



Speaker Demographics Modulate Listeners' Neural Correlates of Spoken Word Processing

Hanlin Wu, Xufeng Duan, and Zhenguang G. Cai

Abstract

■ In language comprehension, listeners expect a speaker to be consistent in their word choice for labeling the same object. For instance, if a speaker previously refers to a piece of furniture as a “couch,” in subsequent references, listeners would expect the speaker to repeat this label instead of switching to an alternative label such as “sofa.” Moreover, it has been found that speakers' demographic backgrounds, often inferred from their voice, influence how listeners process their language. The question in focus, therefore, is whether speaker demographics influence how listeners expect the speaker to repeat or switch labels. In this study, we used ERPs to investigate whether listeners expect a child speaker to be less likely to switch labels compared to an adult speaker, given the common belief that children are less flexible in language use. In the experiment, we used 80 pictures with alternative labels in Mandarin Chinese (e.g., *yi sheng*₁ vs. *dai fu*, “doctor”). Each picture was presented twice over two experimental phases: In the establishment phase, participants listened to an adult or a child naming

a picture with one of the labels and decided whether the label matched the picture they saw; in the test phase, participants listened to the same speaker naming the same picture by either repeating the original label or switching to an alternative label and, again, decided whether the label matched the picture they saw. ERP results in the test phase revealed that, compared to repeated labels, switched labels elicited an N400 effect (300–600 msec after label onset) and a P600 effect (600–1000 msec after label onset). Critically, the N400 effect was larger when listeners were exposed to the child speaker than to the adult speaker, suggesting that listeners found a switched label harder to comprehend when it was produced by a child speaker than an adult speaker. Our study shows that the perceived speaker demographic backgrounds influence listeners' neural responses to spoken words, particularly in relation to their expectations regarding the speaker's label switching behavior. This finding contributes to a broader understanding of the relationship between social cognition and language processing. ■

INTRODUCTION

In language comprehension, listeners expect the speaker to consistently use the same label when referring to the same concept (Shintel & Keysar, 2007; Brennan & Clark, 1996). For instance, once a speaker has referred to an object as a “couch,” listeners anticipate the speaker to continue using “couch” in subsequent references, rather than switching to an alternative label such as “sofa.” Indeed, speakers tend to repeat a previously used label for the same concept (i.e., label repetition). However, they do sometimes switch to other labels for the same concept (e.g., when a concept has synonymous labels), a common linguistic behavior in language communication known as “label switching” (or “precedent breaking”; Kronmüller & Barr, 2007, 2015).

Compared to label repetition, label switching has been found to result in comprehension difficulties for listeners. In a study by Barr and Keysar (2002), they tracked participants' eye movements when they listened to a speaker's instructions to rearrange objects. They found that participants (i.e., listeners) identified objects more quickly if the objects had been named by the speaker (i.e., objects with

an established label) than if they were named for the first time (i.e., objects without an established label). This enhancement in referential search can be interpreted as the benefit of label repetition. Subsequently, Metzger and Brennan (2003) further examined cases where the speaker either repeated the original label (label repetition) or switched to a new label (label switching). They observed that listeners slowed down in object identification when speakers switched to a new label compared to repeating the original label, demonstrating a disruptive effect of label switching on comprehension. Over the years, similar label switching effects have been corroborated in many studies (e.g., Graham, Sedivy, & Khu, 2014; Horton & Slaten, 2012; Matthews, Lieven, & Tomasello, 2010; Brown-Schmidt, 2009; Kronmüller & Barr, 2007; Shintel & Keysar, 2007).

Label Switching Effects Modulated by the Speaker's Individual Identity

Research has looked at whether label switching effects are modulated by the speaker's individual identity. A common method is to involve two experimental phases: In the first phase (i.e., the establishment phase), a speaker uses a label to refer to an object; in the second phase (i.e., the

The Chinese University of Hong Kong

test phase), either the original speaker or a new speaker repeats the label or switches to an alternative label in referring to the same object. It is important to note that, in these studies, the old and new speakers differ in their individual identities but are from the same demographic group (e.g., adult native English speakers). In Metzger and Brennan's (2003) study, they found that the label switching effect (in the test phase) was modulated by the identity of the speaker (old vs. new speaker): When labels were repeated, listeners identified target objects equally quickly regardless of whether the speaker was old or new; when labels were switched, listeners were slower to find the target object with the old speaker than with a new speaker. This finding has been replicated in subsequent studies (e.g., Kronmüller & Barr, 2007, 2015; Horton & Slaten, 2012; Brown-Schmidt, 2009). Modulation of label switching by the speaker's individual identity has also been observed in listeners' neural dynamics. For instance, a magnetoencephalography study by Bögels, Barr, Garrod, and Kessler (2015) showed that there was a significant increase in theta-band neural oscillations (3–7 Hz) around 350–650 msec after listeners encountered an alternative label that deviated from the label originally used by the same speaker. Importantly, this theta-power increase was absent when the alternative label was produced by a new speaker, thereby indicating that the neural dynamics underlying the processing of switched labels are modulated by the speaker's individual identity.

The observation of label switching effects with an old (but not a new) speaker suggests that these effects are not because of long-term lexical priming. If they were, we would expect to see the effect with a new speaker as well. Instead, the effects are likely because of listeners' expectations that speakers will reuse previously used labels. This difficulty with label switching is often linked to the shared knowledge among interlocutors in a conversation, as noted by Clark (1996). Various theoretical perspectives have been proposed to explain the effects of the speaker's individual identity on processing referential labels. For example, some theories suggest that these effects arise from the listener's integration of the speaker's perspective with the objects being referred to (Jara-Ettinger & Rubio-Fernandez, 2021; Heller, Grodner, & Tanenhaus, 2008), whereas others propose that the effects reflect a mechanism of domain-general memory associated with the speaker (Horton & Slaten, 2012; Horton & Gerrig, 2005). Nonetheless, despite these insights, studies comparing individuals from the same demographic group have not thoroughly investigated how the speakers' demographic backgrounds might influence listeners' behavioral or neural responses to label switching.

Sensitivity to Speaker Demographics in Language Comprehension

Research indicates that a speaker's demographic background plays a significant role in affecting how listeners

interpret the speaker's message and language patterns (Cai, 2022; Cai et al., 2017; van Berkum, van Den Brink, Tesink, Kos, & Hagoort, 2008). A speaker's demographic background encompasses the collective attributes and shared characteristics typical of a specific socioeconomic class (Labov, 2006), gender (Coates, 2015), or age group (Walker & Hay, 2011). However, how speaker demographics impact listeners' responses to label switching remains an open question. Consider a monologue scenario where an individual listens to someone naming pictures over two phases. Some pictures may be verbalized with different labels, for example, "couch" or "sofa" for a piece of furniture. The listener might expect the speaker to use the same label for a picture over both phases. However, the listener might be willing to accept an alternative label for a picture in the second phase (although still surprised) if the speaker has a large vocabulary repertoire and is linguistically flexible (e.g., an adult speaker). In contrast, the listener might find it more surprising if the speaker is less flexible in language use (e.g., a child speaker) when the speaker switches to a new label for a concept in the second phase. In the current study, we ask whether listeners are sensitive to speaker demographics (i.e., an adult vs. child speaker) in processing switched versus repeated labels.

There is evidence that listeners use their accumulated experience of interacting with people of different demographic backgrounds to adjust their language comprehension. For instance, when older individuals use words that were more prevalent in the past (e.g., "knitting"), and when younger individuals use words that are more prevalent contemporarily (e.g., "lifestyle"), this congruency between speaker age and word age facilitates listeners' recognition of those words (Kim, 2016; Walker & Hay, 2011). Furthermore, speakers' dialectal accents have been found to modulate listeners' interpretation of words with different dominant meanings in British English and American English. For instance, a word such as "bonnet" is more likely to be interpreted as referring to a car part when spoken in a British accent and to a type of hat when spoken in an American accent (Cai, 2022; Cai et al., 2017).

Speaker demographics have also been shown to modulate listeners' neural correlates of spoken language processing in terms of the sentence message and lexical variation. Early research using EEG indicates that a critical word in a sentence that violates the stereotypical gender assumptions elicits a larger late positive deflection, known as the P600 effect. For instance, Lattner and Friederici (2003) had participants listen to self-referential sentences that conveyed either a stereotypically gendered message (stereotypically masculine such as "I like to play soccer" or stereotypically feminine such as "I like to wear lipstick"). Each sentence was spoken by both male and female speakers. They found that the incongruency between the speaker's gender and the gender stereotypicality of the message elicited a P600 effect at the critical word (e.g., "soccer" spoken by a female speaker or "lipstick"

spoken by a male speaker). As P600 is often assumed to reflect cognitive repair or reanalysis during language comprehension, this result is interpreted as evidence supporting the reintegration of linguistic information and speaker information at a later stage during spoken language comprehension (Lattner & Friederici, 2003; Osterhout, Bersick, & McLaughlin, 1997). Subsequently, van Berkum and colleagues (2008) used a similar paradigm and tested more demographic attributes including age and socioeconomic status as well as gender. For example, they contrasted sentences such as “Every evening I drink some wine before I go to sleep” spoken by an adult speaker versus by a child speaker. Their results showed that the effect of the speaker’s demographic background could be detected as early as around 300 msec after the onset of the critical word “wine,” in a manner similar to the classic N400 effect elicited by semantic anomalies (van Berkum, Hagoort, & Brown, 1999; Kutas & Hillyard, 1980) and world knowledge violation (Hagoort, Hald, Bastiaansen, & Petersson, 2004). These results, instead, suggested a rapid integration of linguistic information and speaker information at a very early stage of spoken language comprehension, arguing for the necessity of social context in interpreting meanings. In a study examining how listeners’ comprehension of lexical variation is modulated by the speaker’s accent, Martin, Garcia, Potter, Melinger, and Costa (2016) had participants listen to speech spoken in a British or American accent. The speech contained words that were more frequently used in either British or American English. For instance, the term “holiday” is more commonly used in British English, whereas “vacation” is the preferred term in American English. Their findings revealed that words incongruent with the speaker’s accent (e.g., British words spoken in an American accent) elicited greater negative EEG deflections 700 msec after the word onset, which was interpreted as a late N400 effect. These findings showed that listeners integrate their knowledge about the speaker’s dialectal background with their lexical usage as speech unfolds. More recent work using similar paradigms reported either a P600 effect (Foucart et al., 2015), a mixture of N400 and P600 (van den Brink et al., 2012), or a mixture of N400 and P3 (Pélessier & Ferragne, 2022). In general, by manipulating the congruency between the speaker’s demographic background and the sentence message or lexical usage, these studies have shown the influence of speaker demographics on how listeners process speech. However, little research has explored whether the speaker’s demographic background influences how listeners process the speaker’s linguistic behavior such as label switching.

