ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202601.00137

The Impact of Virtual Reality Learning Environments on Foreign Language Vocabulary Production: Evidence from Behavioral and EEG Data

Authors: Liu Cong, Liu Qiuxia, Zhu Mengrui, Jiao Lu, Wang Ruiming, Ruiming Wang

Date: 2026-01-17T17:23:49+00:00

Abstract

The rapid development of artificial intelligence technologies such as virtual reality (VR) has offered new approaches to facilitating foreign language vocabulary acquisition, yet no studies to date have examined how VR learning environments influence foreign language vocabulary production. Drawing on VR technology and electroencephalography (EEG) experiments, and using a traditional picture-word environment as a reference, the present study systematically investigates the effects of a VR environment on foreign language vocabulary production and its underlying neural mechanisms. Behavioral results showed that, in the immediate test phase, vocabulary items learned in the VR environment were produced with higher accuracy than those learned in the picture-word environment, whereas no difference was observed between the two environments in the delayed phase. EEG results in the immediate test phase revealed that, compared with the picture-word environment, the VR environment elicited smaller P200 and LPC components; moreover, time-frequency analysis showed more pronounced m-wave suppression in the VR condition, accompanied by reduced q-wave power. Taken together, the findings indicate that, relative to a picture-word environment, a VR environment leverages perceptual–motor simulation to enhance the concreteness of vocabulary representations, thereby facilitating lexical retrieval and production, and providing empirical support for embodied cognition theory.

ChinaRxiv 中X

Full Text

Effects of Virtual Reality Learning Contexts on Foreign Language Word Production: Behavioral and EEG Evidence

LIU Cong[1], LIU Qiuxia[1], ZHU Mengrui[2], JIAO Lu[1], WANG Ruiming[3]

[1]School of Education Science / Brain, Cognition, and Language Learning Laboratory, Qingdao University, Qingdao 266071, China
[2]School of Foreign Languages, Xi' an Jiaotong University, Xi' an 710049, China
[3]School of Psychology / Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631, China

Abstract

The rapid development of artificial intelligence technologies such as virtual reality (VR) has provided new methods for promoting foreign language vocabulary acquisition. However, no studies have yet examined how VR learning contexts affect foreign language word production. This study combined VR technology with EEG experiments, using a traditional picture-word environment as a reference, to systematically investigate the effects of VR contexts on foreign language word production and the underlying neural mechanisms. Behavioral results showed that during the immediate test phase, the production accuracy of words learned in the VR environment was higher than that in the picture-word environment, but no differences emerged in the delayed phase. EEG results from the immediate test phase revealed that, compared to the picture-word environment, the VR environment elicited smaller P200 and LPC components. Time-frequency analysis further showed more pronounced μ-wave suppression in the VR environment, accompanied by weakened θ-wave power. These findings collectively demonstrate that, compared to picture-word environments, VR environments enhance the concreteness of words through perceptual-motor simulation, thereby facilitating lexical retrieval and production, providing empirical support for embodied cognition theory.

Keywords: virtual reality, foreign language vocabulary acquisition, word production, EEG, time-frequency analysis

Vocabulary acquisition forms the foundation of mastering a foreign language, and learning environments significantly affect foreign language vocabulary acquisition outcomes (Linck et al., 2009). In traditional learning environments, learners typically engage in picture-word or word-word paired-associate learning mediated by their native language, yet the learning effects often fail to reach native-language levels. With the rapid advancement of Artificial Intelligence (AI), digital education reform and innovation in learning models have become inevitable trends. Virtual Reality (VR) technology, as an emerging educational tool, provides learners with immersive experiences difficult to obtain in traditional learning environments by simulating real-life contexts, offering new methods for foreign language vocabulary acquisition.

VR technology has garnered widespread attention in foreign language learning and has demonstrated positive effects in orthographic processing, lexical recognition, and semantic integration (Fuhrman et al., 2021; Tai et al., 2022). In VR learning environments, abstract lexical symbols are concretized as 3D objects, allowing learners to grasp and manipulate these objects through virtual equipment, accumulating perceptual-motor experiences (Li & Jeong, 2020). Several studies have found that words learned in VR environments show better recognition effects than those learned in traditional environments. For example, Legault and Zhao et al. (2019) required English native speakers to learn Chinese vocabulary in an immersive VR environment and a word-word paired environment. A four-alternative forced-choice recognition task revealed that VR environments significantly facilitated foreign language vocabulary learning among participants with poorer learning performance. Using a semantic priming paradigm, researchers found, through both behavioral and EEG measures, that Chinese native speakers showed larger semantic priming effects for words learned in VR environments compared to traditional picture-word paired learning environments, indicating that VR environments promote semantic integration of foreign language vocabulary and rely more on automatic processing mechanisms (Jiao, Lin et al., 2025). Additionally, using a vocabulary recall task, an fMRI study demonstrated that the degree of association between brain activation patterns during vocabulary recall and those during the encoding phase affected recall accuracy, and that contextual information provided by VR learning environments facilitated foreign language vocabulary learning and retention (Essoe et al., 2022).

