Evidence for embodied cognition in immersive virtual environments using a second language learning environment

Jack Ratcliffe and Laurissa Tokarchuk
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, UK
{j.ratcliffe@qmul.ac.uk, laurissa.tokarchuk}@qmul.ac.uk

***Abstract*—Immersive virtual environments (IVEs) are increasingly being explored as potential educational tools. However, it is unclear which aspects of IVEs contribute to learning, including hardware modalities and learner responses (e.g. motivation, usability, cognitive load and presence). One IVE hardware modality particularly backed by theory is embodied controls, with their potential for leveraging embodied cognition for enhanced learning outcomes. This paper explores if embodied controls can be leveraged to enhance learning in an IVE by comparing language learning outcomes from an IVE using embodied controls, and a non-embodied control. It explores two words classes - verbs and nouns - to examine if there is a difference in learning outcome for embodied controls with actions (verbs) and object interactions (nouns). This paper also explores co-variables often linked with IVE learning (motivation, presence, cognitive load) to understand why learning gain occurs. It finds that leveraging embodied controls provides better learning outcomes, with no impact on cognitive load. It also finds that the benefit does not correlate with motivation or presence ratings, suggesting that embodiment-induced motivation or immersion is not the cause of the learning enhancements, and therefore this could be evidence for embodied cognition-based learning in IVEs.**

***Index Terms*—Virtual reality, Educational technology, Computer aided instruction**

I. Introduction

Technological advances have seen an increase in the use of immersive virtual environments (IVEs) for educational purposes. A majority of research on these IVEs has focused on IVE design, or comparisons between IVE learning system efficacy and a real-world alternative [1]. This has left a large gap in literature: examinations of which aspects of IVEs contribute to learning; the extent that combinations of these contribute to or detract from learning outcomes; and what causes them to have an impact.

There are a myriad of potential contributory variables to explore. There is already evidence that head-mounted displays play a large role in IVE learning [1]. A less-studied aspect, which has strong theoretical and experimental support for creating positive learning outcomes, is embodied controls. The benefits of embodied controls for learning have been well-explored from an embodied cognition perspective [2]. Experiment results show positive learning benefits related to using iconic gestures with technology [3] and without [4]. The impact is particularly well-defined inside language learning [5] [6], but also exists within wider learning applications [7]. The question is broader than just if embodied controls aid learning in an IVE, however. It is also important to understand how this aspect interacts with other IVE-relevant modalities, as there is evidence that multi-modal learning in an IVE can cause cognitive load issues that harm information retention [8] [9] [10].

Applied Linguistic research informs us that another IVE-input method, spoken input, can have a positive impact on language memorisation. Actively speaking foreign language words while learning causes increased word retention. This is known as the production effect [11], and plays an important role in modern second-language tuition. The positive impact of the production effect on computer-aided language learning (CALL) has been demonstrated experimentally, with automatic-speech recognition systems being some of the most effective types of computer-aided language learning tools [12]. Recent experiments combining gesture and spoken production have also demonstrated positive learning outcomes [13] [11]. However, these experiments did not take place inside IVEs, and whether and how these benefits transfer is yet to be investigated.

In this paper, we investigate the relationship between embodied controls and memorisation in an IVE, as part of a multi-modal interaction environment that also leverages spoken input. We also attempt to identify the cognitive reasons for any difference. To do this, we created an IVE for memorising foreign language words, and split participants between two interaction conditions: embodied controls-and-spoken production; and spoken-production-only. We calculated learning gain from pre- and post-exposure learning outcomes, and examined the learning outcomes separately for verbs, representing embodied actions, and nouns, representing embodied

With thanks to the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1), who funded this research.

978-1-7281-4533-4/20/$31.00 ©2020 Crown

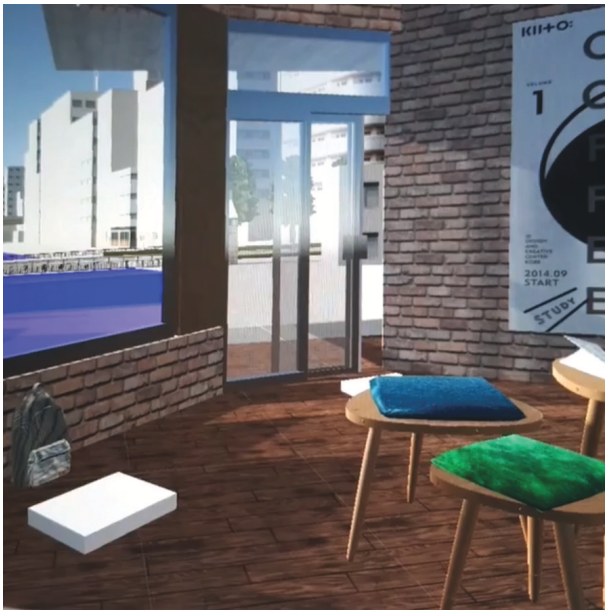

Fig. 1. Image of the virtual learning environment

object interactions. To understand what cognitive processes the embodiment may have affected, we monitored the following co-variables: self-reported cognitive load, system usability, motivation and presence.

