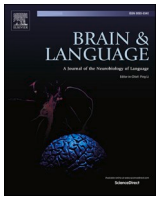
Brain & Language 269 (2025) 105621

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l



Neural mechanisms of face cue predictability and the integration of facial and acoustic cues in native- and nonnative-accented words

Daisy Lei [a,b,*], Yuka Tatsumi [b], Erica Hsieh [b], Cristal Giorio [a,b], Janet G. van Hell [a,b]

[a] Department of Psychology, The Pennsylvania State University, USA
[b] Center for Language Science, The Pennsylvania State University, USA

ARTICLE INFO

Keywords:
Accented speech
Face cue
Word processing
Speech processing
ERP

ABSTRACT

This study examined the integration of face cue and native- and nonnative-accented English speech by manipulating the face cue predictability of a speaker's accent (predictable: one accent (American-accent or Chinese-accented), not predictable: two accents (American-accented and Chinese-accented). Monolingual listeners were first familiarized with each speaker's number and type of accent(s). Then, they completed an EEG-recorded auditory go/no-go animal decision task where the no-go items were critical words and nonwords. Listeners saw a face cue (speaker image) before speech onset and concurrently with the speech. Pre-speech ERP results revealed that listeners processed face cues differently based on face cue predictability. Post-speech ERP analyses revealed N400 lexicality effects for native-accented speech, and face cue predictability effects for nonnative-accented speech. No N400 effects were found for an audio-only experiment. This indicates that the monolingual listeners integrate face cues and auditory cues during real-time nonnative-accented speech processing, but not during native-accented speech processing.

1. Introduction

Nonnative-accented speech and bilingualism are growing more prevalent as globalization increases international travel and immigration, in the United States, European Union, and globally, over the past decades (Ethnologue, 2023; European Commission, 2018; United States Census Bureau, 2022). Worldwide, over 75% of English speakers are second language, nonnative speakers (Ethnologue, 2023).

Importantly, everyday speech comprehension often occurs in a multimodally rich environment and our brains are continuously integrating input from both the auditory and visual modality (A. E. Martin, 2016), including information cueing a speaker's identity. Knowledge related to speaker identity may evoke speaker related top-down processes during listeners' speech perception (e.g., Kutlu et al., 2022; Rubin, 1992). For example, American native-accented speech preceded by an Asian face was rated as more accented and less efficient as a teacher than the same native-accented speech preceded by a White face (Kang & Rubin, 2009). For nonnative-accented speech, participants showed better comprehension when Chinese-accented speech was paired with an Asian face than with a White face (McGowan, 2015). These findings were paralleled in a recent electroencephalogram (EEG) study measuring brain activity during online sentence processing (Grey, Cosgrove, & Van Hell, 2020). When paired with a "congruent" face cue (Asian face) in the face cue group (Grey, Cosgrove, & Van Hell, 2020), the processing of Chinese-accented English sentences containing pronoun violations revealed P600 effects, which have been related to sentence level restructuring and reintegration in response to pronoun errors, while no Event-Related Potential (ERP) effects were found in the audio-only group. The authors interpreted the presence of P600 effects for the face cue group (and lack of such effects in the audio-only group) to suggest that face cue information prepared the listeners that the incoming speech signal may deviate from canonical (native accent) pronunciations, thereby engaging mechanisms associated with sentence level restructuring and reanalysis.

To further understand the role of face cues in processing native-accented and nonnative-accented speech, this study manipulated the predictability of a face cue (predictable: one face one accent, or unpredictable: one face two accents) in cueing the upcoming speech accent.

* Corresponding author.
E-mail address: daisylei415@gmail.com (D. Lei).

https://doi.org/10.1016/j.bandl.2025.105621
Received 1 August 2024; Received in revised form 12 July 2025; Accepted 13 July 2025
Available online 26 July 2025

1.1. Cue integration model

Theories of speech perception have addressed variability in the speech signal by delineating the processing of an acoustic–phonetic signal into lexical recognition and/or phrase/sentence processing (e.g., Cohort theory, Marslen-Wilson & Tyler, 1980; Cue-integration model, A. E. Martin, 2016; TRACE model, McClelland & Elman, 1986). The cue integration model (A. E. Martin, 2016), an interactive speech perception model, holds that linguistic representations are encoded in neural population codes, and that language comprehension involves neural computations of cue combination and cue integration of linguistic representations extracted from auditory and visual input. This model articulates psycholinguistic cues as any internal representation of the environment that is relevant for language processing. Cue combination refers to the summation of cues that do not carry redundant information while cue integration refers to the integration of cues with redundant information regarding the same detail of the environment. The internal representations may be separated into perceptual representations, abstract representations, as well as contextual factors. Perceptual representations include the acoustic signal, phoneme, syllable, word, etc., while abstract representations include syntactic structure, event structure, scope taking, anaphora, etc. Contextual factors include the situation model, discourse context, pragmatic meaning, and semantic memory. A given representation may serve as an input cue to different representations, even if the computations for the given representation are not complete. For example, representations for a word or phrase may receive input from an emerging syllabic representation prior to the complete processing of the acoustic input into the syllabic representation. Upon receiving certain acoustic cues, contextual representations may already be activated and serve as a cue for the phoneme level representations. Each cue is proposed to be weighted by a recent reliability and a latent reliability term. Cue reliability refers to the probabilistic relationship between a cue and an upcoming representation. Recent reliability reflects the reliability of linguistic relationships in a local processing context while latent reliability refers to a more stable global reliability that reflects latent knowledge. The reliability of a cue may be influenced by how consistent a given cue is with what is conveyed by other cues in terms of estimating the representation. Reweighting of cues occurs when certain cues have low cue reliability. In this study, the reliability of the face cue regarding the speaker's accent was manipulated such that a speaker was presented with either a reliable, henceforth predictable, accent (one accent: American or Chinese) or an unreliable, henceforth unpredictable, accent (two accents: American and Chinese). The acoustic signal (which contains the accent information) serves as the acoustic cue listeners are processing. The accent (s) paired with each speaker in the experiment were consistent with what listeners learned about the speakers in the speaker introduction videos. Although the cue integration model does not specify what type of neural or electrophysiological signal (oscillatory activity, ERPs) would encode cue reliability, different cue reliability during language processing is predicted to elicit different ERP signals.

1.2. Face cue in speech perception studies

As previously mentioned, speech perception typically involves more than just the speech signal. Previous accented-speech studies have used photos as a visual cue to depict the social identity of the speaker (e.g., Grey, Cosgrove, & Van Hell, 2020; Hernández-Gutiérrez et al., 2021; Kang & Rubin, 2009; McGowan, 2015; Sala, Vespignani, Casalino, & Peressotti, 2024; Xu, Abdel Rahman, & Sommer, 2022). In Kang & Rubin (2009), although the acoustic–phonetic speech signal was from the same native speaker in both the White face and Asian face conditions, the visual cue of an "incongruent" Asian face speaker negatively impacted the processing of the speech signal. For nonnative-accented speech, McGowan (2015) found that transcription accuracy was higher when Chinese-accented speech was paired with an Asian face than with a White face. This result suggests that nonnative-accented speech paired with a "congruent" visual cue may *facilitate* comprehension. These studies indicate that listeners generate predictions based on visual cues when processing speech.

The facilitation observed in McGowan (2015) was also observed for "congruent" face-accent pairings (native-accented speech produced by a White speaker, Chinese-accented speech produced by Asian speaker) in studies that used audiovisual (video) cues (e.g., McLaughlin, Brown, Carraturo, & Van Engen, 2022; Yi, Phelps, Smiljanic, & Chandrasekaran, 2013). In particular, Yi, Phelps, Smiljanic, & Chandrasekaran (2013) found that participants recognized words in noise better for native-accented speakers than for nonnative-accented speakers, and they performed better in the audiovisual than in the audio-only condition. Crucially, they found a larger increase for audiovisual compared to audio-only for native-accented speakers than for nonnative-accented speakers (also replicated by McLaughlin, Brown, Carraturo, & Van Engen, 2022). In addition to studies that have examined the role of face-cues in accented-speech processing, face (race) cues have been used as a cue in studies on monolingual or bilingual language comprehension and production (e.g., Babel & Russell, 2015; Li, Yang, Suzanne Scherf, & Li, 2013; Ma, Ai, Xiao, Guo, & Pae, 2020; C. D. Martin, Molnar, & Carreiras, 2016; Xu, Abdel Rahman, & Sommer, 2022).

Most relevant to the present study is an ERP study by C. D. Martin, Molnar, & Carreiras (2016) that examined the neural basis of speaker identity predictions during bilingual speech processing using an audiovisual lexical decision task. In the audiovisual lexical decision task, participants decide whether the audio produced by the speaker in the video is a word or not (a lexical decision) via a button press response. As in our study, they examined the lexicality effect in the N400 time window; the N400 is a robust ERP component associated with lexical-semantic access (e.g., Federmeier, 2022; Kutas & Federmeier, 2011). Prior to discussing the C. D. Martin, Molnar, & Carreiras (2016) study, we will explain the lexicality effect. The lexical decision task is a common task used to examine single word processing relative to pseudowords, in both the visual or auditory domains (e.g., Bentin, McCarthy, & Wood, 1985; Holcomb & Neville, 1990; Winsler, Midgley, Grainger, & Holcomb, 2018). A well-established finding is that pseudowords elicit a larger (more negative-going) N400 amplitude relative to words, termed the N400 lexicality effect, indicating increased processing demands for accessing the meaning of pseudowords relative to words (e.g., Bentin, McCarthy, & Wood, 1985; Lei, Liu, & Van Hell, 2022a; for a review, see Kutas & Federmeier, 2011). Words are thought to be processed more efficiently compared to pseudowords because only words are represented in the mental lexicon (i.e., only words are lexicalized in the brain and their forms and meanings can be directly accessed; for reviews, see McNorgan, Chabal, O'Young, Lukic, & Booth, 2015; Woollams, 2015).

