Robots or agents – neither helps you more or less during second language acquisition

Experimental study on the effects of embodiment and type of speech output on evaluation and alignment

Astrid M. Rosenthal-von der Pütten, Carolin Straßmann, and Nicole C. Krämer

University of Duisburg-Essen, Department for Social Psychology: Media and Communication, Forsthausweg 2, 47057 Duisburg, Germany
a.rosenthalvdpuetten, carolin.strassmann, Nicole.kraemer @uni-due.de

Abstract. When designing an artificial tutor, the question arises: should we opt for a virtual or a physical embodied conversational agent? With this work we contribute to the ongoing debate of whether, when and how virtual agents or robots provide more benefits to the user and conducted an experimental study on linguistic alignment processes in HCI in the context of second language acquisition. In our study (n=130 non-native speakers) we explored the influence of design characteristics and investigated the influence of embodiment (virtual agent vs. robot vs. speech based interaction) and system voice (text-to-speech vs. prerecorded speech) on participants' perception of the system, their motivation, their lexical and syntactical alignment during interaction and their learning effect after the interaction. The variation of system characteristics had no influence on the evaluation of the system or participants' alignment behavior.

Keywords: linguistic alignment, second language acquisition, virtual agent, robot, speech output, nonverbal behavior, embodiment, experimental study

1 Introduction

When designing an artificial tutor, numerous decisions have to be made regarding system characteristics. One of the most influential decisions pertains to the question of whether to employ and develop a virtual or physically embodied conversational agent. Virtual agents are comparably cheap and more flexible, but robots are seen to provide an even richer interactive experience, because they can manipulate their environment and actually get in physical contact with users [1]. New approaches opt for migrating both types of artificial entities into one entity represented differently (e.g. as robot or as screen agent) depending on where the user is located [2]. However, it is unclear whether there is a preference of one embodiment form over the other dependent on the specific task the user intends to complete with the help of the system. Results of previous research are somewhat inconclusive (cf. [3] for an extensive review). While quite

a number of studies showed that a robot is more persuasive, receives more attention and is perceived more positively than a virtual agent, there is a general lack of studies examining different behavioral outcomes, particularly, with regard to linguistic behavior. One important future application field for virtual agents and robots are tutoring systems. Artificial tutors could especially be helpful to assist with second language acquisition, because they could help overcome inhibition effects which can occur in human-human interaction due to native speakers linguistically aligning "downward" to non-natives and simplifying their language use. Because people tend to align more strongly to computers than to humans [4], this mechanism might lead to enhanced learning outcomes when computers expose a high standard in the language to learn so that learners are prompted to align upward. Thus, we hypothesize that processes of linguistic alignment in HCI can be exploited to help second language acquisition. However, system characteristics such as embodiment might influence alignment processes. Moreover, alignment processes are supposedly dependent on the learners' comprehension of the speech output of the artificial tutor. Hence, it is important to investigate whether current text-to-speech (tts) software is of sufficient quality to not inhibit alignment and hence also learning processes. Although using prerecorded natural speech would annul comprehension deficits of tts systems, it is more effortful to add study units later on. Thus, we explore whether linguistic alignment can be used in the context of second language acquisition to support learning and whether and if yes how system design characteristics such as embodiment and quality of speech output influence evaluation and learning outcome.

1.1 Effects of differently embodied artificial entities

Virtual agent or robot? This is an essential design decision developers have to make when designing new embodied conversational agents. Both embodiment types provide unique interaction possibilities, but also come along with certain restrictions. In a sense virtual agents are more flexible than robots in that we can easily change a virtual agent's appearance. Hence the appearance can be matched to users' preferences, to the needs of special target groups or to the corporate design of the developing company. Virtual agents can appear on different devices including smartphones. Moreover, they have unlimited degrees of freedom and can perform actions that are not possible in real life. In contrast, robots have limited degrees of freedom, their design cannot be changed easily. Most of the available products are quite stationary and thus can only be used at home or at work. A big advantage of robots is that they are "tangible artifacts" [1], which can be touched, and are able to manipulate their environment by means of physical contact to objects as well as to human interaction partners. Studies comparing robots and virtual agents led to inconsistent results. A majority of findings suggests that robots are superior to virtual representations with regard to the perceived social presence of the entity [5], the evaluation of the entity as entertaining or enjoyable [1, 6], and trustworthy [6]. Furthermore, robots have been demonstrated to be more persuasive [7], elicit more attention [8], and increase user's task performance [9]. On the contrary, other results point to superiority of a virtual representation when the outcome variable is information disclosed to the entity [7]. In fact, there seems to be an interaction effect

