Is imagining a voice like listening to it? Evidence from ERPs

Peiyun Zhou\textsuperscript{a,c,}\textsuperscript{*}, Susan Garner\textsuperscript{b,c}, Kiel Christianson\textsuperscript{a,c}

\textsuperscript{a} Department of Educational Psychology, Urbana, IL, USA
\textsuperscript{b} Department of Psychology, Urbana, IL, USA
\textsuperscript{c} Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

\begin{abstract}
Readers who have seen the Harry Potter movies before reading the novels may “hear” actors’ voices in their heads when they later read the books. This phenomenon of mentally simulating the voice of speakers depicted in texts has been referred to as auditory perceptual simulation (APS). How much is this mental simulation of voices like listening to actual voices? Two event-related potential (ERP) experiments examined the auditory perceptual simulation of native and non-native English speech while participants silently read English sentences containing subject-verb agreement errors or pronoun-case errors. The aim was to compare readers’ ERPs when imagining native and non-native speech to the results of Hanuliková, van Alphen, van Goch, and Weber (2012), who recorded ERPs while participants listened to native and non-native speech and found that native-speaking listeners “forgive” errors (signaled by reduced P600 effects) by non-native speakers. Our participants listened to samples of a native and a non-native English speaker’s speech and were then asked to imagine the voice of either one or the other speaker while reading sentences. Results revealed differences in N400 and P600 waveforms when imagining the non-native speaker’s voice compared to the native speaker’s voice. Importantly, when imagining the non-native speaker committing subject-verb agreement errors, P600 amplitudes were no different from error-free items.
\end{abstract}

1. Introduction

1.1. 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 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 might therefore commit errors when they are producing language, such as failures in subject-verb agreement (e.g., \textit{He walk home every day}) or gender agreement (e.g., Mary saw himself in the mirror). Such errors are therefore also encountered by listeners during language comprehension. The effects that such errors have on a language comprehender may vary, however, depending on who commits the error.

The event-related potential (ERP) method has been used to investigate L1 speakers’ (see reviews in Hagoort, Brown, & Osterhout, 1999; Kutas, Van Petten, & Kluger, 2006) and L2 learners’ (e.g., Gabrielle, Fiorentino, & Baňoš, 2013; Tokowicz & MacWhinney, 2005) sentence processing because it is a sensitive measure of the brain’s electrophysiological responses to language-related events. Two major ERP components, the N400 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 ms (ms) after the onset of words containing lexical/semantic difficulties or anomalies in written, spoken, or signed languages (Friederici, Steinhaus, & Frisch, 1999; Kutas & Hillyard, 1980; see Kutas & Federmeier, 2011, for a review). Researchers have related N400 effects to lexical access or semantic memory (Kutas & Federmeier, 2000; Kutas & Van Petten, 1994), and the ease of semantic integration (Brown & Hagoort, 2000). The amplitude of the N400 can be modulated by the frequency and concreteness of single words, as well as contextual constraints, semantic congruity, and expectancy at the sentence level. Morphosyntactic errors, such as gender or number disagreement, also have been found to affect the amplitude of the N400 (Barber &

\* Corresponding author at: Department of Psychology, Institute for Brain and Learning Sciences, Guthrie Hall (GTH), University of Washington, Seattle, WA 98105, USA.

E-mail address: peiyun@uw.edu (P. Zhou).

https://doi.org/10.1016/j.cognition.2018.10.014

Received 5 September 2017; Received in revised form 3 October 2018; Accepted 16 October 2018
Available online 23 October 2018

0010-0277/ © 2018 Elsevier B.V. All rights reserved.
Carreiras, 2003, 2005). For an example, Spanish sentences that contain gender and number disagreements elicited bigger N400 effects compared to grammatical sentences (Barber & Carreiras, 2003).

The P600 is an increased electrophysiological response in positivity peaking at around 600 ms after the onset of words containing morphosyntactic errors (Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995; Osterhout & Nicol, 1999). Different types of violations modulate the type, latency, and size of the ERP components (Barber & Carreiras, 2005; Molinaro, Barber, & Carreiras, 2011; Steinhauer, Connolly, Stemmer, & Whalnke, 2008). Coulson, King, and Kutus (1998) found larger 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 generated the larger P600 effects.

P600 responses are observed for errors in not only in visual, but also in auditory language input (Hagoort & Brown, 2000). Studies have used ERPs to examine how listeners respond to accented speech (Goslin, Duffy, & Flocia, 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 Helle, 2017; Hanuliková et al., 2012). Of particular interest to the current study, Hanuliková 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 nativeness of the speaker: there was a robust P600 effect in response to the native speaker but not the non-native speaker. The authors interpreted this result as showing that people are more likely to “forgive” grammatical errors in non-native speech. Likewise, Grey and Van Helle (2017) observed a reduced sustained frontal negativity (Nref) effects in listeners’ ERPs to non-native speech with pronoun errors, but not to non-native speech with semantic anomalies. The Nref, like the P600, is associated with certain grammatical errors (Nieuwland, 2014; Nieuwland & Van Berkum, 2006; Srebr, Rösler, & Henninghausen, 1999). They also observed individual differences in responses to native and non-native speech: whether or not listeners succeeded in identifying the non-native accent as an Asian accent influenced their electrophysiological responses to grammatical errors.

