ORIGINAL PAPER

Non-native speaker perception of Intelligent Virtual Agents in two languages: the impact of amount and type of grammatical mistakes

David Obremski[1] · Jean-Luc Lugrin[1] · Philipp Schaper[1] · Birgit Lugrin[1]

Abstract

Having a mixed-cultural membership becomes increasingly common in our modern society. It is thus beneficial in several ways to create Intelligent Virtual Agents (IVAs) that reflect a mixed-cultural background as well, e.g., for educational settings. For research with such IVAs, it is essential that they are classified as non-native by members of a target culture. In this paper, we focus on variations of IVAs' speech to create the impression of non-native speakers that are identified as such by speakers of two different mother tongues. In particular, we investigate grammatical mistakes and identify thresholds beyond which the agents is clearly categorised as a non-native speaker. Therefore, we conducted two experiments: one for native speakers of German, and one for native speakers of English. Results of the German study indicate that beyond 10% of word order mistakes and 25% of infinitive mistakes German-speaking IVAs are perceived as non-native speakers. Results of the English study indicate that beyond 50% of omission mistakes and 50% of infinitive mistakes English-speaking IVAs are perceived as non-native speakers. We believe these thresholds constitute helpful guidelines for computational approaches of non-native speaker generation, simplifying research with IVAs in mixed-cultural settings.

Keywords Intelligent Virtual Agents · Verbal behaviour · Mixed-cultural settings

1 Introduction

Through globalization and migration meeting people from different cultures has become omnipresent in our everyday lives. In the European Union, for example, the number of first and second generation immigrants was 55 million (with an overall population of 507 million) in 2014. Due to long-term or close contact with a new culture, customs are often adopted and communication is adapted. Thus, a large number of people today have mixed-cultural memberships, which means that different cultural believes, values and communication styles might be present in one and the same person at the same time.

Implementing Intelligent Virtual Agents (IVAs) that simulate a mixed-cultural background can be beneficial in many ways. Besides reflecting the cultural diversity of a modern society they can be used for various training settings or to study phenomena occurring in interactions with people with mixed-cultural memberships. Trainings with such IVAs can systematically expose people to specific scenarios and raise cultural sensitivity. Scenarios with trainings for class rooms with children of migrants, help-desks in airports with people of various cultural backgrounds interacting, or for receptionists at clinics with patients from many mixed-cultural backgrounds are just a few examples.

However, designing convincing IVAs with mixed-cultural backgrounds can be resource-consuming as many aspects, e.g. clothing style, visual appearance, language variations, differences in non-verbal behaviour, are supposedly contributing to the fact that they might be perceived as foreign. In this paper we argue that instead of implementing the whole complexity of a mixed-cultural membership, a single aspect

---

[1] https://ec.europa.eu/eurostat/statistics-explained/index.php/First_and_second-generation_immigrants_-_statistics_on_main_characteristics.

✉ David Obremski
david.obremski@uni-wuerzburg.de

Jean-Luc Lugrin
jean-luc.lugrin@uni-wuerzburg.de

Philipp Schaper
philipp.schaper@uni-wuerzburg.de

Birgit Lugrin
birgit.lugrin@uni-wuerzburg.de

[1] Human-Computer Interaction, University Würzburg, Würzburg, Germany

might be sufficient to trigger the impression of a non-native IVA. This might be important, as it could drastically reduce time and cost for the realization of training prototypes. We chose speech as behavioural factor, since it is the primary channel for humans to communicate (e.g., [12,21]) and can be altered for our needs by rather simple inductions of grammatical mistakes.

In this paper, we demonstrate the successful simulation of non-native-speaking IVAs in two different languages by inducing grammatical mistakes at various rates. The first part of this paper is summarising the related work, while the second part is describing the implementation of our *Multi-Cultural IVA Configuration and Simulation* tool. In the third part, we present two studies investigating the perception of IVAs with different grammatical proficiency in both German and English. Finally, we conclude this paper by discussing our results and suggesting topics for future work on the subject.

2 Related work

Culture as a social aspect has been investigated in the domain of Intelligent Virtual Agents (IVAs) for over two decades. To highlight certain culture-related differences or to raise the acceptance of IVAs for certain cultural backgrounds, different aspects of culture-related behaviours have been investigated, such as topic selection in small talk conversations [8], communication management behaviours [6], non-verbal behaviours [7] or rituals [19].

A major area of application of these enculturated IVAs is to teach about culture. Looking at culture on a national level, Valente, Johnson and Vilhjálmsson, for example, introduced a framework called *The Tactical Language and Culture Training System*. The idea of the framework is to teach language, culture and gestures in natural settings [27] by allowing users to multimodally interact with IVAs of a predefined national cultural background. Several scenarios have been introduced in this context, e.g., to prepare US soldiers to successfully interact with people from Iraq or Afghanistan [14], or to help native English speaking business people to interact with Chinese business partners [15]. Another line of research used fantasy cultures that were developed along cultural dichotomies in order to raise a general cultural awareness in users [1,2,20]. Please note that in the approaches introduced above, foreign agents speak and act in a culturally consistent manner.

