MorphHat: A Humanoid Robot Interpreter for Enhancing Multilingual Collaboration

Sandra Müller
Information Systems (WIN)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
sandra.mueller@kit.edu

Martin Feick
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
martin.feick@kit.edu

Alexander Maedche
human-centered systems lab (h-lab)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
alexander.maedche@kit.edu

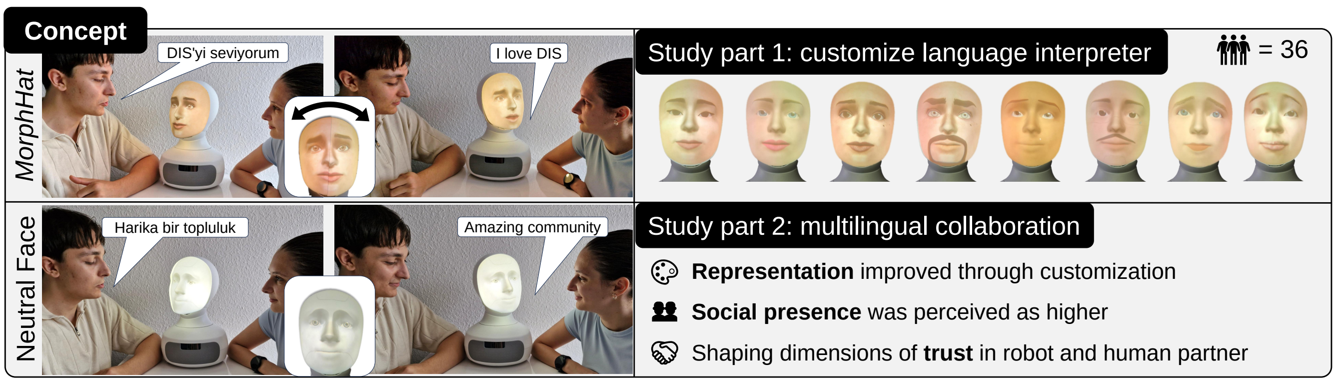

Figure 1: Multilingual collaboration with *MorphHat*. On the left, two participants without a shared language communicate via a humanoid robot interpreter, which either morphs its face to represent the speaker (top) or stays neutral (bottom). Through morphing and customization, *MorphHat* aims to build trust and a feeling of being represented, especially when users depend on translations in unfamiliar languages. On the right, examples of customized, self-resembling robot faces created by participants.

Abstract

Multilingual collaboration is increasingly common as today's world becomes global and culturally diverse. While diversity fosters innovation, language barriers can hinder involvement and effective communication. Prior work has primarily focused on improving translation accuracy with limited attention to how translation systems shape dimensions of trust in interaction. Given that users must rely blindly on technology due to their inability to understand the system's output, this issue becomes a crucial aspect. To address this gap, we introduce *MorphHat*, a co-embodied humanoid robot interpreter featuring a customizable, morphing face that visually represents the active speaker. We evaluated *MorphHat* in multilingual conversations through a study with 36 participants, focusing on key components of successful collaboration. While quantitative results showed no significant differences, qualitative findings suggest that *MorphHat* influenced trust, rapport, and social presence. We discuss these insights as early design implications for future embodied translation systems.

DIS '26, Singapore, Singapore

ACM ISBN 979-8-4007-2563-0/26/06
https://doi.org/10.1145/3800645.3813009

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; **Interaction design**.

Keywords

Human–Robot Interaction; Multilingual Communication; Face Morphing; Humanoid Robot Interpreter; Embodied Machine Translation

ACM Reference Format:
Sandra Müller, Martin Feick, and Alexander Maedche. 2026. MorphHat: A Humanoid Robot Interpreter for Enhancing Multilingual Collaboration. In *Designing Interactive Systems Conference (DIS '26), June 13–17, 2026, Singapore, Singapore.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3800645.3813009

1 Introduction

Multilingual collaboration is increasingly relevant as today's world becomes more global and culturally diverse [24]. One of the driving factors for successful collaboration is effective communication; yet, it is often challenging, as individuals without a shared language rely on technological support to enable communication [11]. While basic translation through mobile apps or ubiquitous devices such as earbuds[1] is often sufficient for simple everyday tasks, such as ordering food or asking people for navigational advice, human-to-human