We found that there was significantly higher memorisation results for those in the embodied control-and-spoken group over those in the control, particularly for embodied actions (verbs) rather than the embodied object interactions (nouns). We also did not find a relationship between embodied controls and cognitive load, system usability, motivation nor presence. This suggests that the memorisation benefits stem from embodiment itself, rather than the embodied controls acting as a mediating influence on other learning-influencing variables. This could be used as evidence in support of both embodied cognition, and leveraging this approach in learning environments inside IVEs.

II. Literature

A. Embodied cognition, iconic gesture and language acquisition

From the embodied cognition perspective, cognitive processes are rooted in the body's interactions with the world [14]. Language is generally considered from this perspective: there is a strong relationship between language and embodied actions or gestures [15]. Of the multiple gesture types, iconic gestures have been considered "fundamental to all languages ... [bridging] the gap between linguistic form and human experience" [16]. Areas of the brain responsible for iconic gestures and physical actions have been shown to activate when associated words are used or heard [5], [17]. For language acquisition, iconic gestures are considered universally important for both first and second language acquisition [18], [19], and have been considered an additional "mode of thinking" [20] for second language learners.

Of Wilson's six claims of embodied cognition [2], the claim that off-line cognition is body-based - "when decoupled from the environment, the activity of the mind is grounded in the mechanisms that evolved for interaction with the environment" - is particularly relevant to the relationship between embodied action gestures and language acquisition. The use of off-line embodiment was operationalised by applied linguists for second language acquisition some decades before embodied cognition theorists began to coalesce around the theory, in the form of the Total Physical Response [21] teaching approach. Asher found that learners of Japanese performed significantly better at recognising spoken words if they performed an action related to the word while learning.

There have been attempts to explain the benefits afforded by embodied actions and gestures for language acquisition outside of embodied cognition. Asher noted that the learning benefits of his approach could be explained through increased learner motivation, while later studies found any light to moderate physical activity during encoding - such as performing actions - is beneficial to vocabulary acquisition and retention [22]. However, there is strong evidence that the positive relationship between iconic gestures and acquisition is not entirely mediated by physical activity or higher motivated. Experiments have shown that iconic gestures relevant to the words being encoded (e.g. jumping while learning the word for "jump"), rather than unrelated gestures (e.g. jumping to learn the word "kick"), have significant retention benefits [23], [5]. If the learning benefits were solely caused by the enhanced motivation provided by learning with physical activity, or merely the effect of the physical activity itself, it would be difficult to explain why the use of related gestures was superior to unrelated ones.

Further evidence for the unique encoding potential of iconic gestures for language learning is found in Macedonia's work [5], which showed that word acquisition related to iconic gestures activated different parts of the brain than word learning with unrelated gestures. The former activates areas associated with the pre-motor cortices that control bodily movement, while the latter activates areas associated with cognitive control.

Whatever the reason for the benefits of using embodied actions or iconic gesture as a tool for language memorisation, experimental results in embodied controls and computer-aided language learning have proved positive: Vasquez [3] used iconic gestures to help with listening skills related to verbs that correspond to the gesture enacted by the learner; Edge [24] had users enact a sequence of movements to complete a foreign-language movement instruction; Macedonia [6] had participants imitate a pedagogical agent's gestures and visually learn words accompanied by gestures; and Repetto [25] found that when recognizing novel words, participants made less errors for words encoded with gestures compared to words encoded with pictures.

Outside of language acquisition, attempts to leverage learn-

ing benefits via immersive embodied controls have seen mixed results. Howard's comprehensive meta-analysis of immersive technologies was unable to show that embodied action-enabling hardware have a meaningful impact on cognitive learning [1]. Howard gives three explanations for this lack of result: (1) Current hardware influences important learning mechanisms, but is underdeveloped at the moment. (2) Hardware influences mechanisms that have little effect on learning. For example, it may influence presence, but presence may have little influence on learning. (3) Specialised hardware may not influence tasks processes enough to have a notable effect over non-specialised (e.g. keyboard and mouse) hardware.