Focusing on different ethnicities in the US American culture, different variations of speaking styles have been modelled [10,13]. In particular, IVAs were designed to match the speaking style of their interaction partner to help African American children to enhance their learning experience. According to research from the communication sciences,

a match of the teacher's and the learner's accent has a positive impact on the learner's performance [11]. Therefore, Iacobelli and Cassell implemented a virtual peer for African American and Caucasian American children in an educational context. They investigated whether it is possible to change the perceived ethnicity of the virtual peer by changing its verbal and non-verbal behaviour. Their findings indicate that children are able to correctly assess whether a virtual peer is an African American or a Caucasian American speaker, when the virtual peer displays the respective verbal and non-verbal behaviour [13]. Based on these findings, Finkelstein et al. used that virtual peer to investigate whether African American children show an improvement in learning when interacting with an agent that matches their ethnicity by speaking African American Vernacular English, as opposed to Mainstream American English. They found a clear improvement in the children's performance when they where interacting with a culturally matching virtual peer [10].

Research that explicitly targets mixed-cultural membership for IVAs is rare. An example includes work by Lugrin et al., who investigated a specific behavioural phenomenon that typically occurs in mixed-cultural settings [18]. While immigrants usually have to take great efforts to adapt to their host culture [4,24], a general adaptation also takes place on the native interlocutors side when interacting with people having a mixed-cultural background. The so-called Foreigner Talk refers to the fact that native speakers tend to adapt their behaviour by using slower and simplified speech, and increased usage of gestures to foster understanding of the non-native speaker [9] (referred to as Adapted Foreigner-directed communication (AFC) in [18] to highlight its multi-modal nature). To investigate whether AFC is also triggered in native speakers who interact with a mixed-cultural IVA, Lugrin et al. altered an IVA's appearance and several aspects of verbal as well as non-verbal behaviour [18]. With it, they successfully showed that both the verbal and non-verbal behaviour of native speakers differ when interacting with the IVA with a simulated mixed-cultural background.

Taking the (mixed-)cultural background of the user into account when interacting with a mixed-cultural IVA, Khooshabbeh et al. [16] found that the cultural background of the listener plays a crucial role in the perception of these agents. In particular, IVAs speaking different accents (using prerecorded human speech) within the US-American language, e.g., Middle Eastern English, were perceived as being foreign by people that do not share the agent's simulated mixed background but triggered a positive effect in terms of perceived shared social identity in people who share a mixed background (e.g. being bi-cultural).

Compared to the works on mixed-cultural IVAs described above, we aim at isolating one behavioural channel to focus on particular aspects therein that qualify for an eventual

automatic generation of behavioural variations that will trigger the impression of a non-native background. We found verbal behaviour particularly suited, as it constitutes the primary channel of human communication and has shown to be effective in previous work, e.g. [16], for prerecorded speech. Particularly identifying thresholds for common grammatical mistakes that are sufficient to trigger the impression of a non-native speaker could help form the basis for a rather simple automatic approach and with it lead to cost-effective generation of non-native IVAs.

3 Implementation

We developed a rapid prototyping tool for easily adapting the appearance and behaviour of an IVA in a virtual setting, using the game engine Unity (2017.2.5).[1] The tool features a variety of previously created content to choose from to adapt an IVA. It is based on the virtual environment used by Lugrin et al. [18] and supports two new modules: I) *Cultural Agent Configuration* and II) *Scenario Builder*. The *Cultural Agent Configuration* is responsible for changing the appearance, as well as the verbal and non-verbal behaviour of an IVA. For instance, the tool allows to choose between two different sets of predefined gestures, either foreign or local gestures, or different ethnicity appearances. The *Scenario Builder* allows to quickly build, save and play a conversation by constructing a temporal sequence of verbal and non-verbal behaviours. Each step in the sequence contains information about the gesture and the speech sample, as well as two optional fields to adjust the synchronicity of speech and gesture execution.

The most crucial part for the work described in this paper are the agent's language settings that are part of the *Agent Cultural Configuration*. They allow to quickly change the agent's language proficiency by inducing two types of common mistakes for two languages. For designing a study using the tool, the experimenter can adjust the level of the respective parameter: For Experiment 1, level 3 represents the *absence* of mistakes, level 2 represents a *moderate* amount of mistakes, and level 1 represents a *high* amount of mistakes. Depending on the selected levels, the system will build the agent's dialogues. For instance, when moving the 'word order' slider to level 3, and the 'infinitive' slider to level 1, a set of synthetic speech templates will be assigned to the IVA's speech output with no mistakes regarding word order, but with a high amount of mistakes in terms of infinitive use. The agent's language settings of the tool were extended for Experiment 2 to contain more levels (five to be precise), and another type of grammatical mistake (omission mistakes). In addition, new speech templates were added.
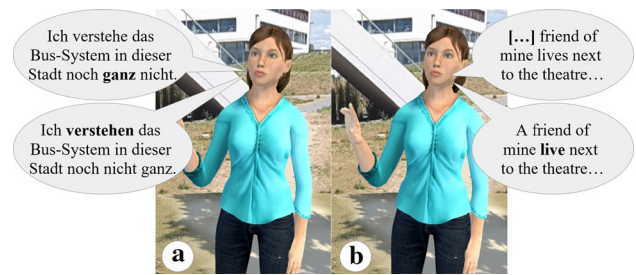

Fig. 1 **a** Example of the word order and infinitive mistakes in Experiment 1, **b** example of the omission and infinitive mistakes in Experiment 2 (mistakes highlighted in bold)

3.1 Agent's language settings for Experiment 1

For the German version used in Experiment 1 the language proficiency of the IVA was altered in two different dimensions: (I) *word order mistakes* and (II) *infinitive mistakes*. According to Csehó these mistakes are commonly conducted by non-native speakers of the German language [5]. Each dimension had two levels: one that represented a low amount of mistakes (10% in word order, 25% in infinitive use) and one that represented a high amount of mistakes (20% in word order, 50% in infinitive use). Additionally, one scene where the agent spoke perfect German with no grammatical mistakes was implemented. The five resulting videos are presented in Table 6.

