Virtual or Physical? Social Robots Teaching a Fictional Language Through a Role-Playing Game Inspired by *Game of Thrones*

Hassan Ali, Shreyans Bhansali, Ilay Köksal, Matthias Möller, Theresa Pekarek-Rosin, Sachin Sharma, Ann-Katrin Thebille, Julian Tobergte, Sören Hübner, Aleksej Logacjov, Ozan Özdemir, Jose Rodriguez Parra, Mariela Sanchez, Nambiar Shruti Surendrakumar, Tayfun Alpay, Sascha Griffiths, Stefan Heinrich, Erik Strahl, Cornelius Weber, and Stefan Wermter

Knowledge Technology, Department of Informatics, Universität Hamburg, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
PHRI1819-extended@informatik.uni-hamburg.de

Abstract. In recent years, there has been an increased interest in the role of social agents as language teachers. Our experiment was designed to investigate whether a physical agent in form of a social robot provides a better language-learning experience than a virtual agent. We evaluated the interactions regarding enjoyment, immersion, and vocabulary retention. The 55 participants who took part in our study learned 5 phrases of the fictional language *High Valyrian* from the TV show *Game of Thrones*. For the evaluation, questions from the Almere model, the Godspeed questionnaire and the User Engagement Scale were used as well as a number of custom questions. Our findings include statistically significant results regarding enjoyment ($p = 0.008$) and immersion ($p = 0.023$) for participants with little or no prior experience with social robots. In addition, we found that the participants were able to retain the *High Valyrian* phrases equally well for both conditions.

Keywords: language learning · social robot · virtual agent · role-play · fictional language.

1 Introduction

Research into technology-enhanced language learning has shown that technology can facilitate communication, reduce anxiety, encourage discussions and lead to an overall increase in learning motivation [1]. However, the effectiveness of language teaching is often dependent on the ability to raise and maintain the interest and enthusiasm of the student, which is easier with a teacher conducting the lessons [2]. Therefore, an embodied agent, for instance a robot, may positively influence how people learn.

In our research, we aim to investigate the impact of having a social robot as a language teacher. We conducted a user study which directly compared interacting with a robot and a virtual agent of the same appearance and capabilities.

Seeing how most people are already accustomed to learning a language with the aid of technology, the virtual agent, which is a virtual version of the robot on a screen, is necessary to measure the effect of the physical agent. We focused on three hypotheses:

1. People enjoy interacting with a physical agent more than with a virtual agent.
2. People have a more immersive interaction with a physical agent than with a virtual agent in a role-playing scenario.
3. People learn better when interacting with a physical agent than with a virtual agent.

Since role-playing is an often deployed and well-received method in language teaching to facilitate engagement [3], we decided to model our interaction in the frame of a role-playing scenario as well. We decided on teaching a fictional language to eliminate any bias caused by prior knowledge of the language in international study participants. The fictional language of our choice was *High Valyrian* from the TV show *Game of Thrones*, which currently boasts over 1 million learners on the language learning platform *Duolingo*[1], making it more popular than naturally evolved languages like Hawaiian (537k learners) or Navajo (311k learners).

2   Social Agents as Second Language Teachers

Human-Robot Interaction (HRI) experiments, which revolve around teaching a language, often benefit from using artificial languages. *Toki Pona* is an artificial language with a small vocabulary size ($\sim$120 words), which was taught to children between the ages of ten and eleven through a dialogue-based interaction with the iCat social robot [4]. Especially in international environments where almost every language could potentially be spoken, creating or using an existing artificial language not only limits bias but also minimizes ambiguity in natural language understanding [5]. We decided on teaching the language *High Valyrian*, which was originally created for the *A Song of Ice and Fire* fantasy book series written by George R. R. Martin and further developed in the context of the famous television series *Game of Thrones*. In addition to the benefits of any fictional language (e.g. similarity to revitalized languages like Hebrew or constructed languages like Esperanto [6]), our intention with selecting *High Valyrian* was using the popularity and interest surrounding the start of the final season of the show to recruit a large number of participants.

