
Imitation of a pedagogical agent's gestures enhances memory for words in second language

Manuela Macedonia^{1,2,*}, Kirsten Bergmann³, Friedrich Roithmayr¹

¹Department for Information Systems, Johannes Kepler University, Linz, Austria

²Research Group Neural Mechanisms of Human Communication, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

³SFB 673, CITEC, Bielefeld University, Germany

Email address:

manuela@macedonia.at (M. Macedonia)

To cite this article:

Manuela Macedonia, Kirsten Bergmann, Friedrich Roithmayr. Imitation of a Pedagogical Agent's Gestures Enhances Memory for Words in Second Language. *Science Journal of Education*. Vol. 2, No. 5, 2014, pp. 162-169. doi: 10.11648/j.sjedu.20140205.15

Abstract: Pedagogical agents (PAs) are virtual characters in computer-based learning environments. PAs can train humans in various domains. Here, a PA cues subjects to learn vocabulary items through enactment, i.e., to perform an illustrative gesture while learning a word. It is well known that enactment impacts memory. Also, imitation is a natural mechanism driving learning. Combining both enactment and imitation could improve memory even more. In a within-subjects study, 44 school children learned 45 vocabulary items according to three conditions: an audio-visual baseline, an observation condition (participants watched the PA during enactment) and an imitation condition (participants imitated the PA's gestures). We documented learning progress by cued recall tests. Over four days, we found that, compared to the baseline and to mere observation, imitation of enactment significantly enhanced memory for words in the foreign language.

Keywords: Pedagogical Agent, Multilingualism, Imitation, Enactment, Gestures, Memory

1. Introduction

Technology is pervading many domains of our lives. Pedagogical agents (PAs) residing in mobile devices may soon become our personal trainers and complement human teachers for certain tasks in learning, for example, vocabulary learning [1]. PAs are embodied virtual characters with an anthropomorphic appearance and multimodal communication skills: They can enunciate words and sentences as well as perform gestures accompanying them, produce facial expressions in line with the communicative content, etc. [2].

However, findings regarding the question of whether employing PAs is beneficial for the users' learning outcome have been inconclusive so far [3]. For the particular case of language learning, some studies demonstrate that pedagogical agents improve learning, while in other studies they do not. Positive effects were shown, for example, in an experiment by Beun and colleagues [4]. They investigated the retention of a short story that was presented by an embodied anthropomorphic agent vs. a cartoon figure. Subjects remembered the story better when it was presented by the embodied agent. A study by Miksatko and colleagues

[5] examined memory performance for English words. Participants learned vocabulary items over eight days with or without an embodied agent. The vocabulary items were displayed on a screen and either enunciated by a text-to-speech system or spoken by an embodied agent that additionally pointed to them with a stick. In this setting, memory performance for the two learning conditions did not differ; the presence of the agent had no effect on learning ("persona-zero" effect). In a recent study on vocabulary learning, Bergmann and Macedonia [6] employed the virtual agent Billie (Figure 1), a child-like character aged 10 or 11 years. He (pardon the personification) enunciated words and performed meaningful gestures accompanying the word to be learned. The authors of the study had him train adults on 45 vocabulary items of Vimmi, an artificial corpus created for experimental purposes [7]. In that study, the main question was whether humans can learn by means of enactment with a virtual agent as well as with a human trainer. In fact, it has been repeatedly demonstrated that enactment leads to better memory for words and phrases than traditional learning in both native [8] and foreign languages [9]. In their within-subjects study, Bergmann and Macedonia [6] subdivided the vocabulary items into three blocks of 15

words. They were trained by videos according to the following conditions: 1) baseline, i.e., audio-visual training (reading the word and hearing it) 2) baseline plus enactment with the agent and 3) baseline plus enactment with a human trainer (2 and 3 were done by means of videos). The overall results showed an enhancement of audio-visual learning through enactment. Most interestingly, there was no significant difference in vocabulary retention for items enacted with the agent and those with the human trainer. A follow-up study investigated the acceptance of the agent Billie as compared to the human trainer during the presentation of vocabulary items by enactment [10]. Naïve subjects were asked to rate the quality of gestures and several attributes of personality (e.g., friendliness, intelligence, competence) for both trainers, as previously done in a study by Fiske and colleagues [11]. Subjects were presented videos showing the trainers enunciating the word in the foreign language and performing a semantically related gesture. The quality of human gestures was rated as significantly higher than the agent's gestures, but the human and the agent did not differ in terms of perceived personality. A similar study on the acceptance of the agent Billie was conducted with 25 children between 10 and 11 [12]. They rated Billie and a human trainer (a boy of the same age) executing the gestures and enunciating the words. Like in the adult study, the children rated the quality of the human gestures higher than the robotic gestures produced by the agent, and again the human trainer and the agent did not show significantly different ratings of personality for the features investigated. In these two studies, adults and children saw significant difference(s) between the human and the PA when enacting words in a foreign language, but in both cases they showed a similar degree of acceptance for the virtual agent and the human.