of embodiment and task. Regarding the evaluation of task attractiveness, Hoffmann and Krämer [10] demonstrated that a robot was better evaluated in a task-related scenario, while a virtual representation was favored for a conversation. Finally, a study by Bartneck [9] yielded no differences between a robot and a screen animation with regard to how entertaining the interaction was evaluated. Also, Kennedy, Baxter, and Belpaeme [8] observed no difference in children's learning increase when interacting with a real NAO robot or an animation of the robot on a screen. Only one study investigated participants' linguistic behavior on the context of differently embodied agents [12]. It was found that verbosity and complexity of linguistic utterances did not differ between a virtual agent or a robot, but participants used more interactional features of language towards the robot such as directly addressing it by its name. The interplay of embodiment and linguistic alignment has not been investigated so far.

1.2 Linguistic alignment in HHI, HCI and in context of second language acquisition

Empirical evidence in human-human-interaction (HHI) research showed that interaction partners align linguistically in conversations on different levels, for instance, regarding accent or dialect [13], speech rhythm [14], lexical choices and semantics [15] as well as syntax [16]. Quite a number of these effects also occur in interactions with artificial entities. Similarly to HHI, users align to computers, for instance, with regard to prosody, lexis, and syntax (for an overview cf. [17]). Studies with virtual agents showed the same tendencies: in the interaction with a virtual tutor users aligned to lay language or medical jargon [18] and to dialect or standard language [19]. However, studies suggest that alignment in HHI and HCI is similar but not the same, since people tend to show stronger alignment with computers [4] presumably to compensate the computers weaker communicative abilities. However, although initial beliefs about their artificial interaction partner are taken into account by human users, recent comparative work showed that "when social cues and presence as created by a virtual human come into play, automatic social reactions appear to override initial beliefs in shaping lexical alignment." [20]. An open question is whether the physical embodiment of the artificial interlocutor strengthens this effect of blurring boundaries or not. This would be especially important to know when designing artificial language tutors for SLA. Native-speakers often adapt to non-natives in order to foster mutual understanding and successful communication, sometimes with the negative outcome of interfering with successful SLA on a native-speaker level. Using artificial tutors could help to overcome this bias. Since users more strongly align to computers in order to ensure communicative success, there is a potential to exploit these alignment processes for SLA. A first study with native and non-native speakers showed that both groups aligned lexically to a virtual tutor. However, alignment was weaker for non-natives [21] due to a substantial lack of fluency. For instance, if people are not able to conjugate a verb or have trouble to pronounce words correctly, they tend to choose easier vocabulary [22]. Hence, participants might not be able to reproduce all linguistic nuances. This might also be due to the speech output quality of the agent since tts systems do not expose perfect pronunciation. Still, alignment is seen as core to language acquisition, thus, also

to SLA [23] and the tendency of non-natives to align to technology in a learning setting could be exploited for SLA. Admittedly, system characteristics have to be taken into account and their potential inhibiting effects need to be explored – especially in the case of speech output quality.

1.3 Research questions and hypotheses

In this work we explore the potential of artificial tutors to avoid inhibition effects and exploit linguistic alignment processes in HCI for SLA. In particular, we examine whether an artificial tutor's embodiment (virtual agent vs. robot vs. speech based interaction) influences participants' evaluation of the tutor, their lexical and syntactical alignment during interaction and their learning effect after the interaction (***RQ1***). Since previous work showed that robots can elicit more positive evaluations than virtual agents (5-9), we propose that the robot will be rated most positively followed by the virtual agent and the solely language-based tutor (***H1***). In contrast to classic language learning software like DVDs or online platforms, most virtual agents and robots do not use prerecorded natural speech, but tts software which could affect listening comprehension and thereby alignment. Thus, quality of speech output is also varied in our study (***RQ2***). Moreover, we want to know whether alignment in dialog results in better performance in a post interaction language test (***RQ3***).

2 Method

2.1 Experimental design and independent variables

In order to determine which system characteristics people prefer in their interaction with an artificial tutor, we chose a 3x2 between-subjects design with ***speech output*** and ***embodiment*** as independent variables. We used three types of embodiment of the artificial tutor. Participants either interacted exclusively language-based (and saw only a blue screen with the text "language learning system"), or they interacted with a virtual version of the Nao robot or the physical present Nao robot. Secondly, we varied the artificial tutor's speech output. Participants were either confronted with speech output generated by tts software or with prerecorded natural speech (ns). Since Nao's tts system is installed on the physical Nao itself and thus is not available for the virtual Nao, we generated wave files by recording the tts speech output. Natural speech was recorded after generating the tts soundfiles. The speaker was instructed to speak similarly, i.e. imitate intonation and speed (sounds examples can be found in the supplementary material). In order to avoid different perceptions of presence due to sound quality (and not type of speech output), we also used the sound files for the people interacting with the physical Nao.

