**Corresponding author:**
Cong Liu
Department of Psychology
Qingdao University, No.308, NingXia R.
Qingdao, 266071, China
Email: congu902@gmail.com

**CAMBRIDGE UNIVERSITY PRESS**

# An ERP study on novel word learning in an immersive virtual reality context

Lu Jiao[1], Mengrui Zhu[1], Zijie Xu[2], Guanzhu Zhou[1], John W. Schwieter[3] and Cong Liu[1] 

[1]Department of Psychology, Normal College & School of Teacher Education, Qingdao University, Qingdao, China; [2]School of Foreign Languages, Qingdao University, Qingdao, China and [3]Language Acquisition, Multilingualism, and Cognition Laboratory / Bilingualism Matters @ Wilfrid Laurier University, Waterloo, Canada

## Abstract

In this behavioral and electrophysiological study, we compare novel word learning, particularly lexical form acquisition, in an immersive virtual reality (VR) context with a picture-word (PW) association context. We also test whether inhibitory control and age of second language acquisition (L2 AoA) have modulating effects. Chinese speakers of L2 English learned two sets of German words, one set in each of the contexts. Behavioral performance from a subsequent recognition task indicated that responses to VR-learned words were faster than PW-leaned words. ERPs revealed that VR-learned words elicited more negative N100 and N200 waveforms than PW-learned words. Moreover, a significant relationship between L2 AoA and N200 amplitude was observed for VR-learned words. Taken together, the results suggest that the multi-sensory, interactive experience simulated by an immersive VR context has a positive effect on early lexical form acquisition of novel words.

## 1. Introduction

It is estimated that approximately 60% of the world's population speaks two or more languages. Indeed, with global economic development and cultural exchanges at home and abroad, the demand for mastering more than one language continues to rise. Research on bi/multilingualism has shown, however, that foreign language learning can present challenges for many individuals, particularly for adults. While some assumptions emphasize a critical period of language learning (Johnson & Newport, 1989), other theories and empirical evidence point out that learning contexts play a key role (Kuhl et al., 2003; Laufer & Hulstijn, 2001; Legault et al., 2019a; Li & Jeong, 2020). For example, compared to traditional classroom settings, immersive study abroad settings offer a contextualized, real-life experience that can enhance foreign language learning outcomes (Jackson & Schwieter, 2019; Klassen et al., 2021; Linck et al., 2009). With the widespread promotion and development of innovative technologies, immersive virtual reality (VR) contexts may have similar effects given their ability to simulate language-enriching experiences through exposure to multiple types of sensory and motor information. However, very little is known about how VR contributes to foreign language learning, especially the early stage of lexical form acquisition. Accordingly, in the present study, we analyze behavioral and electrophysiological data to investigate the effects of immersive VR on lexical form acquisition in a new foreign language, and compare these effects to a traditional picture-word (PW) association context.

### 1.1 Novel word learning in a multisensory context

Current evidence and theories propose that compared to a unisensory context, the presence of complementary information across multiple sensory modalities during learning (e.g., flash cards, video) is beneficial for learning performance (Mayer et al., 2015; von Kriegstein & Giraud, 2006). For the novel word learning domain, learning context is also crucial such that rich sensory information and body movements (e.g., gestures) can facilitate the learning process. In a recent study, Jeong et al. (2021) asked Japanese first language (L1) speakers to learn novel words in Korean in either a multisensory or translation context. Participants in the translation context learned the novel words by hearing them and seeing their written L1 translation. While in the multisensory context with a rich sensory experience, participants watched video clips in which novel words were used in real-life communicative situations. Results revealed that novel word learning in the multisensory context largely recruited brain regions typically associated with social and perception-action-related processing. In addition, word learning with a rich sensory experience was positively correlated to subsequent lexical retrieval performance, suggesting that a learning context involving social, multisensory, action-perception processing may lead to more efficient encoding, retention, and retrieval of novel

words (see also Jeong et al., 2010). One of the theoretical motivations for investigating the role of a rich sensory experience in novel word learning stems from embodied cognition theory which emphasizes that whole-body interactions with an environment shape experience and knowledge (Barsalou, 2008). Mayer et al. (2015) compared novel word learning in a verbal learning condition, a viewed picture condition, and a self-performed gesture condition. The results showed that the gesture condition with multiple sensory information enhanced learning by engaging visual and motor brain regions.

It is widely accepted that novel word learning in multisensory contexts provides the opportunity for learners to make direct links between novel words and the concepts they represent (Jeong et al., 2010; Lan et al., 2015; Mayer et al., 2015). According to social L2 learning theory (SL2, Li & Jeong, 2020), in traditional classroom contexts with very limited sensory experiences, individuals often learn novel words through picture-word or word-word associations in which there is a significant reliance on learning the novel words through their L1 translation equivalents (Kroll & Stewart, 1994). Whereas in multisensory contexts that include rich sensory and motor experiences, learning novel words is reinforced by perceiving and interacting with experiences in the environment. Consequently, learners establish an L1-like lexical-semantic representation in the novel words that contains rich sensory experiences, such as perceptual, spatial, and motor features (Li & Jeong, 2020).

