International Journal of Learning, Teaching and Educational Research
Vol. 7, No. 1, pp. 131-137, August 2014

Do Children Accept Virtual Agents as Foreign Language Trainers?

Manuela Macedonia[1,2], Roland Kern[1] and Friedrich Roithmayr[1]
[1] Department for Information Systems, University of Linz, Austria
[2] Research Group Neural Mechanisms of Human Communication, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Abstract. Virtual (animated software) agents can train humans in vocabulary learning. This has been successfully tested with adults and more recently also with children. However, the question of how children perceive a virtual agent training them had not been investigated. Here we invited 25 children to evaluate their perception of a virtual and a human trainer who presented written words in a foreign language on videos; both the human trainer and the virtual agent additionally performed a semantically related gesture for each word. Subjects rated the trainers for features related to gestures and for their "personalities". Subjects found human gestures better and gave the human trainer higher sympathy scores; however, the overall difference between their perception of virtual and human trainers was not significant.

Keywords: Evaluation; Intelligent Virtual Agent; Enactment; Training; Learning; Foreign Language

Introduction

Evaluation, Intelligent Virtual Agent, Enactment, TVirtual pedagogical agents have been developed in the last decade in order to support learning in different domains (Kim & Baylor, 2006b). Some of them have human looks and can interact with users to a certain extent by appropriate facial expressions, head nodding (Cassell, 2000), and gestures (Bergmann, Kahl, & Kopp, 2013). In different domains, they can successfully support learners (Kim & Baylor, 2006a) and positively influence their attitudes towards the topic to be learned (Johnson, Ozogul, Moreno, & Reisslein, 2013). Also, in future, virtual agents in mobile devices will facilitate multilingualism in remote areas of the world. Agents will support people who are physically or financially disadvantaged to access to foreign language instruction at a low cost (Macedonia, Groher, & Roithmayr, 2014).

In recent studies, the virtual agent Billie (Buschmeier & Kopp, 2011) who looks like a young boy of 11 or 12 years has successfully trained humans on vocabulary learning in a foreign language. Billie is driven by the Asap Realizer (Welbergen, Reidsma, & Kopp, 2012), which enables specification of the agent's behavior in the Behavior Markup Language (BML) (Vilhjálmsson et al., 2007).

BML coordinates speech and gesture as well as gaze, and head and other body movements. Billie trains humans by means of enactment. Enactment pairs the presentation of words (and or phrases) with illustrative gestures (Zimmer, 2001). For nearly four decades, laboratory research has shown that accompanying novel verbal information with gestures enhances its memorability. This is the case for words and phrases in a native language as well as in foreign language (Macedonia & Von Kriegstein, 2012). In order to use enactment as a learning strategy, Billie performs gestures himself.

Billie was recently employed as a vocabulary trainer in a study with two groups of young adults. They were trained by the agent and a human trainer to memorize 45 single words of Vimmi (Macedonia, Müller, & Friederici, 2011), an artificial language created for experimental purposes. Participants learned equally well with both trainers. High performers, however, achieved better scores with the agent than with the human trainer (Bergmann & Macedonia, 2013). In another study, Billie cued school children (average 11.4 years) to learn Vimmi vocabulary. The authors pursued two questions: first, whether children can be trained in an ecologically valid environment, i.e., a classroom, with enactment; second, in order to enhance their memory for the words in the foreign language, whether it suffices for these children to watch the agent enunciating the words and performing the gestures, or whether learners need to perform the gestures themselves. Forty four school children (average 11.2 years) were cued to learn 45 single words of Vimmi by only reading and hearing the words (15 items), by watching the agent making the corresponding gestures (15 items), and by imitating the agent enacting the words he enunciates (15 items). Memory results from cued translation tests show that Billie successfully trained the children in the classroom and that imitating the gestures was the better way to enhance the children's performance (Macedonia, Bergmann, & Roithmayr, 2014). These two experiments demonstrate that the virtual agent Billie can successfully replace human trainers and train both adults and children to learn vocabulary items in a foreign language by means of enactment.