C. D. Martin, Molnar, & Carreiras (2016) examined how faces cueing the speakers' language background affect language processing in Spanish-Basque bilingual listeners. After participants were familiarized with the speakers' language backgrounds (Basque only, Spanish only, or Spanish and Basque bilingual), they completed an EEG-recorded Basque and Spanish audiovisual lexical decision task. The authors found amplitude differences between the monolingual speakers (Basque or Spanish) versus the bilingual Spanish-Basque speakers in the pre-speech time period. This pre-speech amplitude difference was thought to stem from the predictability of the speaker's language (predictable in monolingual but unpredictable in bilingual speakers). Speaker identity (monolingual or bilingual) also modulated the online processing of speech in the lexical decision task. Post-speech results showed a N1 lexicality effect for the monolingual speakers but not for the bilingual speakers. The N1 is an early ERP component typically associated with processing acoustic characteristics of the speech signal (Helenius, Salmelin, Richardson, Leinonen, & Lyytinen, 2002; Kapnoula & McMurray, 2021; for a review, see Getz & Toscano, 2021) and facilitating auditory lexical access (MacGregor, Pulvermüller, van Casteren, & Shtyrov, 2012). N400 lexical effects (in the 500–700 ms time window) were

found for both monolingual and bilingual speakers but were larger for monolingual speakers. ERP mean amplitudes in the 100–200 ms pre-speech time window were correlated with the magnitude of the N400 lexicality effect. Behaviorally, participants responded slower to bilingual speakers than to monolingual speakers, and they responded faster for words than nonwords (lexicality effect). For both speaker conditions, ERP mean amplitudes in the 100–200 ms pre-speech time window were also correlated with the magnitude of the reaction time lexicality effect. Overall, these results suggest that listeners predicted the upcoming language based on speaker identity, and this language prediction impacted later speech processing. These results show that listeners are incorporating top-down information regarding the potential *language* a speaker may produce. With regards to accented speech, are listeners also predicting the upcoming *accent* based on speaker identity? Are listeners integrating face cue information regarding accent(s) and auditory cues during speech processing? To our knowledge, these questions have not been examined in accented speech processing using ERPs.

1.3. Present study

The present study investigated the processing of predictable and unpredictable face cues during native- and nonnative-accented speech processing, with Experiment 1 (control experiment) conducted first to determine the neurocognitive baseline outcomes of the auditory processing of the speech stimuli, and Experiment 2 (main experiment) conducted next to examine the processing of face cues and accented speech. Experiment 1 participants completed an auditory-only go/no-go animal decision task in response to audio-only stimuli used in the Experiment 2 (more information below), but without the familiarization phase and face cues. In Experiment 2, to understand the role of faces cueing different accents and number of accents on the processing of native-accented and nonnative-accented speech, we manipulated the predictability of a speaker in cueing the upcoming accented speech (predictable: one possible accent, unpredictable: two accents). To establish whether listeners processed face cues according to our manipulation, we first asked: 1) Are listeners processing face cues differently based on the predictability of the speaker's accent? If so, how? Our main question was: 2) How does a visual face cue (speaker identity) with differing predictability of accent influence the online neurocognitive processing of words produced in one's (a) native accent and (b) nonnative accent? Experiment 2 participants learned what type (s) of accent(s) was/were associated with each speaker through a familiarization phase. Then, they completed an EEG-recorded face cued go/no-go animal decision task with American-accented and Chinese-accented English words and nonwords. We hypothesized that listeners will process face cues differently, reflecting the predictability of the face cue, and that face cues will impact the subsequent processing of the native- and nonnative-accented speech.

We modeled our study design after C. D. Martin, Molnar, & Carreiras (2016). Instead of the two-choice (yes/no) lexical decision task used in C. D. Martin, Molnar, & Carreiras (2016), we used an animal go/no-go procedure in which participants were asked to press a button when they heard an animal word ("go" trials) to eliminate any button response or response preparation on the critical experimental items (non-animal words and nonwords). Also, previous studies have shown that the go/no-go variant of the lexical decision task has been typically associated with lower response time variability, faster RT, and/or fewer error rates than the yes/no variant of the lexical decision task (Gomez, Ratcliff, & Perea, 2007; Gordon, 1983; Lee, Lee, Tae, & Kwon, 2019; Vergara-Martínez, Gomez, & Perea, 2020). Assuming this behavioral finding can be extended to other button press decision tasks, we decided to use the go/no-go procedure instead of a yes/no procedure for more reliable data. Since we used an animal go/no-go task, we anticipated a reduced lexicality effect in the present experiments compared to experiments that used the explicit yes/no version (e.g., O'Rourke & Holcomb, 2002).

2. Experiment 1: Audio only

In Experiment 1, we used a 2 Accent (American, Chinese) × 2 Lexicality (word, nonword) study design and examined the ERP correlates associated with the processing of American-accented and Chinese-accented spoken words and nonwords.

2.1. Predictions

A main effect of lexicality was expected for both accents such that the ERP amplitude in the N400 time window for nonwords will be significantly less negative than words; the lexicality effect was anticipated to be smaller than what is typically observed in the yes/no version. In line with the cue integration model that predicts neural/electrophysiological signal to encode cue reliability, an Accent by Lexicality interaction was also predicted such that the lexicality effect for the American-accent (native) condition will be greater than for the Chinese-accent (nonnative) condition, reflecting the predicted larger weighting of the American-accented acoustic cue in cueing the word representation.

2.2. Methods

2.2.1. Participants

Thirty-five participants were tested after providing informed consent. This experimental protocol was approved by the Institutional Review Board at The Pennsylvania State University (IRB Number: STUDY00020381). Six participants were excluded from data analysis: two for technical issues, one because of self-reported exposure to Chinese (both Mandarin and Cantonese) at home from a parent, and three had excessive EEG artifact such as eye blinks and/or muscle movement (exclusion criteria: fewer than 60 trials (< 50%) in any condition). The final sample of participants consisted of 29 participants (sex: 6 males, 23 females, 0 non-binary; *M* age: 18.72 years, range = 18–21); this sample size is comparable to other auditory word processing EEG studies (e.g. C. D. Martin, Molnar, & Carreiras, 2016). Participants were recruited from a public university in Pennsylvania, United States, via the Psychology Department's subject pool, and they received course credit for participating. Participants were right handed (Oldfield, 1971), and reported no auditory, language, neurological impairments, and normal or corrected-to-normal vision.

All participants were native speakers of American English with minimal knowledge of other languages. Participants completed a language history questionnaire (Lei, Liu, & Van Hell, 2022b), and if applicable, reported their second language and third language proficiency on a 10-point scale (1 = *very low*, 10 = *perfect*) for speaking, understanding, reading, and writing. Eleven participants reported limited knowledge of a second language (self-reported L2 proficiency: speaking: $M = 1.91$, $SD = 1.58$, understanding: $M = 2.55$, $SD = 2.16$; reading: $M = 2.64$, $SD = 2.42$; writing $M = 1.91$, $SD = 1.58$). Three participants reported limited knowledge of a third language (self-reported L3 proficiency: speaking: $M = 1.00$, $SD = 0$; understanding: $M = 1.33$, $SD = 0.58$; reading: $M = 1.33$, $SD = 0.58$; writing $M = 1.00$, $SD = 0$). Participants reported high daily exposure (scale = 0–100%) to English ($M = 98.31\%$, $SD = 3.25\%$) and minimal daily exposure to other languages ($M = 1.69\%$, $SD = 3.25\%$), and did not grow up exposed to Chinese-accented English at home.

2.2.2. Materials

2.2.2.1. Stimuli. There were 560 stimuli: 240 English words, 240 English nonwords, 80 animal words. The English words were selected from the English Lexicon Project (Balota et al., 2007), and English nonwords were pseudowords selected from the Auditory English Lexicon Project (Goh, Yap, & Chee, 2020). Eighty animal words were selected as the filler "go" items. All stimuli were multisyllabic and contained one

morpheme. Crucially, the word and nonword stimuli contained at least one phoneme (e.g., /θ/-/s/, /ɛ/-/æ/) that typically elicits marked Mandarin-accented English (Flege, 1995; Rogers & Dalby, 2005). See the Supplementary Materials for the stimuli list, average duration of each word type and accent for each speaker, stimuli selection process, and norming studies conducted to validate the stimulus materials. Four pseudo-random lists were created. First, stimuli type (i.e., word, nonword, animal word) were evenly and randomly split into eight groups and assigned to one of the possible speaker-by-accent combinations across the four speakers. The speaker-by-accents assignments were rotated Latin-square wise, so each stimulus was spoken by each possible speaker-by-accent combination across the four lists. In each list, the four predictability and accent conditions (predictable American-accented items, predictable Chinese-accented items, unpredictable American-accented items, and unpredictable Chinese-accented items) were evenly presented (25% for each condition). Participants were presented with an equal number of trials for each speaker and accent, and the trials for the two-accent speakers were split evenly between American and Chinese accents.

2.2.2.2. Speakers. The final set of speakers included four female graduate students: one American-accented English speaker (with a stereotypical White face), one Chinese-accented English speaker (with a stereotypical Asian face), and two speakers who produced American-accented and Chinese-accented English (with stereotypical Asian faces). Accent rating norming studies were conducted to ensure the two-accent speakers' Chinese-accented English was perceived as more accented than their American-accented speech. See Supplementary Materials for details on speaker characteristics (native language, age, strength of accent(s)) and the speaker selection process.

2.2.2.3. Speaker recording setup. Each speaker recorded all 560 experimental stimuli). Speakers were asked to record each item three times. Recordings were made in a sound attenuated booth or room with a Zoom H4n Pro recorder, an external dynamic microphone, and a pop filter. Recorder settings were adjusted to eliminate background noise (60–75 level, 85–98 Hz low pass filter) for each recording session. We did not prevent speakers from displaying idiosyncratic features (e.g., creaky voice) to preserve the speaker's natural speech.

