, $p < 0.1$; structure: z = -1.77, $p < 0.1$).

Only structure ($z = 3.25$, $p < 0.01$) significantly affected skipping on this region, with more skipping in SRCs.

In the second noun region, plausibility significantly predicted first fixation duration ($t = -2.21$, $p < 0.05$), such that readers spent less time reading this region when the sentence was plausible. Plausibility also significantly influenced gaze durations ($t = -2.25$, $p < 0.05$) and go-past times ($t = -2.35$, $p < 0.05$). Plausible sentences yielded shorter gaze durations and faster go-past times. Both structure ($t = -5.26$, $p < 0.001$) and plausibility ($t = -2.50$, $p < 0.05$) significantly affected total reading times in the same directions. Logit mixed-effects models demonstrated that plausibility was a significant predicator of regressions-in ($z = -2.18$, $p < 0.05$), regressions-out ($z = -2.18$, $p < 0.05$), and skipping rate ($z = 2.67$, $p < 0.05$) on the second noun region. Readers were less likely to make regressions-out or in, but more likely to skip the second noun, in plausible sentences.

In the main clause region**,** structure ($t = -2.15$, $p < 0.05$) and plausibility ($t = -2.31$, $p < 0.05$) influenced go-past times with no interaction. Plausibility ($t = -2.04$, $p < 0.05$) was a significant predictor of gaze durations. Plausibility also affected total reading times ($t = -4.14$, $p < 0.001$), which were inflated in implausible sentences. There were no other significant effects on this region.

Response accuracy data

The means and standard deviations of response accuracy in Experiment 1 are presented in Table 2.1. Using the same model selection procedure and criteria as with the linear mixed models, the final best-fitted logit model included structure ($z = 7.12$, $p < 0.001$), plausibility ($z = 7.17$, $p < 0.001$), and trial presentation order ($z = 3.37$, $p <$

0.001), as main effects and random intercepts and slopes for items and participants. There was a two-way interaction between structure and plausibility (z = -3.18, p < 0.001); readers were most likely to misinterpret sentences when they were implausible ORCs. Participants were more likely to interpret items accurately later in the experiment.

Experiment 2

Experiment 2 was designed to test the hypothesis that APS would improve comprehension accuracy of sentences in which semantic plausibility information was most likely to interfere with syntactic structure-building (i.e., implausible object-relatives), which were observed in Experiment 1 to pose the most problems for interpretation, consistent with previous work. We reasoned that APS should buttress the syntactic representation with an elaborated prosodic representation, reducing interferential effects of the semantics. A novel paradigm was developed, in which participants in a silent reading eye-tracking experiment were shown a picture of one of two "speakers." It was hypothesized that these pictures, in combination with initial instructions asking participants to imagine the voices of the "speakers" as they read, would serve as effective prompts for APS.

Methods

Participants

Ninety-eight native English speakers with normal or corrected-to-normal vision were recruited through the Educational Psychology subject pool at University of Illinois at Urbana-Champaign. These participants did not participate in either of the other two experiments. Participants received either one research credit or $7 for their participation. Eight participants' data were not recorded completely due to experimenter error and ten

participants' data were excluded because their experimental sessions were disrupted. The data from the remaining 80 participants were analyzed.

Materials

In order to induce participants to perceptually simulate the speech of distinct speakers, we provided them with examples of both a native and a non-native English speaker's voice. Four 500-word passages were used as the texts for the speech. These narrative passages, balanced for difficulty (Flesch-Kinkaid reading level: text 1= 11.8, text 2= 12.2, text 3=12.5, text 4 =12), were selected from *National Geographic* on topics such as organic fruit. One female native American English speaker and one female Chinese non-native American English speaker of approximately the same age (early 20s) were recruited to read these passages. Reading speeds were compared using a paired *t*-test, confirming that the native English speaker read the passages significantly faster than the non-native English speaker ($t(3) = -72.38$, $p < 0.001$)[2].

The photos of the speakers were not of the actual speakers, but rather stock, non-copyrighted photos found on the Internet and controlled as closely as possible for age, socio-economic status, and expression. The photo of the native speaker was of a blonde, Caucasian woman who appeared to be in her early 30s. The photo of the non-native English speaker was of a Chinese woman who appeared to be in her early 30s. The pictures showed the women from the shoulders up, and both women were dressed in a business jacket and blouse.

The same target sentences from Experiment 1 and ninety-six filler sentences were

[2] The time duration for native speech: text 1 = 2min 50sec; text 2 = 2min 57sec; text 3 = 2min 49sec; text 4 = 2min 57sec. The time duration for non-native speech: text 1 = 4min 08sec; text 2 = 4min 18sec; text 3 = 4min 08sec; text 4 = 4min 20sec.

used in Experiment 2. All sentences had quotation marks to remind participants that they should read them as direct quotations (Stites et al., 2013; Yao & Scheepers, 2011). In addition to the reading task, a social attractiveness scale questionnaire was used to measure participants' attitudes towards the native and non-native speakers and speech. Based on previous studies (Callen, Callois, & Forbes, 1983; Edward, 1977; Giles, 1972; Gass & Varonis, 1984; Vornik, Sharman, & Garry, 2003), fourteen factors were included in the questionnaire, such as comprehensibility, accent, and trustworthiness. (See Appendix B for the complete survey.) Participants rated each speaker on the 14 attributes on a 1-7 Likert scale. The social attractiveness data were collected to ensure that sociolinguistic biases against speakers/accents were not the source of the effects that were observed. Each participant filled out the questionnaire after the eye-tracking portion of the experiment.

Apparatus

The same equipment from Experiment 1 was used in Experiment 2. A separate desktop computer in the same room was used to collect the survey results.

Procedure

The entire experimental session lasted on average 50 minutes. Calibration followed the same procedure as in Experiment 1. Next, the participant listened to two recordings, one by the native English speaker, the other by the non-native English speaker, while the corresponding photograph was displayed on the computer screen. The order of recordings was counterbalanced across participants, with each one presented first half of the time. Participants then read and responded to the six practice items and, after any questions, proceeded to read and respond to the experiment items. One or the other

“speaker” photograph was randomly presented an equal number of times prior to each sentence read in the experiment to cue the “speaker's” voice as participants read the sentences. One concern was that participants' memories for the speakers' voices would fade as the experiment progressed. To address this concern, another recording set of the native and non-native speaker reading similar texts aloud was played halfway through the experiment to remind participants what they sounded like. After the second reading portion of the experiment, the two social attractiveness scale questionnaires (one for each speaker) were administered to measure participants' attitudes towards the speakers.

Experiment 2 Results

Following the same data trimming procedure as in Experiment 1, approximately 3.5% of the data were removed prior to analysis.

Sentence reading time

The same procedure was applied to determine the best-fitted linear mixed-effects model. The final model revealed significant effects of structure (t = -8.59, $p < 0.001$), plausibility (t = -4.08, $p < 0.001$), “speaker” (t = -5.55, $p < 0.001$), and trial presentation order (t = -5.53, $p < 0.001$) on total sentence reading time without any interactions. Participants took longer to read ORCs than SRCs. Implausible sentences were read more slowly than plausible sentences. Sentences attributed to the non-native “speaker” yielded longer reading times than sentences attributed to the native “speaker.” Regardless of structure, plausibility, or “speaker,” participants read sentences presented later faster than sentences presented earlier.

To ensure that participants' memories for the “speakers'“ voices did not fade over time, we presented the reading materials in two blocks, separated by a re-presentation of

another voice sample for each speaker. Interestingly, and consistent with APS effects, we found that the main effect of "speaker" was lacking in post-hoc sub-analyses of the 12 items immediately preceding the re-presentation, but re-emerged in the 12 items immediately following the re-presentation. This pattern was also observed in Experiment 3. The result strongly supports the claim that readers were simulating the voices as they read. See "auditory re-representation effect in Experiment 2" in supplementary file for more information.

Interest areas analysis

In the first noun region, "speaker" condition had a significant effect on gaze duration (t = -2.8, $p < 0.01$). Readers processed the first noun much faster when the target sentence was presented as having been said by the native speaker. Structure (t = -3.67, $p < 0.001$) and "speaker" conditions (t = -4.62, $p < 0.001$) significantly affected total reading time without any interaction. Total reading times were longer in the ORC condition than in SRC. Total reading time was, overall, longer when participants were cued by the non-native speaker's photo. Structure also significantly affected regressions into the first noun region (z = -5.56, $p < 0.001$). Participants made more regressions back to the first noun in ORCs (70.47%) than they did in SRCs (62.28%).

In the relative clause verb region, first fixation durations were significantly inflated in ORCs (t = -7.40, $p < 0.001$) and following the non-native "speaker" photo (t = -1.83, $p < 0.1$). ORC structure also yielded inflated gaze durations (t = -4.97, $p < 0.001$), and go-past times (t = -3.44, $p < 0.001$). Total reading time times on the verb were also significantly greater in ORCs (t = -4.05, $p < 0.001$) and following the non-native "speaker's" photo (t = -4.58, $p < 0.001$). The plausible condition (t = -1.79, $p < 0.1$) led

to marginally shorter total reading times in this region. All three factors, syntactic structure (z = -2.15, p < 0.05), plausibility (z = 3.48, p < 0.001), and "speaker" (z = -2.13, p < 0.05), had significant effects on regressions into the relative clause verb. Participants made more regressions into the verb region in SRCs and when sentences were implausible. The non-native "speaker" also triggered more regressions into this region. Plausibility marginally predicted regressions-out (z = -1.86, p < 0.1), while syntactic structure significantly affected regressions-out (z = 3.5, p < 0.001). Participants made more regressions-out in plausible sentences than implausible ones, and fewer in ORCs than in SRCs. For skipping, only syntactic structure (z = 4.58, p < 0.001) was a significant predictor; more skipping of the relative clause verb occurred in SRCs.

In the second noun region, only sentence structure significantly affected first fixation durations (t = 4.05, p < 0.001), gaze durations (t = 3.07, p < 0.001), and total fixation times (t = -2.81, p < 0.001). First fixation durations and gaze durations on the second noun were significantly longer in SRCs. However, total reading times were shorter in SRCs than ORCs. Plausibility also significantly influenced total reading times (t = -1.85, p < 0.05). Plausible sentences yielded shorter total reading times than implausible sentences. There were significant structure effects on regressions-in (z = -4.96, p < 0.001) and regressions-out (z = -7.22, p < 0.001). Readers made fewer regressions into and out of the second noun in SRCs, demonstrating again the overall greater difficulty of processing ORCs. Structure also significantly affected skipping (z = -2.16, p < 0.05). Readers skipped this region more often in ORCs. ORC sentences yielded longer first fixation durations in the MC region than SRC sentences (t = -2.56, p < 0.01). There was a "speaker" effect on total reading time (t = -3.44, p < 0.001), such that the

native "speaker" condition yielded significantly shorter total times. There were no significant effects on gaze duration, go-past time, regressions-in, regressions-out, or skipping.

Response accuracy data

Using the same model selection procedure and criteria as with the linear mixed models, the final best-fit logit model included structure ($z = 9.53$, $p < 0.001$), plausibility ($z = 6.66$, $p < 0.001$), trial order ($z = 4.46$, $p < 0.001$), and the interaction of structure and plausibility ($z = -3.90$, $p < 0.001$) as main effects and random intercepts and slopes for participants and items. Structure, plausibility and trial order had significant effects on response accuracy with no interactions. Participants made more errors in the ORC condition than SRC condition. They were more likely to misinterpret implausible sentences than plausible sentences. Comprehension improved as the experiment progressed. "Speaker" did not have significant effect on accuracy. There was only a 1% difference in accuracy between APS of native and APS of non-native speech.

See the supplemental online materials and Zhou and Christianson (in preparation) for results and discussion of the social attractiveness survey results.

Experiment 3

Experiment 2 showed that the native and non-native speakers' photos were effective cues for readers to perceptually simulate native and non-native speech rate during silent reading. If the photos did indeed trigger APS, we reasoned that auditory cues should also be effective at triggering readers to perceptually simulate depicted speakers' voices. By way of replication and exploration of the APS effect, in Experiment

3 we replaced the photos of the speakers with recordings of them saying their names. We predicted that these recordings would also prompt APS and drive similar reading patterns to those observed in Experiment 2.

Method

Participants

Ninety-seven native English speakers with normal or corrected-to-normal vision and who did not participate in Experiment 1 or 2 were recruited through the Educational Psychology subject pool at University of Illinois at Urbana-Champaign. Participants received either one research credit or $7 for their participation. Seventeen participants' data were excluded due to bad calibration or excessive drifts. The data from the remaining 80 participants were analyzed.

Materials

The materials were identical to Experiment 2, except that the speakers' photos were not used as APS prompts prior to the presentation of each sentence. Instead, two recordings of the speakers saying their names ("Susan" or "Xiaofu") were used.

Apparatus

The same equipment from Experiment 1 was used in Experiment 3.

Procedure

All procedures were identical to Experiment 2.

Experiment 3 Results

Following the same data trimming procedure in Experiment 1 and 2, approximately 4.2% of the data were removed prior to analysis.

Sentence Reading Time

The same model-fitting procedure as in Experiments 1 and 2 was used, and the best-fitted model for total sentence reading time included plausibility (t = -4.52, p < 0.001), structure (t = -2.93, p < 0.001), "speaker" (t = -3.25, p < 0.001), and trial order (t = -7.23, p < 0.001), along with random slopes and intercepts for participants and items. Plausible sentences yielded significantly shorter total reading times than implausible sentences. SRCs required less time to read than ORCs. Participants read sentences preceded by the native speaker's name faster than sentences preceded by the non-native speaker's name.[3] Items presented later in the experiment were read faster than those presented earlier.

Interest area analysis

The same interest areas as in Experiment 1 and 2 were analyzed in Experiment 3 for the same measures.

In the first noun region, linear mixed-effects results revealed that "speaker" significantly influenced gaze durations (t = -2.27, p < 0.05) on the first noun. When the sentence was preceded by the native speaker's name, gaze durations were shorter on the first noun. Syntactic structure significantly affected the total reading times (t = -2.17, p < 0.05) on the first noun, such that the first noun was read more slowly in ORCs than in SRCs. There were no effects on regressions into this region.

In the relative clause verb region, structure significantly affected first fixation durations (t = -7.17, p < 0.001), gaze durations (t = -4.54, p < 0.001), and regressions-out (z = -3.77, p < 0.001). First fixation durations and gaze durations on the relative clause

[3] We also examined the auditory re-presentation effect in Experiment 3. The results replicated the patterns in Experiment 2. Participants' memory of the "speakers" gradually faded away as the time went on and then returned after re-presentation. See "auditory re-representation effect in Experiment 3" in the supplementary materials for details.

verb were inflated in ORCs compared to SRCs, and more regressions-out of the verb were initiated in ORCs. Structure marginally influenced go-past times (t = -1.82, $p < 0.1$). Total reading times were significantly influenced by structure (t = -4.01, $p < 0.001$) and plausibility (t = -2.52, $p < 0.001$). No interactions were significant. Structure significantly influenced the regressions-out of this region. None of the factors significantly impacted regressions into the relative clause verb region or skipping.

In the second noun region, syntactic structure and "speaker" significantly influenced first fixation durations (structure: t = 2.01, $p < 0.05$; "speaker": t =2.00, p = 0.05), and total times (structure: t = -2.85, $p < 0.001$; "speaker": t = -2.05, $p < 0.05$), and marginally affected the gaze duration (structure: t = 1.71, $p < 0.1$; "speaker": t = -1.92, $p < 0.1$). Readers displayed longer first fixation durations and gaze durations, but significantly shorter total reading times on the second noun when they encountered SRC sentences. They also had longer total reading times when the sentence was preceded by the non-native speaker's name. "Speaker" marginally affected go-past times (t = -1.71, $p < 0.1$). There was no significant effect on skipping, regressions-in or regressions-out in this region.

In the main clause region, only "speaker" marginally influenced total reading times (t = -1.89, $p < 0.1$). Readers spent more time processing this region when the sentence had been preceded by the non-native speaker's name regardless of plausibility or structure. No other measures were significant.

Response accuracy data

The best-fitted logit model included plausibility (z = 10.17, $p < 0.001$), syntactic structure (z = 6.67, $p < 0.001$), trial order (z = 2.54, $p < 0.05$), and the interaction of

plausibility and syntactic structure (z = -4.08, p < 0.001) as fixed effects, and random slopes and intercepts for participants and items. Comprehension was more accurate for plausible sentences than for implausible sentences and for SRCs than for ORCs. Items presented later in the experiment were generally comprehended more accurately than items presented earlier in the experiment. Implausible ORC sentences were harder to comprehend than plausible SRC sentences. "Speaker" was not a significant predictor of accuracy.

Comparison of all three experiments

Experiment 1 examined readers' eye movement patterns and comprehension of the target sentences in normal silent reading. The results largely replicated previous work showing that syntactic structure and plausibility affect reading times and comprehension accuracy. Object-relative clauses were read more slowly than subject-relative clauses (e.g., Gibson, 1998; Traxler, Morris, & Seely, 2002). Implausible sentences yielded longer reading times and lower accuracy than plausible ones (Christianson et al., 2010; Ferreira, 2003; Lim & Christianson, 2013b), with implausible ORCs the most likely to be misinterpreted. Experiments 2 and 3 investigated APS effects during silent reading, using different methods to encourage readers to perform APS: photos of "speakers" in Experiment 2, recordings of "speakers" saying their names in Experiment 3. Results from these two experiments indicated that both cues triggered APS effects, such that readers read faster when they perceptually simulated a native speaker's voice while processing the sentences, reflecting the speaking rates of the respective speakers.

One of the main goals of the present study, however, was to determine how APS affected comprehension, specifically, if inducing a rich prosodic contour would lead to

deeper, less good-enough syntactic processing. In order to answer this question, we compared the reading rate and comprehension accuracy data from Experiments 2 and 3 separately to the data from Experiment 1, in which readers were not prompted to engage in APS. In the APS experiments (2 and 3), half of the time, readers activated the APS of a native speaker's voice and the other half of the time they simulated the non-native speaker's voice. Thus, we compared the results from trials when participants had been cued to simulate native speech in APS experiments to the same items in the no APS data in Experiment 1, and then did the same for items in which participants had been cued to simulate non-native speech. Two new datasets were created and a new variable "experiment" was added to investigate differences across experiment (perceptual simulation of native speech vs. no perceptual simulation; perceptual simulation of non-native speech vs. no perceptual simulation). Linear mixed-effects models were built to analyze reading times and logit mixed-effects models were fit to examine accuracy data.

Native "speaker" in Experiment 2 vs. No APS in Experiment 1

Sentence reading time

The best-fitted model indicated that perceptual simulation (experiment, $t = 2.67$, $p < 0.001$), structure ($t = -6.20$, $p < 0.001$), plausibility ($t = -4.57$, $p < 0.001$), and trial presentation order ($t = -10.28$, $p < 0.001$) all significantly affected sentence reading time. The critical finding is that participants spent significantly *more* time reading sentences in Experiment 1 when they were *not* perceptually simulating native speech than participants who were simulating native speech in Experiment 2.

Response accuracy

Results showed that response accuracy when perceptually simulating native speech

in Experiment 2 ($M = 0.87$, $SD = 0.33$) was significantly higher ($z = -4.9$, $p < 0.001$) than when not simulating any voice in Experiment 1 ($M = 0.80$, $SD = 0.40$).

Non-native "speaker" in Experiment 2 vs. No APS in Experiment 1

Sentence reading times

Results for participants who perceptually simulated the non-native speaker's voice in Experiment 2 were compared with the results from Experiment 1. The best-fitted model included experiment ($t = 2.68$, $p < 0.001$), structure ($t = -6.21$, $p < 0.001$), plausibility ($t = -4.24$, $p < 0.001$), trial presentation order ($t = -10.90$, $p < 0.001$), and the interaction between structure and trial presentation order ($t = 4.40$, $p < 0.001$). Just as with the native speaker APS effects, participants spent significantly *more* time reading sentences when they were *not* perceptually simulating non-native speech.

Response accuracy

Logit mixed model results demonstrated that experiment ($z = -4.07$, $p < 0.001$), structure ($z = 9.01$, $p < 0.001$), plausibility ($z = 8.15$, $p < 0.001$), trial presentation order ($z = 4.03$, $p < 0.001$) and the interaction of structure and plausibility ($z = -2.10$, $p < 0.05$) significantly influenced response accuracy across experiments. No APS during silent reading resulted in lower accuracy than APS of non-native speech. In other words, perceptual simulation of both non-native speech and native speech improved readers' comprehension.

Native "speaker" in Experiment 3 vs. No APS in Experiment 1

Sentence reading time

The best-fitted model demonstrated that plausibility ($t = -7.94$, $p < 0.001$), structure ($t = -6.40$, $p < 0.001$), and trial order ($t = -11.26$, $p < 0.001$) were significant

predicators of sentence reading time. Experiment (native speech APS vs. no APS) was not a significant predictor, despite the numerical trend of shorter reading times in the APS condition in Experiment 3 (M = 4185.53, SD = 2245.45) compared to no APS in Experiment 1 (M = 4519.61, SD = 2435.05).

Response accuracy

APS (z = -2.98, $p < 0.001$) significantly influenced accuracy across experiments in addition to plausibility (z =11.69, $p < 0.001$), structure (z = 12.71, $p < 0.001$), trial presentation order (z = 3.31, $p < 0.001$), and the interaction of plausibility and structure (z = -4.18, $p < 0.001$). Accuracy was significantly higher when readers were perceptually simulating the native speaker's voice in Experiment 3 (M = 0.88, SD = 0.32) compared to normal silent reading in Experiment 1 (M = 0.80, SD = 0.40).

Non-native "speaker" in Experiment 3 vs. No APS in Experiment 1

Sentence reading time

Results showed that plausibility (t = -7.89, $p < 0.001$), structure (t = -6.82, $p < 0.001$), and trial presentation order (t = -11, $p < 0.001$) were significant predicators of sentence reading times. There was no effect of experiment, i.e., no significant speed difference between normal silent reading in Experiment 1 (M = 4519.61, SD = 2435.05) and perceptual simulation of the non-native speaker's voice in Experiment 3 (M = 4304.84, SD = 2152.67), again despite a numerical trend towards an APS speed advantage.

Response accuracy

Experiment (z = -2.61, $p < 0.001$), sentence structure (z = 12.04, $p < 0.001$), plausibility (z = 11.09, $p < 0.001$), trial presentation order (z = 3.69, $p < 0.001$), and the

interaction of structure and plausibility (z = -2.93, p < 0.001) significantly affected response accuracy across experiments. The results indicated that APS of non-native speech (M = 0.88, SD = 0.33) was associated with significantly increased response accuracy compared no perceptual simulation (M = 0.80, SD = 0.40).

APS (Experiments 2 and 3) vs. normal silent reading (Experiment 1) Discussion

Comparisons of total sentence reading times and response accuracy across Experiments 1 and 2 revealed that when APS was triggered by a speaker's photo, readers tended to read more quickly than under normal silent reading conditions. There are two possible explanations for differences in reading time and accuracy across these experiments. The first explanation, which we prefer, is that APS activated phonological and prosodic information that linked to the syntactic representation and strengthened it, facilitating deeper, less good-enough processing of the sentences and resulting in better comprehension. An alternative explanation is that the cues used to encourage APS drew attention to the reading task, leading to more careful reading and more cautious interpretation. We find this alternative explanation less plausible. Under this explanation, faster reading would need to be characterized as "more careful" or "more cautious," and we are not aware of any accounts of reading in which reading more quickly entails reading more carefully. Triggering APS with either the native or non-native speakers' name in Experiment 3 resulted in reading that was numerically, but not statistically, also faster than normal silent reading in Experiment 1. Comparisons between Experiment 1 and 3 also demonstrated that perceptual simulation of either native or non-native speech during silent reading led to better comprehension accuracy, which, we contend, is due to deeper sentence processing when readers are engaged in APS.

