International Journal of English Linguistics; Vol. 12, No. 6; 2022
ISSN 1923-869X E-ISSN 1923-8703
Published by Canadian Center of Science and Education

Effects of English Proficiency on Caucasian Face Gender Perception by Chinese-English Bilinguals: Evidence from ERP

Xinan Zhou[1] & Yifei Ji[2]

[1] School of International Studies, Zhejiang University, Hangzhou, China

[2] Faculty of English Language and Culture, Guangdong University of Foreign Languages, Guangzhou, China

Correspondence: Zhou Xinan, School of International Studies, Zhejiang University, Hangzhou, China. E-mail: 22005042@zju.edu.cn

Received: July 1, 2022 Accepted: August 16, 2022 Online Published: August 18, 2022

doi:10.5539/ijel.v12n6p1 URL: https://doi.org/10.5539/ijel.v12n6p1

Abstract

Chinese and English differ in the encoding of biological gender in the spoken forms of 3rd person singular pronouns. Linguistic relativity theories predict that structural differences across languages are accompanied with differences in non-linguistic cognition. However, the pronoun difference between the two languages seems so trivial that its influence on gender perception is unbelievable except with empirical support. The present study conducted an ERP experiment with native speakers of Chinese learning English as a foreign language and differing in English proficiency. The odd-ball paradigm was used to examine whether L2 proficiency would influence how these Chinese-English bilinguals perform on Caucasian face gender perception. The experiment yielded null effect of L2 proficiency on the vMMN that was elicited for the gender category, as well as the control age category. The results suggest that the difference in the pronoun encoding of biological gender between Chinese and English may not influence gender perception in the nonlinguistic context, although it is not surprising considering the triviality of such cross-linguistic difference and the widespread gender binary opposition in daily life.

Keywords: linguistic relativity, bilingual, L2 proficiency, gender perception, vMMN

1. Introduction

It has been well documented that the language we speak shapes our perception (see Wolff & Holmes, 2011; for recent review, see Athanasopoulos & Casaponsa, 2020). Although cross-linguistic similarities do not always mean similar perception for speakers of different languages and differences do not necessarily create different perception, ample research has shown that cross-linguistic differences may exert a significant impact on non-linguistic cognition, for example, color (e.g., Athanasopoulos et al., 2010a, b) and number (e.g., Athanasopoulos, 2006). One area where languages may differ is the use of pronouns. For example, for Chinese 3rd person singular pronouns, the biological gender information is realized only in their written forms (i.e., '他' vs. '她') but not the spoken form (i.e., tā). While for their English counterparts, the biological gender information has both orthographical and phonological realizations (i.e., he vs. she). Given these cross-linguistic differences, a question arises: Does the subtle linguistic Chinese-English difference in the pronoun encoding of biological gender information affect speakers' gender perception in a non-linguistic context? Since the absence of gender information encoding only occurs in 3rd person singular pronouns in oral Chinese, its influence on gender perception requires empirical support.

A few recent studies have investigated the effect of such a trivial cross-linguistic difference in language processing. Firstly, the difference has been proven to influence the sensitivity to gender information encoded in discourses between Chinese and English native speakers (Chen & Su, 2011). Chen and Su (2011) compared the performance of Chinese and English native speakers in two experiments presented in their L1 respectively. In the first experiment, participants were asked to listen to stories and then answer gender and non-gender related questions. In the second experiment, they were asked to read a sentence first and then choose which of the presented two pictures matched the sentence. The results revealed that the Chinese natives had worse performance than their English counterparts in terms of RT and accuracy when answering gender-related questions or when choosing the correct picture for the gender-related sentences, while they did not differ in the gender-irrelevant conditions. The findings, as the authors put it, seem to support linguistic relativity. That is, the linguistic device associated with a specific language orients its speakers to a particular aspect of the world and results in increased sensitivity to that aspect for the purpose of speaking. Therefore, the lack of gender marker in oral Chinese pronouns probably resulted in Chinese native speakers' relative insensitivity to gender information in language processing.

Secondly, the difference may also impact the L2 pronoun gender agreement processing, which is supported by behavioral (e.g., Antón-Méndez, 2011; Dong & Jia, 2011; Dong et al., 2015) and ERP studies (e.g., Liang et al., 2018; Yu & Dong, 2018, 2019). Findings from behavioral research have often been accounted for by the assumption of L1 transfer, which suggest that when composing L2 preverbal message containing person pronouns, the speaker whose native language employs different gender marker in pronouns may follow their L1-specific procedure and neglect whatever gender information in the L2 discourse or situation context. In other words, the stage of "thinking for speaking" (Slobin, 1996) may be influenced by the speaker's native language. This hypothesis is partly supported by Dong et al. (2015), who conducted two self-paced reading experiments by Chinese-English speakers and investigated whether the reading time for each 'he' or 'she' that matched its antecedent was shorter than that in the corresponding mismatch situation (the mismatch effect). The result is that the mismatch effect was present only when the gender information of the antecedent was enhanced by a human picture. The authors concluded that Chinese native speakers may not habitually process gender information for linguistic purposes, and the mixed use of 'he' and 'she' is probably a result of insufficient processing of gender information. ERP studies (e.g., Liang et al., 2018; Yu & Dong, 2018, 2019) suggest distinct neurocognitive mechanisms being applied to the processing of gender agreement in L1 and L2. Liang et al. (2018) compared participants' ERP responses to L1 and L2 pronouns that were either congruent or incongruent with their antecedents. An LAN-P600 pattern for gender incongruity was observed in L1, while an Earlier Positivity-P600 pattern in L2, which indicates significant variance when bilinguals process L1 and L2 gender incongruity.