Figure 1. Billie, the virtual agent used for this experiment, performing the gesture for “mug” (screenshots from the video).

2. Learning a Second Language with the Body

2.1. Enactment and the Self-Performed Task Effect

Enactment research began in the early 1980s. In their seminal study, Engelkamp and Krumnacker [13] found that actions have an impact on memory for verbal information. They cued subjects to perform actions representing phrases such as “cut the bread”. In a recognition task after learning,

subjects could recognize the phrases better if they had performed the action than if they had only listened to or imagined them. Similarly, Cohen [14] made his subjects learn action phrases either by self-performing or by reading them. In free recall tests, subjects showed enhanced memory performance if they had enacted the phrases. At that time, Cohen coined the term “subject-performed task” (SPT). Accordingly, the effect of gestures on memory for verbal information was called SPT effect. In the 1980s, a considerable number of experiments were conducted with different memory tests and populations that confirmed the SPT effect. For a review see [8].

A main issue in the field is still whether watching the experimenter who performs the action (experimenter-performed task, EPT) leads to the same memory results as a self-performed task (SPT). This issue was already taken up in early enactment research. Whereas Cohen [14] did not find significant differences in memory performance between SPT and EPT, Engelkamp and colleagues demonstrated that SPT leads to superior memory performance in recognition tasks [15, 16]. This controversy was explained later in terms of the quantity of items that had been learned [8]. For short lists (12 items) SPT and EPT do not produce different effects on memory. For longer lists of items, between 24 and 48, SPT lead to better memory results. However, this topic has not been satisfactorily investigated, so that more behavioral research can contribute to clarity in the field.

Results from neuroscientific experiments in the past decade show that self-performing a gesture when learning verbal information leads to the formation of sensorimotor networks that represent and store the words in native [17] and foreign language [7]. Neuroscience provides evidence for the original proposal by Engelkamp that motor information is crucial for memory enhancement for verbal information [18, 19]. Hence if learners do not perform the motor act, they should not form the motor trace in their brains. However, Stefan and colleagues have proven that mere observation of action also leads to the formation of motor memories in the primary motor cortex [20]. Hence the equivalence in memory performance between SPT and EPT as found by Cohen [14] in behavioral experiments could be explained by the possible engagement of the same neural mechanisms involved in both observation and imitation.

2.2. Imitation Grounds Learning

Infants are the best examples of learning through imitation: they simply watch adults and learn how to do things. Children do this without having been instructed to do so [21]. Likewise, animals learn by imitation [22, 23]; this is not only the case for primates, as learning by imitation has been observed in birds [24] and fishes [25]. The basis of modern research on human imitation was laid by Meltzoff and colleagues' study [26] on manual and facial gestures by babies between 12 and 21 days old. In that study, the authors showed that imitation was already present at a very early stage in life and not, as supposed by other developmental psychologists such as Piaget, starting from the age of 8 months [27]. Meltzoff and Moore

[28] replicated the original study a few years later with newborns (mean age 72 hours). Both studies suggested that imitation is not a learned behavior but instead innate. However, in those days the neural mechanisms of imitation were still unknown. They were discovered nearly two decades later in the lab of Giacomo Rizzolatti, an Italian neurophysiologist. The team was investigating grasping actions in macaque monkeys by means of deep electrodes implanted in their motor cortex. They registered the activity of single neurons involved in action control when the monkeys reached for food given to them by the experimenters. Accidentally, the team observed that the neurons involved in grasping became active when the animal reached for the food, but also when the monkey observed the experimenter doing it [29]. The scientists came to the conclusion that these neurons mirror others' actions, hence the term mirror neurons. Despite controversy concerning their function, a large body of evidence has demonstrated over the years that mirror neuron circuits in the brain contribute to learning [30]. Besides action learning, they also subserve the acquisition of social behavior and other cognitive capacities and, most importantly for the present study, language [31]. Interestingly, neurons with mirror functions are also located in different regions of the human brain, including Broca's area, the canonical region for language production. Not surprisingly, motherese, the first form of linguistic interaction between infants and adults [32], involves mirror neuron circuits [33]. We reason that integrating imitation in a language training procedure might be supported by innate learning mechanisms and lead to better results than simply instructing learners to listen to verbal information and to read it.