[1]Apple offers real-time translation functionality with AirPods Pro, see: https://www.apple.com/de/airpods-pro/

collaboration requires a deeper level of involvement. Imagine working professionals that have recently moved to a different country and now faces cultural and communication challenges. They need to collaborate with people to solve problems, but they struggle to feel understood and included, which can make life confusing and intimidating [2]. They experience a common problem that arises when language barriers are involved. Generally, people feel most comfortable in their native language, which strongly influences their ability to engage in collaboration and contribute successfully [11]. Native language use enables emotional expression and clearer communication [3], whereas limited proficiency in a foreign language can even lead to anxiety [29]. People in such collaborative settings often feel misunderstood [35]. This creates a strong dependency on trust that their messages are conveyed accurately and not misinterpreted [21]. Accurate message delivery relies on interpreters or translation systems, requiring speakers to trust that these systems convey their intent, even when they cannot verify the translation themselves. Thus, one of the key aspects of technology that enable multilingual communication need to be designing for trust. Moreover, individuals in multilingual settings often do not feel seen or valued, because language barriers create dependency on others for interpretation, leading to loss of privacy, perceived lower status, and anxiety around conversation [11, 35]. Representation is essential for inclusive communication because it signals presence and fosters a sense of belonging [2, 31]. For those navigating unfamiliar cultural or linguistic environments, feeling represented is critical for building confidence in these situations. One effective way to promote this is through visual representations, which help collaborators feel acknowledged and valued [3]. At the same time, current translation technologies primarily prioritize linguistic accuracy [14, 17, 35, 41, 62], while rarely considering how speaker switching should be shown. Prior systems (e.g., Socibot) often introduce predefined characters rather than visual representations of the human speaker during interaction [22]. Moreover, existing explorations of visual representation remain limited to one speaker, leaving speaker-to-speaker morphing unexplored [71].

Human interpreters are a strong example of effective translation because they not only ensure linguistic accuracy but also foster trust and act as representatives of the speaker. They are widely regarded as the gold standard in language translation, as they go beyond translating words to navigating interpersonal dynamics and cultural nuances [49, 64]. In both business and community contexts, such as healthcare, legal, and social services, their presence is often indispensable [16, 61]. A key strength of interpreters lies in their ability to adapt to the speaker's intent, which helps build trust between conversation partners [25, 50]. This trust is particularly critical in multilingual communication, where even minor misunderstandings or the absence of an interpreter can escalate into strained relationships or reputational harm [12, 57]. Interpreters effectively serve as the speaker's voice in another language, functioning as representatives who embody the speaker's identity, similar to how political interpreters represent their country on the global stage [12]. This representation allows speakers to remain present and engaged in the conversation while creating a buffer from the exact phrasing they cannot fully verify. Despite their effectiveness, human interpreters are not scalable for everyday interactions and remain underutilized, even in contexts requiring high interpersonal sensitivity [37]. Technology-based solutions, by contrast, can support a broader range of languages, making them more adaptable for inclusive communication.

Inspired by human interpreters, we propose *MorphHat*, a humanoid robot interpreter featuring facial customization and dynamic face morphing to represent the active speaker during conversation. Facial customization creates psychological ownership [15], enabling users to project a self-created face onto the robot, making them feel represented in the collaborative setting. By representing the user, *MorphHat* is intended to support feelings of involvement and trust, allowing the robot to effectively "stand in" for the user and convey their message. Furthermore, *MorphHat*'s embodied design may support social presence, overcoming collaborators' feelings of being left on their own and making conversations more engaging [3]. While we focus on multilingual collaboration, *MorphHat* is designed for broader sensitive multilingual communication in professional and community contexts, such as public services, where consultations on health, financial matters, or negotiations require that people feel understood and supported. In this first iteration, we focused on dyadic multilingual collaboration to evaluate our approach. We ground *MorphHat* in the concept of co-embodiment [40], where the robot becomes a shared communicative representative between speakers. **To the best of our knowledge, this represents a novel design approach that combines the scalability of machine-translation systems with the richness of human interpreters.**

To investigate whether *MorphHat* supports these possibilities, we conducted a within-subject design lab study with 36 participants (18 language-diverse dyads). Participants collaborated via a humanoid robot interpreter in two conditions: one with a neutral face, and one with *MorphHat*. Since physical embodiment could split attention between the robot and the human speaker, potentially affecting rapport, i.e., the relationship between collaborators, but also the robot [74], we included it as a control variable between conditions. Our quantitative data showed no statistically significant differences between conditions. However, qualitative insights revealed that the perceived role of the interpreter shifted from a tool (neutral face) to a more supportive presence (*MorphHat*), indicating a trend in feelings of representation, dimensions of trust, rapport, and social presence.

