This article presents the data analyzed in the paper “Is imagining a voice like listening to it? Evidence from ERPs” [1]. The data include individual ERP data when participants were performing auditory imagery of native and non-native English speech during silent reading vs. normal silent reading, and behavioral results from participants performing the Nelson-Denny Reading Comprehension task and Bucknell Auditory Imagery Scale (BAIS). The repository includes the R scripts used to carry out the statistical analyses reported in the original paper.

© 2020 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Data

This paper presents the human experimental data reported in the paper “Is imagining a voice like listening to it? Evidence from ERPs” [1]. Both individual data and the processed data files are available. Individual data include ERP data files, Nelson-Denny Reading Comprehension task data, and Bucknell Auditory Imagery Scale (BAIS) [2] data. Additionally, the R scripts for analyzing the results and generating the figures are also provided. The overall organization of the experimental materials, data, and analysis scripts follow the hierarchical structure presented in Fig. 1.
All data reported in Ref. [1] are organized into the “data,” “R script,” and “Experiment materials” folders. The “data” folder contains four subfolders, including ERP data files, Nelson-Denny Reading Comprehension data, BAIS data, and combined files data for R analysis. The “R script” folder includes the scripts for data analysis in each time window. The “Experiment materials” folder contains the target stimuli for experiment in Excel format, the flowchart for the Auditory Perceptual Simulation (APS) paradigm, and audio files used in the APS paradigm. Fig. 2 presents the procedure of the Auditory Perceptual Simulation (APS) paradigm. A copy of this repository is also available on the Peiyun Zhou’s GitHub account page, at https://github.com/manqi11/Auditory-Imagery.git.

### 2. Experimental design, materials and methods

#### 2.1. Experiment 1

**2.1.1. Participants**

Forty-eight native English speakers (21 females) from the University of Illinois at Urbana-Champaign community participated in Experiment 1. All participants 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. They received cash compensation or course credit for their time.

**2.1.2. Materials and design**

Two types of grammatical errors, subject-verb number agreement and pronoun case mismatch (extending the grammatical errors used by Ref. [16]), 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. Ref. [3]), 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 English L2 learners’ writing [4] (see also [5]) 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 are presented in Table 1 below (see materials in target stimuli). 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

---

There is no subject-verb agreement of any sort in Chinese. Native English speakers may therefore 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 very infrequent in their speech.
number or pronoun case. Eighty of the filler sentences were experimental items for another study [6] 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 2 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.

2.1.3. Nelson-Denny Reading Task

Additionally, participants’ reading speed and comprehension were measured using the comprehension portion of the Nelson-Denny Reading Task (NDRT) [7], which is a standardized paper-and-pencil test for college readers. The comprehension test contains seven reading passages with 38 comprehension probes and lasted 20 minutes. 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.

2.1.4. Bucknell Auditory Imagery Scale (BAIS)

The Bucknell Auditory Imagery Scale (BAIS) [2] 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.

2.1.5. 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.

2.1.6. 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 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.

---

Table 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) |

---

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.
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 (BAIS) [2].

2.1.7. EEG data analysis

EEG and EOG data were analyzed using the EEGLab [8] and ERPLab [9] 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 [10].) 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 [11], linear mixed-effects (LME) models [12] 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 [13]. 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.

2.2. Experiment 2

2.2.1. Participants

Eighty native English speakers (37 females) from the University of Illinois at Urbana-Champaign 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. Participants received cash payment or course credit for their time.

2.2.2. 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 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 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 (counter-balanced for order). The photos showed two women of similar age wearing similar clothing, who were not actually the speakers whose recordings were played (see Refs. [14,15] for full details of the paradigm), 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). During the reading portion of the experiment, participants were shown one of the two photos along with the matching name of the “speaker” whose voice they were supposed to imagine “saying” the sentence that was then presented and read silently by the participant.

The Nelson-Denny Reading Task (NDRT) [7] was used to measure participants’ reading speed and comprehension. Participants’ ability to imagine auditory stimuli was measured with the BIAS [2]. Two social attractiveness surveys measured participants’ attitudes toward both the native and non-native English speakers,3 and another survey collected participants’ demographic information and language background.