Comparison of interest areas across experiments

The semantically implausible and syntactically complex sentences used in this study often led to "good-enough" processing (Ferreira et al., 2002), in which readers reached the wrong interpretation based largely on heuristic semantic information while overlooking syntactic structure. To investigate why readers had better comprehension when they activated APS in Experiments 2 and 3, we examined how readers processed these sentences by comparing their eye-movement patterns in the four interest areas across the three experiments. Table 2. 2 illustrates the means and standard deviations of each variable in the four regions. Table 2. 3 below demonstrates the significant predicators on seven measures in the three experiments. Due to space limitations, we only discuss the eye movement patterns on the second noun region, which was consistent on the whole with the other areas.

In the second noun region, plausibility was the only significant factor that affected all seven measures in Experiment 1. In contrast, structure was the dominant factor that drove eye movements in Experiment 2. The ORC condition yielded longer first fixation durations, gaze durations, and total reading times on the second noun. It also resulted in more regressions into and out of the region as well as fewer skips. In Experiment 3, both structure and "speaker" modulated first fixation durations, gaze durations, and total reading times. "Speaker" also significantly affected go-past times on the second noun. It thus appears, again, that in Experiments 2 and 3, readers were more fully committed to the syntactic processing route when they were perceptually simulating speech during silent reading. In contrast, when not engaged in APS (Experiment 1), readers were more influenced by plausibility heuristics. This effect was not just observed in total sentence

reading times and offline measures, but also in individual interest areas and on early measures of processing. This pattern of results suggests that APS was affecting the moment-by-moment processing, not just the offline memory for the input. In other words, the processing in Experiment 1 more closely resembled good-enough processing (Christianson et al., 2001, 2006, 2010; Ferreira, 2003; Ferreira et al., 2002; Ferreira & Patson, 2007). This difference helps explain why the APS of both native and non-native speech in Experiments 2 and 3 was associated with more accurate comprehension: readers relied more on the syntactic information, which is the more reliable route to the intended meaning of the sentences. In Experiment 1, on the other hand, readers tended to pursue a good-enough interpretation based on plausibility.

Table 2.3

Main Factors that predicated the First Fixation Duration (FFD), Gaze Duration (GD), Go-past time (GP), Total Times (TT), Regressions-in (RI), Regressions-out (RO), and Skipping (SK) in the Four Interest Area Regions across Three Experiments

Region	FFD	GD	GP	TT	RI	RO	SK
FN				plaus			
RCV	struc	struc	struc,plaus, struc x plaus	struc,plaus, struc x plaus	plaus^, struc^	struc^	plaus^,struc
SN	plaus	plaus	plaus	plaus	plaus	plaus	plaus
MC		plaus	plaus, struc	plaus, struc			
FN		"speaker"		"speaker"	struc		
RCV	struc	struc	struc	struc	plaus,struc, "speaker"	struc, "speaker"	struc
SN	struc	struc		struc	struc	struc	struc
MC	struc			"speaker"			
FN		"speaker"		struc			
RCV	struc	struc	struc	struc,plaus		struc	
SN	struc, "speaker"	struc, "speaker"	struc, "speaker"	"speaker"			
MC				"speaker"			

Note. ^$p < 0.1$ (marginally significant)

General Discussion

This study aimed to investigate how auditory perceptual simulation (APS) affects language processing and comprehension during silent reading. We collected online eye movement measures and offline comprehension measures to address the following questions: 1) whether readers would perceptually simulate distinct speakers during silent reading; and 2) what the differences between perceptually simulated speech during silent reading and normal silent reading without perceptual simulation are.

Eye movement data from Experiments 2 and 3 demonstrated that a perceptually simulated “speaker's” voice (which was not physically presented during sentence reading) had significant effects on sentence reading times and response accuracy. Analyses of reading times in the last two experiments demonstrated again and again that participants read sentences that were presented as being spoken by the slower speaker (the non-native speaker) more slowly than those presented as being spoken by the faster speaker (the native speaker). We attribute this effect to APS of the respective speakers' voices during silent reading. In addition, analyses of specific regions also demonstrated that “speaker” had significant effects on earlier and later measures throughout the sentences. Furthermore, split analyses of trials immediately preceding and immediately following “refresher” recordings of native and non-native speaker speech revealed that “speaker” effects weakened over time but then reappeared after exposure to another set of recordings (see supplementary online materials). These results all support previous observations that APS of speech modulates reading processes (Kosslyn & Matt, 1977;

Stites et al., 2013; Yao & Scheepers, 2011).

The most important finding of the current study is that APS facilitated more syntactically-driven, less good-enough processing of the sentences, as evidenced by higher response accuracy in the last two experiments under perceptual simulation conditions. By comparing accuracy between perceptually simulated native and non-native speech in the last two experiments with no APS in Experiment 1, we found that readers derived significantly more accurate interpretations of the sentences when they were perceptually simulating either speaker while reading. Further analyses of the interest areas in these three experiments provided a clear explanation for the findings. Eye movement patterns were significantly affected by the syntactic structure in Experiments 2 and 3: ORC sentences produced longer early and late measures along with more regressions into and out of the four interest areas compared to SRC sentences. In contrast, in Experiment 1 plausibility was the most consistently significant predictor: when sentences were plausible, readers displayed significantly shorter first fixation times, gaze durations, go-past times, and total fixation times on the target regions, and also tended to make fewer regressions into and out of the regions. Taken together, this constellation of effects suggests that readers tended to rely on "fast and frugal" plausibility heuristics (Ferreira et al., 2002) to interpret the sentences in Experiment 1 rather than the algorithmic syntactic processing stream that was more influential in Experiments 2 and 3. Thus, the results suggest that perceptual simulation of either native or non-native speech facilitated deeper sentence processing and thereby improved comprehension.

One possible explanation of this perceptual simulation advantage is that readers activated richer phonological and prosodic information when they engaged in APS as

they read (cf. Breen, 2014). This information, such as the prosodic phrasing of phrasal units, helped solidify the syntactic structure online, and potentially in memory, and reduced the likelihood that the roles of the thematic agents and patients would be reversed. Importantly, the fact that reading speeds were faster in Experiment 2 than Experiment 1, and no slower in Experiment 3 than 1, strongly suggests that cuing APS did not simply cause participants to slow down and "read more carefully." Indeed, the more structurally focused reading in the second two experiments could be considered to be "more careful," but not because more attention was devoted to the text in general. Instead, APS appears to have focused attention on the structural properties of the text, whereas normal silent reading – which certainly also included some degree of typical "inner speech" (Rayner et al., 2012) and "implicit prosody" (Fodor, 2002), but not APS – distributed attention across both structural and, crucially, heuristic lexical-semantic sources of information (cf. Christianson et al., 2010), which are proposed to interfere with one another.

There are two alternative explanations of the results presented here, both of which are predicated on readers' uncertainty about previous material. Neither of these is in principle incompatible with the Good Enough account proposed here, but neither can account for the present results without raising questions that cannot be resolved without future research. The first is the "noisy channel" hypothesis (Gibson, Piantadosi, Brink, Bergen, Lim, & Saxe, 2013). Gibson and colleagues propose that, because meaning is conveyed through "noisy" channels, comprehenders "choose a representation that maximizes meaning recoverability" (p. 1080). In the case of the implausible but reversible relative clauses used here, the hypothesis holds that the more typical thematic

role relations between the nouns and the more canonical Agent-Patient linear order (Townsend & Bever, 2001) should be the ones readers choose under the noisy conditions of implausibility and noncanonical argument order. This account raises a number of unanswered questions, however, and also requires several assumptions. First, the language attitude surveys (see online supplementary materials) revealed that participants were very aware of the non-native speaker's accent, and generally considered the non-native speaker to be less comprehensible, less reliable, less confident, and less likeable than the native speaker. Presumably, this overall attitude would have introduced more "noise" into the channel. If participants were perceptually simulating both voices, we assume that they would also have been simulating the non-native speaker's accent to some extent, not just the speech rate. (Indeed, the difficulty of simulating the accent may partially explain the overall slower reading rate when simulating the non-native speaker's voice; current experiments (Zhou & Christianson, 2014, September) are underway to examine the factor of accent vs. speech rate in APS). The noisy channel hypothesis would predict that reading in the absence of all of this extra "noise" – i.e., in Experiment 1 – should have been faster and/or more accurate. This was clearly not the case. Perhaps, however, the addition of stronger phonological and prosodic "channels" merged with the syntactic "channel" to produce a stronger "mega-channel" and thereby reduce overall noise. We are not aware of research within the noisy channel hypothesis showing that aligning multiple channels reduces noise, and hence uncertainty, but this would certainly be testable.

The second uncertainty-driven account is sketched out by Levy, Bicknell, Slattery, and Rayner (2009), who model how regressive eye movements can be derived from

uncertainty about previous syntactic structure. Although Levy et al. did not address semantic uncertainty, the account could be extended to uncertainty of all kinds, similar to the noisy channel hypothesis. Under this view, some of the data reported here might be interpreted as follows: In Experiment 1, uncertainty derives mostly due to implausibility, and also from noncanonical structure (depending on the region one considers). But in Experiments 2 and 3, the uncertainty stems mostly from syntactic structure. Although plausibility affected three measures in Experiment 2, these effects were marginal, and they did not even approach significance in Experiment 3. Thus, under this view, APS served to reduce overall uncertainty and led to better comprehension. Several aspects of this account do not appear consistent with the data, however. First, by bolstering the prosodic representation through APS, uncertainty about plausibility seems to go away, rather than uncertainty about syntax, as syntax is still a main factor in predicting regressions into and out of most regions in Experiments 2 and 3. It is not clear why strengthening the prosodic representation, which is tightly linked to the syntactic representation (Bader, 1998), should not have instead reduced uncertainty about the syntax in the latter two experiments, leaving behind only effects of plausibility. Second, if participants were simply not sure what the identity of the first noun was, or which noun was which, there should have been more regressions into the first noun region in Experiment 1 than 2 or 3. This was not the case. Also, regressions-out of the final region, which might be predicted if readers were simply unsure what was going on by the time they got to the end of the sentence, were not affected by any of the factors in the three-experiment comparison.

The failure of uncertainty-based approaches to straightforwardly handle multiple aspects of the data here lead us to return to Good Enough Processing: The more veridical prosodic and phonological representation generated when readers performed APS buttressed the syntactic structure against intrusions from a competing, heuristic-based interpretation. ORCs were still more difficult to process than SRCs (for one or more of the reasons proposed by researchers cited in the introduction), but this computation was no longer significantly affected by the added factor of implausibility. Furthermore, we take the results of the interest area analyses as evidence that this buttressing occurred online, during reading.

Summary

This study demonstrated that during silent reading readers engaged in auditory perceptual simulation (APS) of the speech of individual "speakers" to whom texts had been attributed. APS effects were observed in both online reading rates and eye movement patterns, and in offline comprehension measures. This perceptual simulation facilitated sentence processing that was tightly linked to the syntactic representation and and, consequently, improved comprehension. The results of these studies have clear pedagogical implications and should be of significant interest to educators, linguists, and psychologists who are concerned with how to improve reading comprehensions. Our results show that readers read with better comprehension, and no more slowly, when they activate an auditory simulation of another speaker's voice during silent reading. It is possible that if educators and parents instruct children to actively simulate their teachers' or another fluent English speaker's voice during silent reading, children could practice reading more effectively by themselves both at school and home. Future research will

extend these methods to a variety of readers and texts in order to determine if APS can be applied to educational settings. If APS is found to consistently improve reading comprehension across a wider variety of texts and readers (including non-native speakers), research focus can shift towards developing pedagogical methods designed to naturally induce students to undertake APS as part of day-to-day silent reading.

CHAPTER 3

AUDITORY PERCEPTUAL SIMULAITON: SIMULATING SPEECH RATES OR ACCENTS?

Chapter 3 contains my previous published work "Auditory perceptual simulation: Simulating speech rates or accents?".[4]

Auditory Perceptual Simulation (APS) refers to the phenomenon when readers mentally simulate characteristics of either the voices of the characters depicted in texts or the voices of other speakers (including their own) while they read silently (e.g., imagining Daniel Radcliffe's voice from the Harry Potter films when reading the Harry Potter books; see Hubbard, 2010, for a review of auditory imagery). When participants activate auditory perceptual simulation of speech during reading, they generate a rich mental representation of a depicted or imagined speaker "saying" the words in the text. This mental representation could be a more elaborated version of the normal implicit prosody that most skilled readers generate when reading silently (Fodor, 2002). The auditory perceptual simulation representation includes both segmental and suprasegmental information about the depicted or imagined "speaker's" voice, including, e.g., speech rate (Stites, Luke, & Christianson, 2013; Yao & Scheepers, 2011), accent (Filik & Barber, 2011), and characters' perspectives (Drumm, & Klin, 2011; Levine & Klin, 2001; Gunraj & Klin, 2012).

The research on auditory perceptual simulation effects (or auditory imagery) on reading date back to Kosslyn and Matt (1977), who explored whether participants would

[4] Zhou, P., & Christianson, K. (2016b). Auditory perceptual simulation: Simulating speech rates or accents?. *Acta psychologica*, *168*, 85-90.

activate talker-specific auditory imagery while reading aloud after hearing some speakers' recordings. They first familiarized the readers with two speakers, one faster and the other slower, by playing recorded passages. Then, participants were asked to read aloud passages that were purportedly "written" either by the faster or slower speaker. The observed data showed that participants read the faster speaker's text more quickly than the slower speaker's text, suggesting that the participants activated the perceptual features of the corresponding speaker's voice while reading aloud.

Alxendar and Nygaard (2008) extended this finding of engaging in talker-specific auditory imagery during reading aloud to silent reading. They adapted Kosslyn and Matt's paradigm in a silent reading task and manipulated the speech rates of two speakers and text difficulty. Readers were first familiarized with these two speakers' voices (fast vs. slow), then they were told to read the passages that were "written" either by the faster or slower speaker silently and answer comprehension questions. The results demonstrated that even silent reading speeds were modulated corresponding to the speech rates of the speakers. Moreover, readers were more likely to activate auditory imagery of the speakers when the texts were difficult.

Kurby, Magliano, and Rapp (2009) found that auditory imagery could be influenced by familiarity with the speakers and the texts. Participants recognized more words when read by a familiar speaker in a novel script, and repeated exposure to the text strengthened their mental representation of a character's voice, facilitating quicker recognition of the words read by the same character later. Thus, the authors concluded that readers activate perceptually based knowledge while reading even without direct experience of the voice in the particular context.

In recent studies, scholars extended the trigger of the auditory imagery from the voices alone to the photos of the speakers (Woumans, Martin, Vanden Bulcke, Van Assche, Hartsuiker, & Duyck, 2015; Zhou & Christianson, in press). Woumans et al. familiarized Spanish-Catalan bilinguals with two speakers' photos and their corresponding languages, either in Spanish or Catalan, during simulated Skype conversations. In the later language production task, one speaker's photo was presented on a Skype interface while saying a noun in the corresponding language. The subjects were asked to produce the first verb they associated it with and in the same language as the given stimulus. Results demonstrated that subjects responded faster when speaker's photo was congruent with the corresponding language, indicating that the photo on the Skype interface cued bilingual subjects to use the corresponding language. In a follow-up experiment, the finding was replicated among the Dutch-French bilinguals with the same paradigm and task.

Zhou and Christianson (2016) used eye tracking to investigate how auditory perceptual simulation of native and non-native English speech affects sentence processing and comprehension. They also used speakers' photos as the cues to trigger the auditory perceptual simulation effects: an English speaker's photo was matched with the faster native American English speech while a Chinese speaker's photo was matched with the slower Chinese accented English. The researchers first familiarized participants with the native and non-native speech by playing separate recordings while presenting the corresponding native and non-native speakers' photos on the screen. Then, subjects were asked to read sentences (e.g., "*The policeman that chased the thief drove fast.*") and respond to a paraphrase verification probe after each sentence (e.g., *The policeman*

chased the thief. The policeman drove fast. True/False). Before each sentence was presented, one of the speakers' photos appeared on the screen, and the participants were asked to imagine this speaker's voice while reading the upcoming sentence. The total sentence reading time, response accuracy, fixation durations on individual words, and saccade patterns were analyzed.

The results demonstrated that (1) participants read sentences attributed to the slower-speaking non-native speaker more slowly than sentences attributed to the faster-speaking native speaker; 2) participants who were induced to perform auditory perceptual simulation (independently of which speaker's voice had been cued) read sentences faster, in terms of total sentence reading times as well as early and late measures on individual words, than participants (in a separate session) who were not induced to perform auditory perceptual simulation; 3) there were no significant differences in comprehension probe response accuracy between auditory perceptual simulation of native speech and auditory perceptual simulation of non-native speech; and 4) participants who were induced to perform auditory perceptual simulation of either native or non-native speech showed better comprehension overall, most markedly in morphosyntactically complex (object-relative clauses) and semantically implausible sentences, such as "*The bird that the worm ate was small*," compared to those who read under normal silent reading conditions.

Zhou and Christianson argued that the online reading speed differences in the two auditory perceptual simulation conditions were modulated by the different speech rates of the native and non-native speakers—auditory perceptual simulation of the faster native speaker's voice led to faster reading speeds, whereas auditory perceptual simulation of the slower non-native speaker's voice yielded slower reading speeds. However, a large

body of sociolinguistic research has shown that native speakers usually have high standards of acceptability for using their language and judge people based on how they differ from these standards (e.g., Giles & Watson, 2013; Lippi-Green, 1997; Ryan, 1983). Native English speakers often rate non-natively accented speech as less comprehensible, less favorable, less trustworthy, and less persuasive compared to native speech in various settings (Brenna & Brenna, 1981; Callen, Callois, & Forbes, 1983; Edward, 1977; Gass & Varonis, 1982, 1984; Giles, 1972; Giles, Hewstone, Ryan, & Johnson, 1987; Gluszek & Hansen, 2013; Kinzler, Shutts, DeJesus, & Spelke, 2009; Munro & Derwing, 1995, 1998; White, 1991). In Munro and Derwing's (1995) study, when participants were asked to decide the truth-value of statements read either by Chinese-accented English speakers or native English speakers, they spent more time processing the statements read by native Chinese speakers than the ones read by native English speakers. Moreover, listeners have been observed to rate statements (e.g., *A giraffe can go without water longer than a camel can*) read by non-native speakers as less credible than statements read by native speakers (Lev-Ari & Keysar, 2010). Thus, it might be argued that, in the Zhou and Christianson study, readers' slower reading speed when activating auditory perceptual simulation of non-native speech was triggered by difficulty in simulating accented English speech, rather than the speech rate *per se*.

Current Study

In the current study, one eye-tracking experiment was conducted to investigate whether earlier auditory perceptual simulation effects were triggered by the native and non-native speakers' speech rates or by difficulty that the English-speaking readers may have experienced in simulating an unfamiliar accent. The eye-tracking methodology was

applied here for two reasons. First, we wanted to use the same paradigm from Zhou and Christianson's study, which has been shown to reliably trigger auditory perceptual simulation effects. We also wanted to ensure that the results from the current study were comparable to the previous one. Additionally, eye-tracking provides a very accurate measure of readers' fixations and reading times on the target sentences.

This experiment also manipulated the same conditions as in Zhou and Christianson (2016): plausibility, syntactic complexity, and "speaker" identity. Plausibility and structure were manipulated in the experiment because previous studies have shown consistent reading and comprehension patterns for these sentences: implausible sentences are read more slowly than plausible sentences, and object relative clauses are read more slowly than subject relative clauses (Gibson, 2005; Gennari & MacDonald, 2008; Traxler, Morris, & Seely, 2002; Zhou & Christianson, 2016). Readers also usually have more difficulty comprehending object-relative clauses than subject-relative clauses, and implausible sentences are more likely to me misinterpreted than plausible sentences. Importantly, when syntactic complexity and semantic plausibility are crossed in this way, comprehension errors are systematic and predictable. Specifically, readers (and listeners) tend to derive interpretations of implausible sentences with more difficult structures such that the actors (arguments) in these sentences are often reversed. This observation was originally made for passive vs. active sentences (Christianson, Luke, & Ferreira, 2010; Ferreira, 2003; Lim & Christianson, 2013a), and more recently extended to subject- vs. object-relative clauses (Lim & Christianson, 2013b; Zhou & Christianson, 2016). For example, the misinterpretation that is frequently derived from *The bird that the worm ate was small* (1d, below) is that "the bird ate the worm." This

pattern of misinterpretation has been attributed to "good-enough" processing (Christianson, 2016; Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Patson, 2007; Zhou & Christianson, 2016). According to Good Enough theory, these misinterpretations are due to the use of language processing heuristics – such as plausibility and probabilistic word order cues (e.g., Townsend & Bever, 2001) – which operate along with algorithmic morphosyntactic processing. Syntactic structure is fragile, however (Sachs, 1967), and the output of the heuristic-based processing overwhelms the output of the morphosyntactic processing in a significant proportion of trials. Zhou and Christianson (2016) observed a decrease in the occurrence of this sort of misinterpretation when people read with auditory perceptual simulation. We hypothesized that this improved comprehension stemmed from a richer, more robust morphosyntactic representation that was generated with and supported by the richer, more detailed prosodic representation created via auditory perceptual simulation.

For the "speaker" identity condition, an Indian-English speaker and an American-English speaker were recruited to read the texts. Instead of a native Chinese speaker, as had been used in Zhou and Christianson (2016), an Indian-English speaker's voice was used in this experiment to validate that the auditory perceptual simulation effect is not limited to a particular accent. Furthermore, the Indian-English speaker with whose voice participants were familiarized spoke faster than the American-English speaker. In this way, across the previous and current studies, the nativeness and speech rate manipulations were counterbalanced (native vs. non-native; faster vs. slower).

The hypothesis here is that if the auditory perceptual simulation effects reported in Zhou and Christianson (2016) derived from readers' difficulty in simulating an unfamiliar accent, regardless of the speech rate difference, mentally simulating the Indian-English speech here should yield longer sentence reading times compared to perceptual simulation of the American-English speech. If, on the other hand, the previous auditory perceptual simulation effects were triggered by different speech rates rather than accents, readers in this experiment should read no more slowly when they perceptually simulate the faster Indian-English speaker's voice than when they simulate the slower American-English speaker's voice. Readers might read sentences attributed to the faster Indian-English speaker more quickly than sentences attributed to the native American-English speaker; however, there may be an upper limit on how fast people can read and simulate, such that any speed advantage for the faster speaker might be relatively small. Finally, we predict no significant comprehension accuracy differences between auditory perceptual simulation of the two speakers, consistent with Zhou and Christianson (2016).

In addition, here we combine and compare the data from the current experiment to the previous experiment in Zhou and Christianson, where a faster native speaker's voice and a slower non-native speaker's voice were used[5]. The combined data include "speaker" (native vs. non-native) and speech rate (fast vs. slow) as the independent variables to predict reading speed. The prediction is that if accent modulates the auditory perceptual simulation effects, "speaker" but not speech rate, will be the significant predictor of

[5] In Zhou and Christianson (2016), the duration of the native English speech: text 1 = 2min 50s; text 2 = 2min 57s; text 3 = 2min 49s; text 4 = 2min 57s. The duration of non-native speech: text 1 = 4min 08s; text 2 = 4min 18s; text 3 = 4min 08s; text 4 = 4min 20s.

sentence reading speed. Otherwise, speech rate but not "speaker" should be the driving force of the reading speed.