2.3. Language and Gestures Belong Together

Another issue supporting the view that training by means of enactment is more efficient is the tight connection between gestures and language [34, 35]. From an evolutionary point of view, there are a few prominent theories postulating that language developed from gestures [36]. The evolution of spoken language might have gone through different stages gradually: from only gestures and simple vocalization initially, over the combined use of manual and vocal communication, to the production of speech [37]. Furthermore, gestures open the way to the acquisition of first language [38]. In the pre-verbal phase, children communicate by deictic and iconic gestures. Pointing is a well investigated precursor to words in a number of communicative situations between children and care-givers [39]. Also, when adults speak, they tend to gesture in order to express their thoughts in a more comprehensible way and to enhance access to their cognitive resources [40]. Speech and gesture interact in order to enhance communication [41].

3. The Present Study

Considering the results described in the introduction and the issues in the preceding sections, we aim to elucidate in detail gesture-supported training with PAs. In this vein, the present study has two major objectives. First, we seek to expand the

experimental setting already realized for adult participants to school children. They are the most important target audience for supportive technologies in foreign language training. Second, we address the question of whether it suffices for learners to only watch the agent presenting the words or to additionally imitate him by performing the gestures themselves. This question is important for the potential use of PAs as vocabulary trainers. Imagine a class with 25 to 30 children sitting quietly in their chairs, simply watching the agent saying the words to be learned and performing the related gestures. This would be easily feasible. Imagine instead cueing the same group of children to perform the gestures. This could be much more laborious, not only because some learners would not like to perform gestures, but also because certain gestures require locomotion in the room. Children simply watching an agent would be easier to realize, while gesture imitation is supported by arguments on imitation learning. Also, evidence from behavioral experiments conducted so far is contradictory [14,15,16] and more research is needed.

Considering the arguments above and the possible applications of PAs in formal instruction, we tested vocabulary training with the agent Billie. In particular, we aimed to explore the following questions:

- a) Can memory performance for words be enhanced if learners watch the PA performing illustrative gestures to the word, compared to an audio-visual baseline (reading and hearing the word)?
- b) Does imitation of the PA enacting words enhance memory for words compared to audio-visual learning and to mere observation of his gestures?
- c) We worked with school children and trained them in groups in a classroom instead of individually in a lab. From a descriptive point of view, we wonder if this is feasible and whether learners accept our novel way of training vocabulary.

3.1. Method

3.1.1. Participants

Three school classes, altogether 74 school children (mean age 11.2, SD 0.92, 28 females, 46 males) took part in the experiment. They were subdivided into three training groups of 24, 24 and 25 subjects. None of the participants had a prior report of psychiatric or neurological diseases. All parents of the children gave written consent. The experiment was approved by the local Ethics Committee. Subjects were naïve about the aims of the experiment but were informed that they would participate in an experiment on foreign language word learning. The experimenter and three research assistants were present in the classroom in which the experiment took place. They were to ensure that participants attended to the stimuli and did not omit any part(s) of the training, i.e., word repetition or gesture imitation. As in an experiment by Macedonia and Klimesch [42] in which participants were in the same class, we expected high interaction among subjects. This can bias results and should be prevented.

On the first day of training, despite the presence of the

research team, participants often did not fulfill the required tasks. On the second day, we decided to exclude a whole class, 25 subjects. They joked, did not attend to the stimuli and disturbed each other during the stimulation. Also, when asked to imitate, subjects did not do so. On the following days, we excluded single subjects for the same reasons or because they did not show up for one phase of the training or testing. We ended up with two groups and 33 subjects (mean age 11.6, SD 0.87, 18 females, 15 males). We had started with three groups and 74 subjects.

Table 1. Words used in Vimmi for the training with their translation(s) into German. (The English translation is merely for the article's readers and was not used during training).