2.2 Participants and Procedure

One hundred and thirty volunteers (74 female, 56 male) aged between 15 and 53 years (M=26.6; SD=6.87) participated in this study. Seventeen participants had previously interacted with a robot and 26 had interacted with a robot. Participants stem from more than 40 different countries, speak more than 25 different native languages and exposed different levels of German language skills (with a minimum of an intermediate level). Participants were recruited on campus or in German classes in the local adult education center. The study was approved by the local ethics committee. Upon arrival participants read and signed informed consent. They completed two language tests: a test on grammar and reading and listening comprehension and a so called C-Test (www.c-test.de), a cloze test which also addresses language skills with regard to different dimensions. Based on their test results, their country of origin and first language, respectively, participants were distributed equally across conditions where possible and were invited for a second appointment. On the second appointment participants were instructed about the different tasks to be solved with the artificial tutor. Each task was again explained by the tutor during the interaction (cf. Figure 1). Participants were also given a folder with detailed instructions in case they did not understand the tutor. Participants completed five tasks: 1) introducing themselves, 2) describing a picture in detail, 3) playing a guessing game, 4) playing a search game, and 5) again describing a picture. The order of tasks was always the same for all participants. The first two tasks were used to make participants comfortable at speaking loudly to the system. The two structured games (guessing game and search game) were used to analyze alignment processes. We repeated the task of describing a picture to give participants another possibility to speak quite freely at the end of the learning session. This was done to create a more believable training environment for the participants. After the interaction, participants completed a second C-Test as a measure of learning outcome and a questionnaire asking for their experiences and assessment of the interaction. Finally, they were debriefed, reimbursed (€10) and thanked for participation.

2.3 Dependent variables: Self-report

Perception of the artificial tutor. For the person perception of the artificial tutor, we used the Godspeed Questionnaire [24], a semantic differential with 25 bi-polar items which are rated on a 5-point scale. We used the four subscales *Anthropomorphism* (attribution of a human form, characteristics, or behavior to nonhuman things; 5 items, e.g. fake-natural, machinelike-humanlike; Cronbach's $\alpha = .889$), *Animacy* (perception of lifelikeness; 5 items, e.g. dead-alive; stagnant-lively; Cronbach's $\alpha = .880$), *Liking* (5 items, e.g. dislike-like, unfriendly-friendly; Cronbach's $\alpha = .844$), and *Perceived Intelligence* (5 items, e.g. incompetent-competent; Cronbach's $\alpha = .789$).

Social Presence. We assessed participants' sense of co-presence with the Nowak and Biocca Presence Scale [25], which contains 12 items on the concept of "perceived other's co-presence" (Cronbach's $\alpha = .716$) and 6 items on "self-reported co-presence" (Cronbach's $\alpha = .716$), both rated on a 5-point Likert scale.

General evaluation of the interaction. The general evaluation of the interaction was assessed by eight items that asked for the participants' sense of control during the interaction, the enjoyment of the interaction, and whether participants like to use a system like this for other tasks (rated on a 5-point Likert scale; Cronbach's $\alpha = .793$).

Speech output. Additionally, we asked, on a 5-point Likert-scale from "very mechanical" to "very humanlike", how humanlike they perceived the speech output to be.

2.4 Dependent variables: Linguistic alignment

In order to analyze linguistic alignment with the artificial tutor, participants played two structured games (guessing game and search game) in which the tutor and the participant took turns in constructing sentences.

Guessing game. The first structured game was a dialog based game adapted from Branigan et al. [16]. In the original game participants took turns in describing a card and trying to find this card out of a set of cards. The description was originally one sentence. In our adaption of the game participants took turns in guessing the two persons and their interaction on so-called interaction cards (cf. Figure 1) by asking only yes-or-no questions similarly to the "Who am I" guessing game (e.g. "Is the person on the left side female?"; "Is the person on the right side old?", "Is the interaction between the two friendly?"). Questions are asked in a structured manner: first guess who is one the left, then who is on the right and lastly, find out the interaction between the two. By this we created more opportunities to vary lexical and syntactic choices within one round of the guessing game. There were two rounds of guessing in which the system first guessed the participant's card and then the participant guessed the system's card. Between the two rounds, the system changed linguistic choices (e.g. lexical choices (mustache vs. beard); usage of different prepositions, verbs, adjectives or active and passive sentences). Participants´ verbal utterances were analyzed with regard to their lexical choices. A ratio was built for alignment (usage of the same lexical choice (e.g. mustache) / occurrence of the concept (e.g. number of linguistic expression referring to a beard)).