For (1), technological development could make an important difference for any educational subject. For (2), we can investigate the mechanisms that embodied hardware is considered to affect (such as presence), and for (3) it makes sense to examine this on a task-specific basis, because as demonstrated in the distinct learning outcomes for related and non-related gestures, not all types of embodiment are equal. Howard describes this as "specialised input hardware are required to develop certain personal characteristics, such as physical abilities and particular skills". It is quite possible that language learning should be considered as such: if not a "physical ability", than an importantly embodied one.

B. Speaking: the production effect and encoding modalities

Research has demonstrated that speaking a word provides a significant memorisation benefit over reading it silently or listening to it, known as the production effect [11]. It is currently unclear if the production effect produces results for reasons similar to those mentioned under embodied cognition, although research has demonstrated that when non-verbally recalling memorised words, areas of the brain used for speech production are activated [2].

Possible non-embodied explanations for the production effect are that speaking while learning a language typically involves a distance between encoding domains (i.e. reading and then speaking out loud), and memorisation efficacy increases when an input activity requires a translation between processing domains, such as reading to speaking [26]. Or it could simply be that the more modalities used in learning, the better the encoding [6]. Whatever the cause, spoken production has been part of the SLA pedagogical cannon for decades and it is uncontroversial to suggest it plays a large role in helping second language acquisition.

C. Embodied actions + Spoken production = Success?

There is evidence that combining embodied actions, like gesture, with spoken production causes enhanced language memorisation. Gesture and spoken production work together to enhance communication, forming an "an integrated system in language comprehension" [27] with demonstrable benefits in word understanding when gesture and speech are congruent. Growth Point Theory [28] hypothesizes that speech and gesture interact and influence one another throughout the planning and speaking of utterances, with gestures helping speakers to "internalise the abstract via the concrete".

Experimentally, Kelly demonstrated positive Japanese memorisation outcomes by having learners combine gesture with simultaneous, relevant spoken production [13]. Later, Bergmann and Macedonia achieved the same but with sentence learning, rather than singular words [29]. Both of these studies showed that when a learner used gesture with spoken production they achieved better learner outcomes than spoken production alone. Interestingly, these contradict the original Total Physical Response findings, which demonstrated that students' success when attempting to learn both listening and speaking together was significantly decreased [21]. It remains to be discovered how these aspects relate in an immersive virtual environment.

D. Immersive virtual environments

Immersive virtual environments (IVEs) are increasingly explored as a potential tool for learning. They are considered particularly powerful from an embodied cognition perspective because of their ability to provide a situated learning context and provide an environment for offloading cognition onto - both parts of Wilson's embodied cognition definition [2]. The results for embodied controls have been mixed [1], however. Howard's meta-analysis found that the dominant contributing factors to learning are head-mounted displays and interactive environments, with embodied input technologies providing little impact on learning outcomes (although it is possible to argue that the ability to look around a 3D space in a head-mounted display is a type of embodied action). He also noted that many of the benefits of immersive learning could stem from motivational benefits, rather than anything emergent from the interaction itself. Also, while many studies in the field demonstrate significant learning outcomes, few demonstrate learning benefits that outweigh traditional learning methods. Howard's analysis covers a large range of immersive learning studies, however, covering a spectrum of topics, system designs and interactive technologies. His results may be useful for general comments on IVEs and learning, but it remains to be seen if they are applicable to language acquisition, for which the benefits of IVEs have a pedagogical grounding in second language acquisition theories. These groundings include physical language teaching approaches [21], the strong encoding relationship between gestures and verbs [5], and the use of real-world immersion, in which a student travels to and learns inside of a real second language context, such as study abroad or immersion events [30].

Where found, lower learning rates in immersive environments have sometimes been attributed to issues with cognitive load [8] [9] [10], with claims that virtual immersion creates a large cognitive load that detracts from a learner's ability to memorise information. As embodied controls are considered to increase immersion, then according to the above, they should also increase cognitive load. However, Steed et al. found that the use of embodied controls in IVEs actually reduced cognitive load [31]. This suggests that immersion stemming from embodied controls could be different to other types of

Embodied controls + spoken

Spoken

Fig. 2. Showing hand interaction in embodied controls + spoken environment, and animated object mid-animation in spoken environment

immersion and have a different cognitive impact. If Steed is correct, we would see a reduced impact on cognitive load from this study.