Why does VR facilitate foreign language vocabulary acquisition? Embodied Cognition Theory posits that cognitive activities are holistic, situated activities formed through individuals' interactive experiences between their bodies and external environments (Barsalou, 2008). Focusing on foreign language learning, the Social L2 Learning Theory also emphasizes that interaction with environments, objects, and others in real or virtual simulated language contexts can provide rich linguistic and perceptual information, thereby achieving embodied effects similar to native language learning (Li & Jeong, 2020). On the other hand, the Cognitive Affective Model of Immersive Learning suggests that learning effects in VR environments are related not only to factors such as interest, motivation, and cognitive load during the learning process, but also to characteristics of learning content and testing tasks (Makransky & Petersen, 2021).

From the perspective of vocabulary acquisition outcomes, successful acquisition of a foreign language word means that learners can master its form, meaning, and pronunciation—recognizing its orthographic form, understanding its semantics, and articulating it clearly (Wallace, 1982). It should be noted that current experimental evidence on VR and foreign language vocabulary acquisition has only focused on orthographic recognition (Jiao et al., 2024) and semantic comprehension (Legault, Fang et al., 2019; Liu et al., 2024), without addressing how VR learning environments affect the production process of foreign language vocabulary. Vocabulary comprehension differs from vocabulary production: the former emphasizes the process from lexical input to conceptual processing, whereas the latter emphasizes the cognitive process from conceptual processing to lexical output, manifested as learners expressing their intended ideas through speech organs (Peng, 2004). Due to these differences, we cannot simply infer the production process of foreign language vocabulary learned in VR environments based on evidence from the comprehension level.

Based on the above research status, the questions addressed in this study are: Do differences exist in the production effects of foreign language vocabulary learned in VR environments versus picture-word environments? What are the neural mechanisms underlying these differences? First, at the behavioral level, Embodied Cognition Theory and Social L2 Learning Theory consistently suggest that whole-body interactive experiences in VR environments facilitate foreign language vocabulary acquisition, and this facilitative effect should also manifest at the production level. However, this assumption may underestimate the potential interfering effects of redundancy in multimodal sensory information on foreign language vocabulary acquisition (Gao et al., 2016). Cognitive Load Theory posits that learning depends on intrinsic cognitive load (cognitive processing inherent to the learning task itself) and extraneous cognitive load (cognitive processing unrelated to learning objectives) (Sweller, 2011). Some studies have found that, compared to traditional learning environments, redundant information in VR environments may increase learners' extraneous cognitive load, potentially reducing learning effectiveness (Papin & Kaplan-Rakowski, 2024). Therefore, how VR environments affect foreign language word production remains unclear. Second, at the neural mechanism level, current relevant EEG evidence has employed event-related potential (ERP) analysis to examine temporal features of EEG, neglecting neural oscillatory activity during vocabulary processing. Traditional ERP analysis primarily reflects time-locked and phase-locked activity, revealing the temporal course of cognitive processing (Bastiaansen et al., 2012). Time-frequency analysis, by contrast, can better reflect time-locked but non-phase-locked activity, revealing changes in neural oscillatory patterns during cognitive processing (Schneider et al., 2021), which are often canceled out after averaging and cannot be revealed through ERP analysis. Therefore, it is necessary to combine ERP and time-frequency analyses to better reveal the neural mechanisms through which VR environments affect foreign language word production.

Drawing upon research in foreign language processing and memory, this study focuses on the P200 component and Late Positive Component (LPC) at the EEG temporal level. The P200 component is a positive wave occurring approximately 200 ms after stimulus presentation, reflecting early stages of stimulus processing and closely related to selective attention and working memory (Lijffijt et al., 2009). Research on foreign language processing has shown that foreign language words elicit larger P200 amplitudes compared to native language word reading tasks, possibly because foreign language word processing is more difficult and requires greater attentional involvement (Midgley et al., 2009). The LPC component is considered an important indicator of episodic memory retrieval, cognitive resource allocation, and elaborative processing (Che et al., 2021; Rugg & Curran, 2007). Previous foreign language processing studies have found that greater difficulty in retrieving target words during picture naming tasks consumes more cognitive resources, resulting in larger LPC components (Martin et al., 2013).

Neural oscillations reflect the rhythmic responses of brain neurons, and time-frequency analysis is a common method for parsing event-related neural oscillations (Ward, 2003). Currently, only one study has examined neural oscillations in VR environments and foreign language vocabulary acquisition: Zappa et al. (2024) used a passive listening task to compare neural oscillatory strength for foreign language words learned in VR environments versus action conditions, finding no differences between learning conditions in the μ (8-13 Hz) and β (13-30 Hz) bands. However, research using linguistic symbol learning conditions as a control found that multimodal perceptual symbol learning conditions significantly improved foreign language vocabulary acquisition and induced energy changes in the μ (8-13 Hz) and θ (4-7 Hz) bands (Ren et al., 2024). Accordingly, at the EEG time-frequency domain level, this study needs to focus on the μ band, as it is an effective marker of mirror neuron activity, which is considered the neural basis of perceptual simulation (Brunsdon et al., 2019). In this study, the lexical retrieval process during picture naming may be influenced by previously encoded perceptual-motor information, making perceptual simulation stronger for foreign language words learned in VR environments, thereby exhibiting μ-wave suppression. Additionally, previous studies have found that energy changes in θ-band neural oscillations are closely related to memory retrieval (Herweg et al., 2020) and syllable information processing (Jiang et al., 2020).

