EVIDENCE FOR EMBODIED COGNITION: QUANTIFYING THE IMPACT OF GESTURE ON SECOND LANGUAGE ACQUISITION IN AN IMMERSIVE VIRTUAL ENVIRONMENT

Anonymous ICME submission

ABSTRACT

This paper provides evidence that leveraging gesture and spoken interaction in an immersive virtual environment aids language memorisation compared with spoken-only interaction. It finds that this benefit does not correlate with player experience ratings, suggesting that the novelty of gesture-induced motivation is not the cause of the learning enhancements. It also finds no significant cognitive load difference between the two interaction types.

***Index Terms*—** One, two, three, four, five

1. INTRODUCTION

Technological advances have seen an increase in the use of gesture and speech recognition as inputs for computer-aided language learning environments. The majority of research on these interaction methods has focused on exploring a single input modality, such as spoken interaction for automatic speech recognition [1] or gesture recognition for verb memorisation [2]; or examines the input modalities as part of multi-modal systems that incorporate many immersive approaches in an attempt to closely mimic a realistic situated experience [3]. This has left a gap in literature: examinations of how specific immersive inputs interact with each other, and the extent that combinations of these contribute to or detract from learning outcomes. One combination of inputs that has strong theoretical and experimental support for creating positive learning outcomes is that of iconic gesture and spoken production.

The positive impact that actively speaking foreign language words has on second language memorisation is known as the production effect [4], 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 [1].

The benefits of iconic gesture for learning have been well-explored from an embodied cognition perspective [5]. Experiment results demonstrate positive learning benefits related to using iconic gestures with technology [2] and without [6]; inside language learning [4], and for wider learning applications [7].

Recent experiments combining gesture and spoken production have also demonstrated positive learning outcomes [8], [4]. However, these experiments did not take place inside immersive virtual environments (IVEs), and whether and how these benefits transfer into IVEs is yet to be investigated. Generally, learning experiments in IVEs have shown positive results [9], but the cause of the benefits is still under-researched. There are also concerns that multi-modal learning in an IVE might cause cognitive load issues that harm information retention [10] [11], even if IVEs and gesture and spoken interaction are separately useful learning tools.

This experiment was designed to further explore whether gesture, when combined with spoken production, aids language memorization over spoken production alone, and whether this effect carries over into an IVE. It also attempts to understand what aspect of this interaction causes the memorisation benefits, by monitoring co-variables that are commonly attributed to impacting IVE learning, such as system usability, player experience, cognitive load and presence. The results have implications for understanding whether embodied cognition occurs inside IVEs, as well as the choice of user inputs included in future variations of language learning IVEs.

2. LITERATURE

2.1. Embodied cognition, iconic gesture and language acquisition

From the embodied cognition perspective, cognitive processes are rooted in the body's interactions with the world [12]. Language is generally considered from this perspective, with a strong relationship with embodied gesture [13]. Of the multiple gesture types, iconic gestures have been considered "fundamental to all languages … [bridging] the gap between linguistic form and human experience" [14]. Areas of the brain responsible for iconic gestures and physical actions have been shown to activate when associated words are used or heard [15], [16]. For language acquisition, iconic gestures are considered universally important for both first and second language acquisition [17], [18], and have been considered an additional "mode of thinking" [19] for second language learners.

Of Wilson's six claims of embodied cognition [5], one is particularly relevant to the relationship between iconic ges-