Method

Participants

Eighty-seven native American-English speakers with normal or corrected-to-normal vision from the University of Illinois at Urbana-Champaign community participated in the experiment. They received either $7 payment or 1 research credit. Seven subjects' data were excluded due to either fatigue or difficulty in calibrating their eye gaze. Eighty native participants[6] were retained in the data analysis.

Materials

Four 500-words texts, 48 experimental sentences, and the social attractiveness questionnaire used in Zhou and Christianson (2016) were adapted in this experiment. Four texts that were counterbalanced for length and difficulty were used as the auditory materials. A female native Indian-English speaker and a female native American-English speaker, both in their early 20s, were recruited to record the auditory materials. The speaking rates of the speakers were significantly different: the American-English speaker had a significantly slower mean reading rate than the Indian-English speaker ($t(3) = 145$, $p < 0.001$).[7] Photos were presented simultaneously with the recordings. These were not of the actual speakers, but rather non-copyrighted stock photos from the Internet. The photo

[6] Based on Zhou and Christianson (2016), we calculated the sample size (N=48.5). Because the plan from the outset was to compare the results from this experiment to the previous study, however, we ran the same number of subjects as in the previous study (N=80).

[7] Duration for American-English speech: text 1 = 3min 43s; text 2 = 3min 43s; text 3 = 3min 43s; text 4 = 3min 43s. Duration for Indian-English speech: text 1 = 2min 53s; text 2 = 2min 59s; text 3 = 2min 57s; text 4 = 3min 02s.

of the American-English speaker was of a blonde, Caucasian woman who appeared to be in her early 30s. The photo of the Indian-English speaker was of an Indian woman who appeared to be in her early 30s. The photos showed the women from the shoulders up. Both women were smiling and dressed in a business jacket and blouse. These photos were purposely selected to suggest that the two speakers were of similar ages and had similar styles, jobs, and socioeconomic statuses. The Indian-English speaker's voice was always matched to the Indian woman's photo, and the American-English speaker's voice to the Caucasian woman's photo.

Forty-eight target sentences and 96 filler sentences were included, all of which were grammatical. Two factors, structure and plausibility, were manipulated in the target sentences, resulting in a 2 (subject-relative clause, object-relative clause) x 2 (plausible, implausible) x 2 ("speaker" identity) fully-crossed, within-participants and within-items design.

Examples of the four sentence conditions are presented in (1). Before each sentence was presented, one of the two speakers' photos appeared. After each sentence, a paraphrase verification probe was used to measure readers' comprehension. Readers were asked to decide whether the verification probe (1e) was true or false based on the sentence they had just read. The answer was always "True" for experimental items, but the number of True and False correct responses was equal across the experiment. Materials were distributed pseudo-randomly for each participant across four lists in a Latin-square design, so that every participant saw each item only once.

(1) a. "The bird that ate the worm was small." (subject-relative, plausible)

b. “The worm that ate the bird was small.” (subject-relative, implausible)

c. “The worm that the bird ate was small. “ (object-relative, plausible)

d. “The bird that the worm ate was small.” (object-relative, implausible)

e. The bird/worm ate the worm/bird. The bird/worm was small. (T/F)

The social attractiveness survey included fourteen attributes, including accent, speech rate, comprehensibility, conscientiousness, confidence, dependability, education, honesty, intelligence, likability, reliability, pleasantness, prestigious, and sincerity. Participants rated each speaker on a 1-7 Likert scale (e.g., “Please rate the pleasantness of the speaker’s voice on a 1-7 Likert scale”, 1=not pleasant at all, 7=very pleasant. See the complete survey in the Appendix A in Zhou & Christianson in preparation). In addition, each subject reported how often they performed auditory perceptual simulation of the speakers' voices during the eye-tracking portion of the experiment (e.g. “How often were you perceptual simulating the native speaker’s (Judy's) voice when you were asked to? a. All of the time (100%). b. Most of the time (80%). c. Half of the time (50%). d. Sometimes (30%). e. Rarely (10%). f. I was not able to do so (0%)).

Apparatus

Eye movement data and comprehension data were collected using an SR Research EyeLink 1000 remote desktop eye tracker. A chin rest and forehead rest were used to stabilize participants' heads. Sentences were presented in black Courier New monotype font (14pt) on a white background, which was approximately 70cm away from the participants. At this distance, 1° of visual angle subtended approximately three characters. All viewing was binocular, but data were recorded from the right eye. A separate computer was used to collect the survey results.

Procedure

The entire experiment lasted less than one hour. After the participant provided informed consent, he/she was calibrated on the eye tracker using a nine-point calibration procedure to start the eye-tracking portion. After the calibration, the participant listened to two recordings, one by the American-English speaker, the other by the Indian-English speaker, while the corresponding photograph was displayed on the computer screen. The order of recordings was counterbalanced across participants, with each one presented first half of the time. Then participants read and responded to the six practice items and, after any questions, proceeded to read and respond to the experiment items. In each trial, participants first saw either the American-English or Indian-English speaker's photo while listening to a recording of the corresponding speaker saying her name ("Charu" / "Judy"). Then on the next screen, a sentence was presented. Participants were instructed to imagine that the speaker whose photo had just been presented on the previous screen was saying it to them as they read it silently. Another recording set of the American-English and Indian-English speakers reading similar texts aloud was played halfway through the experiment to remind participants what each speaker sounded like. After the eye-tracking portion of the experiment, participants were asked to complete one social attractiveness survey for each speaker (Zhou & Christianson, 2016, in preparation).

Results

We first report the response accuracy for the target sentences and offline survey results. Then, we report the whole sentence reading time data, which is the most important result to address our research question. In addition, we compare whole sentence reading time from the current study to Zhou and Christianson (2016a). Linear

mixed-effects modeling was applied to analyze the continuous eye movement data, and logit mixed-effects modeling was used to analyze the binomial (response accuracy) data. All analyses were performed using maximal random effects structures, i.e., with random slopes and intercepts for participants and items. Table 3.1 provides means and standard deviations of sentence reading times and accuracy in all conditions in the current study and in Zhou and Christianson (2016).

Table 3.1

Means and Standard Deviations (SD) for Sentence Reading Time (in msec) and Accuracy in the Current Experiment.

"Speaker"	Structure	Sentence Reading Time		Accuracy	
		Plausible	Implausible	Plausible	Implausible
Native	Subject-RC	4067 (2211)	4555 (2972)	.96 (.20)	.93 (.25)
	Object-RC	4672 (2573)	5075 (3259)	.88 (.33)	.69 (.46)
Non-native	Subject-RC	4058 (2071)	4279 (2520)	.95 (.22)	.95 (.21)
	Object-RC	4544 (2572)	4874 (2690)	.89 (.32)	.69 (.47)

Response Accuracy

On average, participants achieved 86.64% comprehension accuracy for the target sentences**.** The logit mixed-effects model results demonstrated that comprehension did not vary as a function of which speaker's voice was being simulated (*Est.*= -0.01; *SE* = 0.11; z = -0.05, p = 0.96). Plausibility (Plausible > Implausible; *Est.*= 1.36; *SE* = 0.18; z =7.75, $p < 0.001$), structure (SRC > ORC; *Est.*= 2.2; *SE* = 0.20; z = 10.74, $p < 0.001$), and their interaction (SRC in plausible sentences > ORC in implausible sentences; *Est.*= -1.12; *SE* = 0.31; z = -3.65, $p < 0.05$), as well as trial presentation order (Later trials > Earlier trails; *Est.*= 0.003; *SE* = 0.001; z = 2.28, $p < 0.001$) significantly influenced response accuracy. Table 3.2 presents the details of the LME model for the current study.

Table 3.2

Fixed Effects of Logit Mixed-effects Model for Response Accuracy

Predictor	Estimate	SE	*z* value	*p* value
(Intercept)	0.71	0.18	3.85	< 0.001
Plausibility	1.36	0.18	7.64	< 0.001
Structure	2.2	0.2	10.74	< 0.001
"Speaker"	-0.01	0.11	-0.05	0.96
Trial Order	0.003	0.001	2.28	< 0.05
Plausibility x Structure	-1.12	0.31	-3.65	< 0.001

Social attractiveness survey data

Based on the self-report from the surveys, approximately 66.1% of the time, subjects performed auditory perceptual simulation of the American-English speaker's voice and 59.2% of the time, they performed auditory perceptual simulation of the Indian-English speaker's voice. Fourteen variables were used to evaluate and compare readers' attitudes towards the American-English and Indian-English speakers. ANOVA analysis demonstrated that readers' attitudes were more negative towards the Indian-English speaker than the American-English speaker on all variables except honesty ($F(1, 212) = 3.47$; $p = 0.06$) and dependability ($F(1, 212) = 3.68$; $p = 0.06$), where the differences were only marginal. Participants perceived the accent and speech rate differences between the American-English and Indian-English speakers, but only marginally so ($F(1, 212) = 3.68$; $p < 0.1$). They rated the Indian-English speech as significantly less comprehensible ($F(1, 212) = 64.15$; $p < 0.05$), less confident ($F(1, 212) = 140.9$; $p < 0.05$), less intelligent ($F(1, 212) = 51.60$; $p < 0.05$), less pleasant ($F(1, 212) = 67.08$; $p < 0.05$), less likable ($F(1, 212) = 45.69$; $p < 0.05$), and less reliable ($F(1, 212)$

= 6.57; $p < 0.05$). They regarded the Indian-English speaker as less conscientious ($F(1, 212) = 14.16$; $p < 0.05$), less educated ($F(1, 212) = 37.14$; $p < 0.05$), less sincere ($F(1, 212) = 140.9$; $p < 0.05$), and less prestigious ($F(1, 212) = 55.58$; $p < 0.05$) than the American-English speaker. Nevertheless, "speaker" was not a significant predictor of comprehension accuracy based on the logistic regression model results above, suggesting that readers' biases towards the non-native speech did not affect their comprehension.

Whole sentence reading time

Fixations shorter than 80ms and longer than 1200ms were trimmed before the data analysis. Reading times that were three standard deviations away from the mean of that condition within each subject were excluded. These trimming procedures resulted in removal of 0.06% of the data. Trimmed sentence reading times were centered in the LME analyses. The LME models for all following analyses included the predictors of "speaker," plausibility, structure, and trial order.

Results reveal that auditory perceptual simulation of the American-English speech led to marginally longer sentence reading time than auditory perceptual simulation of the Indian-English speech ($t = -1.86$; $p = 0.06$; 95% CI = -0.104 – 0.002). Plausibility (Implausible > Plausible; $t = -3.12$; $p < 0.001$), structure (ORC > SRC; $t = -4.65$; $p < 0.01$), and trial order (Earlier trials > Later trails; $t = -8.33$; $p < 0.001$) significantly affected sentence reading times. Table 3.3 presents the fixed effects of the linear mixed model for whole sentence reading time.

Table 3.3

Fixed-effects for Linear Mixed Model for the Whole Sentence Reading Time

	Estimate	SE	df	*t* value	*p* value
Intercept	0.42	0.09	134.4	4.47	< 0.001
Plausibility	-0.14	0.05	141.2	-3.12	< 0.01
Structure	-0.21	0.05	147.6	-4.65	< 0.001
"Speaker"	-0.05	0.03	105.7	-1.86	0.06
Trial order	-0.003	0.0003	1408	-8.33	< 0.001

Combined Data Analysis

In the current experiment, we detected a marginally significant difference in the predicted direction between the auditory perceptual simulation of the faster-speaking non-native "speaker" and the slower-speaking native "speaker": auditory perceptual simulation of the faster non-native "speaker" led to faster reading time, despite the non-native accent. We suspect that the marginal ($p = 0.06$) statistical significance may result from a ceiling effect on how fast people can read and perceptually simulate any speech. Alternatively, the difference between slow and fast speech in the current experiment was not quite large enough to yield an effect as clearly as in Zhou and Christianson (2016), in which the difference between speaker rates was larger (t=-5.08, df = 3, $p < 0.05$; see Table 3.4 for the means).

Table 3.4

Means of Recording Time for Faster and Slower Speakers Across Two Experiments (Min:Sec)

	Faster Speaker	Slower Speaker	Differences
Zhou & Christianson (in press)	2:50 (American)	4:14 (Chinese)	1:24
Current Experiment	2:58 (Indian)	3:43 (American)	45

To further investigate whether the auditory perceptual simulation effects were driven by accents or speech rates, we compared the current experiment to the second auditory perceptual simulation experiment[8] in Zhou & Chrisitanson (2016), where the pattern of speech rate effects for the native and non-native speakers were opposite to the current study. In this way, the combined dataset not only included the "speaker" condition (American-English vs. Indian-English), but also the "speaker's" speech rate condition (fast vs. slow). The means and standard deviations for each condition ("speaker," speech rate, plausibility, structure) in the combined dataset are presented in Table 3.5. The same data trimming procedures from the current experiment were applied to the new combined dataset (4.3% of the data were removed). LME models with maximal random effects were built to analyze the results.

Table 3.5

Means and Standard Deviations for Sentence Reading Time (in msec) and Accuracy in the Combined Data (current experiment and Zhou & Christianson, 2015)

"Speaker"	Speed	Structure	Sentence Reading Time		Accuracy	
			Plausible	Implausible	Plausible	Implausible
Native	Fast	Subject-RC	3825 (1764)	3497 (1644)	.95 (.22)	.96 (.19)
		Object-RC	4115 (1970)	4313 (1868)	.70 (.46)	.88 (.32)
	Slow	Subject-RC	4172. (1783)	3842 (1747)	.94 (.24)	.96 (.20)
		Object-RC	4617 (2101)	4447 (2078)	.69 (.46)	.87 (.33)
Non-native	Fast	Subject-RC	3962 (1702)	3754 (1546)	.95 (.21)	.96 (.21)
		Object-RC	4547 (2017)	4198 (1833)	.68 (.47)	.88 (.32)
	Slow	Subject-RC	3973 (1871)	3681 (1678)	.92 (.27)	.96 (.20)
		Object-RC	4645 (1702)	4385 (1546)	.69 (.46)	.87 (.34)

[8] There were two APS experiments in Zhou and Christianson (2016). We only compared to one of them, because the two APS experiments showed the same APS effects on reading time and comprehension, and they differed only in the APS cue (photo vs. a recording of the speaker's name).

The results reveal that speech rates (Fast < Slow; *Est.*= 0.08; *SE* = 0.02; *t* = 3.99, < 0.001), plausibility (Plausible < Implausible; *Est.*= -0.14; *SE*= 0.03; *t* = -4. 59, < 0.001), syntactic structure (SEC < ORC; *Est.*= -0.26; *SE* = 0.03; *t* = -7.93, *p* < 0.001), and trial presentation order (Earlier < Later; *Est.*= -0.002; *SE* = 0.0002; *t* = -11.32, *p* < 0.001), but not "speakers" (*Est.*= 0.03; *SE*= 0.02; *t* = 1.47, *p* > 0.1), were significant predictors of reading times when readers perceptually simulated either native or non-native speaker voices. Readers read faster when they were simulating the faster "voice," regardless of the familiarity of the accent (see Table 3.6 for details).

Table 3.6

Fixed Effects of Linear Mixed-effects Model for Sentence Reading Time in the Combined Dataset

	Estimate	SE	df	*t* value	*p* value
Intercept	0.35	0.06	235.9	5.57	< 0.001
Plausibility	-0.14	0.03	294.2	-4.59	< 0.001
Structure	-0.26	0.03	305.3	-7.93	< 0.001
"Speaker"	0.03	0.02	175.9	1.47	> 0.1
Speech Rate	0.08	0.02	179.1	3.99	< 0.001
Trial order	-0.002	0.0002	1476	-11.03	< 0.001

Discussion

In this study, we employed an auditory perceptual simulation (APS) paradigm with a faster Indian-English speaker's speech and a slower American-English speaker's speech to investigate whether previously observed effects of auditory perceptual simulation of native and non-native speech derived from speech rate differences between speakers or from difficulty simulating an unfamiliar accent. Although social attractiveness survey results clearly demonstrated that readers had more negative attitudes

towards the Indian-English speech, simulation of this (faster) accented speech led to faster silent reading than simulation of the speech of the (slower) American-English speaker. By combining the data from the current experiment and the data from Zhou and Christianson (2016), we found that the auditory perceptual simulation effects on readers' silent reading times were driven by the different speech rates of the voices that were being perceptually simulated, regardless of whether they were native or non-native. Furthermore, accuracy did not differ as a function of which speaker's voice was being simulated. If the online silent reading speeds observed previously (Zhou & Christianson, 2016) and in the present study had been due to difficulty simulating an unfamiliar accent, auditory perceptual simulation of the Indian-English speech in this experiment should have resulted in slower reading times compared to the American-English speech, irrespective of the speech rates. This was not the case. Furthermore, we did not find biases against one type of accent or another to be the source of reading speed differences or comprehension accuracy effects in either study. A sociolinguistic bias account of the results would predict readers to be more likely to choose the paraphrase of implausible sentences as false when engaging in auditory perceptual simulation of non-native or accented speech (e.g., Lev-Ari & Keysar, 2010). Yet, this was not the case: the online reading speed was in the opposite direction of this hypothesis, and auditory perceptual simulation of Indian-English speech did not lead to lower accuracy than the American-English speech.

Based on this evidence, we argue that when readers perform auditory perceptual simulation during silent reading, they focus on the prosodic characteristics of the voice(s) that they are simulating. The mental representations generated by perceptual simulation

my well also contain phonological characteristics, but the relative familiarity of these characteristics do not appear to influence effects of APS on reading speed or comprehension. It is an open question as to what extent phonological characteristics can be perceptually simulated.

As it stands, however, the results reported here are consistent with the proposal by Zhou and Christianson (2016a): Auditory perceptual simulation generates a prosodic representation of the text that is more detailed than the default prosodic contour that is generated by most skilled readers when APS is not cued (Fodor, 2002; Rayner, Pollatsek, Ashby, & Clifton, 2012), including speech rate. Prosodic structure correlates with syntactic and information structure (e.g., Breen, Fedorenko, Wagner, & Gibson, 2010). When a rich prosodic structure is generated via auditory perceptual simulation during reading, it "buttresses" the syntactic structure, which consequently is more likely to be maintained more robustly against intrusions from competing "good-enough" interpretations derived from non-structural heuristics.

In conclusion, the evidence presented here strongly suggests that a central aspect of the perceptual simulation of speech during reading is the speech rate of the speaker whose voice is being simulated. Differences in simulated speech rates in turn drive the reading speed differences observed in previous work and here, rather than the relative difficulty of simulating unfamiliar accents.

CHAPTER 4

L2 ENGLISH SPEAKERS' AUDITORY PERCEPTUAL SIMULATION DURING SILENT READING

My first study demonstrated that APS during silent reading increased native English speakers' reading speed and improved comprehension. If APS could affect L2 English speakers in the same way, it could be applied in classrooms to increase L2 English learners' reading speed and enhance their comprehension. Thus, I extended the investigation of APS to L2 English speakers' sentence processing and comprehension in this chapter. One eye-tracking experiment was run to investigate the processing speed and comprehension accuracy of L2 speakers when they perceptually simulated native and non-native speaker voices. The paradigm from study 1 and 2 was adapted in Experiment 4, using both the photo and voice cues to trigger the APS of native and non-native speakers during silent reading.

Based on the findings from study 1, I hypothesize that L2 English speakers' reading speed will be modulated by the APS of the native/non-native speech. According to the results from study 2, the reading speed rather than the accent of the "speaker's" voice will influence readers' sentence processing speed. Despite the fact that Chinese speakers are familiar with the non-native Chinese accented speech, their reading patterns should be similar to other L2 English speakers: both groups should read the sentences slower when perceptually simulating the slower non-native Chinese accented speech; APS of the faster native speech should yield faster reading times across the sentence.

Experiment

Method

Participants

Forty-four native Chinese speakers and 44 other L2 English speakers from UIUC community participated. They all aged from 18 to 30 years old and had normal to corrected-to normal vision. They received either two research credits or $10 payment for their participation. Due to the experimental mistakes, three native Chinese speakers' data were excluded from the data analysis. Table 4.1 shows the demographic information of the subjects.

Table 4.1

Demographic Information of 41 Native Chinese Speakers and 44 Other L2 English Speakers

Native Language	Female	Male	Averaged Starting age of English Learning	Mean Age
Arabic	2		12	26
Cantonese	1		4	26
Greek	1		5	42
Hindi	3	2	5	24
Japanese	1		12	21
Kazakh	1		12	21
Korean	5	4	10	24
Malay		1	7	21
Mexican		1	12	41
Polish	2		11	36
Russian	1		8	21
Serbian	1		6	36
Spanish	3	6	8	25
Telugu	1	2	4	23
Thai	3		9	30
Turkish		1	14	34
Vietnamese	1		4	19
Yoruba	1		6	21
Chinese	32	9	10	25

Materials

The auditory and reading materials were identical to the ones in Chapter 2, except that the both the speakers' photos and the recordings of the speakers saying their names ("Susan" or "Xiaofu") were used simultaneously as APS prompts prior to the presentation of each sentence, because both photos and voices cues triggered the APS effects and the combination of two may be more effective for L2 English speakers. The same social attractiveness survey was used at the end of experiment to measure participants' attitudes towards the native and non-native speakers. The Nelson-Denny Reading Test (Brown, Fishco, & Hanna, 1993) was used to measure participants' offline reading speed and reading comprehension. The Reading Span Task (Daneman & Carpenter, 1980) was used to measure participants' working memory.

Procedure

The whole experiment lasted less than two hours for all participants. The procedure in the eye-tracking part was the same as the procedure in Experiment 2 of study 1. Participants were cued to activate APS of the native or non-native speech with the speakers' names and photos. After the reading portion, participants were asked to fill the social attractive survey for the native and non-native speaker. Finally, they were asked to complete the Nelson-Denny Reading Test and the Reading Span Task.

Results

Offline behavioral measures

On average, native Chinese speaking participants read 190 words (*SD* = 72) per minute. They achieved 89% accuracy rate (*SD* = 0.00) in the Nelson-Danny Comprehension task. Other L2 English speakers averagely read 179 words per minute (SD = 56) and achieved 85% accuracy rate (*SD* = 0.13) in the Nelson-Danny

Comprehension task. The reading speed differences between the Native Chinese speaking participants and other L2 English speaking participants were marginal ($t = 1.81$; $p = 0.07$). The comprehension accuracy difference was not significant ($t = 0.63$; $p = 0.53$). In the Reading Span Task, other L2 English speaking participants (*Mean* = 62; *SD* = 15) performed significantly better ($t = -4.45$, $p < 0.05$) than the Native Chinese speaking participants (*Mean* = 50; *SD* = 9).

Eye-tracking Results

Fixations shorter than 80ms and longer than 1200ms were trimmed before the data analysis. Reading times that were three standard deviations away from the mean of that condition within each subject were excluded. These trimming procedures resulted in removal of 0.1% of the data. Trimmed sentence reading times were centered in the mixed-effects modeling analyses. Linear mixed-effects modeling (Baayen, 2008) was applied to analyze the continuous eye movement data, such as overall reading times, and logit mixed-effects modeling was used to analyze the binomial data (comprehension accuracy). Speech, plausibility, structure, subjects' native language groups, working memory and offline attitude measures were included in the mixed-effects models as the fixed effects. Subjects and items were included as the random effects. Because not all the subjects were able to take the Nelson-Denny Reading Task within the two-hour experiment, I did not include Nelson-Denny Reading Task results in the analysis. Moreover, because none of the offline attitudes measures significantly predict the online total sentence reading times or response accuracy, they were excluded from the final model. Table 4.2 below shows the means and standard deviations of sentence reading times and accuracy in each condition.