## 1.2 Immersive virtual reality and novel word learning

Although there are crucial benefits of novel word learning in multisensory contexts, it may not be feasible for or accessible to every learner. Nonetheless, recent advances in technology allow researchers to create an immersive virtual reality (VR) environment which can be used to explore novel word learning in simulated real-life environments with rich sensory and motor information (Fuhrman et al., 2021; Legault et al., 2019a; Li & Lan, 2021). Immersive VR technology can provide a high degree of immersion through an interactive environment that is similar to real-life situations (Legault et al., 2019a). Using a head-mounted display, individuals are immersed in a 360° immersive experience in which they can turn and look in any direction by moving their eyes and body.

Based on the evidence examining the role of a rich sensory experience on novel word learning using immersive VR technology, the mainstream opinion proposes that immersive VR can be an effective tool for novel word learning (for a review, see Li & Lan, 2021). One study by Legault et al. (2019a) asked English monolinguals to learn two sets of L2 Chinese words in either a word-word association context or immersive VR context. Behavioral data collected immediately after the treatment revealed a positive effect of immersive VR context on L2 word learning. The authors argued that when learners are exposed to rich sensory and motor information in VR contexts, they perform better at allocating attentional resources and inhibiting irrelevant information, thus having a positive effect on learning performance. Fuhrman et al. (2021) also reported significant effects when participants learned novel words in VR contexts that incorporated motor enactment (e.g., hand gestures and object manipulation).

However, recent studies have revealed inconsistent findings, showing that the higher immersive experience achieved by VR technology may not necessarily lead to better performance of novel word learning (Papin & Kaplan-Rakowski, 2022). One

potential explanation of these divergent results is that when immersed in a VR context, learners bear an excessive cognitive load and their level of distraction is high due to rich complementary information (Papin & Kaplan-Rakowski, 2022; Sweller et al., 2011). The Cognitive Affective Model of Immersive Learning (CAMIL) points out that the increased visual field in an immersive VR context brings an abundant of subtle details that may not be necessary for learning but are processed, leading to increased processing load and inhibition demands (Makransky et al., 2021). Therefore, the individual difference in inhibitory control may be associated with novel word learning in VR context. Specifically, learners with more efficient inhibitory control may allocate more attention to learning targets with fewer distractions (Kapa & Colombo, 2014).

Moreover, some studies have examined the role of prior language experience in novel word learning and note that learning performance may be influenced by prior learning contexts (Bogulski et al., 2019; Hirosh & Degani, 2018). According to the viewpoint of the transfer hypothesis, for the bilingual learners in present study, if the learning context of a novel language (e.g., classroom) resembles the context in which their L2 was learned in (i.e., also in a classroom), this shared context may facilitate learning of the novel language (Nair et al., 2016). Given that all participants of the present study learned their L2 through classroom context with a lack of immersion experience, we selected the age of L2 acquisition (L2 AoA) of bilinguals as reflective of their prior language learning experience. To some extent, L2 learned at a late age with less classroom learning experience is expected to better adapt to an immersive context, while an early age may work against immersion effects. Although preliminary findings on novel word learning in a VR context suggest similar sensitivity to individual differences (Legault et al., 2019a), more research is needed to fully understand these effects by considering inhibitory control and prior language experience.

## 1.3 Present study

The present study used EEG technology to investigate how an immersive VR context affects novel word learning, particularly early lexical form acquisition, and compares these effects with a picture-word (PW) association learning context. A group of Chinese–English bilinguals received three days of training in which they learned novel words in German in both a PW association context and an immersive VR context. After the training, participants completed a recognition task to test whether the target words had been learned or not. During the task, we measured participants' electrophysiological activity using EEG. Moreover, before training, we administered a language background questionnaire and a modified Flanker task to collect data on potential individual differences in L2 AoA and inhibitory control ability (Jiao et al., 2019, 2022).

The high-temporal resolution of EEG data allows us to investigate the exact time-course of novel word learning. Based on previous work and our research objectives, we focus on mean amplitudes of three negative-going ERP components: N100, N200, and N400 (Basirat et al., 2018; Liu & van Hell, 2020). The N100 occurs just after stimulus onset and has been associated with visually selective attention and sensory/perceptual processing (Biau et al., 2018; Vogel & Luck, 2000). The N200 component has a scalp distribution across the fronto-central electrode sites and is widely defined as a marker of early lexical selection and phonological processing (Connolly & Phillips, 1994; Friedrich &

Friederici, 2008; van den Brink et al., 2001), as well as conflict monitoring (Mathalon et al., 2003; van Veen & Carter, 2002). The N400 component, which peaks around 400 ms after stimulus onset (Kutas & Federmeier, 2011), is said to reflect lexical-semantic access of words (Liu & van Hell, 2020).

There are three hypotheses tested in the present study. First, in the behavioral performance, we expect that the multisensory learning context achieved by VR technology will enhance learning outcomes in the recognition task compared with the PW context. Second, with respect to electrophysiological activity, we expect that superior performance of VR-learned words will emerge in the N100 and N200 components, but not in the N400 component. The recognition task in the present study mainly measures the initial formation of lexical representations, while the N400 component is widely associated with semantic access of lexical representations. Third, based on the viewpoints of the CAMIL model and transfer hypothesis, we hypothesize that individual differences in inhibitory control and prior language experience will be related to learning outcomes of VR-learned words. We anticipate that a small flanker effect will be associated with better performance because efficient inhibitory control will decrease distractions in the VR context. Moreover, we expect that the early L2 AoA of bilingual participants may hinder learning performance because individuals who have prior language learning experience in classroom settings may be harder to adapt to an immersive learning context.