III. Experiment

We created an experiment to understand if interacting in an IVE using contextually-accurate embodied controls and spoken production is more effective for language memorisation than spoken production alone. We choose language learning due to the theoretically strong links with embodied actions and learning outcome outlined above. By monitoring co-variables considered related to learning in IVEs, such as cognitive load, usability, motivation and presence, we were able to theorise whether the addition of embodied controls provides benefits due to an implicit advantage of embodied actions on memorisation (potentially stemming from embodied cognition), or whether embodied controls merely trigger a collinear factor from one of the other monitored aspects, each of which, as discussed above, have been linked with learning (cognitive load, motivation, presence, usability).

A. Hypothesis

- H1. Language memorisation occurs when using embodied controls + spoken production in an IVE
- H2. Leveraging embodied controls + spoken production while learning leads to better language learning than spoken production alone
- H3. Cognitive load, presence and motivation do not significantly vary between the two interaction types

B. Procedure

Each participant was assigned to either an embodied controls and spoken production group, or a spoken production-only group. They were then presented with 10 interaction areas inside a virtual coffee shop setting. Each interaction area contained an object and a related action. When a participant navigated to an interaction area, a voice-over introduced the object and explained the possible action in both English and Japanese (e.g. “This is a drink. Drink in Japanese is nomimono. You can pour it. Pour in Japanese is sosogu”).

Depending on their assigned group, when interacting with the object, the participant was asked to either:

- Speak the object and action words, and then complete an accompanying action/gesture by grabbing and moving the related item using their embodied controllers
- Speak the object and action words only, then watch the object complete a corresponding gesture animation

Participants were introduced to each interaction area in sequence, then given 10 minutes to freely explore the environment and attempt to memorise the words.

Each participant only experienced one of the above conditions (between-subject design). The system recognised correct actions and spoken input, and if they were successful, the interaction area ends and a participant may visit the other interaction points. Failed recognition re-prompted users until they correctly performed the spoken utterance or action. Users can also leave an interaction area at any point.

C. Participants

Twenty-four uncompensated participants (15 male, 7 female) were asked to self-report their knowledge of the target language (Japanese) and were pre-tested for their knowledge of the words used in the experiment. Participants were aged in ranges 30-39 (12), 20-29 (8) and 40-59 (4). No participants demonstrated an extensive knowledge of the target learning words during the pre-test (M = .13; SD = .44) nor self-rated their ability as anything above “basic phrases”. Most participants were fluent in more than one language, but we did not find a significant difference in learning outcome between mono-lingual and multi-lingual participants ($t(22)$ = -.84, p = .20; mono-lingual: M = 6.17, SD = 3.18 ; multi-lingual: M = 7.83, SD = 4.25). Twenty-two participants were educated to post-graduate level or above. A visual inspection suggested there was not enough variance in answers related to interest levels in Japanese, Japan, virtual reality and coffee shops to prove useful for further analysis.

D. Corpus

Participants were tested on their knowledge of 10 noun/verb pairs (20 words). Japanese gairaigo words (words imported from other languages, such as “koohii” to mean “coffee”) were specifically avoided to reduce the chance of participants inferring a meaning.

E. Environment

We created a 3D coffee shop environment in Unity to provide a situated context for memorising nouns and verbs related to a coffee shop. The environment was explorable via a head-mounted display and embodied controllers (the Oculus Rift S and Touch controllers). Navigation could be done by moving around the real space, using the analogue controllers, or a combination of both.

F. Evaluation

Participants' knowledge of the Japanese content was measured in three tests: one administered before their exposure to the environment (pre-test); one immediately after (post-test), and one seven days later (week-test). Participants performed the same test each time, listening to a Japanese word and typing the English (or another) language translation if they knew the meaning. The week-test was conducted via the internet and not in controlled conditions. Participants were not given feedback when submitting answers. The maximum score was 20, and a participant's existing knowledge (i.e. correct answers from the pre-test) was removed from the analysis to ensure only acquired knowledge was included in the results. Only three participants had any prior knowledge of the words, to very low levels. The pre-test results for correct answers (out of 20) for each group were M = .21, SD = .568 (embodied controls and spoken production) and M = 0, SD = 0 (spoken-production only).

After using the system, participants were asked to complete a MEEGA+ questionnaire [32] to provide insight on the system usability and motivation. Participants were also asked to self-report their cognitive load on a single-item, 9-point Likert scale as defined by Paas [33], and their level of presence while inside the environment on Slater's single-item, 6-point Likert scale [34]. Asking participants for their subjective evaluation of presence experienced is considered the most direct way of presence assessment [35].

G. Learning Style

Participants were asked to complete the VARK learning preference questionnaire [36] to allow us to determine if learning preference would have an impact on results. However, there was too much homogeneity in the results to allow for segmentation analysis related to different learning preferences and so this was excluded from the analysis.