For the nonwords, speakers were provided with the word each nonword was generated from, the nonword in English alphabet and IPA format, and the nonword's number of syllables. Speakers had the option to listen to the nonword produced by a female American speaker from the AELP database. The two-accent speakers recorded one accent per recording session, with the exception of the final re-recording session.

2.2.2.4. Audio processing. The first author reviewed all three repetitions of an item and selected the best recording. If none of the three repetitions proved satisfactory (e.g., due to mispronunciation, excessive noise, or weak audio signal), speakers were asked to re-record those items. Audio clips were annotated, segmented, padded with a 50 ms silence before and after the audio clip, and normalized to 70 dB in Praat (Boersma & Weenink, 2022). A total of 3,360 audio clips were created.

2.2.3. Procedure

Participants were tested in a single session lasting approximately two hours. First, participants completed a background and language history questionnaire. Then, they completed an EEG recorded go/no-go animal decision task. Participants wore plastic earphones (Etymotic earphones, model ER4S; Elk Grove Village, IL) and the volume was adjusted to a comfortable hearing level. Participants listened to the audio stimulus presented bi-aurally over earphones and were instructed to press a button on a button box when they heard an animal word, and to not press any button for other items. In each trial, a black screen preceded the audio presentation screen. During the audio presentation screen, a visual fixation cross was concurrently presented at the center of the computer screen. Then, a visual 'blink' stimulus replaced the fixation cross for one second, to let the participants blink and rest their eyes. Response times were measured from audio onset. Each participant heard all stimuli, which consisted of 50% native-accented speech and 50% nonnative-accented speech, 50% predictable and 50% unpredictable. Each of the four speakers were presented the same number of times (140 trials each, 560 trials total). After completing the go/no-go task, participants completed a post-experiment survey which included questions that asked participants to describe each speaker's accent and to rate the accent of each speaker. Then, participants completed a language attitude survey, four auditory discrimination tasks assessing sensitivity to duration, formant, pitch, and rise time (from https://www.sla-speech-tools.com), and the English LexTALE task (Lemhöfer & Broersma, 2012).

2.2.4. EEG recording and preprocessing

Participants were seated comfortably in a chair in a sound-attenuated dimly lit room, approximately three feet away from the computer monitor displaying the stimuli. An elastic cap (Brain Products ActiCap, Germany) with a total of 31 active electrodes was placed on the participant's head: five electrodes located along the midline (Fz, FCz, Cz, Pz, Oz) and 26 electrodes on the lateral sites (FP1/2, F7/8, F3/4, FC5/6, FC1/2, T7/8, C3/4, CP5/6, CP1/2, P7/8, P3/4, O1/2, PO9/10). Four electrodes were placed above and below the left eye, and on the canthus of both eyes, to monitor eye movements and blinking. Electrode impedances were kept below 10 kΩ. Electrodes were referenced to a vertex reference (electrode FCz). EEG signals were amplified by a NeuroScan SynampsRT amplifier using a 0.05–100 Hz bandpass filter (first-order Butterworth with a 6 dB/octave roll-off), digitized and continuously sampled at a rate of 500 Hz.

EEG data was preprocessed using the EEGLAB v2024.0 toolbox (Delorme & Makeig, 2004) and ERPLab 10.0.0 (Lopez-Calderon & Luck, 2014). Recorded EEG signals were re-referenced offline to an average of the left and right mastoids (M1/M2). ERPLab was used for data cleaning and processing by applying an offline 30 Hz low-pass filter. EEG data was visually inspected for any flat or extremely noisy channel signal. Artifact rejection thresholds were individually calibrated to each participant: ERPLAB's moving window peak-to-peak artifact detection function was applied to the vertical eye electrode to detect blinks, ERPLAB's step-like artifact detection function was applied to the horizontal eye electrode to detect horizontal eye movements, and ERPLAB's moving window peak-to-peak artifact detection function was applied to all other electrodes to detect channel drift or large amplitude activity. Trials containing artifacts were excluded from further data analysis. ERP data from 'go' trials (animal names) were not analyzed. Participants with less than 60 trials (< 50%) in one of the experimental conditions (American-accented words, American-accented nonwords, Chinese-accented words, Chinese-accented nonwords) were excluded from further ERP analyses ($n = 3$, see section 2.1.1.). For each participant, ERPs from 100 ms prior to stimulus onset to 700 ms after stimulus onset were separately averaged offline in different experimental conditions at each electrode site, by participant, then across all participants. Following C. D. Martin, Molnar, & Carreiras (2016), the mean amplitudes were measured by averaging across the central electrodes: Cz, F3, F4, FC1, FC2, C3, C4, CP1, CP2, P3, Pz, and P4.

2.2.5. Analysis

All analyses were conducted in R (The R Foundation for Statistical Computing, Vienna, Austria, version 4.0). For the animal decision task, separate ANOVAs were conducted on the accuracy and reaction time (RT) with Accent (American, Chinese) as a factor. Accuracy was calculated by dividing the number of correctly identified animal words by total animal items per condition (40). RT (in ms) was measured from the onset of the audio stimulus.

Following C. D. Martin, Molnar, & Carreiras (2016), ERP analyses were conducted in the 500–700 ms time window for the N400

component (for similar time window analysis, see Grey, Cosgrove, & Van Hell, 2020).[1] Repeated measures ANOVA with the variables accent (American, Chinese) and lexicality (word, nonword) involving the central region channels (see above) were conducted. Greenhouse-Geisser correction was applied for violation of sphericity to all F tests with more than one degree of freedom in the numerator.

2.3. Results

2.3.1. Behavioral results

The ANOVA for accuracy revealed a main effect of Accent, $F(1, 28) = 22.26$, $p < 0.001$, $\eta_p^2 = .44$, such that participants were more accurate for the American-accented words ($M = 0.81$, $SD = 0.09$, 95% CI [0.78, 0.84]) than for the Chinese-accented words ($M = 0.72$, $SD = 0.13$, 95% CI [0.68, 0.76]).

The ANOVA on RT showed a main effect of Accent, $F(1, 28) = 279.31$, $p < 0.001$, $\eta_p^2 = .91$, such that participants were faster in responding in the American-accented condition ($M = 1034$ ms, $SD = 110$, 95% CI [1000, 1067]) than in the Chinese-accented condition ($M = 1177$ ms, $SD = 90$, 95% CI [1150, 1205]).

2.3.2. ERP results

The ANOVA in the 500–700 ms ERP time-window indicated no significant main effect of Accent, $F(1, 28) = 3.53$, $p = .07$, $\eta_p^2 = .04$, no main effect of Lexicality, $F(1, 28) = 0.83$, $p = .81$, $\eta_p^2 = .04$, and no significant Accent by Lexicality interaction, $F(1, 28) = 0.99$, $p = .58$, $\eta_p^2 = .13$. Grand mean ERP waveforms of the four conditions are presented in Fig. 1.

2.4. Discussion

In Experiment 1, ERP mean amplitude analyses indicated no lexicality nor accent effects in the N400 time window. Participants did not demonstrate the expected lexicality effect, and online speech processing of single words measured by ERPs was not modulated by accent. Behaviorally, participants were more accurate and faster in their responses to American-accented items compared to Chinese-accented items.

As we wanted to prevent response preparation associated with button presses that would interfere with processing the critical word and nonword items, we used an animal decision go/no-go task. We anticipated that this cover task (press a button when you hear an animal word) may reduce the N400 lexicality effect, but we in fact observed no significant lexicality effect. Still, this outcome can serve as a baseline measure against which the effect of adding a face cue can be interpreted. Furthermore, although the N400 effect of accent failed to reach significance, the effect of accent was significant in the RT and accuracy analyses of the go-items (animal words), indicating that participants did show sensitivity to the accent manipulation in their behavioral animal word identification responses.

[1] In response to a reviewer and based on visual inspection, we also analyzed the 300 – 700 ms time window and found the same pattern of results as reported for the 500 – 700 ms in both experiments. In Experiment 1, there were no significant effects nor interactions. For Experiment 2, there were significant interactions between accent and lexicality, and between accent and predictability. Note that although the posthoc analyses for the accent by lexicality interaction indicated marginal significance with a Bonferroni correction ($p = 0.08$) for American-accented items, the Chinese-accented items were not significant. This suggests that the significant accent by lexicality interaction was driven by the American-accented items. As the ERP results pattern in the 300 – 700 ms time window was similar to that in the 500 – 700 ms time window, we decided to report the ERP results from the 500 – 700 ms time window, to follow C. D. Martin, Molnar, & Carreiras (2016) more closely.

3. Experiment 2: face-cued audio

In Experiment 2, to examine whether listeners tracked speakers, face cue (speaker identity) predictability was manipulated to be either predictable or unpredictable of the upcoming speech accent. We examined the integration of the face cue and accented speech cue with a 2 Predictability (Predictable, Not Predictable) by 2 Accent (American-accented English, Chinese-accented English) by 2 Lexicality (Word, Nonword) study.

3.1. Predictions

The face cue related with accent predictability was expected to hold different weights during speech processing. Prior to the input of any speech, the cue integration model would hypothesize that the listener will begin integrating the visual cue information for the upcoming speech, and brain activity would reflect the differential processing of the face cue based on accent predictability. So, pre-speech analyses in the 100–200 ms and 200–300 ms time windows are predicted to show a main effect of Predictability (predictable, unpredictable) in which the predictable conditions will elicit a more negative-going mean amplitude compared to the unpredictable conditions.

The visual face cue was expected to influence the processing of the auditory word stimulus and was predicted to modulate ERP lexicality effects according to the face cue predictability. For the post-speech analyses, the N1 lexicality effect, reflecting facilitated lexical access, was predicted to be elicited in the predictable condition (one possible accent), but not in the unpredictable condition (two possible accents). The N400 lexicality effect was predicted to be present in all conditions but be modulated by predictability and accent. Specifically, the unpredictable condition was predicted to elicit a reduced N400 lexicality effect relative to the predictable condition, and nonnative-accented speech to elicit a reduced N400 lexicality effect relative to native-accented speech.