## 2. Method

### 2.1 Participants

Thirty-five Chinese learners of L2 English were recruited to take part in the study. Five participants were excluded because of excessive EEG artifacts, leaving data from 30 participants (22 females, 8 males) to form part of the statistical analyses. All participants were right-handed adults (mean age = 20, range = 18–23) with normal or corrected-to-normal vision. Before beginning the formal experiment, the participants were asked to complete a language background questionnaire in which they also rated their L1 and L2 proficiency levels. The participants reported having no experience living abroad and that they had begun learning L2 English on average at 8.76 (SD = 1.35) years. The self-ratings of language proficiency were based on a 7-point scale (1 = very poor, 7 = excellent) and revealed that L1 Chinese proficiency (mean = 6.18, SD = .66) was significantly higher than L2 English (mean = 3.93, SD = .79), $t = 10.72$, $p < .01$. Moreover, all participants reported no prior knowledge of German, the language in which novel words were to be learned in the experiment. The local ethics committee approved the study, and all individuals provided written consent prior to participating in the experiment and received a modest payment for their participation.

### 2.2 Materials

A total of 40 German words were auditorily presented to participants who learned in two learning conditions (20 words learned in an immersive VR condition and 20 words learned in a PW condition). All 40 words included common concepts that could be found in a home setting (e.g., *Schüssel* 'bowl', *Messer* 'knife'). The words were recorded by a highly-proficient Chinese-German female speaker in a sound-proof room. We asked a control group of 21 L1 Chinese speakers with a similar L2 English proficiency level as the participants to assess whether the German words sounded like any words they knew in Chinese or English. Their judgements on a 5-point scale (1 = very dissimilar, 5 = very similar) showed that all German words were considered dissimilar to Chinese or English. The rationale for choosing German as the language of the to-be-learned word is that all participants reported no prior knowledge or experience with German. Moreover, previous work examining Chinese–English bilinguals has also selected German as the target language (Jiao et al., 2021, 2022).

Stimuli for the PW condition consisted of line-drawings from Snodgrass and Vanderwart's (1980) standardized picture database (Zhang & Yang, 2003) and stimuli for the VR condition included a fully immersive environment and colored three-dimensional (3D) objects selected from a standardized database (Peeters, 2018). The immersive VR condition simulated an apartment consisting of a living room, bedroom, and kitchen. This was presented and edited using the software Unity (https://unity.com). To identify which 3D objects to include as experimental materials, based on the standardized 3D object database (Peeters, 2018), we first recruited a group of 108 L1 Chinese speakers from the same population but who did not participate in the formal experience, to assess each 3D object's image-name agreement, familiarity with the image, and visual complexity on a 5-point scale. We chose the target objects considering these ratings and the appropriateness of their fit in the virtual environment of the present study (i.e., an apartment).

### 2.3 Procedure and measures

Participants completed the experimental procedure over four days. Days 1–3 involved learning sessions and Day 4 was the testing session on which a recognition task was administered. Also on Day 1, participants completed a language background questionnaire and a modified Flanker task.

#### 2.3.1 Learning sessions

During the learning sessions, participants were instructed to learn a set of 20 German words in a VR condition equipped with headgear and handsets and an additional 20 German words in a PW condition on a desktop computer. On each of the three days, participants performed one learning session in the VR condition for 15 minutes and one session in the PW condition for 15 minutes. Participants were given a brief break between the two learning conditions to prevent fatigue. The order of learning conditions was counterbalanced across participants (i.e., half of the participants performed the PW condition first and the other half performed the VR condition first). Moreover, the sets of words presented in the two conditions were counterbalanced across participants (i.e., half of the participants learned one set of words in the PW condition, while the other half learned that same set in the VR condition).

In the PW learning condition, individual 2D line drawings were displayed in the center of a computer screen and were accompanied by their spoken name in German. These recordings were played through headphones. Participants then were required to press the response key to move to the next word. During the 15 minutes, participants could repeat through the PWs as many times as they desired. Before starting the PW learning session, participants were administered a practice block of 5 trials in Chinese so that they could familiarize themselves with the experimental procedure.

In the immersive VR learning condition, HTC VIVE headgear and handsets delivered high quality visual fidelity and an engaging experience. Prior to learning the German words, the participants were shown how to use and interact with the VR equipment. During this practice, participants saw the same virtual apartment setting as in the experiment, but when selecting objects, the words were played in Chinese. After they were familiar with the equipment, the 15-minute learning session began in which they physically moved throughout the virtual apartment and used the handset to laser point to 3D objects. Upon selecting the objects, they heard the corresponding German words through the headphones.

After the flexible timetable for learning, all participants were required to complete the word recall task as practice on each day. The recall task was cued by the pictures of target words and participants were asked to name them in German. Their responses were recorded and accuracy was assessed by two experimenters. The data of two participants failed to record due to equipment malfunction. The results of 28 participants showed a gradual increase in accuracy across the three learning sessions (Day 1: PW = 41%, VR = 43%; Day 2: PW = 71%, VR = 75%; Day 3: PW = 86%, VR = 90%).