Table 4.2

Means and Standard Deviations of Sentence Reading Times (in msce) and Accuracy in Each Condition

			Implausible		Plausible	
	Speech	Language Group	ORC	SRC	ORC	SRC
Sentence RT	Native	Chinese	6853 (3872)	6176 (3036)	6016 (3261)	5691 (2849)
		Other L2	6206 (3646)	5607 (3923)	5866 (3128)	5267 (3310)
	Non-native	Chinese	7081 (3943)	6480 (3094)	6132 (3123)	5697 (2876)
		Other L2	6720 (3703)	5970 (3762)	6011 (4470)	5729 (3724)
Accuracy	Native Speech	Chinese	0.76 (0.43)	0.93 (0.25)	0.92 (0.27)	0.95 (0.22)
		Other L2	0.69 (0.46)	0.9 (0.31)	0.87 (0.34)	0.93 (0.26)
	Non-native Speech	Chinese	0.74 (0.44)	0.9 (0.30)	0.91 (0.29)	0.95 (0.22)
		Other L2	0.67 (0.47)	0.9 (0.29)	0.87 (0.34)	0.92 (0.27)

Sentence Reading Time

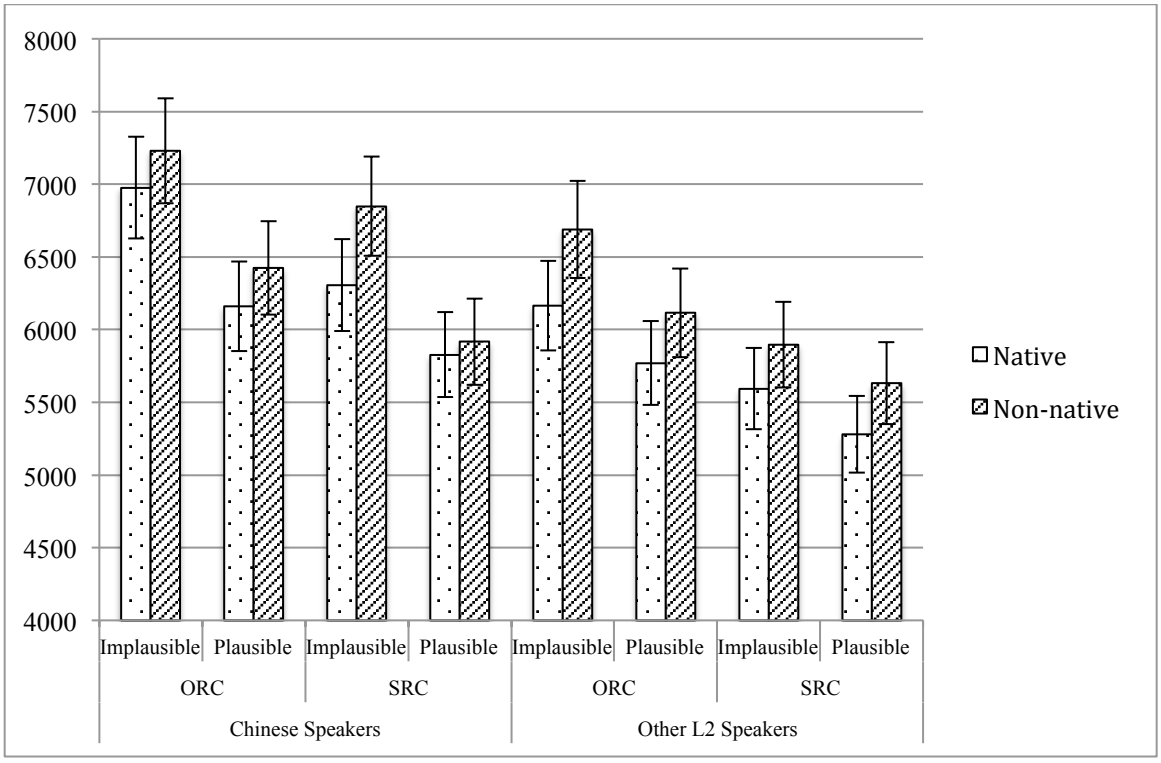


Figure 4.1. Sentence Reading Time (in msec) across different plausibility, syntactic structure, "speaker," and subjects' native language conditions. Error bar stands for the standardized errors.

Figure 4.1 shows total reading times across all conditions, and Table 4.3 presents the model results for total sentence reading times. Auditory Perceptual Simulation of the non-native speech led to longer total sentence reading times than APS of the native speech ($p < 0.05$). Readers processed the plausible sentences faster than the implausible ones ($p < 0.05$), and they read the SRC sentences faster than the ORC sentences ($p = 0.05$). Readers' working memory marginally affected the total reading times—high working memory readers had faster reading speed than the low working memory readers ($p = 0.06$). The main effect of language group was not significant ($p = 0.96$), but the interaction between language group and speech ($p < 0.05$) as well as the interaction between language group and structure were significant ($p < 0.05$). The differences in reading times between the APS of native and non-native speech was bigger among other L2 speaking subjects than the native Chinese speaking subjects. The reading time differences between processing ORCs and SRCs were larger among native Chinese speaking subjects than their non-Chinese L2 English counterparts. Moreover, this difference was more salient in the implausible condition than the plausible one, as indicated by the significant three-way interaction between speech, plausibility, and structure ($p < 0.05$).

Table 4.3

Fixed Effects of Linear Mixed Effects Model for the Whole Sentence Reading Time

		Implausible		Plausible	
Speech	Language Group	ORC	SRC	ORC	SRC

Table 4.3 (cont.)

	Speech	Language Group	Implausible		Plausible	
			ORC	SRC	ORC	SRC
Sentence RT	Native	Chinese	6853 (3872)	6176 (3036)	6016 (3261)	5691 (2849)
		Other L2	6206 (3646)	5607 (3923)	5866 (3128)	5267 (3310)
	Non-native	Chinese	7081 (3943)	6480 (3094)	6132 (3123)	5697 (2876)
		Other L2	6720 (3703)	5970 (3762)	6011 (4470)	5729 (3724)
Accuracy	Native Speech	Chinese	0.76 (0.43)	0.93 (0.25)	0.92 (0.27)	0.95 (0.22)
		Other L2	0.69 (0.46)	0.9 (0.31)	0.87 (0.34)	0.93 (0.26)
	Non-native Speech	Chinese	0.74 (0.44)	0.9 (0.30)	0.91 (0.29)	0.95 (0.22)
		Other L2	0.67 (0.47)	0.9 (0.29)	0.87 (0.34)	0.92 (0.27)

Response accuracy

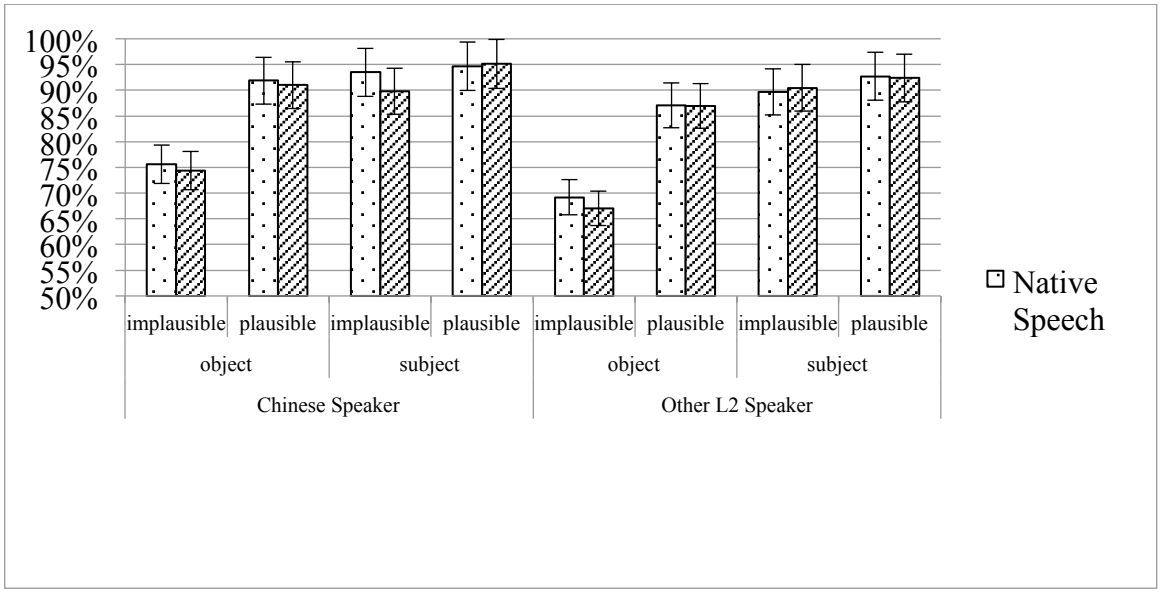


Figure 4.2. Response Accuracy across different plausibility, syntactic structure, "speaker", and subjects' native language conditions. Error bar stands for the standardized errors.

Figure 4.2 shows response accuracy across all conditions, and Table 4.4 shows results for the logit model of response accuracy. Overall, native Chinese speaking subjects achieved better accuracy than other L2 English speaking subjects ($p < 0.05$). The

main effects of plausibility ($p < 0.05$) and syntactic structure ($p < 0.05$) were significant. Although the main effect of speech was not significant ($p = 0.56$), speech significantly affected response accuracy interactively with other factors ($p< 0.05$). Specifically, when native Chinese speakers simulated the non-native speech while reading the implausible SRC sentences, their comprhesion accuracy was significantly lower than when they mentally simualted the native speaker's voice. For other L2 speaking subjects, perceptual simulation of non-native speech led to significnatly lower accuracy in the implausible ORC condition. Moreover, the accuracy difference beween ORCs and SRCs was more prominent among native Chinese speakers than other L2 English speakers, as indexed by the significant interaction between language group and structure ($p < 0.05$).

Table 4.4

Fixed Effects of Logit Mixed Effects Model for Response Accuracy

	Estimate	SE	z value	*p* value
Intercept	1.74	0.3	5.74	< 0.001
Plausibility	2.12	0.39	5.45	< 0.001
Structure	1.85	0.38	4.81	< 0.001
Speech	-0.06	0.11	-0.58	0.56
Group (Chinese vs. Other L2)	-0.88	0.24	-3.65	< 0.001
Working Memory	0.17	0.11	1.49	0.14
Plausibility x Structure	-0.8	0.57	-1.39	0.16
Plausibility x Speech	0.55	0.22	2.52	< 0.05
Structure x Speech	0.6	0.21	2.94	< 0.01
Group x Plausibility	-0.06	0.18	-0.34	0.74
Group x Structure	0.69	0.18	3.76	< 0.001
Group x Speech	0.24	0.14	1.67	< 0.1
Speech x Plausibility x Structure	-1.29	0.35	-3.62	< 0.001
Group x Plausibility x Structure	-0.61	0.31	-1.97	< 0.05
Group x Plausibility x Speech	-0.61	0.27	-2.29	< 0.05
Group x Structure x Speech	-0.97	0.26	-3.66	< 0.001
Group x Structure x Speech x Plausibility	1.22	0.44	2.79	< 0.01

Social attractiveness survey

Fourteen variables were used to evaluate and compare readers' attitudes towards the American-English and Chinese-English speakers. Table 4.5 presents the means and standard deviations of subjects' attitudes towards the native and non-native speech. ANOVA analysis revealed that native Chinese speaking subjects' attitudes were more negative towards the Chinese-English speaker than the American-English speaker on all variables except Education ($F(1, 47) = 6.01$; $p = 0.06$), which is marginal. Other L2 English speakers rated the Chinese-English speaker more negatively than the American-English speaker on all variables except Sincerity ($F(1, 47) = 1.42$; $p = 0.24$), Conscientiousness ($F(1, 47) = 0.78$; $p = 0.38$), Honesty ($F(1, 47) = 0.71$; $p = 0.40$), and Education ($F(1, 47) = 0.28$; $p = 0.60$). Compared to other L2 speakers, Native Chinese speaking subjects perceived the accent of the American-English speaker more positively and rated the accent of the Chinese-English speaker more negatively ($F(1, 94) = 4.30$; $p < 0.05$). Moreover, native Chinese speaking subjects rated the honesty of the native speaker higher ($F(1, 94) = 5.52$; $p < 0.05$) and the intelligence of the non-native speaker lower ($F(1, 94) = 6.37$; $p < 0.05$) than the other L2 speakers did.

Table 4.5

Comparison of attitudes towards native and non-native English speech between 48 native Chinese speakers and 48 non-Chinese L2 English speakers

	Native English Speech				Non-native English Speech			
	Chinese		*L2*		*Chinese*		*L2*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Accent	1.98	1.99	1.56	1.46	5.06	1.55	5.69	1.34
Comprehensibility	6	1.17	6.06	1.59	5.15	1.09	4.48	1.53
Speed	4.77	1.02	4.69	1.07	3.58	0.87	3.46	0.97

Table 4.5 (cont.)

	Native English Speech				Non-native English Speech			
	Chinese		*L2*		*Chinese*		*L2*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Confidence	6.21	1.05	6.38	0.82	4.54	1.38	4.21	1.4
Conscientiousness	5.54	1.01	5.13	1.25	5.02	0.93	4.96	1.11
Dependability	5.4	1.32	5.23	1.24	4.81	1.25	4.69	1.11
Education	5.83	1.02	5.92	0.99	5.52	1.17	5.81	1.27
Honesty	5.77	0.93	5.31	1.11	5.44	1.15	5.44	1.09
Intelligence	5.48	1.15	5.31	1.07	4.94	1.12	5.44	1.05
Likability	5.79	1.13	5.56	1.11	4.92	1.3	4.98	1.51
Pleasantness	5.98	1	5.67	1.08	4.4	1.4	4.23	1.73
Prestige	4.85	1.7	5.25	0.98	4.06	1.54	4.63	1.14
Reliability	5.85	1.05	5.58	1.11	4.81	1.18	4.98	1.08
Sincerity	5.63	1.04	5.48	1.18	5.31	0.99	5.23	1.24

Discussion

This experiment aimed at exploring whether L2 English speakers can perform auditory perceptual simulation and, if so, how APS affects their reading speed and comprehension. Results revealed that APS of different speakers modulated both native Chinese and other L2 English speakers' reading speed in the same direction as observed in studies 1 and 2. Perceptually simulating a faster speaker's voice led to faster sentence reading compare to perceptually simulating a slower speaker's voice, regardless of readers' attitudes and their own accent towards the native and non-native speech in the social attractiveness survey. These results show that APS affects the reading speed of L2 speakers similarly to how it affects the native English readers. Readers' working memory capacity marginally affected the sentence reading time. High working memory readers tended to read faster than low working memory readers. Although both native Chinese speaking subjects and other L2 English speaking subjects perceived the non-native speech more negatively according to the social attractiveness survey results, these offline

attitudes measures did not predict the online reading time or response accuracy, which supports the conclusion in study 2 that the APS effect was driven by speech rate rather than accents. Plausibility and Structure influenced L2 speakers' reading time in the same pattern as they did for native English speakers, suggesting that the materials are reliable.

For response accuracy, surprisingly, APS of different speakers significantly influenced the response accuracy in specific sentence conditions for each language group subjects, which was not the case for native English speakers in studies 1 and 2. When readers perceptually simulated the native English speaker's voice, they were more likely to achieve accurate comprehension of the implausible sentences. Figure 4.2 demonstrated that this speech effect was more salient for accuracy in the implausible SRCs for native Chinese L2 English speakers and implausible ORCs for L2 English speakers from other language groups, but was not significant in the plausible SRCs. One possible explanation is that APS of the non-native speech imposed a bigger working memory demand for other L2 English speakers when they process sentences with implausible meanings and harder syntactic structure, so that they failed to resolve the difficulty to reach the right interpretation. However, the offline working memory capacity was not a significant predictor of the accuracy. Another explanation is that APS of non-native speech changed the processing pattern of the subjects so that they tended to give up reanalyzing the plausibility or resolving the structural complexity. In Chapter 5, I will discuss more about how APS of native and non-native speech changes readers' language processing patterns.

Moreover, although APS of non-native speech slowed the reading speed and was not as beneficial to comprehension compared to the APS of native speech when processing implausible sentences, it is possible that this APS condition is still better than

normal silent reading without APS as the findings in study 1. Additional experiments will be needed to examine whether APS of native or non-native speech can facilitate L2 speakers' reading speed and comprehension compared to normal silent reading.

CHAPTER 5

IS IMAGINING A VOICE LIKE LISTENING TO IT?: EVIDENCE FROM ERPS

Morphosyntactic Processing and Event-Related Brain Potentials (ERPs)

Both native (L1) English speakers and English second-language (L2) learners make morphosyntactic errors in oral and written language production, but sometimes for somewhat different reasons. L1 English speakers' errors are believed to be due to the misapplication of morphological and syntactic rules (e.g., Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001), whereas L2 English speakers' errors are sometimes additionally due to incomplete learning of L2 morphosyntax, sometimes because it is absent in their L1 (Dowens, Vergara, Barber, & Carreiras, 2010; Scherag, Demuth, Rösler, Neville, & Röder, 2004). Both groups often overlook grammatical errors when they are comprehending language, such as failures in subject-verb agreement (e.g., *He walk home every day*) or gender agreement (e.g., *Mary saw himself in the mirror*). Errors such as failures in subject-verb agreement (e.g., *He walk home every day*) or gender agreement (e.g., *Mary saw himself in the mirror*) are therefore frequently encountered during language comprehension. The effects that such errors have on the language comprehender may vary, however, depending on who commits the error.

The electrophysiological response to morphosyntactic errors, measured via electroencephalographic recordings as event-related potentials (ERPs), is typically an increase in positivity peaking at around 600 milliseconds (ms) after the appearance of the offending word, usually called a P600 effect (Osterhout & Mobley, 1995; Osterhout & Nicol, 1999). There are also sometimes effects on other ERP components, including the Anterior Negativity (AN) and the N400, depending on the type of error (see Molinaro,

Barber, & Carreiras, 2011 for a review). Some kinds of errors also elicit more robust responses than others (Barber& Carreiras, 2005; Coulson et al, 1998). Coulson et al (1998) found bigger and more consistent P600 responses to pronouns whose case was wrong for their function in a sentence (e.g., *The plane took we to paradise and back*) than for verbs that disagreed in number with their subject. They argued that the higher salience of the pronoun mismatch errors and the lower frequency of encountering these errors in daily life led to their bigger P600 effects.

Similar P600 responses are observed for errors in visual and auditory sentences (Hagoort & Brown, 2000). There are studies using ERPs to examine how listeners respond to accented speech (Goslin, Duffy, & Floccia, 2012; Romero-Rivas, Martin, & Costa, 2015). In spoken sentences, P600 responses to grammatical errors can be modulated by whether the speaker has a native or non-native accent (Grey & Van Hell, 2017; Hanulíková, van Alphen, van Goch, & Weber, 2012). Hanulíková et al. (2012) compared listeners' electrophysiological responses to both semantically anomalous words and gender agreement errors in Dutch sentences spoken by either a native or a non-native speaker. They found that semantically anomalous words elicited similar N400 effects regardless of the accent of the speaker, but P600 responses to agreement violations were modulated by the type of speaker. There was a robust P600 effect in response to the native speaker but not the non-native speaker. They interpreted this result as showing that people are more likely to "forgive" grammatical errors in non-native speech.

Auditory Perceptual Simulation

Fans of the Harry Potter films might "hear" the actor Daniel Radcliffe's voice when silently reading Harry Potter's lines in the novels. This phenomenon of mentally

simulating a voice, either consciously or unconsciously, while reading silently is called Auditory Perceptual Simulation (APS) (Hubbard, 2010; Zhou & Christianson, 2016a, b). Previous behavioral studies have demonstrated that silent reading is affected by properties of an imagined speaker's voice. For example, reading speed can be modulated by information about the speed or other characteristics of the writer of a text (Drumm, & Klin, 2011; Gunraj & Klin, 2012; Kosslyn & Matt, 1977; Levine & Klin, 2001). Readers read sentences faster when they are told they come from a fast speaker than when they are told they come from a slow speaker, especially when texts are difficult (Alexandar & Nygaard, 2008). Familiarity with a speaker also affects the recognition of words encountered while engaging in auditory imagery. When a familiar speaker's voice is imagined during reading, readers have better recall of content when they imagining a familiar speaker was reading the content to them compared to when imagining a unfamiliar speaker was reading (Kurby, Magliano, & Rapp 2009). Previous neuroimaging studies reveal that imagining sounds activates similar regions of the brain as actually listening to those sounds. For example, similar regions are activated when listening or imagining music (i.e. superior temporal gyrus (STG) supramarginal gyrus, the middle frontal (Mid F) and inferior frontal (Ant Inf F) cortices (Zatorre, Halpern, Perry, Meyer, & Evans, 1996)). Similarly, Yao, Belin, and Scheepers (2011; see also Yao et al., 2012) found that voice-selective regions of readers' auditory cortices are more highly activated during silent reading when people read direct quotes compared to indirect quotes.

Eye-tracking studies (Stites, Luke, & Christianson, 2013; see also Yao & Scheepers, 2011) have revealed that the speed of reading sentences containing direct

quotes is modulated by descriptive adverbs preceding the quote, as illustrated in (1) below. The quotation portion of the sentence was read faster when the adverb preceding it was, e.g., *quickly* than when it was, e.g., *slowly*.

1) John walked into the room and said {*quickly/slowly}* that '*I found my car keys'*.

Zhou and Christianson (2016a, b) further examined how APS of different speakers' voices (faster vs. slower speakers) during silent reading affect reading speed and comprehension. Results revealed that APS of a faster native or non-native English speaker's voice led to faster silent reading compared to APS of a slower native or non-native English speaker's voice. This reading speed difference was triggered by the speech rate of the voices being simulated rather than the nativeness of the speaker's accent. In addition, APS of either a native or non-native English speaker's voice also resulted in the college-aged participants achieving up to 20% better comprehension accuracy for difficult sentences such as in (2a), in which syntactic structure (object-relative clause) and semantics (implausible thematic roles) were manipulated to induce extreme processing difficulty. These findings suggest that APS might modulate the efficiency and accuracy of language processing.

2) a. The bird that the worm ate was small.

b. The worm ate the bird. The bird was small. (True / False)

No study yet has investigated how APS of different speech affect readers' processing of sentences with grammatical errors. Yao et al.'s (2011, 2012) fMRI data

suggest that APS during silent reading is cognitively similar to listening to speech. If this is accurate, then the results of Hanulíková et al.'s (2012) study motivate a straightforward prediction. Recall that Hanulíková and colleagues observed that morphological errors produced verbally by non-native English speakers did not trigger inflated P600 signals in the ERP waveforms of native English speakers. Based on this result, Hanulíková and colleagues suggested that native speakers "forgive" errors that are uttered by non-native speakers, even on a neural level. If APS is indeed cognitively similar to listening to speech, then we would expect to observe a similar attenuation of P600s when native English speakers are silently reading and imagining the voice of a non-native speaker, compared to when they are imagining a native speaker making the same sort of error or when they are not performing APS. Thus, this study measures electrophysiological responses to grammatical errors in sentences that are read silently either with or without imagining a voice to test how similar APS of a voice is to actually hearing that voice.