Theoretical Accounts of Speaker Demographics Effects in Language Comprehension

There are two primary views among researchers to account for the speaker demographics effect in spoken language comprehension: the “acoustic detail account” and the “speaker model account” (for comparison, see

Kapnoula & Samuel, 2019; Cai et al., 2017; Creel & Bregman, 2011; Creel & Tumlin, 2011). The acoustic detail account posits that the speaker’s identity influences spoken language processing by affording a less or more similar acoustic match to listeners’ previous encounters with specific instances of speech. This account aligns with the exemplar-based theories of the mental lexicon (Pufahl & Samuel, 2014; Goldinger, 1996, 1998). In these theories, lexical representations consist of exemplars, each being an experienced token of a word that includes detailed episodic memory traces with phonetic, speaker, and contextual information (Walker & Hay, 2011). When a word is produced by a speaker who has produced it previously, the acoustic features of that word token should better match the listener’s memory (exemplar), as compared to a word token produced by a new speaker (Creel & Tumlin, 2011), leading to a speaker effect¹ in speech perception. For instance, in a two-phase experiment, Goldinger (1996) first had participants exposed to a list of word tokens spoken by various speakers in a study phase and then, in a test phase, presented another list of word tokens, which participants decided whether they had previously heard in the first phase. The results showed that participants were more accurate at identifying a word as being previously heard when the word was spoken by the same speaker between the two phases compared to when it was spoken by different speakers. Subsequent research showed that recognition was even better in cases where word tokens are identical (i.e., the same recording), compared to where word tokens are not identical (different recordings) albeit uttered by the same speaker (Clapp, Vaughn, Todd, & Sumner, 2023).

In contrast, the speaker model account posits that speaker demographics influence language processing via a mental model that the listener constructs to capture the attributes of the speaker (e.g., age, gender, socioeconomic status). Listeners use the speaker model to interpret the speaker’s utterances. For instance, Cai and colleagues (2017) demonstrated that listeners had more access to the American meaning of cross-dialectally ambiguous words (e.g., “flat,” “gas”) when these words were spoken by an American English speaker than by a British English speaker (as evident in the accent). Importantly, this speaker effect did not depend on the accentedness of the word token. For example, listeners still had more access to the American meaning of word tokens that were morphed to be accent neutral as long as they believed that the word tokens were produced by an American English speaker (compared to by a British English speaker). This finding suggests that listeners construct a model of the speaker (their dialectic background in this case), probably during the first instances of exposure to the speaker, and use it to interpret spoken words.

Importantly, the acoustic detail account and the speaker model account differ from each other regarding the mechanics whereby listeners use speaker demographics to constrain language comprehension. The acoustic detail

account assumes that the speaker demographics effect arises in a bottom-up manner, whereby listeners search their lexical memories for the best match for any incoming speech signals to determine the word and the meaning of the speech token. The speaker model account instead assumes a top-down mechanism for the speaker demographics effect, whereby listeners construct a higher-level mental model of the speaker demographics, and the utterances are comprehended against this model. It is important to note that the speaker model can be constructed from various sources such as a short exposure of the speaker's accent at the very beginning of a language task, and it can constrain subsequent language processing even when language is presented in the form of text instead of speech (Foucart, Santamaría-García, & Hartsuiker, 2019; Cai et al., 2017). This suggests that the acoustic details of word tokens may not further contribute to an established speaker model if they do not contradict the model. However, these two accounts are not mutually exclusive. Both acoustic details and speaker models may contribute to spoken language processing in tandem, as suggested by dual-route models (Kapnoula & Samuel, 2019; Sumner, Kim, King, & McGowan, 2014).

The Current Study

Our study used EEG to explore the question of whether and how the speaker's demographic background influences listeners' neural activities during spoken word processing, specifically focusing on the speaker's linguistic behavior of label switching. Our central hypothesis is that listeners generally expect speakers to repeat the label they have previously used when referring to the same concept. Importantly, it has been known that children often exhibit a reluctance to ascribe several different labels to the same object, demonstrating a preference for a singular designation (Piccin & Blewitt, 2007; Markman, 1991). Empirical evidence shows that they sometimes refuse to accept a label provided by someone else, even if they are already familiar with that label. Instead, they tend to persist in using their own chosen label (Clark, 1997). Therefore, we predict that listeners should have a lower expectation of label switching (as compared to label repetition) toward a child speaker than toward an adult speaker.

In addition, it is important to note that our focus lies not on the conversation between interlocutors but rather on the comprehension of speech in a noninteractive setting, resembling a monologue. This line of research is of theoretical interest because of quite a few recent studies demonstrating that even in noninteractive language comprehension, listeners create a mental representation of the speaker and utilize it to enhance speech comprehension (e.g., Cai, 2022; Pélissier & Ferragne, 2022; Cai, Sun, & Zhao, 2021; Foucart & Hartsuiker, 2021; Grant, Grey, & van Hell, 2020; Foucart et al., 2015, 2019; Cai et al., 2017; Martin et al., 2016; Bornkessel-Schlesewsky, Krauspenhaar, & Schlesewsky, 2013; van den Brink et al., 2012; van Berkum

et al., 2008; Lattner & Friederici, 2003). Therefore, our study aims to investigate label switching in noninteractive comprehension, specifically examining whether the cost of label switching is influenced by speaker demographics.

Following the classic protocol (Kronmüller & Barr, 2015), we divided the experiment into the establishment phase and the test phase. During the establishment phase, participants listened to a speaker naming pictures using certain labels, whereas in the test phase, participants listened to the same speaker naming the same set of pictures by either repeating the original labels used in the establishment phase (label repetition) or switching to new labels (label switching). We manipulated the speaker's demographic background by having the picture names spoken by an adult or a child (and by giving the instructions that they were to listen to either an adult or a child naming pictures).

Therefore, we predict increased cognitive effort when participants encounter a switched label compared to a repeated label in the test phase, and the increase in cognitive effort should be greater when a child produces a switched label than when an adult does. In our experiment, we expect such increased cognitive effort to be manifested as larger EEG deflections for the switched labels compared to the repeated labels during the test phase. In other words, if listeners' neural responses to the speaker's label switching are influenced by speaker demographics, EEG deflections that reflect the label switching effect should be modulated by whether the speaker is an adult or a child. In addition, to examine when this speaker effect occurs, we investigated two critical time windows: 300–600 and 600–1000 msec after the label onset based on the typical time windows for N400 (Kutas & Hillyard, 1980) and P600 (Aurnhammer, Delogu, Brouwer, & Crocker, 2023), respectively. Following previous studies (Cai, 2022; Cai et al., 2017), we adopted a between-participant design to manipulate the speaker age (i.e., an adult speaker condition vs. a child speaker condition) to minimize participants' awareness of this manipulation.

METHODS

Design

We adopted a 2 (Label: repeated vs. switched) \times 2 (Speaker: adult vs. child) factorial design. Label was manipulated within participants and between items: During the establishment phase, all participants heard the speaker naming all target pictures using the preferred labels; during the test phase, all participants heard the speaker naming half of the target pictures by repeating the original (preferred) labels, while naming the other half by switching to alternative (dispreferred) labels. The assignment of each item as being repeated or switched was counterbalanced across participants. Speaker was manipulated between participants and within items: Participants listened to either the adult or the child in

the two phases, and the same set of items was used for both speaker conditions.

Participants

Our study recruited 48 neurologically healthy native speakers of Mandarin Chinese (37 women; mean age = 23.17 years, $SD = 1.21$ years). Two participants were subsequently excluded from data analysis (see Data Exclusion), leaving a final total of 46 participants. The sample size was determined based on the number of trials in a condition in reference to previous studies investigating similar topics (Pélessier & Ferragne, 2022; Martin et al., 2016) or using similar paradigms (Bögels et al., 2015; Malins & Joannis, 2012; Desroches, Newman, & Joannis, 2009).² All participants provided their informed consent before the experiment began. Note that, we ensured that, if a participant had taken part in a test/experiment in this study, they would not take part in any other test/experiment of the study. The study protocol was approved by the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee.

Materials

Each stimulus item was composed of a color cartoon picture depicting a person or object and two corresponding Mandarin labels (for target items; e.g., y_1sheng_1 and dai_4fu for the picture of a doctor) or a single label (for filler items; e.g., $xiang_1jiao_1$ for the picture of a banana). To construct the target items, we created 98 items (labels and their corresponding color cartoon pictures) following two constraints. First, both labels are disyllabic noun words; second, paired labels do not have syllabic overlap, avoiding any potential confounds induced by the phonological similarity between the two labels. We then subjected all the labels and their associated cartoon pictures to a norming pretest in a laboratory environment, involving 24 native Mandarin speakers as participants. These pretest participants were shown a picture followed by a disyllabic label (presented in writing as a bicharacter word). They rated, on a 7-point Likert scale, the appropriateness of the word as a label for the picture. We excluded 18 items with at least one of its labels having an average rating lower than 3, resulting in 80 target items for the main experiment (see Appendix). For each of these 80 items, the label with the higher average rating was designated as the preferred label (e.g., y_1sheng_1 , “doctor”), and the one with the lower score was designated as the dispreferred label (e.g., dai_4fu). The preferred label was always used in the establishment phase (i.e., the first time the picture was named). This method, also adopted by Bögels and colleagues (2015), was designed to prevent the participants from otherwise activating an alternative (i.e., the preferred) label for the picture when they were being exposed to a dispreferred label, thereby eliminating potential confounds to the experiment. We also created