### 2.3.2 Testing session

On Day 4, the participants were asked to perform a recognition task in which they were asked to determine whether trials of auditorily presented German words had been among those including in the learning sessions. During the task, participants' behavioral performance and electrophysiological activity were measured. The task consisted of three types of trials (i.e., PW-learned words, VR-learned words, and not-learned words). Not-learned words were real German words recorded by the same human voice, but did not appear in the learning sessions. The entire task consisted of 2 blocks with each block consisting of 60 trials presented in a random order. Each participant performed 40 PW-learned trials, 40 VR-learned trials, and 40 not-learned trials.

Trials began with a fixation cross which was presented at the center of a computer screen for 400 ms. After a blank interval of 200 ms, a target word was played through headphones and participants were required to identify whether it had been learned or not by pressing the left or right response keys, respectively. The response keys were counterbalanced across participants. Once a response was given or after a maximum duration of 8000 ms, a blank screen was presented for 1000 ms prior to the next trial. Before the formal experiment, the participants were presented with 10 practice trials to become familiar with the procedure.

### 2.3.3 Cognitive task

Inhibitory control was measured by a modified Flanker task (Fan et al., 2002; Legault et al., 2019a). In the task, arrows or lines were presented on the screen and participants were instructed to respond to the direction of the center arrow by pressing left or right response key. The task consisted of three blocks, with 48 trials in each block presented in random order. There were three types of trials which were equally presented throughout the task: congruent trials, incongruent trials, and neutral trials. In congruent trials, the flanker arrows pointed towards the same direction as the central target arrow. In incongruent trials, the four flanker arrows pointed towards the opposite direction as the center arrow. In neutral trials, the center arrow was surrounded by lines without direction information. In our analyses, we used the flanker effect (i.e., the performance difference

between incongruent and congruent trials) to index inhibitory control.

### 2.4 Electrophysiological recordings and preprocessing

Electrophysiological data were recorded using 64 Ag/AgCl electrodes placed according to the extended 10–20 positioning system and were online referenced to FCz electrode. All channels were amplified with a band pass of .05–100 Hz and a sampling rate of 1000 Hz. Electrode impedance was kept below 5 kΩ. EEG preprocessing and analyses were conducted using the EEGLAB toolbox. The signal was band-pass filtered at a 1–40 Hz and re-referenced offline to the averaged left and right mastoids (TP9 and TP10). The signal containing eye blinks and other artefacts were corrected for each subject by independent component analysis (ICA). Epochs of 200 ms before to 800 ms after the onset of target word were extracted. Baseline correction was performed in reference to pre-stimulus activity (Liu et al., 2022).

### 2.5 Data analyses

Both behavioral and ERP data from the recognition task were analyzed with linear mixed-effect models in *R* using the *lme4* package (Bates et al., 2014). Each model included a three-level variable of learning type as a fixed effect, with subject and item as random effects. The learning type variable (PW-learned, VR-learned, and not-learned) was coded with orthogonal contrasts in which the first contrast compared not-learned words to learned words (i.e., the average of PW- and VR-learned words) and the second contrast compared PW-learned words to VR-learned words, consistent with our research objectives. We started with a full model including the fixed effect, random intercepts for subjects and items, and random slopes for all predictors (Barr et al., 2013). When the full models failed to converge, we followed a backward-fitting procedure to identify a model that would converge.

Response times (RTs) were measured from the onset of target words. Trials with error responses (2.94%), RTs higher than 3000 ms (3.89%), and RTs more than 2.5 standard deviations (SD) from the mean (3.14%) were removed from the analyses. For the ERP data, we examined the mean amplitude of waveforms across the selected time-windows of N100 (100–150 ms), N200 (250–350 ms), and N400 (400–600 ms) from the recognition task. Based on previous studies examining foreign language learning (Biau et al., 2018; Kutas & Federmeier, 2011), the N100 and N200 components were analyzed at frontocentral electrode sites (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2), and the N400 component was analyzed at central-parietal sites (C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2). Finally, in order to reveal the effect of individual differences, we calculated the correlation between the measures from the recognition task (RT and ERP indicators) and the individual differences indicators (flanker effect and L2 AoA).

## 3. Results

### 3.1 Behavioral results

Figure 1 presents the mean accuracy (ACC, left) and RTs (right) from the recognition task. For ACC, we fit a logistic mixed-effects model with learning type as a fixed effect and by-subject and by-item intercepts as random effects. The variable of learning type was the orthogonal contrast. The results showed that the contrast between not-learned words (M = 95%, SD = 22) and all
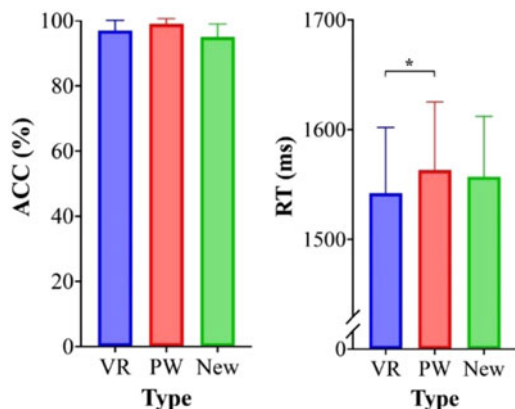
**Figure 1.** Bar plot for ACC (left) and RTs (right) in the recognition task across learning type (VR = virtual reality, PW = picture-word, New = not-learned).