H. Analysis

In order to answer our main research question, does combining embodied controls and spoken production aid learning over spoken production alone (H1 and H2), we used a one-tailed independent t-test on the learning gain from pre-test to post-test of the two groups. The data for each group was normally distributed, and met the requirements of homogeneity.

Previous studies have suggested a distinction between the impact of embodied controls on immediate post-exposure learning, and that of longer-term retention, with learning that leverages embodied action improving retention but not immediate learning [3]. We explored this factor using a mixed ANOVA to examine the learning gains of the post-test and week-tests, and the two conditions.

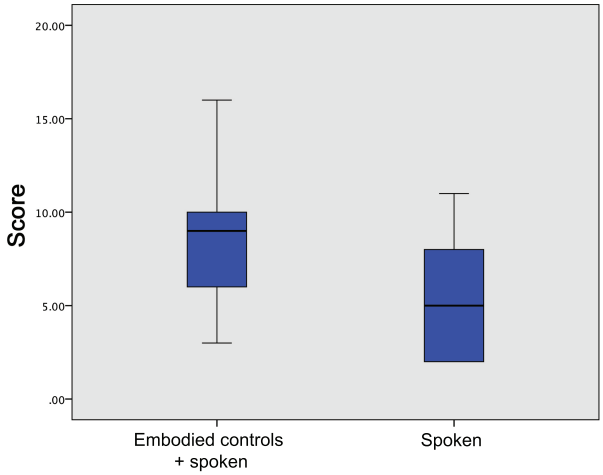

Fig. 3. Showing difference in embodied controls and spoken vs. spoken scores. Note one outlier (o9) in the embodied controls and spoken condition

To understand the impact of the learner's motivation, cognitive load and presence on the results (H3), as well as other factors commonly implicated in IVE and CALL learning (usability), we used multiple linear regression.

IV. Results

A. Embodied controls on immediate retention

An independent-samples t-test was conducted to compare post-test learning gains in embodied controls and spoken interaction, and spoken interaction conditions. Embodied controls and spoken interaction (M = 8.8, SD = 4.1) memorisation was significantly higher than for the spoken-only condition (M = 5.5, SD = 3.2) , with a large effect size ($t(22)$ = 2.03, p = .027, g = .88) (see Fig. 3 and Table. II). This suggests that the addition of embodied controls had a large meaningful benefit to immediate retention over the spoken production-only condition.

When considering results for verbs and nouns separately, the results for verbs remain significant ($t(22)$ = 2.06, p = .026, g = .47), but are no longer significant for nouns $t(22)$ = 1.47, p = .078, g = .66). This suggests that the benefits of embodied controls are only demonstrated for verbs (which we could consider an *embodied action* memorisation), rather than nouns (an *embodied object-interaction* memorisation).

B. Embodied controls on long-term retention

We found no significant interaction between interaction type and retention ($F(1,22)$ = 1.18, p = .29). Therefore while embodied controls and spoken production promote better memorisation, we could not attribute this to any specific retention benefits afforded by embodied controls (Fig. 4).

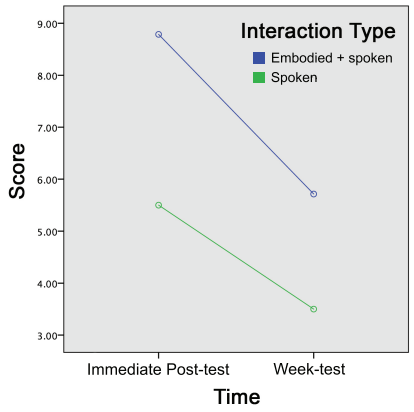

Fig. 4. The average (mean) changes between immediate and one-wee learning gain showed no significant relationship between interaction type and retention over time

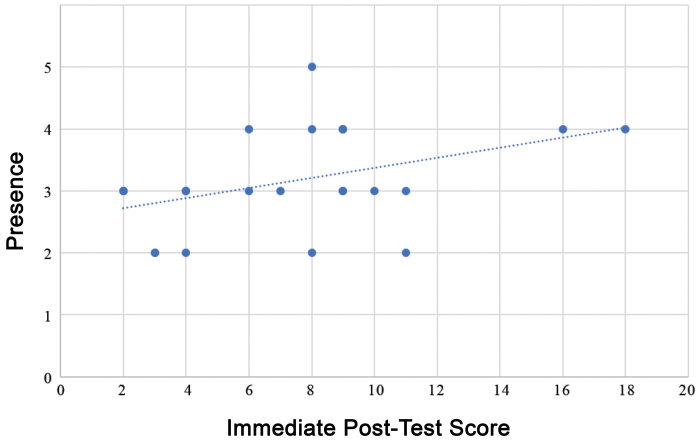

Fig. 5. Linear regression showing the significant relationship between post-test score and presence

C. Cognitive load, motivation, presence and usability

The interaction type (embodied controls and spoken production vs. spoken production only) had no significant impact on reported levels of presence, cognitive load, usability or motivation (See Table I).

