Learning into the Wild: A Protocol for the Use of 360° Video for Foreign Language Learning

Claudia Repetto[1(✉)], Serena Germagnoli[1], Stefano Triberti[1], and Giuseppe Riva[1,2]

[1] Department of Psychology, Catholic University of Sacred Heart, L.go Gemelli 1, Milan, Italy
{claudia.repetto, stefano.triberti, giuseppe.riva}@unicatt.it, serena.germagnoli@gmail.com
[2] Applied Technology for Neuropsychology Laboratory, IRCCS Istituto Auxologico Italiano, Via Magnasco 2, Milan, Italy

Abstract. Learning a second language could be a boring task if accomplished by repeating bilingual words lists. Laboratory research demonstrated that second language learning is more efficient if the material is enriched by pictures or gestures during the encoding phase. Here we want to test the impact of 360° videos on foreign language learning. The 360° videos are spherical videos that allow a lifelike exploration of the environment, if experienced immersively (namely, by means of a Head Mounted Display). The protocol includes ten 360° videos representing natural landscapes, sport performances and adventures; each video has been enriched with a narrative that guides the subject's attention towards the relevant elements named in English (the second language). The goals are twofold: first, we want to investigate whether the videos fruition is able to promote language learning; second, we want to verify the acceptability and perceived utility of this technology by potential users.

Keywords: Virtual Reality · 360° video · Second language learning Embodied Cognition

1 Theoretical Background

Traditional cognitive models of human mind considered our brain as a computer manipulating abstract symbols that follow predefined formal rules [1, 2]. According to this view, superior mental processes, such as problem solving, comprehension, recall of memories, operate on representations detached from sensory-motor experiences. An alternative and increasingly more influential view is the Embodied Cognition approach, which encompasses slightly different theoretical accounts, though associated by a fundamental core position: according to this perspective, the Cartesian dichotomy between mind and body is deemed incorrect, since mental operations are strictly related to, and dependent from, our bodies [3, 4]. Embodied theories of cognition extended the boundaries of anatomical structures to which traditionally a specific function was assigned: the mind is no longer confined to the brain but also includes other body parts,

P. Cipresso et al. (Eds.): MindCare 2018, LNICST 253, pp. 56–63, 2018.
https://doi.org/10.1007/978-3-030-01093-5_8

such as hands, legs, eyes. Moreover, within the brain, the separation between primary areas, recruited for basic sensory and motor processing, and the associative areas, in which more complex processes take place, is not strictly defined anymore: actually, the distinction between low and high level processes drops down in favor of a more integrated model. This new model proposes an interplay that allows the recruitment of primary areas even during "superior" cognitive processes: therefore, according to this account, the neural structures involved in sensory, perceptual or motor areas are also active when we construct concepts, make inferences, recall memories and use language. Although the perceptual and sensory system and their influence on cognitive processing have been investigated [5–7], without a doubt the motor system assumed a very special role in this line of research. The importance of the motor system is well recognizable in Wilson's words [8], "*Cognition is for action. The function of the mind is to guide action, and cognitive mechanisms such as perception and memory must be understood in terms of their ultimate contribution to situation-appropriate behaviour*" (p. 626). Of note, the link between the motor and the cognitive systems is bidirectional: if cognition is for action it is also suggested that action influences cognition [9].

In this study we will focus on the effects of action towards language processing. The idea that language and motor system are not independent, nor free from reciprocal influences is not that recent: the Lieberman's Motor Theory of Speech Perception [10, 11] is one of the first theoretical proposals in this direction. In the last decades, however, a huge amount of experimental proofs demonstrated the crosstalk between the two systems. At a behavioral level, performing an action while processing language has specific consequences on the latter. For example, in the classical Action- sentence Compatibility Effect (ACE) paradigm [12], participants are required to make sensibility judgments on sentences that describe actions towards the body ("John opens the drawer") or away from the body ("John closes the drawer"). Crucially, the response is provided by making an arm movement either away or towards one's body; the effect is underlined by faster reaction times in providing the response when the action performed by the participant matches that described in the sentence.