Previous Study on the Acceptance of Billie

Participants of both studies above were trained by Billie. However, until investigated, it is not clear which attitude humans have towards virtual teachers. Therefore, Billie's acceptance as a language trainer was tested in a recent study by Macedonia (2014). There the author asked 18 adults to rate the agent for his gesture quality and personality and compare him with a human trainer in an online survey. Gesture quality was necessary because enactment plays a major role in the way Billie trains subjects to words in a foreign language. Materials used for the evaluation were the same videos as they were used in the experiment by Bergmann & Macedonia (2013). There adults learned novel words with both, a human and a virtual trainer. Data showed that the agent's gestures, as expectedly, were rated as less natural than the human gestures. However, participants did not perceive a significant difference between their personalities, except for a few traits that were considered better for the human trainer. Methodologically the study had an Achilles heel: raters compared two trainers that were not controlled for age and gender. In fact, the human trainer was an

adult woman, whereas the virtual agent looks like a boy of 11 to 12. Furthermore, subjects were adults rating a peer and a child as vocabulary trainer; differences in acceptance of the virtual trainer might have been biased by the above factors. Furthermore, the fact that many adults are not digital natives can have an influence of their perception of the agent as a trainer.
In the present study, we tested the acceptance of a virtual agent and a human as trainers by controlling for their age and gender. Both trainers were male of approximately the same age (11 to 13) and they were rated by children of the same age. Our aim was to grasp the children's attitude towards a virtual teacher.

Methods

Participants

Twenty five children (16 male, 9 female), mean age 12.5 years (SD 0.65), took part in the study. They were recruited in an Austrian school and participated for free. The participants were naïve of the study's goal and had never seen Billie before. Participants were also interviewed about their interaction with media and the time they spent with them daily. This was done in order to establish possible relationships between their global attitude towards media (Litt, 2013) and their perception of the agent as a trainer.

Stimuli

We used 30 videos (MPEG4) with a length of approximately 5s each, subdivided into two blocks. In one block (15 videos), the agent performed gestures illustrating the words' semantics. In the other block (15 videos), the human trainer did the same with the same words. Additionally, for each word in the foreign language, a translation into the native language of the participants (German) appeared in written form. Note that the human videos were created by copying exactly the agent's videos. Thus both sets of videos were identical for gestural execution (Figure 1).

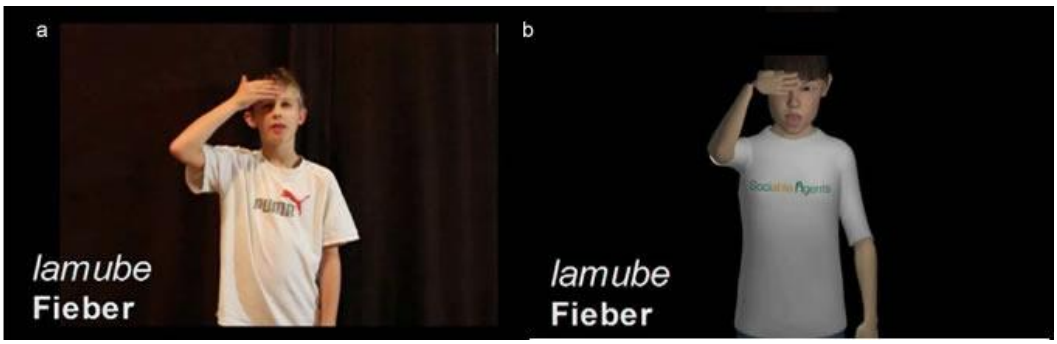

Figure 1: Screenshots from videos used for the online survey for the Vimmi word lamube (English fever). Word enacted a) by the human trainer and b) by the agent.

Sampling procedure

Participants were asked to complete an online survey created with the tool Google Forms (Fuente Valentín, Pardo, & Delgado Kloos, 2009). During this task, they were monitored by the experimenter. He controlled for time and accurate execution of the task. Each child had to complete the survey within 30

minutes. Children carefully watched any single video and rated it according to two subsections of questions. The first regarded gestural features, i.e. iconicity, naturalness and speed of execution (Bergmann, Kopp, & Eyssel, 2010); the second subsection concerned personality traits of the trainers, i.e. sympathy, friendliness and intelligence (Fiske, Cuddy, & Glick, 2007). Furthermore, we acquired data concerning the subjects' gender and their digital nativeness which is their attitude and time spent daily with mobile devices and other media. We did this in order to find possible correlations in this population that might explain preferences for the agent or the human trainer.