3.2. Methods

3.2.1. Participants

Thirty-five new participants from the same population as in Experiment 1 (undergraduate students in the same university) were tested after providing informed consent. The experimental protocol was approved by the same Institutional Review Board (IRB Number: STUDY00020381). Six participants were excluded from data analysis: four because of technical issues, two had excessive EEG artifact such as eye blinks and/or muscle movement (exclusion criteria: fewer than 30 trials (< 50%) in any post-speech condition or fewer than 60 trials (< 50%) in any pre-speech condition). The final sample of participants consisted of 29 participants (sex: 12 males, 17 females, 0 non-binary; M age: 19.62 years, range = 18–25). Participants received either course credits or payment for their participation. Participants were right handed (Oldfield, 1971), self-reported no auditory, language, neurological impairments, and normal or corrected-to-normal vision.

All participants were native speakers of American English with minimal knowledge of other languages. Nine participants reported limited knowledge of a second language (self-reported L2 proficiency: speaking: $M = 1.78$, $SD = 1.64$, understanding: $M = 2.67$, $SD = 1.22$; reading: $M = 3.11$, $SD = 2.03$; writing $M = 2.56$, $SD = 1.5$. One participant reported limited knowledge of a third language (self-reported L3 proficiency: speaking = 2; understanding = 4; reading = 1; writing = 1). Participants reported high daily exposure to English ($M = 96.52\%$, $SD = 97\%$; scale = 0–100%) and minimal daily exposure to other languages ($M = 3.48\%$, $SD = 5.97\%$; scale = 0–100%), and did not grow up exposed to Chinese-accented English at home.

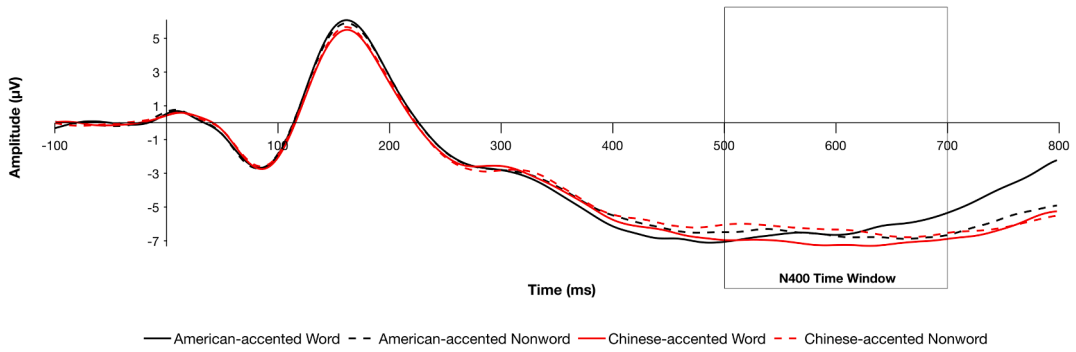


Fig. 1. ERPs time-locked to speech onset. Black waveforms depict ERPs measured for American-accented stimuli, and red waveforms depict ERPs measured for Chinese-accented stimuli (solid: words, dotted: nonwords). A15 Hz low-pass filter was applied to the waveforms for presentation purposes.

3.2.2. Materials

3.2.2.1. Auditory stimuli. The same set of 560 audio stimuli (240 words, 240 nonwords, 80 animal words) per speaker and accent (3,360 total) from Experiment 1 were used.

3.2.2.2. Introductory video. Each speaker recorded a 3.5-minute introductory video for the familiarization phase (see the Supplementary Materials for the transcriptions). Speakers used a pseudonym and talked about their language background guise, their hobbies, interesting stories, and fun facts. Speakers were seated against a white background. As nonnative-accented speakers tend to speak slower than native-accented speakers (e.g., Grey & Van Hell, 2017; Hanulíková, Van Alphen, Van Goch, & Weber, 2012; Romero-Rivas, Martin, & Costa, 2015), we asked the speakers producing American-accented speech to slow down their speech rate so that their video duration matched the Chinese-accented speakers video duration. The introductions ranged from 507 to 619 words. Videos were trimmed and edited, and a noise reduction feature was applied to enhance the audio. Twenty True/False comprehension questions (five per speaker) were created to ensure participants paid attention to the video content. Participants' overall comprehension accuracy was high, $M = 85.69\%$, $SD = 7.76\%$.

3.2.2.3. Photos. Each speaker had a headshot photo taken during their video recording session. Photos were cropped to 400 by 400 pixels in dimension. The photos were presented during the EEG task as face cues and in the debriefing survey.

3.2.3. Procedure

Participants were tested in a single session lasting approximately 2.5 h. First, participants completed a background and language history questionnaire (Lei, Liu, & Van Hell, 2022b). Then, they completed the familiarization phase in which introduction videos of the four speakers were randomly presented. Participants had the option to adjust the volume as needed. Following each video, participants answered the comprehension questions. Next, participants completed the EEG-recorded go/no-go animal decision task. The behavioral go/no-go task was identical to Experiment 1, but for each trial, the speaker's image was presented at the center of the screen for 350 ms prior to and concurrently throughout the whole audio stimulus presentation. Then, a visual 'blink' stimulus appeared on the screen for one second, to let the participants blink and rest their eyes. Response times were measured from audio onset. As in Experiment 1, participants then completed a post-experiment survey, a language attitude survey, the four auditory discrimination tasks assessing sensitivity to duration, formant, pitch, and rise time (from https://www.sla-speech-tools.com), and the English LexTALE task (Lemhöfer & Broersma, 2012).

3.2.4. EEG recording and preprocessing

EEG recording and preprocessing procedures were the same as in Experiment 1, with the exception that the pre-speech time period was also preprocessed. Pre-speech ERP data were time-locked to the image onset. ERPs from 100 ms prior to stimulus onset to 400 ms after stimulus onset were separately averaged offline in different experimental conditions at each electrode site, by participant, then across all participants. For pre-speech data, participants with less than 60 trials (< 50%) in one of the image conditions (e.g., Speaker 1, Speaker 2, Speaker 3, Speaker 4) were excluded from further analyses. Post-speech ERPs were time-locked to speech onset. ERPs from 100 ms prior to stimulus onset to 700 ms after stimulus onset were separately averaged offline by experimental conditions at each electrode site, by participant, then across all participants. For post-speech data, participants with less than 30 trials (< 50%) in one of the experimental conditions (e.g., Predictable American-accented Words) were excluded from further ERP analyses ($n = 2$; see Participants section). Following C. D. Martin, Molnar, & Carreiras (2016), mean amplitudes in the pre-speech and post-speech time windows were averaged across the following electrodes for analyses: Cz, F3, F4, FC1, FC2, C3, C4, CP1, CP2, P3, Pz, and P4.

3.3. Analysis

All analyses were conducted in R (The R Foundation for Statistical Computing, Vienna, Austria, version 4.0). For the animal decision task, separate ANOVAs were conducted on the accuracy and RT data with factors Predictability (Predictable, Unpredictable) and Accent (American, Chinese). Accuracy (on a 0–1 scale) was calculated by dividing the number of correctly identified animal words by total animal items per condition (20). RT (in ms) was measured from the onset of the audio stimulus.

ERP analyses followed C. D. Martin, Molnar, & Carreiras (2016) – pre-speech ERP analyses were conducted in the 0–100 ms, 100–200 ms, and 200–300 ms time windows, and post-speech ERP analyses were conducted in the 100–200 ms for the N1 component and in the 500–700 ms time window for the N400 component. For each pre-speech time window, a repeated measures ANOVA with the variable Predictability (Predictable, Unpredictable) involving the mean amplitude across the central region channels were conducted. For each post-speech time window, a repeated measures ANOVA with the variables Predictability (Predictable, Unpredictable), Accent (American, Chinese), and Lexicality (word, nonword) involving the central region channels were conducted. Greenhouse-Geisser correction was applied for violation of sphericity to all F tests with more than one degree of freedom in the numerator. Follow-up ANOVAs were conducted on significant ($p < 0.05$) interactions involving the factor of Predictability or Accent revealed in the overall ANOVAs.

3.4. Results

3.4.1. Behavioral results

The ANOVA for accuracy revealed a main effect of Accent, $F(1, 28) = 5.82$, $p = 0.01$, $\eta_p^2 = .20$, such that, as in Experiment 1, participants were more accurate for American-accented items ($M = 0.68$, $SD = 0.14$, 95% CI [0.64, 0.72]) than for Chinese-accented items ($M = 0.64$, $SD = 0.14$, 95% CI [0.60, 0.68]). The main effect of Predictability and the interaction of Accent by Predictability were both not significant, $F(1, 28) = 0.74$, $p = .40$, $\eta_p^2 = .03$ and $F(1, 28) = 2.45$, $p = 0.14$, $\eta_p^2 = .08$, respectively.

The ANOVA on RT showed a main effect of Accent, $F(1, 28) = 149.67$, $p < .001$, $\eta_p^2 = .84$, such that participants were faster in responding in the American-accented condition ($M = 972$ ms, $SD = 160$, 95% CI [938, 1007]) than in the Chinese-accented condition ($M = 1128$ ms, $SD = 149$, 95% CI [1095, 1160]), again in line with the findings in Experiment 1. There was no main effect of Predictability, $F(1, 28) = 0.09$, $p = .76$, $\eta_p^2 = .003$, and no significant Predictability by Accent interaction, $F(1, 28) = 4.10$, $p = .05$, $\eta_p^2 = .13$.

3.4.2. ERP results

3.4.2.1. Pre-speech time windows. The ANOVA in the 0–100 ms time window showed no main effect of Predictability, $F(1, 28) = 2.42$, $p = 0.13$, $\eta_p^2 = .08$ (predictable: M μV $= -0.03$, $SD = 0.83$, 95% CI [−0.25, 0.19]; unpredictable: M μV $= 0.13$, $SD = 0.59$, 95% CI [−0.10, 0.35]).

