Journal of Neurolinguistics 78 (2026) 101300

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Research paper

Neural correlates of novel word learning in an immersive virtual reality environment

Cong Liu [a,b], Jia Feng [a], John W. Schwieter [c,d,e], Jingyu Geng [a,*], Lu Jiao [a,b,**]

[a] Department of Psychology, School of Education Science, Qingdao University, Qingdao, China
[b] Brain, Cognition, and Language Learning Laboratory, Qingdao University, Qingdao, China
[c] Language Acquisition, Multilingualism, and Cognition Laboratory / Bilingualism Matters @ Wilfrid Laurier University, Waterloo, Canada
[d] Department of Linguistics and Languages, McMaster University, Hamilton, Canada
[e] School of Education, University College Dublin, Ireland

ARTICLE INFO

Keywords:
Immersive virtual reality
Novel word learning
ERP
Neural oscillations

ABSTRACT

By combining EEG and immersive virtual reality (iVR) technologies, the current study compared novel word learning in an iVR environment and through picture-word (PW) association. During three days of learning sessions, Chinese speakers learned two sets of German words, one set using iVR and the other set through PW association. A recognition task was administered to measure immediate (Day 4) and delayed learning performance (two weeks later). The results of the immediate post-test showed that compared to PW-learned words, there was better behavioral performance on iVR-learned words, along with increased N200 and decreased LPC amplitude. Time-frequency representation analyses further revealed reduced μ power for iVR-learned words relative to PW-learned words. However, the benefits of iVR-learned words did not emerge in the delayed post-test. These findings align with the social second language learning perspective and cognitive theory of multimedia learning, contributing to a deeper understanding of how multimodal learning environments influence word acquisition.

1. Introduction

Learning environments are important factors affecting foreign language vocabulary learning (Escudero et al., 2022). According to embodied language theory, language processes are shaped by the body's interactions with the external environment, grounded in bodily activity that occurs within real-world contexts and supported by sensorimotor experiences (Barsalou, 2008; Glenberg & Kaschak, 2002). Consequently, to facilitate word learning, an increasing number of studies have focused on the role of traditional multimodal environments in foreign language learning, such as video-based learning contexts (e.g., Montero Perez et al., 2018; Pi et al., 2021). However, limited attention has been paid to language learning in an immersive virtual reality (iVR) environment (e.g., Jiao et al., 2024; Legault et al., 2019). Therefore, it is still unclear how language learning proceeds in this innovative context and what neural mechanisms underlie the learning process. In the present study, electroencephalography (EEG) data were collected, and both event-related potential (ERP) and time-frequency representation (TFR) analyses were conducted to investigate how an iVR environment influences foreign language vocabulary learning.

* Corresponding author. Department of Psychology, School of Education Science, Qingdao University, Qingdao 266071, China.
** Corresponding author. Department of Psychology, School of Education Science, Qingdao University, Qingdao 266071, China.
E-mail addresses: 18253103655@163.com (J. Geng), jiaolu902@126.com (L. Jiao).

https://doi.org/10.1016/j.jneuroling.2025.101300
Received 2 July 2025; Received in revised form 9 November 2025; Accepted 25 November 2025
Available online 30 November 2025

1.1. Novel word learning in immersive virtual reality

With the rapid advancement of digital technologies, the use of iVR as an innovative approach of multimodal learning has received increasing attention in the field of foreign language acquisition (Makransky & Mayer, 2022). Li and Jeong (2020) proposed a new perspective on social L2 learning (SL2) within foreign language acquisition and identified iVR as one promising learning method (see also Li & Lan, 2021). SL2 emphasizes that interactions with the environment, objects, and other individuals in both real and virtual language contexts such as iVR provide rich linguistic and perceptual input, thereby facilitating an embodied L2 learning experience comparable to that of first language acquisition (Li & Jeong, 2020; Li & Lan, 2021). This viewpoint is closely related to the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2005), which posits that meaningful learning occurs when learners actively construct mental representations from both verbal and nonverbal information.

In line with SL2 and CTML, some studies have shown that learning novel words using iVR yields beneficial outcomes (Chen et al., 2019; Liu et al., 2024; Xie et al., 2019). For example, Legault et al. (2019) asked monolingual English speakers to learn novel Chinese words either through word-word environment or in an iVR environment. Immediate learning performance was measured by a four-alternative forced choice (4AFC) recognition task in which participants heard a newly-learned word and were asked to recognize the corresponding word in the word-word learning condition or the corresponding 3D picture in the iVR condition. The results revealed that participants performed better when responding to words learned in the iVR condition compared to those learned in the word-word condition, suggesting that iVR has a facilitative effect on novel word learning.

However, the limited capacity assumption of the CTML posits that an individual's information processing capacity is constrained at any given moment (Mayer, 2005). Consequently, the presence of redundant or irrelevant elements in a multimodal environment may impose excessive cognitive demands and hinder novel word learning. Lan et al. (2015) compared the effects of a desktop VR with a traditional learning environment on novel word learning. The procedure included seven learning sessions followed by immediate 4AFC tests, as well as a delayed 4AFC post-test administered later. Results showed that across the first three tests, words learned in the desktop VR yielded lower accuracy than in the traditional learning context, and no significant advantage of the desktop VR was observed in the subsequent immediate tests. These findings suggest that desktop virtual condition may increase learners' cognitive load (Sweller, 1988), potentially limiting the effectiveness for novel word learning.

Overall, whether iVR facilitates L2 word learning remains unclear, particularly in the 4AFC task which assesses lexical-phonological mapping. The present study aims to address this question.

