Words in Motion: Kinesthetic Language Learning in Virtual Reality

Christian D. Vázquez
MIT Media Lab
Cambridge, USA
cdvm@media.mit.edu

Takako Aikawa
MIT GSL
Cambridge, USA
aikawa@mit.edu

Pattie Maes
MIT Media Lab
Cambridge, USA
pattie@media.mit.edu

ABSTRACT

Embodied theories of language propose that the way we communicate verbally is grounded in our body. Nevertheless, the way a second language is conventionally taught does not capitalize on kinesthetic methodologies. The tracking capabilities of room-scale virtual reality systems afford a way to incorporate kinesthetic learning in educational experiences. We explore the potential for kinesthetic learning in virtual environments within the scope of second language vocabulary acquisition. We present *Words in Motion*, a virtual reality application that tracks the user's body movements, engaging students in a game where they perform a sequence of actions while learning new words, thereby reinforcing associations between word-action pairs. Results suggest that the kinesthetic approach in virtual reality has less immediate learning gain in comparison to a text-only condition for equal exposure time. However, kinesthetic learners showed significantly higher retention rates, and showed positive correlation between the number of performed actions and the times a word is remembered, supporting our hypothesis that virtual reality can impact language learning by leveraging kinesthetic elements.

Author Keywords

Virtual Reality; Kinesthetic Learning; Language Learning; Embodied cognition; Embodiment-based learning

ACM Classification Keywords

H.5.1 [H.5.1 Multimedia Information Systems] Artificial, augmented, and virtual realities.

INTRODUCTION

Language has been shown to be linked to our bodily experiences (Barsalou, 2008; Shapiro, 2011). That is, the cognitive processes through which we ground language are directly affected by our physical reality and how it relates to our bodies. Even the way we structure our metaphors in language directly aligns with body-centric associations (e.g. “up” is good, “down” is bad) (Lakoff, 2011). This connection is so intrinsic that brain imaging has shown that sensorimotor regions in the brain associated with carrying out a task light up when words associated with the corresponding actions are used or heard (Masumoto 2006; Watson 2014; Macedonia 2011; Fischer 2008). These effects hold even for abstract words that are not necessarily linked to explicit action (Moseley 2012). Moreover, studies (Dudschig, 2014; De Grauwe 2014) support the notion that this relationship holds beyond our first language. For instance, learners presented with spatially suggestive words (e.g. “sky” strongly relates to the notion of “up”) show no difference in reaction times when asked to either raise or lower their hands in response to the perceived spatial association, suggesting no discrepancy between the bodily encoding of first and second languages (Dudschig, 2014).

CHI’18, May 6–May 11, 2018, Denver, Colorado.
Copyright © 2018 ACM ISBN/14/17
DOI string from ACM form confirmation.

The effect of body action has been widely explored as a memory enhancement tool[x]. Most notably, the enactment effect or subject performed task effect, showed that performing a set of actions can generally increase the capability of subjects to recall these tasks[x]. Given embodiment’s notable effects on memory, its impact on second language learning has also been studied extensively within the context of vocabulary acquisition [x]. Use of iconic gestures or illustrative motions have shown higher learning gain and retention levels than text, visual, or audio centric approaches.

Kinesthetic learning refers to learning that occurs through physical activity, and is deeply rooted in the relationship our mind has with our body. Total Physical Response (TPR) (Asher, 1969) was introduced as a kinesthetic framework that consisted of teachers giving spoken commands to be performed by the student. This technique leveraged the embodied theories of language by creating associations between the spoken order and the subsequent physical response of the learner. However, the framework was not formally evaluated within the field of education and remained theoretical in nature (Macedonia, 2014).

Although the role the body plays in the cognitive processes that involve language and learning has been shown to be relevant, second language education is predominantly audiovisual in nature (Choo, 2012; Graham, 2014). Kinesthetic learning comes with certain constraints that are inherent in its physical characteristics. Limitations in the teaching space, reduce the range of interactions that students can engage in within a classroom environment. Moreover, teachers don’t have tools to observe and provide real-time feedback regarding student actions, making it a scalability problem in formal teaching scenarios. Virtual Reality (VR) has often been proposed as a platform that affords embodied learning. Due to its immersive nature, and body tracking capabilities, VR can allow students to engage in kinesthetic activities that are able to track and understand their movements, provide real-time feedback, and engage them in activities within novel contexts that strongly relate to their physical actions.