The error percentages were calculated as follows: 10% of word order mistakes means that 10% of the words in the entire text are misplaced. 20% of infinitive mistakes means that 20% of the verbs occurring in the text, that should be conjugated, are put in their infinitive form. Albeit the fact that the German language is very robust to changes of word order [3], it is important to note that changing the order of the words can considerably affect the meaning of a sentence. Therefore the selected percentages of the *word order mistake* are lower than the percentages of the *infinitive mistake*, and words were only shifted within a part of the sentence, allowing to preserve the meaning units of the sentence while introducing grammatical mistakes. Fig. 1a) shows an example with a short passage of the text in German. Tables 1 and 2 show the complete transcriptions of the respective texts.

3.2 Agent's language settings for Experiment 2

For the English version used in Experiment 2 the language proficiency of the IVA was altered in two different dimensions: I) *omission mistakes* and II) *infinitive mistakes*. Various studies show that leaving out articles and prepositions, or keeping verbs in the infinitive form where they should be conjugated, are common mistakes of non-native speakers of English [17,23,26]. Each dimension had four levels: One that represented a low amount of mistakes (10% in omission

Table 1 The different texts with induced *Word order mistakes* in German (mistakes are underlined)

Perfect German	Low order mistake	High order mistake
Hallo, ich bin neu hier. Ich verstehe das Bus-System in dieser Stadt noch nicht ganz. Kannst du mir sagen, mit welcher Linie ich in die Stadt komme? Mit der Linie 10 komme ich in die Sanderau, das weiß ich. Aber kannst du mir erklären, wie ich in die Stadt komme?	Hallo, ich bin neu hier. Ich verstehe das Bus-System in dieser Stadt noch ganz nicht. Kannst du mir sagen, mit welcher ich Linie in die Stadt komme? Mit der Linie 10 komme ich in die Sanderau, das ich weiß. Aber du kannst mir erklären, wie in die Stadt ich komme?	Hallo, ich neu hier bin. Ich das Bus-System in dieser Stadt verstehe noch ganz nicht. Du kannst mir sagen, mit welcher ich Linie in die Stadt komme? Linie 10, mit der komme ich in die Sanderau, das ich weiß. Aber du kannst mir yerklären, in die Stadt wie ich komme?

Table 2 The different texts with induced *Infinitive mistakes* in German (mistakes are underlined)

Perfect German	Low infinitive mistake	High infinitive mistake
Hallo, ich bin neu hier. Ich verstehe das Bus-System in dieser Stadt noch nicht ganz. Kannst du mir sagen, mit welcher Linie ich in die Stadt komme? Mit der Linie 10 komme ich in die Sanderau, das weiß ich. Aber kannst du mir erklären, wie ich in die Stadt komme?	Hallo, ich bin neu hier. Ich verstehen das Bus-System in dieser Stadt noch nicht ganz. Kannst du mir sagen, mit welcher Linie ich in die Stadt kommen? Mit der Linie 10 komme ich in die Sanderau, das weiß ich. Aber kannst du mir erklären, wie ich in die Stadt komme?	Hallo, ich bin neu hier. Ich verstehen das Bus-System in dieser Stadt noch nicht ganz. Können du mir sagen, mit welcher Linie ich in die Stadt kommen? Mit der Linie 10 komme ich in die Sanderau, das weiß ich. Aber können du mir erklären, wie ich in die Stadt komme?

Table 3 The text in *Perfect English*

Perfect English
Hi, I'm new in town but I haven't quite figured out how public transport works here. Can you tell me which bus goes into town? A friend of mine lives next to the theatre and I'd like to visit him. I have to be there by 8 o'clock. He told me which bus to get on but I have forgotten. I know that I can get to the train station on the number 10, but can you tell me how to get into town?

and infinitive use), one that represented a medium amount of mistakes (25% in omission and infinitive use), one that represented a high amount of mistakes (50% in omission and infinitive use), and one that represented a very high amount of mistakes (75% in omission and infinitive use). An overview of the nine resulting videos is presented in Table 7.

Analogously to Experiment 1, 20% of infinitive mistakes means that 20% of all verbs that should be conjugated occur in their infinitive form. 20% of omission mistakes means that 20% of all prepositions and articles within the text were omitted. Additionally, one scene where the agent spoke perfect English with no grammatical mistakes was implemented (see Table 3). Since the authors are non-native speakers of English, we cooperated with English native speakers at the English department of the University of Würzburg.

The complete transcriptions of the different texts for both grammatical error types can be found in Tables 4 and 5. In contrast to Experiment 1, in German the same thresholds were applied to each type of mistake because neither of the grammatical mistakes is likely to considerably affect the meaning of the sentence.