In summary, to examine the effects of VR learning contexts on foreign language word production, this study used a traditional picture-word paired learning environment as a reference and adopted a picture naming task. On one hand, behavioral indices were used to explore differences in production effects of foreign language words learned in VR versus picture-word environments. On the other hand, relevant EEG temporal features (P200 and LPC components) and time-frequency features (μ and θ bands) were combined to reveal the underlying neural mechanisms. Based on embodied cognition theory and previous research evidence, this study hypothesized that: (1) VR learning contexts would facilitate foreign language word production, with higher naming accuracy and faster naming speed than the picture-word environment; (2) The foreign language word production process in VR environments would involve less retrieval difficulty, eliciting smaller P200 and LPC amplitudes than the picture-word environment; (3) The perceptual-motor information presented in VR learning contexts would affect foreign language word production, manifesting as μ-wave energy suppression and θ-wave energy enhancement effects.

2.1 Participants

Sample size was calculated using G*Power software. Based on a medium effect size of f = 0.25, with $\alpha = 0.05$ and statistical power of 0.8, a minimum sample size of 24 participants was required for a one-factor two-level within-subjects design (Faul et al., 2007). The experiment actually recruited 35 university students, all right-handed with normal or corrected-to-normal vision. Six participants were excluded due to excessive EEG artifacts, leaving valid data from 29 participants (age: 20.10 $\pm$ 1.20 years, 25 females). All participants were native Chinese speakers with English learning experience but no prior knowledge of German. Participants signed informed consent forms and received compensation after the experiment.

2.2 Experimental Design

This experiment employed a one-factor two-level within-subjects design. The independent variable was learning context, including VR learning context and picture-word learning context. The dependent variable was production effects of words learned in different contexts, with behavioral indices including reaction time and accuracy in both immediate and delayed test phases, and EEG indices including P200 and LPC components elicited during the immediate test phase, as well as μ and θ waves in the time-frequency domain.

2.3 Experimental Materials

The learning materials consisted of 40 German words learned in both VR and picture-word contexts. The immersive VR learning context was created using Unity software (https://unity.com) to simulate daily home life scenes, including a living room, bedroom, and kitchen. The corresponding learning materials were 3D objects selected from a standardized 3D object database (Peeters, 2018). The picture-word learning context was presented via computer screen, with learning materials consisting of black-and-white line drawings selected from a standardized picture database (Zhang & Yang, 2003). Considering the compatibility between learning materials and the VR learning context, all vocabulary words were related to daily home life environments. Additionally, a group of Chinese-English bilinguals with English proficiency equivalent to the experimental participants was recruited to rate whether the selected German words were similar to any Chinese or English words they knew. Ratings on a five-point scale (1 = very dissimilar, 5 = very similar) indicated that the German words in this experiment were not similar to either Chinese or English words ($p < 0.001$), eliminating the influence of existing English vocabulary knowledge on German word learning.

2.4 Procedure

The experimental procedure consisted of a learning phase and a testing phase. The learning phase lasted 1-3 days, with immediate testing on Day 4 concurrently collecting behavioral and EEG data, and delayed testing one week later collecting only behavioral data.

During the learning phase, participants learned 40 German words in two different contexts. The German words were divided into two sets, with one set (20 words) learned in the picture-word context and the other set (20 words) learned in the VR context. Learning materials were counterbalanced across contexts: half of the participants learned the first set of German words in the VR context and the second set in the picture-word context, while the other half learned the first set in the picture-word context and the second set in the VR context. For the VR learning condition, participants needed to familiarize themselves with the VR environment and HTC VIVE equipment. Using a head-mounted display, participants could naturally interact with the VR environment and feel a sense of presence; using handheld controllers, they could select target objects and hear the corresponding object names. After familiarizing themselves with the VR equipment, participants selected target 3D objects via controller operation and then heard the corresponding German words through headphones. In the picture-word learning context, 2D line drawings were presented on a computer screen while the corresponding German words were played through headphones, and participants pressed a key to advance to the next word. Participants spent 15 minutes learning in each context daily, with learning order counterbalanced across participants. To control for fatigue effects, participants took adequate rest between different learning contexts.

[Figure 1: see original paper]

In the testing phase, all participants completed a picture naming task, requiring them to name pictures presented on a computer screen using German words. To ensure consistency between testing and learning phase materials, test pictures for words learned in the picture-word context were identical to those in the learning phase, while test pictures for words learned in the VR context were flat screenshots of the 3D objects. Each participant completed 80 trials, with each target picture presented twice. In each trial, a fixation point was presented for 500 ms, followed by a 200 ms blank screen, then the target picture. Participants needed to speak the corresponding German word for the presented picture, followed by a 1000 ms blank screen before the next trial. The testing task was programmed using E-prime software, with a voice box automatically recording picture naming reaction times. The delayed testing phase used the same task but recorded only behavioral data.

2.5 EEG Recording and Preprocessing

EEG data were collected using a Brain Products system, with EEG recorded from a 64-channel electrode cap based on the extended international 10-20 system. During online recording, the sampling rate was 1000 Hz, impedance between each electrode and scalp was less than 5 kΩ, and the online filter bandpass was 0.05-100 Hz, with FCz as the online reference electrode. During offline analysis, the reference electrode was converted to the average of bilateral mastoids, with a filter bandpass of 0.5-30 Hz. Independent Component Analysis (ICA) was used to correct artifacts such as blinks and electrocardiogram activity, and segments with incorrect responses or amplitudes exceeding $±$80 μV were excluded after segmentation and baseline correction.