In this paper, we present an exploration of virtual reality as a platform to learn second languages kinesthetically with a focus on vocabulary acquisition. We present a platform, *Words in Motion*, that tracks and recognizes user actions in 3D space and provides feedback in the form of the corresponding action verb in the targeted second language. The design of a classroom activity using the *Words in Motion* system is detailed, as well as findings from qualitative interviews with 11 participants regarding the usability and overall experience within the kinesthetic learning activity. A user study with 40 students compares the learning potential of the kinesthetic virtual reality platform with text-only methods, so as to better understand the role it can play in the field of second language education. The work closes with a discussion of our findings and future avenues of research that can build on the insights presented in this paper.

RELATED WORK

Technology enabled kinesthetic learning has been explored as a means to enhance education within different disciplines. The most common usage is learning tasks which are characteristically physical, such as meditation[x], or sport related[x], due to the capability of tracking technology to provide real-time feedback for pose-related activities. However, technology has been exploited to teach activities that are not characteristically physical, leveraging embodied cognition to teach chemistry[x], math[x] ,and anatomy[x], among many others. In this work we focus specifically in the domain of language learning, and do so in the context of vocabulary acquisition.

A large body of work has been dedicated to language learning projects that involve Kinect based tracking to enhance language education. Fundamentally, some projects purely capitalize on the use of the body as a control for the experience, as is the case for JaJan[x] or KinEd[x], systems that create mixed reality environments that can be manipulated with physical gestures to engage in learning activities. Alternatively, a middle ground is covered by projects that leverage movements with weak associations to the linguistic material. An example of this is WordOut[x], a project turned museum exhibit in which children can use their body to match letter shapes as a literacy enhancing activity. Although evidence suggests that experiences in these two categories also benefit to some extent from kinesthetic embedding of knowledge, the experiences presented in this paper are concerned with a third category. That is, technology enhanced kinesthetic activities that explicitly create contextually relevant associations with the language material.

Within the realm of this third category, a variety of systems have been devised and tested on distinct populations. *SpatialEase* was a game developed by Microsoft Research where the participants respond to audio cues in Mandarin with bodily actions. *SpatialEase* was compared with Rosetta Stone software, showing similar gains for vocabulary and grammar acquisition. Similarly, Kuo et al. developed a Kinect based platform to compare the effectiveness of computer enabled TPR versus the classical TPR method. Testing with 50 elementary students showed no significant differences between conditions immediately after the short learning session. Nevertheless, the retention of vocabulary was shown as increased for the technology enhanced group, which is hypothesized as the result of higher engagement. Similar approaches, all of which make use of Kinect or similar tracking methods, are presented in [2,3,4,5], where a kinesthetic approach based on action recognition and word pairing is compared to standardized [x], audiovisual[x], and conventional[x] methodologies for language learning.

The system presented in this paper approaches the kinesthetic learning of language similarly to the aforementioned body of work. However, our platform leverages kinesthetic experiences within fully immersive environments in virtual reality, whereas existing platforms emphasize on non-immersive analogues. That is, experiences where the learner focuses on a monitor in front of them which shows their body pose in relation to a target pose as opposed to being inside a space focusing on their body[x]. Comparisons between computer enabled TPR versus classical TPR suggest that there is a difference between the learning gain, despite both activities requiring the same bodily motions [x]. This implies that the modality through which the kinesthetic learning occurs directly impacts the way material is encoded in bodily motion, and motivates an exploration of virtual reality as a platform for kinesthetic language learning. Although virtual reality is fundamentally an embodied platform, to the best of our knowledge, it has not been used for language learning that explicitly involves the use of the body as a tool to encode contextually relevant vocabulary, which is the focus of this work.

WORDS IN MOTION

Words in Motion is a system designed for the HTC Vive that enables kinesthetic language learning in VR. The platform was developed as a recognition system that allows teachers or students to participate in activities that introduce kinesthetic elements to second language instruction. In this section we describe the system, as well as our design of an activity that leverages *Words in Motion* to enable an interactive classroom experience in a virtual kitchen.

At its core, *Words in Motion* allows the user to embed objects in a virtual environment with verb-action pairs to reinforce associations between new vocabulary words and contextually relevant locomotion. Every object in the virtual environment is supported with a feed-forward neural net trained to recognize actions performed with them. When the learner grabs an object from the environment, he/she can perform the new action to be embedded, multiple times to "teach" the neural net by example to recognize this action. Training a new verb-action pair on the object, creates a motion signifier. This signifier is a dynamically generated animation that displays the path to be performed to trigger a particular verb.