However, using a backward step-wise multiple linear regression, we found that presence, cognitive load, usability and motivation explained a significant amount of the variance in the immediate post-test learning gain ($F(1, 22) = 4.55$, $p < .05$, $R2 = .17$). Analysis showed that usability did not significantly predict post-test outcomes (*Beta* = -.17, $t(19) = -.77$), nor cognitive load (*Beta* = 0.19, $t(20) = .91$); nor motivation (*Beta* = .22, $t(21) = 1.13$).

Self-reported presence significantly predicted immediate post-test learning gain (*Beta* = .355, $t(22) = 1.89$, $p < .05$) to a moderate degree (Fig. 5).

Tests to see if the data met the assumption of collinearity indicated that multi-collinearity was not a concern.

V. Discussion

Our results show that combined embodied controls and spoken production provide significant and notable learning benefits over spoken production alone in an IVE. This replicates findings in the real-world [5] and suggests that embodied learning benefits carry over from the physical space into the virtual one. However, we only found a significant difference for the memorisation of verbs, not nouns. This suggests that embodied actions (verbs) and embodied interactions with objects (nouns) may be importantly distinct from each other to a degree that could effect learning via embodied controls.

Beyond this, the lack of relationship with the co-variables typically associated with computer-aided learning (cognitive load, presence, usability and motivation) provides some insight into how embodied controls aid memorisation. If Asher's assertion, that motivation was the key factor in gesture aiding language acquisition, was correct [21], we would have expected to see higher motivation scores in the embodied controls and spoken production group. However, this was not the case. Similarly, we could not find that embodied controls made the system more usable, reduced cognitive load [31] or increased the feeling of presence in users compared with the control.

Therefore the most likely cause of the benefits of the embodied controls is that embodied cognition played an important part in the success of the memorisation beyond simply increasing motivation, as previously suggested by Macedonia [6]; or that an as-yet unrecorded factor is influencing the learning.

We originally planned to investigate cognitive load to help understand the reasons for a non-significant learning difference. Cognitive load research suggests that IVEs, and IVEs with more modalities, can harm learning through increased cognitive load demands [8]. Our results do not support the hypothesis that adding embodied controls in an IVE increases cognitive load. Nor did we find that adding embodied controls reduces cognitive load, as proposed by [31]. While there was a difference in the mean cognitive load for the embodied controls and spoken group ($M = .71$) and that of the spoken-only group ($M = .50$), it was not significant ($t(22) = .37$, $p = .36$). However, we also found no correlation between reported cognitive load and participants' results, which could also suggest that our measure of cognitive load was not sufficiently sensitive.

Finally, we found that a learner's presence score correlated with their performance, reinforcing a common conception in IVE learning. However, we did not record a significant difference between reported presence in either interaction type, suggesting that embodied controls do not increase the feeling of presence. Therefore while experiencing more presence helped learning, embodied controls did not enhance this feeling, and therefore the benefits of the embodied interaction could not be attributed to this factor.

TABLE I
CORRELATION MATRIX FOR EMBODIED CONTROLS, COGNITIVE LOAD, PRESENCE AND MOTIVATION (NO SIGNIFICANT RESULTS)

	Embodied	Cognitive	Motivation	Presence
Embodied	1			
Cognitive	0.078	1		
Motivation	0.307	0.069	1	
Presence	0.176	0.367	0.189	1

TABLE II
COMPARING SPOKEN, AND EMBODIED ACTION + SPOKEN POST-TEST LEARNING RESULTS

Group	N	Mean	SD	t	Sig
Spoken	10	5.5	4.25	2.03	
Embodied action + Spoken	14	8.78	3.34	2.03	0.027

VI. LIMITATIONS

The environment was designed to maximise the physicality of the learning, with grabbable nouns and verbs as the target acquisitions. Therefore caution should be used in trying to extrapolate these results for more abstract language concepts, such as adjectives, and for the learning of other subjects.

We should also avoid extrapolating these results to language learning generally: the environment and its memorisation objective are non-natural, and present a focus word, rather than language, acquisition. How some of the research's outcomes – such as the benefits of embodied action for verb acquisition – might contribute to other important aspects of second language acquisition, such as communicative competence, is still unclear and not covered in this work.