### 3.2 ERPs Results

Figure 2 shows the grand average ERP waveforms elicited during the recognition task. The mixed-effect model for N100 amplitude included the fixed effect of learning type, and the random effects of by-subject and the by-item intercepts. As in the analyses on behavioral data, the three-level variable of learning type was the orthogonal contrast. The results of N100 amplitude showed that the VR-learned words elicited more negative waveforms than PW-learned words, Estimate $= -.94$, SE $= .36$, $t = -2.58$, $p = .01$. But the comparison between not-learned words and all learned words was not significant, Estimate $= -.19$, SE $= .35$, $t = -.53$, $p = .60$. For N200 amplitude, the model included learning type as a fixed effect with by-subject and by-item intercepts as random effects. The results of N200 amplitude showed that the not-learned words elicited significantly more negative waveforms than learned words, Estimate $= .74$, SE $= .34$, $t = 2.16$, $p = .03$. Moreover, the comparison between the two learning contexts showed more negative waveforms for VR-learned words than PW-learned words, Estimate $= -.74$, SE $= .37$, $t = -2.01$, $p = .04$. However, the model for the N400 component, with the same fixed effect and random effects as in the N200 analyses, revealed no significant differences between all words learned and not-learned words, Estimate $= .27$, SE $= .28$, $t = .97$, $p = .33$, nor between PW and VR, Estimate $= -.13$, SE $= .30$, $t = -.42$, $p = .67$. Overall, the differences between PW- and VR-learned words began at the N100 time window, while significant differences between not-learned and learned words emerged at the N200 time window.

learned words – namely the average ACC of VR-learned (M = 97%, SD = 17) and PW-learned words (M = 99%, SD = 9) – was not significant, Estimate $= .13$, SE $= .43$, $z = .30$, $p = .76$. Moreover, the difference between VR-learned words and PW-learned words was not significant, Estimate $= -.43$, SE $= .42$, $z = -1.03$, $p = .30$.

For the analyses on RTs, the linear mixed-effects model included learning type as a fixed effect and included the by-subject random intercept and the by-item random slope for learning type as random effects. The results showed that there was no significant difference between not-learned words (M = 1557 ms, SD = 302) and all learned words, Estimate = 4.64, SE = 19.75, $t = .23$, $p = .81$. However, the comparison between VR-learned and PW-learned words reached significance, with faster responses for VR-learned words (M = 1542 ms, SD = 329) compared to PW-learned words (M = 1563 ms, SD = 341), Estimate $= -29.63$, SE $= 14.14$, $t = -2.09$, $p = .04$. Further, considering not-learned words as a baseline, the results showed that neither the comparison between not-learned and PW-learned words (Estimate = 19.45, SE = 20.10, $t = .97$, $p = .34$), nor the comparison between not-learned and VR-learned words (Estimate = -10.18, SE = 21.82, $t = -.47$, $p = .64$) reached significance. Overall, in behavioral performance, participants reacted to VR-learned words more quickly as compared to PW-learned words while there was no difference in accuracy.

### 3.3 Correlation results

To examine whether the effects of an immersive VR environment on novel word learning was related to individual differences, we conducted Pearson correlation analyses between performance on the recognition task and participants' inhibitory control (as indexed by the flanker effect) and L2 AoA. We ran the analyses for both behavioral (RTs) and neurophysiological data (N100 and N200 components, but not N400 amplitude as there was no effect of learning context on this component). Results showed no significant effects on RTs (flanker effect: $r = .11$, $p = .57$; L2 AoA: $r = .03$, $p = .86$) nor on the N100 component (flanker effect: $r = -.07$, $p = .72$; L2 AoA: $r = -.10$, $p = .60$). For N200 amplitude
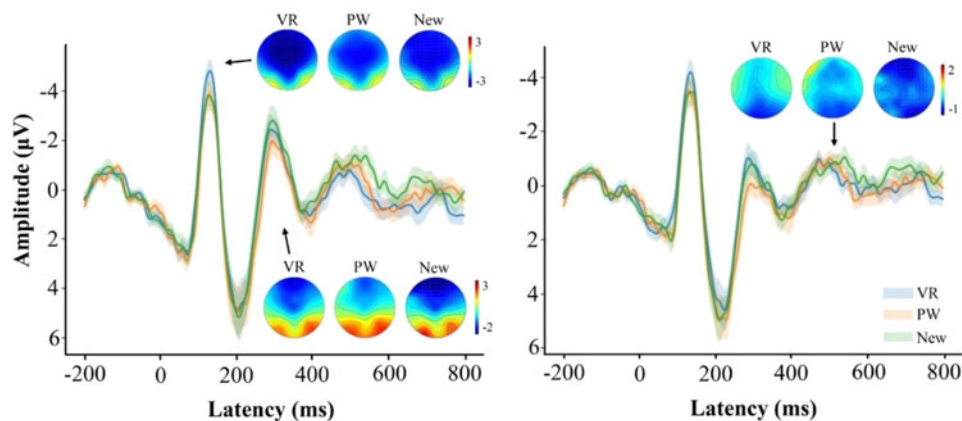


**Figure 2.** Grand average waveforms and topographic maps of recognition task across learning type (VR = virtual reality, PW = picture-word, New = not-learned). Left panel presents N100 and N200 components averaged frontocentral sites (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2); right panel presents N400 component average central-parietal sites (C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2).

in the VR condition, the correlation between N200 and the flanker effect was not significant ($r = -.33$, $p = .07$) but a significant correlation was observed with L2 AoA ($r = -.46$, $p = .01$). These effects can be seen in Figure 3.