Item #	Vimmi	German	English Translation
1	motila	Briefmarke	stamp
2	gelori	Ohrring	earring
3	miruwe	Pfeffermühle	pepper mill
4	lutepa	Petersilie	parsley
5	mebeti	Becher	mug
6	atesi	Treppe	stairs
7	lofisu	Föhn	hair dryer
8	elebo	Flöte	flute
9	siroba	Seife	soap
10	suneri	Geige	violin
11	wiboda	Welle	wave
12	batewo	Zettel	sheet
13	nalefa	Messer	knife
14	sokitu	Krawatte	tie
15	nibesa	Knopf	button
16	sigule	Tempel	temple
17	bekoni	Kaffee	coffee
18	pirumo	Erde	earth
19	giketa	Blume	flower
20	magosa	Shampoo	shampoo
21	uladi	Pullover	pullover
22	nabita	Spülmittel	dishwashing liquid
23	giwupo	Kürbis	pumpkin
24	mesako	Radiergummi	eraser
25	dafipo	Blech	plate
26	gaboki	Spiegel	mirror
27	maloti	Gebiss	denture
28	dotewe	Schwamm	sponge
29	dubeki	Parfüm	perfume
30	mofibu	Fernbedienung	remote control
31	puneri	Handtuch	towel
32	botufe	Taschentuch	handkerchief
33	gasima	Poster	poster
34	nelosi	Reißverschluss	zip
35	gepesa	Besen	broom
36	wugezi	Regal	shelf
37	nowitu	Telefon	telephone
38	guriwe	Faden	thread
39	lamube	Fieber	fever
40	lifawo	Stuhl	chair
41	wasute	Säge	saw
42	gubame	Brücke	bridge
43	serawo	Gießkanne	watering can
44	mewima	Stempel	seal
45	asemo	Krücke	crutch

3.1.2. Experimental Environment

Because we intended to make predictions for foreign language instruction, we opted to train subjects in a classroom in their school instead of in a lab. The school board provided a larger room of about 40 square meters. Tables and chairs were moved to the walls of the classroom. Subjects could stand in the middle of the room and had about a square meter each at their disposal. They could perform the gestures easily when required by the training condition. The groups of subjects came to the training room for four days, one hour daily and thereafter returned to their classrooms.

3.1.3. Training Materials

We used 45 words in Vimmi [7], an artificial corpus used for experiments in vocabulary retention and created in order to avoid associations to words known to the participants (Table 1). Vimmi conforms with Italian phonotactics; hence words sound Italian but they are not. Words were generated by Perl, a programming language used for text manipulation [43], and controlled for word length. All items consisted of three syllables. We assigned each Vimmi item an arbitrary meaning in German. German words were concrete nouns and were controlled for their frequency of use (<http://wortschatz.uni-leipzig.de>). For each item to be learned, the training materials consisted of the written word in Vimmi, its translation into German, the audio file in Vimmi (approximately 0.8 s) and a video (5 s) with the agent performing the corresponding gesture.

3.1.4. Gestures Used

All gestures were performed by the agent Billie. We modelled the agent's gestures according to videos with a human actress executing them. We used the AsapRealizer [44], a framework for behavior synthesis in virtual humans that receives multimodal behavior specifications in the Behavior Markup Language [45] as input. We rendered the agent's gestures into video data. The gestures embodied by the agent were based on arbitrarily chosen feature(s) of the word's semantics (such as object shape) or functions of usage (for additional information see [46]).

3.1.5. Training Procedure

We subdivided the 45 words of Vimmi into three blocks of 15 words each. We then assigned the blocks of 15 items to three learning condition:

- 1) Audio-Visual (AV): Participants were presented the word in written form, heard the corresponding audio file and saw a static picture of the agent;
- 2) Gesture Observation (GO): Participants were presented the word in written form, heard the corresponding audio file, saw the agent performing the corresponding gesture, and were cued only to watch.
- 3) Gesture Imitation (GI): Participants were presented the word in written form, heard the corresponding audio file, saw the agent performing the corresponding gesture, and were cued to repeat the word aloud and to imitate the gesture performed by the agent.

For stimulation, we used the software Presentation (<https://www.neurobs.com>). During stimulation, we randomized the order of presentation of the items within the blocks and grouped them into three smaller blocks of five items each. Every block was presented seven times daily. Learning conditions for the words changed depending on the group (Table 2).

Participants were trained according to the training schedule (Table 3) for four days, from Monday to Thursday. The three training conditions alternated and counterbalanced the experimental conditions daily. Testing occurred on Friday only. We excluded group 1 from the training for reasons explained above.

Table 2. Distribution of the words and the learning condition(s) for group 2 (a) and 3 (b), respectively.