1.2. Neural mechanisms of novel word learning

Very few studies have investigated the cognitive and neural mechanisms underpinning novel word learning in an iVR environment by utilizing neurocognitive techniques. For example, Jiao et al.'s (2024) EEG study compared the effects of picture-word (PW) and iVR environments on novel word learning. After three days of training, a lexical decision task was performed to verify whether target words had been learned or not. The results revealed that novel words learned via the iVR environment elicited more negative early components around the 200 ms time-window, a time frame associated with acoustic processing and early lexical access. The authors argued that the distinct ERP patterns observed across the two learning environments suggest that the immersive sensory experience provided by an iVR environment affects lexical form acquisition of novel words. Moreover, Jiao, Lin, et al. (2025) adopted a similar design using a semantic priming paradigm, and found that novel words learned under the iVR condition elicited stronger semantic effects on the N400 and LPC components compared to the PW condition. As the LPC component is typically associated with controlled, strategic processes underlying lexical retrieval (Bakker et al., 2015; Liu & van Hell, 2020), the present finding offers semantic-level evidence that immersive VR environments, through enriched perceptual input, promote learning and lexicalization of novel words.

However, few EEG studies have investigated the frequency domain of the brain's electrical activity in iVR L2 learning by utilizing a time-frequency representation (TFR) analysis. Different from ERP analyses, which extract time- and phase-locked information, one advantage of TFR analyses is that the power content of EEG signals is assigned to the two-dimensional time-frequency space (Liu et al., 2017). In other words, while ERP analyses yield a time-domain average that reflects static neural responses, TFR analyses allow for a finer-grained depiction of temporal dynamics by capturing oscillatory activity across multiple frequency bands measured by EEG (Keil

Table 1
Participants' language background (N = 29).

Measurements	M	SD	95 %CI
Age of L2 acquisition	9.52	2.59	[8.53, 10.50]
Self-rated proficiency in Chinese			
Listening	6.34	.61	[6.11, 6.58]
Speaking	6.17	.89	[5.83, 6.51]
Reading	6.34	.81	[6.04, 6.65]
Writing	6.21	.72	[5.95, 6.46]
Self-rated proficiency in English			
Listening	3.52	.87	[3.19, 3.85]
Speaking	3.26	.78	[3.12, 3.71]
Reading	4.17	1.04	[3.78, 4.57]
Writing	3.93	.92	[3.58, 4.28]

et al., 2022; Roach & Mathalon, 2008).

Recent research in foreign language learning has identified a link between μ power suppression and novel word learning through the use of time-frequency (TFR) analyses. μ suppression is commonly interpreted as an electrophysiological index of mirror-neuron activity in the sensorimotor cortex (Brunsdon et al., 2019; Pineda, 2005). For example, Ren and colleagues (2024) compared foreign word learning between a linguistic-symbol condition and a perceptual-symbol condition. In the learning session, Chinese characters (linguistic-symbol condition) or spatial cues (perceptual-symbol condition) were first presented as semantic primes to activate participants' linguistic or perceptual representations of the to-be-learned words. Participants then learned the novel words. The results showed that novel words engaging nonverbally perceptual representations elicited more pronounced μ suppression, suggesting that μ power oscillation is associated with perceptual simulation during novel word learning. For iVR-based foreign language learning, to our knowledge, only one study has employed TFR analyses to examine the effects of iVR on word learning (Zappa et al., 2024). In that study, native French speakers were assigned to either a specific action group or a pointing group within iVR environments. During a passive listening task, the TFR analysis of μ (8–13 Hz) power revealed no correlation between sensorimotor activation and learning performance.

Taken together, the neural mechanisms underlying novel word learning have yet to be fully clarified. While previous ERP studies have offered valuable but static insights into how iVR influences novel word learning, time-frequency (TFR) analyses provide a more dynamic perspective by capturing oscillatory brain activity that reflects cognitive operations across distinct frequencies, temporal scales, and cortical sites. Therefore, the present study employed both ERP and TFR approaches to achieve a more comprehensive understanding.

1.3. Present study

In this study, we combine EEG and VR technology to explore the neural correlates of novel word learning in an iVR environment. A group of native Chinese speakers were asked to learn German words, with half of words being learned in an iVR environment, and half via PW association. Following three days of learning phases, we asked participants to complete the 4AFC recognition task to measure novel word learning on Day 4 and two weeks later. In the immediate post-test on Day 4, behavioral and EEG data were recorded, and in the delayed post-test two weeks later, only behavioral data were recorded.

We combine ERP and TFR analyses to explore the dynamic processes involved in novel vocabulary learning across two different learning environments (i.e., iVR and PW association). Building on prior research and our study goals, we examined the mean amplitudes of the N200, P200, and LPC components (Jiao et al., 2024). The early P200 component, peaking around 200 ms after stimulus onset, is closely related to early phonological processing and orthographic extraction in lexical recognition (Barnea & Breznitz, 1998; Coch & Meade, 2016; Liu et al., 2011). The P200 has also been shown to reflect functions such as behavioral inhibition and attention allocation (Lijffijt et al., 2009). The N200 component, primarily observed at frontal-central electrodes, typically occurs 200–350 ms after stimulus onset and is thought to reflect processes related to word form recognition (Liu & Zhang, 2023). The LPC component, emerging between 300 and 600 ms after stimulus onset, has been associated with controlled, strategic processes involved in lexical retrieval (Bakker et al., 2015). Finally, regarding the TFR analyses, we focus on μ power, which is linked to lexical retrieval (van Elk et al., 2010), and whose suppression is identified as a reliable marker of mirror neuron system activation in the sensorimotor cortex (Brunsdon et al., 2019).

Based on the SL2 and CTML (Li & Jeong, 2020; Mayer, 2005), as well as prior findings on novel word learning in iVR (Jiao et al., 2024, 2025), we hypothesize that learning words in an iVR environment will result in more robust and integrated lexical representations compared to PW environment. Specifically, at the behavioral level, we expect that learners will perform better on iVR-learned words than on PW-learned words in both immediate and delayed post-tests. At the neurophysiological level, we predict that novel words learned via iVR will engage distinct neural mechanisms compared to those learned in the PW environment, as reflected in modulations of the P200, N200, and LPC components, as well as μ power activity. Finally, at the brain-behavior correlation level, we anticipate that immediate neural responses will be closely related to delayed behavioral performance for words learned in both learning environments.