Current Study

This study investigated how Auditory Perceptual Simulation (APS) of native and non-native speech affects the reading of ungrammatical sentences (e.g., subject-verb number mismatch, "*He enjoy swimming*") compared to silent reading without APS. Specifically, two research questions were addressed: 1) whether and how APS differs from normal silent reading when processing ungrammatical sentences, and 2) whether APS of native and non-native speech differentially affect native English speakers' recognition of grammatical errors during reading. Two event-related potential (ERP) experiments were conducted to address these questions, because ERPs provide subtle electrophysiological responses to grammatical errors. Experiment 1 measured responses

to ungrammatical sentences containing either subject-verb number disagreements or incorrect pronoun case during normal silent reading (see Table 5.1), to serve as a baseline for Experiment 2. Experiment 2 measured readers' responses to ungrammatical sentences when engaging in APS of either a native or a non-native English speaker's voice. The rationale for conducting two experiments with different participants arose from concern about the possibility of asymmetric carryover in a fully within-participants design that could not be removed by counterbalancing the order of conditions. That is, it seemed likely that participants who were instructed to engage in APS at the beginning might continue to do so despite being later instructed not to.

Experiment 1

Experiment 1 examined readers' electrophysiological responses when they read sentences silently with no instructions about engaging in APS. It was conducted to provide a baseline against which to compare the results of Experiment 2, in which APS instructions would be manipulated.

Method

Participants: Forty-eight native English speakers (21 females) from the university community participated in Experiment 1. All of them were 18-30 years old (M = 20), right-handed, and had normal hearing, normal or corrected-to-normal vision, and no history of neurological or psychiatric disorders.

Materials and Design: Two types of grammatical errors, subject-verb number agreement and pronoun case mismatch, were included for two different reasons. First, many previous ERP studies of morphosyntactic error processing have used sentences containing violations of subject-verb number agreement, and we wanted to be able to

compare our results to those, so we included that error type. However, a few previous studies have found that mismatches in pronoun case elicit larger P600 effects than subject-verb number agreement violations do (e.g., Coulson et al., 1998), and we wanted to include conditions with a strong likelihood of eliciting P600 effects. Another reason for including both kinds of errors is that people are more likely to have heard non-native speakers make subject-verb agreement errors than pronoun case errors. Subject-verb agreement errors are especially common in the speech of native Chinese speakers when they speak English, as Chinese does not have number agreement[9].

Experimental materials included 120 target sentence sets, 92 fillers, and 6 practice items. An example of a target sentence set and the paraphrase verification question that followed it is presented in Table 5.1 below (see all materials in Appendix C). Sentence triplets were constructed with a grammatical version that included both a correct verb and a correct pronoun, which served as the baseline for both types of grammatical error, and the other two triplet members contained errors in either verb number or pronoun case. Eighty of the filler sentences were experimental items for another study investigating how readers process relative clause sentences varying in semantic plausibility (e.g. *The bird that ate the worm was small*), whose results are not included here. The remaining 12 fillers were half plausible and half implausible sentences. Items were counterbalanced over three lists[10] in a Latin square design, such that each participant saw just one version of each item triplet and saw equal numbers of items (40) in each of the conditions. Item

[9] There is no agreement of any sort in Chinese. So native English speakers may be more forgiving of the subject-verb disagreement errors made by native Chinese speakers. Chinese also does not mark case on pronouns, but errors involving pronoun case are nonetheless less common in their speech.

[10] The grammatical sentence serves as the control grammatical sentence for both subject-verb mismatch and pronoun case mismatch. Thus, there were three conditions instead of four in Experiment 1, and six conditions instead of eight in Experiment 2.

order was randomized and adjusted so that the first two items in each block were not target sentences, and target sentences were separated by fillers. Each list was presented in the same order divided into six blocks of 34-38 trials and each participant saw only one list. Each trial included a sentence followed by a paraphrase verification task.

Table 5.1

Design of stimulus materials. Critical verbs and pronouns are underlined

Grammatically correct	*The carpenters chat when they sand the wood.*
Subject-verb mismatch	*The carpenters chats when they sand the wood.*
Pronoun case mismatch	*The carpenters chat when them sand the wood.*
Paraphrase	*The carpenters chatted and worked.* (True/False)

EEG recording: EEG was recorded using an EasyCap® with sintered Ag/AgCl electrodes placed at 28 scalp sites plus the right mastoid, all referenced to the left mastoid and later re-referenced to the average of the left and right mastoids. EOG was recorded with bipolar electrode pairs above and below the right eye and at the external canthi of both eyes to capture vertical and horizontal eye movements. Electrode impedances were maintained below 5kΩ. EEG and EOG were amplified and filtered with a Grass Model 12 amplifier with a band pass of 0.1-30 Hz, and sampled at a rate of 200 Hz using the IWave® software package.

Procedure: Each experimental session lasted 2.5-3.0 hours. Participants were randomly assigned to one of the three stimulus lists. They were instructed to sit in front of a monitor and read sentences silently before responding to a paraphrase verification task after each one, and brainwaves were recorded throughout. The sentences were presented one word at a time in the center of the screen at rate of 400 ms/word (words remained on the screen for 350 ms followed by a blank screen for 50 ms). Next a

paraphrase of the sentence was presented all at once and participants were asked to indicate whether it was correct, by pressing one of two buttons on a Cedrus RB-840 response box. Participants initiated each trial by pushing another button on the response box. Stimuli were presented and behavioral responses were recorded using the Presentation^{Ò} software package.

After the ERP-recording session, participants completed a survey asking about demographic and language background. The Nelson-Denny Reading Task (NDRT) was administered to assess their reading speed and comprehension proficiency, and the vividness of their imagination was measured using the Bucknell Auditory Imagery Scale (BAIS) (Herholz, Halpern, & Zatorre**,** 2012**).**

EEG data analysis

EEG and EOG data were analyzed using the EEGLab (Delorme & Makeig, 2004) and ERPLab (Lopez-Calderon & Luck, 2014) toolboxes in Matlab. Epochs were extracted from 100 ms preceding the onset of the critical word through 1000ms later. Trials contaminated by excessive artifacts (e.g., blinks, eye movements, other muscle activity) were rejected using the ERPLAB toolbox. Participants for whom more than 15% of the trials were lost due to artifacts were dropped from the analyses, leading to 5 dropped subjects, leaving 43 whose data were included in the analyses. 11.1% of the trials from the remaining 43 subjects were excluded from the final analyses due to eyeblinks or movement artifacts.

The N400 and P600 ERP components in response to the critical words were of primary interest. They were quantified as the mean voltage within the time windows 300-500 ms and 600-900 ms after critical word onset, baselined on 100 ms before its onset.

(Visual inspection of the waveforms showed that P600 effects in some conditions were not well captured by a 500-800 ms window, so 600-900 ms was used instead.) To assess the scalp distribution of ERP effects, separate analyses were conducted on the midline electrodes (Fz, Cz, Pz) and on the rest of the electrodes grouped into four quadrants according to laterality and anteriority on the head, each consisting of five electrodes and defined as follows: left-anterior (AF3 F3 FC3 F7 FT7), right-anterior (AF4 F4 FC4 F8 FT8), left-posterior (CP3 P3 T5 P5 PO7), and right-posterior (CP4 P4 T6 P6 PO8). Using the R software package with the lmer() function of the lme4.0 package (Bates, Maechler, & Bolker, 2011), linear mixed-effects models (Baayen, Davidson, & Bates, 2008) were built with the mean voltages for each time window for each participant in each condition as the dependent variable, and grammaticality (grammatical/ungrammatical), word type (verb/pronoun), electrode laterality (left/right), electrode anteriority (anterior/posterior), Nelson-Denny reading speed, and Nelson-Denny comprehension scores as the fixed effects and participants as random effects, using the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). The waveforms were digitally filtered with a bandpass of 0.1-12 Hz to smooth them for the figures, but statistical analyses were conducted before such filtering was applied.

Results

Behavioral Results

Participants achieved an average comprehension accuracy of 94% in the online paraphrase verification task for the target sentences. In the offline Nelson-Denny reading speed and comprehension tasks, average reading speed was 225 words per minute and average comprehension accuracy was 78%. The average vividness of auditory imagery

was 71 out of 98.

EEG data analysis

The target words were either verbs or pronouns (underlined in Table 5.1 above), whose mean response amplitudes were calculated for both the N400 (300-500 ms) and P600 (600-900 ms) time windows, baselined on 100 ms prior to target word onset. Grammaticality and word type were factors in all of the analyses, while electrode location factors differed between the analyses of midline and lateral sites. For midline sites, there was single anteriority factor with three levels (anterior/middle/posterior), while for lateral sites there was an additional electrode laterality factor (right/left) and anteriority had just two levels (anterior/posterior). In addition, in all of the analyses, speed and comprehension accuracy in the Nelson-Denny Reading Test were also included as predictors.

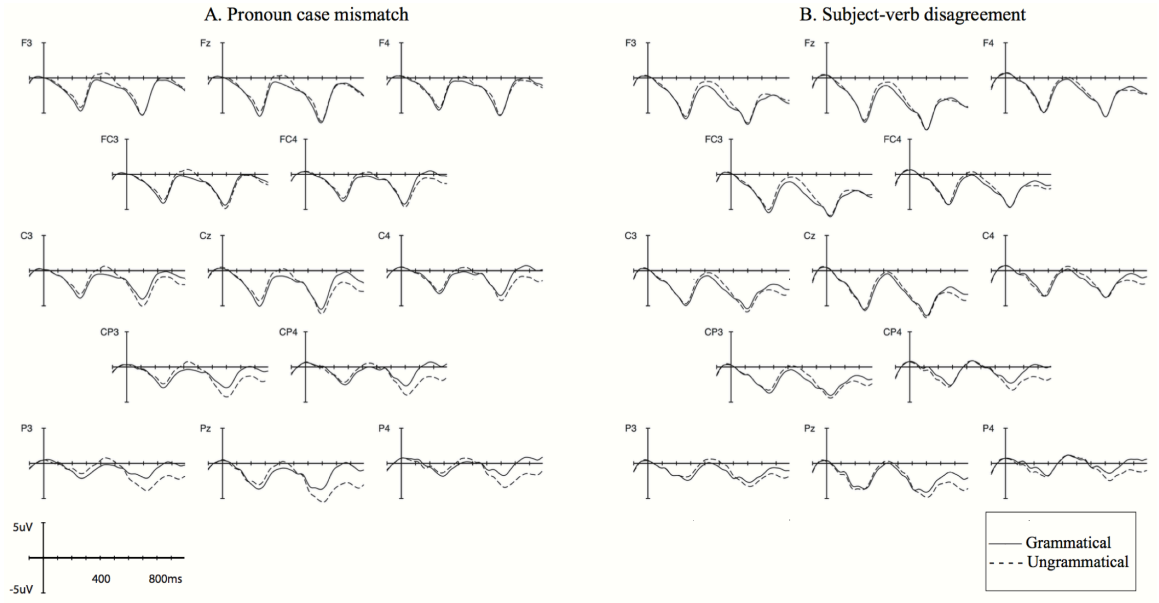


Figure 5.1. *Grand mean ERPs for the target pronouns (A) and verbs (B) in Experiment 1*

Figure 5.1 above shows the grand mean ERP responses to target pronouns (A) and verbs (B), and Tables 5.2-5.5 show the results of the statistical analyses of the P600

(Tables 5.2 and 5.3) and N400 (Tables 5.4 and 5.5) time windows at midline (Tables 5.2 and 5.4) and lateral (Tables 5.3 and 5.5) scalp sites. Visual inspection shows that mean voltages in the ungrammatical conditions were both more positive during the P600 window and more negative during the N400 window than in the grammatical conditions. These grammaticality effects differed across the head, leading to reliable interactions between electrode location factors and grammaticality in most of the analyses.

For the P600 window, the effect of grammaticality was largest at the back of the head, leading to reliable interactions between grammaticality and anteriority in both the midline ($p < 0.001$) and lateral ($p < 0.001$) analyses, rather than a main effect of grammaticality ($p > 0.1$). In the N400 window, in contrast, the effect of grammaticality was consistent enough across the head to yield reliable main effects of grammaticality in both midline ($p < 0.001$) and lateral ($p < 0.001$) analyses, but also different enough between the front and back of the head to yield an interaction between grammaticality and anteriority in the lateral analysis ($p < 0.001$).

Although Figure 1 shows a somewhat larger effect of grammaticality for pronouns than for verbs, especially during the P600 window, the two-way interaction between grammaticality and word type was not reliable in any of the analyses ($p > 0.1$). However, there were reliable three-way interactions between electrode location factors and the grammaticality x word type interaction in both the midline ($p < 0.05$) and lateral ($p < 0.05$) analyses of the P600 window, which arose because the size of the grammaticality effect for the two types of words tended to differ most at the back of the head. There were no similar interactions with scalp location factors for the N400 window ($p > 0.1$).

There were also some overall differences between the responses to pronouns and verbs, shown by both reliable main effects of word type in three of the four analyses (*p*'s < 0.01) and interactions between word type and electrode location factors in all four analyses (*p*'s < 0.01). This pattern arose because responses to the pronouns were more negative overall than to verbs, especially at the front and on the left side of the head.

Table 5.2

Midline analysis for the P600 window in Experiment 1

Predictor	Est	SE	df	*t*	*p*
Intercept	-0.28	0.16	53.4	-1.8	< 0.1
Grammaticality (Gram)	0.12	0.17	68.8	0.7	0.5
Gram x Anteriority$_1$ (Ant$_1$=Fz vs Cz)	**-0.35**	**0.15**	**320**	**-2.41**	**< 0.05**
Gram x Anteriority$_2$ (Ant$_2$=Fz vs Pz)	**-0.75**	**0.15**	**320**	**-5.17**	**< 0.001**
Gram x Word Type (Wtype)	-0.09	0.22	79.2	-0.39	0.7
Gram x Wtype x Ant$_1$	0.24	0.21	320	1.17	0.24
Gram x Wtype x Ant$_2$	**0.47**	**0.21**	**320**	**2.28**	**< 0.05**
Wtype	**0.62**	**0.16**	**74.4**	**3.84**	**< 0.01**
Wtype x Ant$_1$	-0.23	0.15	320	-1.57	0.12
Wtype x Ant$_2$	**-0.67**	**0.15**	**320**	**-4.63**	**< 0.001**
Ant$_1$	0.2	0.1	320	1.99	< 0.05
Ant$_2$	0.39	0.1	320	3.81	< 0.001
NDRT Comprehension	0.03	0.11	38	0.28	0.78
NDRT Reading Speed	0.12	0.11	38	1.12	0.27

Table 5. 3

Lateral analysis for the P600 window in Experiment 1

Predictor	Est	SE	df	*t*	*p*
Intercept	0.04	0.14	47.3	0.29	0.77
Grammaticality (Gram)	0.12	0.14	67.3	0.87	0.39

Table 5.3 (cont.)

Predictor	Est	SE	df	*t*	*p*
Gram x Anteriority (Ant)	**-0.58**	**0.1**	**404**	**-5.98**	**< 0.001**
Gram x Laterality (Lat)	-0.1	0.1	404	-1.03	0.3
Gram x Word Type (Wtype)	0.01	0.19	71.4	0.03	0.97
Gram x Wtype x Ant	**0.31**	**0.14**	**404**	**2.21**	**< 0.05**
Gram x Wtype x Lat	-0.08	0.14	404	-0.6	0.56
Wtype	**0.73**	**0.15**	**66**	**5.04**	**< 0.001**
Wtype x Ant	**-0.66**	**0.1**	**404**	**-6.77**	**< 0.001**
Wtype x Lat	**-0.35**	**0.1**	**404**	**-3.55**	**< 0.001**
Ant	-0.12	0.11	77.6	-1.05	0.3
Lat	-0.04	0.08	150.5	-0.48	0.64
NDRT Comprehension	0.08	0.08	38	1	0.32
NDRT Reading Speed	0.03	0.08	38	0.35	0.73

Table 5.4

Midline analysis for the N400 window in Experiment 1

Predictor	Est	SE	df	*t*	*p*
Intercept	-0.42	0.17	84.5	-2.55	< 0.05
Grammaticality (Gram)	**0.52**	**0.15**	**145.4**	**3.46**	**< 0.001**
Gram x Anteriority$_1$ (Ant$_{1=}$Fz vs Cz)	-0.06	0.16	320	-0.37	0.71
Gram x Anteriority$_2$ (Ant$_{2=}$Fz vs Pz)	-0.08	0.16	320	-0.51	0.61
Gram x Wtype x Ant$_1$	-0.06	0.23	320	-0.24	0.81
Gram x Wtype x Ant$_2$	-0.08	0.23	320	-0.36	0.72
Wtype	**0.59**	**0.2**	**74.8**	**2.9**	**< 0.01**
Wtype x Ant$_1$	-0.14	0.16	320	-0.86	0.39
Wtype x Ant$_2$	**-0.43**	**0.16**	**320**	**-2.64**	**< 0.01**
Ant$_1$	-0.14	0.16	320	-0.86	0.39
Ant$_2$	-0.43	0.16	320	-2.64	< 0.01
NDRT Comprehension	0.01	0.08	38.6	0.11	0.91
NDRT Reading Speed	0.13	0.08	38.6	1.54	0.13

Table 5.5

Lateral analysis for the N400 window in Experiment 1

Predictor	Est	SE	df	*t*	*p*
Intercept	-0.29	0.14	51.2	-2	< 0.1
Grammaticality (Gram)	**0.5**	**0.12**	**125.2**	**4.08**	**< 0.001**
Gram x Anteriority (Ant)	**-0.46**	**0.12**	**404**	**-3.89**	**< 0.001**
Gram x Laterality (Lat)	-0.01	0.12	404	-0.05	0.96
Gram x Word Type (Wtype)	-0.06	0.2	93.2	-0.31	0.76
Gram x Wtype x Ant	-0.11	0.17	404	-0.68	0.49
Gram x Wtype x Lat	-0.1	0.17	404	-0.58	0.56
Wtype	**0.76**	**0.2**	**57.6**	**3.74**	**< 0.001**
Wtype x Ant	-0.2	0.12	404	-1.66	< 0.1
Wtype x Lat	**-0.41**	**0.12**	**404**	**-3.46**	**< 0.001**
Ant	0.09	0.12	317.7	0.77	0.44
Lat	0.07	0.13	221.7	0.54	0.59
NDRT Comprehension	0.02	0.06	38	0.36	0.72
NDRT Reading Speed	0.04	0.06	38	0.62	0.54

Experiment 1 Summary

In sum, readers showed a classic P600 response to ungrammatical words when reading the sentences used here, with a somewhat larger effect for pronouns than for verbs, replicating previous work comparing the two types of grammatical errors (Coulson, King, & Kutas, 1998). Thus, the materials reliably elicited P600 responses, with pronouns doing so more robustly.

There were also smaller effects of grammaticality in the N400 window, which is consistent with some previous studies of sentences containing grammatical errors, which sometimes elicit effects on N400 in addition to (or sometimes even instead of) P600 (Tanner & Van Hell, 2014). There is currently much discussion in the ERP language

literature about how to explain tradeoffs between N400 and P600 responses, with factors such as task differences and individual differences invoked as explanations (e.g., Tanner & Van Hell, 2014). Further discussion of this will be postponed until after the second experiment is presented. For now, what is important is the demonstration that the stimuli produced robust grammaticality effects when people were not given any instructions to engage in Auditory Perceptual Simulation. (It is not possible, though, to know whether any of the participants did so spontaneously.)

Experiment 2

Experiment 2 examines electrophysiological responses to grammatical errors when readers imagine either a native English speaker or a non-native speaker while reading sentences silently. Eye-tracking (Stites et al., 2013; Zhou & Christianson, 2016a,b) and fMRI (Yao et al 2012a, b) studies have shown that engaging in APS affects behavioral responses and brain activation patterns. Readers' reading speed are modulated by texts (quickly vs. slowly) or the speech rates of the speakers (native vs. non-native) that they were perceptually simulated. Readers' APS can be triggered by direct quotation marks during silent reading and their auditory cortex are activated during APS. However, no previous study has investigated whether and how engaging in APS affects electrophysiological responses to grammatical errors.

Previous ERP work has shown that listeners' electrophysiological responses to words in spoken sentences are affected by whether the sentences are spoken in a native or non-native accent (Grey & Van Hell, 2017; Hanulíková et al., 2012). Most directly relevant is Hanulíková and colleagues' (2012) study demonstrating that listeners respond

differently to grammatical errors when they are spoken by native and non-native speakers. Specifically, they found that listeners seemed to forgive grammatical errors made by a non-native speaker. The question addressed in Experiment 2 is whether imagining native and non-native speaker voices while reading is similar enough to actually hearing those voices to yield the same effects.

Method

Participants: Eighty native English speakers (37 females, 48 males) from the university community participated in Experiment 2. All were 18-30 years old (mean =19.8), right-handed, had normal hearing and normal or corrected-to-normal vision, and reported no history of neurological or psychiatric disorders. Eight subjects' data were excluded due to experimental errors, leaving 72 whose data were included in the analyses.

Materials and Design: A 2 x 3 factorial design was created by adding a manipulation of the voice that participants were instructed to imagine while reading silently (native/non-native) to the factors of word type (verb/pronoun) and grammaticality (grammatical/ungrammatical). The same 120 experimental item sets from Experiment 1 were used again and another 60 sets were added. The relative-clause-containing filler sentences from Experiment 1 were not included. The addition of the APS manipulation required counterbalancing across 6 lists in a Latin square design, such that each participant saw just one version of each item triplet imagined in the voice of one speaker and saw equal numbers of items paired with each imagined speaker in each condition (30). Item order was randomized and then adjusted so that the first two items in each block were not target sentences and target sentences were separated by fillers. Each list was presented in the same order divided into six blocks of 34-38 trials and each

participant saw only one list. Each trial consisted of a sentence followed by a verification statement.

To induce participants to imagine the speech of either a native or a non-native speaker of English while reading the sentences, photos of one Caucasian female and one Asian female were shown and short recordings were played of native and non-native women's voices, respectively. The photos showed two women of similar age wearing similar clothing, who were not actually the speakers whose recordings were played (see Zhou & Christianson, 2016) but who were consistent with the voices. The recorded speech examples came from twelve short paragraphs, balanced for length and difficulty, which were read aloud and recorded by both speakers. Each speaker also recorded a common name associated with their native language (English name: Susan, Chinese name: Xiaofu).

Participants' reading speed and comprehension were measured using the Nelson-Denny Reading Task (NDRT), and their ability to imagine auditory stimuli was measured with the Bucknell Auditory Imagery Scale (BAIS; Herholz, Halpern, & Zatorre, 2012). Two social attractiveness surveys measured participants' attitudes toward both the native and non-native English speakers, and another survey collected their demographic and language background.

Procedure

The ERP procedures were the same as in Experiment 1. After electrode application was completed, participants listened to two recordings (e.g., native speaker recording = 29 sec; non-native speaker recording = 37 sec, order counterbalanced across participants) while photographs were displayed on the screen. The sequence of events is

illustrated in Figure 2 below. Participants were instructed to sit in front of a monitor and read sentences silently while imagining that they were spoken by either the native or the non-native speaker they had just seen and heard. Before each trial, one of the speaker pictures was presented for 500 ms along with a recording of that speaker's name, varying randomly across trials. Sentences were presented word by word at the same rate (350 ms/word) as in Experiment 1. Also as in Experiment 1, participants made a paraphrase verification judgment after each sentence. At the beginning of each block, participants listened to recordings of the same native and non-native speakers reading different passages[11] to refresh their memory for the voices.