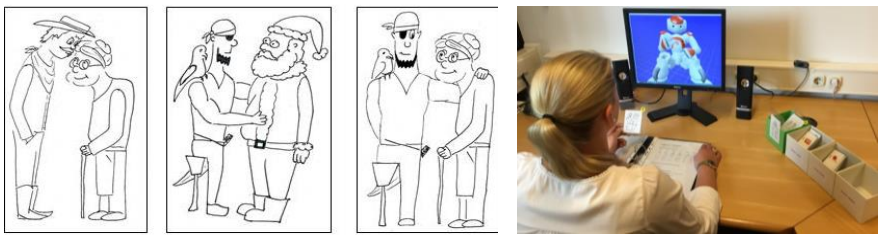

Fig. 1. Left: examples for interaction cards in the guessing game; Right: participant playing the guessing game with the virtual Nao

Search game. The second structured game was also a dialog based game in which the participant and the tutor took turns in describing picture cards to one another. For this game is used the original experimental setup used by Branigan et al. [16] in

order to study syntactic alignment. The cards displayed two characters (e.g. policeman and cowboy) a verb (e.g. to give), and an object (e.g. balloon, cf. Figure 2). Participants had two sets of cards (reading cards and search cards). The task was to take a card from the first card set (reading cards) and to form a sentence based on the characters, verb and object displayed on the card (e.g. the balloon was given to the policemen by the cowboy). The interaction partner's task was to search in their set of "search cards" for this exact card. This means that the system's search cards are identical to participant's reading cards and vice versa. The system began the interaction and built a sentence. The participant had to find the card and put it away and in turn had to take a card from the "reading" set and form a sentence so that the tutor can find the card in its (imagined) search card pool and put it away. In total, the system read out 15 cards, thereby formed 15 sentences in three "blocks". The first block i.e. the first five sentences were formed as passive voice, the second five sentences as prepositional phrase and the last as accusative. A ratio was built for syntactical alignment (usage of the same case / 5 sentences). Since previous research showed that alignment can occur with a delay [26], we also examined whether participants aligned to previously heard syntactic choices, and e.g. in the second block aligned to the first block and in the third block aligned to the second block, respectively.

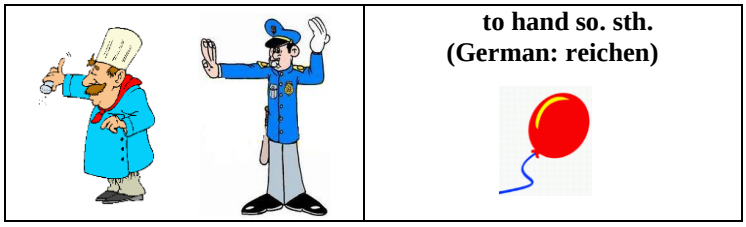

Fig. 2. Example card for the search game

2.5 Dependent variables: Learning outcome

As described in section 2.2., participants exposed different levels of German language skills which were assessed by two language tests: first a standard test on grammar and reading and listening comprehension and second a so called C-Test (www.c-test.de), a cloze test which also addresses language skills with regard to different dimensions. The C-Test was completed again after the interaction with the tutoring system. To explore whether the interaction has a positive effect on participants' language skills we analyzed the results of the C-Tests prior and after the interaction. The C-Test has been used previously for accessing language skills and also improvement in language skills [27]. It usually comprises of five short pieces of self-contained text (ca. 80 words or four to five sentences) in which single words are "damaged". The first sentence is undamaged. Beginning with the second word in the second sentence there are 20 damaged words alternating with undamaged words. In order to reconstruct the sentences, participants have to activate their language fluency. Text pieces were taken from reading exams on an academic language level. Tests are analyzed by true-false answers. Participants could reach 100 points at most.

3 Results

Data were analyzed with ANOVAS and correlation analyses using IBM®SPSS Statistics 21. However, we also estimated Bayes factors using Bayesian Information Criteria [28], comparing the fit of the data under the null hypothesis and the alternative hypothesis using R and the package BayesFactor by Richard D. Morey.