The core recognition software is a heavily modified version of Edwon's ultimate gesture VR system [x]. Users can begin the recognition process by pressing the Vive controller's trigger on the hand holding the virtual object. This provides haptic (a subtle vibration) and visual (a transparent trail) feedback to denote an action is being recorded. When the trigger is released, the path is evaluated and classified. If a path matches a trained action-verb pair with a degree of confidence, the corresponding verb appears floating momentarily in front of the user (Fig. X). Currently, the system is limited to motions with the user's hands, but minimal changes would allow the use of Vive trackers to support actions that involve full body motion.

Activity Design

We designed an activity that leverages the kinesthetic capabilities of *Words in Motion* but also addresses some of the disadvantages noted in purely kinesthetic experiences like TPR: the lack of conversational interactions. The motivation was to engage learners in a game that incorporates both kinesthetic and conversational elements in a way that could be employed as a classroom activity that would be deployed to Kanda University in Tokyo to support English instruction of college level students. A virtual kitchen environment was created and a set of objects were pre-trained with contextual actions-word pairs (e.g. chopping with a knife). The set-up consisted of an HTC Vive in a 15" x 15" space and a pair of external monitors.

The game's goal was to perform a sequence of actions in the virtual kitchen environment with the correct set of objects. The activity was designed with multiple players in mind, where one player is immersed in VR, while the others participate from the real world. This was purposely done to address the fact that many educational facilities have constraints that limit the amount of room-scale devices they can afford or accommodate in a classroom. The participant inside the virtual environment takes the role of the "performer", while the participant(s) outside VR are denoted the observer(s).

The role of the "performer" is to execute the right sequence of actions using the target objects in the kitchen. However, there is no indication within the virtual environment that informs the performer (1) which action needs to be performed, (2) how to perform the action in space, or (3) which object to perform the action with.

Participants outside the virtual environment take on the role of the observers. Observers have two views on external monitors. On one screen, they can monitor the performer's point of view. The other screen displays the motion signifier, an animation that shows the path that the performer must enact, along with an instruction of which object to perform it with. The role of the observers is to communicate verbally with the performer the actions they have to perform, how to perform them, and which object to look for in the virtual kitchen. Observers could also use a computer mouse to rotate the motion indicator, which allowed them to better visualize action paths in 3D space.

There is no method that enforces how communication between the observer and the performer happens. This allows the teacher or moderator to set constraints that give the right measure of difficulty according to the participant's fluency in the target language. This can range from full second language communication, to first language instruction (where only the performer learns the target language by kinesthetic means). This allows teachers to engage the whole classroom as observers that practice conversationally by communicating in the second language with the performers, while taking turns to engage in kinesthetic reinforcement of the material.

The designed scenario also allowed the teacher or observers to select the sequence of action-verb pairs using a desktop interface (Fig. X). These new sequences, which we denote as "challenges", can include pre-trained actions or actions created by the teacher or students.

Activity Trials

The designed *Words in Motion* activity was tested with 11 students with varying conditions. Out of the 11 participants, four acted only as observers while being instructed by a researcher to perform actions. The remaining subjects participated in the activity in pairs, with one participant acting as the performer and the other as the observer.

Participants were to finish a game level that consisted of 5 unique actions, in a set of 6 total actions (one of the actions is required twice.) All of the actions were contextually relevant to the object they had to be carried out with (e.g. chop with a knife). Users with enough proficiency were asked to communicate fully in a second language, but were allowed to revert to English if they became stuck or were not able to transmit the motion path effectively with their limited vocabulary. Participants were interviewed briefly after the experience to obtain feedback on their experience.

All of the performers agreed that the controls were intuitive. The system required little to no training, in order for them to fulfill all the required actions. One of the trials consisted of a participant that had already acted out the role of a performer. As an observer, she was able to explain actions much more efficiently to the other player. This suggests she was able to create familiarity with the system in a relatively short amount of time, such that she was able to transmit that knowledge much more effectively than observers who had never acted out as performers.

