

# Investigating the Perspective of Non-Native Speakers on Foreigner-Directed Speech using Virtual Agents: The Role of Racial Ingroup Affiliation and Language Proficiency on Perception and Comprehension

Ohenewa Bediako Akuffo  
Socially Interactive Agents  
University of Würzburg, Germany  
ohenewa.bediako-akuffo@uni-wuerzburg.de

Birgit Lugin  
Socially Interactive Agents  
University of Würzburg, Germany  
birgit.lugin@uni-wuerzburg.de



**Figure 1: The experimental setup: Experimenter controlling the Wizard of Oz interface (left), the Unity game view showing the Black female IVA in the scene (centre), and the four virtual agents used in the study (right).**

## ABSTRACT

Foreigner-directed speech (FDS) is a speech register in which people modify their language to assist non-native speakers (NNSs). However, little research has focused on NNSs' perceptions of FDS or its actual benefits, partly due to the high costs associated with such studies. Intelligent Virtual Agents (IVAs) are well-suited to investigating multinational and multilingual phenomena due to their high degree of standardisation and adaptability. This study, therefore, examines how NNSs perceive IVAs using FDS, whether FDS aids comprehension, and the impact of the IVA's race and the NNSs' language proficiency. The sample ( $N = 124$ ) consisted of individuals identifying as Black and possessing non-native proficiency in German. Participants interacted with either a Black or White IVA using FDS or native-directed speech (NDS). Results show that NNSs generally rated FDS more negatively than NDS, with FDS not improving comprehension. Negative perceptions of FDS increased with higher language proficiency or when the IVA was perceived as belonging to a racial outgroup. This study highlights the need to consider NNSs' perspectives in FDS research.

## CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI.**



This work is licensed under a Creative Commons Attribution International 4.0 License.

*Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).*

## KEYWORDS

Virtual Agents, Foreigner-Directed Speech, Foreigner-Talk, Non-Native Speaker Perception, Foreign Language Comprehension, In-group Effect, Mixed-Cultural

## ACM Reference Format:

Ohenewa Bediako Akuffo and Birgit Lugin. 2025. Investigating the Perspective of Non-Native Speakers on Foreigner-Directed Speech using Virtual Agents: The Role of Racial Ingroup Affiliation and Language Proficiency on Perception and Comprehension. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025, IFAAMAS*, 10 pages.

## 1 INTRODUCTION

In today's globalised world, intercultural communication has become an integral part of everyday life, particularly in regions experiencing significant immigration. For Black immigrants, the challenge of adapting to a new language is often compounded by their visual distinctiveness, such as skin colour, which can lead to them being more frequently perceived as non-native, regardless of their actual language proficiency. A study by [38] demonstrated that visual attributes influence speech perception in regions with greater ecological diversity, where non-White faces are linked to less intelligible speech. This dual labelling, based on both language ability and physical appearance, may result in Black non-native speakers (NNSs) receiving a disproportionate amount of foreigner-directed speech (FDS). FDS is a speaking style or speech register employed by native speakers (NSs) during interactions with NNSs, where NSs modify their language to meet the linguistic needs of NNSs, thereby

facilitating communication [72, 92]. Systematic research into this phenomenon began in the 1970s with the works of Ferguson [14]. With increasing globalisation and the growing frequency of contact between NSs and NNSs, the relevance of conducting FDS research has grown. As a result, various aspects of FDS have been explored over time, primarily focusing on native interlocutors, their views on FDS [34, 72], and the adaptation strategies they use [59, 62, 64]. However, an aspect that should be central to FDS research has mainly been overlooked — namely, the perception of NNSs and the comprehensive benefits or disadvantages FDS might bring them.

The study of interpersonal phenomena, such as FDS, can be resource-intensive, requiring the involvement of multiple NSs and NNSs. To address these challenges, researchers increasingly turn to intelligent virtual agents (IVAs), as they have been shown to elicit behaviours similar to those observed in human interactions [29, 39, 65]. In addition, they serve as more cost-effective and standardisable alternatives for studying human-human interaction. [40]. The present study is, to our knowledge, the first to investigate how NNSs evaluate FDS when delivered by IVAs. This research contributes to the academic discussion on FDS and offers essential insights that could enhance intercultural communication and support the integration of NNSs.

Furthermore, research on FDS has often been Eurocentric, neglecting the perspectives of people of Sub-Saharan African descent [67]. By including Black immigrants, a more comprehensive understanding of the perception and evaluation of FDS by NNSs that transcends a Western-centric viewpoint may be achieved. Therefore, this study focuses on Black immigrants in Germany. Since the majority of NSs in Germany are racially White, Black immigrants primarily experience FDS from White individuals. However, as little is known about the influence of NSs' racial ingroup or outgroup affiliation with the NNS on the NNS's perception of FDS, our study includes IVAs from both Black and White racial backgrounds. In conclusion, we examine how Black NNSs evaluate FDS delivered by IVAs, its effectiveness in aiding comprehension, and the influence of racial ingroup affiliation and language proficiency.

## 2 THEORETICAL BACKGROUND

As one aim of this study is to present the perspective of people of Sub-Saharan African descent, it is important to clarify the use of certain terminology. In this context, the term "race" describes external characteristics commonly shared by various ethnic groups and is used as a marker of shared ingroup identity [91, 93]. It is important to note that the term "race" is problematic due to its history of misuse and misinterpretation, which has contributed to social inequalities, discrimination, and violence against certain groups [19, 91]. As the concept of ethnicity, defined as the shared cultural identity of a group of people based on common traditions, religions, and languages [4, 93], does not fully encompass the scope of this study, we will use the expression "race". However, we reiterate the problematic nature of this term and the need for reform in this area.

Furthermore, people of Sub-Saharan African descent or IVAs modelled on this group of people are referred to as "Black" with a capital "B". The reason for this is that capitalising certain group labels shows respect for the group [88]. In addition to "Black people", another expression that is accepted by the current consensus, and

which is mainly used in social science and human rights contexts, is "People of Colour" or "Person of Colour" (POC) [80]. It encompasses all non-White individuals who experience systemic racism [32, 80]. However, as our focus is on individuals who identify as Black, the term POC would be too broad. It must be noted that not all people of Sub-Saharan African descent agree with the term "Black", as the use of this term poses a risk of generalising all group members as a monolithic entity [45]. Nonetheless, the lead author of this paper identifies as Black and considers this term appropriate. Moreover, we refer to individuals of European descent or IVAs modelled after individuals of European descent as "White." The term "Caucasian" was explicitly avoided, as it is rooted in outdated racial classifications [77].

### 2.1 Foreigner-Directed Speech

To understand FDS, it is essential to examine its underlying framework, namely Communication Accommodation Theory (CAT). CAT is an interdisciplinary theoretical framework which was first introduced in the 1970s by Howard Giles. This framework shows how and why people instinctively adapt their linguistic behaviour in response to their interlocutors [5, 72, 79]. Therefore, CAT analyses the relationship between language, context, and social identity. The main message of CAT is that individuals modify their linguistic and other communicative behaviours to either converge with or diverge from their interlocutors. Divergence is the deliberate accentuation of differing communication styles to emphasise ingroup-outgroup affiliations [21, 79]. Convergence on the other hand happens when individuals adjust their speech patterns to align themselves with the communicative style of their interlocutor.