One major advantage of one-on-one tutoring is the possible adaptation to the skills of a student by the teacher. A study on the effective role of teachers revealed that enjoyment is highly related to the classroom practices, engaging the students, and a positive attitude [7]. The effect of a personalized experience with a social agent in language tutoring scenarios has been explored in studies like [8],

---

[1] https://www.duolingo.com/courses (accessed 21/08/2019)

where an adaptive response algorithm provided children (three to five years old) with tailored verbal and non-verbal feedback. Even though no significant difference was recorded in regard to learning gains, the long-term valence showed a significant difference in favor of the personalized robot. Thus, we designed our social agent with a focus on facilitating enjoyment and enthusiasm in the adult learner, while providing a personalized learning experience. Additionally, the physical presence of a robot tutor has been shown to have a positive influence on the cognitive learning gains of participants in a game-play scenario [10].

In addition to the teacher's behavior, the setting also plays an important role in learning a foreign language. It has been shown that the use of games, more specifically role-playing games, in language education supports communication and social interaction due to an increase in immersion [3,9]. To improve the learning experience in our study, we designed a role-playing scenario which was simple enough to be easily explained to participants unfamiliar with *Game of Thrones* while still capturing the spirit of the show.

3   Methodology

For our study, we used the Neuro-Inspired COmpanion (NICO), which is a humanoid robot developed at the University of Hamburg as a novel open-source neuro-cognitive research platform [11]. It is equipped with sensors, motors, stereo cameras, and facial expression capabilities to feature human-like perception and interaction. For the sake of the experiment, a virtual version of NICO was built using the Godot[2] game engine. With the thought of unifying the two experimental conditions, the appearance, dialogue, and non-verbal communication of the virtual NICO were designed to be as identical to the physical robot as possible (Figure 1). The participant can interact with the virtual NICO via external speakers, a mouse, an external microphone, and a screen-mounted camera. Grasping real objects (either a bottle or a cup) to present them to the agent was substituted by moving a mouse pointer to the virtual object and clicking it.

3.1   Proposed System

In our experiment, the participant interacts with the social agent mainly through spoken dialogue. Each interactive session with the social agent follows a typical three-phase design [12]:

1. **Presentation**: The participant is welcomed to the role-playing scenario, in which they play an ambassador who is being prepared for a meeting with the Khaleesi[3] by her counselor NICO.
2. **Practice**: 5 phrases consisting of greeting, introduction, presenting two objects (a wine bottle and a golden cup) as gifts to the Khaleesi, and farewell are taught utilizing a simple version of spaced repetition.

---

[2] https://godotengine.org/ (accessed 21/08/2019)

[3] A popular character from *Game of Thrones* speaking *High Valyrian*.

Fig. 1. Experimental setup of the two conditions: the physical robot on the left and virtual agent on the right. Participants sit at the table in front of the respective agent.

3. **Production**: As a final step of the preparation, NICO pretends to be the Khaleesi in order to practice for the imminent meeting. The participant is prompted to recall the taught phrases to test their retention.

The moment the participant steps into the experiment room, the role-playing scenario is initiated and NICO stays in character throughout the whole interaction. We dressed both the physical robot and the virtual agent in the appropriate style for *Game of Thrones* (Figure 1). NICO plays a counselor to the Khaleesi, teaching the participant the *High Valyrian* phrases necessary to navigate the court as an ambassador from a distant land, who might not know the proper etiquette and protocol.

The conversation is modeled through a Hierarchical State Machine control architecture using SMACH[4] for the internal structure and ROS[5] for communicating data. For a lively and human-like experience, the system is enhanced by features like gestures, face tracking, pointing at objects, and object manipulation. Figure 2 shows a snippet of the state machine when teaching the name of an object in the practice phase. The depicted states are repeated for each phrase (Table 1) and define the interaction flow.

Table 1. The *High Valyrian* phrases taught during the interaction with their English translation. *: shortened from "Mother of Dragons", one of the titles for the Khaleesi in the show. **: translates as "coming from the waves", the title given to the participant.