For temporal ERP analysis, continuous EEG data were segmented into epochs from 200 ms before to 800 ms after picture presentation, with baseline correction referencing the -200 ms to 0 ms pre-picture EEG activity. Based on previous research, ERP mean amplitude analysis divided electrodes into five regions: frontal (F1, FZ, F2), fronto-central (FC1, FCZ, FC2), central (C1, CZ, C2), centro-parietal (CP1, CPZ, CP2), and parietal (P1, PZ, P2). Statistical analyses of P200 (150-250 ms) and LPC (400-550 ms) components across different learning contexts were conducted by averaging amplitudes across electrodes within specified regions.

For time-frequency analysis, the short-time Fourier transform method was used with a 200 ms Hanning window to convert EEG signals from all trials into time-frequency representations. First, EEG data were re-segmented from 500 ms before to 800 ms after picture presentation. Then, for each trial, a short-time Fourier transform was performed at each time point (in 1 ms steps) and each frequency (1-30 Hz, in 1 Hz steps), with the resulting spectrograms baseline-corrected at each frequency (Fernandez et al., 2019; Liu et al., 2017). Based on previous research and the current study' s spectrograms, mean energy of μ (8-12 Hz) and θ (4-7 Hz) bands in the 150-250 ms and 400-550 ms time windows was selected for statistical analysis (Ren et al., 2024; Jiao, Schwieter et al., 2025).

For statistical analysis, this experiment employed both traditional hypothesis testing and Bayesian factor analysis. Traditional hypothesis testing was conducted using R software, while Bayesian factor analysis was performed using JASP software (Wagenmakers et al., 2018). Bayesian factor analysis can quantify the strength of evidence supporting the alternative hypothesis (H1) relative to the null hypothesis (H0), rather than merely making a dichotomous decision about H0, thereby addressing limitations of traditional hypothesis testing (Hu et al., 2018). The prior distribution used a Cauchy distribution with γ 0.707 (Jeffreys, 1961).

3.1 Behavioral Results

Immediate test phase: First, incorrect response data (9.18%), extreme data ($<$ 200 ms or $>$ 5000 ms; 0.73%), and reaction time data beyond $±$2.5 SD (3.41%) were removed. Then, paired-samples t-tests were conducted on reaction times and accuracy rates (Figure 2 [Figure 2: see original paper]) for VR and picture-word conditions. Reaction time results showed no significant difference between VR context (1375.66 $\pm$ 227.07, mean $\pm$ SD) and traditional picture-word context (1342.56 $\pm$ 223.25), $t(28) = 1.37$, $p = 0.18$, $d = 0.14$, 90% CI = [-0.03, 0.32]. Bayesian factor analysis further indicated a Bayes factor of

BF01 = 2.17 for reaction time, meaning the likelihood of obtaining the current data under the null hypothesis (assuming no effect) was 2.17 times greater than under the alternative hypothesis (assuming an effect), indicating weak evidence supporting acceptance of H0 (Jeffreys, 1961)—that is, no significant difference in naming reaction times for foreign language words learned in different contexts. Accuracy results showed that participants' naming accuracy for words learned in the VR context (0.92 $\pm$ 0.09) was significantly higher than in the traditional picture-word context (0.89 $\pm$ 0.10), t(28) = 2.40, p = 0.02, d = 0.33, 90% CI = [0.09, 0.56]. Bayesian factor analysis further indicated BF01 = 0.41 for accuracy, showing weak evidence supporting the alternative hypothesis H1—that is, naming accuracy was significantly higher for words learned in VR context compared to picture-word context.

Delayed test phase: Data processing criteria and analysis methods were identical to above, removing incorrect response data (8.96%), extreme data (1.25%), and reaction time data beyond \$±\$2.5 SD (3.15%). Reaction time results showed no significant difference between VR context (1426.05 $\pm$ 199.91) and traditional picture-word context (1447.69 $\pm$ 200.76) after a one-week delay, t(28) = 1.00, p = 0.32, d = 0.10, 90% CI = [-0.07, 0.28]. Bayesian factor analysis also indicated moderate evidence supporting acceptance of H0 for reaction time (BF01 = 3.20) (Jeffreys, 1961), meaning no significant difference in naming reaction times for foreign language words learned in the two contexts. Accuracy results showed no significant difference between naming accuracy for words learned in VR context (0.92 $\pm$ 0.07) and picture-word context (0.90 $\pm$ 0.10), t(28) = 1.90, p = 0.06, d = 0.25, 90% CI = [0.02, 0.47]. Bayesian factor analysis also indicated weak evidence supporting acceptance of H0 for accuracy (BF01 = 1.04), meaning no significant difference in naming accuracy for words learned in the two contexts.

[Figure 2: see original paper]

3.2 ERP Temporal Results

Grand-average ERP waveforms from the immediate test phase are shown in Figure 3 [Figure 3: see original paper]. A two-way repeated measures ANOVA with factors of learning context (2) and electrode location (5) was conducted on mean P200 amplitudes. Results showed a significant main effect of learning context, F(1, 28) = 19.85, p < 0.001, η^2p = 0.42, 90% CI = [0.18, 0.59], with a Bayes factor of BF01 = 0.01, indicating extremely strong evidence supporting the alternative hypothesis H1. This manifested as significantly larger P200 components for words learned in picture-word context compared to VR context. The main effect of electrode location was not significant, F(4, 112) = 2.19, p = 0.07, η^2p = 0.07, 90% CI = [0.00, 0.13], BF01 = 1.00. Importantly, the interaction between the two factors was significant, F(4, 112) = 25.04, p < 0.001, η^2p = 0.47, 90% CI = [0.35, 0.56], with this effect also receiving extremely strong evidence support (BF01 = 3.67\$×10^{-12}\$). Simple effects analysis revealed significant differences in P200 amplitudes between learning contexts in frontal (p < 0.001, η^2p = 0.54), fronto-central (p < 0.001, η^2p = 0.50), central (p <

0.001, η^2p = 0.43), and centro-parietal (p = 0.002, η^2p = 0.30) regions, but not in parietal region (p = 0.09, η^2p = 0.10).