4 Experiment 1: non-native speaker perception in German

To identify a language proficiency threshold below which IVAs are perceived as non-native speakers of German, we conducted a within-subjects online experiment [22]. Participants watched five videos in randomised order, each of which showed the same IVA (*representing a young Caucasian female student on a university campus*) speaking a monologue (consisting of fifty-one words) with five different levels of language proficiency. During her monologue, the IVA asked the participant to explain which bus she should take to get into town.

Independent variable The independent variable was the language proficiency of the IVA shown in the videos, that were created using the prototyping tool. Table 6 shows the different expressions and resulting conditions of the independent variable in Experiment 1.

Dependent variable The participants had to categorise the IVA as a native or non-native speaker by indicating their agreement to the following statement for each video: ***"The agent is a native speaker" Yes or No***. Optional fields were also included that allow to insert a justification for their answer. We decided to limit the participants to these two options, rather than allowing them to choose from various degrees of nativeness (e.g., *"to what degree do you perceive*

Table 4 The different texts with induced *Omission mistakes* in English (mistakes are underlined)

Low omission mistake	Medium omission mistake
Hi, I'm new in town but I haven't quite figured out how public transport works here. Can you tell me which bus goes into town? [...] friend of mine lives next to the theatre and I'd like to visit him. I have to be there by 8 o'clock. He told me which bus to get on but I have forgotten. I know that I can get to the train station on the number 10, but can you tell me how to get into town?	Hi, I'm new in town but I haven't quite figured out how public transport works here. Can you tell me which bus goes into town? [...] friend of mine lives next to the theatre and I'd like to visit him. I have to be there [...] 8 o'clock. He told me which bus to get on but I have forgotten. I know that I can get to [...] train station on the number 10, but can you tell me how to get into town?

High omission mistake	Very high omission mistake
Hi, I'm new [...] town but I haven't quite figured out how public transport works here. Can you tell me which bus goes into town? [...] friend of mine lives next to [...] theatre and I'd like to visit him. I have to be there [...] 8 o'clock. He told me which bus to get [...] but I have forgotten. I know that I can get to [...] train station on the number 10, but can you tell me how to get into town?	Hi, I'm new [...] town but I haven't quite figured out how public transport works here. Can you tell me which bus goes [...] town? [...] friend of mine lives next to [...] theatre and I'd like to visit him. I have to be there [...] 8 o'clock. He told me which bus to get [...] but I have forgotten. I know that I can get to [...] train station on [...] number 10, but can you tell me how to get [...] town?

Table 5 The different texts with induced *Infinitive mistakes* in English (mistakes are underlined)

Low infinitive mistake	Medium infinitive mistake
Hi, I'm new in town but I haven't quite figure out how public transport works here. Can you tell me which bus goes into town? A friend of mine lives next to the theatre and I'd like to visit him. I have to be there by 8 o'clock. He told me which bus to get on but I have forgotten. I know that I can get to the train station on the number 10, but can you tell me how to get into town?	Hi, I'm new in town but I haven't quite figure out how public transport works here. Can you tell me which bus goes into town? A friend of mine lives next to the theatre and I'd like to visit him. I have to be there by 8 o'clock. He tell me which bus to get on but I have forgotten. I know that I can get to the train station on the number 10, but can you tell me how to get into town?

High infinitive mistake	Very high infinitive mistake
Hi, I'm new in town but I haven't quite figure out how public transport works here. Can you tell me which bus go into town? A friend of mine live next to the theatre and I'd like to visit him. I have to be there by 8 o'clock. He tell me which bus to get on but I have forgotten. I know that I can get to the train station on the number 10, but can you tell me how to get into town?	Hi, I'm new in town but I haven't quite figure out how public transport work here. Can you tell me which bus go into town? A friend of mine live next to the theatre and I will like to visit him. I have to be there by 8 o'clock. He tell me which bus to get on but I have forgotten. I know that I can get to the train station on the number 10, but can you tell me how to get into town?

Table 6 Experimental conditions for Experiment 1

Video	Word order mistake (%)	Infinitive mistake (%)	Condition
1	0	0	Perfect German
2	10	0	Low order mistake
3	20	0	High order mistake
4	0	25	Low infinitive mistake
5	0	50	High infinitive mistake

the virtual agent to be native speaker"). This decision was motivated by our intention to identify error thresholds allowing to generate non-native speaking IVAs later.

4.1 Results

Overall 34 participants took part in the first experiment with 13 males, 20 females and one gender neutral person (M_{age} = 26.4, SD_{age} = 8.3). All participants were recruited via e-mail. The results of the *word order mistake* conditions and the *infinitive mistake* conditions are presented in Figs. 2 and 3 respectively. A Cochran's Q test revealed a significant main effect for the *word order mistake* condition ($\chi^2(2) = 54.07$, $p < .001$). A Bonferroni adjusted post-hoc Dunn's test revealed a significant difference between the *perfect German* and each of the two faulty conditions (both $p < .001$) but not between the two faulty conditions. For the *infinitive mistakes* there was also a significant main effect, indicating
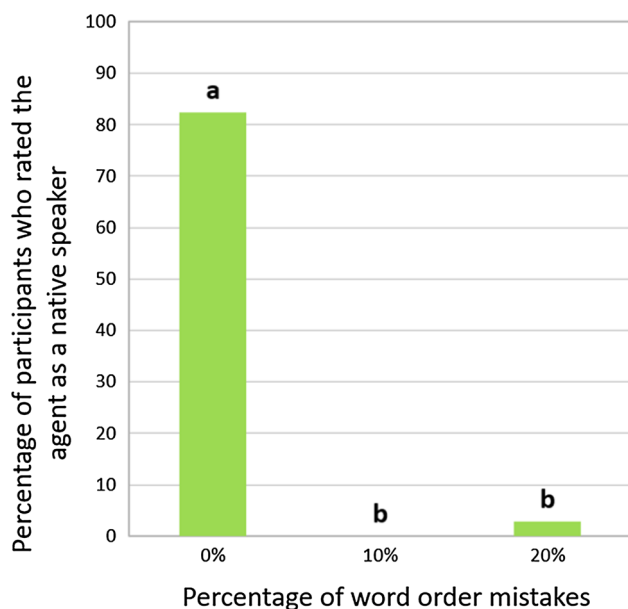
Fig. 2 Results of the *word order mistake* conditions in Experiment 1. Values with different letters are statistically different ($p < .001$)