FDS is a clear speech register which represents a form of convergence that has evolved in response to the need for effective communication with NNSs [14, 72, 92]. The goal of FDS, like all clear speech registers, is to make communication easier and more comprehensible for the target person, in this case, NNSs. This form of convergence is characterised, above all, by its inconsistency, with the degree of accommodation depending on subjective factors such as NNSs' proficiency level [43, 78]. Lynch [43] found that foreign English teachers adjusted their level of accommodation based on the proficiency of the NNS, with the lowest proficiency levels triggering the most accommodation. Given these findings, an important question arises: How do NNSs perceive NSs who use high levels of accommodation strategies, even when such adjustments are unnecessary due to the NNS's good language proficiency? This phenomenon is known as overaccommodation, which occurs when the expectations of both the NS and the NNS regarding the optimal level of accommodation do not align [5, 8, 21]. Even though the level of accommodation varies across NSs, universal accommodation strategies define FDS. According to [27], these strategies can be divided into simplifying and clarifying strategies. Simplifying strategies describe the tendency towards morphological, semantic, and syntactic reduction [27]. Morphological simplification refers to the reduction of the complexity of word formation beyond the grammatical norm [15, 90] like omitting word endings, prepositions, and articles [49, 68, 69]. Syntactic simplification on the other hand is characterised by a short paratactic sentence structure as well as an unusual and ungrammatical arrangement of the sentence elements [26, 60, 68]. Lastly, semantic simplification describes the reduction

of idioms and colloquial expressions and the increased substitution of words or phrases with simpler synonyms or internationalisms. Internationalism has similar meanings in several languages as with the word "control" [26, 62, 69]. Clarifying strategies on the other hand describe the attempt to emphasise and elucidate what is being communicated [27]. These strategies include phonological features such as vowel hyperarticulation, meaning the exaggerated pronunciation of vowel sounds [60, 87], as well as an overall exaggerated intonation with more pauses between the syllables and a lower speech rate [12, 26, 92]. Since NNSs have a lower language proficiency than NSs, it is reasonable to assume that the simplifying and clarifying strategies of FDS might aid their comprehension. [33] examined whether hyperarticulated FDS is more comprehensible than native-directed speech (NDS). Native speakers and highly/low proficient L2 NNSs rated 30-minute recordings of hyperarticulated speech or NDS across four noise levels. Results indicated that hyperarticulated speech was the most intelligible across all groups. Furthermore, [59] investigated whether FDS aids in language acquisition, focusing primarily on word learning and word production. Here, participants in the FDS condition demonstrated faster recognition of newly learnt words, as well as generally improved learning, than those in the NDS condition.

Unfortunately, the perception of FDS by its recipients has been largely overlooked. A meta-analysis by [72] found that of 3,581 FDS studies, only one focused on NNSs, the study by [43]. In a non-controlled experiment with 19 NNSs, Lynch discovered that, although not statistically representative, FDS was generally perceived as friendlier than NDS. This finding is consistent with the study by [8], in which English learners evaluated four audio recordings of different speech accommodation styles. Once again, the speaker using FDS was perceived as friendlier compared to the speaker using casual speech. Interestingly, participants also regarded the speaker as more competent when the conversation was easily comprehensible. These findings support CAT's principle that convergence strategies, such as FDS, lead to more positive perceptions than divergence strategies [21, 79].

However, some researchers suggest that highly proficient NNSs may develop negative attitudes toward FDS due to overaccommodation [8, 43, 95], resulting in them perceiving this type of accommodation as disrespectful or patronising [24, 72, 95]. Although numerous assumptions have been made regarding the perception of proficient NNSs, few studies have directly investigated this. One such study by [71] found a modest positive correlation between the NNSs' proficiency levels and their negative attitudes toward FDS. Nonetheless, negative reactions to overaccommodation have also been observed in other language registers, such as elderspeak, an infantilised speech style directed towards older adults [24, 73].

## 2.2 Social-Identity Theory

Building on the understanding of FDS, it is important to consider how social dynamics, as explained by Social Identity Theory (SIT), further shape these linguistic interactions. While CAT provides the framework for understanding how individuals adapt their speech patterns, SIT offers insight into why these adaptations occur. SIT, was developed in the 1970s by Tajfel and colleagues [7, 81, 82] and is defined by an interpersonal-intergroup spectrum, where people's self-concept and perceptions of others shift along this

continuum [30, 83]. In interpersonal interactions, self-concept is based on individual traits, whereas in intergroup interactions, it is shaped by group membership, with individuals viewing themselves as representatives of their social group [30, 66, 83, 86].

Language plays a crucial role in social identity, particularly in ethnic group identity. NNSs often view their foreign accent, alongside visual markers, as part of their ethnic identity [20]. Simultaneously, NSs may perceive NNSs as part of their linguistic or ethnic outgroup based on NNSs' accents or visual attributes [3, 58]. SIT posits that this 'us' (NS) versus 'them' (NNS) distinction influences the interaction of NS with NNS [30]. This may even lead to ingroup favouritism – the tendency to give preferential treatment to one's ingroup over another group which is driven more by the desire for positive distinctiveness than by antipathy towards the outgroup [16, 57, 63, 85]. Based on SIT, it is therefore reasonable to assume that NNSs, despite forming a linguistic outgroup with NSs, generally rate NSs who share their racial ingroup based on external characteristics more favourably than those belonging to their racial outgroup.

## 2.3 Related Work Regarding Intelligent Virtual Agents

Reeves and Nass [65] suggest that people often approach human-computer interactions similarly to human-human interactions, exhibiting comparable behaviours towards computers as they do towards humans. This effect is even more pronounced when the interaction involves embodied agents, such as IVAs [36]. For instance, [76] demonstrated that IVAs are capable of evoking feelings of shame in individuals, just as human interactions can. A crucial advantage of IVAs is the ability to culturally and racially customise them [42]. This is particularly important, as aligning IVA design and behaviour with the cultural background of users enhances user acceptance [11, 42], overall engagement, and information processing [2, 46]. Given that aspects of human-human interaction can be transferred to IVAs, these agents offer a resource-efficient alternative for studying such phenomena, as their appearance and social behaviour can be adapted very cost-effectively, flexibly and consistently to various situations [6, 41, 51, 70].

Although the cultural tailoring of IVAs is well established in research, studies on mixed-cultural agents and mixed-cultural human-computer interactions, such as NS-NNS interactions and agents with accented speech, are a relatively young field [53, 55, 56]. An even less explored subfield is FDS research with IVAs. However, while FDS research concerning IVAs is still a comparatively unexplored field, it holds promise. For instance, [41] explored whether FDS is used with IVAs. German participants interacted with ethnically distinct agents (White vs. Arabic), with the Arabic agent speaking broken German. The agents' appearance and non-verbal behaviour were culturally tailored. Results showed that participants used FDS with the Arabic, but not the White agent, indicating that FDS is transferable to IVAs. Furthermore, participants rated the ingroup agent (White) as warmer and more competent than the outgroup agent (Arabic), suggesting that SIT applies to IVA interactions. Although evidence suggests that FDS might be transferable to IVAs, there is no known study investigating how NNSs perceive FDS when performed by IVAs.

## 2.4 Contribution

This study presents the implementation and evaluation of an interactive prototype designed to investigate how Black participants perceive IVAs using FDS compared to IVAs that use NDS, and whether FDS aids in language comprehension. Additionally, the study explores the extent to which racial ingroup affiliation between participants and IVAs as well as participants' language proficiency influence these outcomes. Based on the presented research, we formulate the following hypotheses:

- **H1a:** FDS used by an IVA leads to better comprehension by NNSs compared to NDS.
- **H1b:** NNSs' level of proficiency negatively moderates the effect of FDS on comprehension.
- **H2a:** An IVA using FDS is perceived as more helpful by NNSs than an IVA using NDS.
- **H2b:** Race matching between a native-speaking IVA and NNSs positively moderates the effect of FDS on the perceived helpfulness of the IVA.
- **H2c:** NNSs' level of proficiency negatively moderates the effect of FDS on the perceived helpfulness of the IVA.
- **H3a:** An IVA using FDS is perceived as warmer by NNSs than an IVA using NDS.
- **H3b:** Race matching between a native-speaking IVA and NNSs positively moderates the effect of FDS on the perceived warmth of the IVA.
- **H3c:** NNSs' level of proficiency negatively moderates the effect of FDS on the perceived warmth of the IVA.
- **H4a:** An IVA using FDS is perceived as more competent by NNSs than an IVA using NDS.
- **H4b:** Race matching between a native-speaking IVA and NNSs positively moderates the effect of FDS on the perceived competence of the IVA.
- **H4c:** NNSs' level of proficiency negatively moderates the effect of FDS on the perceived competence of the IVA.