Another way to use action in connection with language is through object manipulation. It has been demonstrated that children, who act out a just read sentence by means of toys representing the sentence components, perform better on a comprehension task compared to other children that only read the sentence [13–15]. It has been suggested that the manipulation task forces the child to connect words to particular objects and syntactic relations to actions, resulting in better comprehension of the sentence [16].

An important line of research in the investigation of the effects of action on language is dedicated to the study of second language learning (L2). In the next paragraph the topic will be explored by introducing the concept of enrichment and its beneficial effects on learning.

2 Enrichment in Foreign Language Learning

Studying a foreign language could be a boring task if the learner tries to memorizes bilingual words lists. However, the task could become more engaging and the performance more efficient if the words are enriched by other stimuli. Empirical studies have shown that multisensory enrichment promotes verbal learning [17]. The reason why enrichment enhances memory performance is accounted for by two influential cognitive theories: the Dual Code Theory [18] and the Level of Processing Theory [19]. According to the Dual Code Theory, the combination of visual and verbal information during the encoding phase yields better memory performance. If words are recalled, one of the two codes (either visual or verbal) is available. The Level of Processing Framework (LOP) predicts that memory outcome depends on the "degree of depth" achieved during information encoding. If verbal information (in this case, a word) is encoded in a shallow way (i.e., at the phonological level by only hearing it), the recall of that word will be less efficient than if it were processed deeply, at the semantic level. In this respect, enriching a word, either by adding sensorimotor information to it or by processing it semantically, can enhance its depth of processing, consequently making it more resistant to decay.

Enrichment can be accomplished in different ways: a flash card representing the word's semantics can accompany the novel word during learning. Additionally, the motor system can be recruited to enrich the written word. This is done by performing an action or a gesture connected to the word [20, 21]. Literature in action memory investigates the role of a self-performed action (SPT) in the recall of action sentences (usually short phrases composed of an action verb plus an object) compared to either observing the experimenter performing the action (EPT) or verbal encoding alone (VE). In experiments, participants learning in the SPT condition show better memory performance compared to those in the EPT and VE condition [22]. This indicates that the motor component has a greater impact on retention if the action or gesture is self-performed than if the action is only observed. The superiority of the SPT is often referred to as the "enactment effect" (EE) [23].

A parallel line of research investigates the enactment effect on word learning in a second language (L2). Studies in this field target the impact of different kinds of gestures in word retention. These studies compare the effect of gestures on memory for words to the effect of other enriching methods (e.g., with pictures). Tellier [24] trained 20 French pre-schoolers with eight English words associated with self-performed gestures and with pictures. Memory performance was better for words encoded with gestures than for those encoded with pictures. In a combined fMRI and behavioral study, Macedonia et al. [25] explored the impact of iconic versus meaningless gestures on L2 words in a sample of young adults. This experiment was designed to determine whether the memory enhancement comes from the motor trace created by the action/gesture used in the word's representation or from the gestures' semantics. If the benefit derives from the gestures' semantics, then words encoded with iconic gestures should be remembered better than those encoded with a meaningless gesture. On the other hand, if the motor component per se would impact memory, the two encoding conditions should yield the same results in memory performance. Behavioral data

indicated that better performance relies on the gestural semantics. This finding was confirmed in a study by Mayer et al. [26], where German subjects learned new words under different enrichment conditions: performing an iconic gesture, seeing an associated picture, copying the outline of the picture with the index finger, or no enrichment. After five days of training, participants recalled words encoded with iconic gestures better compared to words encoded in all the other experimental conditions.

In the present protocol we want to test the effect of another type of enrichment, that is the naturalistic exploration of a virtual environment. Virtual Reality is the ideal medium where a multisensory environment can be implemented. It allows the user to feel a high sense of presence within the environment thanks to its tridimensional graphics and, especially, the opportunities of exploration [27]. Previous studies already employed Virtual Reality in the field of second language acquisition [28]. Some decades ago, Rose and Billinghurst [29] set up a fully immersive VR application to teach Japanese prepositions. In their approach, learners were trained with a three-step protocol: first, they freely explored the virtual environment, wherein the objects could be touched in order to hear their name and color from a digitalized voice; second, they watched animation sequences and used vocal commands; third, the played a virtual game where they had to assemble stacks of blocks matching samples shapes presented. More recently, Amoia and collaborators [30] presented I-FLEG a serious game designed to interactively learn French as second language. The game is exploited within Second Life environment (http://secondlife.com/). Connecting to the Allegro Island the user enters a 3D environment populated by different houses, each of whom is a different learning unit. Exploring the houses, the learner can train lexical domains or grammar, and the content varies depending on the individual's proficiency. This game is highly sophisticated as it offers targeted feedbacks and modular learning units, but is not immersive, since it is not designed to be used with a HMD.