1.2. Auditory perceptual simulation

Fans of the Harry Potter films might “hear” the actor Daniel Radcliffe’s voice when subsequently 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) (Alexander & Nygaard, 2008; Hubbard, 2010; Stites, Luke, & Christianson, 2013; Zhou, 2017). Although “inner speech” is widely accepted to be present to some degree in normal, skilled, silent reading (cf. Rayner, Pollatek, Ashby, & Clifton, 2012), APS seems to be an elaborated, prosodically rich version of inner speech, which can be triggered by implicitly or explicitly cueing readers to consciously imagine the voice of a character in a text or of someone saying the text as it is being read (Zhou & Christianson, 2016a, 2016b, Submitted).

Previous behavioral studies have demonstrated that APS can affect silent reading beyond what is observed in normal silent reading by manipulating the characteristics 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; Gunnar & Klin, 2012; Kosslyn & Matt, 1977; Levine & Klin, 2001). Readers read sentences faster when they are told the text was written by a fast speaker than when they are told the text was written by a slow speaker, especially when texts are difficult (Alexander & Nygaard, 2008). Familiarity with a speaker also affects the recognition of words encountered while engaging in APS. Readers have better recall of content when they imagine a familiar speaker was reading the content to them compared to when they imagine an unfamiliar speaker was reading (Kurbay, Magliano, & Rapp, 2009). Previous neuroimaging studies reveal that imagining sounds activates similar regions of the brain as actually listening to those sounds (Perrone-Bertolotti et al., 2012; Zatorre, Halpern, Perry, Meyer, & Evans, 1996; see Perrone-Bertolotti, Rapin, Luchau, Baciu, & LOvenberg, 2014, for a review). For example, similar regions are activated when listening to and imagining music (i.e., superior temporal gyrus (STG) supramarginal gyrus, the middle frontal (MID F), and inferior frontal (Ant Inf F) cortices (Zatorre et al., 1996). Similarly, Yao, Belin, and Scheepers (2011) 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 have revealed that the speed of reading sentences containing direct quotes is modulated by descriptive text preceding the quoted material (Yao & Scheepers, 2011), or even single speech-rate-related adverbs preceding the quote, as illustrated in (1) below (from Stites et al., 2013). 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), “I finally found my car keys.”

Zhou and Christianson (2016a, 2016b) further examined how APS of different speakers’ voices (faster vs. slower speakers) during silent reading affects 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 “nativness” of the speaker’s accent (Zhou & Christianson, 2016b). In addition, APS of either a native or non-native English speaker’s voice also resulted in 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 affects 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 Hanuliková et al.’s (2012) study motivate a straightforward prediction. Recall that Hanuliková and colleagues observed that grammatical gender agreement errors produced verbally by a non-native Dutch speaker did not trigger inflated P600 signals in the ERP waveforms of a native Dutch speaker. Based on this result, Hanuliková 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, in the present study, we measured electrophysiological responses to grammatical errors in sentences that were read silently either with or without imagining a voice to test how similar APS of a voice is to actually hearing that voice.
1.3. 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 measure 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 1), to serve as a baseline control 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.

Table 1

<table>
<thead>
<tr>
<th>Design of stimulus materials</th>
<th>Critical verbs and pronouns are underlined.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatically correct</td>
<td>The carpenters chat when they sand the wood.</td>
</tr>
<tr>
<td>Subject-verb mismatch</td>
<td>The carpenters chas when they sand the wood.</td>
</tr>
<tr>
<td>Pronoun case mismatch</td>
<td>The carpenters chatting when they sand the wood.</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>The carpenters chatted and worked. (True/False)</td>
</tr>
</tbody>
</table>

2. 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.

2.1. Method

2.1.1. 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 years old), right-handed, and had normal hearing, normal or corrected-to-normal vision, and no history of neurological or psychiatric disorders.

2.1.2. 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 a condition 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 Chinese English L2 speakers make subject-verb agreement errors than pronoun case errors. For example, in a large-scale corpus analysis of Chinese L2 English learners’ writing, Chiang and Nesi (2006; see also Dattus & Ching, 2009) found 125 instances of subject-verb disagreement and 0 instances of pronoun case mismatch (out of 4493 total grammatical errors). 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 1 below (see all materials in Appendix A). 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 (Zhou, Toscano, Garms, & Christianson, 2017) investigating how readers process relative clause sentences varying in semantic plausibility (e.g., The bird that are 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 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 order was pseudo-randomized 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.

Nelson-Denny reading task

Additionally, participants’ reading speed and comprehension were measured using the comprehension portion of the Nelson-Denny Reading Task (NDRT), which is a standardized paper-and-pencil test for college readers (Brown, Fishco, & Hanna, 1993). The comprehension test contains seven reading passages with 38 comprehension probes and lasted 20 min. Each participant’s comprehension score was calculated as the number of questions answered correctly out of the 38 questions asked. Raw scores can be transformed into percentile ranks based on reading scores for second-year college students.

Bucknell auditory imagery scale

The Bucknell Auditory Imagery Scale (BAIS) (Herholz, Halpern, & Zatorre, 2012) measures individual differences for both the vividness and control of people’s auditory imagery. Each component (Vividness and Control) contains fourteen 1–7 Likert-scale questions.

2.1.3. 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.

2.1.4. Procedure

Each experimental session lasted 2.5–3.0 h. Participants were randomly assigned to one of the three stimulus lists. They were instructed

1 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.

2 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.
to sit in front of a computer 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 a 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 software package.

After the EEG-recording session, participants completed a survey asking about their 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 auditory imagination was measured using the Bucknell Auditory Imagery Scale (BAIS) (Herholz et al., 2012).