2. Methods

2.1. Participants

Thirty-three native Chinese speakers were recruited to participate in the experiment. Four participants were excluded due to excessive EEG artifacts, leaving 29 participants included in the formal analyses (23 females; mean age = 20.03, range from 18 to 23). This sample exceeded the minimum sample size of 27 calculated by G*Power 3.1 (paired-samples *t*-test, $d = .50$, $\alpha = .05$, power = .80). All participants began learning English as a foreign language from an early age, as is typical in China (mean age of L2 acquisition = 9.5 years old). Self-reported language proficiency ratings (1 = very poor, 7 = excellent) showed that all participants were significantly more dominant in Chinese (Chinese proficiency = 6.27, English proficiency = 3.75; $t = 12.48$, $p < .01$). Moreover, all participants reported no prior knowledge of German, the language in which target words were learned in the present study. The local ethics committee approved the study, and all participants provided their written informed consent before participating in the study (see Table 1).

2.2. Materials

The materials consisted of 40 German words, their corresponding 2D line drawings and 3D objects, and a virtual scene. All 40 to-be-learned words were auditorily presented and represented common concepts typically found in a home setting (e.g., *Schüssel* 'bowl'), with each concept matched to the same real-world referent in both the iVR and PW conditions. The selection of to-be-learned German words considered word length (M = 7.03, SD = 2.73), syllable count (M = 2.03, SD = .66), and semantic category (e.g., Kitchenware, Clothing), as well as the suitability for presentation in the iVR scene. All auditory tokens were recorded by a highly proficient Chinese-German bilingual in a sound-attenuated booth. We recruited another group of Chinese-English bilinguals with similar language proficiency to rate whether each German word resembled any word they knew in their first language or L2 on a 5-point scale (1 = very dissimilar, 5 = very similar). Results showed that all selected German words were perceived dissimilar from Chinese and English.

The 2D picture stimuli consisted of 40 line-drawings which were selected from a standardized picture database (Snodgrass & Vanderwart, 1980; Zhang & Yang, 2003). Accordingly, items were rated 4.60 ± .34 in familiarity, 3.35 ± .58 on image-word agreement, and 2.28 ± .60 in visual complexity. For the iVR learning condition, an iVR setting simulated an apartment (living room, bedroom, kitchen), presented and edited in Unity software (https://unity.com). The 3D objects were selected from a standardized 3D object database (Peeters, 2018). We again recruited a separate group of Chinese-English bilinguals to assess the familiarity (4.49 ± .41), image-word agreement (3.51 ± .61), and visual complexity (2.45 ± .67). An HTC Vive headset provided immersion and allowed participants to move and operate a handheld controller. Using the controller, participants pointed a laser at 3D items to hear their corresponding German words. Before the learning sessions, participants practiced using the equipment to familiarize themselves with the virtual environment and controller operation.

2.3. Procedures and tasks

This study consisted of three learning sessions (one each on Days 1–3), an immediate test session (Day 4), and a delayed test session (two weeks later) (see Fig. 1). During the learning sessions, all participants learned the same 40 German words, half using iVR and half through PW association. Immediate and delayed post-tests were measured by a 4AFC recognition task separately. The immediate post-test collected both behavioral and EEG data, while the delayed post-test only collected behavioral data. Moreover, two participants failed to complete the delayed post-test session, leaving 27 participants included in the delayed post-test.

During the learning sessions, participants were given 15 min to learn 20 words through PW association and 15 min to learn the other half of German words in the iVR environment. To ensure matched exposure across conditions, each participant was required to spend the full 15 min in each learning environment. The learning order and items were counterbalanced across participants.

In the 4AFC recognition task during the immediate and delayed post-tests, participants heard one newly-learned word and were required to choose the correct picture out of four options presented to them on the computer screen. One picture was the correct match, while the other three were pictures of previously learned German words. Participants completed 10 practice trials to familiarize themselves with the procedure. Each novel word was presented twice in 4AFC task, resulting in 40 iVR and 40 PW trials randomized in order, which is within the range commonly adopted in ERP research (e.g., Jiao et al., 2024). For each trial, a fixation cross was presented at the center of the computer screen for 400 ms. After a 200 ms blank screen, a German word was played through the headphones and four pictures were presented on the screen. Participants were required to click on the correct picture corresponding to the German word they heard. Once they made a choice, a 1000 ms blank screen was presented before the next trial.

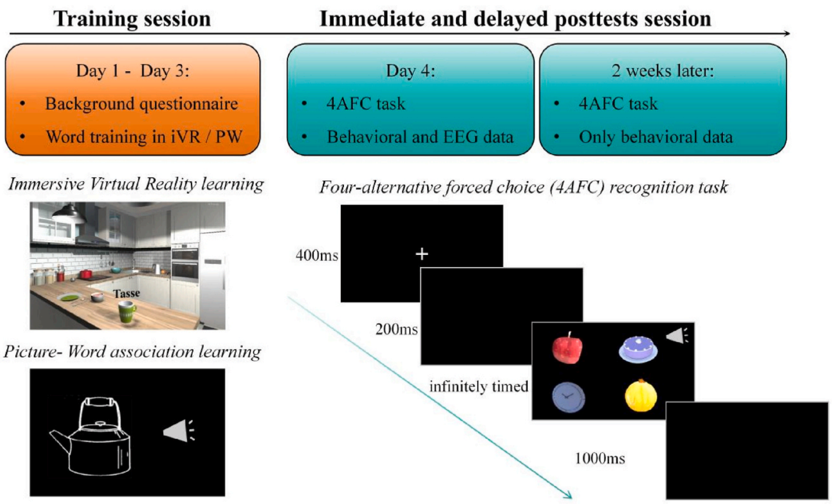

Fig. 1. Schematic overview of the tasks.

2.4. EEG data recording and preprocessing

EEG data were recorded during the immediate post-test stage (i.e., Day 4) using 64 Ag/AgCl electrodes (Brain Products, actiCAP system) placed according to the extended 10–20 positioning system. Electrode impedance was kept below 5 kΩ. EEG data were amplified with a band pass of .05–100 Hz and a sampling rate of 1000 Hz. EEG data were processed using MNE-Python (Gramfort et al., 2013) and EEGLAB (Delorme & Makeig, 2004). The signal was band-pass filtered at .5–40 Hz and re-referenced offline from a common reference to the left and right mastoids. The signals containing eye blinks and other artifacts were corrected for each subject by independent component analysis (ICA). Epochs of 200 ms before to 800 ms after stimuli onset were extracted. Baseline correction was performed in reference to the pre-stimulus activity (Liu et al., 2022). Additionally, EEG data exceeding ±100 μV were automatically removed.