a)				
Day 01	Day 02	Day 03	Day 04	Day 05
welcoming 15 min	test (day 01) 20 min	test (day 02) 20 min	test (day 03) 20 min	test (day 04) 20 min
audio-visual 20 min	gesture observation 20 min	gesture imitation 20 min	audio-visual 20 min	-
break 5 min	break 5 min	break 5 min	break 5 min	-
gesture observation 20 min	gesture imitation 20 min	audio-visual 20 min	gesture observation 20 min	-
break 5 min	break 5 min	break 5 min	break 5 min	-
gesture imitation 20 min	audio-visual 20 min	gesture observation 20 min	gesture imitation 20 min	-
b)				
Day 01	Day 02	Day 03	Day 04	Day 05
welcoming 15 min	test (day 01) 20 min	test (day 02) 20 min	test (day 03) 20 min	test (day 04) 20 min
gesture imitation 20 min	gesture observation 20 min	audio-visual 20 min	gesture imitation 20 min	-
break 5 min	break 5 min	break 5 min	break 5 min	-
gesture observation 20 min	audio-visual 20 min	gesture imitation 20 min	gesture observation 20 min	-
break 5 min	break 5 min	break 5 min	break 5 min	-
audio-visual 20 min	gesture imitation 20 min	gesture observation 20 min	audio-visual 20 min	-

Table 3. Training schedule for group 2 and 3 (group 1 was excluded from training on the 2nd day).

	Words 1-15	Words 16-30	Words 30-45
Group 2	gesture observation	audio-visual	gesture imitation
Group 3	gesture imitation	gesture observation	audio-visual

3.1.6. Testing

Starting from day 2, we assessed daily memory performance by means of cued recall tests from the native (German) into the foreign language (Vimmi) and vice-versa. Participants were given a sheet of paper with a randomized list of words to be translated into the other language. We instructed participants to translate as many words as they could. Each test lasted 10 minutes and the order of translation changed daily. We considered an item as correct if its spelling corresponded to the spelling provided during training (score 1). In Vimmi, for sonorous consonants we made following exception: German natives, because of a phonological rule of German, devoiced sonorous consonants such as [g] and pronounce them as voiceless [k] at the word onset and the

word end. In this case a wrong spelling was not considered as an error. In German, we considered spelling always correct, even if subjects made spelling mistakes common for this school level. For example, a common mistake is [ss] instead of [ß]. Note that in German both [ss] and [ß] have the same sound; however, their use is idiosyncratic and needs to be learned.

4. Results

We averaged memory results for the three conditions through the training time (days 02-05) for the 33 subjects that completed the experiment (Figure 2). In the translation from German (the subject’s native language) into Vimmi, subjects learned slightly better by observing the gestures and by imitating them compared to audio-visual input. However, the variance was high and the differences among conditions were not significant. In the translation from Vimmi into German, the graphics show that subjects benefitted from observing the gesture, and most of all by imitating them, although the difference between these two conditions is not significant.

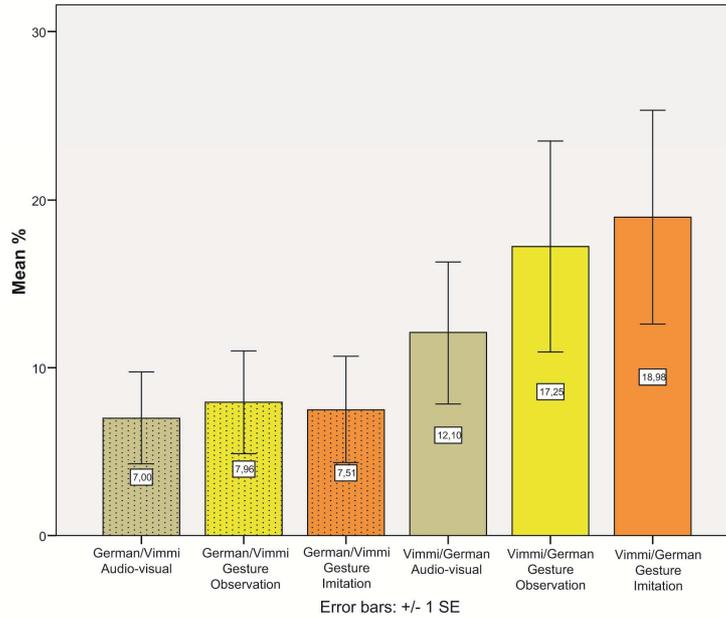


Figure 2. Memory performance in the cued recall test subdivided by translation direction and training.

In order to assess the influence of the agent’s gestures and the imitation of the agent’s gestures on memory’s performance, we computed a 3x2 repeated measures ANOVA for the 33 subjects who completed the training and were present in all four testing phases. In the ANOVA, we considered the factors TRAINING (audio-visual, observation and imitation) and LANGUAGE (German into Vimmi and Vimmi into German). The study yielded significant results for both LANGUAGE and TRAINING, $F(1,3) = 23,79$ $p < 0.05$ and $F(2,6) = 9,21$ $p < 0.05$, respectively. The interaction between LANGUAGE and TRAINING was not significant.