3.1 Participants' self-reported experiences

First, the speech output conditions did not differ regarding how humanlike the voice was perceived (ns: M = 2.80, SD = .70; tts M = 2.72, SD = .75). In order to examine whether embodiment or speech output affects the evaluation of the tutor or the interaction, we conducted ANOVAS with these both factors as independent variables and the dependent variables *general evaluation*, *perceived others co-presence*, *self-reported co-presence*, *likability*, *perceived intelligence*, *anthropomorphism*, and *animacy*. There were no significant differences between the groups nor did we find significant interaction effects (no support for ***H1***, cf. Table 1). An estimated Bayes factor (null/alternative) suggested that the data were between 2.3 and 13.1 times more likely to occur under a model without including an effect of embodiment or speech output, rather than a model with these factors.

Table 1. Means and standard deviations of self-reported dependent variables

	Speech ns[a]	Speech tts[b]	Virtual ns	Virtual tts	Robot ns	Robot tts	BF[c]	BF[c]
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*Emb.*	*SO.*
Gen. Eval.	4.09 (.67)	4.24 (.50)	.4.17 (.53)	4.13 (.53)	4.26 (.71)	4.31 (.45)	4.9	5.3
Perc Oth Co-Pres	2.94 (.27)	3.17 (.28)	3.07 (.28)	3.10 (.38)	3.06 (.33)	3.03 (.31)	11.7	2.3
Self-rep Co-Pres	3.14 (.45)	2.87 (.37)	2.97 (.40)	2.94 (.37)	2.90 (.31)	2.98 (.50)	8.8	4.0
Likability	4.20 (.72)	4.44 (.56)	4.26 (.67)	4.35 (.55)	4.43 (.63)	4.36 (.67)	11.3	4.1
Perc. Intel.	4.13 (.71)	4.07 (.83)	4.05 (.48)	4.08 (.61)	4.08 (.62)	4.06 (.64)	13.1	5.3
Anthro-pomo.	3.32 (1.19)	3.39 (1.13)	2.88 (1.07)	3.20 (.98)	3.19 (1.03)	3.10 (1.00)	6.3	4.8
Animacy	3.44 (1.04)	3.73 (.84)	3.04 (.97)	3.41 (.94)	3.42 (.96)	3.30 (.83)	3.9	3.3

Notes:. a ns = natural speech; b tts = text-to-speech; c BF = Bayes Factor Embodiment; d BF = Bayes Factor Speech Output

3.2 Participants' linguistic alignment

Guessing game. With the guessing game we examined participants' syntactical and semantical alignment during the interaction. Therefore, the system's utterances between the two rounds varied in lexical choices when describing the features of the displayed characters (*age* (old vs. advanced in years), *gender* (male/female vs. a man/ a

woman), *facial hair* (mustache vs. beard)). Moreover, the system used different *verbs* (has vs. wears), *adjectives* (friendly vs. kind) and syntactical constructions (*person on the left side vs. the left person; active vs. passive*). As described above, a ratio was calculated for alignment (usage of the same lexical/syntactical choice (e.g. lexical choice mustache) / occurrence of the concept (e.g. number of expressions referring to a beard)). To examine whether embodiment or speech output affects participants' linguistic alignment, we conducted ANOVAS with both factors as independent variables and the seven ratios for linguistic alignment as dependent variables. There were no significant differences between the groups nor did we find significant interaction effects (cf. Table 2). An estimated Bayes factor (null/alternative) suggested that the data were between 1.7 and 9.4 times more likely to occur under a model without including an effect of embodiment or speech output, rather than a model with these factors.

Table 2. Means and standard deviations for alignment ratios in the guessing game

	Speech ns[a]	Speech tts[b]	Virtual ns	Virtual tts	Robot ns	Robot tts	BF[c]	BF[c]
	MD (*SD*)	*MD (SD)*	*MD* (*SD*)	*MD* (*SD*)	*MD* (*SD*)	*MD* (*SD*)	*Emb.*	*SO.*
Left/Right	.43 (.19)	.37 (.31)	.29 (.16)	.43 (.26)	.36 (.29)	.43 (.18)	9.4	3.0
Age	.52 (.29)	.43 (.15)	.45 (.25)	.60 (.34)	.60 (.32)	.58 (.37)	3.9	4.8
Gender	.64 (.27)	.65 (.24)	.61 (.26)	.58 (.22)	.56 (.30)	.64 (.26)	8.6	4.7
Face hair	.39 (.49)	.44 (.42)	.43 (.41)	.29 (.41)	.21 (.28)	.49 (.47)	3.9	4.8
Verb	.63 (.34)	.44 (.41)	.49 (.24)	.65 (.32)	.64 (.37)	.57 (.32)	7.6	4.3
Adjective	.66 (.40)	.50 (.44)	.46 (.38)	.34 (.40)	.46 (.38)	.38 (.43)	2.2	1.7
Act./Pas.	.39 (.24)	.53 (.30)	.61 (.26)	.38 (.41)	.57 (.33)	.47 (.41)	9.0	4.8