In the 100–200 ms time window, there was also no main effect of Predictability, $F(1, 28) = 3.06$, $p = 0.09$, $\eta_p^2 = .10$ (predictable: M μV $= -1.57$, $SD = 2.33$, 95% CI [−1.74, −1.40]; unpredictable: M μV $= -1.83$, $SD = 2.40$, 95% CI [−2.00, −1.65]).

The ANOVA in the 200–300 ms time window revealed a main effect of Predictability, $F(1, 28) = 18.97$, $p < 0.001$, $\eta_p^2 = .40$, such that predictable images elicited a more negative grand average amplitude, M μV $= -2.56$, $SD = 3.28$, 95% CI [−2.80, −2.31], compared to unpredictable images, M μV $= -1.69$, $SD = 3.17$, 95% CI [−1.93, −1.46]. See Fig. 2 for the plotted ERPs.

3.4.2.2. N1 time window. The omnibus ANOVA in the 100–200 ms time-window indicated no main effect of Predictability, $F(1, 28) = 0.17$, $p = 0.68$, $\eta_p^2 = .01$, no main effect of Accent, $F(1, 28) = 0.00$, $p = 0.97$, $\eta_p^2 < .0001$, and no main effect of Lexicality, $F(1, 28) = 1.17$, $p = 0.23$, $\eta_p^2 = .05$. All interactions were also not significant (all $ps > 0.05$). See Fig. 3.

3.4.2.3. N400 time window. The ANOVA in the 500–700 ms time-window revealed a significant main effect of Accent, $F(1, 28) = 30.21$, $p < 0.001$, $\eta_p^2 = .52$, a main effect of Lexicality, $F(1, 28) = 11.71$, $p = 0.002$, $\eta_p^2 = .29$, a significant Predictability by Accent interaction, $F(1, 28) = 4.88$, $p = 0.04$, $\eta_p^2 = .15$, and an Accent by Lexicality interaction, $F(1, 28) = 13.17$, $p = 0.001$, $\eta_p^2 = .32$. Grand mean ERP waveforms for each condition are presented in Fig. 3.

Posthoc analysis of the Accent by Lexicality interaction showed a significant effect of Lexicality only in the American-accented condition, $F(1, 28) = 17.09$, $p < 0.001$, $\eta_p^2 = .38$, in which the nonwords elicited more negative ERPs than words (nonwords M μV $= -2.24$, $SD = 2.53$, 95% CI [−2.91, −1.58]; words M μV $= -0.75$, $SD = 3.53$, 95% CI [−1.68, 0.18]), reflecting a N400 lexicality effect. There was no significant effect of lexicality in the Chinese accented condition, $F(1, 28) = 0.7$, $p = 0.78$, $\eta_p^2 = .03$ (nonwords M μV $= -2.94$, $SD = 2.90$, 95% CI [−3.70, −2.17]; words M μV $= -2.73$, $SD = 2.51$, 95% CI [−3.39, −2.07]).

Analysis of the Predictability by Accent interaction indicated a significant effect of Predictability for the Chinese-accented condition, $F(1, 28) = 7.52$, $p = .02$, $\eta_p^2 = .21$, such that the predictable condition (M μV $= -3.21$, $SD = 2.57$, 95% CI [−3.88, −2.53]) elicited a more negative N400 amplitude than the unpredictable condition (M μV $= -2.46$, $SD = 2.80$, 95% CI [−3.20, −1.72]). There was no significant effect of Predictability for the American-accented condition, $F(1, 28) = 0.28$, $p = 1.00$, $\eta_p^2 = .01$ (predictable: M μV $= -1.42$, $SD = 3.33$, 95% CI [−2.29, −0.54]; unpredictable: M μV $= -1.57$, $SD = 2.99$, 95% CI [−2.36, −0.79]).

3.4.3. Correlation analyses

C. D. Martin, Molnar, & Carreiras, 2016 found that pre-speech ERP amplitudes correlated with the magnitude of the N400 lexicality effect and the behavioral lexicality effects. So, to examine whether the pre-speech ERP amplitudes were correlated with the post-speech ERP data and behavioral data, the pre-speech and post-speech ERP mean amplitudes with either a significant predictability, lexicality, or accent effect were correlated with the behavioral accuracy data of each accent condition, the magnitude of the accent effect (American – Chinese) in the RT data, the four auditory threshold scores (duration, formant, pitch, rise time), and the LexTale scores. The Pearson correlation matrix revealed a significant correlation between (1) the unpredictable condition's pre-speech mean ERP amplitude and the LexTale scores, $r = 0.39$, $p = 0.04$, (2) the predictable condition's pre-speech 200 – 300 ms mean ERP amplitude and the LexTale scores, $r = 0.39$, $p = 0.04$, (3) the accuracy of American-accented animal words and the accuracy of Chinese-accented animal words, $r = 0.77$, $p < 0.001$, (4) the pitch and duration thresholds, $r = 0.49$, $p = .01$, and (5) pitch and rise time thresholds, $r = 0.52$, $p =$

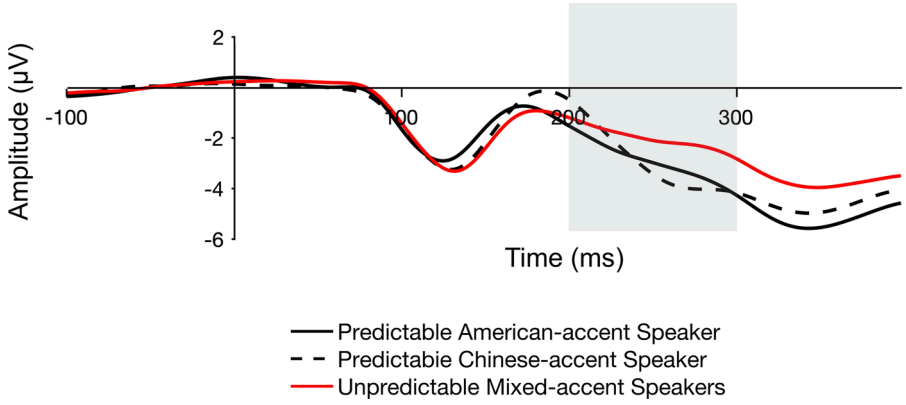


Fig. 2. ERPs time-locked to speaker image onset. Black waveforms depict ERPs measured for predictable speakers (solid: American-accent, dotted: Chinese-accent). The red waveform depicts ERPs measured for unpredictable speakers. A 15 Hz low-pass filter was applied to the waveforms for presentation purposes.

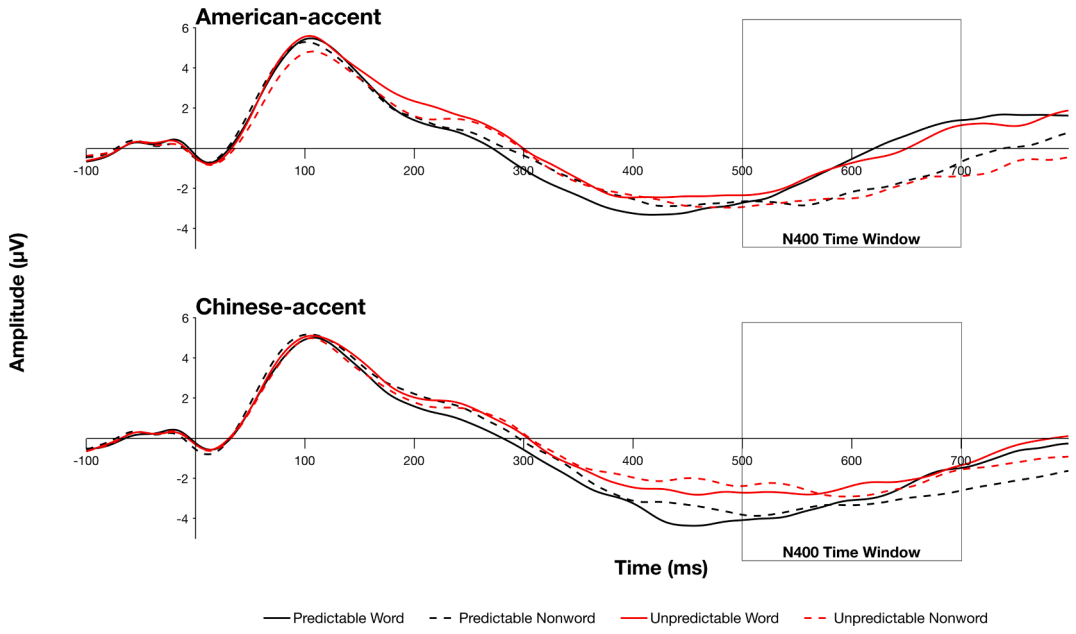


Fig. 3. ERPs time-locked to speech onset for American-accent conditions (top plot) and Chinese-accent conditions (bottom plot). Black waveforms depict ERPs measured in the predictable conditions, and red waveforms depict ERPs measured in the unpredictable conditions (solid: word, dashed: nonword). A 15 Hz low-pass filter was applied to the waveforms for presentation purposes.

.004. There were no other significant correlations between the ERP data and the other measures (all *ps* > 0.05 or higher).

3.4.4. *Cross-experiments analysis*

To confirm whether the lexicality effects in the N400 time window significantly differed between the audio-only (Experiment 1) and the face-cued (Experiment 2) groups, an ANOVA with a between-group factor Experiment (Audio-only, Face-cued), and two within-group factors Accent (American, Chinese) and Lexicality (word, nonword) was conducted.