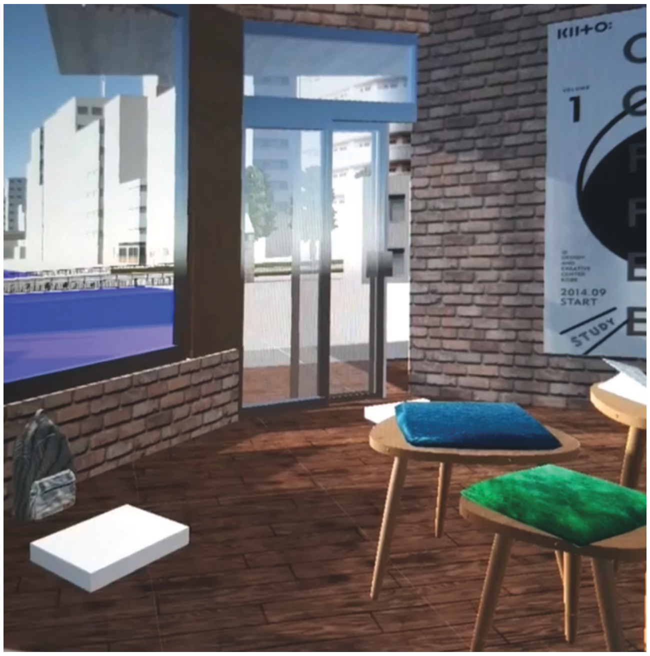


Fig. 1. Image of the virtual learning environment

tures and language acquisition: 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". 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 [20] 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 iconic gesture 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 gestures - is beneficial to vocabulary acquisition and retention [21]. However, there is strong evidence that the positive relationship between iconic gestures and acquisition is not entirely based on it being physically active or highly 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 words (e.g. jumping to learn the word "kick"), have significant word memorisation retention benefits [22], [15]. If the learning benefits were due solely due 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, 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 iconic gesture as a tool for language memorisation, experimental results have proved positive: Vasquez [2] used iconic gestures to help with listening skills related to verbs that correspond to the gesture enacted by the learner; Edge [23] had users enact a sequence of movements to complete a foreign-language movement instruction; Macedonia [24] 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 learning benefits via immersive embodied controls have seen mixed results. Howard's comprehensive meta-analysis of immersive technologies showed that embodied controls did not have a meaningful impact on cognitive learning [9]. This could be because the embodied interactions in Howard's investigations did not distinguish between iconic and non-iconic gesture interventions, nor did it distinguish the relationship between the action and its meaning. It could also be because embodied cognition is especially relevant to language acquisition, rather than more abstract learning topics that Howard included in his analysis, such as learning mathematics or physics.

2.2. 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 [4]. 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 [5].

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 [24]. 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.

2.3. Gestures + Spoken production = Success?

There is evidence that combining gesture and spoken production causes enhanced language memorization. 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. In 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 memorization learning outcomes by having learners combine gesture with simultaneous, relevant spoken production [8]. 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 [20]. It remains to be discovered how these aspects relate in an immersive virtual environment.

2.4. Immersive virtual environments

Immersive virtual environments are increasingly explored as a potential tool for enhancing 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 [5]. The results have been mixed [9], however. Howard found that the dominant contributing factors to learning are head-mounted displays and interactive environments, with embodied interaction technologies providing little impact on learning outcomes. 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 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 theory.

Where found, lower learning rates in immersive environments have been attributed to issues with cognitive load [10] [11], with claims that immersion itself creates a large cognitive load that detracts from a learner's ability to memorise information. If this claim is true, we might see worse memorisation outcomes from speech and gesture than speech alone in this experiment. However, Steed et al. found that the use of embodied controls in IVEs actually reduced cognitive load [30].

3. EXPERIMENT

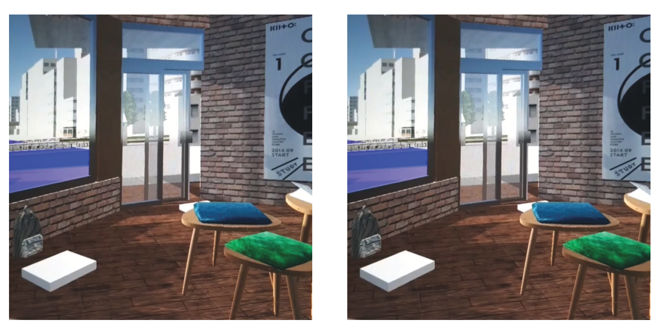
We propose an experiment to understand if interacting in an IVE using gesture and spoken production is more effective for language memorisation than spoken production alone. By monitoring co-variables considered related to learning in IVEs, such as cognitive load, usability, player experience and presence, we will be able to theorise whether the addition of gesture provides benefits due to an implicit advantage of embodied controls (potentially stemming from embodied cognition), or whether it is is because of gesture's impact on one of the other aspects monitored.

