2018 IEEE 18th International Conference on Advanced Learning Technologies

Words in Motion: Kinesthetic Language Learning in Virtual Reality

Christian D. Vázquez, Lei Xia, Takako Aikawa, Pattie Maes
MIT Media Lab
Massachusetts Institute of Technology
Cambridge, MA USA
cdvm@mit.edu, leixia@mit.edu, taikawa@mit.edu, pattie@media.mit.edu

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 modalities. The tracking capabilities of room-scale virtual reality systems afford a way to incorporate kinesthetic learning in language education. We present Words in Motion, a virtual reality language learning system that reinforces associations between word-action pairs by recognizing a student's movements and presenting the corresponding name of the performed action in the target language. Results from a user study involving 57 participants suggest that the kinesthetic approach in virtual reality has less immediate learning gain in comparison to a text-only condition and no immediate difference with participants in a non-kinesthetic virtual reality condition. However, virtual kinesthetic learners showed significantly higher retention rates after a week of exposure than all other conditions and higher performance than non-kinesthetic virtual reality learners. Positive correlation between the times a word-action pair was executed and the times a word was remembered by the subjects, supports that virtual reality can impact language learning by leveraging kinesthetic elements.

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

I. Introduction

Language has been shown to be linked to our bodily experiences [14]. 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) [13]. 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 [8,17,19,26]. These effects hold even for words that are not directly linked to explicit action [22]. Moreover, studies support the notion that this relationship holds beyond our first language [4]. 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 [4].

The enactment effect [7] showed that performing a set of actions can generally increase the capability of subjects to recall these tasks. Given embodiment's notable effects on memory, its impact on second language learning has also been studied within the context of vocabulary acquisition [16,17,21,23]. Use of iconic gestures or illustrative motions have shown higher learning gain and retention than text and audiovisual modalities [11,20,25]. Although the body plays a significant role in language and learning, second language education is predominantly audiovisual [3, 9].

Virtual Reality (VR) has often been proposed as a platform that affords embodied learning. The tracking capabilities of room-scale virtual reality systems can be leveraged to understand a student's movements in space and provide real-time feedback. Moreover, due to VR's immersive nature, it can allow a student to engage in activities within novel contexts that strongly relate to their physical actions (e.g. learning vocabulary related to cooking while preparing a dish in a virtual kitchen), making VR a promising platform for kinesthetic language learning, i.e. learning that occurs through physical activity, and is deeply rooted in the relationship our mind has with our body.

In this paper, we present a virtual reality system for learning second languages kinesthetically with a focus on vocabulary acquisition. Our system, *Words in Motion*, 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 system is used in a user study with 57 students that 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.

II. Related Work

A large body of work has been dedicated to applications that involve kinesthetic elements to enhance language education. To guide our discussion, we divide these into three categories based on the relation between the kinesthetic elements and the learning material.

A. No Relation Between Body Action and Language

Systems in this category use physical action that is not related to the learning content and classifies more as gestural interaction with the material. Ogma [5] is a virtual reality system that uses the MYO armband to allow students to navigate a virtual space and learn new vocabulary words.

2161-377X/18/$31.00 ©2018 IEEE
DOI 10.1109/ICALT.2018.00069

Similarly, PILE [27] creates mixed reality environments that can be manipulated with physical gestures to engage in different aspects of language learning (e.g. moving hands to match pictures with words in the target language). JaJan! [1] takes a similar approach, but involves the presence of a co-located or remote peer to enhance the learning activities.

B. *Indirect Relation Between Body Action and Language*

This category is covered by projects that leverage action that is linked to language but is not strongly associated with the movement's relation to the linguistic component, i.e., there is a weak relationship between meaning and action. An example of this is WordOut [28], a museum exhibit in which children can use their body to match letter shapes as a literacy enhancing activity. Cheng et al. [29] presented a virtual reality game that incorporated body movement related to cultural components of language. The system primarily focused on bowing during social interactions with virtual avatars to enhance the learning of language and culture.