2.1.5. 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 1000 ms post-onset. 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 five dropped participants, leaving 43 whose data were included in the analyses. Approximately 11% of the trials from the remaining 43 participants were excluded from the final analyses due to blinks 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 (Gouveia, Phillips, Kazanina, & Poeppel, 2010). 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 F57), right-anterior (AF4 F4 FC4 F8 F78), left-posterior (CP3 P3 T5 P7 T7), right-posterior (CP4 P4 T6 P8 T8). Using the R software package with the lim4() function of the lim4R package (Bates, Maechler, & Bolker, 2012), linear mixed-effects (LME) 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. The ERP data, cleaned data files, and R scripts are available in Zhou, Garnsey, and Christianson (accepted).

2.2. Results

2.2.1. 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 (BAIS) was 71 out of 98. (The BAIS scores were not included in the LMEs for Experiment 1, as there was no APS involved. BAIS scores were collected in order to compare the participant groups in Experiment 1 and Experiment 2.)

2.2.2. EEG results

The target words were either verbs or pronouns (underlined in Table 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. 

Fig. 1 above shows the grand mean ERP responses to target pronouns (A) and verbs (B), and Tables 2–5 show the results of the statistical analyses of the P600 (Tables 2 and 3) and N400 (Tables 4 and 5) time windows at midline (Tables 2 and 4) and lateral (Tables 3 and 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 Fig. 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, grammaticality, and word type 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 word 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 pronoun case errors and verb agreement errors, shown by both reliable main effects of word type in three of the four analyses ($p < 0.01$) and interactions between word type and electrode location factors in all four analyses ($ps < 0.01$). This pattern arose because responses to the pronoun errors were more negative overall than to the verb agreement errors, especially at the front and on the left side of the head.

### 2.3. Experiment 1 summary

In sum, readers showed a classic P600 response to ungrammaticalities 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 et al., 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, consistent with some previous studies of sentences containing grammatical errors, which sometimes elicit N400 effects in addition to (or sometimes even instead of) P600 effects (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 (Tanner & Van Hell, 2014). Further consideration 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 APS. (It is not possible, though, to know whether any of the participants did so spontaneously.)
3. Experiment 2

Experiment 2 examines electrophysiological responses to grammatical errors when readers imagine either a native English speaker or a non-native English speaker while reading sentences silently. Eye-tracking (Stites et al., 2013; Yao & Scheepers, 2011; Zhou & Christianson, 2016a, 2016b) and fMRI (Yao et al., 2011, 2012) studies have shown that engaging in APS affects behavioral responses and brain activation patterns. Readers’ reading speeds are modulated by texts (“said quickly” vs. “said slowly”) or the speech rates of the speakers that they were perceptually simulating. APS can be triggered by direct quotation marks during silent reading, and readers’ auditory cortices 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; Hanuliková and colleagues, 2012). Most directly relevant is Hanuliková and colleagues (2012) study demonstrating that listeners respond differently to grammatical errors when they are spoken by native vs. 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.

3.1. Method

3.1.1. Participants

Eighty native English speakers (37 females) from the university community participated in Experiment 2. All were 18–30 years old (M = 19.8 years), 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.

3.1.2. Materials and design

A 2 × 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 factor 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 pseudorandomized 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 probe.

To induce participants to imagine the speech of either a native or a non-native speaker of English while reading the sentences, one photo of a Caucasian female and one of an Asian female were shown. Concurrently with the presentation of each photo, a short recording was played of a native and non-native woman’s voice, respectively (counterbalanced for order). The photos showed two women of similar age wearing similar clothing, who were not actually the speakers whose recordings were played (see Zhou & Christianson, 2016a), but who were stereotypically 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).

Fig. 2. The procedure of the paradigm to trigger auditory perceptual simulation in Experiment 2.

The Nelson-Denny Reading Task (NDRT) was used to measure participants’ reading speed and comprehension. Participants’ ability to imagine auditory stimuli was measured with the Bucknell Auditory Imagery Scale (BAIS; Herholz et al., 2012). Two social attractiveness surveys measured participants’ attitudes toward both the native and non-native English speakers, and another survey collected participants’ demographic information and language background.

3.1.3. Procedure

The ERP procedures were the same as in Experiment 1. After electrode application was completed, participants listened to two recordings (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 Fig. 2 below. Participants were instructed to sit in front of a computer monitor and read the upcoming sentences silently while imagining that they were being 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 in the speaker’s voice, 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 to refresh their memory for the voices. After ERP data collection was finished, participants completed the same questionnaires and took the

\(^3\) Participants’ familiarity with Chinese-accented English was not assessed. The University of Illinois has one of the largest populations of Chinese-national students, however, so it is unlikely that the participants had not interacted with Chinese English speakers. Moreover, we assume that if a participant was unfamiliar with a given accent, he or she would be less likely to be able to simulate that accent, and thus less likely to exhibit the key dissociations we observed here, which were dependent on APS of the accentuated speech.

\(^4\) The cartoon pictures here are not the photos we used in the experiment. We used stock, noncopyrighted photos found on the Internet. We controlled for age, expression, and socioeconomic status. The photo of the native English speaker was a blonde, Caucasian woman and the photo of the non-native speaker was a Chinese woman. Both of them were dressed in a business jacket and blouse and appeared to be in their early 30s. The photos showed the women from the shoulders up.

\(^5\) The duration of the native English speech: text 1 = 35 sec; text 2 = 25 sec; text 3 = 27 sec; text 4 = 28 sec; text 5 = 25 sec; text 6 = 24 sec. The duration of non-native speech: text 1 = 37 sec; text 2 = 46 sec; text 3 = 47 sec; text 4 = 40 sec; text 5 = 44 sec; text 6 = 41 sec. The audio files are presented in Zhou, Garnsey, & Christianson (accepted).
same tests of their reading speed, comprehension, and auditory imagery abilities as in Experiment 1. Each experimental session lasted 2.5–3 h.