3.1. Hypothesis

- Language memorisation occurs when using gesture + spoken production in an IVE
- Leveraging gesture + spoken production while learning leads to better language learning than spoken production alone
- Cognitive load does not vary significantly between the two groups

3.2. Procedure

Each participant was assigned to either a gesture 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



Gesture + spoken   Spoken

Fig. 2. Environment interaction differences

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

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 gestures and spoken input, and if they were successful, the interaction area ends and a participant may visit the other interaction points. Failed recognitions re-prompted users until they correctly performed the spoken utterance or action. Users can also leave an interaction area at any point.

3.3. Participants

Participants (n = 24) 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. No participants demonstrated an extensive knowledge of the words 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 difference in learning outcome between mono-lingual and multi-lingual participants. We also failed to notice any notable difference in ability between participants with different interests in Japanese, Japan, virtual reality and coffee shops.

3.4. Corpus

Participants were tested on their knowledge of 10 noun/verb pairs (20 words). Japanese gairaigo (import words) were specifically avoided to reduce the chance of participants' inferring a meaning.

3.5. 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.

3.6. 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 subtracted from later test scores to ensure only acquired knowledge was included in the results.

After using the system, participants were asked to complete a MEEGA+ questionnaire [31] to provide insight on the system usability and their player experience. Participants were also asked to self-report their cognitive load on a single-item, 9-point Likert scale as defined by Paas [32], and their level of presence while inside the environment on Slater's single-item, 6-point Likert scale [33]. Asking participants for their subjective evaluation of presence experienced is considered the most direct way of presence assessment (Ijsselsteijn, de Ridder, Freemann and Acons, 2002).

3.7. Learning Style

Participants were asked to complete the VARK learning preference questionnaire [34] 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.

3.8. Analysis

In order to answer our main research question, does gesture to spoken production aid learning, we used a one-tailed independent t-test on the post-test scores of the two groups. The data for each group was normally distributed, and met the requirements of homogeneity.

To determine if there was a different impact on learning retention between the groups, we used a mixed ANOVA to examine the results of the post-test and week-tests.

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

4. RESULTS

4.1. Gesture on immediate retention

An independent-samples t-test was conducted to compare post-test score in gesture and spoken interaction and spo-

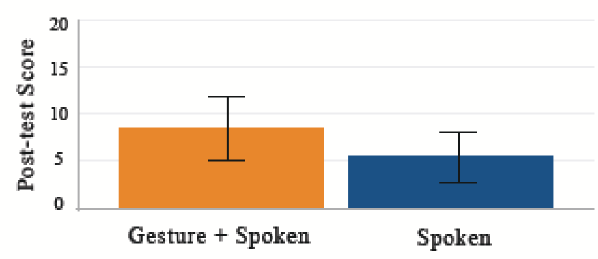


Fig. 3. Showing post-test results for Gesture + Spoken (m = 8.8), and Spoken (m = 5.5) groups, mean and SD.

ken interaction conditions. Gesture and spoken interaction (M=8.8, SD=4.2) memorisation was significantly higher than spoken (M=5.5, SD=3.3) conditions (t(22)=1.55, p=0.046). This suggests that the gesture and spoken production had a meaningful benefit to immediate retention over the spoken production condition.

4.2. Gesture on long-term retention

There was a significant main effect of interaction type on post-test and week-test scores overall (F(1,32) = 225, p = .042). However, there was no significant interaction between interaction type and length of retention (F(1,32) = 0.11, p = .93). Therefore, while gesture and spoken production promote better memorisation in both the short and long terms, the interaction type had no additional impact on user retention.

