In: *Proceedings of the 13th International Conference on Intelligent Virtual Agents (IVA 2013)*, **PREPRINT**
The final publication is available at link.springer.com

A Virtual Agent as Vocabulary Trainer: Iconic Gestures Help to Improve Learners' Memory Performance

Kirsten Bergmann[1] and Manuela Macedonia[2,3]

[1]SFB 673, CITEC, Faculty of Technology, Bielefeld University, Germany
[2]Department for Information Systems, University of Linz, Austria
[3]Research Group "Neural Mechanisms of Human Communication", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
kirsten.bergmann@uni-bielefeld.de
manuela.macedonia@jku.at

Abstract. An important and often laborious task in foreign language acquisition is vocabulary learning. Research has repeatedly demonstrated that performing iconic gestures together with novel words has a beneficial effect on learning performance. Can these findings be transferred onto virtual agents applied in gesture-supported vocabulary training? We present a study investigating whether iconic gestures performed by a virtual agent and imitated by learners have an impact on verbal memory for words in a foreign language. In a within-subject design we compared participants' memory performance achieved with the help of a virtual agent and those achieved with the help of a human trainer regarding both short-term learning effects and long-term decay effects. The overall results demonstrate improved memory scores when participants learned with a virtual agent. Especially high performers could profit from gesture-supported training with a virtual agent.

Keywords: Vocabulary acquisition, iconic gestures, memory performance

1 Introduction

A major challenge in learning foreign languages is the acquisition of novel words. Generations of language learners have struggled with written or spoken materials, e.g., in form of word pair lists. Nowadays, multimodal learning has become more and more popular so that learning materials are enriched by pictures, videos or music to make the acquisition of new words easier. There is also evidence for the beneficial role of gestures in vocabulary acquisition: Performing an iconic gesture, i.e., a gesture whose physical form corresponds with object features [13], together with a novel word, enables faster learning and makes those words more resistant against forgetting [4, 9, 11, 12, 17, 18, for details see section 2.1].

It stands to reason that these findings are of great importance for foreign language pedagogics, but so far they have only barely found their way into class rooms and teaching materials. Among the possibilities to train vocabulary items with gestures, a new generation of trainers has to be considered: embodied conversational agents. The embodied nature of virtual agents allows for multimodal communication. Nonverbal communication skills comprise gestural behavior, but also other behaviors like facial expressions. Both factors play an important role in learning environments [1]. Moreover, virtual trainers are characterized by a high degree of flexibility. In terms of personalization an agent's communicative behaviors, specifically the use of gestures, can easily be adapted to particular needs and preferences of the learner. In addition, unlike a human teacher or tutor, a virtual character is always available and supportive.

Although there is an increasing number of studies dealing with the effects of virtual agents in foreign language learning, nonverbal communication plays only a minor role in agent-supported learning (cf. [6]). With particular regard to gestures, only pointing gestures have been considered so far [1, 14]. And in these studies results with regard to memory performance diverge (for details see section 2.2). Our study presented here investigates whether iconic gestures performed by a virtual agent and imitated by learners have an impact on verbal memory for words in a foreign language. Furthermore, the aim of this study is to compare memory results achieved with the help of a virtual agent and those achieved with the help of a human trainer.

2 Background and Related Work

2.1 Gestures have an Impact on Verbal Memory

A growing body of scientific evidence has demonstrated the so-called *enactment effect* [3] for more than 30 years: action words like 'to go' or phrases like 'play piano' are better recalled if they are performed than if learners listen to them or imagine to perform them. Going beyond verbal memory in native language, several studies have shown that gestures can also have an impact on memory for words and phrases in a foreign language [10]. The first systematic study on this topic was conducted by Quinn-Allen [17]. English-speaking students learned French expressions that simultaneously were accompanied by emblematic gestures. Better recall were achieved as compared to a control group. The study also demonstrated a beneficial long-term effect of gestures on information decay: Eleven weeks after encoding, gesture performers had forgotten less than control subjects. Other studies replicated and extended these findings. Macedonia [11] investigated single word retention for 36 items belonging to different word categories. In a within-subject design, participants learned words nouns, adjectives, verbs and prepositions of an artificial language corpus either by reading and listening to them or by additionally performing gestures. Results showed that retrieval (cued recall) was significantly better in the short- and long-term for enacted items.