C. *Direct Relation Between Body Action and Language*

This category covers projects that teach language with a direct relationship between body action and meaning (e.g. action of cutting is directly associated with the word "cut"). *SpatialEase* was a game developed by Microsoft Research where the participants respond to audio cues in Mandarin with bodily actions [6]. *SpatialEase* was compared with Rosetta Stone software, showing similar gains for vocabulary and grammar acquisition. Similarly, Kuo et al. [12] developed a Kinect based platform to compare the effectiveness of computer enabled Total Physical Response (TPR) versus the classical TPR method, which involves students performing actions issued in a second language. 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. Similar approaches, making use of Kinect or alternate tracking systems, are presented in [2,10].

Although evidence suggests that experiences in the first two categories also benefit from kinesthetic encoding of knowledge, the experiences presented in this paper are concerned with technology enhanced kinesthetic activities that create direct associations between action and language. However, our platform leverages kinesthetic experiences in virtual reality, whereas existing platforms emphasize non-immersive analogues. Comparisons between computer enabled and classical TPR suggest that there is a difference in the learning gain, despite both activities requiring the same bodily motions [12]. 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.

III. WORDS IN MOTION

Words in Motion is a system developed using the HTC Vive that enables kinesthetic language learning in VR. The platform was developed as a recognition system that allows teachers to engage students in activities that introduce kinesthetic elements to second language instruction. In this section, we describe the different components of the system.

A. *Recognizing Actions*

Words in Motion augments objects in a virtual environment with verb-action pairs to reinforce associations between words and the body. Learners can perform actions within the virtual environment that trigger the corresponding name of the action to appear briefly before the student, thereby using the body as a tool to encode new vocabulary. Every object in the virtual environment is supported by a feed-forward neural net trained to recognize actions performed with it. 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-word pair with a degree of confidence, the corresponding word appears floating in front of the user (Figure 1).

B. *Training New Actions*

Teachers can enter the virtual environment to create new actions for students. In order to augment an object with new word-action pairs, a teacher can grab it from the environment and perform the new action multiple times to "teach" the neural net by example to recognize this action. The teacher can perform the action a variable number of times. Training a new verb-action pair for the object, creates a motion signifier (Figure 2). This signifier is a dynamically generated animation that displays the characteristic path that would trigger the word to appear. A virtual orb travels along this path to show students how to perform the new action.

C. *Learning Activities*

The system supports the creation of action lists or "challenges" for students. A challenge consists of a sequence of actions that must be executed by a user with a specific set of objects in the virtual context. A desktop interface allows

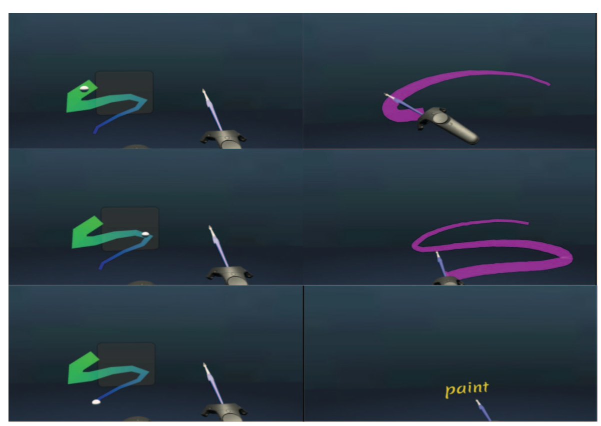

Figure 1. (Left) A motion signifier animation displays the action path to trigger the paired vocabulary word. (Right) A user performs the path to make the associated word appear.

teachers or students to create these challenges to engage in game-like activities.

IV. User Study

A study was conducted with 57 participants to assess the effectiveness of the kinesthetic method in virtual reality in comparison to a group which learned in a text-only condition outside of VR and a group that learned in VR with no kinesthetic component. Both text-only and VR kinesthetic groups consisted of 20 participants who were recruited from the university campus. The non-kinesthetic VR group consisted of 17 participants recruited from campus, as 3 participants dropped out before the final assessment.

A. Word Selection

Participants were exposed to a set of 20 transitive verbs in Spanish. The words were selected from Lifcach [24], a frequency list with over 450 million Spanish words. We selected 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.