Following Wobbrock and Kientz [68] classification of research contributions in HCI, our primary contribution is an artifact contribution supported by a mixed-methods exploratory evaluation, where we emphasize the novelty of using facial morphing for speaker representation in embodied translation. Our contributions can be summarized as:

(1) Developing *MorphHat*, a humanoid robot interpreter featuring dynamic face morphing that supports multilingual dyadic collaboration through psychological ownership and representation.

(2) Investigating how *MorphHat* influences not only human-robot interaction, but also interpersonal dynamics between human participants.

(3) Designing insights for humanoid robot interpreters tailored to multilingual collaborations, emphasizing the role of psychological ownership in shaping dimensions of trust.

2 Related Work

Here, we review prior work on multilingual collaboration, language interpreting and machine translation as well as human-robot collaboration, highlighting insights relevant to the design of *MorphHat*.

2.1 Multilingual Collaboration

As a result of globalization, modern collaboration frequently spans cultural and disciplinary boundaries [26]. This comes with multilingualism, posing challenges that can significantly affect collaboration effectiveness [11]. To this day, language barriers remain a persistent issue, particularly in on-site collaborations where clear and immediate communication is crucial [73]. Individuals feel more confident and expressive in their native language, which can limit their ability to contribute fully to collaboration [2, 11]. Effective collaboration is shaped not only by shared objectives but also by a nuanced blend of social dynamics [26]. Here, trust plays a central role in maintaining a healthy dynamic between collaborators, alongside other essential elements such as effective communication and coordination [8]. When trust is well established, it serves as a foundation for creativity, innovation, and enables seamless collaboration [24]. Equally important are interaction qualities, like rapport, which is closely linked to entrainment, the synchronization of behaviors during interaction, which fosters interpersonal alignment and strengthens collaboration [38]. In multilingual settings, collaboration introduces additional linguistic and cultural complexities that influence how connections are built, decisions are made, and communication unfolds [19, 20]. **The persistent challenge of language barriers in multilingual collaboration calls for solutions that foster trust and rapport.**

2.2 Language Interpreting & Machine Translation

Human interpreters remain essential, enabling multilingual communication in dyadic or triadic settings [16, 61]. Prior work has emphasized their ability to adapt to spatial and emotional dynamics, which contributes to trust and relational depth [25, 50]. These qualities make human interpreters especially effective in sensitive environments, though they are not always scalable or available for everyday use [49]. In diplomatic settings, interpreters play a crucial role in preserving both the accuracy and intent of communication. It is common practice for nations to bring their own interpreters to high-level meetings, as these individuals are not only trusted to convey messages faithfully but also serve as representatives of their country [12, 57]. In the absence of human interpreters, people often turn to mobile translation apps as convenient tools for overcoming language barriers in everyday multilingual communication. These systems support a wide range of languages, making them broadly accessible. However, while machine translation can bridge linguistic gaps across many languages, it also introduces challenges in achieving mutual understanding, because they are not aware of how each message is translated into other languages [69]. Despite recent advances, most systems still prioritize linguistic accuracy over fostering trust in the system and preserving the speaker's representation [14, 17, 35, 41, 62]. Research has already presented technological support tools that can foster equitable participation between people. For example, Chen et al. [11] integrated machine translation and shared displays helps non-native speakers participate more fully in group discussions. While carefully designed machine translation can aid comprehension [21, 30], such tools are rarely embedded in socially embodied systems. With translation technologies and human interpreters playing a crucial role in bridging language gaps, effective multilingual collaboration requires more than accurate translation. It involves supporting inclusive group dynamics [11]. **There is a high need for translation systems that combine the richness of human interpreters with the accessibility of everyday tools.**