Although most studies on cross-linguistic difference in pronoun encoding of gender information concern language processing, corresponding issues concerning non-linguistic cognition are far from clear. To our knowledge, no research up to date has directly investigated *the effect of cross-linguistic difference in the pronoun encoding of gender information on speakers' gender perception in a non-linguistic context*. But the hypothesis may be partly supported by studies on grammatical gender (e.g., Boroditsky et al., 2003; Bassetti, 2014; Cubelli et al., 2011; Maciuszek et al., 2019). The general finding of these studies is that grammatical gender affects speakers' cognition. For example, it influences the way they describe objects: the noun 'key' is marked as feminine in Spanish and is thus often described as 'golden, intricate, little, lovely, shiny, and tiny' by Spanish speakers, while in German it is marked as masculine and is thus often described as 'hard, heavy, jagged, metal, serrated, and useful' by German speakers (Boroditsky et al., 2003). If cross-linguistic difference in the grammatical encoding of gender information affects speakers' perception, the effect of difference in pronoun encoding of gender information also deserves investigation.

Although the concept of linguistic relativity is not new, the exploration of the modulating effect of such a subtle cross-linguistic difference on gender perception is worthwhile. First, the present study looks beyond linguistic level and it is the first attempt to explore the effect of cross-linguistic difference on the pronoun gender processing in the non-linguistic context. Since recent studies have confirmed the effect of this cross-linguistic differences in language processing, an interesting question then arises as to whether such a difference would be expanded to impact the non-linguistic human perception from the perspective of linguistic relativity. The male-female distinction, an important part of the human being's identity, would be directly and deeply influenced and thus deserves investigation. Second, cross-linguistic differences studied in the field of linguistic relativity generally involve the presence of linguistic encoding for a certain concept in one language and the absence of it in another language. However, the targeted encoding difference in the present study is not a simple "presence vs. absence" contrast because the absence of gender information encoding only occurs in 3rd person singular pronouns in oral Chinese. Moreover, both languages have various pairs of lexical units with minimal distinction of biological gender such as boy vs. girl, father vs. mother, Mr. vs. Mrs., etc. The current exploration is like using "the weakest language spear" to pierce "the strongest concept shield". If there is evidence for linguistic relativity in the concept of human gender, the power of language may be greater than what we have expected.

The vMMN of Event-related Potentials. Visual mismatch negativity (vMMN) is a negative-going component of event-related potentials (ERPs) which is usually elicited at posterior sites and peaks around 150–400 ms after the onset of visual stimuli. The most commonly used experimental protocol is the passive oddball paradigm in which vMMN is elicited outside the focus of active attention by infrequent stimuli (deviants) embedded in a sequence of frequent stimuli (standards). Since vMMN is a type of mismatch response, it is obtained by subtracting the ERPs to the standards from the ERPs to the deviants. With regard to language relativity research, Athanasopoulos et al. (2010b) and Thierry et al. (2009) used a more direct method to investigate English and Greek speakers' perception of color and compared the amplitudes of the vMMN elicited by green/blue chunks of deviant luminance. Thierry et al. (2009) found that Greek speakers whose language lexicalizes the light and dark blue by two specific terms, *ghalazio* and *ble*, demonstrated greater and faster perceptual discrimination of the blue chunks of different luminance than English speakers. In a follow-up study, Athanasopoulos et al. (2010b) evenly split the Greek speakers that participated in the study of Thierry et al. (2009) into two groups based on the amount of time they had lived in the UK and found that the length of stay in the L2 country affected early perceptual processing in Greek-English bilinguals. Mo et al. (2011) adopted a less direct method in which the stimuli were presented simultaneously in the right and left visual fields and found that the category-related effect of vMMN only occurred for the stimuli presented in the right visual field, which was interpreted as another evidence for the language effect on color categorization because stimuli presented in the right visual field are processed in the left hemisphere where most language-related processes occur.

The present study used the direct method to investigate language effects on gender perception as in Thierry et al. (2009) and the concept of gender was represented by male-female face pictures. What is different from Thierry et al. (2009) in methodology is that it is not plausible to ask native speakers of the two languages to judge the same face pictures because there may exist "the other-race effect" (see Meissner & Brigham, 2001, for a review). That is, people are generally more sensitive to faces of their own race than to faces of another race. A plausible way is, therefore, to compare groups of native Chinese speakers learning English as a foreign language and differing in English proficiency, and to explore their perception of same-race faces (i.e., Caucasian faces in the present study). What is more, participants' gender might be another confounding factor, because people make a faster judgment about human faces that are of the same gender as their own (Zarate & Smith, 1990), and males are more likely to have a male-biased reading of masculine generic role nouns (Gabriel & Mellenberger, 2004). Thus, the present study recruited the same number of male and female participants and counterbalanced the effect of gender in the experiment procedure. **The hypothesis** is that with higher L2 (English) proficiency, native speakers of Chinese may become more sensitive to the biological gender of Caucasian faces because of their frequent need to distinguish the male-female pair to pick up the right pronoun.

2. Method

The experiment aimed to examine how two groups of Chinese-English bilinguals of different English proficiency perform in the gender perception of Caucasian faces. The odd-ball paradigm was used, with gender as the target variable (male and female faces), and age as the control variable (young and old).

2.1 Participants

Thirty-six Chinese-English bilingual students from Guangdong University of Foreign Studies participated in the experiment for monetary compensation. Twenty of them were first-year postgraduates majoring in English interpreting or English translation (9 female, 9 male), who made up the group of high L2 proficiency. The others were first-year undergraduates majoring in Chinese (9 female, 9 male), who made up the group of low L2 proficiency. Relevant background information was collected by a short-version language history questionnaire (Li et al., 2006). All the participants had normal or corrected-to-normal vision. All of them signed a written consent after the experiment had been fully explained. Table 1 is a summary of the critical details.