B. Evaluation Method

The students were administered tests that required them to provide the English translation of a word after being prompted with the Spanish analogue. The tests 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 were encouraged to guess if a word seemed familiar.

C. Procedure

Subjects were randomly assigned to one of the three learning conditions: a text-only condition, a virtual non-kinesthetic condition, and the virtual kinesthetic condition supported by our system.

The text-only group, would sit and try to learn the words by observing the set of 20 word pairs that cycled on a computer monitor. Each word pair consisted of the target word in Spanish and its corresponding English translation. Each word pair was presented for 15 seconds, with each pair shown twice for a total of 30 seconds worth of exposure to each new vocabulary word.

The kinesthetic learners in virtual reality would be standing in an empty room. Participants were instructed to perform the movement depicted by a motion signifier (Figure 2) at least twice with their right hand. The first movement triggered the associated word pair to appear in front of the participant. 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. Each action was requested twice throughout the learning session, for a total of 30 seconds worth of exposure to each new vocabulary word.

The non-kinesthetic learners in VR would go through the same experience as the VR kinesthetic learners, but instead of performing the actions, they would see the actions performed by the system.

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.

V. Results

Participants from all groups scored low on the pre-test, with the text-only (M=0.65, SD=1.55), virtual non-kinesthetic (M=0, SD=0), and the virtual kinesthetic approach (M=0.25, SD=0.55) knowing almost none of the words prior to the experiment.

The total number of words passively recalled during the immediate post-test differed between the text-only condition (M=14.6, SD=5.14), virtual non-kinesthetic (M=9.41, SD=4.47), and the virtual kinesthetic condition (M=10.8, SD=5.41). A t-test revealed differences between group means, with participants in the text-only condition significantly outperforming non-kinesthetic ($p=0.002$) and kinesthetic learners ($p=0.03$) in VR. 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).

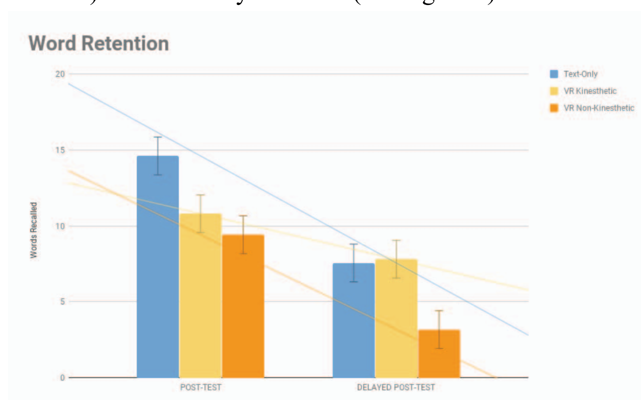

Figure 2. Words recalled by participants in all conditions.

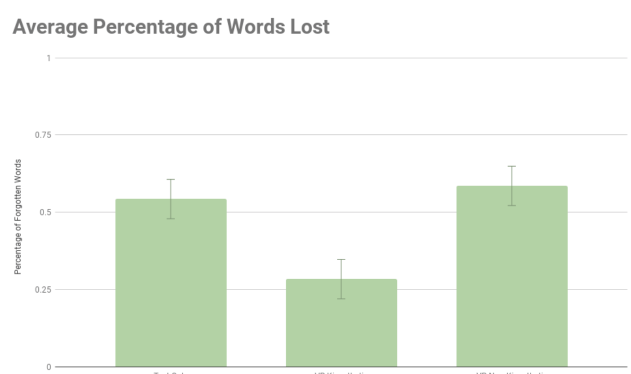

Figure 3. Percentage of words lost between immediate and delayed post-test show higher retention for VR.

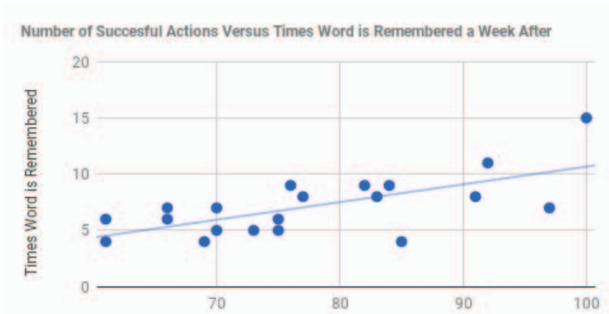

Figure 4. Times a word was remembered one week after exposure versus the times the action was performed.