The ANOVA indicated a significant main effect of Experiment, $F(1, 56) = 36.99, p < 0.001, \eta_p^2 = .40$, a significant main effect of Accent, $F(1, 56) = 24.77, p < 0.001, \eta_p^2 = .31$, and significant interactions of Experiment by Accent, $F(1, 56) = 12.55, p < 0.001, \eta_p^2 = .18$, of Experiment by Lexicality, $F(1, 56) = 10.12, p = .002, \eta_p^2 = .15$, and Accent by Lexicality, $F(1, 56) = 14.06, p < 0.001, \eta_p^2 = .20$. The main effect of Lexicality and the three-way interaction, Experiment by Accent by Lexicality, were not significant (both *ps* > 0.05).

Posthoc analyses conducted for the Experiment by Accent interaction indicated a significant main effect of Accent for Experiment 2, $F(1, 28) = 30.21, p < 0.001, \eta_p^2 = .52$, and not for Experiment 1, $F(1, 28) = 1.29, p = .53, \eta_p^2 = .04$. Posthoc analyses conducted for the Experiment by Lexicality interaction indicated a significant main effect of Lexicality for Experiment 2, $F(1, 28) = 11.71, p < 0.01, \eta_p^2 = .29$, and not for Experiment 1, $F(1, 28) = 1.14, p = .60, \eta_p^2 = .04$.

Note that although the cross-experiments analyses did not indicate a significant Experiment by Accent by Lexicality interaction, the planned analyses in Experiment 2, which included the factors of Accent, Lexicality, and Predictability, showed a significant Accent by Lexicality interaction.

3.5. *Results summary*

In the pre-speech time period, there was a significant effect of Predictability in the 200–300 ms time window. There was no effect of Predictability in the earlier time windows (0–100 ms and 100–200 ms).

In the post-speech time windows, there was no significant effect of Predictability in the N100 time window. In the N400 time window, the Lexicality effect was present in American-accented items, but not in the Chinese-accented items. Also, in the N400 time window, there was a Predictability effect for the Chinese-accented condition, but not the American-accented condition. Cross-experiments analyses confirmed that there were significant effects of Accent and of Lexicality for Experiment 2, but not for Experiment 1.

Behavioral data of the "go" responses revealed that American-accented items were more quickly and accurately identified as animal words than Chinese-accented items, irrespective of predictability of the face cue preceding the items.

4. General discussion

Previous (neuro)cognitive studies presented listeners with speakers associated with one accent (predictable face cues) to examine if face cues affect listeners' accented speech processing (e.g., Grey, Cosgrove, & Van Hell, 2020; Kang & Rubin, 2009; McGowan, 2015; Sala, Vespignani, Casalino, & Peressotti, 2024). The present study introduced a novel manipulation of face cue predictability: Some speakers produced two accents (unpredictable) while others produced one accent (predictable). To examine whether listeners process face cues differently based on predictability, listeners were introduced to the speakers prior to the experimental task so they would have the face cue predictability information from the beginning of the EEG task (Experiment 2). Pre-speech analyses in the 200–300 ms time window revealed a significant predictability effect: Listeners exhibited different neural patterns when they saw a face cue whose accent can be predicted (one-accent speakers) as opposed to seeing a speaker whose accent cannot be predicted (two-accent speakers). Specifically, predictable speakers elicited more negative mean amplitudes than the unpredictable speakers in the 200–300 ms time window. This pre-speech ERP result supports our prediction that a face cue elicits a brain response that reflects linguistic information associated with the speaker's accent predictability information.

We examined our second, and main research question on how face cue predictability influences the online neurocognitive processing of words and nonwords produced in the listener's native accent (American-accented English) and a nonnative accent (Chinese-accented English) by presenting the speaker's face cue–predictable or unpredictable

accent–concurrently with the American-accented or Chinese-accented words and nonwords. Building on the cue-integration model (A. E. Martin, 2016), we predicted N400 lexicality effects (larger N400 for nonwords than for words) for both accents, and that the magnitude of the lexicality effect would vary with the predictability and accent of the speech. However, the N400 analyses revealed a lexicality effect only for the American-accented items and not for the Chinese-accented items. For American-accented items, the lexicality effect was observed irrespective of the face cue predictability. This suggests that face cue predictability had minimal influence (no predictability effect) on the online neurocognitive processing of American-accented words and nonwords – the accent native and familiar to the listeners. In contrast, for Chinese-accented items, a predictability effect was obtained in the N400 analysis: Chinese-accented items spoken by the predictable Chinese-accented speaker elicited a more negative N400 amplitude than Chinese-accented items produced by the unpredictable mixed-accent speakers. This indicates that face cue predictability does affect listeners' online neurocognitive processing of Chinese-accented items – an accent listeners were not experienced with. The overall results in the N400 time window suggest that face cue predictability is integrated during speech processing differently based on the accent of the speaker: Face cue predictability plays a role in Chinese-accented speech processing (the accent listeners were not experienced with), but not in American-accented speech processing (the listeners' own accent).

This interpretation of the Experiment 2 results is supported by the audio-only experiment (Experiment 1). The listeners in Experiment 1 were not introduced to the speakers prior to the go/no task and were not presented with face cues during the go/no task. In Experiment 1, the N400 ERP analysis showed no lexicality and no accent effects, nor a significant interaction. Compared to Experiment 2, where we found a lexicality effect for American-accented items but not for Chinese-accented items, there were no lexicality effects for *both* accent conditions in Experiment 1. Between-experiment ERP analyses confirmed significant experiment differences: there was a main effect of accent and lexicality for Experiment 2, but not Experiment 1. These differences indicate that face cues played a critical role in the online processing of native- and nonnative-accented words and nonwords. Indeed, face cues have been found to influence speech perception at the sentence level in previous ERP and behavioral literature (e.g., Grey, Cosgrove, & Van Hell, 2020; Kang & Rubin, 2009; McGowan, 2015). In these studies, face cues were always predictable (one accent) with respect to native- and nonnative-accented speech. As we will discuss further below, our study adds to this literature by showing that the predictability of the face cue has a differential effect on native- versus nonnative-accented speech.

Although no N400 lexicality or accent effects were found in Experiment 1, its behavioral results did demonstrate accent effects: listeners were more accurate and faster in the animal decision task for the American-accented items than for the Chinese-accented items. So, despite the lack of online accent effects in Experiment 1, the offline behavioral data indicate that participants were sensitive to speech accent. This sensitivity to accent in the go/no-go task in Experiment 1 (which was similar to accent sensitivity in the Experiment 2 go/no-go task), combined with their high accuracy, corroborates that listeners in Experiment 1 were attentive and committed to their task. Hence, the absence of significant N400 ERP effects of accent during listeners' real-time processing of native- and nonnative-accented items does not seem to be driven by task negligence.

What then might explain the absence of an N400 lexicality effect in Experiment 1? In the go/no-go animal decision task, the listeners made a button press response when they heard an animal word, so their attention was on processing the auditory items explicitly for animal words rather than on lexicality judgements. We had expected the task demand of the cover task to reduce lexicality effects in the experiment. Indeed, N400 negativities have been found to be reduced during a passive listening task relative to an explicit yes/no lexical decision task (e.g., O'Rourke & Holcomb, 2002), and we found N400 lexicality effects for American-accented speech in our Experiment 2. Although O'Rourke & Holcomb (2002) did not explicitly analyze N400 lexicality effects (because their research questions were on the latencies in spoken word processing), visual inspection of their Fig. 13 indicates a reduced N400 effect for their Experiment 2 (listening task) relative to their Experiment 1 (lexical decision task). However, in our study, we had not expected the cover task would eliminate the lexicality effect, especially for American-accented speech since earlier ERP studies that used the go/no-go animal decision task have found N400 amplitude differences for cognates and noncognates (Midgley, Holcomb, & Grainger, 2009) or an N400-like negativity in response to English and French translated items (Midgley, Holcomb, & Grainger, 2011). Further research may seek to examine this issue by explicitly manipulating the nature of the decision task. For the purpose of our main question on face cue predictability, however, this outcome can nevertheless serve as a baseline measure against which the effect of adding a face cue can be interpreted, which we will turn to next.

The main question of the present study is how face cue predictability influences the online processing of native- and nonnative-accented words and nonwords. Previous literature has shown processing and intelligibility differences between face-cued (audiovisual or image cueing the speaker identity) and audio-only native- and nonnative-accented speech (e.g., Grey, Cosgrove, & Van Hell, 2020; Kang & Rubin, 2009; McGowan, 2015; McLaughlin, Brown, Carraturo, & Van Engen, 2022; Rubin, 1992; Yi, Phelps, Smiljanic, & Chandrasekaran, 2013). For example, McLaughlin, Brown, Carraturo, & Van Engen (2022) and Yi, Phelps, Smiljanic, & Chandrasekaran (2013) found an audiovisual over audio-only benefit for listeners comprehending native-accented and nonnative-accented speech, but the benefit was smaller for nonnative-accented speech. Also, Grey, Cosgrove, & Van Hell (2020) found that the face cues presented at the beginning of the experiment (a White face for the American-accented English speech and an Asian face for the Chinese-accented English speech) facilitated the processing of both native- and nonnative-accented speech. More specifically, for their face-cued group, a P600 component, associated with syntactic violations and sentence level restructuring and reintegration, was found for both native- and nonnative-accented speech processing of pronoun mismatches, but this was not found in their audio-only group (Grey & Van Hell, 2017). In the present study, face cues were also facilitatory as there were between-experiment accent and lexicality differences for Experiment 2 (face-cued audio) and Experiment 1 (audio-only). Experiment 2 (face-cued) listeners showed N400 lexicality processing for American-accented speech while Experiment 1 (audio-only) listeners did not.

Extending this earlier work, our novel manipulation of face cue predictability (predictable: one accent, unpredictable: two accents) reflects a more naturalistic listening context as the accents of our interlocutors are sometimes unpredictable. Face cue predictability effects were found for Chinese-accented speech, but not for American-accented speech processing. As our monolingual participants were American-English speakers themselves and not familiar with Chinese-accented speech (i.e., did not grow up exposed to or speak this accent), this result suggests that face-cue predictability is particularly pertinent for the processing of nonnative-accented speech in listeners unfamiliar with that accent. An interesting follow-up question would be to examine the role of listeners' experience with mixed-accent speakers. Would these listeners be less sensitive to face-cue predictability in processing accented speech? Based on Fernandez (2019), we hypothesize this would be the case. Fernandez (2019) observed that Chinese-American listeners, who, based on lived experiences, were familiar with incongruences between face cues (here: Asian faces) and acoustic cues (American-accented speech), were not sensitive to a face cue-accent congruency/incongruency manipulation, as opposed to White listeners with minimal familiarity with nonnative-accented speech.