Tellier [18] presented common words like house, swim, cry etc. to French children learning English. Half of the items were associated with a picture and the other half were illustrated by a gesture that the children saw in a video and thereafter performed. Enacted items were better memorized than items enriched visually by the pictures. Moreover, Macedonia and Knösche [12] demonstrated that enactment enhances memory performance not only for concrete, but for also for abstract words (nouns, verbs, and adverbs). In an additional transfer test, participants of this study were asked to produce new (non-canonical) sentences with the words they had learned during the training. Enacted items were recruited significantly more often than words learned audio-visually.

Other studies investigated whether the type of gestures performed while learning has an effect on word learning. Kelly et al. [4] trained young adults on Japanese verbs conveying common everyday meanings. The words were presented according to four modes: (i) speech, (ii) speech + congruent gesture, (iii) speech + incongruent gesture, and (iv) repeated speech. The results showed that participants memorized the largest number of words in the speech + congruent gesture mode, followed by the repeated speech mode.The least number of words was memorized when they were accompanied by an incongruent gesture. Another study controlling for the type of gestures was conducted by Macedonia et al. [9]. Subjects were trained on 92 concrete nouns. Half of the items were encoded with iconic gestures. They depicted some aspect of each words semantics and enriched the word with a plausible sensorimotor connotation. The other half of the items was learned with meaningless gestures. The results showed better memory performance for iconic gestures than for meaningless gestures in the short- and long-term (after 60 days).

2.2 Virtual Agents: Effects on Memory Performance

Due to the sparse number of experiments on this topic, the question whether agents have an impact on the memory performance of learners is still a matter of debate and we expand our view to virtual teachers, trainers or tutors in general learning contexts here. Beun and colleagues [2] compared two types of virtual agents (realistic and cartoon) with a no-agent condition in a text comprehension task. Subjects were provided with a short story they had to re-tell directly after the stimulus presentation. The presence of an embodied agent had a positive effect on the retainability of information. On the contrary, Moundridou and Virvou [15] did not find a beneficial effect of the presence of a virtual agent when students were solving math problems. Similarly, Mulken et al. [16] investigated whether the presence of a virtual agent was helpful for participants recall rate of technical material (information about machines) as well as non-technical material (introduction of fictitious people). Researchers compared an agent using deictic gestures with a pointer stick against a no-agent condition in which only the pointer sticks were visible. Recall tests directly after stimulus presentation showed that there was neither a positive nor a negative effect on recall.

In contrast to the aforementioned studies, Mitsako et al. [14] evaluated the effect of a gesticulating virtual agent as vocabulary trainer on learning performance in *repeated* interactions. German speaking participants learned English words either with or without a virtual agent in four sessions over eight days. In the no-agent condition the words were displayed on a screen and spoken by a text-to-speech system, whereas in the agent condition, a female agent was additionally present. She featured some idle movements and pointing gestures referring to the words to be learned. The presence of the agent had no effect on learning performance ('persona-zero' effect). Altogether, previous related studies revealed rather controversial effects of virtual agents on learners memory performance. This holds for the mere presence of embodied agents as well for use of nonverbal behavior. Nevertheless, given that existing implementations of nonverbal behavior are often by far not as natural as human behavior, researchers are still convinced that nonverbal communication is "a huge potential for agent systems" [6, p. 80]. Given that experiments have only tested the impact of deictic gestures on learning behaviors, iconic gestures still remain a field to be explored. In fact, it has been repeatedly demonstrated that they support verbal memory (see section 2.1). Another issue that has not been evaluated sufficiently are effects of virtual agents on long term memory performance. Mulken et al. [16] explicitly point to this and speculate that other effects might be found on a long term performance.