Fig. 3 Results of the *infinitive mistake* conditions in Experiment 1. Values with different letters are statistically different ($p < .001$)

a difference between the three conditions ($\chi^2(2) = 37.27$, $p < .001$). Again, a Bonferroni adjusted post-hoc Dunn's test revealed differences between the *perfect German* and each of the two faulty conditions (both $p < .001$), while no significant difference between the two faulty conditions was found.

4.2 Discussion

Overall, the results indicate that even with small modifications of the language proficiency, the impression of a non-native IVA can be achieved in the German language. This implies that in order to implement a non-native speaking IVA slight changes of word order or few infinitive mistakes are sufficient, and a larger amount of mistakes seems unnecessary.

Although there are no significant differences between the faulty conditions, our results show a smaller difference between the faulty word order conditions (10% and 20%), relative to the faulty infinitive mistake conditions (25% and 50%), c.f. Figs. 2 and 3. This suggests that a clear threshold can be identified below which IVAs are perceived to be non-native speakers. However, a finer grated placement of grammatical mistakes might have shown that the classification of nativeness in a language is ambiguous in certain scenarios.
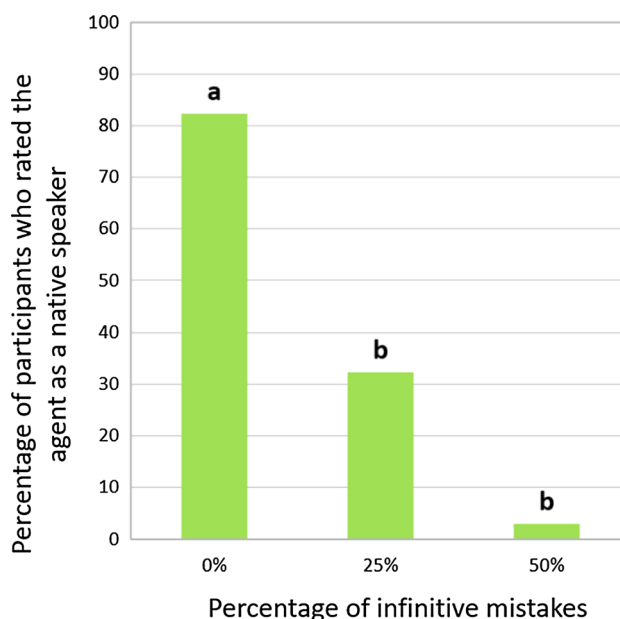
5 Experiment 2: non-native speaker perception in English

Experiment 1 was replicated for the English language to generate results with a broader impact for the IVA community. We also added finer grated levels in grammatical mistakes, a larger pool of participants, and other minor improvements. Fig. 1b displays an example with a short passage of the English text.

Additional levels contain finer variations in the number of grammatical mistakes in order to find out whether we can identify a clear threshold above which IVAs are perceived as non-native speakers of English. Analogously to Experiment 1 participants watched nine videos in randomised order, each of which showed the same IVA speaking a monologue (consisting of eighty-seven words) with nine different levels of language proficiency.

To be able to find potential small effects between the different conditions, we decided on the continued use of a within-participant design with randomized presentation of stimuli to maximize statistical power. We weighed the benefit of potentially finding differences between previously not statistically differing groups against participants guessing at the aim of the study. Because we did not expect adverse effects from the latter, but considered group differences as critical for identification of thresholds, a within-design was used.

Participants were recruited using an established online recruitment platform (prolific.co). The site uses a variety of filters and methods to prevent people from creating mul-

Table 7 Experimental conditions for Experiment 2

Video	Omission mistake (%)	Infinitive mistake (%)	Condition
1	0	0	Perfect English
2	10	0	Low omission mistake
3	25	0	Medium omission mistake
4	50	0	High omission mistake
5	75	0	Very high omission mistake
6	0	10	Low infinitive mistake
7	0	25	Medium infinitive mistake
8	0	50	High infinitive mistake
9	0	75	Very high infinitive mistake

tiple accounts. All participants stated to live in the United Kingdom and stated having English as their first language. To verify the participant's statement, four questions testing the language proficiency were included at the beginning of the survey. These questions were selected from a language assessment test in consultation with the English department of the University of Würzburg. Participants were excluded from the study if they failed at two or more of these four native-speaker questions. *Independent variables* The independent variable was the language proficiency of the IVA shown in the videos, that were created using the prototyping tool. Table 7 shows the different expressions and resulting conditions of the independent variable in Experiment 2.