2.5. Data analyses

The data analyses examined behavioral data, EEG data (ERP and TFR analyses), and correlations. We focused on response times (RTs) and accuracy by conducting mixed-effects model with subjects and items as crossed random effects in R (*lme 4* package, Baayen et al., 2008; Jiao et al., 2021). In the RTs analyses, error trials and trials with response times exceeding 3000 ms were first excluded. Subsequently, within each condition, the mean and standard deviation (SD) were computed, and trials with RTs beyond M ± 2 SD were removed. Because both immediate and delayed post-tests collected behavioral data, the best-fitting model included test session, environment, and their interaction as fixed effects, as well as the by-subject and by-item random effects. Both variables were centered (session: immediate = −.5, delayed = .5; environment: PW = −.5, iVR = .5). We started with a full model including all fixed effects, random intercepts for participants and items, and random slopes for all predictors (Barr et al., 2013), and when models did not converge, we followed a backward-fitting procedure to identify a model that would converge.

In accordance with previous work and the theoretical interests of the present study, the ERP analyses focused on the P200 (150–250 ms) and N200 (250–350 ms) components, computed as the mean amplitude averaged across frontal-central electrodes (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2), and on the LPC (350–500 ms), computed as the mean amplitude averaged across frontal-parietal electrodes (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2). Because EEG data was only collected during the immediate post-test session, we calculated the mean amplitudes across the selected time-windows and conducted paired-samples *t*-tests between the PW and iVR learning environments. For the TFR analyses, oscillatory power for μ band (8–12 Hz) was obtained from single trial EEG epochs using morlet transform (Jiao, Schwieter, & Liu, 2025; Mellem et al., 2012; Piai & Zheng, 2019). TFR analyses examined the range of 2–28 Hz in steps of 1 Hz. The oscillatory power value was calculated for each epoch and then averaged across epochs for the PW and iVR learning environments. In the TFR analyses, the time windows of interest corresponded to those used in the ERP analyses (i.e., 150–250, 250–350, and 350–500 ms). Paired-samples *t*-tests on mean power were conducted for each time window. Finally, in order to explore the predictive effect of learning environment on the delayed performance, we calculated the Pearson correlation between the neural responses of the immediate post-test and the behavioral performance on the delayed post-test for both PW and iVR learning environments (Jiao et al., 2022).

3. Results

3.1. Behavioral performance for PW- and iVR-learned words

Fig. 2A presents the estimated marginal mean (EMM) values for accuracy from the 4AFC task for PW- and iVR-learned words. The mixed-effects model for accuracy only showed a significant main effect of test session (Estimate = −.68, SE = .25, $z = -2.71$, $p = .007$), with higher accuracy on the immediate test (M = .99, SD = .01) than the delayed test (M = .98, SD = .02). Neither the main effect of learning environment (Estimate = −.12, SE = .25, $z = -.48$, $p = .63$), nor the interaction was significant (Estimate = .34, SE = .50, $z = .70$, $p = .49$).

Fig. 2B displays the EMMs values of RTs from the 4AFC task for PW- and iVR-learned words. The mixed-effects model for RTs

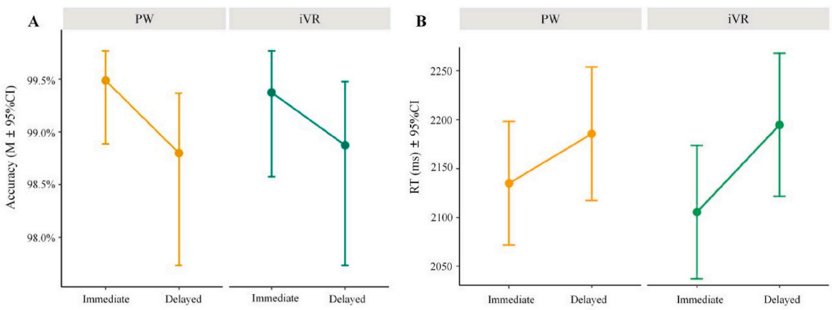

Fig. 2. Estimated marginal means (EMMs) with 95 % confidence intervals for accuracy (A) and response times (B) in the 4AFC task, shown separately for PW and iVR across the immediate and delayed post-tests. Colored dots denote EMMs; error bars indicate 95 % CIs.

showed that the main effect of test session was significant (Estimate = 63.72, SE = 18.37, $t = 3.47$, $p = .002$), with faster responses on the immediate post-test (M = 2097 ms, SD = 111 ms) than the delayed post-test (M = 2160 ms, SD = 132 ms). Despite the main effect of learning environment not reaching significance (Estimate = −5.78, SE = 21.86, $t = -.26$, $p = .79$), the interaction between learning environment and test session was significant (Estimate = 47.45, SE = 22.91, $t = 2.07$, $p = .04$). Further analyses revealed that participants responded faster to iVR-learned words (M = 2080 ms, SD = 128 ms) than PW-learned words (M = 2114 ms, SD = 114 ms) on the immediate post-test (Estimate = −31.72, SE = 15.12, $t = -2.1$, $p = .03$), but not on the delayed post-test (PW = 2152 ms, iVR = 2167 ms; Estimate = 16.12, SE = 17.83, $t = .90$, $p = .36$).

3.2. Neural responses to PW- and iVR-learned words

Fig. 3 displays the grand average ERP waveforms elicited by the 4AFC task during the immediate post-test session. Paired-samples *t*-tests were conducted for EEG data because this data were collected only during the immediate post-test. There were three time-windows selected for the ERP analyses: P200 (150–250 ms), N200 (250–350 ms), and LPC (350–500 ms). Results showed that for the P200 component, the waveform elicited by iVR-learned words was less positive than PW-learned words, $t = 3.78$, $p = .001$. Similarly, the comparison of the LPC time-window was also significant, with less positive waveforms for iVR-learned words than for PW-learned words, $t = 3.95$, $p < .001$. Finally, during the N200 time-window, the two conditions were significantly different, with more negative waveforms for iVR-learned words than for PW-learned words, $t = 4.01$, $p < .001$.