3 Study

We assess the effect of iconic gestures performed by a virtual agent compared to gestures performed by a human on memory performance for words in a foreign language. Given the embodied nature of virtual agents and their capabilities of expression, we expect that gesture-supported vocabulary training with a virtual agent will be as successful as training with equivalent gestures presented by a human. Second, we investigate both short-term effects of learning and long-term effects of information decay. From research on video stimuli of humans for the same purpose [4, 9, 11, 12, 17, 18] we expect that gesture-based training with a virtual agent will also be beneficial for both.

The study employed a within-subject design manipulating the type of training: (1) Gesture-based training with video stimuli of a human (human gestures; HG), (2) Gesture-based training with video stimuli of a virtual agent (agent gestures; AG), and (3) a control condition without any gestures (Con). Participants trained vocabularies on three consecutive days. Their (short-term) learning performance was measured the next day prior to the next training session, respectively (day1, day2, day3). The long-term effect of information decay was measured additionally four weeks after training was finished (day30).

3.1 Participants

A total of 32 native speakers of German, aged from 20 to 40 (M=25.3, SD=3.5), participated in the study. 15 participants were female and 17 participants were male. All of them were either paid or received credits for their participation. Subjects were randomly assigned to two training groups to counterbalance training conditions and items.

3.2 Materials

The training material comprised 45 nouns in 'Vimmi', an artificial corpus created for experimental purposes [9, 12]. It aims to avoid associations and to control for different factors that, in natural languages, can favor the memorization of particular vocabulary items. The Vimmi items were created according to Italian phonotactic rules. First they were randomly generated by Perl and thereafter adjusted to avoid tautological occurrence of syllables, high frequency of particular consonants or vowels, the appearance of strings sounding unusual to German-speaking subjects, association with words from European languages taught at school (English, French, Italian, and Spanish), and with proper nouns comprising names of products available on the German market. The artificial words were assigned common meanings like bridge and suitcase. Word familiarity was controlled for using the word frequency counter of German[1]. The 45 words were recorded in 45 single audio files, with each file having a length of approximately 0.8s.

[1] http:// Wortschatz.Uni-Leipzig.de

(a) Gesture for the word 'staircase'

(b) Gesture for the word 'mirror'

Fig. 1. Example stimuli performed by the virtual agent and human actress.

The gestures mirrored some feature of the semantics of the word that was arbitrarily chosen from a range of possibilities discussed in the team. For instance, for the word 'staircase', the actress modeled several steps of different height on after another (figure 1a). For the human gesture stimuli we video-taped a human actress. The agent's gestures were modeled to match the human gestures as much as possible. They were specified in the Multimodal Utterance Representation Markup Language (MURML; [7]) and realized with the Articulated Communicator Engine (ACE; [5]), a toolkit for building animated embodied agents that are able to generate human-like multimodal utterances. The agent's gestures were also rendered into video data. For examples of both human and agent gesture stimuli see figure 1.

3.3 Procedure

Participants were informed that they took part in an experiment on foreign language learning with the goal to memorize as many words as possible. Participants also were informed that their performance would be assessed at different time points through different kinds of written tests. The training lasted approximately 45min per day on three consecutive days. Training was performed in groups. The total of 45 words were subdivided into three blocks of 15 words each, in which the three training conditions daily alternated and counterbalanced the experimental conditions. In each block the items were randomly subdivided into three smaller blocks of five items each. A block was first shown and participants were instructed to watch it. Thereafter, the block was played again six times and participants were cued to imitate the gesture and/or to repeat the word in Vimmi after seeing and hearing it. All the words were randomized within the blocks. In total, every vocabulary item was presented seven times every day. After each training block, a break of five minutes followed. The software used for the training was Presentation[2].

[2] v16.3 by Neurobehavioral Systems, Albany, CA

3.4 Tests

Memory performance was assessed daily starting from the second experiment day measuring the learning outcome of the first training day etc. (day 1-3) and on day 30. Participants were administered a free and thereafter a cued recall test. In the free recall test participants were provided with an empty sheet. They were instructed to write as many items as possible in both languages. Items could be loose (i.e., only German or only Vimmi) or matched (i.e., Vimmi and German). In the cued recall test participants were given a randomized list of the 45 trained items to be translated from German into Vimmi and then a further randomized list of the same words to be translated from Vimmi into German (duration 5 min), with the instructions to translate the items from one language into the other. The order of the translation from one into the other language alternated daily. Items were considered correct if their spelling corresponded 100% to the word spelling provided during training (score of 1). Partial correctness was only considered in case of minor mistakes, e.g., interchanged letters like 'asemo-aseno' for the Vimmi words or nominalization effects like 'to pipe-pipe' in German. Here a score of 0.5 was given. All other items were considered wrong and given a score of 0.