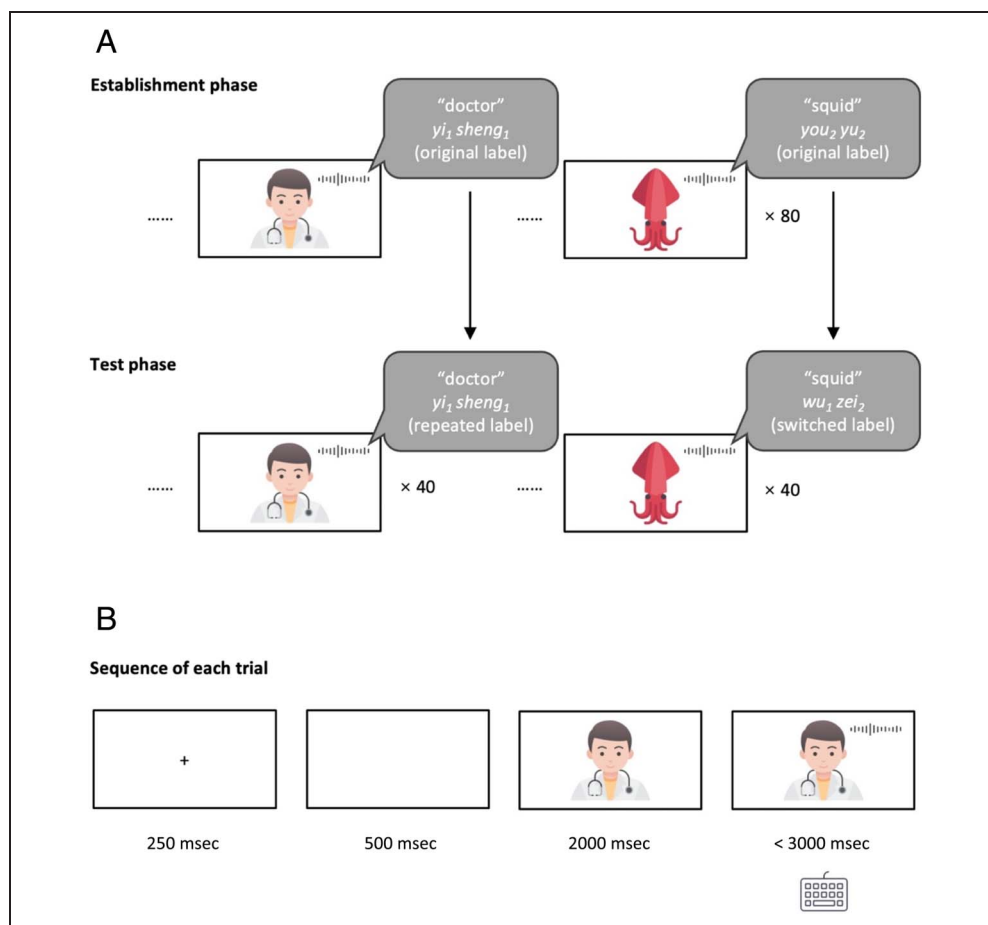
120 filler items, with color cartoon pictures that could be named using a disyllabic or trisyllabic label.

We then generated audio recordings of the labels using both an adult voice and a child voice. To minimize any differences between the adult and the child audio recordings other than the manipulated demographic attribute of age, and to better control for potential confounds such as accents, volume, and speech rate, which are often inevitable if we use human speakers, we used iFLYTEK text-to-speech technology, which provided a realistic voice simulation, to generate two sets of audio files: one set in an adult voice and another in a child voice. The adult voice was designed to mimic a man in his 30s; and the child voice, to resemble a primary-school boy. All stimuli were normalized for their durations and sound volumes; each disyllabic word token had a duration of 650 msec, whereas each trisyllabic one had a duration of 750 msec. To validate the effectiveness of our speaker age manipulation, we conducted an online pretest involving 100 participants (10 of whom were later excluded from analysis for failing to complete the test). Participants listened to all the word tokens in the adult and child voice (speaker voice manipulated between participants) and supplied a number (from 1 to 99) to estimate the perceived age of the speaker. The results showed that participants estimated the adult-voice tokens to be produced by someone with an age of 32.43 ± 6.55 years and the child-voice tokens by someone with an age of 11.83 ± 5.20 years, with a significant difference between the two voices ($\beta = -20.60$, $t = -16.29$, $p < .001$). To further test the naturalness of the word tokens, we conducted an online posttest involving another 100 participants (not included in the age pretest; 13 of them were later excluded from analysis for failing to complete the test); they listened to all the word tokens in the adult and child voice (speaker age again manipulated between participants) and rated how natural they thought a word token was on a 7-point Likert scale (1 = *absolutely unnatural*, 7 = *absolutely natural*). The result showed that the adult voice had a rating of 5.33 ± 1.14 , whereas the child voice had a rating of 5.33 ± 1.17 , showing no significant difference between the two speaker conditions ($\beta = 0.00$, $t = 0.00$, $p = 1$).³ All materials used in this study are available at osf.io/2gvkt/.

Procedure

Participants were individually tested in a soundproof booth designed for EEG signal acquisition. The procedure of the experiment is depicted in Figure 1. Before the start of the establishment phase, we first introduced the “speaker” (either an adult man or a young boy) to participants with a profile photo displayed on the screen. Participants were informed that the “speaker” had been invited to the laboratory, shown pictures, and asked to name each picture using any word he preferred, with his responses recorded. Participants were then required to listen to these audio recordings, each paired with a picture displayed on the screen. Their task was to decide whether

Figure 1. (A) Experimental setup. The experiment comprised two phases: In the establishment phase, listeners heard the speaker naming all pictures using the preferred labels; in the test phase, they heard the speaker naming half of the pictures by repeating the original labels (repeated label condition) and the other half by switching to alternative labels (switched-label condition). (B) Trial structure. A spoken label was played 2000 msec after the picture onset. The EEG response was time-locked to the onset of the spoken label. Participants listened to the spoken label and decided whether it matched the picture or not by pressing the corresponding keys.



the word produced by the “speaker” matched the picture shown on the screen. In the test phase, all pictures were shown again. In half of the target trials, the “speaker” named a picture with the same label he used in the establishment phase (repeated label condition), whereas in the remaining half, the “speaker” switched to an alternative label when naming the picture (switched label condition).

In each phase, the spoken label matched the picture in all 80 target trials as well as 20 filler trials; the label mismatched the picture in the remaining 100 filler trials (thus, in each phase, there were 100 matching trials and 100 mismatching trials). As Label was manipulated between items, we constructed four versions of item lists and ensured that each label used in the establishment phase had an equal chance of being repeated or switched in the test phase. Each participant was randomly assigned to one of the four versions, and the trial orders during both phases were randomized. Each trial followed the following sequence (as shown in Figure 1). First, a fixation cross was presented at the center of the screen for 250 msec, followed by a blank of 500 msec. Then, a picture appeared on the screen; 2000 msec after the picture onset, a spoken label was played, with the picture remaining on the screen, either until the participant made a response or for 3000 msec if no response had been detected. The experiment was conducted using E-Prime 2.0 software (Psychology Software Tools).

Data Exclusion

Two participants were excluded from the final analysis, one in the child speaker condition for noncompliance with task instructions and the other in the adult speaker condition for an excessive error rate (more than 20% inaccurate responses) in the test phase. A final sample of 46 participants comprised 22 in the adult speaker condition and 24 in the child speaker condition.

EEG Recording and Preprocessing

The EEG was recorded during both the establishment phase and the test phase, using 29 Ag–AgCl scalp electrodes mounted on an EasyCap (Brain Products), each referred to CPz. These electrodes were positioned to offer an optimal equidistant selection of 10% positions in the 10/20 system. Three of these electrodes were placed at midline sites (Fz, Cz, and Pz), and 13 pairs were placed at lateral sites (FP1/FP2, F3/F4, F7/F8, FC1/FC2, FC5/FC6, C3/C4, T7/T8, CP1/CP2, CP5/CP6, TP9/TP10, P3/P4, P7/P8, O1/O2). In addition, vertical EOG and horizontal EOG were recorded bipolarly from electrodes placed above and below the left eye and at the outer left and right canthi. Signals were recorded using a Neuroscan SynAmps 2 amplifier and digitized at a sampling rate of 1000 Hz. All

electrode impedances were maintained below 5 kΩ throughout the experiment.

EEG data preprocessing and analyses were performed separately for the establishment phase and the test phase using customized scripts and the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) in MATLAB. For each phase, EEG data were bandpass-filtered offline at 0.1–30 Hz (Tanner, Morgan-Short, & Luck, 2015), rereferenced to the average of all 29 scalp electrodes, and segmented from 1300 msec before and 2000 msec after the onset of the audio in each target trial. Trials with inaccurate responses were excluded (2.2% in the establishment phase and 4.9% in the test phase). Independent component analysis was performed to identify artifacts caused by eye blinks and eyeball movements, with an average of 2.089 components (in the establishment phase) and 2.088 components (in the test phase) identified and removed from the data of each participant. The data were then epoched from 200 msec before to 1000 msec after the onset of the audio, and epochs in which the EEG amplitudes exceeded $\pm 100 \mu\text{V}$ were considered to contain artifacts and thus excluded (3.4% in the establishment phase and 2.2% in the test phase). The data of the remaining epochs were baseline-corrected by subtracting the mean amplitude from 200 msec to 0 msec before the audio onset. EEG and behavioral data for all participants are available at osf.io/2gvkt/.

RESULTS

Behavioral Results

Logit and linear mixed-effects (LME) modeling were conducted on trial-level response accuracy (ACC: correct vs. incorrect responses) and RTs in the test phase, respectively. Label (repeated = -0.5 , switched = 0.5) and

Table 1. LME Models for Behavioral Measures in the Test Phase

Predictor	β	SE	z/t	p
Response ACC				
Intercept	3.79	0.19	19.56	<.001
Label	-1.83	0.21	-8.57	<.001
Speaker	0.37	0.31	1.17	.243
Label: Speaker	0.04	0.43	0.10	.923
RT (log-transformed)				
Intercept	2.88	0.01	219.65	<.001
Label	0.11	0.00	23.33	<.001
Speaker	-0.02	0.03	-0.81	.424
Label: Speaker	0.00	0.01	0.34	.738

The model for the ACC analysis: $\text{ACC} \sim \text{Label} * \text{Speaker} + (1|\text{Participant}) + (\text{Speaker} + 1|\text{Item})$. The model for the RT analysis: $\text{RT} \sim \text{Label} * \text{Speaker} + (\text{Label} + 1|\text{Participant}) + (1|\text{Item})$. Inaccurate responses were excluded for the RT analysis.

Table 2. LME Models for Region Analyses on Amplitudes in the Test Phase

Predictor	β	SE	t	p
<i>Anteriority analysis</i>				
N400 (300–600 msec)				
Intercept	0.09	0.03	2.93	.003
Label	0.00	0.06	0.01	.995
Anteriority	1.21	0.11	10.83	<.001
Label: Anteriority	-0.89	0.12	-7.56	<.001
P600 (600–1000 msec)				
Intercept	0.01	0.03	0.42	.677
Label	0.00	0.07	-0.05	.959
Anteriority	2.06	0.12	17.17	<.001
Label: Anteriority	0.58	0.13	4.45	<.001
<i>Laterality analysis</i>				
N400 (300–600 msec)				
Intercept	0.09	0.03	2.86	.004
Label	0.00	0.06	0.01	.995
Laterality	-0.18	0.13	-1.35	.184
Label: Laterality	0.03	0.12	0.29	.774
P600 (600–1000 msec)				
Intercept	0.01	0.03	0.40	.691
Label	-0.00	0.07	-0.05	.963
Laterality	0.03	0.13	0.19	.845
Label: Laterality	0.07	0.14	0.52	.602

The models for the anteriority analysis: $\text{N400} \sim \text{Label} * \text{Anteriority} + (1|\text{Participant}) + (\text{Anteriority} + 1|\text{Item})$, $\text{P600} \sim \text{Label} * \text{Anteriority} + (1|\text{Participant}) + (\text{Anteriority} + 1|\text{Item})$. The models for the laterality analysis: $\text{N400} \sim \text{Label} * \text{Laterality} + (\text{Laterality} + 1|\text{Participant}) + (1|\text{Item})$, $\text{P600} \sim \text{Label} * \text{Laterality} + (\text{Label} + \text{Laterality} + 1|\text{Participant}) + (1|\text{Item})$.