Dependent variable Analogously to Experiment 1, participants were asked whether they think the IVA was a native speaker of English. Participants had to indicate their agreement with the following statement for each video: ***"The agent is a native speaker" Yes or No***. Again, optional fields to insert a justification for their answer were included. In case the participant categorised the agent as a non-native speaker of English, an additional input field appeared where the participant was asked to guess the agent's mother tongue. This field was added to investigate whether the different grammatical mistakes just evoke the general impression of a non-native speaker of English, or whether they might lead to the assumption of a speaker from a certain country.

5.1 Results

Overall 197 participants took part in the study with 79 males and 118 females (M_{age} = 37.94, SD_{age} = 12.09). Out of the 219 participants that initially signed up for the study, 22 were excluded because they failed at native speaker test. The detailed results of the *omission mistake* conditions and the *infinitive mistake* conditions are presented in Figs. 4 and 5 respectively. A Cochran's Q test revealed a significant main effect for the *omission mistake* conditions ($\chi^2(4)$ = 422.83, $p < .001$). A Bonferroni adjusted post-hoc Dunn's test

revealed significant differences in native speaker ratings between each of the conditions (all $p < .001$), except for 50% versus 75% omission mistakes.

For the *infinitive mistakes* a Cochran's Q test also revealed a significant difference between the different conditions ($\chi^2(4)$ = 359.66, $p < .001$). A Bonferroni adjusted post-hoc Dunn's test revealed significant differences in native speaker ratings in eight of the ten pairwise comparisons ($p = .014$ for 25% versus 75%, $p < .001$ for all other comparisons). The post-hoc tests showed no significant difference between the 25% and the 50% infinitive mistakes conditions, and no significant difference between the 50% and the 75% infinitive mistakes conditions.

When asked for the assumed agent's mother tongue, after classifying her as a non-native speaker of English, the majority of participants stated that they were unable to indicate any. In the optional comment field at the end of the study, eighteen participants explicitly stated that it was impossible for them to guess the agent's mother tongue just based on the grammatical mistakes. Several of these participants added that it would have been easier if the agent had had a certain accent.

5.2 Discussion

The results of Experiment 2 suggest that with 50% mistakes and above, IVAs are consistently perceived as non-native speakers of English, as no more significant difference in classification as a non-native speaker can be observed from there. Interestingly, the finer gradation of the amount of errors integrated in the IVA's speech results in a linear progression of the classification of the IVA as a non-native speaker. The more mistakes are integrated in the IVA's speech, the more likely it is classified as a non-native speaker.

This suggests that also a finer categorisation of non-native speaker perception in terms of language proficiency could be possible. For example, people might differentiate between "not native, but at a very high level" and "not native, at a beginner level".
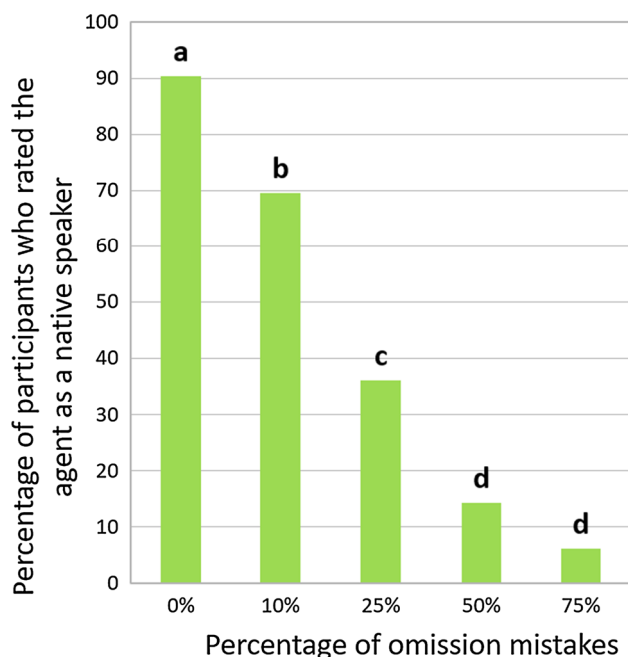
Fig. 4 Results of the *omission mistake* conditions in Experiment 2. Values with different letters are statistically different ($p < .001$)
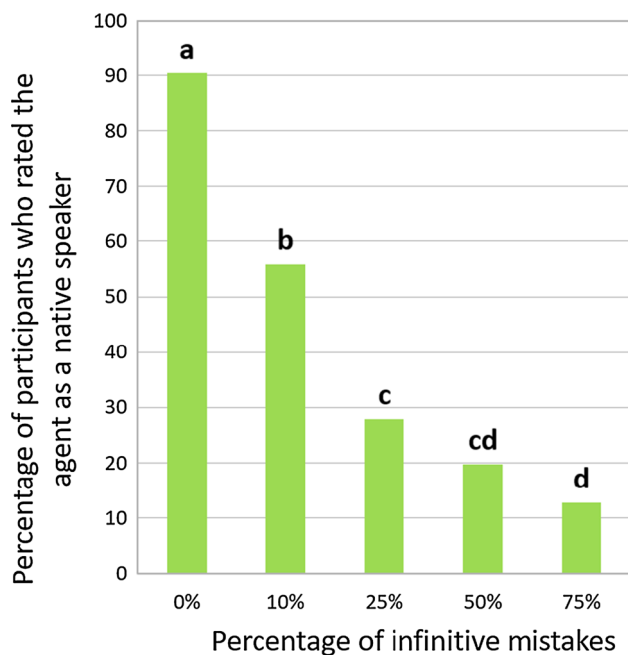