## 4. Discussion

Combining behavioral performance and electrophysiological activity, the present study examined how an immersive VR context influences novel word learning and compares it with a picture-word (PW) association context. Recent research suggests that rich sensory experiences (e.g., video, VR technology) offer benefits for the lexical meaning access of novel words – however, less attention has been given to early lexical form acquisition (Legault et al., 2019a, 2019b; Li & Jeong, 2020; Makransky & Petersen, 2021). Using the high-temporal resolution offered by EEG technology, we compared novel word learning in an immersive VR context and a PW association context among a group of Chinese–English speakers. Overall, our findings on a recognition task showed a positive effect of the rich sensory experience of an immersive VR context on novel word learning. We will elaborate on these findings in the next subsections.

One main finding of the present study revealed how a rich sensory experience of a VR context affects lexical form acquisition of novel words by comparing it with a unisensory PW context. The behavioral performance demonstrated that responses to words learned in the immersive VR context were faster than to words learned through PW context. This finding is consistent with previous studies demonstrating that enriched perceptual and sensorimotor experiences benefit L2 word learning (Ibáñez et al., 2011; Johnson-Glenberg et al., 2014; Legault et al., 2019a, 2019b). For example, Legault et al. (2019a) asked English speakers to learn Mandarin Chinese words through immersive VR and word-word association. Semantic access of novel words was examined by an alternative forced-choice recognition task and results revealed a beneficial effect of the VR learning context compared to the word-word association context. Different from the study of Legault et al. (2019a), the task in the present study was to recognize the pronunciation of words and only judge whether they had been learned or
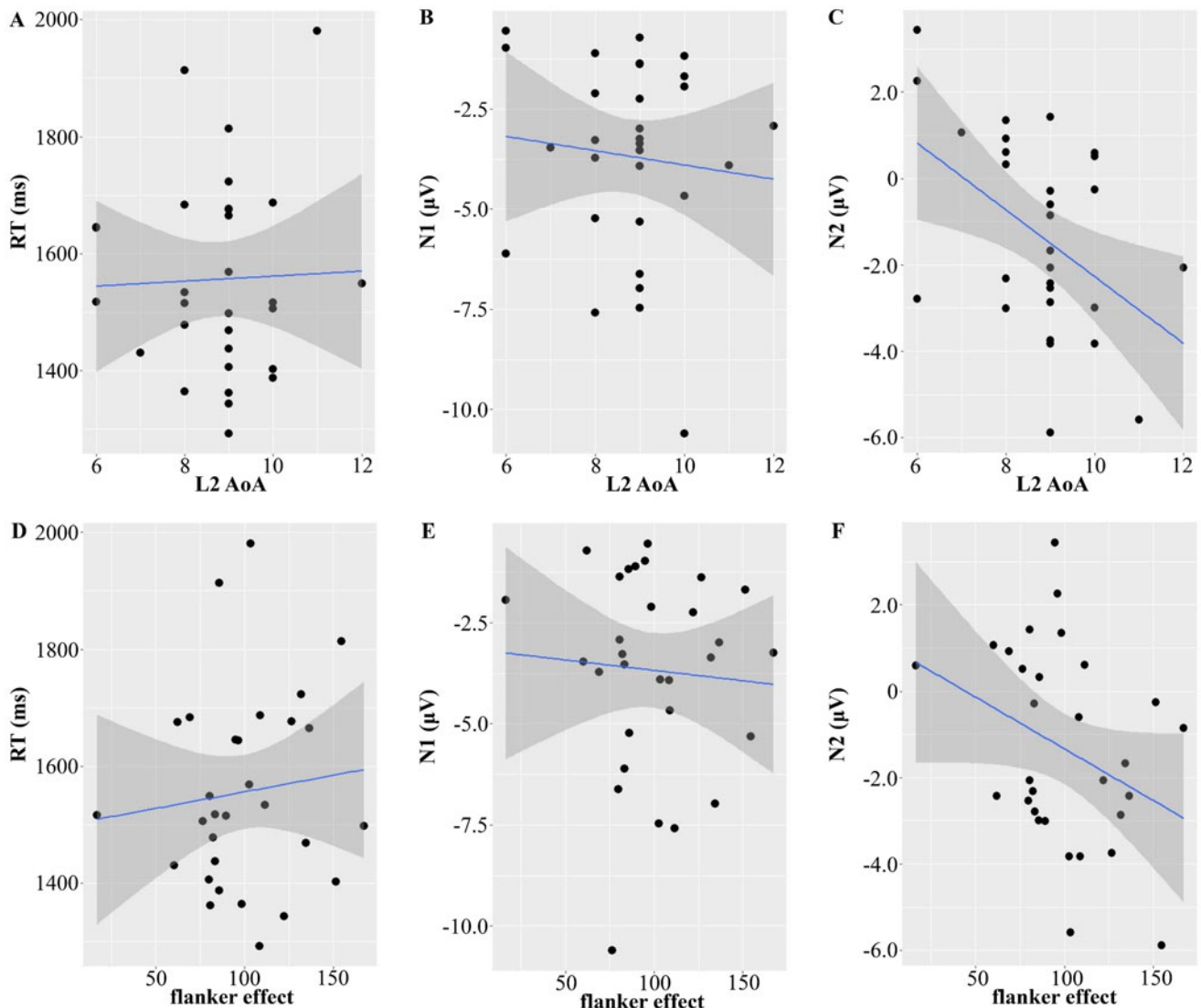