In this study we aim to employ enriched 360° videos displayed on the smartphone and experienced immersively by means of a cardboard, to improve second language learning in high school students.

3 Experimental Protocol

3.1 Material

Ten 360° videos have been selected, among those available on the web, with the purpose of catching the adolescents' attention. 360° videos are spherical videos recorded by means of special camera with omnidirectional lenses. During the playback, a user, after having worn a cardboard, can control the viewing direction by means of the head movement in a very realistic way (looking up one can see the sky/roof, looking down one can see the floor/ground, and if one wants to see what happens on the left/right, it is sufficient to turn the head accordingly, as in the real life situations). Of note, within 360° videos it is not possible to select the direction of the navigation, nor to interact with the elements of the environment.

Selected videos represent natural landscapes, sport performances, adventures and interior environments.

Each video was carefully inspected to identify the relevant elements included and the actions performed by the characters. Scenes were extracted that represented elements whose label are both at high and low frequency in the second language (Mean frequency = 168.87; sta. dev. = 316.48, according to COLFIS, Corpus e Lessico di Frequenza dell'Italiano Scritto [31]). High frequency words are supposed to be already known by the students, but still they are useful to make the user feel confident with the task and to improve his/her self-efficacy. After having selected the scenes, objects and actions displayed within them were listed and become the target words to be learned in the second language (namely English, for the Italian native speakers). Afterwards, a narrative was created that connected them into a coherent story. The narrative has the purpose to guide the user's attention and focus it towards the target elements named. An English mother tongue woman audio recorded the narratives, and the audio file was added to the video in the post-production phase, synchronizing the speech to the images displayed. Duration ranges from 30 s to 1 min and 30 s. In the end, the user playing the video can experience the environment as if (s)he was in a foreign country with a native speaker talking to him/her and illustrating the environment in which both are navigating.

3.2 Procedure

This study is targeted to high school students studying English as second language.

The protocol includes two sessions in the classroom during regular lessons, with one month of delay in the between. An outline of the research protocol is represented in Fig. 1.

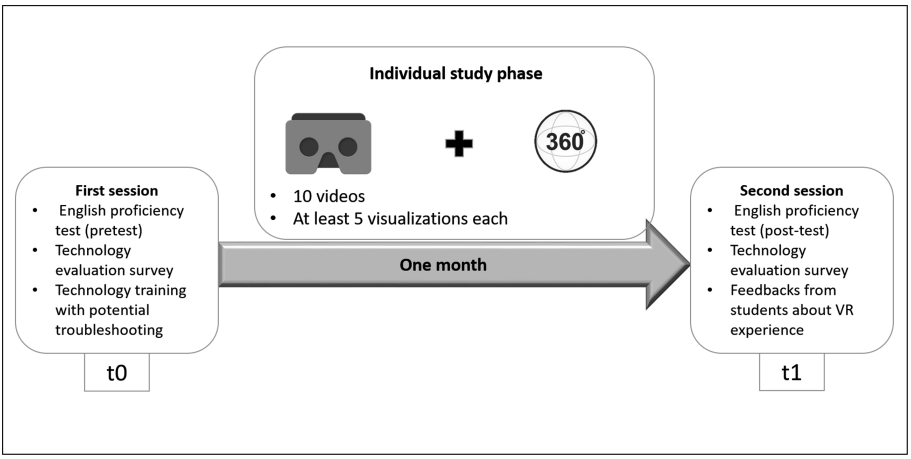

Fig. 1. The protocol of the research

First Session (T0) → During this session a series of pretests will be conducted. First, the baseline English knowledge will be assessed by presenting the students with the lists of item trained within the videos and asking them to translate from Italian into

Table 1. Items of the technology evaluation survey used during the first session.