Fig. 4 illustrates the oscillatory activity elicited by the 4AFC task during the immediate post-test session. In accordance with our study's goal, the TFR analyses focused on μ power (8–12 Hz) over the midline electrodes (Fz, FCz, Cz, CPz) (Zhao et al., 2018) across three different time-windows (150–250, 250–350, and 350–500 ms). Paired-samples *t*-tests showed that the μ power of iVR-learned words was significantly smaller than PW-learned words during the 350–500 ms time-window, $t = 2.19$, $p = .03$. But these differences were not significant during the other two time windows (*ps* > .05).

3.3. Predictive effect of neural responses to delayed learning performance

When considering the ERP and TFR results, there were clear cognitive processing differences between PW- and iVR-learned words in both mean amplitude and oscillatory activity during the 350–500 ms time-window. All *p*-values reported below were FDR-corrected unless otherwise stated. Thus, we concentrated on this particular time-window and conducted correlation analyses between neural responses (i.e., LPC amplitude and μ power value) and RTs during the delayed post-test session to examine the relationship between neural responses during the immediate post-test and behavioral performance two weeks later in the delayed post-test. For PW-learned words, we did not observe significant correlations between the neural responses (LPC amplitude vs. RTs of delayed session: $r = -.25$, $p = .44$; μ power vs. RT of delayed session: $r = -.03$, $p = .88$) and the RTs in the delayed session (see Fig. 5A and 5B). However, for iVR-learned words, despite there being no significant correlation between μ power ($r = .16$, $p = .43$) and the RTs in the delayed session (see Fig. 5D), the correlation with LPC amplitude was marginally significant ($r = .42$, $p = .06$) (see Fig. 5C). Notably, the uncorrected *p*-value for this LPC-RT correlation was .03, indicating statistical significance before correction. This positive correlation indicates that the larger LPC amplitude elicited by iVR-learned words during the immediate session was associated with slower response speeds during the delayed session.

4. Discussion

This study combined EEG and iVR technology to examine the effects of learning environments on novel word acquisition among Chinese speakers who learned German words. After three days of learning words, some in an iVR environment and some through PW association, an immediate post-test on Day 4 showed that participants responded faster to the iVR-learned words than to PW-learned

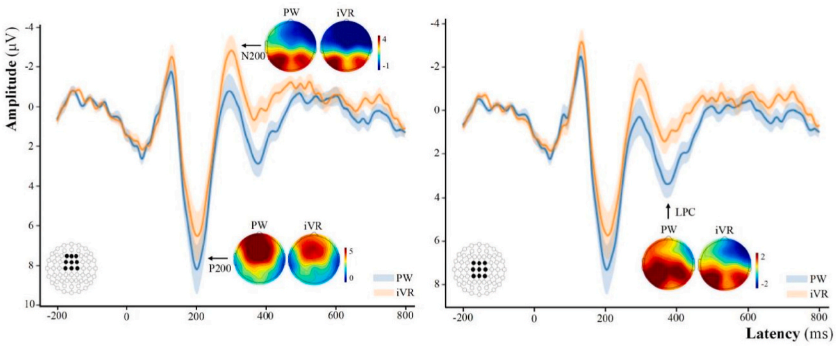

Fig. 3. Grand averaged ERPs (P200 and N200) at the frontal-central cluster (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) across PW and iVR environments (left), and Grand averaged ERP (LPC) at the frontal-parietal cluster (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2) (right) from the immediate post-test. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

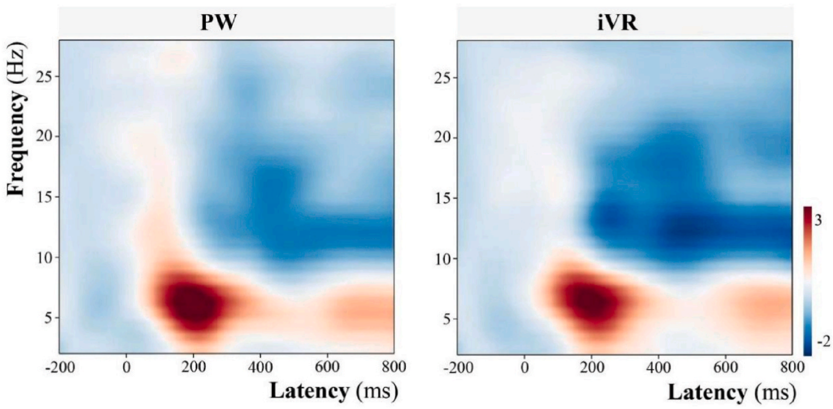

Fig. 4. Time-frequency distribution plots for oscillatory activity for PW- (left) and iVR- (right) learned words on the immediate post-test.

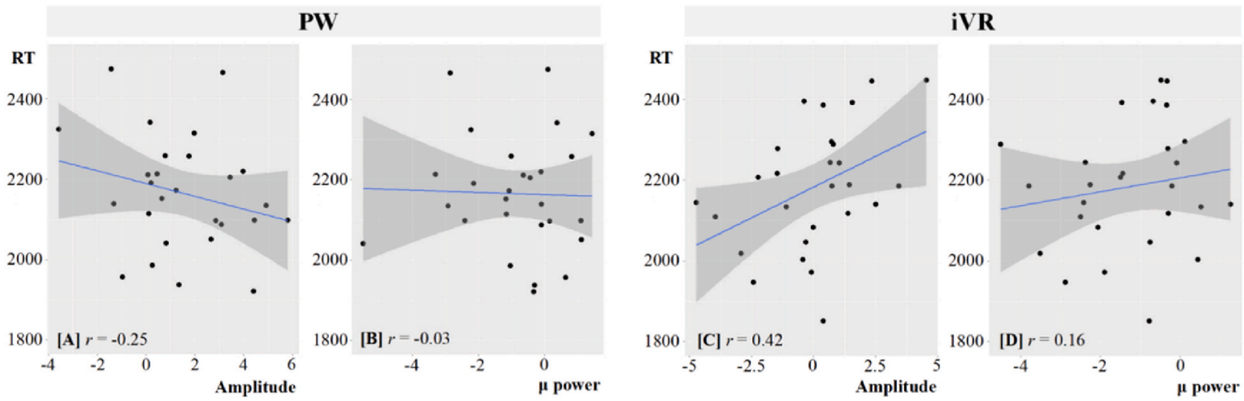

Fig. 5. Correlations between neural responses (i.e., mean amplitude and μ power of the 350–500 time-window) during the immediate post-test and RTs during the delayed post-test.