Speaker (adult = -0.5 , child = 0.5) were used as interacting predictors. Participant and Item were coded as categorical variables and were used as random factors. In all LME analyses conducted in our study, we used the maximal random-effect structure justified by the data and determined by forward model comparison ($\alpha = .2$; see Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

As shown in Table 1, a significant main effect of Label was observed for both ACC and RTs. Responses were more accurate and faster when a label in the test phase was repeated compared to when it was switched (ACC: 0.98 vs. 0.92; RT: 702 vs. 900 msec). Neither the main effect of Speaker nor the interaction between Label and Speaker reached statistical significance.

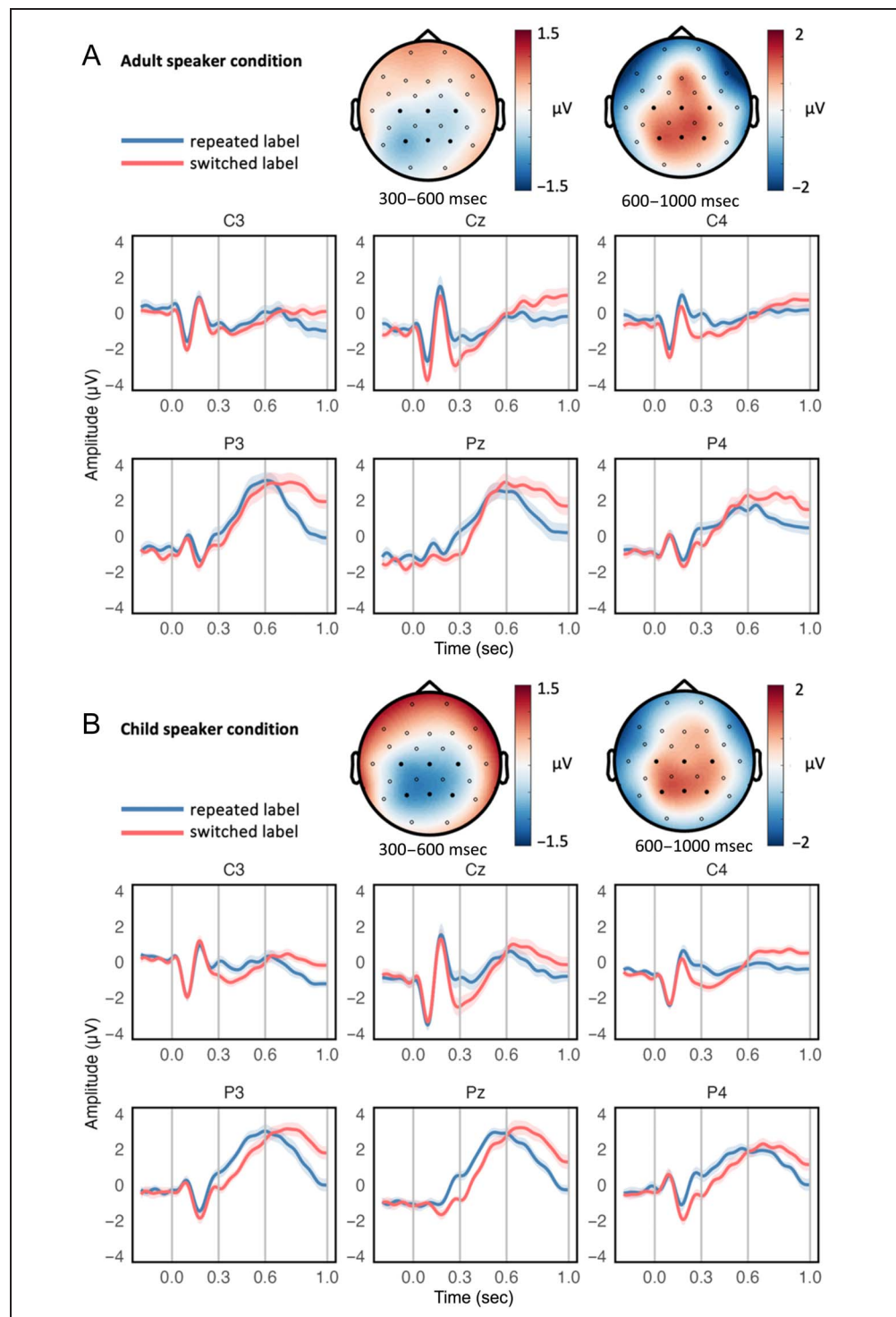
Waveform Analysis

Following the classic time windows of N400 (e.g., Kutas & Hillyard, 1980) and P600 (e.g., Lattner & Friederici, 2003), we focused our analyses on 300–600 and 600–1000 msec after the audio onset for N400 and P600, respectively. We performed waveform analyses by fitting LME models to the mean amplitudes over the target time windows in each

trial (Nieuwland et al., 2018). LME-based methods are suggested to yield more robust results than traditional ANOVA-based methods in ERP amplitude analyses (Heise, Mon, & Bowman, 2022).

To explore the topographies of label switching effects in the test phase, we conducted region analyses of anteriority and laterality, respectively. Scalp sites were divided into four ROIs, with mean amplitudes collapsed across all

Figure 2. Grand-averaged ERPs (with *SE* in shaded areas) elicited by the repeated label (blue line) or the switched label (red line) during the test phase under the adult speaker condition (A) or the child speaker condition (B). Waveforms are illustrated with six centro-parietal sites, namely, C3, Cz, C4, P3, Pz, and P4. Topographical maps show the effect magnitudes (switched label condition – repeated label condition) across the 300- to 600-msec and 600- to 1000-msec time windows.



electrodes in each ROI. The left-anterior sites included Fp1, F3, F7, FC1, and FC5; right-anterior sites included Fp2, F4, F8, FC2, and FC6; left-posterior sites included CP1, CP5, PO3, P7, and O1; right-posterior sites included CP2, CP3, PO4, P8, and O2 (see van Berkum et al., 2008, for a similar selection of ROIs).

As shown in Table 2, for the anteriority analysis, we fit models with Label (repeated = -0.5, switched = 0.5) and Anteriority (anterior = -0.5, posterior = 0.5) as interacting predictors. A significant main effect of Anteriority

and a significant interaction between Anteriority and Label were observed for the time windows of 300–600 and 600–1000 msec. For the laterality analysis, we fit models with Label and Laterality (left = -0.5, right = 0.5) as interacting predictors. Neither a main effect nor an interaction was detected in either the 300- to 600-msec or 600- to 1000-msec window. Combining the results of both the anteriority analysis and the laterality analysis, we confirmed the presence of label switching effects during 300–600 and 600–1000 msec after the audio onset, with larger effects

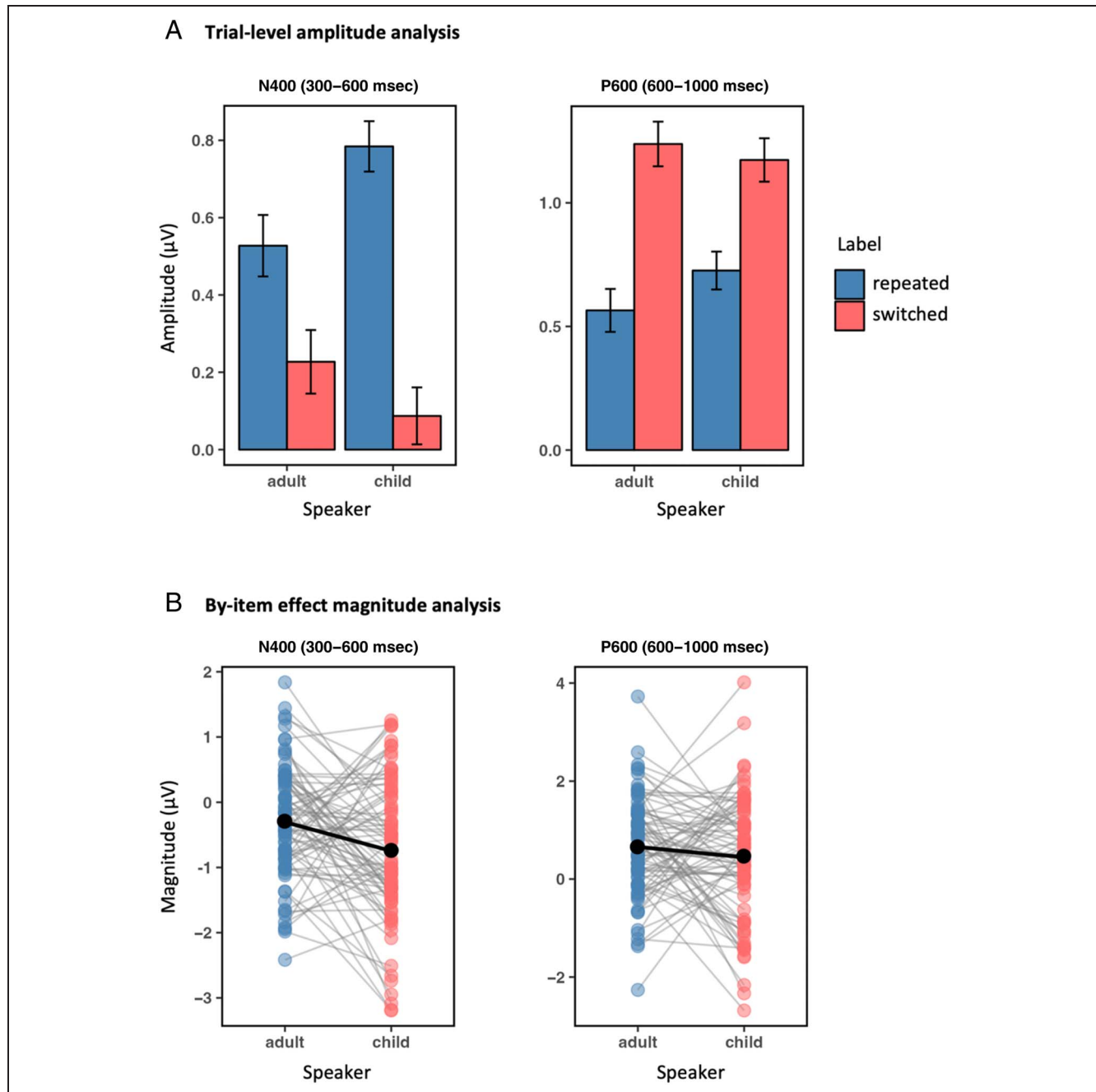
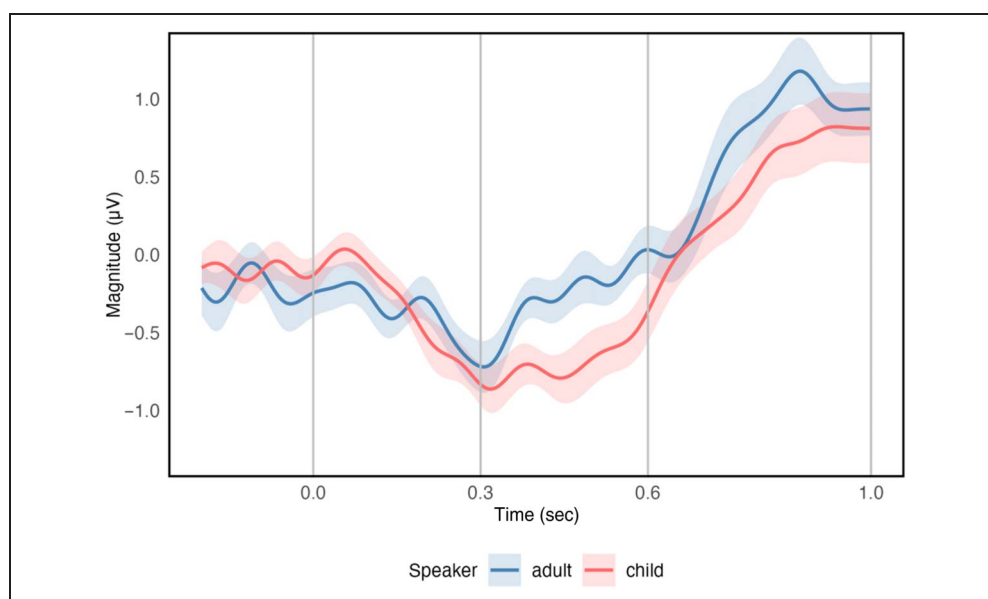