Although the participants described the controls as intuitive, a common theme emerged from conversations with the immersed subjects. Out of the 8 performers, 6 noted that the way that actions were requested did not match their mental model of how that action was performed in the real world. One subject mentioned that giving an action a name made it harder for him to perform the action if it didn't converge with his mental model of it. Another participant expressed the realization that the game was, "not about performing actions as he envisions them but actions as the observer wants them." This made actions harder to perform correctly to the users, which in turn detracted from the vocabulary association and focused attention primarily on the action path. This posits an interesting question regarding the effectiveness of the kinesthetic approach that is recognized automatically, as opposed to qualitatively judged by a teacher or observer.

For language learning, placement of the second language word and the feedback given when the correct action is performed is important. In the experience, a success sound was played every time the action was performed. The corresponding word in the second language also popped up in front of the user, momentarily floating away from the hand with which the action was performed. Most participants, report ignoring the success sound and relying on the popped up word for confirmation that the correct action was performed. However, several participants expressed the desire to see the word before they performed the action as opposed to briefly exposing it right after. Others expressed a desire to see all the possible actions that could be performed with an object when they grabbed it, and formally highlighting the performed action-word pair when it is executed successfully.

USER STUDY

A study was conducted with 40 participants to preliminary asses the capabilities of the kinesthetic method in virtual reality in comparison to a control group which learned in a text-only condition. Each group consisted of 20 participants who were recruited from the MIT campus.

Participants were exposed to a set of 20 transitive verbs in Spanish. The words were selected from Subtletx-Esp[x], a frequency list that analyzed dialogue from news articles and popular media. We divided the list into frequency bins and handpicked words in low frequency bins in order to ensure the task would be challenging even to participants with prior exposure to Spanish. An action was then trained in the *Words in Motion* platform for each word. This action would directly match the meaning of the paired word. Finally, we removed any English cognates from our selection to reduce influence of prior knowledge on the experimental results.

Participants were invited to join a training session to learn new vocabulary in Spanish. The students were administered a pre-test to passively recall the English translation of a word after being prompted with the Spanish analogue. Passive recall was selected, as opposed to active recall, to balance the difficulty of the task[x].

The pre-test consisted of 25 words (which included words that the participant would not learn throughout the experience). Participants were instructed to leave an answer blank if they did not know it, but where encouraged to guess if a word seemed familiar. After administering the pre-test, each participant was subjected to a training session. Subjects were randomly assigned to one of two groups: a text-only control condition and the virtual kinesthetic condition supported by *Words in Motion*.

The control group, would sit and try to learn the words by observing slides that cycled through the set of 20 word pairs. Each slide consisted of two words: the target word in Spanish and its corresponding English translation. Each word pair was presented for 15 seconds, with each target word shown twice for a total of 30 seconds worth of exposure to each new vocabulary word in Spanish.

The kinesthetic learners in virtual reality would be standing in an empty room. An indicator attached to their left arm displayed an orb moving along a path. Participants were instructed to perform the movement depicted by the orb at least twice with their right hand. The first movement "triggered" the associated word pair to appear in front of the participant in a similar fashion to the way the word is presented to the control group on the computer screen. The second movement was requested to create a direct association between the word pair and the body action. The word pair remained visible for 15 seconds, before the next action was queued by the indicator. Each action was requested twice throughout the learning session, for a total of 30 seconds worth of exposure to each new vocabulary word in Spanish. Additional metadata about the interactions with the VR system was collected. This includes the amount of successful and failed attempts to perform an action for a particular word.

All groups were tested both immediately after the training session and exactly one week after, using the same test administered before being exposed to the experimental condition. An additional questionnaire was also collected for all participants with information related to language learning background, and engagement through the experience. A 5-point Likert scale was used to assess questions regarding usability and engagement of the VR platform.

RESULTS

Objective Metrics

Participants from both groups scored low on the pre-test, with the control (M=0.65, SD=1.55) and the virtual kinesthetic approach (M=0.25, SD=0.55) knowing almost none of the words prior to the experiment. A t-test was performed to determine whether the difference between group means is significant. There was no significant difference between groups(p<0.05). No participant knew more than 5 words, which held true across participants who had prior experience with Spanish, and can be explained by the selection of vocabulary which was purposely low frequency words that would require a high degree of fluency.

The total number of words passively recalled during the post-test differed between the text-only condition (M=14.6, SD=5.14) and the virtual kinesthetic condition (M=10.8, SD=5.41). A two-sample t-test revealed differences between group means, with participants in the text-only condition significantly outperforming kinesthetic learners in VR (p<0.05). However, one week later, the amount of words still remembered by the kinesthetic learners (M=7.8, SD=5.38) and participants in the control condition (M=7.56, SD=5.31) were virtually the same (see Figure 1).