4.3. Presence, Cognitive load, Usability and experience

The interaction type (gesture and spoken production vs. spoken production only) had no impact on reported levels of presence, cognitive load, usability and or player experience.

However, using the backward step-wise method for a multiple linear regression, we found that presence, cognitive load, usability and player experience explain a significant amount of the variance in the overall post-test scores (F(1, 22) = 4.55, p ¡ .05, R2 = .17, R2Adjusted = 0.134). Analysis showed that usability did not significantly predict post-test outcomes (Beta = -0.17, t(19) = -0.77). Further iterations showed that cognitive load (Beta = 0.19, t(20) = 0.91) and player experience (Beta = 0.22, t(21) = 1.13) did not predict post-test outcomes.

Self-reported presence significantly predicted post-test scores (Beta = .37, t(22) = 1.89, p ¡ .05) to a moderate degree.

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

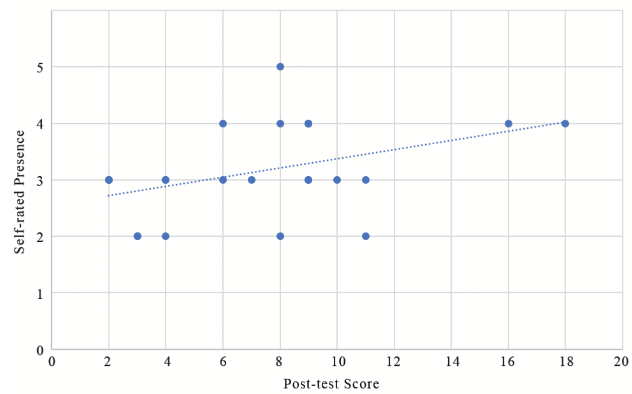


Fig. 4. Linear regression showing relationship between post-test score and presence

5. DISCUSSION

Our results show that combined gesture and spoken production provides significant learning benefits over spoken production alone in an IVE. This replicates findings in the real-world [15] and suggests that embodied learning benefits carry over from the physical space into the virtual one.

Beyond this, the lack of relationship between either interaction method with the co-variables typically associated with computer-aided learning (cognitive load, presence, usability and player experience) provides some insight into how gesture interaction aids memorisation. If Asher's assertion that motivation was the key factor in gesture aiding language acquisition was correct ([20], we would have expected to see higher player experience scores in the gesture and spoken production group. However, this was not the case. Similarly, the embodied controls did not make the system more usable, reduce cognitive load ([30]) or increase the feeling of presence in users compared with the control.

Therefore the most likely cause of the benefits of the gesture and spoken production interaction method is that embodied cognition played an important part in the success of the memorisation beyond simply motivation, as suggested by Macedonia [24]; or that an as-yet unrecorded factor is influencing the learning.

We originally planned to investigate cognitive load to help us 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. Our results do not support the hypothesis that adding gesture control in an IVE contributes to cognitive load demands [10], but nor did it reduce it [30]. While there was a difference in the mean for the gesture and spoken group (m = 0.71) and that of the spoken-only group (m = 0.50), it was not significant (t(22) = 0.37, p = 0.36). However, we also found that cognitive load had no correla-

Table 1. Comparing Spoken, and Gesture + Spoken post-test learning results

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

tion with 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 (R2 = 0.17), 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 have a notable impact on feelings of presence. Therefore while users feeling present helps their learning, embodied controls did not enhance this feeling, and therefore the benefits of the embodied interaction could not be attributed to this factor.

6. 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 other learning subjects. We should also avoid extrapolating these results to language learning generally. This environment and its memorisation objective are non-natural, limited methods of word - rather than language - acquisition. How some of the potential outcomes – such as the promotion of gesture and embodiment – 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. We also should be careful about using player experience as a perfect representation of motivation, as there are still many methods and definitions of motivation and its relationship to learning.