3.1.4. EEG recording

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

3.1.5. EEG data analysis

Using the same procedures as in Experiment 1 for eliminating artifact-contaminated data, four participants were dropped, and approximately 9% of the trials for the remaining participants (N = 68) 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 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. Specifically, the dependent variables in linear mixed-effects (LME) models were the mean voltages for each time window for each participant in each condition. The independent predictors included APS conditions (no APS in Experiment 1 vs. native APS in Experiment 2 in the first set of analyses; native APS in Experiment 2 vs. non-native APS in Experiment 2 in the second set of analyses), grammaticality (grammatical/ungrammatical), word type (verb/pronoun), electrode laterality (left/right), electrode anteriority (anterior/posterior), and reading speed and comprehension accuracy in the Nelson-Denny Reading Test as the fixed effects and participants as random effects, using the maximal random effects structure (Barr et al., 2013). Because Nelson-Denny reading speed and comprehension accuracy were not significant in any of the analyses, they are not presented in the results tables. The ERP data, cleaned data files, and R scripts are available in Zhou, Garnsey, and Christianson (accepted).

3.2. Results

3.2.1. Behavioral results

Participants achieved an average comprehension accuracy of 93% in the paraphrase verification task for the target sentences, with no difference between the native-APS and non-native-APS conditions. 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. There were no significant differences in the Nelson-Denny comprehension accuracy (p = 0.17) or vividness of auditory imagery (p = 0.81) between the participants in Experiment 1 and Experiment 2. Social attractiveness survey results showed that participants perceived significant accent (p < 0.05) and speech rate (p < 0.05) differences between the native and non-native speakers. Participants showed more positive attitudes towards the native speech than the non-native speech on all measures except honesty and likability. Although these survey differences did not prove to be significant predictors of the ERP measures reported below, we include them here as evidence of multifaceted perceived differences between the two voices and speakers that were to be simulated. Although no data yet exist that demonstrate exactly how distinct voices need to be in order to successfully simulate each one without confusion, it seems likely that the more distinct the voices being imagined, the larger the APS effects will be (e.g., Kurby et al. (2009) used a male and a female voice to make them maximally distinct).

3.2.2. EEG results

No previous study has examined whether or how engaging in APS while reading sentences affects electrophysiological responses to grammatical errors, so it is first necessary to determine whether adding APS altered the pattern or magnitude of 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 in Experiment 2 will be compared with the native-speaker-APS responses, also from Experiment 2.

3.2.2.1. Comparison of no-APS (Experiment 1) and native-speaker-APS (Experiment 2). Fig. 3 shows the grand mean ERP waveforms for grammatical and ungrammatical pronouns (A) and verbs (B) in the native-speaker-APS conditions. Tables 6–9 show the results of the statistical analyses of the P600 (Tables 6 and 7) and N400 (Tables 8 and 9) time windows at midline (Tables 6 and 8) and lateral (Tables 7 and 9).
and 9) scalp sites. Visual inspection of Fig. 3 reveals a very similar overall pattern of responses as in Experiment 1, when there were no instructions to imagine a voice. However, comparing Fig. 3 to Fig. 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.

Just as in Experiment 1, verb agreement errors and pronoun case errors 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 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 anterior interactions in midline analyses), but more so for pronouns than for verbs (p < 0.05 for grammaticality x word type x anterior interactions in midline analyses). There was also an interaction between 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 (ps < 0.01), as well as interactions between word type and electrode location factors (ps < 0.01). These 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,
Fig. 4. 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.

Fig. 5. Grand mean ERP waveforms for the Non-Native-Speaker-APS conditions for pronouns (A) and verbs (B) in Experiment 2.
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 ($p < 0.05$).

Fig. 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 were 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. Most importantly, the grammaticality effect in the P600 window disappeared for the verb agreement errors. There was still a grammaticality effect for the verbs, but it manifested as increased negativity in the N400 window instead.

### Table 11
Native-APS vs non-native-APS, P600 window, lateral sites.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.32</td>
<td>0.12</td>
<td>193.7</td>
<td>2.78</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Grammaticality (Gram)</td>
<td>-0.27</td>
<td>0.11</td>
<td>1790</td>
<td>-2.51</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Gram x Anteriority (Ant)</td>
<td>0.25</td>
<td>0.12</td>
<td>1790</td>
<td>2.07</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Gram x Latency (Lat)</td>
<td>0.02</td>
<td>0.12</td>
<td>1790</td>
<td>0.17</td>
<td>0.87</td>
</tr>
<tr>
<td>Gram x Word Type (Wtype)</td>
<td>0.25</td>
<td>0.15</td>
<td>1790</td>
<td>1.66</td>
<td>&lt; 0.1</td>
</tr>
<tr>
<td>Wtype x Ant</td>
<td>0.02</td>
<td>0.17</td>
<td>1790</td>
<td>0.12</td>
<td>0.91</td>
</tr>
<tr>
<td>Wtype x Lat</td>
<td>0.4</td>
<td>0.12</td>
<td>666.06</td>
<td>3.31</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>APS x Gram</td>
<td>0.08</td>
<td>0.15</td>
<td>1790</td>
<td>0.53</td>
<td>0.59</td>
</tr>
<tr>
<td>APS x Wtype</td>
<td>0.26</td>
<td>0.15</td>
<td>1790</td>
<td>1.74</td>
<td>&lt; 0.1</td>
</tr>
<tr>
<td>APS x Gram x Wtype</td>
<td>0.42</td>
<td>0.21</td>
<td>1790</td>
<td>-1.96</td>
<td>0.15</td>
</tr>
<tr>
<td>APS x Ant</td>
<td>-0.16</td>
<td>0.12</td>
<td>1790</td>
<td>1.3</td>
<td>0.19</td>
</tr>
<tr>
<td>APS x Lat</td>
<td>0.01</td>
<td>0.12</td>
<td>1790</td>
<td>0.1</td>
<td>0.92</td>
</tr>
<tr>
<td>APS x Gram x Ant</td>
<td>-0.33</td>
<td>0.17</td>
<td>1790</td>
<td>-1.89</td>
<td>&lt; 0.1</td>
</tr>
<tr>
<td>APS x Gram x Lat</td>
<td>0.07</td>
<td>0.17</td>
<td>1790</td>
<td>0.38</td>
<td>0.7</td>
</tr>
<tr>
<td>APS x Wtype x Ant</td>
<td>0.12</td>
<td>0.17</td>
<td>1790</td>
<td>0.02</td>
<td>0.99</td>
</tr>
<tr>
<td>APS x Wtype x Lat</td>
<td>0.12</td>
<td>0.17</td>
<td>1790</td>
<td>0.68</td>
<td>0.49</td>
</tr>
<tr>
<td>APS x Gram x Wtype x Ant</td>
<td>0.51</td>
<td>0.25</td>
<td>1790</td>
<td>2.08</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Ant</td>
<td>-0.12</td>
<td>0.1</td>
<td>555</td>
<td>-1.27</td>
<td>0.21</td>
</tr>
<tr>
<td>Ant x Wtype x Lat</td>
<td>-0.42</td>
<td>0.42</td>
<td>1790</td>
<td>-4.05</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