Table 1. Chinese-English bilingual participants' language background (means with SDs in brackets)

	High L2 group N = 18 (9 male)	Low L2 group N = 18 (9 male)	*t* value
L2 proficiency	21.06/28 (2.36)	16.11/28 (2.30)	6.364***
Daily L2 use (hours)	8.58 (4.74)	2.82 (1.94)	4.761***
Daily L1 use (hours)	6.64 (4.24)	9.36 (3.72)	-2.049*
Formal L2 instruction (years)	11.83 (2.28)	11.72 (1.84)	
AOA for L2	11.00 (2.50)	6.89 (2.11)	5.334**

Note. 1) *p < .05, **p < .01, ***p < .001; 2) Means of L2 proficiency are scores out of the total score of 28 (7 points for each of the four commons skills of listening, speaking, reading and writing); 3) AOA: age of acquisition; 4) no significance in blank cell.

As Table 1 shows, the two participant groups differed in the critical factor of L2 proficiency ($t_{(34)} = 6.364$, $p < .001$). As for daily L2 use and daily L1 use, they are the two sides of the same coin, and the between-group differences in these factors were consistent with that in L2 proficiency. That is, more proficient L2 speakers were generally speakers who used the L2 more frequently, especially for our participants who spent most of their time learning either Chinese or English. As to L2 acquisition age, the two groups differed ($t_{(34)}$ = 5.334, p = .003), which to some extent explains why there was no significant difference in formal English instruction between groups. Since our hypothesis was that Chinese-English bilinguals of higher L2 proficiency may become more sensitive to face gender perception probably because of their frequent need to distinguish the male-female pair to pick up the right pronoun, we took for granted that L2 proficiency (generally a consequence of more frequent L2 use) played a more critical role than the other factors.

2.2 Materials

The stimuli were 60 Caucasian faces (20 young male, 20 young female, 10 old male and 10 old female) selected from the face database employed by Zhao and Bentin (2008) in which all the faces were grayscale and equated for luminance, brightness and contrast. As in Kecskes-Kovacs et al. (2013), a dark grey cropping mask (size: 1024 × 768 pixels) was created and applied to each face so as to eliminate the physical individual differences concerning hair style and face shape (see Appendix A for a few examples of the facial stimuli in the experiment).

Because the cropped faces lost certain prominent features indicating gender (e.g., hair style), it is necessary to ensure their gender saliency which is one of the basic requirements to evoke gender-related vMMN. Apart from that, it is also important to examine the facial emotion of each face because the difference in facial emotion can also lead to vMMN (e.g., Stefanics et al., 2012). Although the neutral emotion of each selected face had been confirmed by the author, it is necessary to check through a test whether the emotional expressions have been balanced between standards and deviants, the difference between which is supposed to be the pure gender-related effect. Taking these two aspects into consideration, a gender discriminability test and a facial emotion assessment test were conducted.

The two tests recruited respectively 17 undergraduate students from Guangdong University of Foreign Studies who did not participate in the formal experiment. In the gender discriminability test (an active two-stimulus oddball experiment), participants were asked to judge whether the pictures they saw were male or female, while in the facial emotion assessment test, they were instructed to score the emotional expression for each presented face based on a three-point scale: negative facial expression (1st value), neutral facial expression (2nd value) and positive facial expression (3rd value). Among the faces assessed, altogether 10 male and 10 female faces for gender category, as well as another 20 male (10 young, 10 old) and 20 female (10 young, 10 old) for age category were selected. For the 20 faces selected for the gender category, the mean hit rate was 96.88% (SD = 0.01), and the independent samples t-test yielded no significant difference in emotion-related scores between male and female faces ($t_{(18)}$ = -.616, p = .546). For either the male or female version of the age category, young and old faces were comparable in emotion-related scores ($t_{(18)}$ = -.714, p = .484 for male faces, and $t_{(18)}$ = -1.206, p = .243 for female faces).

2.3 Procedure

The experiment was conducted in the Bilingual Cognition and Education Lab. Participants were seated in front of a 17” LCD (Lenovo L1710D, 60-Hz refresh rate) with a 1.2 m viewing distance. Stimuli were flashed for 300 ms with an inter-stimulus interval of 400 ms. There were altogether four blocks, each of which had 300 trials. Two of the blocks were critical ones consisting of the repetitions of 20 young faces for the gender category (10 male and 10 female). The other two were control blocks comprised of the repetitions of 10 old and 10 young faces, with half of the participants viewing only male faces, and the other half only female faces. See Figure 1 for the block samples. Within each block, only one type of stimuli was frequent: gender standard (male or female faces, 80%) or age standard (old or young faces, 80%) and the other type of stimuli was infrequent: gender deviant (faces of gender opposite to that of standards, 20%) or age deviant (faces of age opposite to that of standards, 20%). The deviants were ordered pseudo-randomly in presentation, with randomly two to nine standards between two deviants. The successive stimuli were never physically identical.

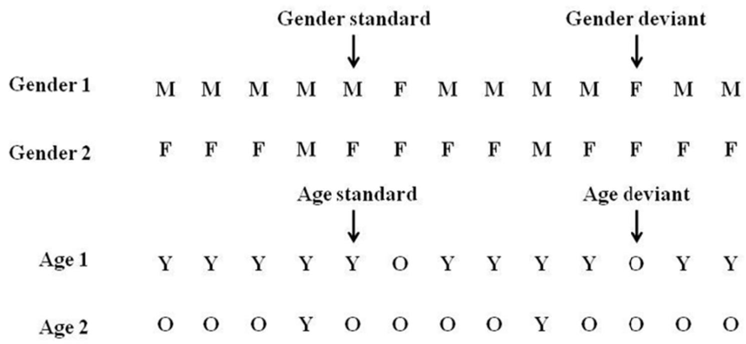


Figure 1. Examples of experimental blocks

Note. Gender 1, Gender 2, Age 1 and Age 2 represent the four blocks each participant viewed in the experiment. “M” and “F” refer to a young male or female face; “Y” and “O” refer to a young or old face (with half of the participants viewing male faces and the other half, female faces).