Rytsas, mhysa.	Iksan pelone.	Averilla\Aeksion maghan.	Ilas daria.
Hello, mother*.	I am pelone**.	I bring wine\a golden cup.	Farewell, my queen.

---

[4] http://wiki.ros.org/smach (accessed 21/08/2019)
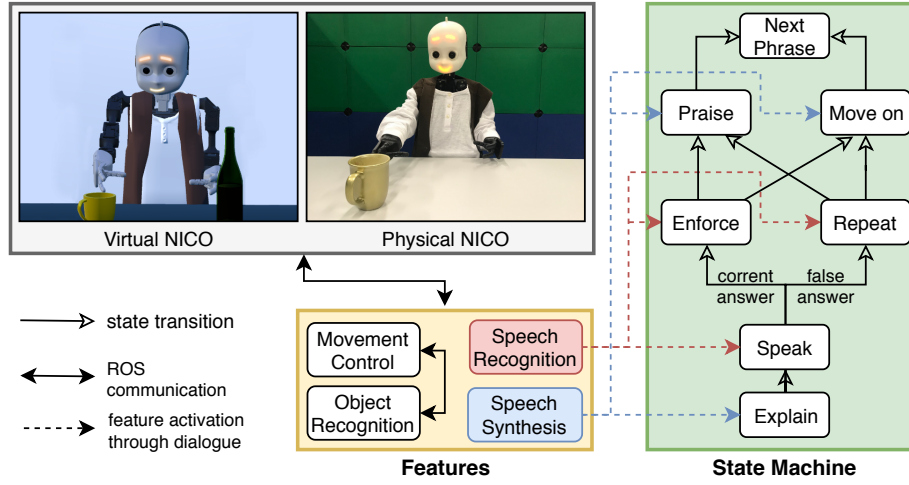
[5] http://www.ros.org/ (accessed 21/08/2019)

Fig. 2. The interaction flow while teaching an object name in the practice phase. The red and blue arrows show in which states speech recognition or speech synthesis is used. 1) NICO says the phrase ("Explain") and points at the object. 2) The participant repeats it ("Speak"). 3) The "Enforce" state ensures an equal number of repetitions among all participants. 4) NICO provides motivational feedback depending on the correctness of the participant's answer ("Praise"\"Move on") before proceeding to the next phrase.

Phrases spoken by NICO are generated with a speech synthesizer built using Amazon Polly[6]. The Text-to-Speech system supports monolingual (only English or *High Valyrian*) and bilingual (mixed) phrases. To recognize the participant's utterances of *High Valyrian*, an Automatic Speech Recognition (ASR) system has been built on top of DOCKS [13]. The system achieved a precision of 0.81, a recall rate of 0.66, and an F1 score of 0.73.

To achieve an immersive conversation, NICO has to be able to show awareness to the surroundings and interact with them. For example, NICO should be able to look and point at an object simultaneously when teaching its name (Figure 2). To be able to perform such gestures, NICO needs to be able to detect the positions of the objects on the table in front of it. While the object has a fixed position in the virtual environment, Darknet CNN (YOLOv3-416) [14] is used to detect the objects for the physical robot. The positions are then mapped to the closest match in a previously recorded data set in order to efficiently generate correct joint movements.

3.2   Experimental Procedure

59 participants, recruited through flyers around town and the University of Hamburg, took part in the experiment, of which 4 samples had to be excluded due

---

to their insufficient English skills, leaving a total of 55 participants (21 female, 34 male). 70.9 % reported little or no prior experience with social robots, while 76.4 % had used a language learning application before. Before the experiment, 63.7 % reported being comfortable with the idea of interacting with a robot in an educational context. The majority of the participants were native German speakers between the ages of 18 and 29.

The experimental procedure consisted of two interactions, one with the physical and one with the virtual NICO, as part of a within-subject design, with a number of questionnaires in between. The informed consent of each participant was acquired before the start of the experiment. First, they were asked to fill out a questionnaire inquiring their experience with language learning applications, social robots, and their preference regarding real or virtual teachers in a language learning context. The order of the interaction with the two agents was randomized for each participant while keeping an equal number of samples for each condition.