The percentage of words lost between the immediate and delayed post-tests between control (M=0.54 SD=0.27) and VR participants (M=0.28, SD=0.32) showed that those involved in kinesthetic training retained more words than those in the text-only condition. Two sample t-test shows these differences to be significant (p<0.05). It is worth nothing that this loss was not always positive, given 2 participants in the kinesthetic group actually remembered more words one week later than immediately after the experience.

Correlation was calculated between the amount of words a user remembered immediately after the test and and the number of words that are lost after a week showing virtually no correlation. No correlation was found between the test results and the amount of languages spoken or studied by the participant.

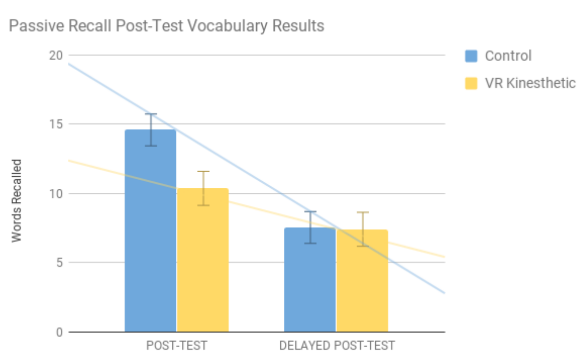

Figure 1. Number of words recalled by participants in both conditions immediately after exposure and one week after.

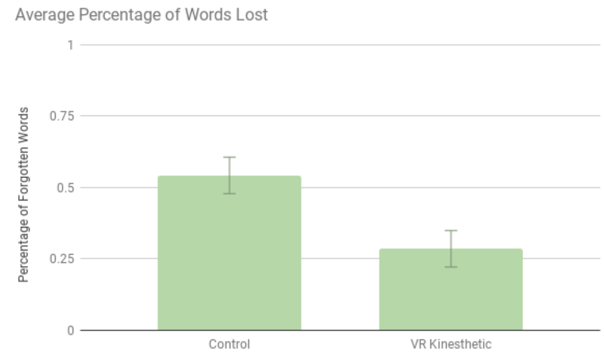

Figure 2. Number of words recalled by participants in both conditions immediately after exposure and one week after.

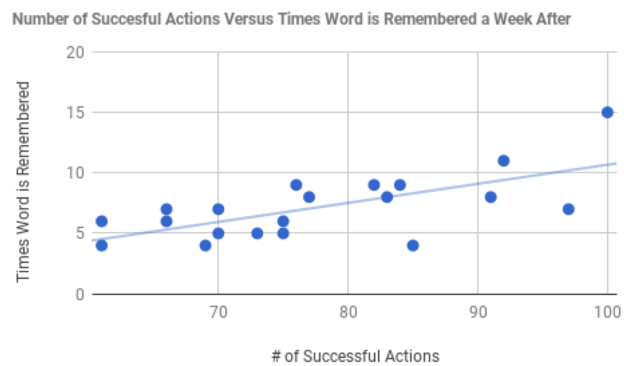

Figure 3. Number of words recalled by participants in both conditions immediately after exposure and one week after.

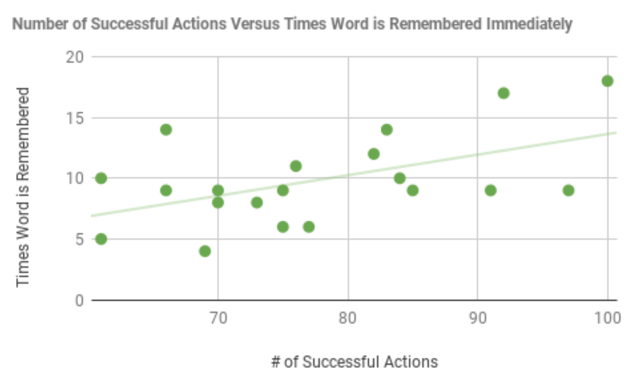

Figure 4. Number of words recalled by participants in both conditions immediately after exposure and one week after.