7. CONCLUSION

This study showed that using gesture 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 gesture interaction has a positive effect on language memorisation that is not explained by enhanced motivation levels or increased presence response, and therefore could provide evidence for embodied cognition theories of learning and in immersive virtual environments.

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

Future work should continue to explore the applications of embodied cognition in immersive learning environments for more complicated and natural types of language acquisition, such as communicative competence inside spoken dialogue systems.

8. REFERENCES

[1] Ewa M Golonka, Anita R Bowles, Victor M Frank, Dorna L Richardson, and Suzanne 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.

[2] Christian Vázquez, Lei Xia, Takako Aikawa, and Pattie 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.

[3] Rahul R Divekar, Yalun Zhou, David Allen, Jaimie Drozdal, and Hui Su, "Building human-scale intelligent immersive spaces for foreign language learning," *iLRN 2018 Montana*, p. 94, 2018.

[4] Colin M MacLeod, Nigel Gopie, Kathleen L Hourihan, Karen R Neary, and Jason D Ozubko, "The production effect: Delineation of a phenomenon.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 36, no. 3, pp. 671, 2010.

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

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

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

[8] Spencer D Kelly, Tara McDevitt, and Megan 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.

[9] Matt 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.

[10] Claudia Schrader and Theo 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.

[11] Sarah Van Der Land, Alexander P Schouten, Frans Feldberg, Bart Van Den Hooff, and Marleen 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.

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

[13] Martin H Fischer and Rolf 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.

[14] Robin L Thompson, David P Vinson, Bencie Woll, and Gabriella Vigliocco, "The road to language learning is iconic: Evidence from british sign language," *Psychological science*, vol. 23, no. 12, pp. 1443–1448, 2012.

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

[16] Kouhei Masumoto, Masahiko Yamaguchi, Kouichi Sutani, Satoru Tsuneto, Ayako Fujita, and Mitsuo Tonoike, "Reactivation of physical motor information in the memory of action events," *Brain research*, vol. 1101, no. 1, pp. 102–109, 2006.

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

[18] Brenda Zinober and Margaret 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.

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

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

[21] Maren Schmidt-Kassow, Marie Deusser, Christian Thiel, Sascha Otterbein, Christian Montag, Martin Reuter, Winfried Banzer, and Jochen Kaiser, "Physical exercise during encoding improves vocabulary learning in young female adults: a neuroendocrinological study," *PloS one*, vol. 8, no. 5, pp. e64172, 2013.

[22] De-Fu Yap, Wing-Chee So, Ju-Min Melvin Yap, Ying-Quan Tan, and Ruo-Li Serene Teoh, "Iconic gestures prime words," *Cognitive science*, vol. 35, no. 1, pp. 171–183, 2011.

[23] Darren Edge, Kai-Yin Cheng, and Michael Whitney, "Spatialease: learning language through body motion," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013, pp. 469–472.

[24] Manuela Macedonia, Claudia Repetto, Anja Ischebeck, and Karsten Mueller, "Depth of encoding through observed gestures in foreign language word learning," *Frontiers in psychology*, vol. 10, 2019.

[25] Claudia Repetto, Elisa Pedroli, and Manuela Macedonia, "Enrichment effects of gestures and pictures on abstract words in a second language," *Frontiers in psychology*, vol. 8, pp. 2136, 2017.

[26] Martin A Conway and Susan 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] Spencer D Kelly, Aslı Özyürek, and Eric 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] David McNeill, "Gesture, gaze, and ground," in *International workshop on machine learning for multimodal interaction*. Springer, 2005, pp. 1–14.

[29] Kirsten Bergmann and Manuela 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] Anthony Steed, Ye Pan, Fiona Zisch, and William 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.

[31] Giani Petri, C Gresse von Wangenheim, and Adriano Ferretti Borgatto, "Meega+: an evolution of a model for the evaluation of educational games," *INCoD/GQS*, vol. 3, 2016.

[32] Fred 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, pp. 429, 1992.

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

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