Virtual non-kinesthetic learners (M=3.18, SD=3.08) performed significantly lower than text-only (p=0.008) and virtual kinesthetic (p=0.002) groups a week after.

The percentage of words lost between the immediate and delayed post-tests between text-only (M=0.54 SD=0.27), VR non-kinesthetic (M=0.59, SD=0.34), and VR kinesthetic (M=0.28, SD=0.32) participants showed that those involved in kinesthetic training retained more words than those in the text-only (p=0.006) and non-kinesthetic virtual conditions (p=0.01). No significant difference was found between the retention rate of text-only and non-kinesthetic virtual conditions (p = 0.79).

Correlation analysis was performed on the metadata obtained from the VR kinesthetic 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 an 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).

VI. Discussion

Participants in the control case significantly outperformed those who underwent training in both VR conditions 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 [13,22]. These results suggest that participants in VR are initially hindered or distracted by the novelty of the system. Given both kinesthetic and non-kinesthetic VR groups are outperformed by the text-only group, we can assume that the kinesthetic component is not the confounding element. 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.

A study in a VR platform called Ogma [5] that explored vocabulary acquisition among participants in VR versus subjects using flashcards, revealed trends almost identical to the ones presented in our work. There are two 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. Authors of the Ogma platform also 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.

Despite the initial advantage that the text-only method has over the VR kinesthetic approach, one week after, there is no significant difference between the performance of both groups. However, there is a significant difference in the amount of words recalled between participants who engaged in kinesthetic modalities in VR and those who did not. 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 a subject performed an action in VR, the more likely that subject will remember the word associated to it. This correlation is stronger after a week, supporting the notion that the kinesthetic component plays a strong role in the retention of vocabulary, and explains why participants performing actions in VR remember more words than those in non-kinesthetic interactions. This aligns with the embodied theory of language and results from previous experiments that compare kinesthetic and non-kinesthetic methods.

Although VR kinesthetic subjects performed similarly to the text-only group a week after, the rate at which participants forgot the words they learned was significantly lower for VR kinesthetic learners and similar between participants in the text-only and VR non-kinesthetic conditions. In other words, VR kinesthetic experiences were more memorable and helped participants retain a larger number of words, despite any confounding elements that hindered their initial learning gain.

VII. Conclusion

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. The *Words in Motion* platform was used to carry out an experiment that compared the characteristics of virtual reality kinesthetic learning against non-kinesthetic virtual reality and text-only modalities with 57 students recruited from campus.

Results showed that participants in the text-only condition initially outperform virtual kinesthetic and virtual non-kinesthetic subjects for equal exposure time to the material. Similar trends in prior body of work suggest this is characteristic of the immersive nature of VR and its novelty as a tool for learning in comparison to traditional modalities.

Both VR groups performed similarly immediately after exposure. However, a week after exposure, subjects in the virtual kinesthetic group significantly outperformed those in the virtual non-kinesthetic group and showed no difference to participants in the text-only group. 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 evaluations. In other words, performing actions in VR has a positive effect on the retention of words when learning new vocabulary.

Although participants in kinesthetic and text-only groups 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. The fact that the retention rate between text-only and virtual non-kinesthetic are not different further supports the hypothesis that this effect is due to the kinesthetic component, and support the virtual kinesthetic modality as more effective for retaining new vocabulary words in comparison to non-kinesthetic VR and text-only modalities.

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. Nevertheless, virtual kinesthetic learning shows the same enhanced retention effect as real-world and non-immersive analogues. Given the positive effect on 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.

REFERENCES

[1] Aikawa and Hunger. (2013). Language Learning in a Shared Virtual Space. Proceedings of the Asian Conference on Society, Education and Technology 2013.

[2] Chao, Kuo-Jen, et al. "Embodied play to learn: Exploring Kinect-facilitated memory performance." British Journal of Educational Technology 44.5 (2013).