Results

Variables were rated on a five-point Likert scale, with 1 standing for the worst evaluation and 5 for the best. The human execution of the gestures achieved better scores in iconicity ($F(14,336) = 1.91$, $p < 0.05$) and naturalness ($F(14,33) = 2.76$, $p < 0.005$), see Figure 2.

We tested the quality of gestures by conducting a repeated measures ANOVA, 3x2 with the factor quality of gestures (iconicity, naturalness and speed of execution) and the factor trainer (human vs. agent). The results show significant effects for both experimental factors, i.e., quality of gestures ($F(2,48)=24,32$, $p <0.001$, Figure 2) and trainer ($F(2,48)=1.57$, $p=<0.217$, Figure 3). These two results are not surprising, as Billie's gestures are definitely not as fluent as human gestures.

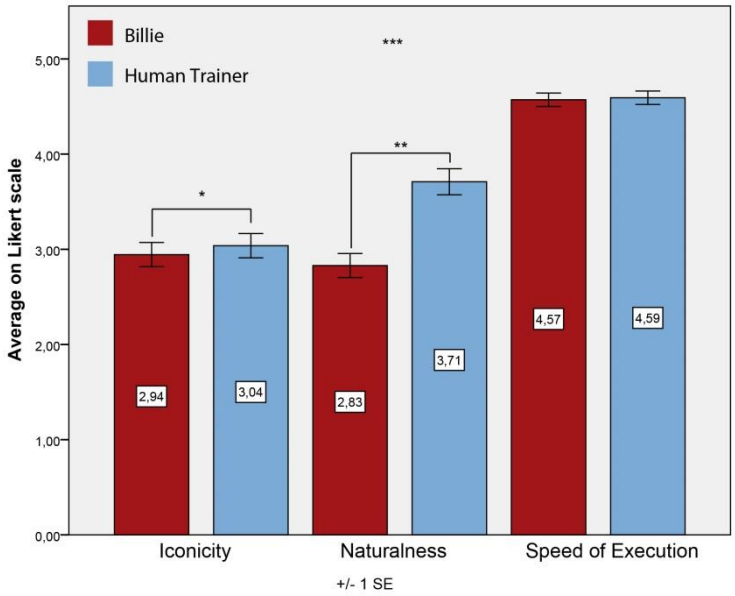

Figure 2: Children's perception of the quality of the gestures.

The factor "personality" aggregated sympathy, friendliness and intelligence. The human trainer achieved higher scores only for sympathy ($F(1,24) = 10.90$, $p < 0.005$).

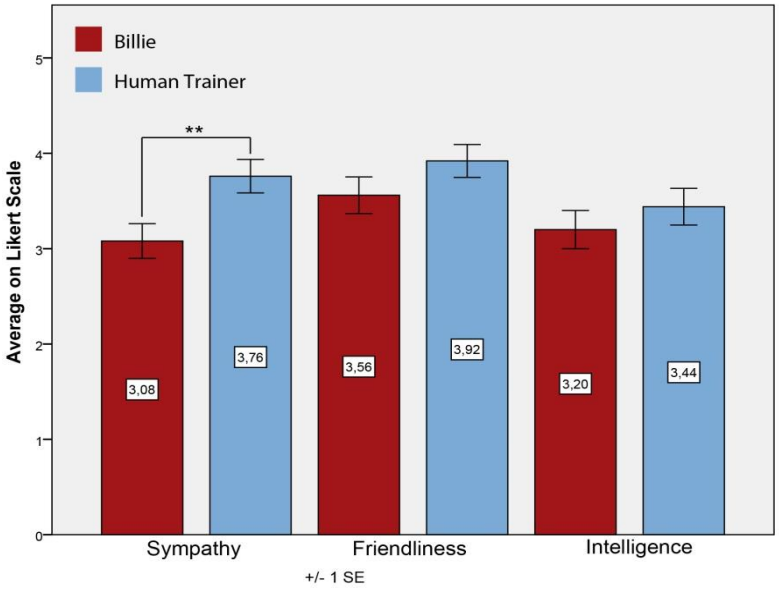

Figure 3: Children's perception of trainer personalities.