After ERP data collection was finished, participants completed the same questionnaires and took the same tests of their reading speed, comprehension, and auditory imagery abilities as in Experiment 1. Each experimental session lasted 2.5-3 hours.

EEG recording

The equipment and recording procedures were the same as Experiment 1.

EEG data analysis

Using the same procedures as in Experiment 1 for eliminating artifact-contaminated data, 4 participants were dropped and approximately 9% of the trials for the remaining participants were excluded from the final analyses. The same mixed-effects modeling as in Experiment 1 was used to analyze the results, with the addition of a factor

[11] The duration of the native English speech: text 1 = 35sec; text 2 = 25sec; text 3 = 27sec; text 4 = 28sec, text 5= 25sec, text 6= 24sec. The duration of non-native speech: text 1 = 37sec; text 2 = 46sec; text 3 = 47sec; text 4 = 40sec, text 5= 44sec, text 6= 41sec.

coding APS condition. Two sets of analyses were conducted, one comparing the native-APS condition in Experiment 2 to the no-APS data from Experiment 1, and the second comparing the native- and non-native-APS conditions in Experiment 2.

Results

Behavioral tasks

Participants achieved an average comprehension accuracy of 93% in the paraphrase verification task for the target sentences. Average reading speed was 267 words/min and average comprehension accuracy was 81.5% in the Nelson-Denny reading speed and comprehension tasks. Average vividness on the auditory imagery test was 72 out of 98.

Event-Related Potentials

No previous study has examined whether and how engaging in APS while reading sentences affects electrophysiological responses to grammatical errors, so it is first necessary to determine whether adding APS changes the responses observed in Experiment 1. Thus, electrophysiological responses to morphosyntactic errors while imagining a native speaker's voice will first be compared to the responses from Experiment 1 where there were no instructions to engage in APS while reading. After that, the responses in the non-native-speaker-APS condition will be compared with native-speaker APS responses.

Comparison of no-APS (Experiment 1) and native-speaker-APS (Experiment 2)

Figure 5.2 shows the grand mean ERP waveforms for grammatical and ungrammatical pronouns (A) and verbs (B) in the native-speaker-APS conditions. Tables 5.6-5.9 show the results of the statistical analyses of the P600 (Tables 5.6 and 5.7) and

N400 (Tables 5.8 and 5.9) time windows at midline (Tables 5.6 and 5.8) and lateral (Tables 5.7 and 5.9) scalp sites. Visual inspection of Figure 5.3 reveals a very similar overall pattern of responses as in Experiment 1 when there were no instructions to imagine a voice. However, comparing Figure 5.2 to Figure 5.1 suggests that grammaticality effects were smaller overall in the native-speaker-APS condition in Experiment 2 than when no voice was imagined in Experiment 1.

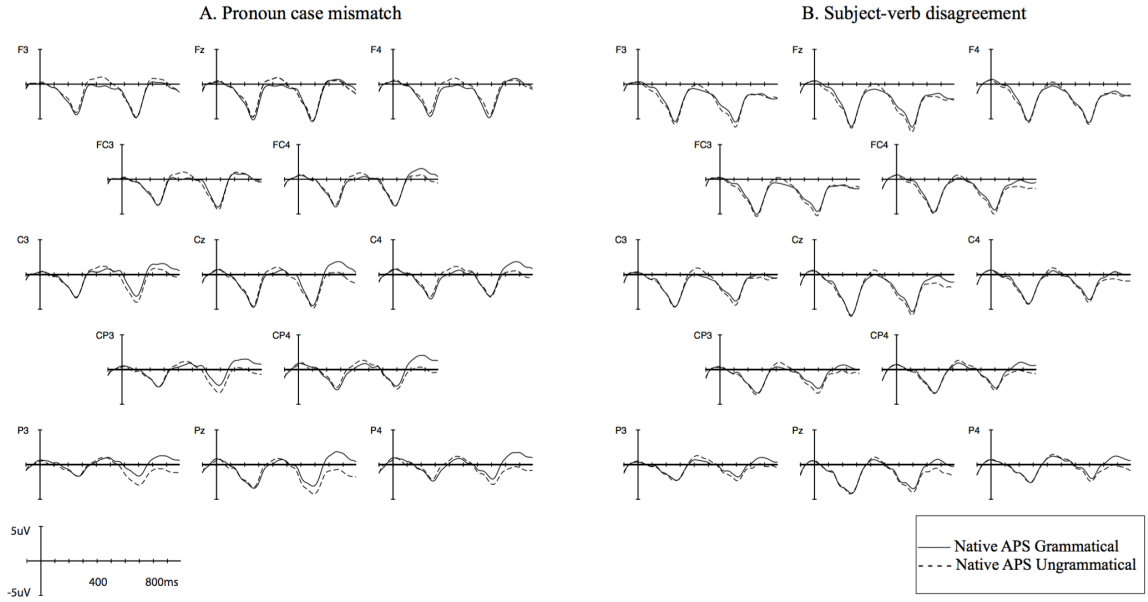


Figure 5. 2. *Grand mean ERP waveforms for the native-speaker-APS conditions for pronouns (A) and verbs (B) in Experiment 2*

Table 5.6

No-APS (Expt 1) vs Native-Speaker-APS (Expt 2), P600 window, midline sites

Predictor	Est	SE	df	*t*	*p*
Intercept	-0.31	0.11	111.6	-2.95	< 0.05
Grammaticality (Gram)	0.01	0.12	382.1	0.1	0.92
Gram x Anteriority$_1$ (Ant$_1$=Fz vs Cz)	**-0.28**	**0.13**	**888.9**	**-2.2**	**< 0.05**
Gram x Anteriority$_2$ (Ant$_2$=Fz vs Pz)	**-0.55**	**0.13**	**888.9**	**-4.37**	**< 0.001**

Table 5.6 (cont.)

Predictor	Est	SE	df	*t*	*p*
Gram x Word Type (Wtype)	-0.16	0.13	888.9	-1.24	0.22
Gram x Wtype x Ant_1	**0.01**	**0.21**	**888.9**	**0.05**	**< 0.05**
Gram x Wtype x Ant_2	**-0.01**	**0.21**	**888.9**	**-0.03**	**< 0.001**
Wtype	**0.63**	**0.12**	**352.4**	**5.29**	**< 0.001**
Wtype x Ant_1	-0.24	0.13	888.9	-1.89	0.06
Wtype x Ant_2	**-0.69**	**0.13**	**888.9**	**-5.45**	**< 0.001**
APS	**0.4**	**0.19**	**143**	**2.11**	**< 0.05**
APS x Gram	0.06	0.19	382.1	0.33	0.75
APS x Wtype	-0.05	0.2	352.4	-0.24	0.81
APS x Gram x Wtype	0.05	0.21	888.9	0.26	0.79
APS x Ant_1	0.16	0.15	782.8	1.05	0.29
APS x Ant_2	0.18	0.16	528.5	1.07	0.29
APS x Gram x Ant_1	-0.07	0.21	888.9	-0.34	0.74
APS x Gram x Ant_2	-0.2	0.21	888.9	-0.98	0.33
APS x Wtype x Ant_1	0.01	0.21	888.9	0.05	0.96
APS x Wtype x Ant_2	-0.01	0.21	888.9	-0.03	0.98
APS x Gram x Wtype x Ant_1	0.01	0.29	888.9	0.03	0.98
APS x Gram x Wtype x Ant_2	0.03	0.29	888.9	0.12	0.91
Ant_1	0.04	0.09	782.8	0.38	0.7
Ant_2	0.2	0.1	528.5	2.02	< 0.05

Table 5.7

No-APS (Expt 1) vs Native-Speaker-APS (Expt 2), P600 window, lateral sites

Predictor	Est	SE	df	*t*	*p*
Intercept	0.03	0.11	151.9	0.29	0.77
Grammaticality (Gram)	0.2	0.19	151.9	1.09	0.28
Gram x Anteriority (Ant)	0.08	0.15	286.1	0.53	0.59
Gram x Laterality (Lat)	0.05	0.13	440.4	0.4	0.69
Gram x Word Type (Wtype)	0.1	0.18	365.4	0.59	0.56
Gram x Wtype x Ant	-0.02	0.15	1229	-0.14	0.89
Gram x Wtype x Lat	**-0.35**	**0.15**	**1229**	**-2.33**	**< 0.05**
Wtype	**0.6**	**0.11**	**365.4**	**5.61**	**< 0.001**

Table 5.7 (cont.)

Predictor	Est	SE	df	*t*	*p*
Wtype x Ant	**-0.67**	**0.09**	**1229**	**-7.21**	**< 0.001**
Wtype x Lat	0	0.09	1229	-0.02	0.99
APS	-0.04	0.1	438	-0.39	0.7
APS x Gram	0.13	0.17	438	0.79	0.43
APS x Wtype	-0.11	0.11	1229	-0.99	0.32
APS x Gram x Wtype	0.1	0.19	1229	0.52	0.6
APS x Ant	-0.43	0.09	1229	-4.67	**< 0.01**
APS x Lat	0.06	0.09	1229	0.65	0.51
APS x Gram x Ant	-0.17	0.15	1229	-1.1	0.27
APS x Gram x Lat	-0.18	0.15	1229	-1.21	0.23
APS x Wtype x Ant	**0.35**	**0.13**	**1229**	**2.67**	**< 0.01**
APS x Wtype x Lat	-0.04	0.13	1229	-0.28	0.78
APS x Gram x Wtype x Ant	0	0.21	1229	0.02	0.98
APS x Gram x Wtype x Lat	-0.01	0.21	1229	-0.05	0.96
Ant	-0.09	0.08	440.4	-1.1	0.27
Lat	**-0.21**	**0.09**	**286.1**	**-2.33**	**< 0.05**

Table 5.8

No-APS (Expt 1) vs Native-Speaker-APS (Expt 2), N400 window, midline sites

Predictor	Est	SE	df	*t*	*p*
Intercept	-0.41	0.11	232.5	-3.62	< 0.01
Grammaticality (Gram)	**0.37**	**0.13**	**460.7**	**2.86**	**< 0.01**
Gram x Anteriority$_1$ (Ant$_1$, Fz vs Cz)	-0.19	0.15	999	-1.28	0.2
Gram x Anteriority$_2$ (Ant$_2$, Fz vs Pz)	-0.22	0.15	999	-1.5	0.13
Gram x Word Type (Wtype)	-0.19	0.15	999	-1.29	0.2
Gram x Wtype x Ant$_1$	0.17	0.21	999	0.8	0.42
Gram x Wtype x Ant$_2$	0.09	0.21	999	0.44	0.66
Wtype	**0.58**	**0.15**	**299**	**3.9**	**< 0.01**
Wtype x Ant$_1$	-0.15	0.15	999	-1.05	0.29
Wtype x Ant$_2$	**-0.44**	**0.15**	**999**	**-3.01**	**< 0.01**
APS	0.24	0.19	232.5	1.31	0.19
APS x Gram	0.08	0.21	460.7	0.37	0.71

Table 5.8 (cont.)

Predictor	Est	SE	df	*t*	*p*
APS x Wtype	-0.08	0.25	299	-0.32	0.75
APS x Gram x Wtype	0.03	0.24	999	0.14	0.89
APS x Ant_1	0.09	0.17	999	0.52	0.6
APS x Ant_2	0.01	0.17	999	0.09	0.93
APS x Gram x Ant_1	0.13	0.24	999	0.55	0.58
APS x Gram x Ant_2	0.14	0.24	999	0.56	0.57
APS x Wtype x Ant_1	-0.01	0.24	999	-0.04	0.97
APS x Wtype x Ant_2	0.02	0.24	999	0.07	0.94
APS x Gram x Wtype x Ant_1	-0.21	0.34	999	-0.6	0.55
APS x Gram x Wtype x Ant_2	-0.15	0.34	999	-0.44	0.66
Ant_1	0	0.1	999	0.02	0.99
Ant_2	0.05	0.1	999	0.45	0.66

Table 5.9

No-APS (Expt 1) vs Native-Speaker-APS (Expt 2), N400 window, lateral sites

Predictor	Est	SE	df	*t*	*p*
Intercept	-0.3	0.1	324.8	-3.11	< 0.01
Grammaticality (Gram)	**0.31**	**0.12**	**640.5**	**2.61**	**< 0.01**
Gram x Anteriority (Ant)	-0.17	0.12	1451	-1.43	0.15
Gram x Laterality (Lat)	0.11	0.14	1451	0.79	0.43
Gram x Word Type (Wtype)	-0.23	0.15	1451	-1.56	0.12
Gram x Wtype x Ant	0.23	0.17	1451	1.34	0.18
Gram x Wtype x Lat	-0.15	0.17	1451	-0.89	0.37
Wtype	**0.74**	**0.14**	**350.4**	**5.19**	**< 0.001**
Wtype x Ant	**-0.61**	**0.12**	**1451**	**-5.09**	**< 0.001**
Wtype x Lat	-0.12	0.12	1451	-1.04	0.3
APS	0.16	0.16	324.8	1	0.32
APS x Gram	0.15	0.2	640.5	0.78	0.43
APS x Wtype	-0.08	0.23	350.4	-0.33	0.74
APS x Gram x Wtype	0.18	0.24	1451	0.73	0.47
APS x Ant	-0.11	0.14	1451	-0.82	0.41
APS x Lat	0.11	0.14	1451	0.79	0.43
APS x Gram x Ant	0.15	0.2	1451	0.77	0.44
APS x Gram x Lat	-0.3	0.2	1451	-1.54	0.12

Table 5.9 (cont.)

Predictor	Est	SE	df	*t*	*p*
APS x Wtype x Ant	0.16	0.2	1451	0.83	0.41
APS x Wtype x Lat	-0.26	0.2	1451	-1.34	0.18
APS x Gram x Wtype x Ant	0.08	0.28	1451	0.3	0.76
APS x Gram x Wtype x Lat	-0.3	0.28	1451	-1.09	0.28
Ant_1	-0.11	0.09	1451	-1.31	0.19
Ant_2	0.07	0.09	1451	0.82	0.41

Just as in Experiment 1, ungrammatical verbs and pronouns elicited more positivity during the P600 window and more negativity during the N400 window than their grammatical counterparts. Also just as in Experiment 1, this led to a reliable main effect of grammaticality during the N400 window in both the midline and lateral analyses ($p < 0.01$), but during the P600 window there were instead interactions among grammaticality, word type, and electrode location factors. These arose because the effect of grammaticality was again bigger at the back of the head ($p < 0.05$ for grammaticality x anteriority interactions in midline analyses), but more so for verbs than for pronouns ($p < 0.05$ for grammaticality x word type x anteriority interactions in midline analyses). There was also an interaction among grammaticality, word type, and electrode laterality in the lateral analysis of the P600 window ($p < 0.05$) because grammaticality effects were slightly larger over the right hemisphere than over the left.

Also just as in Experiment 1, there were some overall differences between the two kinds of words in all of the analyses ($p < 0.01$), as well as interactions between word type and electrode location factors ($p < 0.01$), which arose because differences between the word types were larger at the front of the head. Finally, there were also some overall differences between the no-APS and native-speaker-APS conditions, (a main effect in the

midline analysis for the P600 window, $p < 0.05$, and some interactions with scalp location factors in the lateral analysis of the P600 window, $p < 0.01$). Most importantly, though, imagining a native English speaker's voice while reading silently did not change the pattern of responses elicited by ungrammatical words compared to their grammatical counterparts ($p > 0.1$ for all interactions involving APS and grammaticality). Figure 4 summarizes the grammaticality effects in the two APS conditions in each time window for each word type at the Pz electrode site, showing that the overall pattern was the same whether or not a voice was imagined. Figure 5.3 also shows, however, that all of the grammaticality effects were smaller when a voice was imagined. That diminishment was not sufficient to yield interactions between grammaticality and whether a voice was imagined, though it did produce a main effect of APS condition in the P600 window at midline sites only.

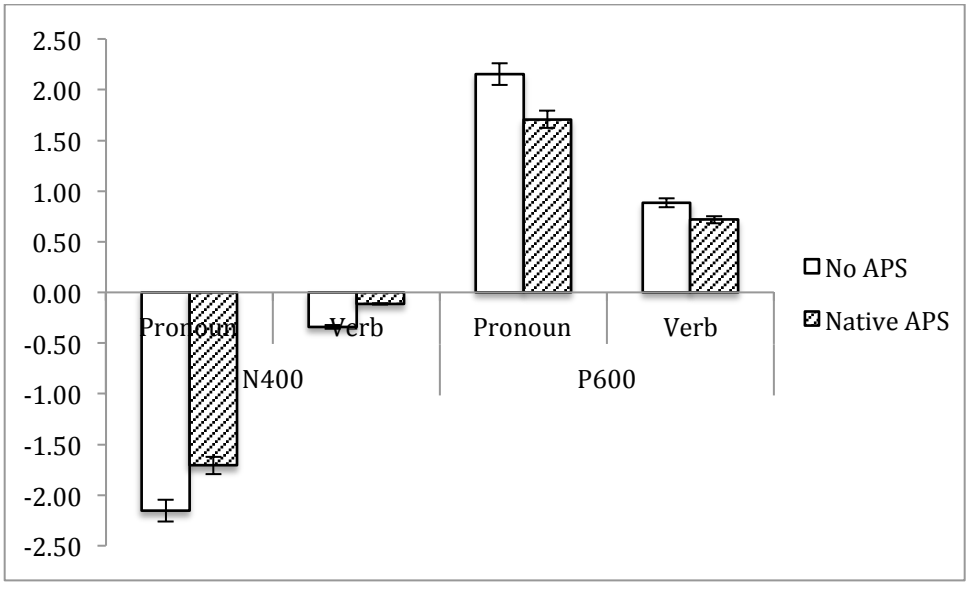


Figure 5.3. Grammaticality effect (ungrammatical minus grammatical) in N400 and P600 time windows at electrode Pz for the No-APS (Expt 1) and Native-Speaker-APS (Expt 2) conditions; error bars represent standard errors.

The finding that imagining a voice did not eliminate the usual electrophysiological response to grammatical errors in sentences is crucial to addressing the most important question here, which is whether readers "forgive" some grammatical errors if they imagine a non-native speaker's voice while they read, just as listeners do when listening to a voice with a non-native accent making those same kinds of errors (Hanulíková et al., 2012). Before that question could be addressed, it was first necessary to determine whether and how simply imagining any voice at all would affect electrophysiological responses to grammatical errors in sentences. The fact that imagining the voice of a native speaker diminished but did not extinguish such responses means that it provides the most appropriate baseline to compare responses to grammatical errors imagined in a non-native speaker's voice to, to which we turn next.

Comparison between native- and non-native-speaker APS in Experiment 2

We applied linear mixed effects modeling with maximal random effect structure, including APS type (native APS/non-native APS), grammaticality (grammatical/ ungrammatical), word type (pronoun/verb), and scalp distributions (anteriority in both midline and lateral analysis and laterality in just the lateral analyses) as the fixed effects, as well as subjects as a random effect. Tables 5.10-5.13 show the results of the statistical analyses of the P600 (Tables 5.10 and 5.11) and N400 (Tables 5.12 and 5.13) time windows at midline (Tables 5.10 and 5.12) and lateral (Tables 5.11 and 5.13) scalp sites.

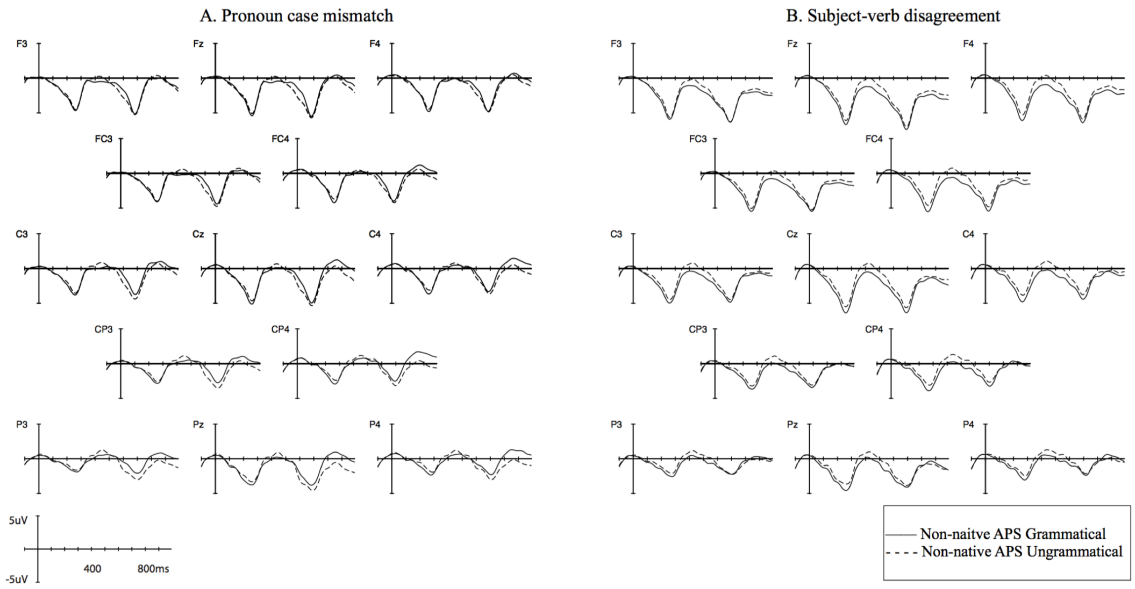


Figure 5.4. Grand mean ERP waveforms for the non-native-speaker-APS conditions for pronouns (A) and verbs (B) in Experiment 2.

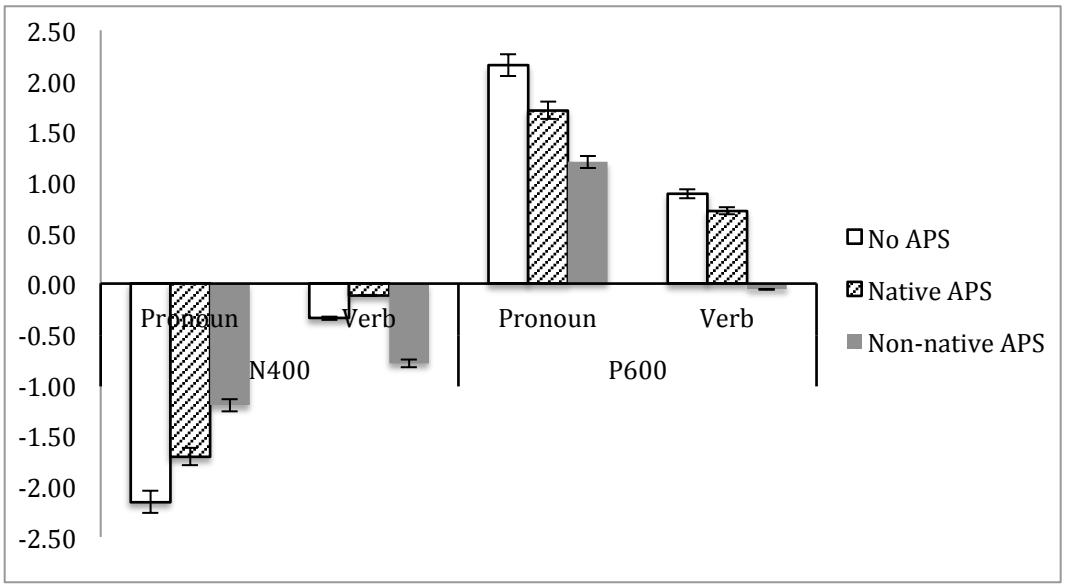


Figure 5.5. Grammaticality effect (ungrammatical minus grammatical) in N400 and P600 time windows at electrode Pz for the No-APS (Experiment 1), Native-Speaker-APS (Experiment 2), and Non-native-Speaker-APS (Experiment 2) conditions; error bars represent standard errors.