Figure 3. (A) Means and *SEs* of trial-level amplitudes for N400 (300–600 msec) and P600 (600–1000 msec) for repeated label condition and switched label condition in the test phase. (B) Magnitudes of by-item label-switching N400 and P00 effects (switched label condition – repeated label condition); the black dots represent the mean effect magnitudes.

Figure 4. Mean difference waves (switched label condition – repeated label condition) across the centro-parietal sites (Cz, Pz, C3, C4, CP1, CP2, CP5, CP6, P3, P4, P7, P8) for the adult (blue line) and child (red line) speaker conditions in the test phase; shaded areas represent *SEs*.



over the posterior than the anterior regions and no significant difference between the two hemispheres. These results were consistent with the classic time windows and topographies of N400 and P600 in language processing (Figure 2).

Our primary focus was to compare the two speaker conditions with respect to the N400 and P600 effects. To this end, we selected an ROI that comprised all centro-parietal sites (Cz, Pz, C3, C4, CP1, CP2, CP5, CP6, P3, P4, P7, P8) and fit models to the mean amplitudes across these sites with Label and Speaker as interacting predictors. As shown in Figure 3A, Figure 4, and Table 3, in the time windows of 300–600 and 600–1000 msec, significant main

effects of Label were observed, confirming the occurrence of the N400 and P600 effects elicited by switched labels compared to repeated labels. Crucially, a significant interaction between Label and Speaker was observed during 300–600 msec, indicating that the N400 effect was larger in the child speaker condition (0.70 μV) than in the adult speaker condition (0.30 μV). However, this interaction did not reach significance during 600–1000 msec, suggesting that the P600 effects were comparable between the child and adult speaker conditions (0.45 vs. 0.67 μV). These results suggested that switched labels elicited an N400 effect and a P600 effect, and only the N400 effect was further modulated by speaker demographics (i.e., adult vs. child).

In addition, to explore whether label preference influences label switching effects and potentially interacts with the observed modulation effect of speaker demographics, we conducted a post hoc by-item analysis and examined whether the difference in preference ratings between the original (preferred) label and the alternative (dispreferred) label could predict label switching effects on N400 and P600, respectively. We calculated the preference difference for each item by subtracting the rating of the alternative (dispreferred) label from that of the original (preferred) label. We also calculated the N400 and P600 effect magnitudes for each item by subtracting the mean amplitude in the repeated label condition from that in the switched label condition over 300–600 and 600–1000 msec in the test phase. We then fit general linear models to the effect magnitudes with Speaker and Preference difference (a scaled continuous variable) as interacting predictors. As shown in Figure 3B and Table 4, the results confirmed the findings from our trial-level analyses, revealing a significant main effect of Speaker during 300–600 msec but not during 600–1000 msec. More importantly, neither the interaction between Speaker and Preference difference

Table 3. LME Models for Amplitude Analysis across the Centro-parietal Sites in the Test Phase

Predictor	β	<i>SE</i>	<i>t</i>	<i>p</i>
N400 (300–600 msec)				
Intercept	0.39	0.13	2.94	.005
Label	−0.49	0.07	−7.11	<.001
Speaker	0.05	0.25	0.19	.850
Label: Speaker	−0.40	0.14	−2.92	.004
P600 (600–1000 msec)				
Intercept	0.90	0.15	5.88	<.001
Label	0.57	0.12	4.62	<.001
Speaker	0.04	0.30	0.12	.908
Label: Speaker	−0.27	0.25	−1.08	.288

The model for the N400 analysis: $\text{N400} \sim \text{Label} * \text{Speaker} + (1|\text{Participant}) + (1|\text{Item})$; the model for P600 analysis: $\text{P600} \sim \text{Label} * \text{Speaker} + (\text{Label} + 1|\text{Participant}) + (1|\text{Item})$.

nor Preference difference alone predicted the effect magnitudes in either the 300- to 600 msec or 600- to 1000-msec time window. To further test the null effects of Preference difference and the interaction, we turned to Bayes factor (BF) analysis, which allowed us to determine how likely the null effects were true based on the observed data. Following Wagenmakers (2007; see also Wagenmakers, Verhagen, & Ly, 2016), we made use of the Bayesian information criterion (BIC) from a full regression model (here with Speaker and Preference difference as interacting predictors) and that from a regression model without the effect under examination (i.e., the main effect of Preference difference or the interaction between Speaker and Preference difference); we then used the difference in BICs ($\Delta\text{BIC} = \text{BIC}_{\text{full}} - \text{BIC}_{\text{reduced}}$) to compute the BF in support of the null hypothesis regarding an effect, using the formula $\text{BF} = e^{\Delta\text{BIC}/2}$ (see Cai, Pickering, Wang, & Branigan, 2015, for a similar application). The results showed that the models without the main effect of Preference difference were favored over the full models (N400: $\text{BF}_{01} = 6.14$; P600: $\text{BF}_{01} = 5.81$), and the models without the interaction between Speaker and Preference difference were favored over the full models (N400: $\text{BF}_{01} = 9.08$; P600: $\text{BF}_{01} = 11.29$). These results suggested that the label switching effects were not contingent on how people prefer the original labels over the alternative labels.

To further explore whether people's preferences toward each label would predict the ERP amplitudes in N400 and P600 time windows, we additionally analyzed the trial-level data in the establishment phase. The establishment-phase data of one participant from the adult speaker condition were excluded from this analysis because of the loss of EEG data during the establishment phase. We fit LME models to the trial-level amplitudes with Speaker and Preference rating (a scaled continuous

Table 4. General Linear Models for By-item Analysis of Preference Difference for Label Pairs

Predictor	β	SE	<i>t</i>	<i>p</i>
N400 (300–600 msec)				
Intercept	−0.53	0.08	−6.72	<.001
Speaker	−0.44	0.16	−2.83	.005
Preference difference	−0.09	0.08	−1.19	.236
Speaker: Preference difference	−0.13	0.16	−0.81	.422
P600 (600–1000 msec)				
Intercept	0.55	0.09	6.08	<.001
Speaker	−0.21	0.18	−1.17	.242
Preference difference	0.11	0.09	1.23	.219
Speaker: Preference difference	0.09	0.18	0.47	.638

Table 5. LME Models for Amplitude Analysis on Preference Ratings for Labels in the Establishment Phase

Predictor	β	SE	<i>t</i>	<i>p</i>
N400 (300–600 msec)				
Intercept	0.32	0.12	2.56	.014
Speaker	0.39	0.24	1.60	.117
Preference rating	0.14	0.06	2.44	.017
Speaker: Preference rating	−0.12	0.11	−1.08	.285
P600 (600–1000 msec)				
Intercept	0.26	0.14	1.89	.065
Speaker	−0.17	0.27	−0.63	.535
Preference rating	−0.06	0.06	−0.97	.336
Speaker: Preference rating	−0.16	0.11	−1.45	.146

The model for the N400 analysis: $\text{N400} \sim \text{Speaker} * \text{Preference rating} + (1|\text{Participant}) + (\text{Speaker} + 1|\text{Item})$. The model for the P600 analysis: $\text{P600} \sim \text{Speaker} * \text{Preference rating} + (1|\text{Participant}) + (1|\text{Item})$.

variable) as interacting predictors. As shown in Table 5, Preference rating significantly predicted the amplitudes during 300–600 msec but not during 600–1000 msec. Specifically, lower ratings were associated with more negative amplitudes (i.e., greater N400 effects). Furthermore, the interaction between Speaker and Preference rating was not significant for either 300–600 or 600–1000 msec. To further test the null effects of the interaction between Speaker and Preference rating, we, again, performed BF analyses and showed that the models without the interaction between Speaker and Preference rating were favored over the full models (N400: $\text{BF}_{01} = 121.32$; P600: $\text{BF}_{01} = 74.32$). These results suggested that although participants' preferences toward each label modulated the EEG amplitudes in the N400 time window (300–600 msec), this modulation effect remained comparable between the adult speaker condition and the child speaker condition.