Pearson correlation analysis was performed on the metadata obtained from the virtual reality group. Due to minor issues with the telemetry collecting module, only metadata from 14 participants was analyzed as the data for 6 participants was partial or corrupted. This data included telemetry of the successful and failed attempts at performing the associated action. A moderate positive correlation was found between the amount of successfully performed actions and the times this word is remembered in both the immediate post-test and a week after (see Figure 3). This correlation is stronger a week later (p=0.67) than immediately after the training session (p=0.52). Analysis revealed no correlation between the number of failed attempts at performing a certain action with the amount of times the associated word is remembered in subsequent tests.

Subjective Metrics

Participants in the virtual reality condition completed a survey specifically addressing self reported engagement during the experience and perceived intuitiveness of the system. Users of the *Words in Motion* platform reported moderate to high engagement in a 5-point Likert scale (see Figure 6). Likewise, the system's intuitiveness was perceived moderate to high amongst participants in the virtual kinesthetic condition (see Figure 7).

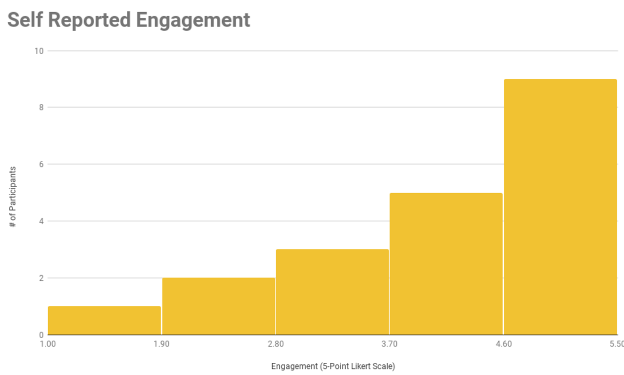

Figure 6. Histogram of self reported engagement on a 5-point Likert scale from subjects in the virtual kinesthetic condition.

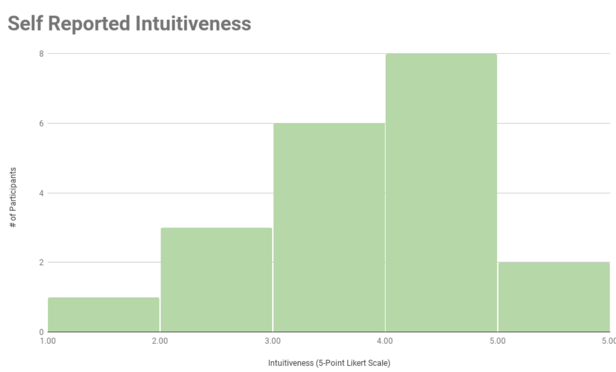

Figure 5. Number of words recalled by participants in both conditions immediately after exposure and one week after.

Open commentary from participants in both groups was collected regarding their experiences throughout the experimental procedure.

In the Virtual Reality settings several users pointed out that auditory feedback or hearing the word would have made the task easier for them. One subject pointed out that he was not very convinced by the experience, but was surprised a week after when he was able to remember so many words. Some participants did not understand what the actions were for, pointing out that they felt like "something else to memorize". They felt actions were distracting them from learning the words. Particularly, the same recurring theme from the kitchen scenario surfaced in open commentary: participants became confused or frustrated when the action they had to perform did not match their mental model of how that action would normally be executed by them. This separation or distance was described by subjects as distracting and causing their focus to shift towards the action instead of the words to be learned. Nevertheless, positive feedback related to the actions was also observed among subjects, where they pointed out "visualizing themselves performing the actions" to remember the words.

Participants in the text-only flashcard condition report the experience as not very engaging or memorable. Many felt it more like a cramming or memorization task, as opposed to a learning task given the lack of context. Nonetheless, they point out it is not "terrible" to learn the vocabulary words, commenting that although it wasn't the most interactive, they were able to remember words in the set.

ANALYSIS

Participants in the control case significantly outperform those who underwent training in the VR condition immediately after they are exposed to the learning experience. This result differs from prior literature, where kinesthetic learners often match or outperform non-kinesthetic methods[x]. Despite this initial advantage that the flash-card method has over the VR kinesthetic approach, one week after, there is no significant difference between the performance of both groups. This suggests that the VR kinesthetic condition is more memorable and helps participants retain a larger amount of words that they learned during the first exposure. We can see this when we compare the percentage of words lost between both tests, showing that the VR participants lose significantly less words. One may argue that learning more words initially means there are more words to be lost in the course of one week. However, no correlation between the amount of words learned and words forgotten was found in the data, and the percentage loss measure creates a relatively fairer comparison. Nevertheless, we acknowledge that subsequent testing at more intervals should be carried out in order to see if this trend continues over a longer period of time, albeit prior body of work has shown higher retention rates two[x] and three[x] weeks after exposure.