## 3 IMPLEMENTATION

Using the *Unity game engine* (version 2021.3.31)<sup>1</sup>, a Wizard of Oz scenario was implemented in which participants interacted virtually via *Zoom*<sup>2</sup> with an IVA controlled by an experimenter. The Wizard of Oz interface comprised a series of buttons, each corresponding to a specific verbal or non-verbal behaviour of the IVA. After each participant's response, the experimenter selected the appropriate button, initiating the IVA's reaction. The environment was designed to resemble a German outdoor festival, as seen in Figure 1 (centre), where the participants engaged in small talk with the seemingly autonomous agent. The IVA was positioned in front of food and drink stalls. The background noises of a bustling crowd enhanced the illusion of a real festival without requiring the implementation of additional background characters.

The agent began the conversation with a monologue, briefly introducing itself. This monologue was presented in NDS across all study conditions to prevent participants from assuming that the racially Black IVA spoke broken German due to its own low

language proficiency and also to help participants familiarise themselves with the IVA. During the conversation, the agent reaffirmed its German nativeness as well. Depending on the study condition, the IVA either switched to FDS or continued in NDS, after the participant spoke a few words themselves. A condition in which the agent consistently used FDS from the start was not included, as this would likely be perceived as non-native speech rather than FDS, shifting the focus to a different research topic. Introducing a switch ensured that the non-standard speech patterns were perceived as intentional. The study was designed for the IVA to guide the conversation by asking questions about the participant's name, country of origin, and preferred cuisine, maintaining the appearance of autonomy. The experimenter could select from two to four predefined response options, which meant that the agent could only provide a limited number of possible answers. To address unforeseen questions or responses from participants, a series of emergency buttons were implemented, which triggered generic responses such as "yes", "no", or "how nice". The setup consisted of two screens. One monitor displayed the game view, i.e., the IVA and the environment. This display was also shared with the participants via Zoom. The other one exhibited the Wizard of Oz interface, with the buttons used to control the agents and the interaction, as seen in Figure 1 (left).

### 3.1 Agents

The IVAs were created using the web-based 3D design and animation tool *Autodesk Character Generator*<sup>3</sup>. As the focus is on Black participants as well as ingroup-outgroup effects, the agents were divided into two racially distinct groups, Black and White. The Black IVAs were modelled after the average facial features of West African people, while the White IVAs were based on European features<sup>4</sup>. The West African models were chosen as it was known prior to recruitment that most participants would originally be from West and Central Africa. The Black agents were designed to have darker skin and eyes [47], shorter and broader nasal bridges with wider nostrils [31, 61], fuller lips [18, 50], wider faces [31, 61], and tightly curled hair [13]. In contrast, the White agents exhibited lighter skin, lighter hair and eye colour, thinner lips, and less textured hair [13, 18]. To minimise bias, all agents were dressed similarly in short-sleeved T-shirts and jeans, with short hair. Four agents in total were created: two from each racial group, with one male and one female per race. This setup ensured that participants only interacted with IVAs of the same gender, aligning with the cultural norms of the target population [9, 23], and thus, eliminating gender as a variable in the evaluation. The IVAs are shown in Figure 1 (right).

### 3.2 Speech

The study contains two language conditions: FDS and NDS. The text-to-speech (TTS) software *Amazon Polly*<sup>5</sup> was used to generate the agents' speech, with the standard German voices "Vicki" for female agents and "Hans" for male agents. This approach was chosen based on previous studies suggesting that, while human voices

<sup>1</sup><https://unity.com/de/releases/editor/whats-new/2021.3.31>

<sup>2</sup><https://www.zoom.com/de>

<sup>3</sup><https://charactergenerator.autodesk.com/>

<sup>4</sup><https://www.artfido.com/this-is-what-the-average-person-looks-like-in-each-country/>

<sup>5</sup><https://aws.amazon.com/de/polly/>

are sometimes preferred, synthetically generated voices can yield equivalent results [1, 54, 56]. The NDS files were generated without any additional modifications.

A two-stage process was used to create the FDS speech files. First, nine native German students (seven female, two male) recorded speech samples based on Ferguson’s study [14], simulating how native speakers communicate with low-proficient NNSs. Two independent coders, one with a linguistics master’s degree, analysed these recordings according to FDS theories. Their analyses were merged into a single codebook. Instead of relying solely on FDS theories, recordings were incorporated to ensure the IVAs’ FDS aligned with German speech patterns, as most FDS studies focus on English and accommodation strategies can vary across languages [12, 28]. The FDS characteristics most frequently listed in the codebook were implemented in the second step using speech synthesis markup language (SSML) tags. For instance, vowel hyperarticulation was achieved by doubling vowels and at the same time emphasising the entire word. For each language condition (FDS or NDS), two MP3 files – one male and one female voice – were generated. The audio files were then divided into smaller segments and integrated into the Wizard of Oz interface.

### 3.3 Non-Verbal Behaviour

The non-verbal behaviour of the IVAs included facial expressions as well as gestures, blinking and gaze behaviour. The facial expressions, blinking behaviour and lip synchronisation were generated by the *SALSA LipSync Suite toolset* by Crazy Minnow Studio <sup>6</sup> and were identical across all conditions. The gestures were downloaded from *Mixamo* <sup>7</sup> and integrated into the Unity environment. For the NDS condition, only beat gestures and occasional deictic gestures were employed to imitate natural conversational behaviour. Based on human-human FDS research [84], mainly iconic and deictic gestures were used for the FDS condition. Deictic gestures were utilised whenever the IVA asked the participants a question, and whenever the IVA made a statement about itself or a place. Iconic gestures or emblems were used each time the IVA described a relevant action. Two NSs and one NNS tested the stimuli for each condition. For NDS, participants rated speech on a 1–5 scale based on how native-like the speaker sounded. For FDS, an additional question addressed noticeable changes in speech over time. Language files were deemed authentically NDS if rated above 3 and authentically FDS if rated below 3 on native-like speech. The environment was iteratively adjusted based on feedback from the three respondents.

## 4 METHOD

A 2 (FDS, NDS) × 2 (Black IVA, White IVA) between-subjects design examined how Black non-native German speakers rate IVAs of their racial ingroup and outgroup using FDS or NDS, and whether FDS aids language comprehension. Participants were randomly assigned to one of four conditions. As an online presence study, participants registered via a virtual calendar and were informed about the procedure in advance. After registration, they received an automated confirmation and a Zoom link the day before the study. Upon joining the meeting, participants were informed that their

camera and microphone had to remain on to ensure independent completion. Communication with the experimenter was possible in German, English, French, or Twi. After explaining the procedure and answering any questions, participants gave informed consent before the experimenter deactivated their camera and microphone to begin screen sharing the Unity game. The interaction started with the virtual agent’s monologue. Depending on the condition, participants interacted with either a Black or White IVA performing FDS or NDS. They could ask questions at any time, and the agent could repeat answers if needed. The interaction ended with the IVA mentioning they had to catch the bus. Participants then received a link to the online questionnaire via Zoom chat, which they completed while remaining in the meeting. In addition to German, the questionnaire was available in English and French as well, as these two languages are among the most widely spoken in Africa, reflecting the continent’s colonial history [94]. Any remaining questions were addressed during the debriefing session. The study was approved by the ethics committee of the Institute of Human-Computer-Media at the University of Würzburg (reference number: 010823)

### 4.1 Measures

All respondents completed the same questionnaire across conditions, which included an image of the agent they interacted with for reference. This resulted in four versions of the questionnaire, corresponding to the visually distinct IVAs: male Black, female Black, male White, and female White. The study description specified that only people identifying as racially Black could participate. To ensure this and serve as a manipulation check, a yes/no question was included to confirm whether respondents categorised the Black agent as part of their racial ingroup and the White agent as not. Incorrect answers lead to an exclusion from further analysis.