Figure 5.4 shows the grand mean waveforms for grammatical and ungrammatical

pronouns (A) and verbs (B) in the non-native-speaker-APS condition, and Tables 5.9-5.12 show the results of statistical analyses comparing them to the native-speaker-APS condition. Visual inspection of Figure 5.4 suggests that the usual grammaticality effect is present for the pronouns, though it appears to be smaller than in the no-APS condition (see Figure 1). For the verbs, however, the P600 component of the grammaticality effect seems to be entirely absent. There is still an effect of grammaticality, but it is largely restricted to the N400 time window, with the ungrammatical condition being more negative.

Just as was found for the no-APS and native-speaker-APS conditions, grammaticality effects were larger for pronouns than for verbs, as indicated by reliable interactions between grammaticality and word type in the midline analysis of the P600 window ($p < .05$) and in both the midline and lateral analyses of the N400 window (p's$<$ 0.05). Different from the comparison of no-APS and native-speaker-APS, however, there were reliable overall differences during both time windows depending on whether the imagined voice was a native or a non-native speaker. When readers imagined a non-native speaker's voice, waveforms were more negative overall, leading to a main effect of APS condition in both the midline and lateral analyses of the N400 window (p's < 0.01) and in the midline analysis of the P600 window ($p < 0.05$). The difference between APS conditions was bigger for verbs than for pronouns in the N400 time window, as indicated by a reliable interaction between APS type and word type in both midline and lateral analyses of the N400 window (ps$<$ 0.05). Most importantly, the effect of APS condition on the size of the grammaticality effect was larger for verbs than for pronouns, as indicated by reliable three-way interactions among APS type, word type, and

grammaticality in both midline and lateral analyses of the N400 time window (*p*s < 0.05) and midline analyses of the P600 time window (p < 0.05). In the one analysis in which this three-way interaction was not reliable, i.e., the lateral analysis of the P600 window, it was the four-way interaction with anteriority that was reliable instead (p < .05).

Figure 6 summarizes the grammaticality effects in all three APS conditions in each time window for each word type at the Pz electrode site. This figure shows that the pattern of grammaticality effects was somewhat different when a non-native speaker's voice was imagined. In the P600 window, imagining the non-native speaker's voice led to a decrease in the size of grammaticality effects, just as imagining the native speaker's voice did in the comparison of the no-APS and native-speaker-APS conditions (though the decrease was not reliable in that comparison). Most importantly, the grammaticality effect in the P600 window disappeared for the verbs. There was still a grammaticality effect for the verbs, but it manifested as more negativity in the N400 window instead.

Table 5.10

Native-APS vs Non-native-APS, P600 window, midline sites

Predictor	Est	SE	df	*t*	*p*
Intercept	0.05	0.12	162.7	0.46	0.64
Grammaticality (Gram)	-0.21	0.13	761.3	-1.63	0.1
Gram x Anteriority$_1$ (Ant$_1$, Fz vs Cz)	-0.06	0.17	1197	-0.37	0.71
Gram x Anteriority$_2$ (Ant$_2$, Fz vs Pz)	-0.16	0.17	1197	-1	0.32
Gram x Word Type (Wtype)	**0.36**	**0.17**	**1197**	**2.17**	**< 0.05**
Gram x Wtype x Ant$_1$	0.1	0.23	1197	0.43	0.66
Gram x Wtype x Ant$_2$	0.01	0.23	1197	0.04	0.97
Wtype	**0.44**	**0.13**	**551**	**3.29**	**< 0.01**
Wtype x Ant$_1$	-0.23	0.17	1197	-1.41	0.16
Wtype x Ant$_2$	**-0.52**	**0.17**	**1197**	**-3.12**	**< 0.01**

Table 5.10 (cont.)

Predictor	Est	SE	df	*t*	*p*
APS	**-0.28**	**0.13**	**558.6**	**-2.07**	**< 0.05**
APS x Gram	0.13	0.17	1197	0.8	0.42
APS x Wtype	0.23	0.17	1197	1.41	0.16
APS x Gram x Wtype	**-0.49**	**0.23**	**1197**	**-2.1**	**< 0.05**
APS x Ant_1	0.11	0.17	1197	0.64	0.53
APS x Ant_2	0.24	0.17	1197	1.45	0.15
APS x Gram x Ant_1	-0.25	0.23	1197	-1.07	0.28
APS x Gram x Ant_2	-0.43	0.23	1197	-1.86	0.06
APS x Wtype x Ant_1	-0.06	0.23	1197	-0.25	0.81
APS x Wtype x Ant_2	-0.27	0.23	1197	-1.16	0.25
APS x Gram x Wtype x Ant_1	0.21	0.33	1197	0.62	0.54
APS x Gram x Wtype x Ant_2	0.58	0.33	1197	1.76	< 0.1
Ant_1	-0.04	0.12	997.8	-0.37	0.71
Ant_2	0	0.12	820.1	0.02	0.99
NDRT Comprehension	-0.04	0.07	61.1	-0.59	0.56
NDRT Reading Speed	-0.02	0.07	61.1	-0.35	0.73

Table 5.11

Native-APS vs Non-native-APS, P600 window, lateral sites

Predictor	Est	SE	df	*t*	*p*
Intercept	0.32	0.12	139.7	2.78	< 0.01
Grammaticality (Gram)	**-0.27**	**0.11**	**1709**	**-2.51**	**< 0.05**
Gram x Anteriority (Ant)	**0.25**	**0.12**	**1709**	**2.07**	**< 0.05**
Gram x Laterality (Lat)	0.02	0.12	1709	0.17	0.87
Gram x Word Type (Wtype)	0.25	0.15	1709	1.66	< 0.1
Gram x Wtype x Ant	-0.27	0.17	1709	-1.58	0.11
Gram x Wtype x Lat	0.02	0.17	1709	0.12	0.91
Wtype	**0.4**	**0.12**	**660.6**	**3.31**	**< 0.001**
Wtype x Ant	**-0.49**	**0.12**	**1709**	**-4**	**< 0.001**
Wtype x Lat	-0.04	0.12	1709	-0.36	0.72
APS	-0.17	0.11	983.8	-1.5	0.13
APS x Gram	0.08	0.15	1709	0.53	0.59

Table 5.11 (cont.)

Predictor	Est	SE	df	*t*	*p*
APS x Wtype	0.26	0.15	1709	1.74	< 0.1
APS x Gram x Wtype	-0.42	0.21	1709	-1.96	0.5
APS x Ant	-0.16	0.12	1709	-1.3	0.19
APS x Lat	0.01	0.12	1709	0.1	0.92
APS x Gram x Ant	-0.33	0.17	1709	-1.89	< 0.1
APS x Gram x Lat	0.07	0.17	1709	0.38	0.7
APS x Wtype x Ant	0	0.17	1709	-0.02	0.99
APS x Wtype x Lat	0.12	0.17	1709	0.68	0.49
APS x Gram x Wtype x Ant	**0.51**	**0.25**	**1709**	**2.08**	**< 0.05**
APS x Gram x Wtype x Lat	-0.15	0.25	1709	-0.61	0.54
Ant_1	-0.12	0.1	555.3	-1.27	0.21
Ant_2	**-0.42**	**0.1**	**372.1**	**-4.05**	**< 0.001**
NDRT Comprehension	-0.02	0.05	61	-0.3	0.76
NDRT Reading Speed	-0.02	0.05	61	-0.33	0.74

Table 5.12

Native-APS vs Non-native-APS, N400 window, midline sites

Predictor	Est	SE	df	*t*	*p*
Intercept	0.02	0.11	287.4	0.14	0.89
Grammaticality (Gram)	0.05	0.14	927	0.32	0.75
Gram x $Anteriority_1$ (Ant_1, Fz vs Cz)	0.03	0.19	1197	0.18	0.86
Gram x $Anteriority_2$ (Ant_2, Fz vs Pz)	0.15	0.19	1197	0.77	0.44
Gram x Word Type (Wtype)	**0.41**	**0.19**	**1197**	**2.13**	**< 0.05**
Gram x Wtype x Ant_1	0.02	0.27	1197	0.07	0.95
Gram x Wtype x Ant_2	-0.24	0.27	1197	-0.88	0.38
Wtype	0.17	0.15	580.2	1.14	0.26
Wtype x Ant_1	-0.11	0.19	1197	-0.56	0.57
Wtype x Ant_2	-0.19	0.19	1197	-0.99	0.33
APS	**-0.38**	**0.15**	**800.1**	**-2.59**	**< 0.01**
APS x Gram	0.27	0.19	1197	1.39	0.16
APS x Wtype	**0.44**	**0.19**	**1197**	**2.27**	**< 0.05**
APS x Gram x Wtype	**-0.54**	**0.27**	**1197**	**-1.98**	**< 0.05**
APS x Ant_1	0.16	0.19	1197	0.83	0.41

Table 5.12 (cont.)

Predictor	Est	SE	df	*t*	*p*
APS x Ant_2	0.3	0.19	1197	1.58	0.11
APS x Gram x Ant_1	-0.23	0.27	1197	-0.86	0.39
APS x Gram x Ant_2	-0.36	0.27	1197	-1.32	0.19
APS x Wtype x Ant_1	-0.09	0.27	1197	-0.34	0.73
APS x Wtype x Ant_2	-0.32	0.27	1197	-1.19	0.23
APS x Gram x Wtype x Ant_1	0.19	0.38	1197	0.5	0.62
APS x Gram x Wtype x Ant_2	0.38	0.38	1197	0.99	0.32
Ant_1	-0.14	0.14	1165.4	-1.03	0.3
Ant_2	-0.24	0.14	1006.1	-1.76	< 0.1
NDRT Reading Speed	0.04	0.05	61.7	0.75	0.45
NDRT Comprehension	-0.05	0.05	61.7	-0.93	0.36

Table 5.13

Native-APS vs Non-native-APS, N400 window, lateral sites

Predictor	Est	SE	df	*t*	*p*
Intercept	0.02	0.11	287.4	0.14	0.89
Grammaticality (Gram)	0.05	0.14	927	0.32	0.75
Gram x Anteriority (Ant)	0.03	0.19	1197	0.18	0.86
Gram x Laterality (Lat)	0.15	0.19	1197	0.77	0.44
Gram x Word Type (Wtype)	**0.41**	**0.19**	**1197**	**2.13**	**< 0.05**
Gram x Wtype x Ant	0.02	0.27	1197	0.07	0.95
Gram x Wtype x Lat	-0.24	0.27	1197	-0.88	0.38
Wtype	0.17	0.15	580.2	1.14	0.26
Wtype x Ant	-0.11	0.19	1197	-0.56	0.57
Wtype x Lat	-0.19	0.19	1197	-0.99	0.33
APS	**-0.38**	**0.15**	**800.1**	**-2.59**	**< 0.01**
APS x Gram	0.27	0.19	1197	1.39	0.16
APS x Wtype	**0.44**	**0.19**	**1197**	**2.27**	**< 0.05**
APS x Gram x Wtype	**-0.54**	**0.27**	**1197**	**-1.98**	**< 0.05**
APS x Ant	0.16	0.19	1197	0.83	0.41
APS x Lat	0.3	0.19	1197	1.58	0.11
APS x Gram x Ant	-0.23	0.27	1197	-0.86	0.39
APS x Gram x Lat	-0.36	0.27	1197	-1.32	0.19

Table 5.13 (cont.)

Predictor	Est	SE	df	*t*	*p*
APS x Wtype x Ant	-0.09	0.27	1197	-0.34	0.73
APS x Wtype x Lat	-0.32	0.27	1197	-1.19	0.23
APS x Gram x Wtype x Ant	0.19	0.38	1197	0.5	0.62
APS x Gram x Wtype x Lat	0.38	0.38	1197	0.99	0.32
Ant	**-0.45**	**0.1**	**1772**	**-4.37**	**< 0.001**
Lat	0.03	0.1	1772	0.31	0.75
NDRT Comprehension	0.01	0.04	61	0.29	0.77
NDRT Reading Speed	-0.04	0.04	61	-1.02	0.31

Discussion

Two ERP experiments have compared the readers' brain responses to grammatical errors during sentence processing in three reading modes: Non-APS, APS of native speech, and APS of non-native Speech. Our hypothesis is that if auditory perceptual simulation is analogous to "listening" to one's speech, then the electrophysiological responses pattern should be similar to the results of Hanulíková et al. (2012), in which native speech led to P600 effects, while non-native speech resulted in no P600 effects. Thus, the P600 effects should be observed in both non-APS (normal silent reading) and APS of native speech conditions with no significant difference. However, when readers activate the APS of non-native speech, they will tend to "forgive" the morphosyntactic violation and there will be no P600 effects in their electrophysiological responses to the ungrammatical words.

Results demonstrated that APS of speakers' voices affected electrophysiological responses to morphosyntactic errors in silent reading. Imagining a native speaker's voice decreased the size of grammaticality effects in the both N400 and P600 window for both

verbs and pronouns compared to not imagining any voice, though that decrease was not reliable in the statistical analyses. A possible explanation is that having to imagine a voice while reading reduces the resources available for processing morphosyntactic errors, leading to smaller effects of grammaticality. Imagining a non-native speaker's voice reduced the size of the grammaticality effects even further.

Most importantly, the grammaticality effect in the P600 window disappeared when readers imagined a non-native speaker's voice. Hanulíková et al. (2012) also found that hearing similar morphosyntactic errors spoken by a non-native speaker eliminated P600 effects, compared to listening to a native speaker. Our results differed from theirs, however, in that we found a reliable grammaticality effect in the N400 window, while they did not. However, examination of their (unanalyzed) waveforms in the N400 window revealed a very similar increase in N400 amplitude as we found. Thus, both the disappearance of the P600 effect and the increase in the N400 effect in response to ungrammatical verbs replicates the pattern found in their study. (They did not include a second type of morphosyntactic error, so we cannot compare our results for the pronoun conditions to their study, but they did include sentences with semantic anomalies and found that the N400 elicited by those was the same regardless of the accent of the speaker.) It appears that imagining native and non-native voices while reading silently affects responses to morphosyntactic errors in much the same way as actually hearing such voices speak such errors, which suggests that imagining a voice during silent reading is similar in at least some ways to hearing a similar voice.

The finding of an increase in N400 rather than P600 amplitude in response to subject-verb number disagreement when imagining a non-native speaker's voice is

relevant to recent literature about tradeoffs between N400 and P600 effects. Tanner and Van Hell (2014) found that there are individual differences in response to subject-verb disagreement, with some readers tending to show an N400 effect and others tending to show a P600 effect. They argued that this difference might arise from different processing strategies, with some readers relying more on semantic information and others more on syntactic information to understand sentences. Grey & Van Hell (2017) also observed an increased N400 and reduced P600 when subjects listened to the grammatical errors read by a speaker with accented speech. However, we found different kinds of responses in the same people, depending on whose voice they imagined and what kind of error they encountered. When they encountered a pronoun error, it elicited effects in both the N400 and P600 time windows regardless of whose voice they imagined. In contrast, when they encountered an error in verb number, if they were imagining a native speaker's voice, it elicited the usual combination of effects in both the N400 and P600 windows, but when they imagined a non-native speaker's voice, the effect in the P600 window disappeared while the effect in the N400 window increased. This suggests that verb errors imagined in a non-native voice triggered a different kind of processing. It would make sense to use a processing strategy of relying more on semantic information when trying to understand non-native speakers, which might lead to N400 rather than P600 effects. Oines and Kim (2014) have demonstrated that using tasks that emphasize either meaning or form can elicit either N400 or P600 responses to the same sentences in the same subjects. It is notable here that the same shift to an N400-dominant response did not happen for the pronoun case errors when imagined in a non-native voice. Maybe those errors were so unusual that their ungrammaticality was too salient to allow

suspension of the kind of processing that yields P600 effects. One possibility that it is important to consider is that there was actually no increase in N400 amplitude itself in response to verb errors imagined in a non-native voice, but rather that it was not partially cancelled by the beginning of a partially overlapping P600. If that was the case, the absence of the P600 effect still suggests a shift in processing mode, but not necessarily that there was more reliance on semantic processing.

CHAPTER 6

GENERAL DISCUSSION AND FUTURE DIRECTIONS

Together, the four studies reported in this dissertation investigated how auditory perceptual simulation of a particular speaker's voice (native or non-native speech) affects native and L2 English speakers' sentence processing and comprehension, as well as native English speakers' electrophysiological responses to grammatical errors in sentence processing.

Study 1 showed that readers can perceptually simulate a speaker's voice during silent reading. This simulation can be cued by either the picture or voice of a given speaker following very brief examples of the speaker's voice. When readers simulated the faster-speaking native English speaker's voice, they read the online sentences faster. When they simulated the slower-speaking non-native speaker's voice, their online reading slowed down. Although readers perceived the non-native speech more negatively, they still benefited from APS of both native and non-native speech during silent reading—readers read more quickly and comprehended more correctly in the APS conditions compared to normal silent reading.

Study 2 manipulated the speech rates of the native and non-native speakers' voices to investigate whether readers' online reading speed is modulated by the accent or speech rate of the speaker, using the same paradigm from Study 1. When readers engaged in APS of the faster non-native speaker's voice, they read the sentences online faster. In contrast, APS of a slower native speaker's voice led to slower online reading speed, despite the fact that readers perceived the slower native speaker's voice more positively than the faster non-native speaker's voice. Taken together, the experiments in Study 1 and Study 2 demonstrate that speakers' speech rates, rather their accents, modulate

readers' online reading.

Study 3 extended the investigation of APS to L2 English speakers. Results revealed that L2 English speakers showed similar reading patterns as native English speakers. L2 English speakers' reading speed was modulated by the APS of native and non-native speech— L2 English speakers read faster when they activated the faster native English speaker's voice simultaneously. Interestingly, their reading speed was also influenced by their working memory: higher working memory readers tended to have faster reading speeds. Moreover, imagining the non-native speaker's voice lowered L2 English speakers' comprehension accuracy in implausible sentence conditions.

Study 4 is the first to explore readers' electrophysiological responses using ERPs during APS (Yao et al., 2011, used fMRI). The results demonstrated that APS of a native English speaker's voice yields similar brain responses to grammatical errors as normal silent reading. The smaller ERs amplitudes in APS in the native speech condition compared to the non-APS condition suggest that APS requires extra attention, such as retrieving the speaker's voice from memory, in addition to normal silent reading. APS of non-native speech changed readers' processing strategies on subject-verb agreement errors, resulting in a grammaticality effect in the N400 window, but not in the P600 window, which is similar to the findings by Hanulíková et al. (2012). Results indicated that auditory perceptual simulation of the native and non-native speech is similar to listening to the native and non-native speech, where readers are more likely to forgive grammatical errors read by a non-native English speaker.

Collectively, the results revealed that both native English speakers and L2 English speakers can activate APS of a native or non-native speech during silent reading. APS of

native and non-native speech benefit native English speakers' online reading speed and comprehension of syntactically complex sentences by generating a prosodic representation that buttresses the syntactic representation and promoting readers' to rely on this syntactic presentation to reach the correct interpretation (Breen, 2011; Fodor, 2002). These benefits may also be derived from the activation of production during APS. When readers simulated a native/non-native speaker's voice, they may bridge the connection between comprehension and production and create a "continuity in comprehension," which involves both fluent production and "listening." Because production generates prediction (dual path model, Dell & Chang, 2014; forward model, Pickering & Garrod, 2013a,b), readers' reading speed was facilitated by better prediction triggered by the generation of Auditory Perceptual Simulation.

Additionally, however, it should be noted that the APS effects on comprehension differ between native English speakers and L2 English speakers in the way that APS of different speech modulated the comprehension accuracy of implausible sentences among L2 English speakers, but not native English speakers. For L2 speakers, APS of native speech significantly improved comprehension accuracy, probably because APS of the non-native speech changed readers' processing strategies as it did in the ERP experiment 2. Rather than solving the complex syntactic structures, L2 English speakers reached a quick incorrect interpretation of the implausible sentences using the superficial semantic information.

Although APS of native or non-native speech benefits readers' reading speed and comprehension, it is not cost-free. The working memory load of retrieving the speaker's voice and the attention necessary to maintaining two tasks (silent reading & APS)

simultaneously influence readers' processing patterns and judgment. In Study 4, when APS was activated, readers' ERP amplitudes were, smaller overall than they were in the non-APS condition. Moreover, it seems that readers "forgave" the ungrammatical errors when they imagined the non-native speech, which is similar to results of "listening" to non-native speech (Hanulíková et al., 2012). This cancellation of P600 effects suggests readers shift their processing strategy during APS of non-native speech. One possibility is that subject-verb disagreement errors are grammatical mistakes that native Chinese speakers make frequently because Chinese does not have morphemes for plural. The frequency of encountering this type of mistakes in listening to Chinese- accented English is quite high in daily life. Thus, when readers "listen" to them during APS, they just ignore this type of mistakes.

The results of these studies have important pedagogical implications that promise to benefit children and adults who experience difficulty in reading comprehension. Instructional methods that incorporate active mental simulation of a fluent native speaker's voice during silent reading could improve processing and enhance comprehension. Further research can extend the exploration of APS to young readers, especially to those with reading disabilities. Another important extension of my work would be to explore the extent that auditory perceptual simulation can aid autistic children and facilitate them to perceptually simulate their own or another person's (e.g., their mother's) voice in comprehension. It will also be interesting to explore the effects of these different reading modes in another language to see if the comprehension benefits could be generalized in another language. The benefits that Auditory Perceptual Simulation confers on reading speed and comprehension may well vary as a function of

where readers lie on the Autism Spectrum Disorders (ASD) spectrum. The very ability to perform Auditory Perceptual Simulation at all could conceivably be a marker for an individual's relative position on the spectrum. One might imagine that simulating a given speaker's prosody or affect would be challenging for readers along the spectrum (e.g., “What is that?” vs. “What is THAT?”). Eye-tracking could also be used to monitor autistic children’s attention using instances of when they hear different sounds (familiar vs. unfamiliar) compared to when they imagine these sounds (i.e., auditory imagery).

REFERENCES

Alexander, Alexander, J. D. & Nygaard, L. C. (2008). Reading voices and hearing text: Talker specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 446–459.

Ashby, J., & Clifton Jr., C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, *9*(3), 89-100.

Ashby, J., Sanders, L. D., & Kingston, J. (2009). Skilled readers begin processing sub-phonemic features by 80ms during visual word recognition: Evidence from ERPs. *Biological psychology*, *80*(1), 84-94.

Baddeley, A., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *The Quarterly Journal of Experimental Psychology*, *33*(4), 439-454.

Bergen, B. & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain & Language, 112*, 150–158.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 6*8(3), 255-278.

Brennan, E. M., & Brennan, J. S. (1981). Accent scaling and language attitudes: Reactions to Mexican American English speech. *Language and Speech*, *24*(3), 207-221.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137-153.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some

attractions of verb agreement. *Cognitive Psychology*, *43*(2), 83-128.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660.

Barsalou, L. & Hale, C. (1993). Components of conceptual representation. From feature lists to recursive frames. In I. Van Mechelen, J. Hampton, R. Michalski & P. Theuns (eds.), *Categories and Concepts: Theoretical Views and Inductive Data Analysis* (pp. 97–144). Academic Press.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. *Reanalysis in Sentence Processing Studies in Theoretical Psycholinguistics*, 21, 1-46.

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, *8*(2), 37-50.