1. I think that Virtual Reality will help me to improve my skills
2. I think that Virtual Reality-based training will improve my English proficiency
3. I think that Virtual Reality is useful
4. I think that learning to use Virtual Reality will be easy for me
5. I think it will be easy for me becoming a Virtual Reality expert
6. I think that Virtual Reality is easy to use
7. I'm going to use Virtual Reality regularly

English the words they know. Second, the attitude of the students towards the technology employed will be evaluated. To this purpose a brief survey will be administered, in which one has to claim the degree of agreement with the sentences presented. According to the Technology Acceptance Model [32], sentences cover two main topics: perceived ease of use and perceived utility. The complete list of sentences is reported in Table 1.

The session ends with a training illustrating the use of the cardboard, the videos' downloading and fruition by means of commercial Applications, and a potential troubleshooting phase to fix issues arisen with the personal mobile device.

Individual study → it lasts about 1 month, in which the students are requested to watch the videos individually, at home. A detailed program of visualization is provided, that combines the need to maximize the learning opportunities and the need to prevent side effects (each video should be watched at least five times, no more than 5 min of training are recommended, at least 2 h should pass between two consecutive trainings).

Second Session (T1) → During this session a series of post-tests will be conducted. First, memory performance for the trained words will be assessed: the students will undergo the same bilingual translation task administered in the First Session. Second, students will compile the technology evaluation survey claiming their opinion about the virtual experience and reporting potential side effects encountered. Free comments and suggestions will be collected for further tuning of the technologies or the protocol. Also, students will be asked to report, as an approximate percentage, how much of the expected program of visualization they think to have actually completed: this would allow to compare actual use with users' previous expectations and evaluations, according to the Technology Acceptance Model [32]. Third, students will be asked to evaluate the sense of presence they felt within each 360° video experience. It should be taken into account that 360° videos do not allow interaction except for visually exploring the environment, and interaction (or the perceived possibility for it) is considered one of the main factors impacting on sense of presence [27]. Anyway, the sense of presence is still considered one of the main aspects to monitor in order to evaluate the efficacy of immersive technologies [27]. In order to provide a "presence score" which would make available for researcher to compare single videos' effectiveness, students will be familiarized with the concept of presence by means of a short debriefing; then, they will be asked to rate the level of presence they experienced in each 360° video, as previously done in the literature [33].

4 Expected Results

This protocol will allow us to collect several information about the potential use of the 360° video for second language learning. First, by comparing the students' performance in the pre and post sessions we can test the efficacy of the protocol in terms of new words learned: we expect a significant improvement of the vocabulary after the training. More, we are interested in investigating the impact of the personal attitudes towards technology on the learning process: we expect that students having positive expectations about the ease of use and the utility of the videos will learn more efficiently, indicating that the technology is not "preferable" per se, but strongly interacts with motivation and beliefs. However, it is possible that the post sessions evaluations of the technology are better predictors of the learning performance, indicating that, despite the initial skepticism about the potential usefulness and ease of use, a positive experience is able to promote an efficient learning. Finally, by analyzing presence rates, it will be possible to identify a set of videos more likely to constitute an effective resource for second language learning.

Acknowledgments. This work was partially supported by Fondazione Cassa di Risparmio di Alessandria.