*Comprehension:* To assess participants’ understanding of the conversation, both subjective (s-) and objective (o-)comprehension measures were used. The subjective measure asked respondents to rate their understanding of the IVA interaction on a scale from 1 to 5, with 1 indicating “not at all well” and 5 “very well”. The objective measure consisted of a quiz [74], where participants had to identify three statements made by the agents from three possible answers. The questions included the agent’s original city, their favourite food, and the reason for ending the conversation.

*Helpfulness:* The two items “helpful” and “useful” from [48]’s 12-item scale was used to measure the perceived helpfulness of IVAs. The scale, containing only these two items, was highly reliable with Cronbach’s  $\alpha = .91$ .

*Warmth and Competence:* To measure the perceived warmth and competence of the IVAs, the questionnaire by [17] was used, featuring five items for competence and four for warmth. The warmth items included descriptors like “tolerant” and “warm” while the competence items included items such as “competent” and “confident”. Participants rated how well these items, along with those measuring perceived helpfulness, characterised the IVA on a scale from 1 = “not at all well” to 5 = “very well.” The warmth and competence scale showed high reliability, with Cronbach’s  $\alpha = .84$ .

*Level of Proficiency:* An extract from the LEAP-Q [44] was used to assess the respondents’ German language proficiency. Participants

<sup>6</sup><https://crazyminnowstudio.com/>

<sup>7</sup><https://www.mixamo.com/>

rated their proficiency in speaking, understanding and reading German from 1 = "none" to 9 = "perfect". Furthermore, respondents indicated whether German was their native, second, or other language to ensure only NNSs were included in the final analysis. This scale demonstrated high reliability, with Cronbach's  $\alpha = .94$ .

**Demographic Data:** The demographic data collected in the study consisted of gender, age, country of origin, highest level of education and the device used to complete the study.

## 4.2 Participants

The sample was restricted to individuals who identified as racially Black. Additionally, participants were required to have learned German later in life, indicating it was not their first language [8, 87]. Given the challenging recruitment process, most participants were drawn from referrals from former participants. Additionally, two African churches in Hamburg, Germany, promoted the study during services. A total of 132 people participated in the study, but eight were excluded from further analysis for failing to recognise the Black IVA as an ingroup member, identifying the White IVA as an ingroup member, or indicating that German was their native language. Ultimately,  $N = 124$  participants (61 female, 62 male, 1 diverse who wanted to be assigned to a female IVA) were included in the analysis. Their ages ranged from 18 to 64 years, with the majority (44.4%) being between 25 and 34 years old. Mean age reporting was not appropriate as ages were recorded in categories, making them ordinal. Regarding country of origin, most participants (91.1%) identified an African country, with the majority of them (62.9%) being from West Africa. The mean German proficiency level was  $M = 6.08$  ( $SD = 1.87$ ).

## 5 RESULTS

All analyses were carried out using the statistical software *IBM SPSS Statistics 27*<sup>8</sup> with a significance level of .05. Additionally, the add-on *macro PROCESS*<sup>9</sup> was employed to conduct moderation analyses. The prerequisites for the calculated tests were fulfilled, unless otherwise stated.

### 5.1 Self-Reports

Levene's tests showed homogeneity of variances across all variables with  $p > .05$ , except for helpfulness ( $p < .05$ ). Additionally, Shapiro-Wilk tests indicated a deviation from normality for all variables. Based on these results, a Welch ANOVA was conducted for the variables comprehension, perceived warmth, and competence. The results of the Welch ANOVA can be found in Table 2. Due to the violation of variance homogeneity and normality, a Kruskal-Wallis test was performed to assess perceived helpfulness. The means of the latent variables are presented in Table 1.

**Comprehension:** Regarding the o- and s-comprehension measure no significant differences were found between the conditions.

**Helpfulness:** Participants' perceptions of the IVAs' helpfulness differed significantly when comparing the conditions using a Kruskal-Wallis test,  $H(3) = 14.27, p = .003$ . Subsequent post-hoc Dunn-Bonferroni tests revealed statistically significant differences between the FDS-White and NDS-Black agents ( $z = -2.92, p = .021$ )

**Table 1: Descriptive Values (Means) of the Latent Variables.**

| Latent Variables | FDS                   |                       | NDS                   |                       |
|------------------|-----------------------|-----------------------|-----------------------|-----------------------|
|                  | Black<br><i>M(SD)</i> | White<br><i>M(SD)</i> | Black<br><i>M(SD)</i> | White<br><i>M(SD)</i> |
| O-Comprehension  | 2.61(.50)             | 2.77(.43)             | 2.80(.55)             | 2.72(.58)             |
| S-Comprehension  | 4.10(.91)             | 4.00(.97)             | 4.57(.86)             | 4.38(.94)             |
| Helpfulness      | 3.95(.88)             | 3.39(1.16)            | 4.20(.86)             | 4.34(.71)             |
| Warmth           | 4.22(.59)             | 3.98(.79)             | 4.47(.54)             | 4.44(.52)             |
| Competence       | 3.81(.68)             | 3.61(.75)             | 4.04(.69)             | 4.18(.70)             |

*Note.*  $N = 124$

and between the FDS-White and NDS-White agents ( $z = -3.50, p = .003$ ), with FDS-White agents being perceived as less helpful than both the NDS-White and NDS-Black agents.

**Warmth:** In terms of warmth, significant differences were found between the conditions. Again, post-hoc analysis using the Games-Howell test ( $p < .05$ ) revealed that these differences were specifically between the FDS-White and NDS-White conditions ( $-0.45, 95\% CI[-0.90, -0.01]$ ) as well as the FDS-White and NDS-Black conditions ( $-0.48, 95\% CI[-0.94, -0.03]$ ), with the FDS-White agents being rated significantly less warm than the NDS-White and NDS-Black agents.

**Competence:** Similar to perceived warmth, significant differences were found for agent competence between the groups. Here, a Games-Howell post-hoc test ( $p < .05$ ) showed that the difference was specifically between the FDS-White and NDS-White agents ( $-0.57, 95\% CI[-1.05, -0.08]$ ), with the FDS-White agents being rated as significantly less competent.

**Table 2: Results of the Analysis of Variance (Welch ANOVA)**

|                 | df1 | df2   | <i>F</i> | <i>p</i> | $\eta^2$ |
|-----------------|-----|-------|----------|----------|----------|
| O-Comprehension | 3   | 66.18 | 0.84     | .478     | 0.019    |
| S-Comprehension | 3   | 66.64 | 2.46     | .070     | 0.057    |
| Warmth          | 3   | 66.02 | 3.40     | .023*    | 0.092    |
| Competence      | 3   | 66.58 | 3.71     | .016*    | 0.090    |

*Note.*  $N = 124$

\* $p < .05$

### 5.2 Influence of Race

Moderation analyses were conducted to examine whether the relationship between the speech condition (FDS, NDS) and the perceived helpfulness, warmth, and competence of the agents was moderated by the agents' race. The interaction terms were added to the respective regression models after entering the main effects of the independent and moderator variables.