How do the ERP results in Experiment 2 relate to C. D. Martin, Molnar, & Carreiras (2016) who manipulated face cue predictability by having Spanish-Basque bilinguals listen to monolingual Spanish and

monolingual Basque speakers (predictable) and Spanish-Basque bilingual speakers (unpredictable)? In our pre-speech time windows, a predictability effect was present only in the 200–300 ms time window, while in C. D. Martin, Molnar, & Carreiras (2016) a predictability effect was already present in the earlier 100–200 ms time window and in the 200–300 ms time window. However, the direction of the predictability effect in the 200–300 ms time window is consistent across our study and C. D. Martin, Molnar, & Carreiras (2016): the predictable speakers (one-accent speakers in the present study and one-language speakers in C. D. Martin et al.) elicited more negative mean amplitudes than the unpredictable speakers (two-accent speakers in the present study and bilingual speakers in C. D. Martin et al.). One possible explanation for the onset difference for the predictability effect in the two studies is that our predictability manipulation pertained to the face cue (i.e., speaker) predictability regarding speech accent, and our monolingual listeners were unfamiliar to the nonnative-accent variety. In contrast, in C. D. Martin, Molnar, & Carreiras (2016) the predictability pertained to the language(s) of the speakers (Basque and Spanish), and the Spanish-Basque bilingual listeners were fluent and familiar with both languages. These bilingual listeners thus had ample experience with two languages and with speakers who do and do not switch between languages. These language experiences might lead to differences in the weight assigned to the visual face cues, and possibly leading to earlier pre-speech predictability effects.

While we did not find any N1 lexicality effects, C. D. Martin, Molnar, & Carreiras (2016) found N1 lexicality effects in their (predictable) monolingual speaker condition and not in their (unpredictable) bilingual speaker condition with their bilingual listeners. If we were to expect any lexicality effects in the N1 time window, it would most likely emerge in the single-accent predictable American-accented speech context. One possible explanation for the lack of N1 lexicality effects in Experiment 2 is that C. D. Martin, Molnar, & Carreiras (2016) used a yes/no lexical decision task whereas we used a go/no-go animal decision task. As discussed earlier, task procedure can influence the magnitude of the N400 effects (O'Rourke & Holcomb, 2002), so a task not explicitly asking for a lexical decision on the critical words and nonwords (as we used) can reduce or eliminate N1 lexicality effects. Another possible explanation for the lack of N1 lexicality effects is that our nonwords were derived from existing words by changing the second syllable of the word – the first syllable always remained intact. So, it is not until listeners hear the second syllable that the item can unfold as a nonword. The N1 time period, which is extremely early (100–200 ms post stimulus onset), may be too early for the second syllable of the nonword to be processed. C. D. Martin, Molnar, & Carreiras (2016) had created pseudowords by changing up to 2 phonemes in 2–4 syllable words, but did not report in which syllable phonemes were changed. If their phoneme changes were on the first syllable already, this may have facilitated the emergence of the earlier N1 lexicality effect. Lastly, the listeners in C. D. Martin, Molnar, & Carreiras (2016) were fluent Spanish-Basque bilingual speakers, so that may increase their sensitivity to the acoustic characteristics between the words and nonwords in both languages, in contrast to our listeners who were monolingual native American-accented English listeners unfamiliar to Chinese-accented English.

In the N400 time window, C. D. Martin, Molnar, & Carreiras (2016) found N400 lexicality effects for both monolingual and bilingual speaker conditions, but the magnitude of the lexicality effect was larger in the (predictable) monolingual speaker condition. We found lexicality effects in the American-accented speech, in both predictable and unpredictable contexts. Our listeners were monolingual American-English speakers with minimal familiarity with Chinese-accented English, and their lack of familiarity to Chinese-accented English may explain why they did not demonstrate N400 lexicality effects for Chinese-accented English speech in Experiment 2, and only for American-accented English (and irrespective of the face cue predictability). Interestingly, although no lexicality effects were found for Chinese-accented items, this condition did yield N400 predictability effects, suggesting that listeners were relying on the predictability of the face cues during the speech processing of Chinese-accented speech, the accented variant they were not familiar with.

How can the results in this study be interpreted within the framework of the cue integration model (A. E. Martin, 2016)? According to the cue integration model, any relevant psycholinguistic cue, whether visual or auditory, may be reweighted based on the reliability of the cue and integrated during speech comprehension. In Experiment 2, the speaker familiarization videos had informed listeners about each speaker's language background and their accent(s), which provided them with top-down knowledge to make predictions when presented with a face cue of the speaker during the EEG task. The pre-speech predictability effect observed in Experiment 2 is in line with the cue integration model, which predicts that cue reliability is encoded in the neural or electrophysiological signal, because the visual face cue contained information regarding the possible accent(s) of the speakers' English speech. So, although there was no speech signal yet, listeners have begun processing the visual cue as it pertains to the upcoming speech signal. Prior to the onset of the speech, the face cue informed the listeners of the possible accent(s) that they can expect.

During online speech processing, Experiment 2 revealed that the predictability effects were modulated by accent, such that it impacted listeners' processing of Chinese-accented speech but not of American-accented speech. This modulation could be interpreted as a reweighting of the face cue upon hearing the accent of the auditory cue. For the monolingual American-English listeners, their internal representation of American-accented English is robust and reliable. So, possibly, in the American-accented speech processing context, the face cue information is redundant and unnecessary, which might lead to minimal or no weight for the face cue and minimal integration during speech processing. In contrast, these listeners did not have a strong internal representation of the Chinese-accented auditory cues, so in the Chinese-accented speech processing context, the weight of the visual cue increased to support speech comprehension. Moreover, it is likely that the Chinese-accented auditory cues were processed in terms of listeners' American-accented English representations. For these listeners, the face cue weight varied based on the speech accent rather than on face cue predictability. For nonnative-accented speech, that the monolingual listeners were not familiar with, the nonnative audio cue weight decreased as it does not align with the listeners' internal representation. So, with a decreased weight of the audio cue, the face cue weight increased during speech processing. Although the face cue contained linguistic information for both accents, listeners did not use or integrate that information in all speech processing contexts, only in case of processing nonnative-accented speech they were not familiar with.

In Experiment 2, listeners' speech processing reflects that they used the top-down information they learned about the speaker from the introduction videos. Interestingly, we found differential processing of face cues and native- and nonnative- accented speech in which predictability played a role only during nonnative-speech accent, but not native-accented speech. This may in part be due to the participants' weaker or less developed internal representations of Chinese-accented English speech. Here, instead of listeners opting out of the communicative burden when hearing nonnative-accented speech (Lippi-Green, 1997), it appears that listeners remained engaged but processing nonnative-accented speech was more effortful. Extending this finding to other dialectal varieties and language contexts, we expect monolingual speakers who have minimal experience with a particular dialect (e.g., Southern American English) or nonnative-accented speech (e.g., Spanish-accented English) to rely on top-down information to aid in language comprehension of an unfamiliar accent or dialect. Listeners may also form language-related predictions based on speaker identity features within race, such as personas (e.g., D'Onofrio, 2019). Rather than generating expectations regarding the speaker across a racial line, persona portrayed by the speaker modulated the listeners' expectations of nonnative-accented speech (D'Onofrio, 2019).

Future ERP research may also seek to unravel how visual face cues impact listeners' processing of familiar dialectal and accented speech processing. For listeners familiar with both accents, dialect, and/or language, face cues may play less of a role. Indeed, in a behavioral study that examined face cue (Asian face, Caucasian face) and L2 accented speech (American-accent versus Chinese-accent) comprehension, the authors reported that, irrespective of the face cues, Chinese-English bilingual participants showed semantic effects for American-accented word pairs in accuracy and RT, and semantic effects in RT and reversed semantic effects in accuracy for Chinese-accented word pairs (Ma, Ai, Xiao, Guo, & Pae, 2020).

Our findings related to the behavioral cover task point to some possible limitation of this task in EEG research. We decided to use the go/no-go animal decision task to eliminate button response and response preparation on the critical experimental items, to minimize response-related activities to impact the EEG signal. We had expected reduced N400 lexicality effects since the no-go items were the critical words and nonwords. Indeed, previous research has reported reduced N400 lexicality effects in a passive listening task of words and nonwords (O'Rourke & Holcomb, 2002). Yet, we did not expect no lexicality effects in Experiment 1, especially for the American-accented items. Future research may implement a yes/no lexical decision task in both audio-only and face-cued experiment groups to capture the impact of face cue predictability on word processing in a more explicit lexical processing context. Behavioral data might then be correlated with the ERP data and support the interpretation of the ERP results.

Also, listeners in the two experiments were all monolingual native-accented speakers of English. They had minimal experience and minimal familiarity with nonnative-accented speech. Future work can examine listeners with a different language experience background to explore how different language experiences influence the role of face cue predictability during accented speech processing. Testing different language experiences can include listeners who are familiar with the nonnative-accented speech such as individuals who grew up hearing Chinese-accented speech in their daily life or speakers of Chinese-accented English. These individuals would have a more robust internal representation of the Chinese-accented acoustic cues, especially when compared to the individuals who are unfamiliar with Chinese-accented English as tested in the present study.