Fig. 5 Results of the *infinitive mistake* conditions in Experiment 2. Values with different letters are statistically different ($p < .015$)

6 General discussion and future work

Both experiments suggest that with rather simple variations in speech, the impression of a non-native speaker can be triggered. In particular, the results of Experiment 1 indicate that even slight changes of the word order (10% of mistakes), and

only few infinitive mistakes (25%) result in the impression that the agent is a non-native speaker of German. The results of Experiment 2 indicate that by increasing the number of grammatical mistakes, the number of participants who classify the IVA as a non-native speaker of English increases respectively. We suggest a threshold of above 50% for both omission mistakes and infinitive mistakes for sufficient non-native speaker perception for the English language.

When comparing Experiment 2 to Experiment 1, the identified threshold above which an IVA is considered as a non-native speaker is noticeably higher. It seems that native speakers of English are more tolerant towards violations of grammar than native speakers of German are. This could be due the fact that native speakers of English are more used to listening to foreigners speaking in their mother tongue than Germans are, since there are many countries such as India or Pakistan where English is not the primary but still an official language.[2] This can lead to a modified version of English that differs from British and American English in grammar and vocabulary [25]. In total there are approximately one billion people that speak English as their first or second language[3] while there are only around 130 million people that speak German as their first or second language.[4] The substantially higher number of English speaking people and the resulting variations of the English language alone could explain the observed higher tolerance towards grammatical mistakes in English.

A limitation of our two experiments lies in the fact, that we have shown videos of IVAs instead of providing an interactive setting. Nevertheless, we were able to identify thresholds and trigger the effect of non-native speaker perception in both languages. In the future, we will test whether our results hold in an interactive setting with the user having a conversation with the IVA in a virtual reality environment.

In order to evaluate the impact of certain grammatical mistakes on the native speaker perception of an IVA, we isolated different grammatical mistakes and induced them in an IVA's speech. However, in a real world mixed-cultural setting grammatical mistakes are likely to appear in combination. In future research, we will also investigate the impact of combined grammatical mistakes instead of isolated ones.

There are also other types of speech errors that could be implemented to result in IVAs which are identified as non-native speakers. The error types chosen seemed particularly suited as they were identified as being common in the respective language, and can be applied to any text with rather

---

simple automatic approaches in the future, quickly enabling us to generate future experiments.

Considering that particularly Experiment 2 indicated that a finer categorisation of non-nativeness could potentially be identified, we will continue our series of experiments, allowing for more classifications of native vs. non-native. Because non-native speakers differ from native speakers not only by grammatical mistakes but also by incorrect pronunciation, we also plan to investigate the impact of synthetic accents on the perception of IVAs. Inducing synthetic accents could potentially lead to a way of generating non-native IVAs that need to represent a speaker of a specific mother-tongue. Multiple participants in Experiment 2 stated that it is impossible to guess an IVA's native language by just listening to its grammatical mistakes but also noted that a synthetic accent could give more insight into the IVA's intended mother-tongue.

As an overall goal, we aim at automatically generating speech for IVAs that are perceived as non-native speakers. The thresholds we have identified in this contribution serve as a valuable guideline on how to build such a computational model. In such a model, grammatical mistakes need to be carefully inserted in existing text to ensure to not break the meaning units. Having such a generative model at hand will allow us to dynamically create non-native speaker simulations for a variety of contexts where an efficient implementation of a foreign IVA is crucial.

7 Conclusion

In this paper, we argue that implementing Intelligent Virtual Agents (IVAs) that simulate a mixed-cultural background can be beneficial in many ways, for example for training scenarios, or to study behavioural phenomena occurring in mixed cultural settings.

While previous research has shown that it is possible to generate convincing mixed-cultural IVAs by altering the IVA's appearance and verbal as well as non-verbal behaviour, or using different accents by means of pre-recorded speech, the implementation of such approaches is resource consuming. In this paper, we present an approach to evoke the impression of a non-native speaking IVA by solely altering its verbal behaviour by inducing grammatical mistakes. In two experiments, we showed that we can trigger the perception of a non-native IVA in two different languages (German and English). Particularly in this paper, we identified thresholds of grammatical mistakes above which IVAs are consistently perceived as non-native speakers.

With these thresholds, we will be able to implement a rather simple approach to generate non-native IVAs in the future. With it, we aim to not only facilitate the cost-effective design of non-native speaking IVAs for mixed-cultural settings to be used in a variety of cultural training systems or other systems that take place in mixed cultural settings. Our findings could also be of interest to the gaming industry, where there may be a need for characters who speak the player's language but appear to be from another country.