4 Results

Data from 29 participants was analyzed for both short-term and long-term effects. Three participants had to be removed from the data set. This was due to data storage problems in two cases. In one case it turned out that the participant did not fit into the group of young adults we aimed to investigate with our study.

4.1 Short-term effects

In order to assess the influence of training on memory performance in free and cued recall respectively, a repeated-measures ANOVA with the within-subject factors TIME (day1, day2, day3) and TRAINING (HG, AG, Con) was conducted. Means and standard deviations are summarized in table 1.

Table 1. Short term training results for the aggregated measure of free and cued recall: Mean scores of proportional recall and standard deviations in parentheses.

		HG	AG	Con
Free Recall	day1	15.56 (6.64)	18.05 (9.23)	11.92 (7.65)
	day2	33.10 (12.02)	34.79 (16.02)	28.24 (16.21)
	day3	46.09 (18.11)	46.59 (17.38)	43.98 (19.04)
Cued Recall	day1	9.66 (11.13)	11.78 (13.77)	13.16 (11.24)
	day2	41.44 (25.05)	43.10 (27.80)	41.03 (28.66)
	day3	60.98 (25.35)	63.28 (25.69)	60.11 (29.67)

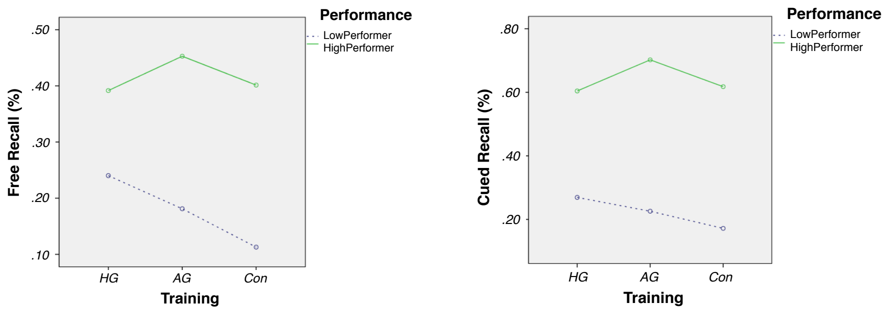
Free Recall For the aggregated measure of free recall Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of TIME (χ^2=20.45, p<.001). Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ϵ=.65). There was a significant main effect of TIME one free recall rate (F(2,56)=187.26, p<.001). Contrasts revealed that memory performance increased significantly from day2 to day3 (F(1,28)=160.23, p<.001). More interestingly, there was also a main effect of TRAINING one free recall rate at a significant level (F(2,56)=4.24, p=.019). Memory performance was significantly better for words trained with the agent gestures (AG) as compared to words that had been trained in the control condition (F(1,28)=6.25, p=.019).

Cued Recall For the aggregated measure of free recall Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of TIME (χ^2=19.55, p<.001). Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ϵ=.66). There was a significant main effect of TIME on free recall rate (F(2,56)=139.69, p<.001). Contrasts revealed that memory performance increased significantly from day2 to day3 (F(1,28)=146.79, p<.001).

4.2 Long-term effects

For the analysis of long-term effects on memory, we considered another variable, namely learning performance across learners in vocabulary. We subdivided participants in high and low performers. This we determined by a median-split of the group. The average retrieval performance for the 29 participants over the four time points showed a mean value of 34.57% (SD 15.24). The median value of 30.6% split the group into two subgroups: 13 low performers (4 females and 9 males) with a mean performance of 21.20% and 16 high performers (10 females and 6 males) with a mean performance of 45.43%. In order to assess the influence of training on memory performance in free and cued recall respectively, a mixed design ANOVA with the within-subject factor TRAINING (HG, AG, Con) and the between-subject factor PERFORMANCE (high performers, low performers) was conducted.