**Figure 3.** RTs, N1, and N2 correlations between inhibitory control/L2 AoA and novel word learning in the VR learning context.

not. This allowed us to examine the lexical form acquisition, rather than semantic acquisition of novel words.

The comparison between VR and PW in the electrophysiological data showed enhanced N100 and N200 amplitudes elicited by the VR context, but not in N400 amplitude. Compelling evidence for the effect of the VR context on novel words is the N100 component showing that the amplitude on words learned via immersive VR was significantly more negative compared to words learned in PW context. Given that the N100 component has been linked to early acoustic processing (Biau et al., 2018; Vogel & Luck, 2000), our results suggest that the VR-learned and PW-learned words are different in the early stage of phonetic processing and recognition. The results of the N200 component, which reflects early lexical selection and phonological processing (Connolly & Phillips, 1994; van den Brink et al., 2001), revealed that the word-form recognition during initial processing was distinct between VR-learned and PW-learned words. For words learned in the VR context, their sound information could link both to the corresponding 3D object though visual perception and to the surrounding environment through physical interaction. Whereas for words learned in the PW context, their sound information only linked to visual perception (i.e., corresponding picture). Hence, these results suggest that the rich sensory experience afforded by VR affects lexical form acquisition of novel words. This hypothesis is particularly supported by the differences in the N100 effect between VR and PW contexts.

In addition, given that word-form recognition processes can only occur in learned words, it is less likely that the pronounced N200 changes we observed in the not-learned words reflects phonological processes. Therefore, we speculate that the enhanced N200 effect elicited by not-learned words originated from cognitive rather than phonological processes. Previous studies have shown that N200 effects are sensitive to conflict monitoring and attention control (Mathalon et al., 2003; van Veen & Carter, 2002). Drawing on this background, the N200 changes we found in not-learned words might imply that this condition placed differential demands on the monitoring/controlling for conflicting information compared to words learned. The involvement of the N200 component during lexical form acquisition and discrimination is worthy of a more comprehensive investigation.

The theoretical support for the effect of VR context is closely related to embodied cognition theory (Barsalou, 2008; Li & Jeong, 2020). This theory argues that the interaction between a learner's body and their learning environment plays a key role in defining their experience and the extent to which knowledge is acquired (Barsalou, 2008). To some extent, through traditional picture-word/word-word paired association, learners rely on rote memorization in which they are presented novel words alongside their L1 translation equivalents. Contrarily, immersive learning context supported by VR technology provides a multiple sensory experience which allows learners to move around in and interact with a contextualized environment, thus leading to superior learning outcomes. Overall, the rich sensory experience involved in VR contexts enables learners to better connect novel words and their lexical forms, pronunciation, and conceptual representations, and thus, improves their recognition (Malt et al., 2015; Zinszer et al., 2014).

In our analyses, we also explored potential modulating effects of L2 AoA. We observed a relationship between L2 AoA and N200 amplitude, but not with inhibitory control. Based on the transfer hypothesis and our objectives, we selected L2 AoA as an indicator of prior L2 learning experience because it reflects the amount of language learning exposure in traditional classroom contexts. To some extent, an earlier L2 AoA among participants in the present study represents an accumulation of language learning experiences in classroom contexts (i.e., non-immersive environments). We speculate that such accumulated experience may benefit novel word learning through similar learning methods (e.g., picture-word/word-word paired association), but not through immersive contexts. This finding was demonstrated in the negative relationship between L2 AoA and N200 amplitude. Given that there are few EEG studies examining word learning in VR contexts, we approached the potential effect of L2 AoA from a rather explorative perspective. A limitation of the present study is that we only examined one individual difference (i.e., L2 AoA). Future studies should consider other aspects of prior language experience, such as language proficiency, language dominance, and other linguistic skills.

We acknowledge that there were some unexpected findings in the present study. First, we failed to observe a pronounced difference between not-learned and novel-learned words in the behavioral performance: as novel word learning is a gradual process, the link between novel words and the lexical network is not fully established (Liu & van Hell, 2020); thus, the recognition of novel-learned words from not-learned words is likely not automatic, and fails to be detected by the less sensitive measure of RTs. Second, the gender makeup of our sample with the female majority. Although some previous studies have also included a similar distribution in their samples, there is no direct evidence revealing whether gender may affect learning outcomes in a VR context. Therefore, future studies focusing on immersive VR learning should also consider the potential role of demographic factors, e.g., gender, age, etc.