References

1. Aylett R, Hall L, Tazzyman S, Endrass B, André E, Ritter C, Nazir A, Paiva A, Hofstede G, Kappas A (2014) Werewolves, cheats, and cultural sensitivity. In: International conference on autonomous agents and multiagent systems (AAMAS 2014). IFAAMAS, pp 1085–1092

2. Aylett R, Vannini N, Andre E, Paiva A, Enz S, Hall L (2009) But that was in another country: agents and intercultural empathy. In: 8th international conference on autonomous agents and multiagent systems (AAMAS 2009). IFAAMAS, pp 329–336

3. Bader M, Haeussler J (2010) Word order in German: a corpus study. Lingua 120(3):717–762. https://doi.org/10.1016/j.lingua.2009.05.007

4. Berry JW (1997) Immigration, acculturation, and adaptation. Appl Psychol 46(1):5–34

5. Csehó T (2009) Zum Einfluss des Foreigner Talk auf die Entstehung des Migrantendeutsch. Germanistische Studien 7:153–168

6. Endrass B, André E, Huang L, Gratch J (2010) A data-driven approach to model culture-specific communication management styles for virtual agents. In: van der Hoek W, Kaminka GA, Lespérance Y, Luck M, Sen S (eds) 9th international conference on autonomous agents and multiagent systems (AAMAS 2010), vol 1. IFAAMAS, pp 99–108. http://dblp.uni-trier.de/db/conf/atal/aamas2010.html#EndrassAHG10

7. Endrass B, André E, Rehm M, Nakano Y (2013) Investigating culture-related aspects of behavior for virtual characters. Auton Agents Multi Agent Syst 27(2):277–304

8. Endrass B, Nakano YI, Lipi AA, Rehm M, André E (2011) Culture-related topic selection in small talk conversations across Germany and Japan. In: Vilhjálmsson HH, Kopp S, Marsella S, Thórisson KR (eds) 11th international conference on Intelligent Virtual Agents (IVA 2011), LNCS, vol 6895. Springer, pp 1–13. http://dblp.uni-trier.de/db/conf/iva/iva2011.html#EndrassNLRA11

9. Ferguson CA (1975) Toward a characterization of English foreigner talk. Anthropol Linguist 17(1):1–14

10. Finkelstein S, Yarzebinski E, Vaughn C, Ogan A, Cassell J (2013) The effects of culturally congruent educational technologies on student achievement. In: Lane HC, Yacef K, Mostow J, Pavlik P (eds) Artificial intelligence in education. Springer, Berlin, pp 493–502

11. Gill MM (1994) Accent and stereotypes: their effect on perceptions of teachers and lecture comprehension. J Appl Commun Res 22(4):348–361. https://doi.org/10.1080/00909889409365409

12. Gumperz JJ (2009) The speech community. In: Duranti A (ed) Linguistic anthropology: a reader. Wiley Linguistic anthropology: a reader, vol 1, p 66

13. Iacobelli F, Cassell J (2007) Ethnic identity and engagement in embodied conversational agents. In: 7th international conference on Intelligent Virtual Agents (IVA '07). Springer, pp 57–63

14. Johnson WL, Valente A (2008) Collaborative authoring of serious games for language and culture. In: Proceedings of SimTecT, vol 2008

15. Johnson WL, Zaker SB (2012) The power of social simulation for Chinese language teaching. In: Proceedings of TCLT7

16. Khooshabeh P, Dehghani M, Nazarian A, Gratch J (2017) The cultural influence model: when accented natural language spoken by virtual characters matters. AI Soc 32:9–16

17. Lee J, Seneff S (2008) An analysis of grammatical errors in non-native speech in English. In: 2008 IEEE spoken language technology workshop, pp 89–92. https://doi.org/10.1109/SLT.2008.4777847

18. Lugrin B, Eckstein B, Bergmann K, Heindl C (2018) Adapted foreigner-directed communication towards virtual agents. In: 18th international conference on Intelligent Virtual Agents, IVA '18. ACM, pp 59–64. https://doi.org/10.1145/3267851.3267859

19. Mascarenhas S, Dias J, Afonso N, Enz S, Paiva A (2009) Using rituals to express cultural differences in synthetic characters. In: Proceedings of The 8th international conference on autonomous agents and multiagent systems, vol 1, pp 305–312

20. Mascarenhas S, Silva A, Paiva A, Aylett R, Kistler F et al (2013) Traveller: an intercultural training system with intelligent agents. In: Proceedings of 12th international conference on autonomous agents and multiagent systems (AAMAS 2013)

21. Nass CI, Brave S (2005) Wired for speech: how voice activates and advances the human–computer relationship. MIT Press, Cambridge

22. Obremski D, Lugrin JL, Schaper P, Lugrin B (2019) Non-native speaker generation and perception for mixed-cultural settings. In: Proceedings of the 19th ACM international conference on Intelligent Virtual Agents, IVA '19. ACM, New York, NY, USA, pp 105–107. https://doi.org/10.1145/3308532.3329427.

23. Richards J (1971) Error analysis and second language strategies. The text of an invited lecture given at Indiana University, Bloomington, p 28

24. Rumbaut RG (2015) Assimilation of immigrants. James D. Wright (editor-in-chief). Int Encycl Soc Behav Sci 2:81–87

25. Sedlatschek A (2009) Contemporary Indian English: variation and change. John Benjamins. https://www.jbe-platform.com/content/books/9789027290120

26. Ting SH, Mahadhir M, Chang SL (2010) Grammatical errors in spoken English of university students in oral communication course. GEMA Online® J Lang Stud 10(1):53–70

27. Valente A, Johnson WL, Vilhjálmsson HH (2006) The tactical language and culture training system: a demonstration. In: AAAI, pp 1955–1957