Similar to the study conducted by Kecskes-Kovacs et al. (2013), participants were instructed to perform a simple visual detection task in which they should detect the change of a cross which lasted for a random interval between 4 and 8 seconds. The target cross was presented in a grey square subtending 0.67° of visual angle which was located in the foreground at the center of the screen. The cross was comprised of two lines one of which may be shorter than the other. A response was required once the lengths of the two lines were reversed. Participants were asked to respond by pressing a button as quickly and as correctly as possible. See Figure 2 for the procedure.

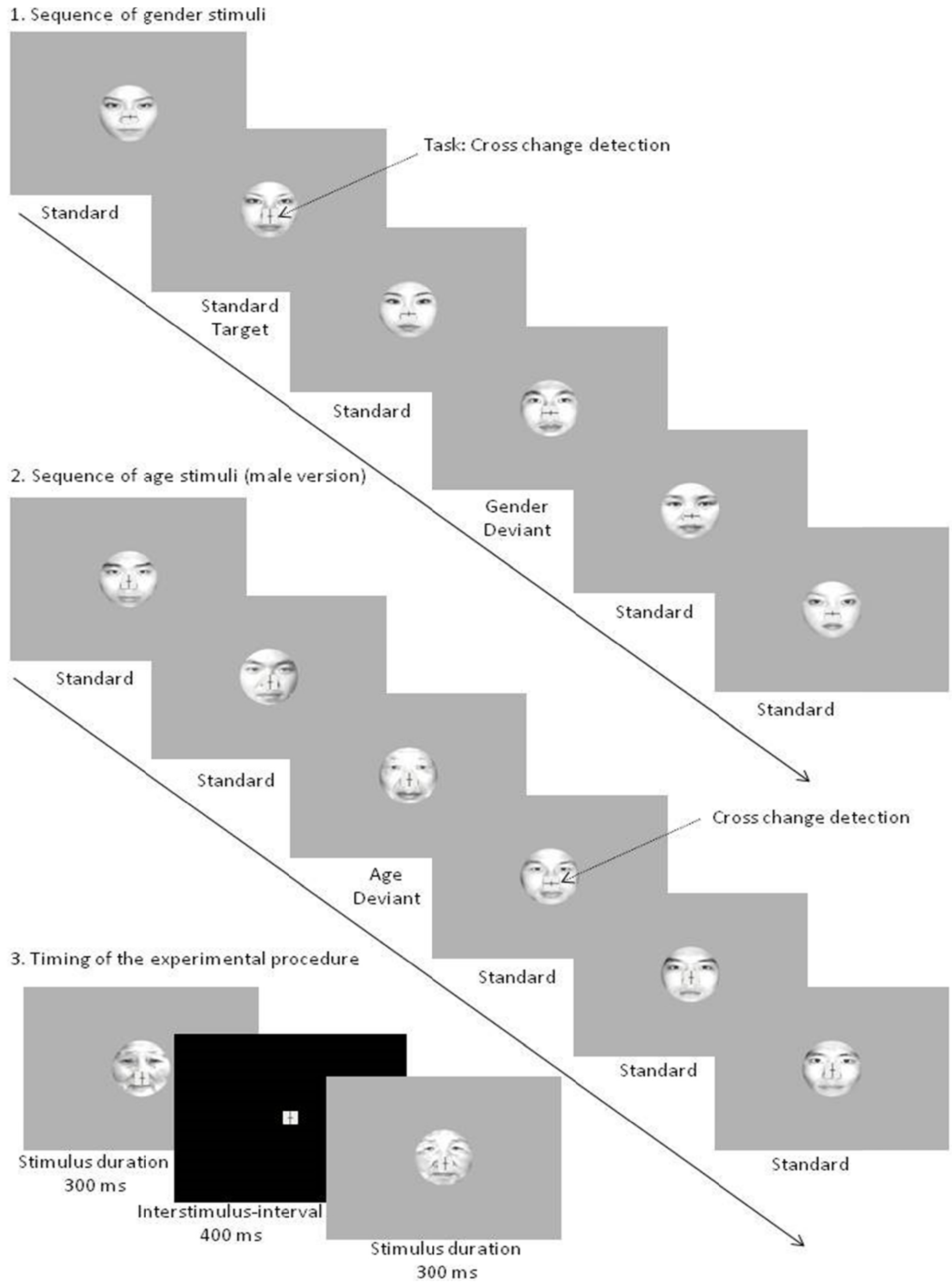


Figure 2. Examples of Gender-related and Age-related sequences and timing of the experimental procedure

2.4 EEG Recording and Offline Processing

Electroencephalogram (EEG) was continuously recorded (band pass 0.05–100 Hz, sampling rate 1000 Hz) with Neuroscan DC Amplifier, using an electrode cap with 64 Ag/AgCl electrodes mounted according to the extended international 10–20 system. The vertical electrooculogram (VEOG) reflecting eye-blinks and the horizontal electrooculogram (HEOG) indexing horizontal eye movements were recorded with two pairs of electrodes, one placed above and below the left eye, and the other 1 cm from the bilateral canthi. The EEG recording was referenced online to the tip of the nose. Impedances were kept below 10 kΩ.

EEG signals affected by artifacts including HEOG, drifting, and myoelectricity were manually rejected. VEOG artifacts were corrected using a function provided by the Neuroscan software. Then a digital low pass filter of 30 Hz (24 dB/octave) was applied to the EEG signals. After that, the EEG signals were segmented in epochs of 800 ms, time-locked to faces' onset and including a 100 ms pre-stimulus interval. The mean voltage of the pre-stimulus interval was used as the baseline to correct the segments, after which the segments with an amplitude change exceeding ±70 µV on any EEG channel were rejected from further analysis. The offline processing resulted in a data loss of 7.5% of the trials.