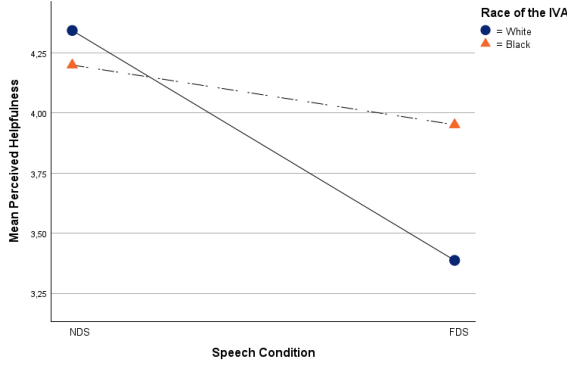
**Helpfulness:** Regarding helpfulness, the overall model was significant,  $F(3, 120) = 5.39, p = .002$ , with an  $R^2$  of .142. The results showed a significant interaction effect,  $\beta = 0.71, p = .037$ , suggesting that the relationship between the agents' speech condition (FDS, NDS) and their perceived helpfulness varies depending on the agents' race,  $F(1, 120) = 4.47, p = .037, \Delta R^2 = .033$ . Simple slopes analysis revealed that White agents in the FDS condition

<sup>8</sup><https://www.ibm.com/de-de/>

<sup>9</sup><https://processmacro.org/download.html>



were perceived as significantly less helpful than in the NDS condition,  $\beta = -0.96, p = .0002, 95\% - CI[-1.45, -0.47]$ . Figure 2 shows the moderation effect. For Black agents, the effect was less negative and not significant ( $\beta = -0.25, p = .274$ ).



**Figure 2: Results of the Moderation Effect of IVA's Race on the Relationship between Speech Condition and Perceived Helpfulness of the IVA.**

*Warmth:* The overall model was found to be significant for perceived warmth,  $F(3, 120) = 3.36, p = .021, R^2 = .092$ . However, the interaction effect between speech condition (FDS, NDS) and agents' race was not significant ( $\beta = 0.20, p = .367$ ). The change in  $R^2$  due to the interaction was minimal ( $F(1, 120) = 0.82; p = .367, \Delta R^2 = .006$ ).

*Competence:* In terms of competence, the overall model was found to be significant,  $F(3, 120) = 3.67, p = .014, R^2 = .090$ . However, race was not a considerable moderator of the relationship between speech condition and perceived competence of the IVA ( $F(1, 120) = 1.69, p = .196, \Delta R^2 = .013$ ). Nevertheless, the main effects of speech condition on both perceived warmth ( $\beta = -0.45, p = .009, 95\%CI[-0.79, -0.11]$ ) and perceived competence ( $\beta = -0.57, p = .003, 95\%CI[-0.94, -0.20]$ ) were significant.

### 5.3 Influence of Language Proficiency

Moderation analyses were calculated to analyse the influence of language proficiency on the relationship between the speech condition of the agents (FDS, NDS) and the comprehension of the conversation as well as the perceived helpfulness, warmth, and competence.

*Comprehension:* Regarding s-comprehension, the overall model was significant,  $F(3, 120) = 20.65, p < .001, R^2 = .308$ . The interaction between speech condition and participants' language proficiency was also significant with  $\beta = -0.218, p = .021, \Delta R^2 = .042$ . Simple slopes analysis showed that the use of FDS led to a decrease in subjective understanding at moderate ( $\beta = -0.373, p = .013$ ) and high language proficiency levels ( $\beta = -0.782, p < .001$ ). However, the effect was not significant at low proficiency levels ( $\beta = 0.035, p = .891$ ). With o-comprehension, although the overall model was significant ( $F(3, 120) = 5.18, p = .002, R^2 = .202$ ), adding the interaction term to the model could not explain a considerable proportion of the variance,  $\beta = -.13, p = .078$ .

*Helpfulness:* The regression analysis examining the effect of speech condition on the perceived helpfulness of the agents, moderated by language proficiency, yielded a significant overall model,  $F(3, 120) = 7.27, p = .0002, R^2 = .181$ . The interaction between speech condition and language proficiency was significant,  $\beta = -0.199, p = .016, \Delta R^2 = .033$ . Further analysis of the interaction effect revealed, with increasing proficiency levels ( $\beta = -0.993, p = .0001$ ), speech condition significantly affected perceived helpfulness, with FDS leading to lower perceived helpfulness scores.

*Warmth and Competence:* In terms of warmth ( $F(3, 120) = 3.61, p = .0153, R^2 = .105$ ) and competence ( $F(3, 120) = 6.37, p = .0005, R^2 = .129$ ) both overall models were significant. However, the interaction between speech condition and language proficiency in the case of warmth ( $\beta = -0.090, p = .147$ ) and competence ( $\beta = -0.110, p = .1235$ ) was not statistically significant.

## 6 DISCUSSION

This study aimed to investigate the perception of Black non-native German speakers on FDS performed by an IVA and whether FDS facilitates conversational comprehension using an interactive prototype. Furthermore, the impact of participants' language proficiency and the IVA's racial affiliation was examined.

For H1a, no significant differences were found between the FDS and NDS conditions regarding comprehensibility; therefore, H1a is rejected. However, in terms of subjective comprehensibility of the conversation, there was a trend for participants in the NDS condition to rate the comprehensibility of the IVA higher than those in the FDS condition, contradicting the findings of [33] and [59]. Instead, our results may align with [10]'s alternative explanation that simplifying strategies may confuse language learners by increasing cognitive load rather than enhancing comprehensibility. Notably, the studies by [33] and [59], which reported comprehension advantages for FDS, only included clarifying strategies such as hyperarticulation. Since we used both clarifying and simplifying strategies based on NNS recordings and established FDS studies [27, 49, 68, 92], it's unclear if the higher comprehensibility in the NDS condition really supports Chaudron's explanation. Future research should examine the specific effects of different FDS strategies on comprehension.

H2a predicted that an IVA using FDS would be perceived as more helpful than one using NDS. However, our results show that White IVAs using FDS were rated significantly less helpful than both White and Black IVAs using NDS. While Black FDS agents were also perceived as less helpful than NDS agents, this difference was not statistically significant. These findings contradict our hypothesis, leading to the rejection of H2a. A possible explanation is that, since no significant differences in comprehensibility were found, though there was a tendency towards lower comprehensibility in the FDS condition, agents using FDS may have been perceived as less helpful for this reason.

We hypothesised that IVAs in the FDS condition would be perceived as warmer (H3a) and more competent (H4a) than those in the NDS condition. However, the FDS agents were rated significantly less warm and competent than those in the NDS condition, contradicting previous studies [8, 43] and leading to the rejection of H3a and H4a. It's important to note that both [8] and [43] focused

on ratings of audio recordings, without participants engaging in direct interactions with a person or IVA. As such, visual factors that can influence a multimodal interaction were absent. To our knowledge, this is the first study to examine NNSs' perceptions of FDS in an IVA interaction. Given the limited research in this area and the reliance of most socio-psychological studies on non-interactive experiments, our findings offer valuable insights. Additionally, [8] found that participants perceived interlocutors as more competent when they were perceived as friendlier. Since our FDS agents were rated less friendly than the NDS agents, this could explain why they were also viewed as less competent.

Regarding H2b, results showed that race significantly moderated the perception of helpfulness. All FDS agents were seen as less helpful than NDS agents, with the effect being more pronounced for White agents. This aligns with SIT and the concept of ingroup favouritism, where White outgroup agents were rated less favourably, likely as a way for participants to maintain a positive self-concept [16, 30, 82]. Although H3b and H4b followed the hypothesised direction, a significant effect was only found for the relationship between language condition and perceived helpfulness, but not for perceived warmth (H3b) or competence (H4b), leading to their rejection. Overall, the racial modelling of IVAs partly triggered responses consistent with SIT, as participants perceived White agents as part of their racial outgroup. Since IVAs provide a resource-efficient alternative to studying human interactions [6, 40], the study design can be adapted to explore how other racial groups interact with IVAs, fostering inclusivity in research.