After each interaction, the participants were asked to fill in a questionnaire which was composed of 1) the Godspeed questionnaire [15], measuring anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety, 2) the perceived enjoyment and usefulness scale of the Almere Model [16], and 3) the User Engagement Scale (UES) [17] which measures perceived usability, aesthetic appeal and focused attention. An additional post-experiment questionnaire contained some demographic questions regarding age, gender, the participant's familiarity with the role-playing scenario, and lastly, their preference for one of the two conditions. Afterward, we briefly interviewed each participant to gain further insight into the reasons behind their preference of one agent over the other.

While NICO itself interacted autonomously with the participants, the interaction was supervised by two researchers hidden behind a wall, who were responsible for evaluating the accuracy of the ASR system and, in case of system errors, operated a Wizard of Oz interface for a seamless flow of the experiment. It was generally triggered due to microphone failures caused by noise and unintelligible speech (coughing, sneezing, etc.) with a failure rate of 3.6% per condition.

4   Results

According to our hypotheses we expect 1) increased enjoyment, 2) increased learnability and 3) increased immersion when interacting with the physical agent. For the evaluation of the results of the 5-point semantic differential scales of the Godspeed questionnaires and the 5-point Likert scales of the Almere Model and the UES, a Mann-Whitney-U test was used with an expected α value of 0.05.

4.1   User Experience

While the results of the Godspeed questionnaires proved to be partially inconclusive for our pool of participants, we found statistically significant results

$(p < 0.05)$ for both enjoyment and immersion in the subgroup of participants, who had no or little prior experience with social robots $(n = 39)$.

Enjoyment: Comparing the score of the perceived enjoyment scale of the Almere model shows that participants with little or no experience with robots experienced higher enjoyment while interacting with the robot. They perceive the robot as more enjoyable, more fascinating and less boring than the virtual agent. There was no significant difference in the perceived usefulness of the agents (Table 2), even though the virtual agent scored slightly higher in that regard.

Table 2. The individual items of the Perceived Enjoyment scale. For the subgroup of participants with no or little prior experience with robots $(n = 39)$, the analysis of the Perceived Enjoyment scale of the Almere model shows a significantly higher mean for the robot $(p < 0.05)$. While the Virtual Agent scores higher in regard to Perceived Usefulness, the difference is not statistically significant. *: reverse-scored.

Item (Almere)	Robot	Virtual	p
Enjoyable	3.95 ± 0.89	3.53 ± 0.86	0.014
Fascinating	4.05 ± 0.86	3.28 ± 0.99	<0.001
Boring*	4.26 ± 0.79	3.69 ± 1.10	0.011

Scale (Almere)	Robot	Virtual	p
Perceived Enjoyment	3.80 ± 0.75	3.41 ± 0.71	0.008
Perceived Usefulness	3.00 ± 1.00	3.09 ± 0.88	0.379

For the entire pool of participants $(n = 55)$, the Godspeed questionnaires showed a significant difference in anthropomorphism $(p = 0.03)$ and a tendency of animacy $(p = 0.08)$ in favor of the robot, especially for the participants who interacted with the robot after interacting with the virtual agent. Regarding the likeability, perceived intelligence and perceived safety of the two agents, no significant differences $(p > 0.05)$ could be discerned, which indicates that both agents were perceived as equally likeable $(M_R = 4.03 \pm 0.83, M_V = 4.01 \pm 0.63)$, and competent $(M_R = 3.57 \pm 0.63, M_V = 3.61 \pm 0.55)$. Participants felt equally safe $(M_R = 2.33 \pm 1.04, M_V = 2.44 \pm 0.91)$ regardless of the observed condition. In general, the robot seems to be perceived as more enjoyable and people seem to be more willing to attribute human-like qualities to the physical NICO.

Regardless of the condition, the second interaction was enjoyed less by the participants, with a measurable decrease in the perceived likeability $(p = 0.038)$, friendliness $(p = 0.009)$, and kindness $(p = 0.003)$ of the agent. This could be linked to a growing familiarity, as anxiety $(p = 0.003)$, agitation $(p < 0.001)$, surprise $(p < 0.001)$ and confusion $(p = 0.023)$ also decreased for the second interaction. The fact that the enjoyment decreased regardless of the condition shows that a similar enough design was accomplished.