3. Results

For behavioral performance, the mean hit rate for all the participants in the task was 94.35% (SD = 4.57) and the mean RT was 427.84 ms (SD = 32.65). There was no difference in performance between conditions or between groups. Before analyzing the effects of L2 proficiency on face gender perception, we had to make sure that the vMMN component of event-related potentials was successfully elicited for the gender and age blocks (i.e., effects of stimulus type in the present design).

3.1 Effects of Stimulus Type

To examine whether the component of vMMN was successfully elicited for the gender and age categories across five ROIs, comparisons of ERP responses to deviants and standards (effects of stimulus type) were conducted at the intervals of 150–250 ms and 300–400 ms respectively for all the participants.

Comparison at the interval of 150–250 ms. The ANOVA of the amplitude values at the 150–250 ms interval with repeated-measures factors of Stimulus Type (standard vs. deviant), ROI (left temporal vs. left occipital vs. middle occipital vs. right temporal vs. right occipital), and Category (gender vs. age) yielded no main effect of Stimulus Type ($F_{(1, 35)} = .097$, $p = .757$, $\eta^2 = .003$). The two-way interaction between Stimulus Type and ROI was significant ($F_{(4, 140)} = 4.407$, $p = .009$, $\eta^2 = .112$, $\varepsilon = .637$). The interaction between Stimulus Type and Category approached significance ($F_{(1, 35)} = 3.267$, $p = .079$, $\eta^2 = .085$). Besides, significant main effects of ROI and Category were observed ($F_{(4, 140)} = 37.014$, $p < .001$, $\eta^2 = .514$ for ROI; $F_{(1, 35)} = 4.866$, $p = .034$, $\eta^2 = .122$ for Category). The interaction between ROI and Category was significant ($F_{(4, 140)} = 4.613$, $p = .003$, $\eta^2 = .116$, $\varepsilon = .835$). All the others were not significant.

Simple effect analyses were conducted to probe the two significant interactions involving Stimulus Type. For the Stimulus Type × ROI interaction, the results of simple effect analysis show that except for the Stimulus Type effect in the right occipital ROI ($F_{(1, 35)} = 2.36$, $p = .134$), those in the other ROIs didn't approach significance (Fs < 1, ps > .1). For the Stimulus Type × Category interaction, the results show that the Stimulus Type effect approached significance only for the age category ($F_{(1, 35)} = 2.99$, $p = .093$) and no significant effect of Stimulus Type was observed for the gender category ($F_{(1, 35)} = .93$, $p = .341$). To examine the Stimulus Type effect in each ROI for both categories, a planned simple effect analysis for Category × Stimulus Type × ROI was carried out (see Table 2 for details). *The results show that the component of vMMN was only successfully elicited for the age category in bilateral occipital ROIs.*

Comparison at the interval of 300–400 ms. The same ANOVA as conducted for the previous time window yielded a marginally significant main effect of Stimulus Type ($F_{(1, 35)} = 3.030$, $p = .091$, $\eta^2 = .080$) for the 300-400 ms interval. The two-way interaction between Stimulus Type and ROI was significant ($F_{(4, 140)} = 5.782$, $p = .003$, $\eta^2 = .142$, $\varepsilon = .588$). Besides, a main effect of ROI was observed ($F_{(4, 140)} = 16.012$, $p < .001$, $\eta^2 = .314$). All the others were not significant (ps > .10). To examine the Stimulus Type effect at each ROI for both categories, a planned simple effect analysis for Category × Stimulus Type × ROI was carried out (see Table 2 for results). The results show that *for the age category, the Stimulus Type effect was significant in the bilateral occipital ROIs while for the gender category, the Stimulus Type effect only approached significance in the right occipital ROI.*

Table 2. Simple effect analysis on the effect of Stimulus Type (standard vs. deviants) for Gender and Age categories across five ROIs (Experiment 2).

ROIs	150–250 ms interval		300–400 ms interval	
	Gender	Age	Gender	Age
Left occipital		4.95*		5.30*
Left temporal				
Middle occipital				
Right occipital		8.58**	3.33△	7.23*
Right temporal				

Note. △p <.10; *p < .05, **p < .01, ***p < .001; no significance in the blank cell.

To sum up, the vMMN component was successfully elicited for the age category in the bilateral occipital ROIs at both 150–250 ms and 300–400 ms intervals. As to the gender category, the Stimulus Type effect was only marginally significant in the right occipital ROI at the later time window. It therefore could be concluded that for Caucasian faces, the vMMN for gender was not elicited. Since there was a marginal significance for the right occipital ROI at the 300–400ms interval for gender, further analysis for the participant groups at this later time window would be conducted in case there might be some effect of participant groups.

3.2 Effects of Participant Groups

To probe the L2 effect on the Deviant Gender-minus-Standard Gender difference potentials, comparisons of ERP responses between the high and low L2 groups were conducted. Figure 3 shows the grand-average waveforms for both Gender and Age categories from the two participant groups in two of the five ROIs which had elicited some vMMN for either gender or age category.