For LPC mean amplitudes, the two-way repeated measures ANOVA showed a significant main effect of electrode location, F(4, 112) = 26.48, $p < 0.001$, η^2p = 0.48, 90% CI = [0.37, 0.56], with Bayes factor BF01 = 2.82×10^{-13}, indicating extremely strong evidence supporting the alternative hypothesis H1 (Jeffreys, 1961). The main effect of learning context was not significant, F(1, 28) = 3.17, p = 0.08, η^2p = 0.10, 90% CI = [0.00, 0.29], with Bayesian factor analysis showing BF01 = 0.90, also indicating weak evidence supporting the alternative hypothesis H1. Additionally, the interaction between the two factors was significant, F(4, 112) = 9.63, $p < 0.001$, η^2p = 0.26, 90% CI = [0.13, 0.35]. Simple effects analysis showed that LPC amplitudes for words learned in picture-word context were significantly larger than those for VR context in frontal (p = 0.03, η^2p = 0.15), fronto-central (p = 0.03, η^2p = 0.14), and central (p = 0.04, η^2p = 0.13) regions, but not in centro-parietal (p = 0.17, η^2p = 0.06) or parietal (p = 0.59, η^2p = 0.01) regions. Bayesian factor analysis for the interaction yielded BF01 = 0.86×10^{-4}, indicating extremely strong evidence supporting the alternative hypothesis H1 (Jeffreys, 1961).

[Figure 3: see original paper]

3.3 Time-Frequency Results

Time-frequency analysis is shown in Figure 4 [Figure 4: see original paper]. Repeated measures ANOVA on μ-band energy in the 150-250 ms time window revealed a significant main effect of learning context, F(1, 28) = 14.54, $p < 0.001$, η^2p = 0.34, 90% CI = [0.12, 0.53], with Bayes factor BF01 = 0.02, indicating very strong evidence supporting the effect of learning context (alternative hypothesis H1). The main effect of electrode location was significant, F(4, 112) = 15.45, $p < 0.001$, η^2p = 0.35, 90% CI = [0.23, 0.48], BF01 = 3.23×10^{-8}. Moreover, the interaction between the two factors was significant, F(4, 112) = 11.50, $p < 0.001$, η^2p = 0.29, 90% CI = [0.16, 0.38], BF01 = 9.05×10^{-6}, indicating extremely strong evidence supporting the alternative hypothesis H1 (Jeffreys, 1961). Simple effects analysis showed that μ-wave energy for words learned in VR context was significantly lower than for picture-word context in frontal ($p < 0.001$, η^2p = 0.38), fronto-central ($p < 0.001$, η^2p = 0.39), central (p = 0.001, η^2p = 0.32), and centro-parietal (p = 0.03, η^2p = 0.15) regions, but not in parietal region (p = 0.26, η^2p = 0.04).

Repeated measures ANOVA on μ-band energy in the 400-550 ms time window revealed a significant main effect of learning context, F(1, 28) = 6.61, p = 0.01, η^2p = 0.19, 90% CI = [0.02, 0.39], BF01 = 0.27, indicating moderate evidence supporting the learning context effect. The main effect of electrode location was not significant, F(4, 112) = 0.52, p = 0.72, η^2p = 0.01, 90% CI = [0.00, 0.04], BF01 = 29.47. The interaction between the two factors was significant, F(4, 112) = 4.80, p = 0.001, η^2p = 0.15, 90% CI = [0.04, 0.23], BF01 = 0.03, indicating that the significant interaction received very strong evidence support (Jeffreys, 1961). Simple effects analysis showed that μ-wave energy for words learned in VR context was significantly lower than for picture-word context in frontal ($p = 0.006$, $\eta^2p = 0.24$), fronto-central ($p = 0.009$, $\eta^2p = 0.22$), and central ($p = 0.03$, $\eta^2p = 0.15$) regions, but not in centro-parietal ($p = 0.36$, $\eta^2p = 0.03$) or parietal ($p = 0.12$, $\eta^2p = 0.08$) regions.

Repeated measures ANOVA on θ-band energy in the 150-250 ms time window revealed a significant main effect of learning context, $F(1, 28) = 10.64$, $p = 0.003$, $\eta^2p = 0.27$, 90% CI = [0.07, 0.47], BF01 = 0.07. The main effect of electrode location was significant, $F(4, 112) = 12.58$, $p < 0.001$, $\eta^2p = 0.31$, 90% CI = [0.18, 0.40], BF01 = 1.21$\times10^{-6}$. Additionally, the interaction between the two factors was significant, $F(4, 112) = 8.62$, $p < 0.001$, $\eta^2p = 0.23$, 90% CI = [0.11, 0.33], BF01 = 3.10$\times10^{-4}$, indicating extremely strong evidence supporting the alternative hypothesis H1. Simple effects analysis showed that enhanced θ-wave energy for words learned in picture-word context was primarily observed in frontal ($p = 0.003$, $\eta^2p = 0.27$), fronto-central ($p = 0.002$, $\eta^2p = 0.29$), and central ($p = 0.002$, $\eta^2p = 0.28$) regions, with no significant differences in θ-wave energy between the two conditions in other brain regions ($ps > 0.05$).