We predicted that participants' language proficiency would negatively moderate the relationship between language condition and comprehensibility (H1b), perceived helpfulness (H2c), warmth (H3c), and competence (H4c) of the IVA. The results partially support H1b, with higher proficiency leading to lower subjective comprehensibility ratings in the FDS condition, though this wasn't reflected in the o-comprehension test. Additionally, FDS agents were seen as significantly less helpful with higher proficiency, supporting H2c. No significant effects were found for H3c or H4c, though the trends suggest a moderating effect, warranting further investigation with a larger sample. These findings align with research on overaccommodation [5, 24, 71].

## 6.1 Limitations

As previously stated, the FDS of the IVAs in this study was modelled on recordings of NSs who imagined interacting with NNSs. Nevertheless, [75] have already observed that the method of imagined FDS can potentially bias the results of studies, as imagined FDS is similar but not identical to real FDS. Consequently, future studies could base the FDS of the IVAs on recordings of real-life interactions between NSs and NNSs. This method could ensure that the behaviour of the IVAs is as similar as possible to that of humans. Additionally, the recordings on which the FDS behaviour of the IVAs was based were independently coded by two individuals for this study. Morphological and syntactic strategies employed during FDS can be easily discerned by humans. However, errors may occur during the coding process regarding more subtle phonetic attributes. Thus, some studies that focus on the phonetic attributes of FDS have used software for phonetic analyses, such as *Praat* [25, 35]. To achieve more accurate coding, it would therefore be

advantageous for future studies to utilise said software in addition to human coders.

Furthermore, in the FDS condition, a switch was implemented from NDS (used in the monologue) to FDS. Since FDS is typically not adopted before recognising a conversational partner as non-native, this switch was necessary to reflect both behavioural adaptation and its trigger. However, the study design resulted in a switch (FDS) and a non-switch (NDS) condition. Although implementing the switch was crucial, we cannot determine its specific impact on the results. Future studies could address this by introducing a control condition in which IVAs use FDS from the outset. Moreover, the stimuli in this study were iteratively adapted based on feedback from three observers. While it was important to include the perspectives of both NSs and NNSs in the study design, a more extensive pre-test with a bigger sample size should be conducted in future studies to enhance the validity of the stimuli.

As the first known study to assess NNSs' perspectives on FDS using IVAs, our experiment employed a Wizard of Oz design via Zoom. This approach ensured greater control and standardisation compared to fully autonomous conversational AI by predefining system responses, reducing variability, and enhancing internal validity. Additionally, Zoom facilitated access to a geographically diverse sample. Future studies could automate interactions with AI to enable more natural conversations and facilitate large-scale implementation [22, 37, 89]. Laboratory studies in virtual reality can further strengthen results due to greater immersion and sense of presence [41, 52], although potentially at the expense of the sample size.

## 7 CONCLUSION

We investigated how Black non-native participants view an intelligent virtual agent (IVA) using foreigner-directed speech (FDS) compared to an IVA using native-directed speech (NDS), and whether FDS aids in language comprehension. Additionally, we explored the effects of racial dynamics and participants' language proficiency on these outcomes. Therefore, we conducted a 2(FDS, NDS) x 2(Black IVA, White IVA) between-subject design in the form of an interactive prototype, where, depending on the condition, participants engaged in small talk with an IVA. We found that FDS did not enhance the understanding of non-native speakers (NNSs); in fact, it may have worsened comprehension especially for high-proficient NNSs. Additionally, IVAs using FDS were perceived as less helpful, less friendly, and less competent than those using NDS, with these negative perceptions becoming stronger as NNSs' proficiency increased and when the IVA belonged to the racial outgroup. While most of the findings contradict the initial hypotheses, the study provides valuable insights into the perspective of Black people on interaction with IVAs and at the same time highlights the need to consider NNSs' views in FDS research. Concurrently, to our knowledge, this study is the first to investigate NNSs' perception of FDS using IVAs, laying the groundwork for future studies. Overall, this study contributes to research on successful communication between NSs and NNSs, as well as human-computer interaction, autonomous agent research and linguistics. Furthermore, the results suggest that findings from human-human NS-NNS interactions are indeed transferable and applicable to interactions with IVAs.