DISCUSSION

Our study investigated how the demographic background of a speaker (an adult vs. a child) modulated listeners' electrophysiological responses when the speaker named a picture using an expression (a label) that differed from the one that they had used before (i.e., label switching). The experimental rationale was based on the common expectation that speakers tend to repeat their expressions for the same concept. Therefore, compared to the repeated labels, processing the switched labels would demand more cognitive effort from the listener, which should be reflected by larger EEG deflections. Furthermore, a child, with lower flexibility in word use than an adult, is expected to repeat expressions more frequently. If listeners integrate this perception of the speaker's characteristics during language processing, we should expect larger deflections when a child switches a

label than when an adult does. Consistent with our prediction, we demonstrated that switched labels (compared to repeated labels) elicited an N400 effect and a P600 effect. Crucially, the N400 effect was larger in the child speaker condition, despite P600 being comparable between the two speaker conditions. These results provide evidence for a modulation effect of speaker demographics on spoken word processing.

Time Course of Speaker Demographics Effects in Spoken Word Processing

Our results regarding label switching converge with time course evidence of previous studies. A meta-analysis of eye-tracking studies shows that the disruptive effect triggered by switched labels begins its trend at 200 msec and becomes reliable by 400 msec (Kronmüller & Barr, 2015). Similar results have also been reported by magnetoencephalography research showing that switched labels elicit an increase in the power of theta-band neural oscillations from 350 to 650 msec (Bögels et al., 2015). More interestingly, the time course of speaker demographics modulation on the label switching effect diverges from that reported by some existing studies. We revealed that the speaker effect occurs at an early stage, from 300 to 600 msec after the speaker articulated the label, coinciding with the main effect of label switching but contrasting with earlier findings of a later onset of the speaker effect. For instance, in an eye-tracking study that used a similar two-phase setup as ours, Kronmüller and Barr (2007) contrasted cases where labels were either repeated or switched by the original speaker versus a new speaker. They found that listeners' eye fixation patterns were comparable in the early moments of processing in both the original speaker condition and the new speaker condition. The speaker effect only emerged in a stage later than the main effect of label switching. Therefore, they proposed a hypothesis of "recovery from preemption," suggesting that when the speaker uses a new label to name the object, the mapping of the new label to the object is initially preempted by the original label, and listeners use the speaker information to inhibit the original label at a later stage.

There is an important difference between Kronmüller and Barr's (2007) study and our study though. Kronmüller and Barr (2007) contrasted different speakers with the same demographic background (Adult Speaker A vs. Adult Speaker B) who either had produced or had not produced a label (the original label) for a picture in the establishment phase. Thus, listeners tended to engage in reanalysis when a speaker broke the consistency in their word use and to search for a reason why the current speaker would switch to a new label for that object. It would be easier for listeners to reconcile this inconsistency if the new label is articulated by a new speaker who has little reason to know how this object was previously labeled by the original speaker. This reanalysis requires time and occurs after

the listener realizes the new label is in fact appropriate for the referent.

Conversely, our study contrasted speakers from different demographic backgrounds (an adult speaker vs. a child speaker), both of whom had produced a label for a picture in the establishment phase. Therefore, it targeted the speaker demographics effects, which emerged from listeners' life experiences with specific social groups. Given that speaker demographics are essentially world knowledge (Creel & Tumlin, 2011), when listeners encounter difficulty in integrating a speaker's demographic background and their language use, one would expect a type of neural response similar to semantic or world knowledge violations (Kutas & Federmeier, 2011; Hagoort et al., 2004; van Berkum et al., 1999). The speaker effect on the N400 we observed is consistent with studies that report an N400 effect when listeners encounter speech that violates their expectations based on speaker demographics (Pélissier & Ferragne, 2022; Martin et al., 2016; van den Brink et al., 2012; van Berkum et al., 2008). It is also consistent with studies showing that a speaker's social attributes modulate the N400 effects elicited by world knowledge violations (Grant et al., 2020; Bornkessel-Schlesewsky et al., 2013).

Speaker Modeling in Language Comprehension

Revisiting the comparison between the speaker model account and the acoustic detail account for the speaker effect in the Introduction, given that we made careful control for all audio stimuli, there is little reason to assume that the acoustic differences between the original labels and the new labels differ systematically between the two speaker conditions. Thus, according to the acoustic detail account, there should be no difference between the two speaker conditions in the label switching effect. Yet, we still observed larger label switching effects on N400 in the child speaker condition, which demonstrates a modulation mechanism that influences language processing from outside the acoustic-phonetic system. This mechanism is likely to be the modulating role of a speaker model that incorporates the speaker's demographic background including age, from which listeners may infer other attributes related to the current task such as linguistic ability (Cai et al., 2021; Suffill, Kutasi, Pickering, & Branigan, 2021; Branigan, Pickering, Pearson, McLean, & Brown, 2011). Our finding of the speaker effect on the N400 but not the P600 processing stage indicates that the speaker information is integrated with meaning at an early stage during language processing. This early integration is possible because a speaker model (e.g., of age) can be quickly constructed from the speaker's voice details, probably during the first few exposures to the audio (see also Cai et al., 2017, for a discussion), on top of the introduction of the speaker before the experiment began. Therefore, very early on in the experiment, listeners should have

already built a model of the speaker against which they interpreted labels.

Why do listeners experience more difficulty with label switching by children compared to adults? We have argued that listeners make use of the belief that children are linguistically less flexible than adults. This explanation indicates that the speaker's demographic background not only influences how listeners process the sentence message (e.g., van Berkum et al., 2008; Lattner & Friederici, 2003) and lexical variation (e.g., Martin et al., 2016) but also influences how they process the speaker's linguistic behavior such as label switching.

However, it is also possible that people expect that children, compared to adults, are less likely to use dispreferred labels in naming; hence, they experienced a larger "surprise" when a child used a dispreferred label in the test phase in our study than when an adult did. This account predicts an interaction between Preference rating and Speaker, which, however, was disconfirmed in the two post hoc analyses. In the first post hoc analysis, we showed that it is not the case that a more dispreferred or rare label (as compared to the preferred one) leads to a larger difference in the label switching effect between the child and adult speakers. This finding thus suggests that participants did not feel "more surprised" when the child, as compared to the adult, used a label (in the test phase) that had a larger difference in preference with its original counterpart (in the establishment phase). In the second post hoc analysis, we conducted a trial-level amplitude analysis on the establishment-phase data to test whether the label's preference rating (a continuous predictor) could predict brain potential amplitudes in listeners and whether it interacted with the speaker conditions. The results showed a significant main effect of Preference rating on the amplitudes for N400, which indicated that the less preferred a label is, the more cognitive effort is required to process it (Rugg, 1990; Van Petten & Kutas, 1990). However, we did not observe an effect of Speaker or an interaction between Speaker and Preference rating in either the N400 or P600 time window. These results suggested that when the child uttered a less preferred label (as compared among all labels in the establishment phase), it did not elicit a larger N400 effect than when the adult did so.

Label Comprehension across Contexts

Our research primarily investigated the comprehension of monologues from a speaker (see also Cai et al., 2017; Martin et al., 2016; van Berkum et al., 2008). Although monologue comprehension can be quite common in daily life (e.g., watching a video or listening to a podcast), they do not embody the dynamics of dialogues, which involve two or more individuals interacting with each other. For dialogues, the communication accommodation theory proposes that interlocutors converge their linguistic and communicative behavior toward each other, and

this proposal has received much empirical support (for a review, see Zhang & Giles, 2017). It might be reasonable to expect that the speaker demographics effect we observed in monologue comprehension may also apply to dialogue scenarios, with listeners being more "surprised" when a child interlocutor (compared to an adult interlocutor) switches a linguistic label for a concept, despite the possibility that interlocutors may adapt their label usage and recalibrate their expectation of label consistency for each other during interactions.

In addition, although we recruited native Mandarin Chinese speakers who all reported being originally from the Chinese mainland as participants (in both the pretest and the main experiment), they were not necessarily monodialectal or monolingual. Different labels might have different usage frequencies in different dialects, which might increase the variance of responses among participants. On this note, we used preference rating instead of word frequency to better capture the label using habits of the participant population (native Mandarin-speaking students studying in Hong Kong), in addition to the fact that preference rating should be a better indicator of how well a label fits the context of the object picture, whereas word frequency only reflects how often people use the word without a specific context. Nevertheless, it is also worth noting that, in the pretest, participants rated written labels according to how they would use the written label to name this object and might thus not take into account the speaker's demographic background (e.g., adult, child).

Furthermore, in daily scenarios, speakers switch labels in a discourse context that has been built up by the speaker's previous utterances (in monologues) or verbal interactions (in dialogues). In our study, the previous context was created in the form of a picture of an object paired with a label provided by the speaker. Although this context effectively formed an operational simulation of a speech scenario, it might not capture the full characteristics of label switching. Future research is encouraged to explore paradigms with more natural settings for better testing of label switching effects and the potential speaker effect on the processing of label switching.

Conclusion

Our study demonstrates that the demographic background of the speaker modulates listeners' neural correlates of spoken word processing. Specifically, it influences how listeners process the speaker's linguistic behavior of label switching. When a speaker refers to an object with a specific label, listeners expect the speaker to consistently use the same label. A switch to a less common label for the same object violates this expectation, leading to more substantial negative deflections in listeners' brain potentials. These deflections are modulated by the listeners' perception of the speaker's demographic background of age. Our finding contributes to a broader understanding of the interplay between social cognition and language processing.