Repeated measures ANOVA on θ-band energy in the 400-550 ms time window revealed a significant main effect of learning context, $F(1, 28) = 9.16$, $p = 0.005$, $\eta^2p = 0.25$, 90% CI = [0.05, 0.45], BF01 = 0.11, indicating moderate evidence supporting this effect. The main effect of electrode location was not significant, $F(4, 112) = 0.53$, $p = 0.71$, $\eta^2p = 0.02$, 90% CI = [0.00, 0.04], with Bayes factor BF01 = 27.91, indicating strong evidence supporting acceptance of H0 (no electrode location effect). The interaction between the two factors was significant, $F(4, 112) = 3.67$, $p = 0.008$, $\eta^2p = 0.11$, 90% CI = [0.02, 0.19], BF01 = 0.16, indicating moderate evidence supporting the alternative hypothesis H1 (Jeffreys, 1961). Simple effects analysis showed that words learned in picture-word context exhibited more significant θ-wave energy enhancement in frontal ($p = 0.01$, $\eta^2p = 0.20$), fronto-central ($p = 0.01$, $\eta^2p = 0.21$), and central ($p = 0.005$, $\eta^2p = 0.25$) regions.

[Figure 4: see original paper]

This study used a picture naming task to investigate the effects of immersive VR learning contexts on foreign language word production. Behavioral results revealed that words learned under VR conditions showed higher accuracy in immediate testing. ERP analysis in the immediate test phase found that, compared to picture-word conditions, words learned in VR conditions elicited smaller P200 and LPC components. Time-frequency analysis revealed more pronounced μ-wave suppression and accompanying θ-wave energy reduction in VR conditions compared to picture-word conditions. These results provide EEG evidence for the effects of VR contexts on foreign language word production and support embodied cognition theory.

First, the behavioral results of this study indicate that, compared to traditional picture-word contexts, foreign language words learned in VR contexts show clear advantages in word production, manifested as higher picture naming accuracy in immediate testing, consistent with previous research. For example, Tai et al. (2022) compared vocabulary learning effects between 2D video and VR conditions, finding better test performance for words learned in VR conditions. Fuhrman et al. (2021) found that when participants learned vocabulary by performing relevant actions in VR contexts, accuracy in picture-word matching tasks was significantly higher than in conditions with irrelevant actions. These studies suggest that multimodal information integration promotes vocabulary acquisition. In immersive virtual environments, foreign language words are concretized as real objects, allowing learners to enhance physical immersion through vision, audition, and touch, achieving immersive, situated learning. Simultaneously, learners can explore virtual environments from a first-person perspective, manipulate 3D objects corresponding to foreign language words, and deeply integrate linguistic and non-linguistic information related to the words (Li & Jeong, 2020). It should be noted that previous research on VR and foreign language learning has focused on vocabulary comprehension, emphasizing that multimodal perceptual-motor information improves foreign language vocabulary comprehension—such as shallow-level word recognition (Fuhrman et al., 2021; Liu et al., 2026) and deep-level semantic integration (Jiao, Lin et al., 2025; Liu et al., 2024). The present study, however, focused on the production level of foreign language vocabulary. In this study, the observed differences in foreign language word production during immediate testing may be related to the immersive and interactive features of VR contexts. Compared to traditional picture-word learning contexts, the visual, auditory, and motor non-linguistic information provided by VR contexts may enhance memory and representation of foreign language vocabulary knowledge, enabling participants to achieve more accurate lexical retrieval in naming tasks. This speculation aligns with previous empirical research (Legault, Fang et al., 2019) and embodied cognition theory (Barsalou, 2008).

Second, regarding the EEG temporal features of the word production process, this study found that words learned in VR contexts elicited smaller P200 and LPC components, primarily in frontal, fronto-central, and central regions. Previous research indicates that the P200 component is related to selective attention, with larger P200 components meaning greater processing difficulty and increased attentional resource demands (Midgley et al., 2009). In this study, changes in P200 amplitude may be related to information encoding during the learning phase. Compared to single-modal information encoding in picture-word contexts, VR learning contexts may facilitate integration of multimodal information with semantic networks, strengthening connections between lexical information and non-linguistic sensory information (pictures, actions, etc.). Therefore, when pictures are presented during the testing phase, words learned in picture-word contexts can only activate abstract native language symbols or single picture information, whereas words learned in VR contexts may activate more non-linguistic information, reducing lexical retrieval and production difficulty, manifested as reduced P200 amplitude. It should be noted that since this study did not collect neural activity during the learning encoding phase, this speculation warrants further investigation. Regarding the late LPC component that reflects episodic memory retrieval and strategic processing (Kutas & Federmeier, 2011; Rugg & Curran, 2007), this study found that words learned in VR conditions elicited significantly smaller LPC components than picture-word conditions. In VR conditions, the foreign language word production process could utilize previously encoded episodic memory cues to achieve more automatic lexical processing, whereas in picture-word conditions, participants could not engage in episodic memory encoding, making word production more dependent on strategic processing and requiring more attentional resources for semantic retrieval and lexical extraction. Consistently, previous foreign language vocabulary acquisition research in traditional contexts has also found that semantic retrieval and integration effects are manifested in LPC components when learning foreign language words through word-word/picture-word matching, indicating involvement of strategic processing mechanisms (Lei et al., 2022).