## REFERENCES

- [1] Amal Abdulrahman and Deborah Richards. 2022. Is natural necessary? Human voice versus synthetic voice for intelligent virtual agents. *Multimodal Technologies and Interaction* 6, 7 (2022), 51.
- [2] Mashael AlSaleh and Daniela Romano. 2015. Culturally appropriate behavior in virtual agents: A review. In *Proceedings of the AAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 11. 69–74.
- [3] Nicholas Aoki and Georgia Zellou. 2023. Speakers talk more clearly when they see an East Asian face: Effects of visual guise on speech production. In *Proceedings of the 20th International Congress of Phonetic Sciences*. Guarant International, 2294–2298.
- [4] Timothy Baumann. 2004. Defining ethnicity. *The SAA archaeological record* 4, 4 (2004), 12–14.
- [5] Leslie M Beebe and Howard Giles. 1984. Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language* (1984).
- [6] Timothy Bickmore, Dhaval Parmar, Everlyne Kimani, and Stefan Olafsson. 2021. Diversity Informatics. In *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents*. ACM.
- [7] Michael Billig and Henri Tajfel. 1973. Social categorization and similarity in intergroup behaviour. *European journal of social psychology* 3, 1 (1973), 27–52.
- [8] Susan C Bobb, Kristin Mello, Emily Turco, Larissa Lemes, Erika Fernandez, and Kathrin Rothermich. 2019. Second language learners' listener impressions of foreigner-directed speech. *Journal of Speech, Language, and Hearing Research* 62, 9 (2019), 3135–3148.
- [9] Silvia Bruzzi. 2017. Islam, Gender and Leadership. In *Islam and Gender in Colonial Northeast Africa*. Brill, 43–62.
- [10] C Chaudron. 1983. Foreigner talk in the classroom-An aid to learning. *Classroom oriented research in second language acquisition/Newbury House* (1983).
- [11] Birgit Endrass. 2014. *Cultural diversity for virtual characters: Investigating behavioral aspects across cultures*. Springer Science & Business Media.
- [12] Kapitolina Fedorova. 2015. *Foreigner Talk: A Register or Registers?* Finnish Literature Society, 138–149. <https://doi.org/10.2307/j.ctvgx2qk.11>
- [13] Cynthia Feliciano. 2016. Shades of race: How phenotype and observer characteristics shape racial classification. *American Behavioral Scientist* 60, 4 (2016), 390–419.
- [14] Charles A Ferguson. 1975. Toward a characterization of English foreigner talk. *Anthropological linguistics* (1975), 1–14.
- [15] Charles A Ferguson. 1982. Simplified registers and linguistic theory. *Exceptional language and linguistics* (1982), 49–66.
- [16] Samantha Finkelstein, Evelyn Yarzabinski, Callie Vaughn, Amy Ogan, and Justine Cassell. 2013. The effects of culturally congruent educational technologies on student achievement. In *Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16*. Springer, 493–502.
- [17] Susan T Fiske, Amy JC Cuddy, Peter Glick, and Jun Xu. 2018. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. In *Social cognition*. Routledge, 162–214.
- [18] Livia Maria Andrade de Freitas, Karina Maria Salvatore de Freitas, Arnaldo Pinzan, Guilherme Janson, and Marcos Roberto de Freitas. 2010. A comparison of skeletal, dentoalveolar and soft tissue characteristics in white and black Brazilian subjects. *Journal of Applied Oral Science* 18 (2010), 135–142.
- [19] Herbert J Gans. 2005. Race as class. *Contexts* 4, 4 (2005), 17–21.
- [20] Elizabeth Gatbonton, Pavel Trofimovich, and Michael Magid. 2005. Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation. *TESOL quarterly* 39, 3 (2005), 489–511.
- [21] Howard Giles, Tania Ogay, et al. 2007. Communication accommodation theory. *Explaining communication: Contemporary theories and exemplars* (2007), 293–310.
- [22] David Griol, Araceli Sanchis, José Manuel Molina, and Zoraida Callejas. 2019. Developing enhanced conversational agents for social virtual worlds. *Neurocomputing* 354 (2019), 27–40.
- [23] Noah Haber, Paul Jake Robyn, Saidou Hamadou, Gervais Yama, Herve Hien, Davy Louvouezo, and Günther Fink. 2018. Surveyor Gender Modifies Average Survey Responses: Evidence from Household Surveys in Four Sub-Saharan African Countries. *arXiv preprint arXiv:1810.01981* (2018).
- [24] Jake Harwood, Howard Giles, Susan Fox, Ellen Bouchard Ryan, and Angie Williams. 1993. Patronizing young and elderly adults: Response strategies in a community setting. *Journal of Applied Communication Research* (1993), 211–226.
- [25] Valerie Hazan, Maria Uther, and Sonia Grunlund. 2015. How does foreigner-directed speech differ from other forms of listener-directed clear speaking styles? *18th International Congress of Phonetic Sciences* (2015).
- [26] Vera M Henzl. 1979. Foreign talk in the classroom. *International review of applied linguistics* 17, 2 (1979), 159–167.
- [27] Volker Hinnenkamp. 1981. "Türkisch Mann, Du?": Sprachverhalten von Deutschen gegenüber Gastarbeitern. In *Mehrsprachigkeit in der Stadtregion: Jahrbuch 1981 des Instituts für deutsche Sprache*. Schwann, 171–193.
- [28] Hans Henrich Hock and Brian D Joseph. 2009. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Mouton de Gruyter.
- [29] Laura Hoffmann, Nicole C Krämer, Anh Lam-Chi, and Stefan Kopp. 2009. Media equation revisited: do users show polite reactions towards an embodied agent?. In *Intelligent Virtual Agents: 9th International Conference, IVA 2009 Amsterdam, The Netherlands, September 14-16, 2009 Proceedings 9*. Springer, 159–165.
- [30] Matthew J Hornsey. 2008. Social identity theory and self-categorization theory: A historical review. *Social and personality psychology compass* 2, 1 (2008), 204–222.
- [31] James M Jeffries III, Barry DiBernardo, and Gregory E Rauscher. 1995. Computer analysis of the African-American face. *Annals of plastic surgery* 34, 3 (1995), 318–322.
- [32] Anita Kalunta-Crumpton. 2020. The inclusion of the term 'color' in any racial label is racist, is it not? *Ethnicities* 20, 1 (2020), 115–135.
- [33] Jayanthiny Kangatharan, Maria Uther, and Fernand Gobet. 2022. The effect of hyperarticulation on speech comprehension under adverse listening conditions. *Psychological Research* 86, 5 (2022), 1535–1546.
- [34] Jayanthiny Kangatharan, Maria Uther, and Fernand Gobet. 2023. The effect of clear speech to foreign-sounding interlocutors on native listeners' perception of intelligibility. *Speech Communication* 150 (2023), 66–72.
- [35] Monja A Knoll and Lisa Scharrer. 2007. Acoustic and affective comparisons of natural and imaginary infant-, foreigner- and adult-directed speech. In *INTER-SPEECH*. 1414–1417.
- [36] Nicole Krämer and Arne Manzeschke. 2021. Social reactions to socially interactive agents and their ethical implications. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*. Association for Computing Machinery, 77–104.
- [37] Sheetal Kusal, Shruti Patil, Jyoti Choudrie, Ketan Kotecha, Sashikala Mishra, and Ajith Abraham. 2022. AI-based conversational agents: a scoping review from technologies to future directions. *IEEE Access* 10 (2022), 92337–92356.
- [38] Ethan Kutlu, Mehrgol Tiv, Stefanie Wulff, and Debra Titone. 2022. Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications* 7, 1 (2022), 7. <https://doi.org/10.1186/s41235-022-00354-0>
- [39] Tze Wei Liew and Su-Mae Tan. 2016. Virtual agents with personality: Adaptation of learner-agent personality in a virtual learning environment. In *2016 Eleventh International Conference on Digital Information Management (ICDIM)*. IEEE, 157–162.
- [40] Birgit Lugin. 2021. Introduction to socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*. Association for Computing Machinery, 1–20.
- [41] Birgit Lugin, Benjamin Eckstein, Kirsten Bergmann, and Corinna Heindl. 2018. Adapted foreigner-directed communication towards virtual agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 59–64.
- [42] Birgit Lugin and Matthias Rehm. 2021. Culture for socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*. Association for Computing Machinery, 463–494.
- [43] AJ Lynch. 1988. Speaking up or talking down: Foreign learners' reactions to teacher talk. *ELT Journal* 42, 2 (1988), 109–116.
- [44] Viorica Marian, Henrike K Blumenfeld, and Margarita Kaushanskaya. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of speech, language, and hearing research* (2007).
- [45] Ben L Martin. 1991. From Negro to Black to African American: The power of names and naming. *Political Science Quarterly* 106, 1 (1991), 83–107.
- [46] Prasanth Murali, Ameneh Shamekhi, Dhaval Parmar, and Timothy Bickmore. 2020. Argumentation is more important than appearance for designing culturally tailored virtual agents. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 1940–1942.
- [47] Piya Parth Naik and Syed Nadir Farrukh. 2022. Influence of ethnicities and skin color variations in different populations: a review. *Skin pharmacology and physiology* 35, 2 (2022), 65–76.
- [48] Clifford Nass, Youngme Moon, and Paul Carney. 1999. Are people polite to computers? Responses to computer-based interviewing systems 1. *Journal of applied social psychology* 29, 5 (1999), 1093–1109.
- [49] Dana Kristine Nelson. 1992. The Foreigner Talk of a Family Physician: An Observational Study. *ERIC Documents* (1992).
- [50] Joe Niamtu. 2010. Lip reduction surgery (reduction cheiloplasty). *Facial Plastic Surgery Clinics* 18, 1 (2010), 79–97.
- [51] David Obremski, Ohenewa Bediako Akuffo, Leonie Lücke, Miriam Semineh, Sarah Tomiczek, Hanna-Finja Weichert, and Birgit Lugin. 2023. Reducing racial bias by interacting with virtual agents: An intervention in virtual reality. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multi-agent Systems*. 747–755.