A study in virtual reality[x] that explored vocabulary acquisition among 18 participants in virtual environments versus 18 participants using the flash-card method, revealed trends almost identical to the ones presented in our work. They show a significant difference favoring the flashcard method immediately after exposure, no significant difference between methods one week after, and a significant difference in the retention rate favoring virtual reality. This experiment was reproduced at a later time, showing the same trends.

There are three main differences between this study and ours. First, we focused on second language acquisition using Spanish as the target language, whereas the other study taught participants in Swedish. Second, participants in the aforementioned study performed no actions to learn the vocabulary words other than pointing at the object in the virtual environment. Finally, the virtual scenario in which Ogma subjects learned was context heavy (a virtual house where objects were positioned naturally in the environment), as opposed to our study where context was limited to the object with which the action had to be performed.

Spanish and Swedish are two languages from different linguistic roots. Therefore, similarity in the results lightly suggests that the trends are not characteristic of the language that was used in the experiments but more related with the characteristics of virtual reality as a platform. Moreover, despite the discrepancies in context between both experiences, the trends remain similar. This similarity raises the question, how much of this effect is due to immersion in VR and how does the kinesthetic component play a role in the learning of the words?

The correlation analysis between the number of correct actions performed for a word and the amount of times it was remembered both immediately and one week after suggests a relevant effect from the kinesthetic nature of our platform. Namely, the more times you perform an action in VR, the more likely you are to remember the word associated with it. This correlation is even stronger after a week, supporting the notion that the kinesthetic component plays a stronger role in the retention of new vocabulary. This aligns with the embodied theory of language and results from previous experiments that compare kinesthetic versus non-kinesthetic methods. However, there is very little to say about the initial advantage flash-card methods hold over VR in both kinesthetic and non-kinesthetic approaches.

Authors of the Ogma platform attribute this to the familiarity students have with flash-card style learning, and the fact that immersive environment can be more distracting to the subject. The finding that this initial disadvantage occurs in an experiment where no physical action was asked of the participants, supports the notion that the kinesthetic component was not a defining factor that distracted the participants, but that there is a stronger component associated with immersive environments themselves. The fact that the number of failed attempts at performing an action has virtually no correlation (and more importantly no negative correlation) with the times the word is remembered in subsequent tests, further support this statement.

Although it happened only in two cases, participants in the virtual reality condition actually remembered more words the week after they were tested. This may very well be a fluke, but we note that participants in the Ogma experiment also came across two participants that showed the same results. This "fluke" seems to have only happened in virtual reality conditions in both *Ogma* and *Word in Motion* experiments, which might indicate an interesting effect for a small population of participants.

CONCLUSION

In this work we presented Words in Motion, a kinesthetic language learning platform for virtual reality that leverages the connection between body and mind to enhance second language vocabulary acquisition. This platform was used to create a learning activity for the classroom to be used in Kanda University in Tokyo as a new tool for English teachers to engage their students. The resulting activity involves students in a communication-heavy game where people outside VR instruct an immersed student using the target language to carry out a set of actions in a virtual kitchen scenario. The *Words in Motion* platform was used to carry out an experiment that compared the characteristics of virtual reality kinesthetic learning against the more common cramming flashcard method with 40 students recruited from campus.

Results showed that participants in the flashcard condition initially outperform virtual kinesthetic subjects for equal exposure time. Similar trends in prior body of work suggest this as characteristic of the immersive nature of VR, albeit not heavily influenced by the kinesthetic component of the *Words in Motion* platform. Although participants in kinesthetic and flashcard methods showed no significant difference a week after they were exposed to the material, the retention rate was significantly higher for subjects in the virtual kinesthetic condition. Moreover, the amount of times a word was remembered was directly correlated to the number of times the action associated with that word was performed both in immediate and delayed evaluation. In other words, performing actions in virtual reality has a positive effect in the retention of words when learning new vocabulary.