We also conducted a 2x2 repeated measures ANOVA with the factors agent's features (quality of gestures and trainer's personality) and trainer (agent vs. human). Our aim was to answer the final question of whether altogether subjects had perceived the two trainers in significantly different way besides the differences reported above. For this analysis, we yielded no significant results.

The digital nativeness (Prensky, 2001) of the subjects, i.e., the expertise and the propensity to use digital media, was computed by acquiring data on the frequency of interaction with the following media: smartphone, desk or laptop, tablet, radio, television, interactive television. Children rated how frequently they used the media on a five-point Likert scale with 1 for the lowest and 5 for the highest frequency in interaction. All media used were aggregated and averaged for each child. We found an average interaction of 3.54 (SD .49) out of 5.

We further computed Pearson correlations with the variables score of digital nativeness and perception of the trainers' quality of gestures and trainers' personalities. In both cases we failed to find positive correlations ($rs = -.080$, $p = .705$ and $rs = -.050$, $p = .813$, respectively). Interestingly, the correlation between the gender of the subjects (16 males and 9 females) and their perception of the gestures and personalities of the trainers yielded significant results: for the agent's quality of gestures $rs = -.539^{**}$, $p = .005$ and trainer's personality $rs = -.579^{**}$, $p = .002$. Gender mattered: girls liked the agent better than their human peer as a trainer.

Conclusion

We conducted the present study with the aim to investigate children's acceptance of the virtual agent Billie as a vocabulary trainer. We designed the

study as a comparison between Billie, a virtual agent that looks like a boy of about 12 and a human boy of 12. Both vocabulary trainers were rated by school children of the same age. Raters agreed that gestures were better if performed by the human, particularly regarding naturalness. Speed did not differ for both trainers. We attribute this result to the fact that when the child's videos were realized, the young actor was instructed to first watch the virtual trainer and thereafter to perform the same gestures. Even if not instructed to do so, the boy did not only imitate the shape of the gesture but also the speed to which the agent performed it. Hence, raters could not see any differences in the speed of execution.

The raters had more sympathy for the human trainer, as previously also reflected in an adult study (Macedonia, 2014). As participants had no interaction with the trainers, sympathy might simply be related to human appearance. In fact, preference for species is influenced by similarity. In her study, Batt (2009) found out that humans like other species on the basis of shared bio-behavioral traits. In our study, because of his gestures and several other features, the agent is still not a boy, despite his anthropomorphic looks. This possibly lead participant to give higher sympathy scores to the child.

For the variable intelligence, subjects detected no difference between the trainers. This could be related to the trainers' task during the experiment. They simply performed gestures and presented words in an unknown language. What they did had nothing intelligent per se and it does not surprise that the raters could not see any difference between both trainers. Also, we speculate that subjects aged of 12 might not have a clear representation of the concept of intelligence. Subjects might not have abstracted that the machine must be less intelligent than the human, as adults did in the study by Macedonia (2014). Correlations with participants' gender show that girls liked the agent better than the human. It is unclear why girls do. Further research is needed in order to confirm and / or clarify this result. Altogether, our present study confirms the acceptance of a virtual agent while controlling for age and gender. We reason that future improvements in the software behind virtual agents, particularly in gesture execution, will further increase their acceptance and facilitate the widespread use of virtual agents in foreign language learning.

References

Batt, S. (2009). Human attitudes towards animals in relation to species similarity to humans: a multivariate approach. Bioscience Horizons, 2(2), 180-190.

Bergmann, K., Kahl, S., & Kopp, S. (2013). Modeling the Semantic Coordination of Speech and Gesture under Cognitive and Linguistic Constraints. In R. Aylett, B. Krenn, C. Pelachaud & H. Shimodaira (Eds.), Intelligent Virtual Agents (Vol. 8108, pp. 203-216): Springer Berlin Heidelberg.