There are also some questions about the sensitivity of the validated questions we used to understand participant presence and cognitive load. We used very condensed questioning, which may not be as robust as more comprehensive surveying, or if also paired with other quantitative measures.

VII. CONCLUSION

This study showed that using embodied controls and spoken production to interact with an immersive virtual environment aided second language memorisation over spoken production alone. By examining co-variables typically associated with computer-aided learning, it provides evidence that embodied controls have a positive effect on language memorisation that is not explained by enhanced motivation levels or increased presence response, and therefore provides evidence for embodied cognition theories of learning, and the efficacy of those approaches in immersive virtual environments.

The use of embodied controls and spoken production had no effect on the perceived cognitive load of participants; their motivation; nor their experience of presence. However, the study found that greater feelings of presence correlated with better learning outcomes.

The distinction between the significant result for verbs and the non-significant result for nouns poses interesting further questions: could the benefits of embodiment be limited to the movements your body makes, and not aid memorisation related to the wider, spatially-embodied environment? One way to investigate this is through better understanding the distinction between embodied action and embodied object interactions in IVEs, as the mechanism through which we consider movements in IVEs is still unclear: are we physically gesturing in the real world, with reactions from a virtual one, or are we taking "physical" actions in a virtual space? Also, if noun learning is not benefiting from this type of embodiment, perhaps we need to examine how virtual objects are considered in relation to ones in the physical world. We could also explore this area from an observational perspective: do the embodied actions we perceive being performed by virtual agents have an impact on learners, compared with less-embodied agents, and to what extent?

ACKNOWLEDGMENT

The authors would like to thank all subjects who participated in this study, and our anonymous reviewers. This work is supported by the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1).

REFERENCES

[1] M. C. Howard, "Virtual reality interventions for personal development: A meta-analysis of hardware and software," *Human–Computer Interaction*, vol. 34, no. 3, pp. 205–239, 2019.

[2] M. Wilson, "The case for sensorimotor coding in working memory," *Psychonomic Bulletin & Review*, vol. 8, no. 1, pp. 44–57, 2001.

[3] C. Vázquez, L. Xia, T. Aikawa, and P. Maes, "Words in motion: Kinesthetic language learning in virtual reality," in *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 2018, pp. 272–276.

[4] S. W. Eskildsen and J. Wagner, "Embodied l2 construction learning," *Language Learning*, vol. 65, no. 2, pp. 268–297, 2015.

[5] M. Macedonia and T. R. Knösche, "Body in mind: How gestures empower foreign language learning," *Mind, Brain, and Education*, vol. 5, no. 4, pp. 196–211, 2011.

[6] M. Macedonia, C. Repetto, A. Ischebeck, and K. Mueller, "Depth of encoding through observed gestures in foreign language word learning," *Frontiers in psychology*, vol. 10, 2019.

[7] M. C. Johnson-Glenberg, C. Megowan-Romanowicz, D. A. Birchfield, and C. Savio-Ramos, "Effects of embodied learning and digital platform on the retention of physics content: Centripetal force," *Frontiers in psychology*, vol. 7, p. 1819, 2016.

[8] C. Schrader and T. J. Bastiaens, "The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games," *Computers in Human Behavior*, vol. 28, no. 2, pp. 648–658, 2012.

[9] S. Van Der Land, A. P. Schouten, F. Feldberg, B. Van Den Hooff, and M. Huysman, "Lost in space? cognitive fit and cognitive load in 3d virtual environments," *Computers in Human Behavior*, vol. 29, no. 3, pp. 1054–1064, 2013.

[10] G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learning and Instruction*, vol. 60, pp. 225–236, 2019.

[11] C. M. MacLeod, N. Gopie, K. L. Hourihan, K. R. Neary, and J. D. Ozubko, "The production effect: Delineation of a phenomenon." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 36, no. 3, p. 671, 2010.

[12] E. M. Golonka, A. R. Bowles, V. M. Frank, D. L. Richardson, and S. Freynik, "Technologies for foreign language learning: a review of technology types and their effectiveness," *Computer assisted language learning*, vol. 27, no. 1, pp. 70–105, 2014.

[13] S. D. Kelly, T. McDevitt, and M. Esch, "Brief training with co-speech gesture lends a hand to word learning in a foreign language," *Language and cognitive processes*, vol. 24, no. 2, pp. 313–334, 2009.

[14] M. Wilson, "Six views of embodied cognition," *Psychonomic bulletin & review*, vol. 9, no. 4, pp. 625–636, 2002.