### Table 12
Native-APS vs non-native-APS, N400 window, midline sites.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.02</td>
<td>0.11</td>
<td>297.4</td>
<td>0.14</td>
<td>0.89</td>
</tr>
<tr>
<td>Grammaticality (Gram)</td>
<td>0.05</td>
<td>0.14</td>
<td>927</td>
<td>0.32</td>
<td>0.75</td>
</tr>
<tr>
<td>Gram x Anteriority (Ant)</td>
<td>0.03</td>
<td>0.19</td>
<td>1197</td>
<td>0.18</td>
<td>0.86</td>
</tr>
<tr>
<td>Gram x Anteriority (Ant)</td>
<td>0.15</td>
<td>0.19</td>
<td>1197</td>
<td>0.77</td>
<td>0.44</td>
</tr>
<tr>
<td>Gram x Word Type (Wtype)</td>
<td>0.41</td>
<td>0.19</td>
<td>1197</td>
<td>2.13</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Wtype x Ant</td>
<td>-0.24</td>
<td>0.27</td>
<td>1197</td>
<td>-0.88</td>
<td>0.38</td>
</tr>
<tr>
<td>Wtype</td>
<td>0.17</td>
<td>0.15</td>
<td>580.2</td>
<td>1.14</td>
<td>0.26</td>
</tr>
<tr>
<td>Wtype x Ant</td>
<td>-0.11</td>
<td>0.19</td>
<td>1197</td>
<td>-0.56</td>
<td>0.57</td>
</tr>
<tr>
<td>APS</td>
<td>-0.38</td>
<td>0.15</td>
<td>800.1</td>
<td>-2.59</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>APS x Gram</td>
<td>0.27</td>
<td>0.19</td>
<td>1197</td>
<td>1.39</td>
<td>0.16</td>
</tr>
<tr>
<td>APS x Wtype</td>
<td>0.44</td>
<td>0.19</td>
<td>1197</td>
<td>2.27</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>APS x Gram x Wtype</td>
<td>0.54</td>
<td>0.21</td>
<td>1197</td>
<td>0.98</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>APS x Ant</td>
<td>0.16</td>
<td>0.21</td>
<td>1197</td>
<td>0.67</td>
<td>0.41</td>
</tr>
<tr>
<td>APS x Ant</td>
<td>0.3</td>
<td>0.19</td>
<td>1197</td>
<td>1.58</td>
<td>0.11</td>
</tr>
<tr>
<td>APS x Ant</td>
<td>0.36</td>
<td>0.27</td>
<td>1197</td>
<td>1.32</td>
<td>0.19</td>
</tr>
<tr>
<td>APS x Wtype x Ant</td>
<td>0.09</td>
<td>0.27</td>
<td>1197</td>
<td>-0.34</td>
<td>0.73</td>
</tr>
<tr>
<td>APS x Ant x Wtype</td>
<td>0.52</td>
<td>0.27</td>
<td>1197</td>
<td>-1.19</td>
<td>0.23</td>
</tr>
<tr>
<td>APS x Gram x Wtype x Ant</td>
<td>0.38</td>
<td>0.19</td>
<td>1197</td>
<td>0.5</td>
<td>0.62</td>
</tr>
<tr>
<td>Ant</td>
<td>0.24</td>
<td>0.14</td>
<td>1006.1</td>
<td>-1.76</td>
<td>&lt; 0.1</td>
</tr>
</tbody>
</table>