The current research consists of speech processing of a single word/nonword, but in everyday communication and natural speech processing settings, we use sentences. The processing of sentences involves more complex language processes and cognitive effort. In sentence processing, the continuous acoustic signal is parsed and activates words that need to be integrated in a semantically and syntactically well-formed sentence structure. The hierarchical sentence structure also needs to be constructed while holding multiple units of information. Furthermore, suprasegmental aspects of speech (e.g., prosody, speech rate) that characterize nonnative-accented speech are less explicitly present in word processing. Future research can examine accented-speech processing and face cue predictability in a more natural sentence processing context, and extend studies that have examined predictable face cues and accented sentence processing (e.g., Grey, Cosgrove, & Van Hell, 2020; Sala, Vespignani, Casalino, & Peressotti, 2024).

In conclusion, listeners introduced to the speakers' accent(s) processed face cues differently based on the predictability of the speaker's accent. For native-accented speech, N400 lexicality effects were found regardless of face cue predictability. For nonnative-accented speech, no N400 lexicality effects were found, but there was a face cue predictability effect. Altogether we found neural activity reflecting the integration of face cue predictability and nonnative-accented acoustic cues. Listeners without extensive experience with the nonnative accent integrated speaker face cues during the online word processing of nonnative-accented speech, but not native-accented speech produced by the same speaker.

CRediT authorship contribution statement

Daisy Lei: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yuka Tatsumi:** Writing – review & editing, Software, Data curation. **Erica Hsieh:** Writing – review & editing, Software, Data curation. **Cristal Giorio:** Writing – review & editing, Software, Data curation. **Janet G. van Hell:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing 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.

Acknowledgement

This work was supported by grants from the National Science Foundation (NSF) Ling-DDRI BCS 2215183 awarded to Daisy Lei and Janet G. van Hell, GRFP DGE 1255832 awarded to Daisy Lei, GRFP DGE 1255832 awarded to Cristal Giorio, and BCS 2041081, OISE 1545900, and DGE NRT 2125865 awarded to Janet G. van Hell.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2025.105621.

Data availability

Data will be made available on request.

References

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America, 137*(5), 2823–2833. https://doi.org/10.1121/1.4919317

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*(4), 343–355. https://doi.org/10.1016/0013-4694(85)90008-2

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer [Computer program]*. (Version 6.3.02) [Computer software]. http://www.praat.org/.

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

D'Onofrio, A. (2019). Complicating categories: Personae mediate racialized expectations of non-native speech. *Journal of Sociolinguistics, 23*(4), 346–366. https://doi.org/10.1111/josl.12368

Ethnologue (2023). How Many Languages Are There in the World? https://www.ethnologue.com/insights/how-many-languages/.

European Commission (2018). 65% Know at Least One Foreign Language in the EU. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Foreign_language_skills_statistics.

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology, 59*(1), Article e13940. https://doi.org/10.1111/psyp.13940

Fernandez, C.B. (2019). *The role of speaker identity on the processing of native and foreign-accented speech* [Dissertation, The Pennsylvania State University]. Penn State University Libraries.

Flege, J. (1995). *Second language speech learning: Theory, findings and problems* (pp. 229–273).

Getz, L. M., & Toscano, J. C. (2021). The time-course of speech perception revealed by temporally-sensitive neural measures. *Wiley Interdisciplinary Reviews: Cognitive Science, 12*(2), e1541. https://doi.org/10.1002/wcs.1541

Goh, W. D., Yap, M. J., & Chee, Q. W. (2020). The Auditory English Lexicon Project: A multi-talker, multi-region psycholinguistic database of 10,170 spoken words and nonwords. *Behavior Research Methods, 52*(5), 2202–2231. https://doi.org/10.3758/s13428-020-01352-0

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General, 136*(3), 389–413. https://doi.org/10.1037/0096-3445.136.3.389

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 24–44. https://doi.org/10.1016/S0022-5371(83)80004-8

Grey, S., Cosgrove, A. L., & Van Hell, J. G. (2020). Faces with foreign accents: An event-related potential study of accented sentence comprehension. *Neuropsychologia, 147*, Article 107575. https://doi.org/10.1016/j.neuropsychologia.2020.107575

Grey, S., & Van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics, 42*, 93–108. https://doi.org/10.1016/j.jneuroling.2016.12.001

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103

Helenius, P., Salmelin, R., Richardson, U., Leinonen, S., & Lyytinen, H. (2002). Abnormal auditory cortical activation in dyslexia 100 msec after speech onset. *Journal of Cognitive Neuroscience, 14*(4), 603–617. https://doi.org/10.1162/08989290260045846

Hernández-Gutiérrez, D., Muñoz, F., Sánchez-García, J., Sommer, W., Abdel Rahman, R., Casado, P., Jiménez-Ortega, L., Espuny, J., Fondevila, S., & Martín-Loeches, M. (2021). Situating language in a minimal social context: How seeing a picture of the speaker's face affects language comprehension. *Social Cognitive and Affective Neuroscience, 16*(5), 502–511. https://doi.org/10.1093/scan/nsab009

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*(4), 281–312. https://doi.org/10.1080/01690969008407065

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology, 28*(4), 441–456. https://doi.org/10.1177/0261927X09341950

Kapnoula, E. C., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language, 223*, Article 105031. https://doi.org/10.1016/j.bandl.2021.105031

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*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2022). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications, 7*(1), 7. https://doi.org/10.1186/s41235-022-00354-0

Lee, H., Lee, Y., Tae, J., & Kwon, Y. (2019). Advantage of the go/no-go task over the yes/no lexical decision task: ERP indexes of parameters in the diffusion model. *PLOS ONE, 14*(7), Article e0218451. https://doi.org/10.1371/journal.pone.0218451

Lei, D., Liu, Y., & Van Hell, J. G. (2022a). Novel word learning with verbal definitions and images: Tracking consolidation with behavioral and event-related potential measures. *Language Learning, 72*(4), 941–979. https://doi.org/10.1111/lang.12502

Lei, D., Liu, Y., & Van Hell, J. G. (2022b). Questionnaires. Materials from "Novel word learning with verbal definitions and images: Tracking consolidation with behavioral and event-related potential measures" [Questionnaire]. [Dataset]. IRIS Database. https://doi.org/10.48316/59kt-ah37.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Li, Y., Yang, J., Suzanne Scherf, K., & Li, P. (2013). Two faces, two languages: An fMRI study of bilingual picture naming. *Brain and Language, 127*(3), 452–462. https://doi.org/10.1016/j.bandl.2013.09.005

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

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00213

Ma, F., Ai, H., Xiao, T., Guo, T., & Pae, H. K. (2020). Speech perception in a second language: Hearing is believing—seeing is not. *Quarterly Journal of Experimental Psychology, 73*(6), 881–890. https://doi.org/10.1177/1747021820911362

MacGregor, L. J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications, 3*(1), 711. https://doi.org/10.1038/ncomms1715

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1–71. https://doi.org/10.1016/0010-0277(80)90015-3

Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00120

Martin, C. D., Molnar, M., & Carreiras, M. (2016). The proactive bilingual brain: Using interlocutor identity to generate predictions for language processing. *Scientific Reports, 6*(1), 26171. https://doi.org/10.1038/srep26171

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech, 58*(4), 502–521. https://doi.org/10.1177/0023830914565191

McLaughlin, D. J., Brown, V. A., Carraturo, S., & Van Engen, K. J. (2022). Revisiting the relationship between implicit racial bias and audiovisual benefit for nonnative-accented speech. *Attention, Perception, & Psychophysics, 84*(6), 2074–2086. https://doi.org/10.3758/s13414-021-02423-w

McNorgan, C., Chabal, S., O'Young, D., Lukic, S., & Booth, J. R. (2015). Task dependent lexicality effects support interactive models of reading: A meta-analytic neuroimaging review. *Neuropsychologia, 67*, 148–158. https://doi.org/10.1016/j.neuropsychologia.2014.12.014

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2009). Language effects in second language learners and proficient bilinguals investigated with event-related potentials. *Journal of Neurolinguistics, 22*(3), 281–300. https://doi.org/10.1016/j.jneuroling.2008.08.001

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience, 23*(7), 1634–1647. https://doi.org/10.1162/jocn.2010.21463

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

O'Rourke, T. B., & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology, 60*(2–3), 121–150. https://doi.org/10.1016/S0301-0511(02)00045-5

Rogers, C. L., & Dalby, J. (2005). Forced-choice analysis of segmental production by Chinese-accented English speakers. *Journal of Speech, Language, and Hearing Research, 48*(2), 306–322. https://doi.org/10.1044/1092-4388(2005/021)

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience, 9*, 167. https://doi.org/10.3389/fnhum.2015.00167

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511–531. https://doi.org/10.1007/BF00973770

Sala, M., Vespignani, F., Casalino, L., & Peressotti, F. (2024). I know how you'll say it: Evidence of speaker-specific speech prediction. *Psychonomic Bulletin & Review, 1–13*. https://doi.org/10.3758/s13423-024-02488-2

United States Census Bureau (2022). American Community Survey. https://data.census.gov/table/ACSDP5Y2022.DP02.

Vergara-Martínez, M., Gomez, P., & Perea, M. (2020). Should I stay or should I go? An ERP analysis of two-choice versus go/no-go response procedures in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(11), 2034. https://doi.org/10.1037/xlm0000942

Winsler, K., Midgley, K. J., Grainger, J., & Holcomb, P. J. (2018). An electrophysiological megastudy of spoken word recognition. *Language, Cognition and Neuroscience, 33*(8), 1063–1082. https://doi.org/10.1080/23273798.2018.1455985

Woollams, A. M. (2015). Lexical is as lexical does: Computational approaches to lexical representation. *Language, Cognition and Neuroscience, 30*(4), 395–408. https://doi.org/10.1080/23273798.2015.1005637

Xu, J., Abdel Rahman, R., & Sommer, W. (2022). Who speaks next? Adaptations to speaker identity in processing spoken sentences. *Psychophysiology, 59*(1), Article e13948. https://doi.org/10.1111/psyp.13948

Yi, H. G., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America, 134*(5), Article EL387–EL393. https://doi.org/10.1121/1.4822320