We further aggregated both translation directions into three conditions and plotted the graphics in Figure 3 to indicate

overall memory results related to the training condition.

Despite the poor performance and the high subject dropout, best memory results were achieved by imitation compared to the other learning conditions. The difference between audio-visual learning and gesture observation was slight. In the repeated measures ANOVA we found a significant main effect for the factor TRAINING $F(2,6) = 6,23$ $p < 0.05$. Note that best memory results (both cued recall tests aggregated) were on average only about 20% of the learning target, i.e., about 9 words out of 45 after 4 hours of training. Compared with other studies with adult subjects [7, 47, 48], school children performed poorly.

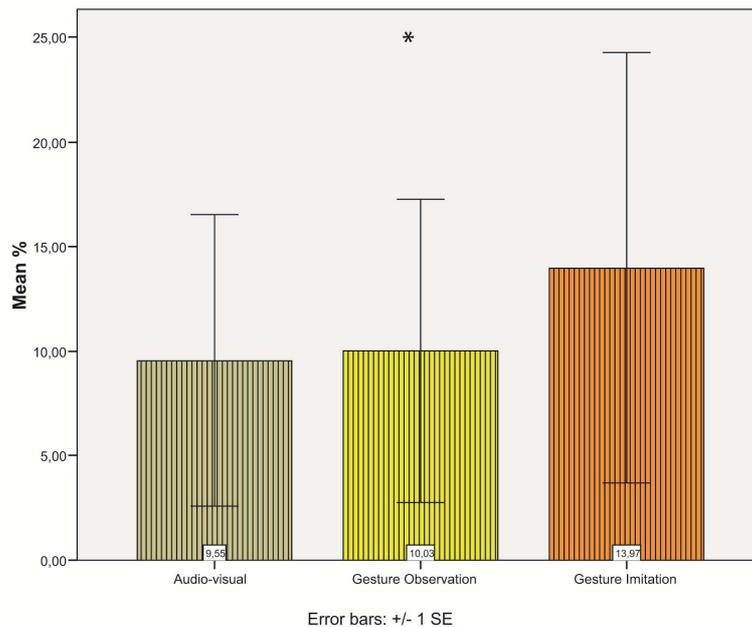


Figure 3. Memory performance subdivided by training procedure after aggregation of translation direction.

5. Discussion and Conclusion

This experiment was designed to assess whether school children can be successfully trained in L2 vocabulary by a PA. Our main research question was whether observing semantically related gestures performed by the PA is as efficient as imitation, i.e., self-performance of those gestures. Our participants learned Vimmi items under three conditions: audio-visually, by watching a PA performing gestures related to the words' semantics, and through imitation of the gestures produced by the PA. The data displayed in the preceding section provide the following answers to our research questions.

First, compared to an audio-visual baseline, memory for words is enhanced if subjects observe and/or imitate gestures performed by an agent. However, our data show that imitation of a PA performing gestures is the training that leads to best memory results for single words in L2. Thus our results support the SPT hypothesis [15] and not the EPT hypothesis [14]. However, the statistical significance documenting the superiority of imitation (SPT) was not given in all analyses and particularly the results of the cued recall tests from German into Vimmi do not confirm the hypothesis (Figure 2). We can only speculate about the reasons for this. It is possible that in this translation direction results might be connected with the difficulty of the task. If subjects would get more intensive training, either for a longer session daily or for longer than four days, subjects might achieve better results. This was observed in other studies [7, 47] in which subjects started to benefit of enactment on the 3rd day of training. However the training in those studies lasted much longer than in the present one (2 to 3 hours daily for 5 days and not only 1 hour daily for 4 days).

The school children were trained together in a classroom. This is feasible, but it was a challenge for the experimenters' team to maintain discipline and to make sure that participants attended to the stimuli and followed the instructions. We imagine that teachers working with the class regularly might achieve better interaction. Also, participants were naïve about the goals of the study. This might have influenced their motivation to cooperate as they might not have seen any sense in performing gestures when learning words of a non-existing language. Hence for future experiments, we will provide participants with more information about the experiment and teach them words from a living language. Also, we might provide some reward for good discipline and cooperation, but not for memory performance, as this might bias results.

Altogether, these data indicate that audio-visual learning of vocabulary items can be enhanced by means of enactment. However, enactment works better if not only performed by an actor and observed by learners but also performed by the learners themselves. Beyond the limitations discussed above concerning the cooperation of our subjects, more research is needed in order to clarify this issue.

Acknowledgements

The authors thank Christopher Hackl, Oliver Kern and Judith Etzelshofer for assistance during the experiment and Christopher Hackl for data collection. The authors are grateful to Bob Bach for helpful discussion.