words. This behavioral evidence was further supported by observed changes in amplitude on three key ERP components (P200, N200, and LPC) and in μ oscillations. Specifically, the iVR-learned words elicited more negative-going N200 amplitude, whereas PW-learned words elicited more positive P200 and LPC effects. iVR-learned words also triggered greater μ-band suppression in the later time window. Moreover, correlation analyses revealed that for iVR-learned words, but not for PW-learned words, LPC amplitude during the 350–500 time-window in the immediate post-test was marginally associated with behavioral performance in the delayed post-test. These findings highlight the positive effects that a multisensory learning environment such as iVR has on foreign language learning.

4.1. An iVR environment positively impacts novel word learning

The behavioral results of the immediate post-test indicated that responses to iVR-learned words were faster than to PW-learned words, a finding which is consistent with our hypothesis. According to the SL2 and CTML (Li & Jeong, 2020; Mayer, 2005), an iVR-based learning environment can provide rich nonverbal information (e.g., spatial layout, surrounding environment) that facilitates novel word learning in a foreign language. The facilitative effect observed in the present study aligns with prior work (Legault et al., 2019; Liu et al., 2024) suggesting that an immersive environment is an effective tool in improving foreign language learning, at least with respect to novel word learning. Moreover, the enhanced N200 negativity and greater μ-band suppression observed for iVR-learned words suggests that iVR learning engages stronger perceptual, semantic, and multimodal processing, which in turn, leads to faster responses to iVR-learned words than to PW-learned words.

However, our findings are inconsistent with some of the results reported by Lan et al. (2015). In their study, which compared desktop VR with PW context, the results showed that words learned in the desktop VR yielded lower accuracy than in the PW context. The authors argued that the desktop VR context imposes greater cognitive load on the learners than the PW context, particularly during the early stages of learning. These distinct findings, therefore, may indicate that the limited capacity of working memory may be one critical factor in iVR learning research (Mayer, 2009). This said, in the study by Lan et al. (2015), learners only had limited exposure to the desktop VR context during early stages of word learning, which may not have been sufficient to integrate nonverbal information into to-be-learned knowledge.

Interestingly, the initial stronger learning outcomes from the iVR context in the immediate post-test did not emerge in the delayed post-test, i.e., there were similar RTs for both iVR- and PW-learned words. To some extent, this finding was unexpected. One possible explanation is that the rich nonverbal information and activated motivation might only act on memory encoding in the short term, and thus may not assist memory retention. Given that the present study failed to consolidate newly-learned words before the delayed post-test, future research should examine the delayed effects of iVR environment on novel word learning.

4.2. Neural patterns of novel word learning in an iVR environment

In addition to the behavioral analyses, we examined how an iVR environment impacts novel word acquisition through ERP and TFR analyses. With respect to ERP patterns, a more negative N200 was elicited by iVR-learned words compared to PW-learned words, suggesting that novel words learned via iVR required more cognitive resources for lexical recognition and phonological processing (Liu & Zhang, 2023). This N200 effect is in line with a lexical recognition study by Jiao et al. (2024). We also observed decreased LPC amplitude for iVR-learned words compared to PW-learned words, indicating that iVR-learned words may impose less demands on attentional resources and cognitive control during a later stage. Previous research on children with attention deficit hyperactivity disorder has demonstrated that VR can effectively enhance attention and cognitive control (Bashiri et al., 2017; Parsons et al., 2007). Thus, one possible explanation is that iVR may sharpen attention to target information and strengthen control over irrelevant input during recognition (indexed by N200), thereby reducing later-stage processing demands (indexed by LPC). Otherwise, increased LPC amplitude for PW-learned words may imply greater cognitive demands on their processing.

Given SL2's claim that iVR affords embodied neural representations (Li & Jeong, 2020) and that μ suppression serves as an index of sensorimotor engagement (Ren et al., 2024), our TFR analyses focused on μ power (8–12 Hz). The findings revealed a significant effect of iVR on μ power in the 350–500 ms time window (i.e., corresponding to the LPC component). On the one hand, based on previous studies, increased μ power is associated with slower RTs in L2 word recognition and greater difficulty in lexical retrieval (van Elk et al., 2010; Vukovic & Shtyrov, 2014). Therefore, in present study, the increased μ power of PW-learned words may imply that lexical retrieval is more challenging when learning is through PW association than via iVR. The lexical representations of iVR-learned words have integrated multiple verbal and nonverbal experiences, which facilitates easier retrieval during a lexical recognition task. On the other hand, given that μ power is linked to the activity of mirror neurons in the sensorimotor cortex (Brunsdon et al., 2019), the decreased μ power of iVR-learned words suggests that the perceptual information provided by the iVR environment facilitates word learning and processing. However, inconsistent with our findings, Zappa et al. (2024) examined μ and β power activity and did not observe a VR effect. This null finding may be attributed to insufficient embodiment in the VR environment or limited statistical power in detecting differences in μ and β power.

In addition, a series of correlational analyses were conducted to explore the relationship between neural patterns (ERPs and TFR) during the immediate post-test and behavioral performance during the delayed post-test. We found that, for iVR-learned words, LPC amplitude in the immediate post-test showed a negative-trending association with RTs in the delayed post-test, suggesting that smaller LPC amplitude predicated faster RTs to iVR-learned words two weeks later. However, no such relationship was found for PW-learned words. Although the behavioral results indicated no advantage of the iVR learning environment in the delayed post-test, this significant correlation may suggest that, to some extent, the iVR environment facilitates vocabulary learning. Overall, our findings provide support at both the behavioral and neural levels for the positive role that iVR and its multisensory richness and interactivity play in word learning.