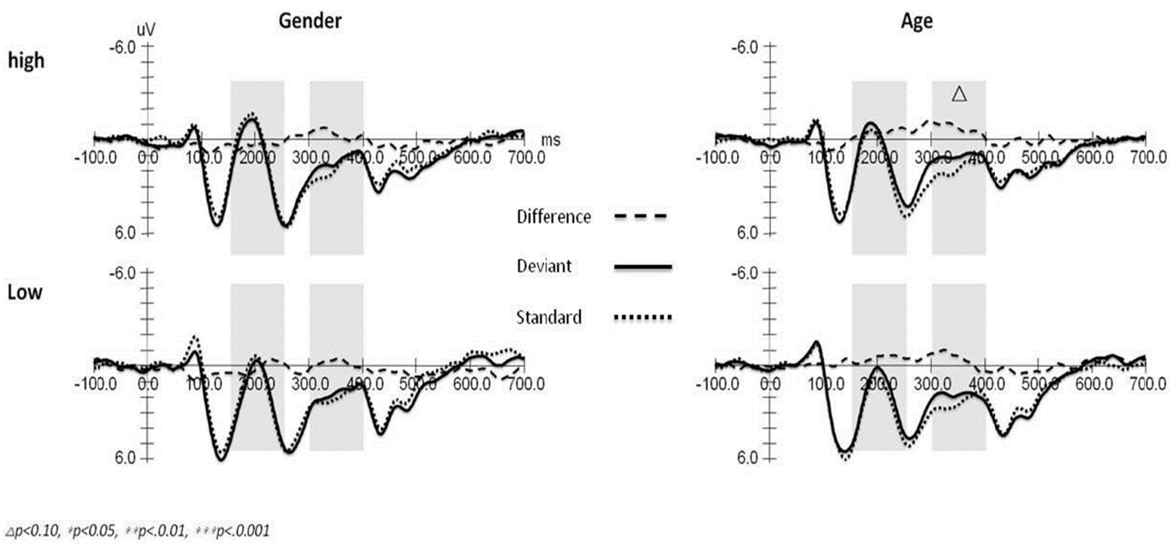


Figure 3a. Averaged ERP responses from high and low L2 groups for Gender and Age categories in the left occipital ROI, with the triangle indicating some marginally significant vMMN

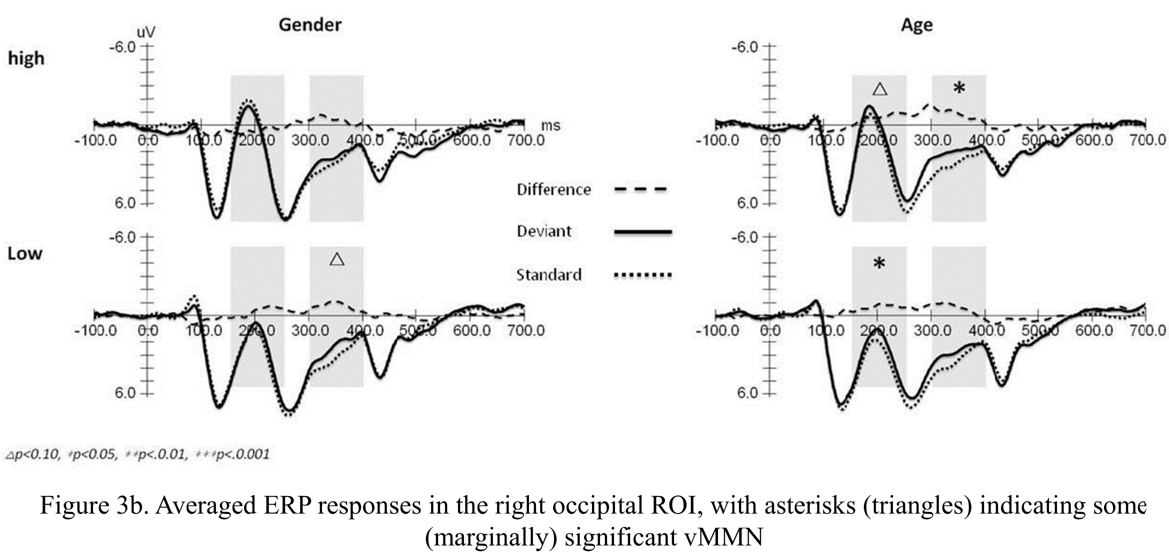


Figure 3b. Averaged ERP responses in the right occipital ROI, with asterisks (triangles) indicating some (marginally) significant vMMN

Comparison at the interval of 300–400 ms (Note 1). The mixed ANOVA of the amplitude values at the 300–400 ms interval with factors of Stimulus Type (standard vs. deviant), Category (gender vs. age), and L2 proficiency (high vs. low) yielded no significant Stimulus Type × Category × L2 Proficiency interaction in any ROI (Fs < 1, ps > .1).

To confirm whether there was no L2 effect on the amplitudes of Deviant Gender-minus-Standard Gender difference waveforms, mixed ANOVAs with factors of Stimulus Type, ROI, and L2 proficiency were performed for the two categories separately. *Neither the two-way interaction of Stimulus × L2 nor the triple interaction of Stimulus × ROI × L2 approached significance for either category* (Stimulus × L2: $F_{(1,34)} = .003$, $p = .959$, $\eta^2 < .001$ for gender, and $F_{(1,34)} = .852$, $p = .363$, $\eta^2 = .024$ for age; Stimulus × ROI × L2: $F_{(4,136)} = 1.665$, $p = .185$, $\eta^2 = .047$, $\varepsilon = .667$ for gender; and $F_{(4,136)} = 1.396$, $p = .253$, $\eta^2 = .039$, $\varepsilon = .589$ for age). That is, *there was no significant difference between the two participant groups in their perception of either gender or age in any of the ROIs in either time window.*

To sum up, with Caucasian faces, the experiment failed to find the L2 effect on the Gender vMMN. The primary cause is probably the predicted failure to elicit the Gender vMMN in Caucasian faces: only a marginally significant effect of Stimulus Type was observed for the gender category in the right occipital ROI at the interval of 300–400 ms.