Notes:. a ns = natural speech; b tts = text-to-speech

Table 3. Means and standard deviations for alignment ratios in the search game

	Speech ns[a]	Speech tts[b]	Virtual ns	Virtual tts	Robot ns	Robot tts	BF[c]	BF[c]
align-ment	*MD* (*SD*)	*MD (SD)*	*MD* (*SD*)	*MD* (*SD*)	*MD* (*SD*)	*MD* (*SD*)	*Emb.*	*SO.*
direct	.34 (.17)	.36 (.16)	.38 (.18)	.46 (.19)	.34 (.19)	.37 (.16)	2.4	5.2
delayed	.23 (.19)	.18 (.21)	.25 (.24)	.26 (.22)	.18 (.23)	.20 (.20)	11.7	5.7

Notes:. [a] ns = natural speech; [b] tts = text-to-speech

Search game. The search game focused on the syntactical alignment. Regarding all 15 sentences, participants most often used accusative (M =6.61, SD = 5.08), followed by prepositional phrases (M = 3.49, SD = 3.91) and passive voice (M = 3.01, SD = 3.73). In order to examine whether embodiment or speech output affects participants' syntactical alignment, we conducted ANOVAS with both these factors as independent variables and the alignment ratio. There were no significant differences between groups nor did we find significant interaction effects. Since studies have shown that alignment can

occur with a delay [26], we also analyzed whether participants aligned to the previous blocks. Again there were no significant differences between the experimental groups with regard to embodiment and speech output nor interaction effects (cf. Table 3). An estimated Bayes factor (null/alternative) suggested that the data were between 2.4 and 11.7 times more likely to occur under a model without including an effect of embodiment or speech output, rather than a model with these factors.

3.3 Language skills & learning effect

To explore whether the interaction has a positive effect on participants' language skills we analyzed the results of the C-Tests prior and after the interaction. Thus, we conducted split-plot ANOVAS with the group factors embodiment and speech output and repeated measures for the C-Test. Two main effects emerged. First, participants' C-Test scores were worse after the interaction (M = 51.90, SD = 17.09) than at their first appointment to assess their language proficiency level (M = 55.77, SD = 19.43; $F(124,1)=29.97$; $p<.001$, $\eta p^2=.195$). Moreover, the system's embodiment influenced participants' C-Test scores after the interaction ($F(124,2)=6.24$; $p=.003$, $\eta p^2=.091$). The descriptive data suggests that participants interacting with a robot had lower test results than those interacting with a virtual agent or only language-based (cf. Table 4). The factor speech output showed no effect, nor did we find interaction effects. One goal of this study was to explore the potential of artificial tutors to exploit linguistic alignment processes in HCI for SLA. Thus, we correlated participants' alignment ratios with their C-Test results after the interaction, but did not find a significant correlation.

Table 4. Means and standard deviations for the pre and post C-Test (language skill test)

	Speech ns[a]	Speech tts[b]	Virtual ns	Virtual tts	Robot ns	Robot tts
	MD (*SD*)	*MD (SD)*	*MD* (*SD*)	*MD* (*SD*)	*MD* (*SD*)	*MD* (*SD*)
pre C-Test	61.95 (22.97)	53.81 (16.44)	57.15 (20.44)	57.72 (17.23)	49.35 (21.23)	54.18 (18.16)
post C-Test	55.00 (19.28)	50.09 (13.83)	51.25 (18.00)	52.20 (15.68)	51.85 (18.58)	51.00 (18.48)

Notes:. [a] ns = natural speech; [b] tts = text-to-speech

4 Discussion

With this work we contribute to the ongoing debate of whether virtual agents or robots provide more benefits to the user. Previous work predominantly found that robots were more persuasive, entertaining, enjoyable, and trustworthy. Moreover, they elicit more attention, and increase user's task performance (cf. [3] for an overview). However, there is a lack of research on behavioral effects, particularly, with regard to linguistic behavior. Moreover, some studies found interaction effects of type of embodiment and type of task showing that robots were preferred for (physical) tasks and virtual agents for conversations [10]. We conducted an experimental study on linguistic