Finally, despite offering novel insight into the field of iVR and novel word learning, the present study has some limitations. First, although we ensured matched learning duration across conditions, we did not record the number of clicks each participant made for individual words in the iVR and PW environments, which may have introduced variability in their exposure to words. Second, EEG data were only collected during the immediate post-test, and we did not track neural activity during the delayed post-test. Future research should address this limitation by incorporating neurophysiological measures at multiple time points.

5. Conclusion

In the present study, we examined whether and how iVR impacts novel word learning, providing new insight into foreign language learning by using a combination of EEG and iVR technology. The results demonstrated that behavioral responses to iVR-learned words were faster than PW-learned words in an immediate post-test, but such advantage was not observed two weeks later. Moreover, in the immediate post-test, iVR-learned words triggered different ERPs patterns as evidenced by the P200, N200, and LPC components, and μ power. Furthermore, for iVR-learned words, LPC amplitude in the immediate post-test showed a trend-level correlation with behavioral performance in the delayed post-test. Overall, our findings offer theoretical contributions that support the SL2 and CTML. They also provide implications for practitioners who may wish to consider the adoption of immersive technologies such as iVR in their instructional approaches.

CRediT authorship contribution statement

Cong Liu: Writing – review & editing, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jia Feng:** Writing – original draft, Visualization, Formal analysis. **John W. Schwieter:** Writing – review & editing. **Jingyu Geng:** Writing – review & editing, Supervision. **Lu Jiao:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgements

This work was supported by the Social Science Planning Project of Shandong Province (grant number: 22DJYJ06).

Data availability

Data will be made available on request.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia, 79*(Pt A), 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020

Barnea, A., & Breznitz, Z. (1998). Phonological and orthographic processing of Hebrew words: Electrophysiological aspects. *The Journal of Genetic Psychology, 159*(4), 492–504.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bashiri, A., Ghazisaeedi, M., & Shahmoradi, L. (2017). The opportunities of virtual reality in the rehabilitation of children with attention deficit hyperactivity disorder: A literature review. *Korean Journal of Pediatrics, 60*(11), 337–343. https://doi.org/10.3345/kjp.2017.60.11.337

Brunsdon, V. E. A., Bradford, E. E. F., & Ferguson, H. J. (2019). Sensorimotor mu rhythm during action observation changes across the lifespKor. J. Pediatricsan independently from social cognitive processes. *Developmental Cognitive Neuroscience, 38*, Article 100659. https://doi.org/10.1016/j.dcn.2019.100659

Chen, Y., Smith, T. J., York, C. S., & Mayall, H. J. (2019). Google earth virtual reality and expository writing for young English learners from a funds of knowledge perspective. *Computer Assisted Language Learning, 33*(1–2), 1–25. https://doi.org/10.1080/09588221.2018.1544151

Coch, D., & Meade, G. (2016). N1 and P2 to words and wordlike stimuli in late elementary school children and adults. *Psychophysiology, 53*(2), 115–128. https://doi.org/10.1111/psyp.12567

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Escudero, P., Smit, E. A., & Angwin, A. J. (2022). Investigating orthographic versus auditory cross-situational word learning with online and laboratory-based testing. *Language Learning, 73*(2), 543–577. https://doi.org/10.1111/lang.12550

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558–565. https://doi.org/10.3758/BF03196313

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hamalainen, M. (2013). MEG and EEG data analysis with MNE-python. *Frontiers in Neuroscience, 7*, 267. https://doi.org/10.3389/fnins.2013.00267

Jiao, L., Duan, X., Liu, C., & Chen, B. (2022). Comprehension-based language switching between newly learned languages: The role of individual differences. *Journal of Neurolinguistics, 61*, Article 101036. https://doi.org/10.1016/j.jneuroling.2021.101036

Jiao, L., Lin, Y., Schwieter, J. W., & Liu, C. (2025). Learning novel words in an immersive virtual reality context: Tracking lexicalization through behavioral and event-related potential measures. *Language learning*. Advance online publication. https://doi.org/10.1111/lang.12707

Jiao, L., Liu, C., Schwieter, J. W., & Chen, B. (2021). Switching between newly learned languages impacts executive control. *Psychophysiology, 58*(10), Article e13888. https://doi.org/10.1111/psyp.13888

Jiao, L., Schwieter, J. W., & Liu, C. (2025). Comprehension-based language switching in experienced and newly learned languages: Evidence from induced brain oscillations. *Journal of Neurolinguistics, 73*, Article 101232. https://doi.org/10.1016/j.jneuroling.2024.101232

Jiao, L., Zhu, M., Xu, Z., Zhou, G., Schwieter, J. W., & Liu, C. (2024). An ERP study on novel word learning in an immersive virtual reality context. *Bilingualism: Language and Cognition, 27*(1), 128–136. https://doi.org/10.1017/s136672892300038x

Keil, A., Bernat, E. M., Cohen, M. X., Ding, M., Fabiani, M., Gratton, G., Kappenman, E. S., Maris, E., Mathewson, K. E., Ward, R. T., & Weisz, N. (2022). Recommendations and publication guidelines for studies using frequency domain and time-frequency domain analyses of neural time series. *Psychophysiology, 59*(5), Article e14052. https://doi.org/10.1111/psyp.14052

Lan, Y.-J., Fang, S.-Y., Legault, J., & Li, P. (2015). Second language acquisition of Mandarin Chinese vocabulary: Context of learning effects. *Educational Technology Research & Development, 63*(5), 671–690. https://doi.org/10.1007/s11423-015-9380-y

Legault, J., Fang, S.-Y., Lan, Y.-J., & Li, P. (2019). Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain and Cognition, 134*, 90–102. https://doi.org/10.1016/j.bandc.2018.09.004

Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., & Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages, 4*(1), 13. https://doi.org/10.3390/languages4010013

Li, P., & Jeong, H. (2020). The social brain of language: Grounding second language learning in social interaction. *Npj Science of Learning, 5*, 8. https://doi.org/10.1038/s41539-020-0068-7