Despite the virtual kinesthetic experimental method and kitchen scenario show moderate to high levels of self-reported intuitiveness and engagement, a common concern emerged from open commentary and interviews with participants in both the experimental and learning activity trials. The discrepancy between the way an action is expected by the system and the user's prior conceptions of how an action is performed can collide, and subjects often report confusion or frustration when they do. What happens when there is a mismatch between the actions taught by the teacher and the student's mental models for those actions? We leave the question open, but suggest further research in this venue, suggesting these discrepancies may hold challenges for future kinesthetic education of language.

The findings in this paper support the hypothesis that virtual reality can benefit from explicit kinesthetic elements to enhance language learning activities. However, they also highlight that virtual reality kinesthetic learning is characteristically different from real-world and non-immersive technology enhanced kinesthetic learning, where exposure tends to result in higher immediate learning gains when compared to other audiovisual modalities. Nevertheless, virtual kinesthetic learning showcases the same enhanced retention effect as real-world and non-immersive analogues. Given the high levels of self reported engagement and the positive effect on vocabulary retention, this work suggests that with additional exposure and conditioning to the effect of "novelty", kinesthetic language learning in virtual reality can positively impact language education.

FUTURE WORK

The purpose of this paper was to present a platform that enabled kinesthetic language learning in virtual reality and establish a preliminary assessment of how kinesthetic elements impact learning within virtual environments. Given our findings, it would be worthwhile to understand how to condition subjects to the novelty of VR so as to reduce distractions that may deter learning gains from exposure. The same experiment should be replicated with a higher number of participants, and with delayed post-tests at multiple intervals to profile trends in the retention rate of the proposed method in comparison to other methodologies.

The kinesthetic platform presented in this paper, only takes advantage of gestural motions with the learner's hands. Additional tracking and inverse kinematics would allow a more complete kinesthetic experience that takes advantage of the whole body. Moreover, it would be interesting to devise embodied methods by which language learning could be enhanced to cover more complex elements of language that go beyond vocabulary by taking advantage of the body.

ACKNOWLEDGMENTS

This work was a collaboration between the MIT Global Languages and Studies, MIT Media Laboratory, and Kanda University of International Studies (KUIS), Japan. We would like to thank KUIS for its continuing support of this project. We also thank Sang-won Leigh for insightful discussions regarding the topics presented in this work.

REFERENCES

1. N D Leitan and G Murray. 2014. The mind-body relationship in psychotherapy: grounded cognition as an explanatory framework. Frontiers in psychology. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24904486
2. Lawrence A. Shapiro. 2011. Embodied cognition. Routledge/Taylor & Francis Group, London.
3. George Lakoff and Mark Johnson. 2011. Metaphors we live by:Univ. of Chicago Press, Chicago, Ill.
4. Masumoto, Kouhei, et al. "Reactivation of physical motor information in the memory of action events." Brain research 1101.1 (2006): 102-109.
5. Macedonia, M., and Knösche, T. R. (2011). Body in mind: how gestures empower foreign language learning. Mind Brain Educ. 5, 196–211. doi: 10.1111/j.1751-228X.2011.01129.x
6. Watson, C. E., Cardillo, E. R., Ianni, G. R., and Chatterjee, A. (2013). Action concepts in the brain: an activation likelihood estimation meta-analysis. J. Cogn. Neurosci. 25, 1191–1205. doi: 10.1162/jocn_a_00401
7. Fischer, M. H., and Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. Q. J. Exp. Psychol. 61, 825–850. doi: 10.1080/17470210701623605
8. Dudschig, C., De La Vega, I., and Kaup, B. (2014). Embodiment and second-language: automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop paradigm. Brain Lang. 132, 14–21. doi: 10.1016/j.bandl.2014.02.002
9. Rachel Moseley, Francesca Carota, Olaf Hauk, Bettina Mohr, Friedemann Pulvermüller; A Role for the Motor System in Binding Abstract Emotional Meaning, *Cerebral Cortex*, Volume 22, Issue 7, 1 July 2012, Pages 1634–1647, https://doi.org/10.1093/cercor/bhr238
10. De Grauwe, S., Willems, R. M., Rueschemeyer, S.-A., Lemhöfer, K., and Schriefers, H. (2014). Embodied language in first- and second-language speakers: neural correlates of processing motor verbs. Neuropsychologia 56, 334–349. doi: 10.1016/j.neuropsychologia.2014.02.003
11. Macedonia, Manuela. "Bringing back the body into the mind: gestures enhance word learning in foreign language." *Frontiers in psychology* 5 (2014).
12.