### Table 13
Native-APS vs non-native-APS, N400 window, lateral sites.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.02</td>
<td>0.11</td>
<td>287.4</td>
<td>0.14</td>
<td>0.89</td>
</tr>
<tr>
<td>Grammaticality (Gram)</td>
<td>0.05</td>
<td>0.14</td>
<td>927</td>
<td>0.32</td>
<td>0.75</td>
</tr>
<tr>
<td>Gram x Anteriority (Ant)</td>
<td>0.03</td>
<td>0.19</td>
<td>1197</td>
<td>0.18</td>
<td>0.86</td>
</tr>
<tr>
<td>Gram x Anteriority (Ant)</td>
<td>0.15</td>
<td>0.19</td>
<td>1197</td>
<td>0.77</td>
<td>0.44</td>
</tr>
<tr>
<td>Gram x Word Type (Wtype)</td>
<td>0.41</td>
<td>0.19</td>
<td>1197</td>
<td>2.13</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Wtype x Ant</td>
<td>-0.24</td>
<td>0.27</td>
<td>1197</td>
<td>-0.88</td>
<td>0.38</td>
</tr>
<tr>
<td>Wtype</td>
<td>0.17</td>
<td>0.15</td>
<td>580.2</td>
<td>1.14</td>
<td>0.26</td>
</tr>
<tr>
<td>Wtype x Ant</td>
<td>-0.11</td>
<td>0.19</td>
<td>1197</td>
<td>-0.56</td>
<td>0.57</td>
</tr>
<tr>
<td>APS</td>
<td>-0.38</td>
<td>0.15</td>
<td>800.1</td>
<td>-2.59</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>APS x Gram</td>
<td>0.27</td>
<td>0.19</td>
<td>1197</td>
<td>1.39</td>
<td>0.16</td>
</tr>
<tr>
<td>APS x Wtype</td>
<td>0.44</td>
<td>0.19</td>
<td>1197</td>
<td>2.27</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>APS x Gram x Wtype</td>
<td>0.54</td>
<td>0.21</td>
<td>1197</td>
<td>0.98</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>APS x Ant</td>
<td>0.16</td>
<td>0.21</td>
<td>1197</td>
<td>0.67</td>
<td>0.41</td>
</tr>
<tr>
<td>APS x Ant</td>
<td>0.3</td>
<td>0.19</td>
<td>1197</td>
<td>1.58</td>
<td>0.11</td>
</tr>
<tr>
<td>APS x Ant</td>
<td>0.36</td>
<td>0.27</td>
<td>1197</td>
<td>1.32</td>
<td>0.19</td>
</tr>
<tr>
<td>APS x Wtype x Ant</td>
<td>0.09</td>
<td>0.27</td>
<td>1197</td>
<td>-0.34</td>
<td>0.73</td>
</tr>
<tr>
<td>APS x Ant x Wtype</td>
<td>-0.32</td>
<td>0.27</td>
<td>1197</td>
<td>-1.19</td>
<td>0.23</td>
</tr>
<tr>
<td>APS x Gram x Wtype x Ant</td>
<td>0.38</td>
<td>0.19</td>
<td>1197</td>
<td>0.5</td>
<td>0.62</td>
</tr>
<tr>
<td>Ant</td>
<td>0.38</td>
<td>0.21</td>
<td>1772</td>
<td>-4.37</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Lat</td>
<td>0.03</td>
<td>0.1</td>
<td>1772</td>
<td>0.3</td>
<td>0.75</td>
</tr>
</tbody>
</table>

### 4. Discussion

Two ERP experiments were conducted to compare 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 was that if auditory perceptual simulation (APS) is analogous in certain respects to “listening” to speech, then the electrophysiological response patterns should be similar to the results of Hamulčiková et al. (2012), in which listening to agreement errors in native speech led to P600 effects, while listening to the same errors in non-native speech resulted in no P600 effects. Thus, we predicted that the P600 effects should be observed in both non-APS (normal silent reading) and APS of native speech conditions with no significant
Fig. 6. 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.

differences. However, when readers activate APS of non-native speech, they should tend to “forgive” the morphosyntactic violation, and there will be no P600 effects in their electrophysiological responses to the ungrammatical words.

Results confirmed our hypothesis that APS of different speakers’ voices affected electrophysiological responses to morphosyntactic errors during silent reading differentially. Critically, ERP waveforms reflecting the APS of the native speaker in Experiment 2 showed the same pattern of effects as the non-APS group in Experiment 1. APS of the non-native speaker in Experiment 2, however, triggered a different ERP pattern. When readers were instructed to imagine a native speaker’s voice in Experiment 2, we observed similar grammaticality effects in both the N400 and P600 windows for both verb agreement errors and pronoun case errors compared to not imagining any voice in Experiment 1. The size of these grammaticality effects decreased numerically from no-APS to native APS, but that decrease was not reliable in the statistical analyses. It is possible that population differences were causing this difference, or that any APS at all encourages some amount of grammatical “leeway.” Future research should address these possibilities.

Most importantly, the grammaticality effect in the P600 window for subject-verb agreement violations disappeared when readers imagined a non-native speaker’s voice in Experiment 2. Hanuliková et al. (2012) also found that hearing grammatical gender agreement errors in Dutch spoken by a non-native Dutch speaker eliminated P600 effects compared to listening to a native Dutch speaker. Our results differed from theirs, however, in that we found a reliable grammaticality effect in the N400 window, while they did not. Examination of Hanuliková et al.’s (2012) waveforms in the N400 window appears to reveal a similar increase in N400 amplitude to what we found, but their N400 effects were not significant. Thus, the disappearance here of the P600 effect in response to subject-verb disagreement is very similar to the pattern found in their study with gender disagreement. (Hanuliková et al. 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 therefore appears that imagining native and non-native voices while silently reading affects responses to morphosyntactic errors in much the same way as actually hearing voices articulate errors, which suggests that imagining a voice during silent reading is similar in at least some respects to hearing a 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 and Van Hell (2017) also observed an increased N400 and reduced N60 when participants listened to pronoun 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 subject-verb agreement, 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 subject-verb agreement 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 participants. 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 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 canceled by the beginning of a partially overlapping P600. If this 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.