References

- [1] M. Macedonia, I. Groher, and F. Roithmayr, "Intelligent Virtual Agents as Language Trainers Facilitate Multilingualism". *Frontiers in Psychology*, 2014. 5.
- [2] J. Cassell, *Embodied conversational agents*. 2000, Cambridge, Mass. ; London: MIT Press.
- [3] S. Heidig and G. Clarebout, "Do pedagogical agents make a difference to student motivation and learning?". *Educational Research Review*, 2011. 6(1): p. 27-54.
- [4] R.-J. Beun, E. Vos, and C. Witteman, "Embodied Conversational Agents: Effects on Memory Performance and Anthropomorphisation", in *Intelligent Virtual Agents*, T. Rist, et al., Editors. 2003, Springer Berlin Heidelberg. p. 315-319.
- [5] J. Miksatko, K. Kipp, and M. Kipp, "The Persona Zero-Effect: Evaluating Virtual Character Benefits on a Learning Task with Repeated Interactions", in *Intelligent Virtual Agents*, J. Allbeck, et al., Editors. 2010, Springer Berlin Heidelberg. p. 475-481.
- [6] K. Bergmann and M. Macedonia, "A Virtual Agent as Vocabulary Trainer: Iconic Gestures Help to Improve Learners' Memory Performance", in *Proceedings of the 13th International Conference on Intelligent Virtual Agents.*, T.I.C.o.I.V.A.I. 2013), Editor 2013. p. submitted.
- [7] M. Macedonia, K. Müller, and A.D. Friederici, "The impact of iconic gestures on foreign language word learning and its neural substrate". *Human Brain Mapping*, 2011. 32(6): p. 982-998.
- [8] H.D. Zimmer, *Memory for action: a distinct form of episodic memory?* 2001, Oxford: Oxford University Press. ix, 205 p.
- [9] M. Macedonia and K. Von Kriegstein, "Gestures enhance foreign language learning". *Biolinguistics*, 2012. 6(Special Issue "Embodiment"): p. 393-416.
- [10] M. Macedonia, "A bizarre pedagogical virtual trainer enhances memory for words in a foreign language". *International Journal of Computer Science and Artificial Intelligence*, 2014. 4(2): p. 24-34.
- [11] S.T. Fiske, A.J.C. Cuddy, and P. Glick, "Universal dimensions of social cognition: warmth and competence". *Trends in Cognitive Sciences*, 2007. 11(2): p. 77-83.
- [12] M. Macedonia, R. Kern, and F. Roithmayr, "Do children accept virtual agents as foreign language trainers?". *International Journal of Learning, Teaching, and Educational Research*, 2014. 7(1): p.131-137.
- [13] J. Engelkamp and H. Krumnacker, "Imaginale und motorische Prozesse beim Behalten verbalen Materials". *Zeitschrift für experimentelle und angewandte Psychologie*, 1980(27): p. 511-533.