APPENDIX

Table A1. Preference Ratings for Labels of Experimental Items

Item	Preferred Label				Dispreferred Label			
	Chinese Name		Rating	Chinese Name		Rating		
1	acne	痘痘	<i>dou₄dou</i>	6.39	粉刺	<i>fen₃ci₄</i>	4.09	
2	air conditioner	空调	<i>kong₁tiao₂</i>	6.78	冷气	<i>leng₃qi₄</i>	3.74	
3	bald	光头	<i>guang₁tou₂</i>	6.09	秃子	<i>tu₁zi</i>	4.00	
4	bandage	绷带	<i>beng₁dai₄</i>	5.48	纱布	<i>sha₁bu₄</i>	5.22	
5	banknote	钞票	<i>chao₁piao₄</i>	5.13	纸币	<i>zhi₃bi₄</i>	4.52	
6	bonsai	盆栽	<i>pen₂zai₁</i>	5.83	绿植	<i>lü₄zhi₂</i>	4.26	
7	boxed meal	盒饭	<i>he₂fan₄</i>	5.43	便当	<i>bian₄dang₁</i>	4.96	
8	broom	扫把	<i>sao₄ba₃</i>	6.13	笤帚	<i>tiao₂zhou</i>	5.17	
9	brush	毛笔	<i>mao₂bi₃</i>	5.78	画刷	<i>hua₄shua₁</i>	3.17	
10	building	大厦	<i>da₄sha₄</i>	5.00	楼房	<i>lou₂fang₂</i>	4.91	
11	bus	公交	<i>gong₁jiao₁</i>	5.87	巴士	<i>ba₁shi₄</i>	4.30	
12	card	扑克	<i>pu₁ke₄</i>	6.70	纸牌	<i>zhi₃pai₂</i>	4.57	
13	CD	光盘	<i>guang₁pan₂</i>	6.13	影碟	<i>ying₃die₂</i>	3.00	
14	cell phone	手机	<i>shou₃ji₁</i>	6.00	电话	<i>dian₄hua₄</i>	4.96	
15	cheese	奶酪	<i>nai₃lao₄</i>	5.96	芝士	<i>zhi₁shi₄</i>	5.74	
16	cloak	斗篷	<i>dou₃peng₂</i>	5.96	披风	<i>pi₁feng₁</i>	4.83	
17	coat	大衣	<i>da₄yi₁</i>	5.70	外套	<i>wai₄tao₄</i>	5.13	
18	cookie	饼干	<i>bing₃gan₁</i>	6.13	曲奇	<i>qu₃qi₂</i>	5.83	
19	corn	玉米	<i>yu₂mi₃</i>	6.52	苞谷	<i>bao₁gu₃</i>	3.48	
20	couple	情侣	<i>qing₂li₃</i>	6.26	恋人	<i>lian₄ren₂</i>	5.39	
21	crystal	水晶	<i>shui₃jing₁</i>	5.74	宝石	<i>bao₃shi₂</i>	4.26	
22	doctor	医生	<i>yi₁sheng₁</i>	6.78	大夫	<i>dai₄fu</i>	4.74	
23	doll	玩偶	<i>wan₂ou₃</i>	4.70	公仔	<i>gong₂zai₃</i>	3.74	
24	door bolt	插销	<i>cha₁xiao₁</i>	4.65	门闩	<i>men₂shuan₁</i>	3.13	
25	evening dress	长裙	<i>chang₂qun₂</i>	5.13	晚装	<i>wan₃zhuang₁</i>	3.30	
26	fence	围栏	<i>wei₂lan₂</i>	5.35	篱笆	<i>li₂ba</i>	4.43	
27	fireworks	礼花	<i>li₃hua₁</i>	4.09	彩炮	<i>cai₃pao₄</i>	3.13	
28	freezer	冰箱	<i>bing₁xiang₁</i>	5.96	冷柜	<i>leng₃gui₄</i>	3.61	
29	ghost	幽灵	<i>you₁ling₂</i>	5.74	鬼魂	<i>gui₃hun₂</i>	4.96	
30	grenade	手雷	<i>shou₃lei₂</i>	5.22	炸弹	<i>zha₄dan₄</i>	4.74	
31	hammer	锤子	<i>chui₂zi</i>	6.22	榔头	<i>lang₂tou</i>	4.30	
32	high-speed train	高铁	<i>gao₁tie₃</i>	6.35	动车	<i>dong₄che₁</i>	5.35	
33	hoodie	卫衣	<i>wei₄yi₁</i>	6.43	帽衫	<i>mao₄shan₁</i>	3.65	
34	hotel	酒店	<i>jiu₃dian₄</i>	5.83	宾馆	<i>bin₁guan₃</i>	4.48	
35	knife	小刀	<i>xiao₃dao₁</i>	5.91	匕首	<i>bi₃shou₃</i>	4.35	

Table A1. (continued)

Item	Preferred Label				Dispreferred Label			
	Chinese Name		Rating	Chinese Name		Rating		
36	lady	女士	nǚ ₃ shì ₄	5.78	小姐	xiao ₃ jiě ₃	4.48	
37	lawn	草坪	cao ₃ ping ₂	6.09	绿地	lǜ ₄ dì ₄	3.65	
38	lipstick	口红	kou ₃ hong ₂	6.30	唇膏	chun ₂ gao ₁	5.30	
39	locust	蚂蚱	ma ₄ zha	5.52	蝗虫	huang ₂ chong ₂	5.17	
40	man	男人	nan ₂ ren ₂	6.04	先生	xian ₁ sheng	5.26	
41	microphone	话筒	hua ₄ tong ₃	5.22	麦克	mai ₄ ke ₄	4.43	
42	monk	和尚	he ₂ shang ₄	6.74	僧人	seng ₁ ren ₂	4.04	
43	motorcycle	摩托	mo ₂ tuo ₁	6.48	机车	ji ₁ che ₁	3.13	
44	pepsi cola	可乐	ke ₃ le ₄	6.17	百事	bai ₃ shi ₄	4.96	
45	pill	胶囊	jiao ₁ nang ₂	5.65	药丸	yao ₄ wan ₂	5.48	
46	pineapple	菠萝	bo ₁ luo ₂	6.74	凤梨	feng ₄ li ₂	3.57	
47	police	警察	jing ₃ cha ₂	6.70	公安	gong ₁ an ₁	3.78	
48	popsicle	雪糕	xue ₃ gao ₁	6.00	冰棒	bing ₁ bang ₄	4.52	
49	professor	老师	lao ₃ shi ₁	6.22	教授	jiao ₄ shou ₄	4.35	
50	railway	铁路	tie ₃ lu ₄	5.61	轨道	gui ₃ dao ₄	5.39	
51	rat	老鼠	lao ₃ shu ₃	6.74	耗子	hao ₄ zi	4.09	
52	restaurant	餐厅	can ₁ ting ₁	6.13	饭店	fan ₄ dian ₄	5.43	
53	ring	钻戒	zuan ₄ jie ₄	5.70	指环	zhi ₃ buan ₂	3.13	
54	rubber band	皮筋	pi ₂ jin ₁	5.22	头绳	tou ₂ sheng ₂	4.39	
55	scarf	围脖	wei ₂ bo ₂	4.65	丝巾	si ₁ jin ₁	3.87	
56	shower head	花洒	hua ₁ sa ₃	6.13	喷头	pen ₁ tou ₂	4.17	
57	singer	歌手	ge ₁ shou ₃	5.96	明星	ming ₂ xing ₁	4.96	
58	sink	水槽	shui ₃ cao ₂	5.22	碗池	wan ₃ chi ₂	3.35	
59	socket	插座	cha ₁ zuo ₄	6.43	电源	dian ₄ yuan ₂	4.17	
60	soldier	士兵	shi ₄ bing ₁	5.74	军人	jun ₁ ren ₂	5.70	
61	speaker	音响	yin ₁ xiang ₃	6.52	喇叭	la ₃ ba	3.91	
62	spoon	勺子	shao ₂ zi	6.65	调羹	tiao ₂ geng ₁	3.26	
63	squid	鱿鱼	you ₂ yu ₂	6.30	乌贼	wu ₁ zei ₂	4.35	
64	staircase	楼梯	lou ₂ ti ₁	6.61	台阶	tai ₂ jie ₁	5.35	
65	steamed bun	馒头	man ₂ tou	5.87	蒸馍	zheng ₁ mo ₂	3.04	
66	stick	棍子	gun ₄ zi	5.09	木棒	mu ₄ bang ₄	3.96	
67	stool	板凳	ban ₃ deng ₄	4.65	马扎	ma ₃ zha ₂	4.22	
68	suit	西服	xi ₁ fu ₂	5.22	正装	zheng ₄ zhuang ₁	5.04	
69	sweet potato	红薯	hong ₂ shu ₃	6.13	地瓜	di ₄ gua ₁	4.35	
70	tableware	刀叉	dao ₁ cha ₁	6.43	餐具	can ₁ ju ₄	5.65	
71	tattoo	纹身	wen ₂ shen ₁	6.78	刺青	ci ₄ qing ₁	3.87	
72	taxi	的士	di ₁ shi ₄	5.70	出租	chu ₁ zu ₁	4.78	
73	thief	小偷	xiao ₃ tou ₁	6.48	盗贼	dao ₄ zei ₂	3.87	

Table A1. (continued)

Item	Preferred Label				Dispreferred Label			
	Chinese Name		Rating	Chinese Name		Rating		
74	toad	蛤蟆	<i>ba₂ma</i>	6.35	蟾蜍	<i>chan₂chu₂</i>	3.70	
75	toast	面包	<i>mian₄bao₁</i>	5.96	吐司	<i>tu₃si₁</i>	5.09	
76	toilet	马桶	<i>ma₃tong₃</i>	6.87	坐便	<i>zuo₄bian₄</i>	3.35	
77	towel	毛巾	<i>mao₂jin₁</i>	6.13	抹布	<i>ma₁bu₄</i>	4.70	
78	vest	马甲	<i>ma₃jia₃</i>	5.91	背心	<i>bei₄xin₁</i>	4.30	
79	wallet	钱包	<i>qian₂bao₁</i>	6.52	皮夹	<i>pi₂jia₁</i>	3.48	
80	wheel	车轮	<i>che₁lu₂</i>	5.30	轱辘	<i>gu₁lu</i>	3.61	

Acknowledgments

The authors thank Bei Xiao for her assistance during data collection.

Corresponding author: Zhenguang G. Cai, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, or via e-mail: zhenguangcai@link.cuhk.edu.hk.

Data Availability Statement

The stimuli and data set of this study are available at osf.io/2gvkt/.

Author Contributions

Hanlin Wu: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Visualization; Writing—Original draft; Writing—Review & editing. Xufeng Duan: Formal analysis; Resources; Software. Zhenguang G. Cai: Conceptualization; Funding acquisition; Supervision; Writing—Original draft; Writing—Review & editing.

Funding Information

This work was supported by the General Research Fund, grant number: 14600220 to Zhenguang G. Cai, University Grants Committee (<https://dx.doi.org/10.13039/501100001839>), Hong Kong.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549,

W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be M/M = .389, W/M = .315, M/W = .111, and W/W = .185.

Notes

1. In this article, we use “speaker effects” to refer to the effects elicited by either the speaker's individual identity or the speaker's demographic background. We use “speaker demographics effects” to specifically refer to the effects elicited by the speaker's demographic background.
2. To illustrate, the study by Martin and colleagues (2016) involved 45 participants (with 40 remaining after data exclusion), each participating in 19 trials per condition, culminating in a total of 760 trials for each condition. In contrast, our study included 23 participants in the adult speaker group (22 remaining after exclusion) and 25 in the child speaker group (24 after exclusion), with each participant engaging in 40 trials per condition. This resulted in 880 trials for the adult speaker group and 960 trials for the child speaker group per condition. As such, despite the apparent discrepancy in participant numbers, the overall volume of data collected in our study exceeds that of Martin and colleagues (2016) when considering the total number of trials per condition. In addition, it should be noted that most of the previous studies used a within-participant design (including Martin et al., 2016), whereas we used a mixed design.
3. Although the lack of difference between adult and child speech generated by AI provided reassurance that any potential effects of unnaturalness on the processing of the spoken labels should be consistent across both conditions, it should be noted that we did not include a human-generated speech baseline in the naturalness test.