Finally, we note a recent behavioral study by Gibson et al. (2017), who also conducted a conceptual replication of Hanuliková et al. (2012) and found that native English speakers were likely to “forgive” semantic errors committed by non-native English speakers. Specifically, they found that sentences like “The mother gave the candle the daughter” were more likely to be understood as “The mother gave the candle to the daughter” when uttered by a non-native English speaker than when uttered by a native English speaker. Gibson and colleagues accounted for this finding within the “noisy channel” model of language comprehension (Gibson et al., 2013). Within this model, a Bayesian equation is applied to all language input that takes into account prior language input from a given source (or speaker) and continually adjusts comprehension processes to determine the most likely intended interpretation of new input. Interestingly, the equation applied by Gibson et al. (2013, 2017) does not differentiate between temporally distal and local language “prior,” and as such, appears to encounter difficulty accounting for the pattern of results observed here. Note that in Experiment 2, the sentences imagined in the both the native and non-native speaker were equally likely to contain subject-verb agreement errors and pronoun case errors (in equal portions). Therefore, the local priors — i.e., the likelihood of encountering errors, or “noisy” input, in the context of the experimental stimuli — were equal for both speakers’ voices that were being simulated. Fine, Jaeger, Farmer, and Qian (2013) claimed that local expectations about input might be rapidly updated during the course of an experiment. If this claim is correct, under the noisy channel account, one might have expected equally errorful non-native and native “production” (really, imagined production) to result in equally diminished P600s for both the two speakers and the two error types. In order for the noisy channel to predict the present results, however, a distinction would need to be made between local priors (local to the experimental stimuli) and distal priors (previous experience with Chinese-L1/English-L2 speakers), in which subject-verb agreement errors are more frequent than pronoun-case errors. Furthermore, some sort of weighting of distal over local priors (or local
over distal, in the case of Fine et al., 2013) would need to be determined
to make accurate predictions.

An alternative but related language processing framework, Good
Enough Processing (Christianson, 2016; Ferrera, Bailey, & Ferraro,
2002), was proposed by Zhou and Christianson (2016a) to account for
APS effects observed in that study. Briefly, what was proposed was that
APS generated an elaborated prosodic representation that bolstered the
syntactic representation of input that was read while performing APS.
The resulting syntactic representation was more resilient to interference
from semantic, or heuristic, influences on interpretation and therefore
more faithful to the actual input. In the present study, the pronoun-case
errors involved free morphemes (pronouns), which contribute to the
prosodic contour of the input sentence. The verb-agreement errors, on
the other hand, consisted only of the bound morpheme -s, which does
not contribute to the sentence's prosodic contour. Under the Good En-
ough account, then, it is the verb-agreement errors that should be more
likely to be shifted to a semantic processing route, signaled by the in-
creased N400. In contrast, the pronoun-case errors, as part of the pro-
sodic/syntactic representations, should be more likely to continue to
disrupt syntactic processing, irrespective of whose voice they are asso-
ciated with. We leave it to future research to adjudicate between
these alternative accounts of the APS effects observed here.

5. Conclusion

This ERP study is the first to explore readers' electrophysiological
responses when they imagine different voices during silent reading. The
results demonstrated that APS of a native English speaker's voice yields
similar brain responses to grammatical errors as normal silent reading.
In contrast, APS of a non-native voice showed electrophysiological
responses to subject-verb agreement errors, resulting in a grammati-
cally effect in the N400 window, but not in the P600 window, showing
similarity to the findings of Hanuličková et al. (2012). Results indicate
that actively imagining speakers' voices during silent reading is similar,
at least in certain ways, to listening to the voices.

Acknowledgement

We thank the Beckman Institute Language and Brain Lab members,
Keqi Wei, Young Jae Lee, Laeh Ragans, Shaolingyun Guo for their va-
luable help in collecting data. The project was funded by the research
grant awarded to Kiel Christianson and Peiyun Zhou from the Confucius
Institute at University of Illinois Urbana-Champaign, United States as
well as a University of Illinois Graduate Dissertation Fellowship and a
Beckman Graduate Fellowship to Peiyun Zhou.