Free Recall For the aggregated measure of free recall the main effect of TRAINING was significant (F(2,54)=3.68, p=.032). Compared to the control condition the amount of correctly recalled items was significantly higher in both the training condition with human gestures (F(1,27)=4.72, p=.039) as well as the training condition with agent gestures (F(1,27)=9.17, p=.005). Not surprisingly, there was also a significant effect of PERFORMANCE, indicating that high performers had a significantly enhanced memory performance as compared to low performers (F(1,27)=19.39, p<.001). We also found a significant interaction effect between TRAINING and PERFORMANCE (F(2,54)=4.37, p=.017) indicating that the different training conditions affected low and high performers differently. In order to break down this interaction, we compared the different training groups across high and low performers. This analysis revealed significant interactions when between the gesture-based training with human stimuli and the control condition across high and low performing participants (F(1,27)=6.41, p=.017). The graph in figure 2 plotting the above results suggests that human gestures were helpful to low performers while the supporting effect was lower for high performers. Interestingly, the gestures performed by the agent had an impact on the high performing group.



(a) Interaction of TRAINING and PERFORMANCE in free recall.

(b) Interaction of TRAINING and PERFORMANCE in cued recall.

Fig. 2. Interaction effects of TRAINING and PERFORMANCE after 30 days.

Cued Recall In this measure we found a significant main effect of TRAINING (F(2,54)=3.49, p=.038). Gesture-based training with the agent led to significantly better retrieval than the audiovisual training (F(1,27)=5.90, p=.022). There was also a significant main effect of PERFORMANCE indicating that high performers had a significantly enhanced memory performance as compared to low performers (F(1,27)=58.06, $p<.001$). In addition, there was a significant interaction effect between TRAINING and PERFORMANCE (F(2,54)=3.99, p=.024) indicating that memory scores achieved in the different training conditions were not the same for low and high performers. Again, contrasts were used to break down this interaction. They revealed significant interactions between high and low performers' training with human gestures and the control condition (F(1,27)=4.19, p=.05). The graph in figure 2 shows that the human gestures were more helpful to low performing participants than for high performers. By contrast, the agent's gestures had an impact and were more helpful for the high performing group.

5 Discussion and Conclusions

The present study had two goals: first to test the impact of gestural training provided by a virtual agent on memory for words in a foreign language. Second, this study compared memory performance after a training with an agent and memory performance after a training with a human. We predicted that gestural training would enhance memory for words and that training provided by an human and by an agent would not differently affect memory performance. From a virtual agent perspective this hypothesis could actually be exceeded: For most of the tests we employed, training with a virtual agent led to better memory performance than training with a human. For both types of long-term measures (free and cued recall) memory performance for items learned in the AG condition outperformed memory performance for items learned in the control condition. The same effect was present for short-term measures of free recall. However, training with human gestures did not reach significance in all tests which contradicts empirical evidence of multiple studies (see Sect. 2.1). In terms of an explanation, we speculate that participants of our study only trained for three days. In other studies with the same material training spanned over a longer period of time [9, 12]. Nevertheless, in absence of significant benefits of human gestures, the beneficial effects of a virtual agents gestures is even more striking.

To further elucidate the effect of agent-based training being more successful than human-based training, we took participants overall performance into account for the long term effects. Interestingly, we found that agent-based training was particularly successful for high-performers. For low-performers the human-based training resulted in higher memory scores. How can we explain this difference? Here we assume that the virtual agents gestures were less natural than the video recordings of the human actress. This difference in naturalness might bring a hindered processing about. Observers of these gestures might need a few more moments of thinking to clarify what exactly the gesture means, how it relates to the word to be learned and how to imitate it. It may be that this additional thinking further consolidates the information in high performers. For low performers, on the contrary, this further processing might be too much so that they do not benefit from the agents gestures as much as from natural human gestures. Leutner et al. [8, p.83] raise cognitive load theory in this context: “If an agent behaves in a way that is perceived as somewhat strange for the learners, this thinking about strangeness requires cognitive resources, and consequently imposes extraneous cognitive load in working memory”. Interpreted in the light of this context our results suggest that high performers, who have stronger language acquisition skills [9], are in a better position to handle this additional load, and they even seem to profit from it.