Third, the effects of the two learning contexts on word production were also reflected at the EEG time-frequency level, manifested as energy changes in μ and θ bands. On one hand, compared to picture-word contexts, foreign language words learned in VR contexts showed more pronounced μ-wave suppression, consistent with previous research on perceptual symbol effects (Ren et al., 2024). Energy changes in the μ band are related to mirror neuron activity, reflecting cognitive processing involving perceptual simulation (Brunsdon et al., 2019). In this study, participants engaged in situated multimodal cognitive processing through perceptual-motor simulation during VR vocabulary learning, such as real-time interaction with 3D objects using head-mounted VR equipment. This learning process was accompanied by active neural mirror neurons, which may have led to deep binding between perceptual simulation information and foreign language word information, activated during word production, thereby inducing μ-wave suppression. However, foreign language words learned in picture-word contexts could not achieve sufficient perceptual simulation during encoding, with weaker mirror neuron activity, making μ-wave suppression less pronounced. On the other hand, this study did not observe θ-band energy enhancement effects in VR contexts, contrary to our expectations. In previous vocabulary comprehension tasks, researchers found that learning conditions rich in perceptual information produced θ-wave energy enhancement effects, rather than the θ-wave energy reduction observed for VR conditions in this study. This inconsistent result may be due to differences in cognitive demands of foreign language word processing: the word production task in this study required correct reporting of foreign language words corresponding to pictures, focusing on phonological information output, whereas previous vocabulary comprehension tasks required recognition of foreign language words, focusing on phonological or orthographic information input (Ren et al., 2024). Of course, this study did not simultaneously examine the neural mechanisms of VR contexts on both foreign language word production and comprehension, and this speculation requires further verification.

Finally, at the theoretical contribution level, on one hand, this study reveals the production advantages of foreign language words learned in VR contexts, providing new empirical support for embodied cognition theory (Barsalou, 2008) and social L2 learning theory (Li & Jeong, 2020). The perceptual-motor interaction in VR environments allows learners to deeply integrate foreign language word information with multimodal perceptual-motor information, strengthening connections between lexical representations and sensory features, thereby promoting foreign language vocabulary acquisition and production (Shi et al., 2022; Yang & Wang, 2023). This inference aligns with previous related brain imaging research. For example, Legault and Fang et al. (2019) found that foreign language learning effects in VR contexts were closely related to cortical thickness in the right superior parietal lobule, a key brain region responsible for perceptual-action integration and embodied representation. These research findings collectively suggest that perceptual-motor interaction may be an important mechanism through which VR contexts facilitate foreign language word production. On the other hand, this study helps us deepen our understanding of the role patterns of cognitive load in VR learning. Although Cognitive Load Theory (Sweller, 2011) suggests that multimodal redundant information in learning contexts may increase extraneous cognitive load and interfere with learning, the immersion and interactivity of VR contexts can enhance learners' interest, motivation, and self-efficacy, thereby increasing intrinsic cognitive load and positively affecting learning (Makransky & Petersen, 2021). When intrinsic cognitive load exceeds extraneous cognitive load, learners can allocate more cognitive resources to elements relevant to the learning content itself, thereby forming meaningful generative learning (Mayer, 2005). Therefore, we speculate that the effects of VR contexts are not simply about increasing or decreasing cognitive load, but rather result from the combined action of different types of cognitive load. Future research should directly investigate the effects of different types of cognitive load on foreign language vocabulary learning in VR contexts.

Additionally, at the practical application level, with the deepening implementation of China' s digital education strategy, applying AI technologies such as VR to the educational and teaching process is an inevitable trend. This study comprehensively employs research methods from AI, brain science, and cognitive science to provide scientific evidence for the digital transformation of foreign language education. Currently, while some studies recognize the limitations of traditional teaching environments, research on foreign language acquisition in real-world environments is scarce (Liu et al., 2021), and its conclusions are difficult to generalize because not all learners have opportunities to immerse themselves in real environments for foreign language acquisition (e.g., living in Korea to learn Korean). With advances in AI technology, VR technology has evolved from desktop environments to immersive environments, from single-hand control to full-body tracking, and from single-user to multi-user interaction, providing rich possibilities for empirical research and pedagogical applications in foreign language acquisition. However, it should be noted that more multimodal information in VR environments is not necessarily better. In line with Cognitive Load Theory (Sweller, 2011), VR learning contexts contain more redundant information than picture-word contexts, requiring greater inhibitory control of irrelevant information and imposing heavier cognitive load, which may prevent learners from effectively selecting and integrating target information, thereby limiting foreign language vocabulary acquisition. Some empirical research has found that low-immersion VR learning environments using desktop equipment are more conducive to foreign language vocabulary acquisition than high-immersion VR learning environments using head-mounted displays, with better vocabulary memory effects (Papin & Kaplan-Rakowski, 2024). In this regard, Makransky and Petersen (2021) propose that immersive learning is influenced not only by embodied factors but also by multiple factors including interest, motivation, self-efficacy, and cognitive load. In this study, we focused on the multimodal information provided by VR technology, particularly the effects of embodied factors on foreign language word production, without considering the synergistic mechanisms between VR-assisted embodied factors and other factors. Future research should comprehensively investigate the combined effects of cognitive and affective factors in VR contexts and consider how to apply these findings in real teaching situations.