2.2.3. 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 photographs4 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 passages5 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.

2.2.4. EEG recording

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

2.2.5. 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

---

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 accented 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 = 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.
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 [13]. Because Nelson-Denny reading speed and comprehension accuracy were not significant in any of the analyses, they are not presented in the results tables.

Acknowledgements

The research was funded by a research grant awarded to Kiel Christianson and Peiyun Zhou from the Confucius Institute at University of Illinois as well as a University of Illinois Graduate Dissertation Fellowship and a Beckman Graduate Fellowship to Peiyun Zhou.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2020.105242.

References

[1] P. Zhou, S. Garnsey, K. Christianson, Is imagining a voice like listening to it? Evidence from ERPs, Cognition 18 (2019) 227–241, https://doi.org/10.1016/j.cognition.2018.10.014.
[2] S.C. Herholz, A.R. Halpern, R.J. Zatorre, Neuronal correlates of perception, imagery, and memory for familiar tunes, J. Cognit. Neurosci. 24 (2012) 1382–1397, https://doi.org/10.1162/jocn_a_00216.
[3] S. Coulson, J.W. King, M. Kutas, Expect the unexpected: event-related brain response to morphosyntactic violations, Lang. Cognit. Process. 13 (1998) 21–58, https://doi.org/10.1080/016909698386582.
[4] F.-Y. Chuang, H. Nesi, An analysis of formal errors in a corpus of L2 English produced by Chinese students, Corpora 1 (2006) 251–271, https://doi.org/10.3366/cor.2006.1.2.251.
[5] S. Darus, K.H. Ching, Common errors in written English essays of form one Chinese students: a case study, Eur. J. Soc. Sci. 10 (2009) 242–253.
[6] P. Zhou, J. Toscano, S. Garnsey, K. Christianson, in: Structural Complexity Modulates “Semantic P600” in Sentence Processing, Poster Presented at the 30th CUNY Conference, Boston, MA, 2017.
[7] J.I. Brown, V.V. Fishco, G. Hanna, Nelson-denny Reading Test: Manual for Scoring and Interpretation, forms G & H. Riverside Publishing Company, 1993.
[8] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (2004) 9–21, https://doi.org/10.1016/j.jneumeth.2003.10.009.
[9] J. Lopez-Calderon, S.J. Luck, ERPLAB: an open-source toolbox for the analysis of event-related potentials, Front. Hum. Neurosci. 8 (2014) 213, https://doi.org/10.3389/fnhum.2014.00213.
[10] A.C. Gouvea, C. Phillips, N. Kazanina, D. Poeppel, The linguistic processes underlying the P600, Lang. Cognit. Process. 25 (2010) 149–188, https://doi.org/10.1080/10603660902965951.
[11] D. Bates, M. Maechler, B. Bolker, lme4: Linear Mixed-Effects Models Using S4 Classes. R Package Version 0, ScienceOpen.com, 2011, 999375–42.
[12] R.H. Baayen, D.J. Davidson, D.M. Bates, Mixed-effects modeling with crossed random effects for subjects and items, J. Mem. Lang. 59 (2008) 390–412, https://doi.org/10.1016/j.jml.2007.12.005.
[13] D.J. Barr, R. Levy, C. Scheepers, H.J. Tily, Random effects structure for confirmatory hypothesis testing: keep it maximal, J. Mem. Lang. 68 (2013) 255–278, https://doi.org/10.1016/j.jml.2012.11.001.
[14] P. Zhou, K. Christianson, I “hear” what you’re “saying.” Auditory perceptual simulation, reading speed, and comprehension, Q. J. Exp. Psychol. 69 (2016) 972–995, https://doi.org/10.1080/17470218.2015.1018282.
[15] P. Zhou, K. Christianson, Auditory perceptual simulation: simulating speech rates or accents? Acta Psychol. 168 (2016) 85–90, https://doi.org/10.1016/j.actpsy.2016.04.005.
[16] A. Hanušková, P.M. van Alphen, M.M. van Goch, A. Weber, When one person’s mistake in another’s standard usage: the effect of foreign accent on syntactic processing, J. Cognit. Neurosci. 24 (2012) 878–887, https://doi.org/10.1162/jocn_a_00103.