## 5. Conclusion

In the present study, Chinese–English bilinguals learned novel words in German, a language with which they had no prior experience, in an immersive VR context and compared learning outcomes with a PW context. The results of a recognition task revealed faster responses and enhanced N100 and N200 for words learned in the VR context compared to words learned in the PW context, suggesting that learning context affects early lexical form acquisition. Moreover, we found that L2 AoA was related to N200 amplitude of VR-learned words. The findings of the present study provide the first evidence that an immersive VR context with a rich sensory experience can have facilitative effects on early lexical form acquisition. Our study also provides neural evidence for embodied cognition theory by considering immersive learning. It will be highly beneficial for our understanding of novel word learning if future studies continue to systematically investigate the effects of multisensory learning environments.

## References

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.

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

Basirat, A., Brunellière, A., & Hartsuiker, R. (2018). The role of audiovisual speech in the early stages of lexical processing as revealed by the ERP word repetition effect. *Language Learning*, *68*, 80–101.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7), 1–23.

Biau, E., Fromont, L. A., & Soto-Faraco, S. (2018). Beat gestures and syntactic parsing: An ERP study. *Language Learning*, *68*, 102–126.

Bogulski, C. A., Bice, K., & Kroll, J. F. (2019). Bilingualism as a desirable difficulty: Advantages in word learning depend on regulation of the dominant language. *Bilingualism: Language and Cognition*, *22*(5), 1052–1067.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347.

Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *NeuroReport: For Rapid Communication of Neuroscience Research*, *19*(18), 1757–1761. https://doi.org/10.1097/WNR.0b013e328318f014

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), 672–683.

Hirosh, Z., & Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin & Review*, *25*(3), 892–916.

Ibáñez, M. B., García, J. J., Galán, S., Maroto, D., Morillo, D., & Kloos, C. D. (2011). Design and implementation of a 3D multi-user virtual world for language learning. *Journal of Educational Technology & Society*, *14*(4), 2–10.

Jackson, J., & Schwieter, J. W. (2019). Study abroad and immersion. In J. W Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 727–750). Cambridge University Press.

Jeong, H., Sugiura, M., Sassa, Y., Wakusawa, K., Horie, K., Sato, S., & Kawashima, R. (2010). Learning second language vocabulary: Neural dissociation of situation-based learning and text-based learning. *Neuroimage*, *50*(2), 802–809.

Jeong, H., Li, P., Suzuki, W., Sugiura, M., & Kawashima, R. (2021). Neural mechanisms of language learning from social contexts. *Brain and Language*, *212*, 104874.

Jiao, L., Zhang, Y., Plummer, P., Liu, C., & Chen, B. (2019). The influence of bilingual language experience on executive control: An ERPs study. *Journal of Neurolinguistics*, *51*, 42–52.

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., Duan, X., Liu, C., & Chen, B. (2022). Comprehension-based language switching between newly learned languages: The role of individual differences. *Journal of Neurolinguistics*, *61*, 101036.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology*, *106*(1), 86–104.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Kapa, L. L., & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language*, *76*, 237–252.

Klassen, G., Ferreira, A., & Schwieter, J. W. (2021). The role of study abroad in the acquisition and processing of L2 gender agreement. *Applied Linguistics Review* [Ahead-of-Print], 1–23.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101.

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

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 and Development*, *63*(5), 671–690.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26.

Legault, J., Fang, S. Y., Lan, Y. J., & Li, P. (2019a). 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. (2019b). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages*, *4*(1), 1–32.

Li, P., & Jeong, H. (2020). The social brain of language: Grounding second language learning in social interaction. *npj Science of Learning*, *5*(1), 1–9.

Li, P., & Lana, Y. J. (2021). Digital language learning (DLL): Insights from behavior, cognition, and the brain. *Bilingualism: Language and Cognition*, 1–18.

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, 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.

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.

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 Psychology Review*, *33*(3), 937–958.

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, *113*(4), 719–735.

Malt, B. C., Li, P., Pavlenko, A., Zhu, H., & Ameel, E. (2015). Bidirectional lexical interaction in late immersed Mandarin-English bilinguals. *Journal of Memory and Language*, *82*, 86–104.

Mathalon, D. H., Whitfield, S. L., & Ford, J. M. (2003). Anatomy of an error: ERP and fMRI. *Biological Psychology*, *64*(1–2), 119–141.

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology*, *25*(4), 530–535.

Nair, K. K., Biedermann, B., & Nickels, L. (2016). Consequences of late bilingualism for novel word learning: Evidence from Tamil–English bilingual speakers. *International Journal of Bilingualism*, *20*, 473–487. https://doi.org/10.1177/1367006914567005

Papin, K., & Kaplan-Rakowski, R. (2022). A study of vocabulary learning using annotated 360° pictures. *Computer Assisted Language Learning*, 1–28.

Peeters, D. (2018). A standardized set of 3-D objects for virtual reality research and applications. *Behavior Research Methods*, *50*(3), 1047–1054.

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 and Memory*, *6*(2), 174–215.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–985.

van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, *77*(4–5), 477–482.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203.

von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS biology*, *4*(10), e326. https://doi.org/10.1371/journal.pbio.0040326

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

Zinszer, B. D., Malt, B. C., Ameel, E., & Li, P. (2014). Native-likeness in second language lexical categorization reflects individual language history and linguistic community norms. *Frontiers in Psychology*, *5*, 1203.