Li, P., & Lan, Y.-J. (2021). Digital language learning (DLL): Insights from behavior, cognition, and the brain. *Bilingualism: Language and Cognition, 25*(3), 361–378. https://doi.org/10.1017/s1366728921000353

Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., Gerard Moeller, F., & Swann, A. C. (2009). P50, N100, and P200 sensory gating: Relationships with behavioral inhibition, attention, and working memory. *Psychophysiology, 46*(5), 1059–1068. https://doi.org/10.1111/j.1469-8986.2009.00845.x

Liu, B., Jin, Z., Qing, Z., & Wang, Z. (2011). The processing of phonological, orthographical, and lexical information of Chinese characters in sentence contexts: An ERP study. *Brain Research, 1372*, 81–91. https://doi.org/10.1016/j.brainres.2010.11.068

Liu, H., Liang, L., Zhang, L., Lu, Y., & Chen, B. (2017). Modulatory role of inhibition during language switching: Evidence from evoked and induced oscillatory activity. *International Journal of Bilingualism, 21*(1), 57–80. https://doi.org/10.1177/1367006915600800

Liu, C., Mao, Y., Wang, X., Schwieter, J. W., & Jiao, L. (2024). Sleep-dependent consolidation effects on foreign language word acquisition in a virtual reality environment. *Memory & Cognition, 52*(2), 302–311. https://doi.org/10.3758/s13421-023-01461-z

Liu, Y., & van Hell, J. G. (2020). Learning novel word meanings: An ERP study on lexical consolidation in monolingual, inexperienced foreign language learners. *Language Learning, 70*(S2), 45–74. https://doi.org/10.1111/lang.12403

Liu, C., Wang, H., Timmer, K., & Jiao, L. (2022). The foreign language effect on altruistic decision making: Insights from the framing effect. *Bilingualism: Language and Cognition, 25*(5), 890–898. https://doi.org/10.1017/s1366728922000128

Liu, J., & Zhang, Y. (2023). Language experience modulates the visual N200 response for disyllabic Chinese words: An event-related potential study. *Brain Sciences, 13* (9). https://doi.org/10.3390/brainsci13091321

Makransky, G., & Mayer, R. E. (2022). Benefits of taking a virtual field trip in immersive virtual reality: Evidence for the immersion principle in multimedia learning. *Educational Psychology Review, 34*(3), 1771–1798. https://doi.org/10.1007/s10648-022-09675-4

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning, 41*(1), 31–48.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mellem, M. S., Bastiaansen, M. C., Pilgrim, L. K., Medvedev, A. V., & Friedman, R. B. (2012). Word class and context affect alpha-band oscillatory dynamics in an older population. *Frontiers in Psychology, 3*, 97. https://doi.org/10.3389/fpsyg.2012.00097

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1–2), 1–26. https://doi.org/10.1080/09588221.2017.1375960

Parsons, T. D., Bowerly, T., Buckwalter, J. G., & Rizzo, A. A. (2007). A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology, 13*(4), 363–381. https://doi.org/10.1080/13825580600943473

Peeters, D. (2018). A standardized set of 3-D objects for virtual reality research and applications. *Behavior Research Methods, 50*(3), 1047–1054. https://doi.org/10.3758/s13428-017-0925-3

Pi, Z., Zhu, F., Zhang, Y., & Yang, J. (2021). An instructor's beat gestures facilitate second language vocabulary learning from instructional videos: Behavioral and neural evidence. *Language Teaching Research, 28*(5), 1997–2025. https://doi.org/10.1177/13621688211039023

Piai, V., & Zheng, X. (2019). Speaking waves: Neuronal oscillations in language production. *Psychology of Learning and Motivation, 71*, 265–302. https://doi.org/10.1016/bs.plm.2019.07.002

Pineda, J. A. (2005). The functional significance of mu rhythms: Translating "seeing" and "hearing" into "doing.". *Brain Research Reviews, 50*(1), 57–68. https://doi.org/10.1016/j.brainresrev.2005.04.005

Ren, W., Yang, T., & Wang, H. (2024). Different effects of linguistic and perceptual symbolic representations on foreign language vocabulary learning: Evidence from behavioral and EEG data. *Acta Psychology Sinica, 56*(5), 542–554. https://doi.org/10.3724/SP.J.1041.2024.00542

Roach, B. J., & Mathalon, D. H. (2008). Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia Bulletin, 34*(5), 907–926. https://doi.org/10.1093/schbul/sbn093

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285. https://doi.org/10.1207/s15516709cog1202_4

van Elk, M., van Schie, H. T., Zwaan, R. A., & Bekkering, H. (2010). The functional role of motor activation in language processing: Motor cortical oscillations support lexical-semantic retrieval. *NeuroImage, 50*(2), 665–677. https://doi.org/10.1016/j.neuroimage.2009.12.123

Vukovic, N., & Shtyrov, Y. (2014). Cortical motor systems are involved in second-language comprehension: Evidence from rapid mu-rhythm desynchronisation. *NeuroImage, 102*(Pt 2), 695–703. https://doi.org/10.1016/j.neuroimage.2014.08.039

Xie, Y., Chen, Y., & Ryder, L. H. (2019). Effects of using mobile-based virtual reality on Chinese L2 students' oral proficiency. *Computer Assisted Language Learning, 34* (3), 225–245. https://doi.org/10.1080/09588221.2019.1604551

Zappa, A., Bolger, D., Pergandi, J. M., Fargier, R., Mestre, D., & Frenck-Mestre, C. (2024). The neural correlates of embodied L2 learning: Does embodied L2 verb learning affect representation and retention? *Neurobiology of Language, 5*(2), 360–384. https://doi.org/10.1162/nol_a_00132

Zhang, Q., & Yang, Y. (2003). The determiners of picture-naming latency. *Acta Psychology Sinica, 35*(4), 447–454.

Zhao, G., Zhang, Y., & Ge, Y. (2018). Frontal EEG asymmetry and middle line power difference in discrete emotions. *Frontiers in Behavioral Neuroscience, 12*. https://doi.org/10.3389/fnbeh.2018.00225