[3] Choo, L. E. E. B., Tan Ai Lin, D., and Pandian, A. (2012). Language learning approaches: a review of research on explicit and implicit learning in vocabulary acquisition. Procedia Soc. Behav. Sci. 55, 852–860. doi:10.1016/j.sbspro.2012.09.572

[4] 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

[5] Ebert, Dylan, Sanika Gupta, and Fillia Makedon. "Ogma: A Virtual Reality Language Acquisition System." Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2016.

[6] Edge, Darren, Kai-Yin Cheng, and Michael Whitney. "SpatialEase: learning language through body motion." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2013.

[7] Engelkamp, J., and Krumnacker, H. (1980). Imaginale und motorische Prozesse beim Behalten verbalen Materials. Z. Exp. Angew. Psychol. 28, 511–533.

[8] 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

[9] Graham, S., Santos, D., and Francis-Brophy, E. (2014). Teacher beliefs about listening in a foreign language. Teach. Teach. Educ. 40, 44–60. doi:10.1016/j.tate.2014.01.007

[10] Hwang, Wu-Yuin, et al. "Recognition-based physical response to facilitate EFL Learning." Journal of Educational Technology & Society 17.4 (2014): 432.

[11] Kelly, S. D., Mcdevitt, T., and Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. Lang. Cogn. Process. 24,313–334. doi: 10.1080/01690960802365567

[12] Kuo, Fan-Ray, et al. "The effects of Embodiment-based TPR approach on student English vocabulary learning achievement, retention and acceptance." Journal of King Saud University-Computer and Information Sciences 26.1 (2014): 63-70.

[13] Lakoff and Johnson. 2011. Metaphors we live by:Univ. of Chicago Press, Chicago, Ill.

[14] Lawrence A. Shapiro. 2011. Embodied cognition. Routledge/Taylor & Francis Group, London.

[15] Leitan and 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

[16] Macedonia, M. (2003). Sensorimotor Enhancing of Verbal Memory through "Voice Movement Icons" During Encoding of Foreign Language (German: Voice Movement Icons. Sensomotorische Encodierungsstrategie zur Steigerung der quantita- tiven und qualitativen Lerneffizienz bei Fremdsprachen). Ph.D. thesis, University of Salzburg, Salzburg.

[17] 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.x228X.2011.01129.x

[18] Macedonia, Manuela. "Bringing back the body into the mind: gestures enhance word learning in foreign language." Frontiers in psychology 5 (2014).

[19] Masumoto, Kouhei, et al. "Reactivation of physical motor information in the memory of action events." Brain research 1101.1 (2006): 102-109.

[20] Mayer, K. M., Yildiz, I. B., Macedonia, M., and Von Kriegstein, K. (2014). Motor and visual brain areas support foreign language word learning. Curr. Biol. (in press).

[21] Quinn-Allen, L. (1995). The effects of emblematic gestures on the development and access of mental representations of french expressions. Mod. Lang. J. 79, 521–529. doi: 10.1111/j.1540-4781.1995.tb05454.x

[22] 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

[23] Radonvilliers, C. F. L. D. (1768). De la ManieÌre d'Apprendre les Langues. Paris: Saillant.

[24] Sadowsky, Scott, & Martínez-Gamboa, Ricardo. 2012. LIFCACH 2.0: Word Frequency List of Chilean Spanish (Lista de Frecuencias de Palabras del Castellano de Chile), version 2.0. Zenodo.

[25] Tellier, M. (2008). The effect of gestures on second language memorisation by young children. Gesture 8, 219–235. doi: 10.1075/gest.8.2.06tel

[26] 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

[27] Yang, Jie Chi, Chih Hung Chen, and Ming Chang Jeng. "Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning." Computers & Education 55.3 (2010): 1346-1356.

[28] Yap, Kelly, et al. "Word out!: learning the alphabet through full body interactions." Proceedings of the 6th Augmented Human International Conference. ACM, 2015.

[29] Cheng, Alan, Lei Yang, and Erik Andersen. "Teaching Language and Culture with a Virtual Reality Game." Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.