alignment processes in HCI in the context of SLA and varied the artificial tutor's embodiment (virtual agent vs. robot vs. speech based interaction). Moreover, we argued that tts systems might be problematic in SLA since they do not always pronounce words correctly and therefore we varied whether participants heard tts or prerecorded natural speech. We found that neither embodiment nor quality of speech output influenced participants' perception of the system or their lexical and syntactical alignment during interaction. There were no differences in perceived human-likeness of the speech output. Regarding language skills, participants performed worse in the language test after the interaction compared to the first appointment where participants' language skills were assessed in order to distribute them equally across conditions. We did not find the hoped for positive influence of the system on language skills. This might be due to the different workload of the two appointments. In the initial appointment participants only completed these language tests. On the second appointment they first interacted with the system for about 50 minutes and then completed the language test. Since participants had to concentrate on the tutoring system and interact with it by speaking German this means constant cognitive effort. Hence, participants were probably less concentrated and more exhausted in the posttest than in the pretest. However, there was one effect regarding embodiment: Participants interacting with a robot (regardless of quality of speech output) performed worse in the posttest. Although there were no evaluation differences, we observed that participants were more excited meeting the physical Nao than meeting the virtual or speech-based system. Some of them explored the Nao or even took a picture. Supposedly, at least some participants paid less attention to the actual tasks ahead and concentrated on the robot itself. In sum, the variation of system characteristics had barely influence on the evaluation of the system (***H1)*** or participants' alignment behavior – neither for embodiment (***RQ1***) nor for quality of speech output (***RQ2***). In this study we kept the appearance of the system between the virtually embodied and physically embodied condition consistent with the virtual version of the actual Nao robot. It could be that the recorded speech might be evaluated differently when matched with a human form like a humanlike virtual agent. It is, however, striking that prerecorded speech and tts did also not differ in perceived human-likeness in the language-based only conditions when there is no possible match or mismatch with the appearance of the tutor.

There are several implications relevant for designers of artificial tutors. First, at least in the domain of language learning with predominantly conversation based tasks the type of embodiment or more precisely embodiment itself in whatever form did not result in more positive evaluation effects or different linguistic behavior. Hence, developers can opt for the more flexible and inexpensive virtual agent or solely speech-based system. Second, tts systems have a sufficient quality to be used for SLA purposes. Astonishingly, the tts was perceived only marginally less humanlike than the actual human voice. Since the usage of prerecorded speech is more expensive, harder to implement and to change later on (e.g. extend learning system with new learning situations / games etc.), it is good news that tts systems are perceived equally positive.

Moreover, we wanted to know whether alignment in dialog results in better performance in the post interaction language test (***RQ3***). We found that although participants aligned to the artificial tutor in all conditions comparably to previous studies [18,20,21]

this did not significantly contribute to the post interaction test performance. Maybe participants would need more learning sessions to benefit from the system and to transfer the alignment into a learning progress. Moreover, the descriptive analysis of the seven alignment ratios showed that participants aligned differently strongly. For instance, they aligned more when referring to *gender* than to *facial hair*. Moreover, alignment is generally lower for any passive construction, because they are also rarely used in everyday conversations. Moreover, we observed that participants with lower initial language skills had trouble in producing sentences. This may have confounded the process of alignment which is at least in part an unconscious process based on priming [15-17]. If words or constructions are not known, they cannot be easily activated by primes thereby eliciting alignment. Hence, linguistic alignment in SLA might only be effective for very advanced language learners. Future work should explore whether repeated tutoring sessions accumulate in a learning effect that might be moderated by alignment during the interaction sessions. Moreover, more distinct groups of participants regarding their initial language skills might give insight into the question of whether alignment contributes effectively to SLA only for advanced learners.

5 Acknowledgements

The noALIEN (Using linguistic alignment in German language promotion for immigrants on the basis of human-technology interaction**)** project was funded by the German Federal Ministry of Education and Research.

6 References

1. Wainer J, Feil-Seifer DJ, Shell D et al. (2007) Embodiment and human-robot interaction: A task-based perspective. In: 16th IEEE International Conference on Robot & Human Interactive Communication. IEEE, Piscataway, N.J., pp 872–877
2. Aylett R, Kriegel M, Wallace I et al. (2013) Memory and the design of migrating virtual agents. In: Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, pp 1311–1312
3. Li J (2015) The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. Int J Hum-Comput St 77: 23–37. doi: 10.1016/j.ijhcs.2015.01.001
4. Branigan HP, Pickering MJ, Pearson J et al. (2004) Beliefs about mental states in lexical and syntactic alignment: Evidence from human-computer dialogs. In: Proceedings of the 17th CUNY Conference on Human Sentence Processing
5. Fasola J, Mataric M (2013) A socially assistive robot exercise coach for the elderly. Journal of Human-Robot Interaction 2(2): 3–32
6. Kidd CD, Breazeal CL (2004) Effect of a robot on user perceptions. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004). IEEE, Piscataway, N.J., pp 3559–3564