This study has certain limitations that should be addressed in future research. First, although the VR learning context in this study could effectively simulate real environments, there were some differences in color richness and visual complexity between 3D objects in VR contexts and line drawings in picture-word contexts. Future studies should consider using real photographs as control conditions to control for material differences affecting foreign language vocabulary learning in VR contexts. Meanwhile, given the encoding specificity principle in memory research (Tulving & Thomson, 1973), the current study controlled material consistency between testing and learning phases as much as possible, but this created differences in test pictures between VR and picture-word conditions. Future research should set up different types of test pictures to deeply investigate the specificity and stability of how learning contexts affect foreign language word production. Additionally, due to sampling limitations, this study only had Chinese native speakers learn German, with some degree of gender imbalance. Future studies should expand sampling ranges and combine demographic variables (e.g., gender, age) and foreign language vocabulary types (e.g., Japanese words, pseudowords) to further explore the neural mechanisms through which VR contexts affect foreign language vocabulary learning.

This study, based on EEG and VR technology, reveals the effects of immersive VR contexts on foreign language word production and their neural mechanisms. Behavioral results show that, compared to traditional picture-word learning contexts, foreign language words learned in VR contexts demonstrate certain production advantages in immediate testing. Related neural indices indicate that VR contexts enhance embodied information processing related to foreign language words by integrating multimodal information through perceptual-motor experiences, thereby facilitating lexical retrieval and production. This study provides behavioral and EEG evidence for embodied cognition theory from the perspective of foreign language word production. The study has limitations in participant sampling and material matching, and future research should combine different demographic and cognitive factors to further explore the effects of VR learning contexts on foreign language word production.

References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59(1), 617-645.

Bastiaansen, M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: oscillatory neuronal dynamics. *The Oxford handbook of event-related potential components*, 31-50.

Brunsdon, V. E., Bradford, E. E., & Ferguson, H. J. (2019). Sensorimotor mu rhythm during action observation changes across the lifespan independently from social cognitive processes. *Developmental Cognitive Neuroscience*, 38, 100659.

Che, X., Xu, H., Wang, K., Zhang, Q., & Li, S. (2021). Precision requirement of working memory representations influences attentional guidance. *Acta Psychologica Sinica*, 53(7), 694-713.

Essoe, J. K. Y., Reggente, N., Ohno, A. A., Baek, Y. H., Dell' Italia, J., & Rissman, J. (2022). Enhancing learning and retention with distinctive virtual reality environments and mental context reinstatement. *npj Science of Learning*, 7(1), 31.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* Behavior Research Methods*, 39(2), 175-191.

Fernandez, C. B., Litcofsky, K. A., & Van Hell, J. G. (2019). Neural correlates of intra-sentential code-switching in the auditory modality. *Journal of Neurolinguistics*, 51, 17-41.

Fuhrman, O., Eckerling, A., Friedmann, N., Tarrasch, R., & Raz, G. (2021). The moving learner: Object manipulation in virtual reality improves vocabulary learning. *Journal of Computer Assisted Learning*, 37(3), 699-712.

Gao, Y., Liu, D. J., Huang, Z. Z., & Huang, R. H. (2016). The core factors and challenges of virtual reality technology enhanced learning. *Journal of Educational Technology & Research*, 37(10), 77-87+103.

Herweg, N. A., Solomon, E. A., & Kahana, M. J. (2020). Theta oscillations in human memory. *Trends in Cognitive Sciences*, 24(3), 208-227.

Hu, C. P., Kong, X., Wagenmakers, E. J., Ly, A., & Peng, K. (2018). The Bayes factor and its implementation in JASP: A practical primer. *Advances in Psychological Science*, 26(6), 951-965.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Jiang, Y., Cai, X., & Zhang, Q. (2020). Theta band (4~8 Hz) oscillations reflect syllables processing in Chinese spoken word production. *Acta Psychologica Sinica*, 52(10), 1199-1211.

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*.

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, 101232.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62(1), 621-647.

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.

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.

Lei, D., Liu, Y., & van Hell, J. G. (2022). Novel word learning with verbal definitions and images: Tracking consolidation with behavioral and event-related potential measures. *Language Learning*, 72(4), 941-979.

Li, P., & Jeong, H. (2020). The social brain of language: grounding second language learning in social interaction. *npj Science of Learning*, 5(1), 8.

Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., ···& Swann, A. C. (2009). P50, N100, and P200 sensory gating: relationships with behavioral inhibition, attention, and working memory. *Psychophysiology*, 46(5), 1059-1068.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, 20(12), 1507-1515.

Liu, C., de Bruin, A., Jiao, L., Li, Z., & Wang, R. (2021). Second language learning tunes the language control network: a longitudinal fMRI study. *Language, Cognition and Neuroscience*, 36(4), 462-473.

Liu, C., Feng, J., Schwieter, J. W., Geng, J., & Jiao, L. (2026). Neural correlates of novel word learning in an immersive virtual reality environment. *Journal of Neurolinguistics*, 78, 101300.

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.

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.

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (Camil): A theoretical research-based model of learning in immersive virtual reality. *Educational

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.