- [52] David Obremski, Eva Brucker, Paula Friedrich, and Birgit Lugin. 2023. Behavioural adaptation towards foreign virtual agents in VR-the impact of non-native speech. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. 1–3.
- [53] David Obremski, Paula Friedrich, Nora Haak, Philipp Schaper, and Birgit Lugin. 2022. The impact of mixed-cultural speech on the stereotypical perception of a virtual robot. *Frontiers in Robotics and AI* 9 (2022), 983955.
- [54] David Obremski, Helena Babette Hering, Paula Friedrich, and Birgit Lugin. 2022. Exploratory study on the perception of intelligent virtual agents with non-native accents using synthetic and natural speech in German. In *Proceedings of the 2022 International Conference on Multimodal Interaction*. 15–24.
- [55] David Obremski, Jean-Luc Lugin, Philipp Schaper, and Birgit Lugin. 2019. Non-native speaker generation and perception for mixed-cultural settings. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. 105–107.
- [56] David Obremski, Jean-Luc Lugin, Philipp Schaper, and Birgit Lugin. 2021. Non-native speaker perception of Intelligent Virtual Agents in two languages: the impact of amount and type of grammatical mistakes. *Journal on Multimodal User Interfaces* 15 (2021), 229–238.
- [57] Amine Ouazad. 2014. Assessed by a teacher like me: Race and teacher assessments. *Education Finance and Policy* 9, 3 (2014), 334–372.
- [58] Maria Petrou and Marko Dragojevic. 2024. “Where Are You From?” Language Attitudes and (Non) Accommodation During Native–Nonnative Speaker Interactions in Germany. *Journal of Language and Social Psychology* 43, 3 (2024), 353–375.
- [59] Giorgio Piazza, Marina Kalashnikova, and Clara D Martin. 2023. Phonetic accommodation in non-native directed speech supports L2 word learning and pronunciation. *Scientific Reports* 13, 1 (2023), 21282.
- [60] Giorgio Piazza, Clara D Martin, and Marina Kalashnikova. 2022. The acoustic features and didactic function of foreigner-directed speech: A scoping review. *Journal of Speech, Language, and Hearing Research* 65, 8 (2022), 2896–2918.
- [61] Jennifer Parker Porter and Krista L Olson. 2001. Anthropometric facial analysis of the African American woman. *Archives of facial plastic surgery* 3, 3 (2001), 191–197.
- [62] Valentijn Prové, Bert Oben, and Julien Perrez. 2022. Foreigner directed gesture: larger, faster, longer. *Foreigner Directed Gesture: larger, faster, longer* 8 (2022), 1–31.
- [63] Jacob M Rabbie and Gerard Wilkens. 1971. Intergroup competition and its effect on intragroup and intergroup relations. *European Journal of Social Psychology* 1, 2 (1971), 215–234.
- [64] R Ramamurti. 1980. Strategies involved in talking to a foreigner. *PENN Review of Linguistics* 4 (1980), 84–93.
- [65] Byron Reeves and Clifford Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK* 10, 10 (1996), 19–36.
- [66] Katherine Reynolds. 2015. Social Identity in Social Psychology. *International Encyclopedia of the Social Behavioral Sciences* 2 (2015). <https://doi.org/10.1016/B978-0-08-097086-8.24064-6>
- [67] William Jamal Richardson. 2018. Understanding Eurocentrism as a structural problem of undone science. *Decolonising the university* (2018), 231–247.
- [68] Jörg Roche. 1998. Variation in Xenolects. *sociolinguistica* (1998), 117–139.
- [69] Jörg Matthias Roche. 2006. Wie Deutsche mit Ausländern sprechen und was man daraus für den Spracherwerb und den Sprachunterricht lernen kann. *Informationen Deutsch als Fremdsprache* 33, 5 (2006), 395–410.
- [70] Brent Rossen, Kyle Johnsen, Adeline Deladisma, Scott Lind, and Benjamin Lok. 2008. Virtual humans elicit skin-tone bias consistent with real-world skin-tone biases. In *Intelligent Virtual Agents: 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008. Proceedings* 8. Springer, 237–244.
- [71] Kathrin Rothermich, Rose Baker, Sharon Falkins, Jaeyoung Kum, Madison Capps, and Susan C Bobb. 2021. Everyday prevalence of foreigner-directed speech. <https://doi.org/10.31235/osf.io/jwefk>
- [72] Kathrin Rothermich, Havan L Harris, Kerry Sewell, and Susan C Bobb. 2019. Listener impressions of foreigner-directed speech: A systematic review. *Speech Communication* 112 (2019), 22–29.
- [73] Ellen Bouchard Ryan, Janet M Hamilton, and Sheree Kwong See. 1994. Patronizing the old: How do younger and older adults respond to baby talk in the nursing home? *The International Journal of Aging and Human Development* 39, 1 (1994), 21–32.
- [74] Joanna Sankowska, María Luisa García Lecumberri, and Martin Cooke. 2011. Interaction of intrinsic vowel and consonant durational correlates with foreigner directed speech. *Poznań Studies in Contemporary Linguistics* 47, 1 (2011), 109.
- [75] Rebecca Scarborough, Olga Dmitrieva, Lauren Hall-Lew, Yuan Zhao, and Jason Brenier. 2007. An acoustic study of real and imagined foreigner-directed speech. *Journal of the Acoustical Society of America* 121, 5 (2007), 3044.
- [76] Tanja Schneeberger, Mirella Scholtes, Bernhard Hilpert, Markus Langer, and Patrick Gebhard. 2019. Can social agents elicit shame as humans do? In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 164–170.
- [77] Luwi J Shamambo and Tracey L Henry. 2022. Rethinking the use of “Caucasian” in clinical language and curricula: a trainee’s call to action. *Journal of General Internal Medicine* 37, 7 (2022), 1780–1782.
- [78] Catherine E Snow, Roos Eeden, and Pieter Muysken. 1981. The interactional origins of foreigner talk: Municipal employees and foreign workers. *International Journal of the Sociology of Language* (1981).
- [79] Jordan Soliz and Howard Giles. 2014. Relational and identity processes in communication: A contextual and meta-analytical review of communication accommodation theory. *Communication yearbook* 38 (2014), 107–143.
- [80] Paul Starr. 2023. The re-emergence of “People of Color”. *Du Bois Review: Social Science Research on Race* 20, 1 (2023), 1–20.
- [81] Henri Tajfel. 1979. An integrative theory of intergroup conflict. *The social psychology of intergroup relations/Brooks/Cole* (1979).
- [82] Henri Tajfel, Michael G Billig, Robert P Bundy, and Claude Flament. 1971. Social categorization and intergroup behaviour. *European journal of social psychology* 1, 2 (1971), 149–178.
- [83] Henri Tajfel and Alan L Wilkes. 1963. Classification and quantitative judgement. *British journal of psychology* 54, 2 (1963), 101–114.
- [84] Marion Tellier, Gale Stam, and Alain Ghio. 2021. Handling language: How future language teachers adapt their gestures to their interlocutor. *Gesture* 20, 1 (2021), 30–62.
- [85] John C Turner, Rupert J Brown, and Henri Tajfel. 1979. Social comparison and group interest in ingroup favouritism. *European journal of social psychology* 9, 2 (1979), 187–204.
- [86] John C Turner and Katherine J Reynolds. 2011. Self-categorization theory. *Handbook of theories in social psychology* 2, 1 (2011), 399–417.
- [87] Maria Uther, Monja A Knoll, and Denis Burnham. 2007. Do you speak E-NG-LI-SH? A comparison of foreigner- and infant-directed speech. *Speech communication* 49, 1 (2007), 2–7.
- [88] Robert S Wachal. 1994. The Capitalization of Color Names and Other Terms for Ethnic Groups. *Publication of the American Dialect Society* 78, 1 (1994), 81–90.
- [89] Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, and Elisabeth André. 2021. “Let me explain!”: exploring the potential of virtual agents in explainable AI interaction design. *Journal on Multimodal User Interfaces* 15, 2 (2021), 87–98.
- [90] Davis P. Wilkins. 1996. *Morphology*. Walter de Gruyter, 109–117. 10.1515/9783110132649.1
- [91] Howard Winant. 2000. Race and race theory. *Annual review of sociology* 26, 1 (2000), 169–185.
- [92] Blue Wooldridge. 2001. ‘Foreigner talk’: An important element in cross-cultural management education and training. *International review of administrative sciences* 67, 4 (2001), 621–634.
- [93] Frank C Worrell. 2014. Culture as Race/Ethnicity. *The Oxford handbook of identity development* (2014), 249–268.
- [94] Jan Záhofík. 2010. Languages in Sub-Saharan Africa in a broader socio-political perspective. *Acta Orientalia Vilnensia* 11, 2 (2010), 77–90.
- [95] Jane Zuengler. 1991. Accommodation in native-nonnative interactions: Going beyond the “what” to the “why” in second-language research. *Contexts of accommodation: Developments in applied sociolinguistics* (1991), 223–244.