In future work we aim to further pursue this explanation by collecting ratings of naturalness and the degree of iconicity for the different gesture stimuli employed in our study. Moreover, as we could show in the present study that virtual agents using iconic gestures are actually supportive for vocabulary training, we can think of more comprehensive language training tools based on embodied characters. A virtual trainer permanently at the users disposal on mobile devices can offer individualized training, not only in second language education but also in other domains like language rehabilitation.

Acknowledgements

This research is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center 673 “Alignment in Communication” and the Center of Excellence 277 “Cognitive Interaction Technology” (CITEC).

References

1. Baylor, A.L., Kim, S.: Designing nonverbal communication for pedagogical agents: When less is more. Computers in Human Behavior 25, 450–457 (2009)
2. Beun, R., Vos, E., Witteman, C.: Embodied conversational agents: Effects on memory performance and anthropomorphisation. In: Proceedings of IVA 2003. pp. 315—319. Springer, Berlin/Heidelberg (2003)
3. Engelkamp, J., Krumnacker, H.: Imaginale und motorische Prozesse beim Behalten verbalen Materials. Zeitschrift für experimentelle und angewandte Psychologie pp. 511–533 (1980)
4. Kelly, S.D., McDevitt, T., Esch, M.: Brief training with co-speech gesture lends a hand to word learning in a foreign language. Language and Cognitive Processes 24, 313—334 (2009)
5. Kopp, S., Wachsmuth, I.: Synthesizing multimodal utterances for conversational agents. Computer Animation and Virtual Worlds 15, 39–52 (2004)
6. Krämer, N., Bente, G.: Personalizing e-learning. The social effects of pedagogical agents. Educational Psycholoy Review 22, 71–87 (2010)
7. Kranstedt, A., Kopp, S., Wachsmuth, I.: MURML: A multimodal utterance representation markup language for conversational agents. In: AAMAS'02 Workshop Embodied conversational agents-Let's specify and evaluate them! (2002)
8. Leutner, D., Wirth, J.: Commentary on: Personalizing e-learning. The social effects of pedagogical agents. Educational Psycholoy Review 22, 71–87 (2010)
9. Macedonia, M., Müller, K., Friederici, A.: Neural correlates of high performance in foreign language vocabulary learning. Mind, Brain and Education 4(3), 125—134 (2010)
10. Macedonia, M., von Kriegstein, K.: Gestures enhance foreign language learning. Biolinguistics 6(3-4), 393—416 (2012)
11. Macedonia, M.: Voice Movement Icons' during encoding of foreign language. Ph.D. thesis, Salzburg University (2003)
12. Macedonia, M., Knösche, T.R.: Body in mind: How gestures empower foreign language learning. Mind, Brain, and Education 5, 196–211 (2011)
13. McNeill, D.: Hand and Mind—What Gestures Reveal about Thought. University of Chicago Press, Chicago (1992)
14. Miksatko, J., Kipp, K., Kipp, M.: The persona zero-effect: Evaluating virtual character benefits on a learning task. In: Allbeck et al., J. (ed.) Proceedings of the 10th International Conference on Intelligent Virtual Agents. pp. 475–481. Springer, Berlin/Heidelberg (2010)
15. Moundridou, M., Virvou, M.: Evaluating the persona effect of an interface agent in a tutoring system. Journal of Computer Assisted Learning 18, 253—261 (2002)
16. Mulken, S., André, E., Müller, J.: The persona effect: How substantial is it? People and Computers pp. 53—66 (1998)
17. Quinn-Allen, L.: The effects of emblematic gestures on the development and access of mental representations of french expressions. The Modern Language Journal 79, 521—529 (1995)
18. Tellier, M.: The effect of gestures on second language memorisation by young children. Gesture 8, 219—235 (2008)