References

1. Fodor, J.A.: The Modularity of Mind. MIT press, Cambridge (1983)
2. Fodor, J.A.: The Language of Thought. Harvard University Press, Cambridge (1975)
3. Barsalou, L.W.: Grounded cognition. Annu. Rev. Psychol. **59**, 617–645 (2008)
4. Shapiro, L.A.: Embodied Cognition. Routledge, Abingdon (2011)
5. Martin, A., Haxby, J.V., Lalonde, F.M., Wiggs, C.L., Ungerleider, L.G.: Discrete cortical regions associated with knowledge of color and knowledge of action. Science (80-) **270**, 102–105 (1995)
6. Goldberg, R.F., Perfetti, C.A., Schneider, W.: Perceptual knowledge retrieval activates sensory brain regions. J. Neurosci. **26**, 4917–4921 (2006)
7. Thompson-Schill, S.L.: Neuroimaging studies of semantic memory: inferring 'how' from 'where'. Neuropsychologia **41**, 280–292 (2003)
8. Wilson, M.: Six views of embodied cognition. Psychon. Bull. Rev. **9**, 625–636 (2002)
9. Glenberg, A.M.: What memory is for. Behav. Brain Sci. **20**, 1–55 (1997)
10. Galantucci, B., Fowler, C.A., Turvey, M.T.: The motor theory of speech perception reviewed. Psychon. Bull. Rev. **13**, 361–377 (2006)
11. Liberman, A.M., Mattingly, I.G.: The motor theory of speech perception. Cognition **21**, 1–36 (1985)
12. Glenberg, A.M., Kaschak, M.P.: Grounding language in action. Psychon. Bull. Rev. **9**, 558–565 (2002)
13. Glenberg, A.M., Gutierrez, T., Levin, J.R., Japuntich, S., Kaschak, M.P.: Activity and imagined activity can enhance young children's reading comprehension. J. Educ. Psychol. **96**, 424–436 (2004)
14. Glenberg, A.M., Brown, M., Levin, J.R.: Enhancing comprehension in small reading groups using a manipulation strategy. Contemp. Educ. Psychol. **32**, 389–399 (2007)

15. Marley, S.C., Levin, J.R., Glenberg, A.M.: Improving Native American children's listening comprehension through concrete representations. Contemp. Educ. Psychol. **32**, 537–550 (2007)
16. Glenberg, A.M., Goldberg, A.B., Zhu, X.: Improving early reading comprehension using embodied CAI. Instr. Sci. **39**, 27–39 (2011)
17. Shams, L., Seitz, A.R.: Benefits of multisensory learning. Trends Cogn. Sci. **12**, 411–417 (2008)
18. Paivio, A., Csapo, K.: Concrete image and verbal memory codes. J. Exp. Psychol. **80**, 279 (1969)
19. Craik, F.I.M., Tulving, E.: Depth of processing and the retention of words in episodic memory. J. Exp. Psychol. Gen. **104**, 268–294 (1975)
20. Macedonia, M., von Kriegstein, K.: Gestures enhance foreign language learning. Biolinguistics **6**, 393–416 (2012)
21. Zimmer, H.D., et al.: Memory for Action: A Distinct for of Episodic Memory?. Oxford University Press, Oxford (2001)
22. Nilsson, L.G.: Remembering actions and words. In: Tulving, E., Craik, F.I.M. (eds.) The Oxford Handbook of Memory, pp. 137–148. Oxford University Press (2000)
23. Engelkamp, J.: Memory for Actions. Psychology Press/Taylor & Francis, Hove/Abingdon (1998)
24. Tellier, M.: The effect of gestures on second language memorisation by young children. Gesture **8**, 219–235 (2008)
25. Macedonia, M., Knösche, T.R.: Body in mind: how gestures empower foreign language learning. Mind Brain Educ. **5**, 196–211 (2011)
26. Mayer, K.M., Yildiz, I.B., Macedonia, M., Von Kriegstein, K.: Visual and motor cortices differentially support the translation of foreign language words. Curr. Biol. **25**, 530–535 (2015)
27. Triberti, S., Riva, G.: Being present in action: a theoretical model about the 'interlocking' between intentions and environmental affordances. Front. Psychol. **6**, 2052 (2016)
28. Schwienhorst, K.: The state of VR: a meta-analysis of virtual reality tools in second language acquisition. Comput. Assist. Lang. Learn. **15**, 221–239 (2002)
29. Rose, H., Billinghurst, M.: Zengo Sayu: an immersive educational environment for learning Japanese (1995)
30. Amoia, M., et al.: A serious game for second language acquisition in a virtual environment. J. Syst. Cybern. Inform. **10**, 24–34 (2012)
31. Bertinetto, P., et al.: Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS) (2005). http://linguistica.sns.it/CoLFIS/Home.htm
32. Davis, F.D.: Perceived usefulness, perceived ease of use and user acceptance information technology. MIS Q. **13**, 319–339 (1989)
33. IJsselsteijn, W., de Riddera, H., Hamberga, R., Bouwhuisa, D., Freeman, J.: Perceived depth and the feeling of presence in 3DTV. Displays **18**, 207–214 (1998)