Breen, M. & Clifton Jr., C. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of memory and language*, *64*(2), 153-170.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and cognitive processes*, *25*(7-9), 1044-1098.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, U.K.: Cambridge University Press.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, *56*(1), 65-85.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, *40*(2), 141-201.

Carlson, K., Clifton Jr., C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *45*(1), 58-81.

Christianson, K. (2008). Sensitivity to changes in garden path sentences. *Journal of Psycholinguistic Research*, 37, 391-403.

Christianson, K. & Luke, S. G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, *15*(2), 136-166.

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *36*, 538- 544.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42,* 368–407.

Christianson, K., Williams, C.C., Zacks, R.T., & Ferreira, F. (2006). Misinterpretations of garden-path sentences by older and younger adults. *Discourse Processes, 42*, 205-238.

Chen, L., Shu, H. U. A., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(02), 161-174.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes*, *13*(1), 21-58.

Callen, V., Callois, C., & Forbes, P. (1983). Evaluation to accented English. *Journal of Cross-Cultural Psychology, 14,* 407-426.

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394.

Drumm, A. M. & Klin, C. M. (2011). When story characters communicate: readers' representations of characters' linguistic exchanges. *Memory & cognition*, *39*(7), 1348-1357.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

De Villiers, J. G., Tager, H. B., Flusberg, T., Hakuta, K., & Cohenl, M. (1979). Children's comprehension of relative clauses. *Journal of psycholinguistic Research, 8*, 499-518.

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

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870-1887.

Edfelt, A. W. (1960). *Silent Speech and Silent Reading*. Chicago: University of Chicago Press.

Edwards, J. R. (1977). Students' reaction to Irish regional accent. *Language & Speech, 20,* 280-286.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47,* 164–203.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291-325.

Ford, M. (1983). A method for obtaining measures of local parsing complexity Throughout sentences. *Journal of verbal learning and verbal behavior*, *22*(2), 203-218.

Fodor, J. D. (2002, April). *Psycholinguistics cannot escape prosody*. Paper presented at the meeting of Speech Prosody, Aix-en-Provence, France.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica*, *128*(3), 528-536.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, *6*(2), 78-84.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183-192.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition*, *27*(3), 438-453.

Filik R, & Barber E (2011) Inner Speech during Silent Reading Reflects the Reader's Regional Accent. *PLoS ONE*, 6(10): e25782. doi:10.1371/journal.pone.0025782

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

Ferreira, F. & Patson, N. D. (2007). The 'good-enough' approach to language comprehension. *Language and Linguistics Compass, 1,* 71–83.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, *54*(4), 541-553.

Gluszek, A., & Hansen, K. (2013). Language attitudes in the Americas. *The social meanings of language, dialect and accent. International perspectives on speech styles*, 26-44.

Glenberg, A. M. & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9,* 558–565.

Gabriele, A., Fiorentino, R., & Bañón, J. A. (2013). Examining second language development using event-related potentials: a cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism*, *3*(2), 213-232.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, *29*(2), 261-290.

Gunraj, D. N., & Klin, C. M. (2012). Hearing story characters' voices: auditory imagery during reading. *Discourse Processes*, *49*(2), 137-153.

Gibson, E. A. F. (1998). Linguistic complexity: locality and syntactic dependencies. *Cognition, 68,* 1–76.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy channel account of crosslinguistic word-order variation. *Psychological Science*, *24*(7), 1079-1088.

Giles, H. (1972). Evaluation of personality content from accented speech as a function of listeners' social attitudes. *Perceptual and Motor Skills, 34,* 168-170. doi: 10.2466/pms.1972.34.1.168

Giles, H., Hewstone, M., Ryan, E. B., & Johnson, P. (1987). Research on language attitudes. *Sociolinguistics: An international handbook of the science of language and society*, *1*, 585-597.

Giles, H., & Watson, B. (2013). The social meanings of language, dialect and accent. *New York, NY: Peter Lang*.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51,* 97-114.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 1304-1321.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13,* 425-430.

Gennari, S. P. & MacDonald, M. C. (2008). Semantic indeterminacy and relative clause comprehension. *Journal of Memory and Language, 58,* 161–187.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language learning*, *34*(1), 65-87.

Glenberg, A. M. & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language, 43*, 379–401.

Gass, S. & Varonis, E. (1984). The effect of familiarity on comprehension of the non-native speech. *Language and Speech. 34,* 65-89.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and language*, *122*(2), 92-102.

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93-108.

Huey, E. B. (1968). *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press. (Original work published 1908)

Hubbard, T. L. (2010). Auditory Imagery: Empirical Findings. *Psychological Bulletin*, *136*, 302–329.

Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. Brown & P. Hagoort (Eds), *Neurocognition of Language* (pp. 273–316). Oxford, UK: Oxford University Press.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531-1549.

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, *16*(1), 38-50.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension.

Cognitive Neuroscience, Journal of, *18*(4), 665-679.

Hale, A. D., Hawkins, R. O., Sheeley, W., Reynolds, J. R., Jenkins, S., Schmitt, A. J., & Martin, D. A. (2011). An investigation of silent versus aloud reading comprehension of elementary students using maze assessment procedures. *Psychology in the Schools*, *48*(1), 4-13.

Hanulíková, A., van Alphen, P.M., van Goch, M.M., & Weber, A. (2012). When one person's mistake in another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*, 878–887.

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of cognitive neuroscience*, *24*(6), 1382-1397.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59,* 434–446.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205-225.

Kaan, E. (2007). Event‐related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571-591.

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and language*, *100*(3), 257-261.

Kolk, H. H., Chwilla, D. J., Van Herten, M., & Oor, P. J. (2003). Structure and limited

capacity in verbal working memory: A study with event-related potentials. *Brain and language*, *85*(1), 1-36.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*(5), 539-550.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. A. Gernsbacher & M. Traxler (Eds.), *Handbook of Psycholinguistics* (pp. 659–724). New York, NY: Elsevier.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition*, 112 457–461.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social cognition*, *27*(4), 623.

Kim, J. H. & Christianson, K. (2013). Sentence complexity and working memory effects in ambiguity resolution. *Journal of Psycholinguistics Research, 42*, 393-411.

Kluender, R., & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*(2), 196-214.

Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language*

and cognitive processes, *8*(4), 573-633.

Kosslyn, S. M. & Matt, M. C. (1977). If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. *Bulletin of the Psychonomic Society, 9,* 250–252.

Levy, B. A. (1975). Vocalization and suppression effects in sentence memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(3), 304-316.

Lim, J. H. & Christianson, K. (2013a). Integrating meaning and structure in L1–L2 and L2–L1 translations. *Second Language Research, 29*, 233-256.

Lim, J-H. & Christianson, K. (2013b). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, *16*, 518-537.

Louwerse, M. M. & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science 35,* 381–398.

Louwerse, M. M. & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition 114*, 96–104.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086-21090.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, *8*, 213.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093-1096.

Levine, W. H., & Klin, C. M. (2001). Tracking of spatial information in narratives. *Memory & Cognition*, *29*, 327-335.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*,*45*(1), 73-97.

Munro, M. J., & Derwing, T. M. (1998). The Effects of Speaking Rate on Listener Evaluations of Native and Foreign-Accented Speech. *Language Learning*, *48*(2), 159-182.

MacWhinney, B. (1982). Basic syntactic processes. *Language acquisition*, *1*, 73-136.

MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, *29*(2), 95-141.

Mecklinger, A., Schriefers, H., Steinhauder, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: an analysis with event-related potentials. *Memory & Cognition, 23,* 477-494.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *cortex*, *47*(8), 908-930.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, *36*(4), 569-587.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain

potentials. *Journal of cognitive Neuroscience*, *3*(2), 151-165.

Oines, L. & Kim, A. (2014). *Integrate or repair? ERP responses to semantic anomalies depend on choice of processing strategy.* Presented at the Architectures and Mechanisms for Language Processing Conference, Edinburgh, Scotland.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739-773.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, *14*(3), 283-317.

Pecher, D. & Zwaan, R. A. (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. New York: Cambridge University Press.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, *14*, 119 -124.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation *Psychonomic Bulletin & Review, 11,* 164-167.

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,

35(1), 280.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. E., Jr. (2012). *The psychology of reading. New York: Psychology Press. ISBN: 978-1-84872-975-9.*

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of memory and language*, *57*(3), 348-379.

Pickering, M. J., & Garrod, S. (2013a). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347.

Pickering, M. J., & Garrod, S. (2013b). Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, *36*(04), 377-392.

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in human neuroscience*, *9*, 167.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition, 36*(1), 201 216.

Solomon, K. O. & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition 2004, 32,* 244-259.

Stites, M., Luke, S. G., & Christianson, K. (2013) The psychologist said quickly, "Dialogue descriptions modulate reading speed!" *Memory and Cognition, 41*, 137-151.

Slowiaczek, M. L. & Clifton Jr., C. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning and Verbal Behavior*, *19*(5), 573-582.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437-442.

Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, *93*(3), B97-B108.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, *25*(2), 249-271.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104-120.

Townsend, D. J. & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules.* Cambridge, MA: MIT Press.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on "Grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex*, *66*, 149-155.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289-301.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: evidence from eye movements. *Journal of Memory and Language, 47*, 69-90.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to

violations in second language grammar: An event-related potential investigation. *Studies in second language acquisition*, *27*(2), 173.

Varonis, E. M., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in second language acquisition*, *4*(02), 114-136.

Van Dyke, J. A. & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory & Language, 55,* 157-166.

Vornik, L., Sharman, S., & Garry, M. (2003). The power of the spoken word: Sociolinguistics cues influence the misinformation effect. *Memory, 11,* 101-109.

van Gompel, R. P., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, *55*(3), 335-362.

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

Woumans, E., Martin, C. D., Bulcke, C. V., Van Assche, E., Costa, A., Hartsuiker, R. J., & Duyck, W. (2015). Can faces prime a language?.*Psychological science*, *26*(9), 1343-1352.

Weckerly, J. & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology, 36,* 559–570.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive Psychol*ogy, *58,* 250–271.

Yao, B. & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition, 121,* 447–453.

Zhou, P., & Christianson, K. (2014, September). *Mechanisms underlying auditory perceptual simulation: reading speed or speech biases?* Poster presented at Architectures and Mechanisms for Language Processing. Edinburg, UK.

Zhou, P. & Christianson, K. (2016a). I "hear" what you're "saying". Auditory perceptual simulation, reading speed, and comprehension. *Quarterly Journal of Experimental Psychology, 69, 972-995.*

Zhou, P., & Christianson, K. (2016b). Auditory perceptual simulation: Simulating speech rates or accents?. *Acta psychologica*, *168*, 85-90.

Zhou, P. & Christianson, K. (in preparation). Comparison of attitudes towards native and non-native English speech.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 44, pp. 35–62). San Diego, CA: Academic.

Zwaan, R. A. (2008). Experiential traces and mental simulations in language comprehension. In M. DeVega, A. M. Glenberg, & A.C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 165–180). Oxford, UK: Oxford University Press.

Zwaan, R. A. (2009). Mental simulation in language comprehension and social cognition. *European Journal of Social Psychology, 7*, 1142–1150.

Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *Journal of*

Cognitive Neuroscience, *8*(1), 29-46.

APPENDIX A.

EXPERIMENTAL STIMULI USED IN EYE-TRACKING STUDIES

Plausible and subject-relative clause sentences:

1. The police officer that arrested the citizen was handsome.
2. The chef that ruined the food was in the kitchen.
3. The cop that pursued the thief was driving a car.
4. The cat that chased the mouse was fast.
5. The ghost that scared the boy was hiding behind a curtain.
6. The bird that ate the worm was small.
7. The hunter that shot the deer was in the Rocky Mountains.
8. The lawyer that sued the doctor was smart.
9. The coach that scolded the player won the championship twice.
10. The owner that fed the cat was sitting on a sofa.
11. The man that walked the dog was in the park.
12. The father that scolded the teenager was in the living room.
13. The guide that led the tourist liked Europe a lot.
14. The waiter that served the guest was tall.
15. The soldier that protected the villager was brave.
16. The fly that ate the frog was green.
17. The doctor that treated the patient was female.
18. The teacher that quizzed the student was in the classroom.
19. The angler that caught the fish was in the middle of the ocean.
20. The detective that investigated the suspect was very tired.
21. The dog that bit the man was in the yard.

22. The golfer that hit the ball was in the shade.

23. The politician that the voter deceived was Korean.

24. The mother that bathed the child smelled nice.

25. The bird that protected the chick was in the big tree.

26. The reporter that interviewed the actress was at the coffee shop.

27. The fan that admired the Hollywood star loved to wear big sunglasses.

28. The dog that herded the sheep was very furry.

29. The grandmother that dressed the child had a beautiful smile.

30. The tutor that taught the student solved the math problem.

31. The consultant that advised the client was very clever.

32. The homeowner that the gardener paid loved the garden.

33. The kids that obeyed the teacher enjoyed the summer break.

34. The criminal that kidnapped the girl was on CNN news.

35. The boss that fired the worker was unhappy.

36. The volunteer that helped the blind was very handsome.

37. The fan that cheered for the baseball player wore a red shirt.

38. The parent that raised the twins lived in New York.

39. The parents that punished the child went to church.

40. The dean that awarded the student a prize was very famous in the school.

41. The nuns that took care of the orphans live in a small village.

42. The lawyer that defended the client was worried about the result.

43. The secretary that assisted the President was in a conference room.

44. The boy that petted the puppy had cute eyes.

45. The terrorist that held the hostage was located in the building.

46. The wrecker that towed the car drove at a high speed.

47. The conductor that led the orchestra was pleased with the performance.

48. The guard that locked up the prisoner regretted past decisions.

APPENDIX B

SOCIAL ATTRACTIVENESS SURVEY

1. Subject ID_____
2. Your gender_____
3. Your age _____
4. Comprehensibility of the native English speaker (1=completely INCOMPHREHENSIBLE, 7=completely comprehensible)

   1 2 3 4 5 6 7
5. Accent of the native English speaker (1=no foreign accent at all, 7=very strong foreign accent)

   1 2 3 4 5 6 7
6. Intelligence of the native English speaker (1=very unintelligent, 7=very intelligent)

   1 2 3 4 5 6 7
7. Reliability of the native English speaker (1=very unreliable, 7=very reliable)

   1 2 3 4 5 6 7
8. Pleasantness of the native English speaker (1=very unpleasant to listen to, 7=very pleasant to listen to)

   1 2 3 4 5 6 7
9. Confidence of the native English speaker (1=not confident at all, 7=very confident

   1 2 3 4 5 6 7
10. How sincere is the native English speaker? (1=not sincere at all, 7=very sincere)

1 2 3 4 5 6 7

11. How conscientious is the native English speaker? (1=Not conscientious at all,7=Very conscientious)

    1 2 3 4 5 6 7

12. How honest is the speaker? (1=Not honest at all, 7=Very honest)

    1 2 3 4 5 6 7

13. How prestigious is the native English speaker? (1=Not prestigious at all, 7=Very prestigious)

    1 2 3 4 5 6 7

14. How educated is the native English speaker? (1= not well educated, 7=well educated)

    1 2 3 4 5 6 7

15. How dependable is the native English speaker? (1=Not dependable at all, 7=Very dependable)

    1 2 3 4 5 6 7

16. How likable is the native English speaker? (1=Not likable at all, 7=Very likable)

    1 2 3 4 5 6

APPENDIX C

EXPERIMENTAL STIMULI USED IN ERP STUDY

1	His fellow classmates dislike him since they are jealous.	Grammatical
	His fellow classmates dislike he since they are jealous.	Pronoun case mismatch
	His fellow classmates dislikes him since they are jealous.	Subject-verb disagreement
2	The lab assistant orders beakers for her lab every month.	Grammatical
	The lab assistant orders beakers for she lab every month.	Pronoun case mismatch
	The lab assistant order beakers for her lab every month.	Subject-verb disagreement
3	Jimmy's parents worry because he plays with dolls.	Grammatical
	Jimmy's parents worry because him plays with dolls.	Pronoun case mismatch
	Jimmy's parents worries because he plays with dolls.	Subject-verb disagreement
4	The kidnapper apparently lures them into his van.	Grammatical
	The kidnapper apparently lures they into his van.	Pronoun case mismatch
	The kidnapper apparently lure them into his van.	Subject-verb disagreement
5	The pirate attacks the merchants and seizes their ship.	Grammatical
	The pirate attacks the merchants and seizes they ship.	Pronoun case mismatch
	The pirate attacks the merchants and seize their ship.	Subject-verb disagreement
6	Birds migrate to south with their family when winter approaches.	Grammatical
	Birds migrate to south with they family when winter approaches.	Pronoun case mismatch
	Birds migrates to south with their family when winter approaches.	Subject-verb disagreement
7	I practice my trombone every day.	Grammatical
	I practice me trombone every day.	Pronoun case mismatch
	I practices my trombone every day.	Subject-verb disagreement
8	Her mechanic always treats her with respect.	Grammatical

	Her mechanic always treat she with respect.	Pronoun case mismatch
	Her mechanic always treat her with respect.	Subject-verb disagreement
9	The founder is indicted for his insider trading practices.	Grammatical
	The founder is indicted for him insider trading practices.	Pronoun case mismatch
	The founder is indict for his insider trading practices.	Subject-verb disagreement
10	Vince feels nervous whenever he punts the football.	Grammatical
	Vince feels nervous whenever him punts the football.	Pronoun case mismatch
	Vince feels nervous whenever he punt the football.	Subject-verb disagreement
11	Lily prints documents for her colleagues everyday.	Grammatical
	Lily prints documents for she colleagues everyday.	Pronoun case mismatch
	Lily print documents for her colleagues everyday.	Subject-verb disagreement
12	Their strange uncle frightens them with ghost stories.	Grammatical
	Their strange uncle frightened they with ghost stories.	Pronoun case mismatch
	Their strange uncle frighten them with ghost stories.	Subject-verb disagreement
13	The kids go fishing in their parents' boat.	Grammatical
	The kids go fishing in they parents' boat.	Pronoun case mismatch
	The kids goes fishing in their parents' boat.	Subject-verb disagreement
14	Alex cleans his father‘s car for some pocket money.	Grammatical
	Alex cleans he father‘s car for some pocket money.	Pronoun case mismatch
	Alex clean his father‘s car for some pocket money.	Subject-verb disagreement
15	The accountant submits the financial report to her supervisor.	Grammatical
	The accountant submits the financial report to she supervisor.	Pronoun case mismatch
	The accountant submit the financial report to her supervisor.	Subject-verb disagreement
16	My boyfriend forgets how we met each other.	Grammatical
	My boyfriend forgets how us met each other.	Pronoun case mismatch
	My boyfriend forget how we met each other.	Subject-verb disagreement

17	We all really rely on her for support on this project.	Grammatical
	We all really rely on she for support on this project.	Pronoun case mismatch
	We all really rellies on her for support on this project.	Subject-verb disagreement
18	Actors spend hours memorizing their lines.	Grammatical
	Actors spend hours memorizing they lines.	Pronoun case mismatch
	Actors spends hours memorizing their lines.	Subject-verb disagreement
19	Our rich neighbour brings us paintings from Europe.	Grammatical
	Our rich neighbour brings we paintings from Europe.	Pronoun case mismatch
	Our rich neighbour bring us paintings from Europe.	Subject-verb disagreement
20	My pet bites my sofa when I am out.	Grammatical
	My pet bites my sofa when me am out.	Pronoun case mismatch
	My pet bite my sofa when I am out.	Subject-verb disagreement
21	On weekend Ike parks his car in back.	Grammatical
	On weekend Ike parks he car in back.	Pronoun case mismatch
	On weekend Ike park his car in back.	Subject-verb disagreement
22	The writer revises his work several times before publishing.	Grammatical
	The writer revises him work several times before publishing.	Pronoun case mismatch
	The writer revise his work several times before publishing.	Subject-verb disagreement
23	Kay and Jimmy are good friends since they share common interests.	Grammatical
	Kay and Jimmy are good friends since them share common interests.	Pronoun case mismatch
	Kay and Jimmy are good friends since they shares common interests.	Subject-verb disagreement
24	They need to drink water when they feel thirsty.	Grammatical
	They need to drink water when them feel thirsty.	Pronoun case mismatch
	They needs to drink water when they feel thirsty.	Subject-verb disagreement
25	Posters of musicians cover her bedroom wall.	Grammatical
	Posters of musicians cover she bedroom wall.	Pronoun case mismatch
	Posters of musicians covers her bedroom wall.	Subject-verb disagreement
26	Alex turns off the light when I replace the bulb.	Grammatical
	Alex turns off the light when me replace the	Pronoun case mismatch

	bulb.	
	Alex turn off the light when I replace the bulb.	Subject-verb disagreement
27	Gideon has a problem feeding his family during the famine.	Grammatical
	Gideon has a problem feeding he family during the famine.	Pronoun case mismatch
	Gideon have a problem feeding his family during the famine.	Subject-verb disagreement
28	The taxi driver stops the car when he notices a stop sign.	Grammatical
	The taxi driver stops the car when him notices a stop sign.	Pronoun case mismatch
	The taxi driver stop the car when he notices a stop sign.	Subject-verb disagreement
29	Ray stays late to finish his math homework.	Grammatical
	Ray stays late to finish he math homework.	Pronoun case mismatch
	Ray stay late to finish his math homework.	Subject-verb disagreement
30	Few colleges truly prepare students for their future careers.	Grammatical
	Few colleges truly prepare students for they future careers.	Pronoun case mismatch
	Few colleges truly prepares students for their future careers.	Subject-verb disagreement
31	Larry receives a letter that his fiance sends from Paris monthly.	Grammatical
	Larry receives a letter that him fiance sends from Paris monthly.	Pronoun case mismatch
	Larry receive a letter that his fiance sends from Paris monthly.	Subject-verb disagreement
32	The babysitter ends her work at 7pm on the weekend.	Grammatical
	The babysitter ends she work at 7pm on the weekend.	Pronoun case mismatch
	The babysitter end her work at 7pm on the weekend.	Subject-verb disagreement
33	He calls the dentist when his teeth hurt	Grammatical
	He calls the dentist when him teeth hurt.	Pronoun case mismatch
	He call the dentist when his teeth hurt.	Subject-verb disagreement
34	The ugly guy protects us from the rain.	Grammatical
	The ugly guy protects we from the rain.	Pronoun case mismatch
	The ugly guy protect us from the rain.	Subject-verb disagreement
35	I learn a lot about her at the summer camp.	Grammatical
	I learn a lot about she at the summer camp.	Pronoun case mismatch

	I learns a lot about her at the summer camp.	Subject-verb disagreement
36	When a student comes to class he brings his textbook.	Grammatical
	When a student comes to class he brings he textbook.	Pronoun case mismatch
	When a student come to class he brings his textbook.	Subject-verb disagreement
37	Jim plants some roses on his roof to get the sunshine.	Grammatical
	Jim plants some roses on he roof to get the sunshine.	Pronoun case mismatch
	Jim plant some roses on his roof to get the sunshine.	Subject-verb disagreement
38	The witch curses the prince and turns him into a donkey.	Grammatical
	The witch curses the prince and turns he into a donkey.	Pronoun case mismatch
	The witch curse the prince and turns him into a donkey.	Subject-verb disagreement
39	He botches up his errand thoroughly.	Grammatical
	He botches up him errand thoroughly.	Pronoun case mismatch
	He botch up his errand thoroughly.	Subject-verb disagreement
40	Jim's sister never offers to shake my hand.	Grammatical
	Jim's sister never offers to shake I hand.	Pronoun case mismatch
	Jim's sister never offer to shake my hand.	Subject-verb disagreement