Immersion: The results of the user engagement scale (UES) show that the robot was generally perceived as more aesthetically appealing ($p = 0.012$) than the virtual agent, while there was no measurable difference for Focused Attention, Perceived Usability, and Reward Factor ($p > 0.05$). However, participants with little or no experience with robots ($n = 39$) were more immersed and therefore more attentive while interacting with the robot. They also found the robot more appealing (Table 3). The subgroup of participants who were familiar with *Game of Thrones* ($n = 23$), more frequently "lost themselves in the experience" when interacting with the robot ($p = 0.03$), tended to find the robot more "aesthetically appealing" ($p = 0.09$) and more "appealing to the senses" ($p = 0.03$). They were in agreement that the experiment and the subsequent role-playing scenario fit the series well.

Table 3. The subgroup of participants familiar with *Game of Thrones* ($n = 23$) more frequently "lost themselves in the experience" while interacting with the robot. Participants with minimal experience with robots ($n = 39$) perceived the robot as more attentive, aesthetically appealing and the interaction as more rewarding.

Item (UES)	Robot	Virtual	p
Lost themselves	3.17 ± 0.65	2.70 ± 0.93	0.035
Aesthetically appealing	2.70 ± 1.11	2.26 ± 0.86	0.090
Appealing to the senses	3.17 ± 1.07	2.70 ± 0.82	0.034

Scale (UES)	Robot	Virtual	p
Focused Attention	3.50 ± 0.89	3.12 ± 0.96	0.023
Perceived Usability	2.79 ± 0.93	2.79 ± 0.81	0.500
Aesthetic Appeal	3.04 ± 0.88	2.62 ± 0.90	0.017
Reward Factor	3.92 ± 0.71	3.63 ± 0.92	0.088

4.2   Language Retention

While there was only little difference in the retention rate after the first interaction with both the robot (48.5%) and the virtual agent (43.9%), participants on average were able to recall more phrases with the use of the hint system ($R = 34.6\%$, $V = 30.7\%$) than without ($R = 13.9\%$, $V = 12.2\%$).

When interacting with the robot, participants who described themselves as comfortable while interacting with robots used for educational purposes performed better with regards to retention and learnability than participants who did not. This group of participants ($n = 35$) used a lower number of hints, while still performing significantly better ($p = 0.002$) than those that explicitly reported themselves as not comfortable ($n = 8$). This implies that a positive predisposition towards robots might be necessary to achieve better learning results with social robots.

In the post-experiment interviews, some participants pointed out that it was difficult to recall the newly learned phrases in the practice phase because it

was only taught through dialogue, which could explain the overall low retention rate. They specifically mentioned visual clues or written hints as possible improvements. Due to the physical limitations of the robot, and thus the virtual agent, additional assistance could not be given in our case, but further improving the teaching methods could lead to a higher retention rate.

5   Conclusion

In this paper, we have investigated the impact of having a physical robot as a language teacher, compared to interacting with its virtual counterpart on a computer screen. We started with our three hypotheses measuring the enjoyment, immersion, and retention of the participant and in order to compare them, we built a language tutor to teach a handful of *High Valyrian* words and phrases to our participants. Teaching a fictional language gave us the opportunity to make the role-playing scenario interesting, helping the participants immerse themselves in the interactions and enjoy the experience.

Analyzing the results of our user study shows that the perceived enjoyment is significantly higher in the physical condition for the subgroup of participants who report little or no experience with robots. These participants are also significantly more attentive while interacting with the robot, hence, they are more immersed during the interaction with the robot. There is no significant difference between the two conditions regarding how much has been learned, which mirrors the results of previous similar studies comparing robots and virtual agents in a language teaching scenario [18]. However, the evaluation of our results strongly suggests that the physical robot as a language teacher facilitates a better learning experience for the participant. One reason, in particular, could be the physical presence of the robot which prompts people to devote themselves more to the learning process. Another reason could be that robots are usually not devices present in our daily lives, which could foster curiosity and focused attention.