7. Kiesler SB, Powers A, Fussell SR et al. (2008) Anthropomorphic Interactions with a Robot and Robot-Like Agent. Soc Cognition 26(2): 169–181. doi: 10.1521/soco.2008.26.2.169
8. Kennedy J, Baxter P, Belpaeme T Comparing Robot Embodiments in a Guided Discovery Learning Interaction with Children. Int J of Soc Robotics 7(2): 293–308. doi: 10.1007/s12369-014-0277-4
9. Bartneck C (2003) Interacting with an embodied emotional character. In: Proceedings of the International Conference on Designing Pleasurable Products and Interfaces. ACM Press, New York, NY, pp 55–60
10. Hoffmann L, Krämer NC (2013) Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts. Int J Hum-Comput St 71(7-8): 763–774. doi: 10.1016/j.ijhcs.2013.04.007
11. Bainbridge WA, Hart JW, Kim ES et al. (2011) The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. Int J of Soc Robotics 3(1): 41–52. doi: 10.1007/s12369-010-0082-7
12. Fischer K, Lohan KS, Foth K (2012) Levels of embodiment: Linguistic analyses of factors influencing HRI. In: Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI'12), pp 463-470
13. Giles H (1973) Accent mobility: A model and some data. Anthropol Linguis 15: 87–105
14. Szczepek Reed B (2010) Speech rhythm across turn transitions in cross-cultural talk-in-interaction: Special Issue: Pragmatic Perspectives on Parliamentary Discourse. J of Pragmatics 42(4): 1037–1059. doi: 10.1016/j.pragma.2009.09.002
15. Brennan SE, Clark HH (1996) Conceptual pacts and lexical choice in conversation. J of Exp Psychol: Learning, Memory, and Cognition 22: 1482–1493
16. Branigan HP, Pickering MJ, Cleland AA (2000) Syntactic co-ordination in dialogue. Cognition 75(2): B13-25
17. Branigan HP, Pickering MJ, Pearson J et al. (2010) Linguistic alignment between people and computers. J Pragmatics 42(9): 2355–2368. doi: 10.1016/j.pragma.2009.12.012
18. Rosenthal-von der Pütten AM, Wiering L, Krämer NC et al. (2013) Great minds think alike. Experimental study on lexical alignment in human-agent interaction. i-com 12(1): 32–38. doi: 10.1524/icom.2013.0005
19. Kühne V, Rosenthal-von der Pütten AM, Krämer NC (2013) Using Linguistic Alignment to Enhance Learning Experience with Pedagogical Agents: The Special Case of Dialect. In: Intelligent Virtual Agents, vol 8108. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 149–158
20. Bergmann K, Branigan HP, Kopp S (2015) Exploring the Alignment Space – Lexical and Gestural Alignment with Real and Virtual Humans. Front. ICT 2: 7. doi: 10.3389/fict.2015.00007
21. Wunderlich H (2012) Talking like a machine?! Linguistic alignment of native-speakers and non-native speakers in interaction with a virtual agent (bachelor theses), anonymized for review
22. Costa A, Pickering MJ, Sorace A (2008) Alignment in second language dialogue. Lang Cognitive Proc 23(4): 528–556. doi: 10.1080/01690960801920545

23. Atkinson D, Churchill E, Nishino T et al. (2007) Alignment and Interaction in a Sociocognitive Approach to Second Language Acquisition. Mod Lang J 91(2): 169–188. doi: 10.1111/j.1540-4781.2007.00539.x
24. Bartneck C, Kulić D, Croft E et al. (2009) Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int J of Soc Robotics 1(1): 71–81. doi: 10.1007/s12369-008-0001-3
25. Nowak KL, Biocca F (2003) The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. PRESENCE-Teleop Virt 12(5): 481–494. doi: 10.1162/105474603322761289
26. Reuter M (2011) Linguistic alignment with virtual agents. Bachelor Thesis. University of Duisburg-Essen
27. Baur RS, Meder G (1994) C-Tests zur Ermittlung der globalen Sprachfähigkeit im Deutschen und einer Muttersprache bei ausländischen Schülern in der Bundesrepublik Deutschland. In: Grotjahn R (ed) Der C-Test. Theoretische Grundlagen und praktische Anwendungen.
28. Wagenmakers EJ (2007) A practical solution to the pervasive problems of p values. Psychonomic Bulletin & Review 14(5): 779-804