Bergmann, K., Kopp, S., & Eyssel, F. (2010). Individualized Gesturing Outperforms Average Gesturing – Evaluating Gesture Production in Virtual Humans. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud & A. Safonova (Eds.), Intelligent Virtual Agents (Vol. 6356, pp. 104-117): Springer Berlin Heidelberg.

Bergmann, K., & Macedonia, M. (2013). Proceedings of the 13th International Conference on Intelligent Virtual Agents.: T. I. C. o. I. V. A. I. 2013).

Buschmeier, H., & Kopp, S. (2011). Towards Conversational Agents That Attend to and Adapt to Communicative User Feedback. In H. Vilhjálmsson, S. Kopp, S. Marsella & K. Thórisson (Eds.), Intelligent Virtual Agents (Vol. 6895, pp. 169-182): Springer Berlin Heidelberg.

Cassell, J. (2000). Embodied conversational agents. Cambridge, Mass. ; London: MIT Press.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. Trends in cognitive sciences, 11(2), 77-83.

Fuente Valentín, L., Pardo, A., & Delgado Kloos, C. (2009). Using Third Party Services to Adapt Learning Material: A Case Study with Google Forms. In U. Cress, V. Dimitrova & M. Specht (Eds.), Learning in the Synergy of Multiple Disciplines (Vol. 5794, pp. 744-750): Springer Berlin Heidelberg.

Johnson, A. M., Ozogul, G., Moreno, R., & Reisslein, M. (2013). Pedagogical Agent Signaling of Multiple Visual Engineering Representations: The Case of the Young Female Agent. Journal of Engineering Education, 102(2), 319-337.

Kim, Y., & Baylor, A. (2006a). Pedagogical Agents as Learning Companions: The Role of Agent Competency and Type of Interaction. Educational Technology Research and Development, 54(3), 223-243.

Kim, Y., & Baylor, A. (2006b). A Social-Cognitive Framework for Pedagogical Agents as Learning Companions. Educational Technology Research and Development, 54(6), 569-596.

Litt, E. (2013). Measuring users' internet skills: A review of past assessments and a look toward the future. New Media & Society, 15(4), 612-630.

Macedonia, M. (2014). A bizarre pedagogical virtual trainer enhances memory for words in a foreign language. International Journal of Computer Science and Artificial Intelligence, 4(2), 24-34.

Macedonia, M., Bergmann, K., & Roithmayr, F. (2014). Imitation of a pedagogical agent's gestures enhances memory for words in second language. Under Revision.

Macedonia, M., Groher, I., & Roithmayr, F. (2014). Intelligent Virtual Agents as Language Trainers Facilitate Multilingualism. [Perspective]. Frontiers in Psychology, 5.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. Human Brain Mapping, 32(6), 982-998.

Macedonia, M., & Von Kriegstein, K. (2012). Gestures enhance foreign language learning. [Review]. Biolinguistics, 6(Special Issue "Embodiment"), 393-416.

Prensky, M. (2001). Digital natives, digital im migrants part 1. On the horizon, 9(5), 1-6.

Vilhjálmsson, H., Cantelmo, N., Cassell, J., E. Chafai, N., Kipp, M., Kopp, S., et al. (2007). The Behavior Markup Language: Recent Developments and Challenges. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis & D. Pelé (Eds.), Intelligent Virtual Agents (Vol. 4722, pp. 99-111): Springer Berlin Heidelberg.

Welbergen, H., Reidsma, D., & Kopp, S. (2012). An Incremental Multimodal Realizer for Behavior Co-Articulation and Coordination. In Y. Nakano, M. Neff, A. Paiva & M. Walker (Eds.), Intelligent Virtual Agents (Vol. 7502, pp. 175-188): Springer Berlin Heidelberg.

Zimmer, H. D. (2001). Memory for action: a distinct form of episodic memory? Oxford: Oxford University Press.

Acknowledgments

The authors thank Bob Bach for helpful discussion.