REFERENCES

- Aurnhammer, C., Delogu, F., Brouwer, H., & Crocker, M. W. (2023). The P600 as a continuous index of integration effort. *Psychophysiology*, 60, e14302. <https://doi.org/10.1111/psyp.14302>, PubMed: 37042061

- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391–418. <https://doi.org/10.1006/jmla.2001.2815>
- Bögels, S., Barr, D. J., Garrod, S., & Kessler, K. (2015). Conversational interaction in the scanner: Mentalizing during language processing as revealed by MEG. *Cerebral Cortex*, *25*, 3219–3234. <https://doi.org/10.1093/cercor/bhu116>, PubMed: 24904076
- Bornkessel-Schlesewsky, I., Krauspenhaar, S., & Schlesewsky, M. (2013). Yes, you can? A speaker's potency to act upon his words orchestrates early neural responses to message-level meaning. *PLoS One*, *8*, e69173. <https://doi.org/10.1371/journal.pone.0069173>, PubMed: 23894425
- Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*, 41–57. <https://doi.org/10.1016/j.cognition.2011.05.011>, PubMed: 21723549
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493. <https://doi.org/10.1037/0278-7393.22.6.1482>, PubMed: 8921603
- Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*, 171–190. <https://doi.org/10.1016/j.jml.2009.04.003>, PubMed: 20161117
- Cai, Z. G. (2022). Interlocutor modelling in comprehending speech from interleaved interlocutors of different dialectic backgrounds. *Psychonomic Bulletin & Review*, *29*, 1026–1034. <https://doi.org/10.3758/s13423-022-02055-7>, PubMed: 35106731
- Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., et al. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, *98*, 73–101. <https://doi.org/10.1016/j.cogpsych.2017.08.003>, PubMed: 28881224
- Cai, Z. G., Pickering, M. J., Wang, R., & Branigan, H. P. (2015). It is there whether you hear it or not: Syntactic representation of missing arguments. *Cognition*, *136*, 255–267. <https://doi.org/10.1016/j.cognition.2014.11.017>, PubMed: 25506775
- Cai, Z. G., Sun, Z., & Zhao, N. (2021). Interlocutor modelling in lexical alignment: The role of linguistic competence. *Journal of Memory and Language*, *121*, 104278. <https://doi.org/10.1016/j.jml.2021.104278>
- Clapp, W., Vaughn, C., Todd, S., & Sumner, M. (2023). Talker-specificity and token-specificity in recognition memory. *Cognition*, *237*, 105450. <https://doi.org/10.1016/j.cognition.2023.105450>, PubMed: 37043968
- Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition*, *64*, 1–37. [https://doi.org/10.1016/S0010-0277\(97\)00010-3](https://doi.org/10.1016/S0010-0277(97)00010-3), PubMed: 9342930
- Clark, H. (1996). *Using language*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511620539>
- Coates, J. (2015). *Women, men and language: A sociolinguistic account of gender differences in language*. London: Routledge. <https://doi.org/10.4324/9781315645612>
- Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass*, *5*, 190–204. <https://doi.org/10.1111/J.1749-818X.2011.00276.X>
- Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*, 264–285. <https://doi.org/10.1016/j.jml.2011.06.005>
- Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*, 1893–1906. <https://doi.org/10.1162/jocn.2008.21142>, PubMed: 18855555
- Foucart, A., Garcia, X., Ayguasanosa, M., Thierry, G., Martin, C., & Costa, A. (2015). Does the speaker matter? Online processing of semantic and pragmatic information in L2 speech comprehension. *Neuropsychologia*, *75*, 291–303. <https://doi.org/10.1016/j.neuropsychologia.2015.06.027>, PubMed: 26115602
- Foucart, A., & Hartsuiker, R. J. (2021). Are foreign-accented speakers that 'incredible'? The impact of the speaker's indexical properties on sentence processing. *Neuropsychologia*, *158*, 107902. <https://doi.org/10.1016/j.neuropsychologia.2021.107902>, PubMed: 34052231
- Foucart, A., Santamaría-García, H., & Hartsuiker, R. J. (2019). Short exposure to a foreign accent impacts subsequent cognitive processes. *Neuropsychologia*, *129*, 1–9. <https://doi.org/10.1016/j.neuropsychologia.2019.02.021>, PubMed: 30858062
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166–1183. <https://doi.org/10.1037/0278-7393.22.5.1166>, PubMed: 8926483
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279. <https://doi.org/10.1037/0033-295X.105.2.251>, PubMed: 9577239
- Graham, S. A., Sedivy, J., & Khu, M. (2014). That's not what you said earlier: Preschoolers expect partners to be referentially consistent. *Journal of Child Language*, *41*, 34–50. <https://doi.org/10.1017/S0305000912000530>, PubMed: 23398907
- Grant, A., Grey, S., & van Hell, J. G. (2020). Male fashionistas and female football fans: Gender stereotypes affect neurophysiological correlates of semantic processing during speech comprehension. *Journal of Neurolinguistics*, *53*, 100876. <https://doi.org/10.1016/j.jneuroling.2019.100876>
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441. <https://doi.org/10.1126/science.1095455>, PubMed: 15031438
- Heise, M. J., Mon, S. K., & Bowman, L. C. (2022). Utility of linear mixed effects models for event-related potential research with infants and children. *Developmental Cognitive Neuroscience*, *54*, 101070. <https://doi.org/10.1016/j.dcn.2022.101070>, PubMed: 35395594
- Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*, 831–836. <https://doi.org/10.1016/j.cognition.2008.04.008>, PubMed: 18586232
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142. <https://doi.org/10.1016/j.cognition.2004.07.001>, PubMed: 15925573
- Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: The influence of speaker-specific memory associations on reference resolution. *Memory and Cognition*, *40*, 113–126. <https://doi.org/10.3758/s13421-011-0135-7>, PubMed: 21818693
- Jara-Ettinger, J., & Rubio-Fernandez, P. (2021). Quantitative mental state attributions in language understanding. *Science Advances*, *7*, eabj0970. <https://doi.org/10.1126/sciadv.abj0970>, PubMed: 34788100
- Kapnoula, E. C., & Samuel, A. G. (2019). Voices in the mental lexicon: Words carry indexical information that can affect access to their meaning. *Journal of Memory and*

- Language*, 107, 111–127. <https://doi.org/10.1016/j.jml.2019.05.001>
- Kim, J. (2016). Perceptual associations between words and speaker age. *Laboratory Phonology*, 7, 18. <https://doi.org/10.5334/labphon.33>
- Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, 56, 436–455. <https://doi.org/10.1016/j.jml.2006.05.002>
- Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, 83, 1–19. <https://doi.org/10.1016/j.jml.2015.03.008>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>, PubMed: 20809790
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205. <https://doi.org/10.1126/science.7350657>, PubMed: 7350657
- Labov, W. (2006). *The social stratification of English in New York City*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511618208>
- Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing—Evidence from event-related brain potentials. *Neuroscience Letters*, 339, 191–194. [https://doi.org/10.1016/S0304-3940\(03\)00027-2](https://doi.org/10.1016/S0304-3940(03)00027-2), PubMed: 12633885
- Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, 50, 2032–2043. <https://doi.org/10.1016/j.neuropsychologia.2012.05.002>, PubMed: 22595659
- Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 72–106). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511983689.004>
- Martin, C. D., Garcia, X., Potter, D., Melinger, A., & Costa, A. (2016). *Holiday or vacation?* The processing of variation in vocabulary across dialects. *Language, Cognition and Neuroscience*, 31, 375–390. <https://doi.org/10.1080/23273798.2015.1100750>
- Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. *Developmental Psychology*, 46, 749–760. <https://doi.org/10.1037/a0019657>, PubMed: 20604599
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49, 201–213. [https://doi.org/10.1016/S0749-596X\(03\)00028-7](https://doi.org/10.1016/S0749-596X(03)00028-7)
- Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., et al. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7, e33468. <https://doi.org/10.7554/eLife.33468>, PubMed: 29631695
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869. <https://doi.org/10.1155/2011/156869>, PubMed: 21253357
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition*, 25, 273–285. <https://doi.org/10.3758/BF03211283>, PubMed: 9184479
- Péllissier, M., & Ferragne, E. (2022). The N400 reveals implicit accent-induced prejudice. *Speech Communication*, 137, 114–126. <https://doi.org/10.1016/j.specom.2021.10.004>
- Piccin, T. B., & Blewitt, P. (2007). Resource conservation as a basis for the mutual exclusivity effect in children's word learning. *First Language*, 27, 5–28. <https://doi.org/10.1177/0142723707067544>
- Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology*, 70, 1–30. <https://doi.org/10.1016/j.cogpsych.2014.01.001>, PubMed: 24480453
- Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition*, 18, 367–379. <https://doi.org/10.3758/BF03197126>, PubMed: 2381316
- Shintel, H., & Keysar, B. (2007). You said it before and you'll say it again: Expectations of consistency in communication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 357–369. <https://doi.org/10.1037/0278-7393.33.2.357>, PubMed: 17352617
- Suffill, E., Kutasi, T., Pickering, M. J., & Branigan, H. P. (2021). Lexical alignment is affected by addressee but not speaker nativeness. *Bilingualism: Language and Cognition*, 24, 746–757. <https://doi.org/10.1017/S1366728921000092>
- Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, 4, 1015. <https://doi.org/10.3389/fpsyg.2013.01015>, PubMed: 24550851
- Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, 52, 997–1009. <https://doi.org/10.1111/psyp.12437>, PubMed: 25903295
- van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11, 657–671. <https://doi.org/10.1162/089829999563724>, PubMed: 10601747
- van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580–591. <https://doi.org/10.1162/jocn.2008.20054>, PubMed: 18052777
- van den Brink, D., van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., et al. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience*, 7, 173–183. <https://doi.org/10.1093/scan/nsq094>, PubMed: 21148175
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18, 380–393. <https://doi.org/10.3758/BF03197127>, PubMed: 2381317
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, 14, 779–804. <https://doi.org/10.3758/BF03194105>, PubMed: 18087943

Wagenmakers, E.-J., Verhagen, J., & Ly, A. (2016). How to quantify the evidence for the absence of a correlation. *Behavior Research Methods*, 48, 413–426. <https://doi.org/10.3758/s13428-015-0593-0>, PubMed: 26148822

Walker, A., & Hay, J. (2011). Congruence between ‘word age’ and ‘voice age’ facilitates lexical access. *Laboratory*

Phonology, 2, 219–237. <https://doi.org/10.1515/labphon.2011.007>

Zhang, Y. B., & Giles, H. (2017). Communication accommodation theory. In K. Y. Yun (Ed.), *The international encyclopedia of intercultural communication* (pp. 1–14). <https://doi.org/10.1002/9781118783665.ieicc0156>