Future work could entail using a different scenario that is specifically tailored to learning. The three-phase structure we used is sufficiently flexible to be used for other teaching scenarios. Moreover, a different environmental setup may enhance the user experience even more.

Overall, our results show that interacting with a robot as a language teacher is already perceived positively. Improvements in making robots more human-like, more interactive, and more intelligent can certainly pave a way for robots as language teachers in the perceivable future.

References

1. Yang, S., Chen, Y.-J.: Technology-enhanced language learning: A case study. Computers in Human Behavior, **23**(1), 860–879 (2007)
2. Lucardie, D.: The impact of fun and enjoyment on adults learning. Procedia - Social and Behavioral Sciences **142**, 439– 446 (2014)
3. Livingstone, C.: Role Play in Language Learning. 1st edn. Longman, Harlow (1983)

4. Saerbeck, M., Schut, T., Bartneck, C., Janse, M. D.: Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In: Proc. of the 28th Int. Conf. on Human Factors in Computing, pp. 1613–1622. ACM: New York, Georgia (2010)

5. Mubin, O., Bartneck, C., Feijs, L.: What you say is not what you get: Arguing for Artificial Languages Instead of Natural Languages in Human Robot Speech Interaction. Spoken Dialogue and HRI Workshop at IEEE (RO-MAN), Japan (2009)

6. Adams, M.: From Elvish to Klingon: Exploring Invented Languages. 1st edn. Oxford University Press, Oxford (2011)

7. Dewaele, J. M., Witney, J., Saito, K., Dewaele, L.: Foreign language enjoyment and anxiety: The effect of teacher and learner variables. Language Teaching Research **22**(6), 676–697 (2018)

8. Gordon, G., Spaulding, S., Kory Westlund, J. M., Lee, J. J., Plummer, L., Martinez, M., Breazeal, C.: Affective personalization of a social robot tutor for childrens second language skills. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence (AAAI-16), pp. 3951–3957. ACM: New York, Arizona (2016)

9. Peterson, M.: Massively multiplayer online role-playing games as arenas for second language learning. Computer Assisted Language Learning **23**(5), 429–439 (2010).

10. Leyzberg, D., Scassellati, B., Spaulding, S., Toneva, M.: The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains. In: 34th Annual Conf. of the Cognitive Science Society, pp. 1882–1887. Japan (2012)

11. Kerzel, M., Strahl, E., Magg, S., Navarro-Guerrero, N., Heinrich, S., Wermter, S.: NICO - Neuro-Inspired COmpanion: A Developmental Humanoid Robot Platform for Multimodal Interaction. In: 26th Proc. of the IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN). pp. 113–120. Portugal (2017)

12. Akalin, N., Uluer, P., Kose, H., Ince, G.: Humanoid robots communication with participants using sign language: An interaction based sign language game. In: Proc. of the IEEE Workshop on Advanced Robotics and its Social Impacts, pp. 181–186. IEEE, Japan (2013)

13. Twiefel, J., Baumann, T., Heinrich, S., Wermter, S.: Improving domain-independent cloud-based speech recognition with domain-dependent phonetic post-processing. In: Proc. of the 28th AAAI Conf. on Artificial Intelligence (AAAI-14), pp. 1529–1535. AAAI Press: US, Canada (2014)

14. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 779–788. (2016)

15. Bartneck, C., Croft, E., Kulic, D.: Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int. Journal of Social Robotics **1**(1), 71–81 (2009)

16. Heerink, M., Kröse, B., Evers, V., Wielinga, B.: Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. Int. Journal of Social Robotics **2**(4), 361–375 (2010)

17. O'Brien, H., Cairns, P. Hall, M.: A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. International Journal of Human-Computer Studies (112), 28–39 (2018)

18. Rosenthal-v. d. Pütten, A. M., Straßmann, C., Krämer, N. C.: Robots or Agents - Neither Helps You More or Less During Second Language Acquisition. In: Proc. of the Int. Conf. on Intelligent Virtual Agents (IVA), pp. 256–268. Springer: Cham, California (2010)