- [14] R.L. Cohen, "On the generality of some memory laws.". *Scand J Psychol*, 1981. 22: p. 267-281.
- [15] J. Engelkamp and H. Zimmer, "Zum Einfluß von Wahrnehmen und Tun auf das Behalten von Verb-Objekt-Phrasen". *Sprache & Kognition*, 1983(2): p. 117-127.
- [16] J. Engelkamp, et al., "Memory of self-performed tasks: Self-performing during recognition". *Memory and Cognition*, 1994. 22(1): p. 34-39.
- [17] K. Masumoto, et al., "Reactivation of physical motor information in the memory of action events". *Brain Res*, 2006. 1101(1): p. 102-9.
- [18] J. Engelkamp and H.D. Zimmer, "Motor program information as a separable memory unit". *Psychological Research*, 1984(46): p. 283-299.
- [19] J. Engelkamp and H.D. Zimmer, "Motor programs and their relation to semantic memory". *German Journal of Psychology*, 1985(9): p. 239-254.
- [20] K. Stefan, et al., "Formation of a Motor Memory by Action Observation". *J. Neurosci.*, 2005. 25(41): p. 9339-9346.
- [21] N.E. Miller and J. Dollard, *Social learning and imitation*. 1979, Westport, Conn.: Greenwood Press. xiv, 341 p.
- [22] B.G. Galef and K.N. Laland, "Social Learning in Animals: Empirical Studies and Theoretical Models". *BioScience*, 2005. 55(6): p. 489-499.
- [23] C.M. Heyes and B.G. Galef, *Social learning in animals : the roots of culture*. 1996, San Diego: Academic Press. xix, 411 p.
- [24] T.R. Zentall, "Action imitation in birds". *Learn Behav*, 2004. 32(1): p. 15-23.
- [25] K. Thonhauser, et al., "Social learning in Cartilaginous fish (stingrays *Potamotrygon falkneri*)". *Animal Cognition*, 2013. 16(6): p. 927-932.
- [26] A.N. Meltzoff and M.K. Moore, "Imitation of facial and manual gestures by human neonates". *Science*, 1977. 198(4312): p. 74-78.
- [27] J. Piaget, H.E. Gruber, and J.J. Vonèche, *The essential Piaget*. 1977, London: Routledge and K. Paul. xlii, 881 p.
- [28] A.N. Meltzoff and M.K. Moore, "Newborn-Infants Imitate Adult Facial Gestures". *Child Development*, 1983. 54(3): p. 702-709.
- [29] V. Gallese, et al., "Action recognition in the premotor cortex". *Brain*, 1996. 119(2): p. 593-609.
- [30] G. Rizzolatti and L. Craighero, "The mirror-neuron system". *Annu Rev Neurosci*, 2004. 27: p. 169-92.
- [31] H. Théoret and A. Pascual-Leone, "Language Acquisition: Do as You Hear". *Current Biology*. 12(21): p. R736-R737.
- [32] M. Pelaez, J. Virues-Ortega, and J.L. Gewirtz, "Reinforcement of vocalizations through contingent vocal imitation". *J Appl Behav Anal*, 2011. 44(1): p. 33-40.
- [33] D. Lenzi, et al., "Neural basis of maternal communication and emotional expression processing during infant preverbal stage". *Cereb Cortex*, 2009. 19(5): p. 1124-33.
- [34] L.W. Barsalou, "Grounded cognition". *Annual Review of Psychology*, 2008. 59: p. 617-645.
- [35] V. Gallese and G. Lakoff, "The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge". *Cogn Neuropsychol*, 2005. 22(3): p. 455-79.
- [36] D. McNeill, *How language began : gesture and speech in human evolution*. xiv, 264 pages.
- [37] M. Gentilucci and M.C. Corballis, "From manual gesture to speech: A gradual transition". *Neuroscience & Biobehavioral Reviews*, 2006. 30(7): p. 949-960.
- [38] J.M. Iverson and S. Goldin-Meadow, "Gesture paves the way for language development". *Psychological Science*, 2005. 16(5): p. 367-371.
- [39] S. Ozcaliskan and S. Goldin-Meadow, "Gesture is at the cutting edge of early language development". *Cognition*, 2005. 96(3): p. B101-B113.
- [40] S. Goldin-Meadow and M.W. Alibali, "Gesture's Role in Speaking, Learning, and Creating Language". *Annual Review of Psychology*, 2013. 64(1): p. 257-283.
- [41] S.D. Kelly, A. Ozyurek, and E. Maris, "Two sides of the same coin: speech and gesture mutually interact to enhance comprehension". *Psychol Sci*, 2010. 21(2): p. 260-7.
- [42] M. Macedonia and W. Klimesch, "Long term effects of gestures on memory for foreign language words trained in the classroom". *Mind, Brain and Education*, 2014. 8(1): p. In Press.
- [43] R. Reißweber, *Programmieren mit PEARL*. 1988, München, Wien R. Oldenbourg Verlag.
- [44] H. Welbergen, D. Reidsma, and S. Kopp, "An Incremental Multimodal Realizer for Behavior Co-Articulation and Coordination", in *Intelligent Virtual Agents*, Y. Nakano, et al., Editors. 2012, Springer Berlin Heidelberg. p. 175-188.
- [45] H. Vilhjálmsón, et al., "The Behavior Markup Language: Recent Developments and Challenges", in *Intelligent Virtual Agents*, C. Pelachaud, et al., Editors. 2007, Springer Berlin Heidelberg. p. 99-111.
- [46] M. Macedonia, "Learning a Second Language Naturally: The Voice Movement Icon Approach". *Journal of Educational and Developmental Psychology*, 2013. 3(2): p. 102-116.
- [47] M. Macedonia and T.R. Knösche, "Body in Mind: How Gestures Empower Foreign Language Learning". *Mind, Brain, and Education*, 2011. 5(4): p. 196-211.
- [48] K. Bergmann and M. Macedonia, "A Virtual Agent as Vocabulary Trainer: Iconic Gestures Help to Improve Learners' Memory Performance", in *Intelligent Virtual Agents*, R. Aylett, et al., Editors. 2013, Springer Berlin Heidelberg. p. 139-148.