4. Discussion

The present study aimed to examine whether Chinese-English bilinguals of different L2 proficiency show differential sensitivity to the perception of biological gender in non-linguistic context. With the odd-ball paradigm, the experiment explored this issue with Caucasian faces and found that participants of higher L2 proficiency were not more sensitive to the male-female distinction than participants of lower L2 proficiency. What's more, almost no significant vMMNs were elicited for gender perception, and only a weak vMMN was elicited in two of the five ROIs for age perception. The results suggest that, for native speakers of Chinese learning English as a foreign language, English proficiency may not be associated with perception sensitivity of biological gender, at least for the perception of Caucasian faces.

Compared with previous studies on color perception (e.g., Athanasopoulos et al., 2010b; Thierry et al., 2009), the present study failed to obtain the language effect although similar experimental designs and the same ERP component were adopted. This is reasonable if two aspects are taken into consideration. First, compared with the dark/light distinction within the concept of color, the male/female distinction within the concept of gender is much more difficult to be modulated by language because as long as we are in touch with the outside world, our sensory organs are occupied with various instantiations of the male/female distinction, such as voices, faces, clothes and so on. In a world where the gender binary opposition is so deeply rooted, the effect of a trivial cross-linguistic difference on gender perception cannot be strong. Second, the targeted cross-linguistic encoding difference for the concept of color is at the lexical level while that for the concept of gender is at the mere phonological level. The latter is much subtler than the former. To sum up, given the two differences, the null effect of language obtained in our study is within expectation.

The present study has made a major contribution to the literature. The difference between the present study and similar previous studies on color perception is the critical statistical interactions on which the major conclusions were based. The most critical interaction, in Athanasopoulos et al. (2010b) and Thierry et al. (2009) for example, is the interaction of Stimulus Type (standard vs. deviant) × Color (blue vs. green) which was significant for one group of participants (i.e., Greek natives/Short-stay bilinguals) but not for the other (i.e., English natives/Long-stay bilinguals). The exact meaning of this interaction is in fact that Greek natives or short-stay bilinguals were more sensitive to the dark/light distinction in blue than in green while English natives or long-stay bilinguals did not have such differential sensitivity regarding blue and green. It would be better if there was evidence supporting the assumption that the two participant groups' sensitivity to the dark/light distinction in green was not different. In the present study, however, the most critical interaction was the interaction of Stimulus Type (standard vs. deviant) × L2 Proficiency (high vs. low) which was predicted to be significant for the concept of gender but not for the concept of age. The exact meaning of this interaction is that for the concept of gender, participants of different L2 proficiency had differential sensitivity to the perception of gender, while this was not true for the concept of age. A very important reason for our choice of this interaction is that the contrast of color is quite different from the contrast of gender and age. In the studies on color perception (e.g., Athanasopoulos et al., 2010b; Thierry et al., 2009), blue standards/deviants were comparable with green standards/deviants in chroma (color purity) and hue based on the Munsell color system. The difference between standards and deviants was the same in luminance for green and blue stimuli (Note 2). But in the present study, neither the male/female distinction nor the young/old distinction could not be as nicely graded, and therefore the two distinctions were not as perfectly comparable. As shown in the examination of the Stimulus type effect in the experiment, the vMMN elicited for the age category was much larger than that elicited for the gender category, which means that the interaction of Stimulus type × Category (gender vs. age) could be the same for both high and low L2 groups. On the other hand, the interaction of Stimulus type × L2 proficiency for either gender or age solved the problem, which is a methodological suggestion for future studies of similar nature.

There remain some questions to be pursued in future studies. First, the present study chose Caucasian faces, but results might be different if the participants were presented with Oriental faces. According to "the other-race effect", people are easier to identify an individual face of one's own race than that of another race. One explanation would be *Perceptual Learning* defined by Gibson (1969) which involves "an increase in the ability to extract information from the environment, as a result of practice and experience with stimulation coming from it" (p. 3). According to this theory, people may be able to discriminate own-race faces more accurately due to their frequent use of some invariant cues of their own-race faces in daily life. However, cues used for own-race faces may not be appropriate when attempting to discriminate other-race faces, and thus performance would worsen when attempting to discriminate such unfamiliar stimuli. In the present study, Caucasian faces were unfamiliar to the Chinese participants, and thus "the other-race effect" might override the L2 effect. It would be better if there were studies comparing the participants' unconscious discrimination performance for both the Oriental and Caucasian faces. The second class of questions to be pursued is whether the minor difference in the perception of biological gender brings about any real difference in people's everyday life. Our intuition is "No", but we certainly need more empirical evidence, especially when it mixes up with the other-race effect.

5. Conclusion

The present study explores the effect of cross-linguistic difference in the pronoun encoding of gender information among Chinese-English bilinguals. Though there are a few studies related to the assumption that such a trivial cross-linguistic difference may affect language processing, no empirical study up to date has investigated the effect in a non-linguistic context. Hence, the present study investigated the effect on gender perception and the concept of gender was represented by male-female face pictures. The results showed that participants of higher L2 proficiency were not more sensitive to the male-female distinction than participants of lower L2 proficiency. Our study suggests that the difference may not influence gender perception in the nonlinguistic context, although it is not surprising considering the triviality of such cross-linguistic difference and the widespread gender binary opposition in daily life.

Our findings may contribute to a better understanding of the power of language that influence human perception. We also provide a methodological suggestion for future ERP linguistic relativity studies when distinctions are not perfectly comparable, such as the male/female and young/old distinction. Given "the other-race effect", further studies may compare the participants' unconscious discrimination performance for both the Oriental and Caucasian faces. Moreover, more studies are needed to explore whether the cross-linguistic difference makes any real difference in people's everyday life.