[15] M. H. Fischer and R. A. Zwaan, "Embodied language: A review of the role of the motor system in language comprehension," *The Quarterly Journal of Experimental Psychology*, vol. 61, no. 6, pp. 825–850, 2008.

[16] R. L. Thompson, D. P. Vinson, B. Woll, and G. Vigliocco, "The road to language learning is iconic: Evidence from british sign language," *Psychological science*, vol. 23, no. 12, pp. 1443–1448, 2012.

[17] K. Masumoto, M. Yamaguchi, K. Sutani, S. Tsuneto, A. Fujita, and M. Tonoike, "Reactivation of physical motor information in the memory of action events," *Brain research*, vol. 1101, no. 1, pp. 102–109, 2006.

[18] Ş. Özçalışkan and S. Goldin-Meadow, "Gesture is at the cutting edge of early language development," *Cognition*, vol. 96, no. 3, pp. B101–B113, 2005.

[19] B. Zinober and M. Martlew, "Developmental changes in four types of gesture in relation to acts and vocalizations from 10 to 21 months," *British Journal of Developmental Psychology*, vol. 3, no. 3, pp. 293–306, 1985.

[20] S. G. McCafferty, "Space for cognition: Gesture and second language learning," *International Journal of Applied Linguistics*, vol. 14, no. 1, pp. 148–165, 2004.

[21] J. J. Asher, "The total physical response approach to second language learning," *The modern language journal*, vol. 53, no. 1, pp. 3–17, 1969.

[22] M. Schmidt-Kassow, M. Deusser, C. Thiel, S. Otterbein, C. Montag, M. Reuter, W. Banzer, and J. Kaiser, "Physical exercise during encoding improves vocabulary learning in young female adults: a neuroendocrinological study," *PloS one*, vol. 8, no. 5, p. e64172, 2013.

[23] D.-F. Yap, W.-C. So, J.-M. Melvin Yap, Y.-Q. Tan, and R.-L. S. Teoh, "Iconic gestures prime words," *Cognitive science*, vol. 35, no. 1, pp. 171–183, 2011.

[24] D. Edge, K.-Y. Cheng, and M. Whitney, "Spatialease: learning language through body motion," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013, pp. 469–472.

[25] C. Repetto, E. Pedroli, and M. Macedonia, "Enrichment effects of gestures and pictures on abstract words in a second language," *Frontiers in psychology*, vol. 8, p. 2136, 2017.

[26] M. A. Conway and S. E. Gathercole, "Writing and long-term memory: Evidence for a "translation" hypothesis," *The Quarterly Journal of Experimental Psychology*, vol. 42, no. 3, pp. 513–527, 1990.

[27] S. D. Kelly, A. Özyürek, and E. Maris, "Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension," *Psychological Science*, vol. 21, no. 2, pp. 260–267, 2010.

[28] D. McNeill, "Gesture, gaze, and ground," in *International workshop on machine learning for multimodal interaction*. Springer, 2005, pp. 1–14.

[29] K. Bergmann and M. Macedonia, "A virtual agent as vocabulary trainer: iconic gestures help to improve learners' memory performance," in *International workshop on intelligent virtual agents*. Springer, 2013, pp. 139–148.

[30] B. F. Freed, N. Segalowitz, and D. P. Dewey, "Context of learning and second language fluency in french: Comparing regular classroom, study abroad, and intensive domestic immersion programs," *Studies in second language acquisition*, vol. 26, no. 2, pp. 275–301, 2004.

[31] A. Steed, Y. Pan, F. Zisch, and W. Steptoe, "The impact of a self-avatar on cognitive load in immersive virtual reality," in *2016 IEEE Virtual Reality (VR)*. IEEE, 2016, pp. 67–76.

[32] G. Petri, C. G. von Wangenheim, and A. F. Borgatto, "Meega+: an evolution of a model for the evaluation of educational games," *INCoD/GQS*, vol. 3, 2016.

[33] F. G. Paas, "Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach." *Journal of educational psychology*, vol. 84, no. 4, p. 429, 1992.

[34] M. Slater and M. Usoh, "Representations systems, perceptual position, and presence in immersive virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 2, no. 3, pp. 221–233, 1993.

[35] W. A. IJsselsteijn, H. De Ridder, J. Freeman, and S. E. Avons, "Presence: concept, determinants, and measurement," in *Human vision and electronic imaging V*, vol. 3959. International Society for Optics and Photonics, 2000, pp. 520–529.

[36] N. D. Fleming, *Teaching and learning styles: VARK strategies*. IGI global, 2001.