Appendix A

Experimental stimuli used in the study

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>His fellow classmates dislike him since they are jealous.</td>
</tr>
<tr>
<td></td>
<td>His fellow classmates dislike he since they are jealous.</td>
</tr>
<tr>
<td></td>
<td>His fellow classmates dislikes him since they are jealous.</td>
</tr>
<tr>
<td>2</td>
<td>The lab assistant orders beakers for her lab every month.</td>
</tr>
<tr>
<td></td>
<td>The lab assistant orders beakers for she lab every month.</td>
</tr>
<tr>
<td></td>
<td>The lab assistant order beakers for her lab every month.</td>
</tr>
<tr>
<td>3</td>
<td>Jimmy's parents worry because he plays with dolls.</td>
</tr>
<tr>
<td></td>
<td>Jimmy's parents worry because him plays with dolls.</td>
</tr>
<tr>
<td></td>
<td>Jimmy's parents worries because he plays with dolls.</td>
</tr>
<tr>
<td>4</td>
<td>The kidnapper apparently lures them into his van.</td>
</tr>
<tr>
<td></td>
<td>The kidnapper apparently lures they into his van.</td>
</tr>
<tr>
<td></td>
<td>The kidnapper apparently lure them into his van.</td>
</tr>
<tr>
<td>5</td>
<td>The pirate attacks the merchants and seizes their ship.</td>
</tr>
<tr>
<td></td>
<td>The pirate attacks the merchants and seizes they ship.</td>
</tr>
<tr>
<td></td>
<td>The pirate attacks the merchants and seize their ship.</td>
</tr>
<tr>
<td>6</td>
<td>Birds migrate to south with their family when winter approaches.</td>
</tr>
<tr>
<td></td>
<td>Birds migrate to south with they family when winter approaches.</td>
</tr>
<tr>
<td></td>
<td>Birds migrates to south with their family when winter approaches.</td>
</tr>
<tr>
<td>7</td>
<td>I practice my trombone every day.</td>
</tr>
<tr>
<td></td>
<td>I practice me trombone every day.</td>
</tr>
<tr>
<td></td>
<td>I practices my trombone every day.</td>
</tr>
<tr>
<td>8</td>
<td>Her mechanic always treats her with respect.</td>
</tr>
<tr>
<td></td>
<td>Her mechanic always treat she with respect.</td>
</tr>
<tr>
<td></td>
<td>Her mechanic always treat her with respect.</td>
</tr>
<tr>
<td>9</td>
<td>The founder is indicted for his insider trading practices.</td>
</tr>
<tr>
<td></td>
<td>The founder is indicted for him insider trading practices.</td>
</tr>
<tr>
<td></td>
<td>The founder is indict for his insider trading practices.</td>
</tr>
<tr>
<td>10</td>
<td>Vince feels nervous whenever he punts the football.</td>
</tr>
<tr>
<td></td>
<td>Vince feels nervous whenever him punts the football.</td>
</tr>
<tr>
<td></td>
<td>Vince feels nervous whenever he punt the football.</td>
</tr>
<tr>
<td>11</td>
<td>Lily prints documents for her colleagues everyday.</td>
</tr>
<tr>
<td></td>
<td>Lily prints documents for she colleagues everyday.</td>
</tr>
<tr>
<td></td>
<td>Lily print documents for her colleagues everyday.</td>
</tr>
<tr>
<td>12</td>
<td>Their strange uncle frighten them with ghost stories.</td>
</tr>
<tr>
<td></td>
<td>Their strange uncle frighten they with ghost stories.</td>
</tr>
<tr>
<td></td>
<td>Their strange uncle frighten them with ghost stories.</td>
</tr>
</tbody>
</table>
The kids go fishing in their parents' boat.
The kids go fishing in their parents' boat.
The kids go fishing in their parents' boat.
Alex cleans his father's car for some pocket money.
Alex cleans his father's car for some pocket money.
Alex cleans his father's car for some pocket money.
The accountant submits the financial report to her supervisor.
The accountant submits the financial report to her supervisor.
The accountant submits the financial report to her supervisor.
My boyfriend forgets how we met each other.
My boyfriend forgets how we met each other.
My boyfriend forgets how we met each other.
We all really rely on her for support on this project.
We all really rely on her for support on this project.
We all really rely on her for support on this project.
Actors spend hours memorizing their lines.
Actors spend hours memorizing their lines.
Actors spend hours memorizing their lines.
Our rich neighbour brings us paintings from Europe.
Our rich neighbour brings us paintings from Europe.
Our rich neighbour brings us paintings from Europe.
My pet bites my sofa when I am out.
My pet bites my sofa when I am out.
My pet bite my sofa when I am out.
On weekend Ike parks his car in back.
On weekend Ike parks his car in back.
On weekend Ike parks his car in back.
The writer revises his work several times before publishing.
The writer revises his work several times before publishing.
The writer revises his work several times before publishing.
Kay and Jimmy are good friends since they share common interests.
Kay and Jimmy are good friends since they share common interests.
Kay and Jimmy are good friends since they share common interests.
They need to drink water when they feel thirsty.
They need to drink water when they feel thirsty.
They need to drink water when they feel thirsty.
Posters of musicians cover her bedroom wall.
Posters of musicians cover her bedroom wall.
Posters of musicians cover her bedroom wall.
Alex turns off the light when I replace the bulb.
Alex turns off the light when I replace the bulb.
Alex turns off the light when I replace the bulb.
Gideon has a problem feeding his family during the famine.
Gideon has a problem feeding his family during the famine.
Gideon has a problem feeding his family during the famine.
The taxi driver stops the car when he notices a stop sign.
The taxi driver stops the car when he notices a stop sign.
The taxi driver stops the car when he notices a stop sign.
Ray stays late to finish his math homework.
Ray stays late to finish his math homework.
Ray stays late to finish his math homework.
Few colleges truly prepare students for their future careers.
Few colleges truly prepare students for their future careers.
Few colleges truly prepare students for their future careers.
Larry receives a letter that his fiancé sends from Paris monthly.
Larry receives a letter that his fiancé sends from Paris monthly.
Larry receives a letter that his fiancé sends from Paris monthly.
The babysitter ends her work at 7 pm on the weekend.
The babysitter ends her work at 7 pm on the weekend.
The babysitter ends her work at 7 pm on the weekend.
He calls the dentist when his teeth hurt
He calls the dentist when his teeth hurt.
He call the dentist when his teeth hurt.
The ugly guy protects us from the rain.
The ugly guy protects us from the rain.
The ugly guy protect us from the rain.
I learn a lot about her at the summer camp. I learn a lot about her at the summer camp.
When a student comes to class he brings his textbook.
When a student comes to class he brings his textbook.
Jim plants some roses on his roof to get the sunshine.
Jim plants some roses on his roof to get the sunshine.
The witch curses the prince and turns him into a donkey.
The witch curses the prince and turns him into a donkey.
He boffs up his erand thoroughly.
He boffs up his erand thoroughly.
Jim’s sister never offers to shake my hand.
Jim’s sister never offers to shake my hand.

References

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of