References

Antón-Méndez, I. (2011). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, *14*(3), 318–331. https://doi.org/10.1017/S1366728910000325

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *1*, 89. https://doi.org/10.1017/S1366728905002397

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, *37*(5), 393–412. https://doi.org/10.1080/02643294.2020.1769050

Athanasopoulos, P., Damjanovic, L., Krajciova, A., & Sasaki, M. (2010a). Representation of colour concepts in bilingual cognition: The case of Japanese blues. *Bilingualism: Language and Cognition*, *14*(01), 9–17. https://doi.org/10.1017/S1366728909990046

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J., & Thierry, G. (2010b). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition*, *116*(3), 437–443. https://doi.org/10.1016/j.cognition.2010.05.016

Bassetti, B. A. (2014). Is grammatical gender considered arbitrary or semantically motivated? Evidence from young adult monolinguals, second language learners, and early bilinguals. *British Journal of Psychology*, *105*(2), 273–294. https://doi.org/10.1111/bjop.12037

Boroditsky, L., Schmidt, L., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 61–79). Londyn: MIT Press

Chen, J. Y., & Su, J. J. (2011). Differential sensitivity to the gender of a person by English and Chinese speakers. *Journal of Psycholinguistic Research*, *40*(3), 195–203. https://doi.org/10.1007/s10936-010-9164-9

Cubelli, R., Paolieri, D., Lotto, L., & Job, R. (2011). The effect of grammatical gender on object categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 449–460. https://doi.org/10.1037/a0021965

Czigler, I., & Sulykos, I. (2010). Visual mismatch negativity to irrelevant changes is sensitive to task-relevant changes. *Neuropsychologia*, *48*(5), 1277–1282. https://doi.org/10.1016/j.neuropsychologia.2009.12.029

Dong, Y., & Jia, T. (2011). Composition of preverbal message: Gender errors of English pronouns by Chinese learners of English. *Modern Foreign Languages*, *34*(3), 279–286.

Dong, Y., Wen, Y., Zeng, X. et al. (2015). Exploring the Cause of English Pronoun Gender Errors by Chinese Learners of English: Evidence from the Self-paced Reading Paradigm. *Journal of Psycholinguistic Research*, *44*, 733–747. https://doi.org/10.1007/s10936-014-9314-6

Gabriel, U., & Mellenberger, F. (2004). Exchanging the generic masculine for gender-balanced forms—The impact of context valence. *Swiss Journal of Psychology*, *63*(4), 273. https://doi.org/10.1024/1421-0185.63.4.273

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Appleton-Century-Crofts.

Kecskes-Kovacs, K., Sulykos, I., & Czigler, I. (2013). Is it a face of a woman or a man? Visual mismatch negativity is sensitive to gender category. *Frontiers in Human Neuroscience*, *7*, 532. https://doi.org/10.3389/fnhum.2013.00532

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–210. https://doi.org/10.3758/BF03192770

Liang, L., Wen, Y., & Dong, Y. (2018). Gender constraint in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals. *Journal of Neurolinguistics*, *45*, 1–12. https://doi.org/10.1016/j.jneuroling.2017.08.001

Maciuszek, J., Polak, M., & Świątkowska, N. (2019). Grammatical gender influences semantic categorization and implicit cognition in Polish. *Frontiers in Psychology*, *10*, 2208. https://doi.org/10.3389/fpsyg.2019.02208

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Mo, L., Xu, G., Kay, P., & Tan, L. H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(34), 14026–14030. https://doi.org/10.1073/pnas.1111860108

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70–96). New York, NY: Cambridge University Press.

Stefanics, G., Csukly, G., Komlosi, S., Czobor, P., & Czigler, I. (2012). Processing of unattended facial emotions: a visual mismatch negativity study. *Neuroimage*, *59*(3), 3042–3049. https://doi.org/10.1016/j.neuroimage.2011.10.041

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(11), 4567–4570. https://doi.org/10.1073/pnas.0811155106

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104

Yu, Z., & Dong, Y. (2018). The Neurocognitive mechanism of gender agreement processing in L2 pronoun resolution by Chinese EFL learners. *Modern Foreign Languages*, *41*(4), 467–480.

Yu, Z., & Dong, Y. (2019). The neurocognitive mechanisms of semanticgender processing in L1 and L2 personal nouns by Chinese-English bilinguals. *Language, Cognition and Neuroscience*, *34*(7), 901–915. https://doi.org/10.1080/23273798.2019.1591472

Zarate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition*, *8*(2), 161. https://doi.org/10.1521/soco.1990.8.2.161

Zhao, L., & Bentin, S. (2008). Own- and other-race categorization of faces by race, gender, and age. *Psychonomic Bulletin & Review*, *15*(6), 1093−1099. https://doi.org/10.3758/PBR.15.6.1093

Notes

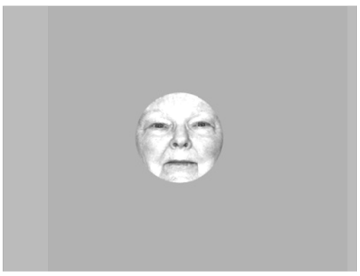
Note 1. As expected, similar statistical analyses for the 150–250 ms interval revealed no group differences in any ROI for either the age or gender category on the amplitudes of Deviant Gender-minus-Standard Gender difference waveforms.

Note 2. The following Munsell colours were used in the experiments of Thierry et al. (2009): dark blue: 5PB/value 4, light blue: 5PB/value 7, dark green: 5G/value 4, light green: 5G/value 7. The number of value demonstrates the amount of the luminance.

Appendix A

Facial stimuli in the experiment

The following are a few examples of the facial stimuli in the experiment. A dark grey cropping mask (size: 1024 × 768 pixels) was created and applied to each face so as to eliminate the physical individual differences concerning hair style and face shape.

	
Caucasian old female	Caucasian old male
	
Caucasian young female	Caucasian young male

Copyrights