2.3 Human-Robot Collaboration

By leveraging human-like behaviors and aligning socially through visual cues, robots have been shown to enhance trust, rapport, and social presence in diverse conversational settings [56, 74]. Features like matched ethnicity and head pose increase perceived usefulness and sociability [32, 52], while interactive capabilities boost social presence [34, 54, 59]. Robots with human-like characteristics (i.e., humanoids) are often viewed as more likable and emotionally engaging than virtual agents [63]. Recent work expands human-robot-human collaboration beyond physical tasks [18] to cognitive collaboration, emphasizing team dynamics over individual interactions [53]. While this presents great potential, in the context of language interpreting, humanoid robots are currently used primarily for sign language, offering multimodal communication by integrating gestures and visual cues [28]. These systems help reduce cognitive load during interactions [39] and enable expressive, real-time communication. **To pioneer multilingual collaboration, there is an opportunity for embodied humanoid robots that foster connection beyond words.**

2.4 Face Displaying and Morphing

Robot faces can visually represent speakers and shape perceived presence during interaction. Systems such as SociBot [22] display predefined avatar characters onto a robotic face to create a sense of co-presence. Similarly, early design explorations, such as WorldHat, employed a static neutral robot face for interpretation, without adapting to individual speakers [45]. However, these approaches rely on static character displays and offer no dynamic user influence. In telepresence designs, for example, a remote speaker can be projected onto a 3D, face-shaped display [44], or shown on a screen that replaces a robot's head, increasing perceived anthropomorphism compared to a fixed robot face [72]. Morphing and transforming faces extend these approaches by enabling adaptive speaker representation. Prior work shows that morphing a speaker's face, for example by altering facial expressions to convey emotional states during video conferencing can increase creativity during interaction [47]. Morphing can also be used between a human and robot face, as presented in Tele-Vestige [71]. Another system goes even further by placing a live, motion-mapped avatar onto a humanoid robot to better align the robot's appearance with the speaker, demonstrating that resemblance can shape control perceptions [23]. **Despite these efforts, no prior work has used dynamic identity morphing to continually indicate who is speaking in real-time conversation.**

3 MorphHat

We propose *MorphHat*, a customizable humanoid robot interpreter designed to support multilingual collaboration. The system functions as a socially co-embodied entity, providing a solution for scenarios where participants share the same physical space.

3.1 Design

MorphHat draws inspiration from how human interpreters represent speakers in interaction, functioning as an interpreter that not only translates but also provides a visual representation for who is currently speaking. Our design draws on prior work showing that customizing can influence how people relate to digital characters [9]. Rather than assuming specific psychological effects, we use customization as a way to let users select how they prefer to be visually represented. This customization process allows users to design a face that reflects their self-perception, giving them creative freedom over its appearance. Visual similarity may also matter for some users, as suggested by homophily research [55], though its role in embodied translation remains open. However, designing for accurate homophily raises ethical considerations that must be carefully managed [9]. Users may want to customize their representation in ways that don't necessarily look like themselves physically, but still feel representative [31]. *MorphHat* leverages this by allowing users to customize a robot face that reflects their self-perception, rather than replicating their exact appearance.

Coordination. Aligning and synchronizing actions between collaborators is an essential part of human collaboration [8], and *MorphHat* supports this process by structured turn-taking inspired by the behavior of human interpreters in dyadic conversations. Using eye and head gaze shifts and speech timing, the robot maintains a coordinated interaction flow between participants. Figure 2 illustrates the interaction sequence. The robot begins in a neutral resting state. When a user starts talking to *MorphHat*, it turns to the user with the customized face of the conversational partner. *MorphHat* listens, translates the message, and turns to the other person to speak, thereby morphing its face to represent the sender of the message. The face morphing process occurs when the robot shifts its gaze from the center of the room (or from the other participant) to the active speaker. The gaze transition and face morphing are synchronized so that the customized face is fully visible once the robot has completed its turn toward the user. If no speech input is received within 2 seconds, *MorphHat* returns to its resting state while retaining the last displayed face. A neutral resting face was introduced for two reasons: (1) it signals errors, for example, if MorphHat stops facing a participant and re-centers, users know the system did not register input; and (2) it avoids social bias by staying neutral when no one is speaking, reducing pressure to respond immediately. The 2-second return interval was informed by pilot testing, which showed that providing quick feedback enabled users and evaluators to detect issues promptly without introducing awkward silence or immediate pressure to respond.

Social Cue Rendering. Speech was generated using Furhat's built-in say state, which manages a speech synthesis queue and supports asynchronous execution, enabling gestures and behaviors to be integrated mid-utterance for natural prosody. High-precision lip-sync

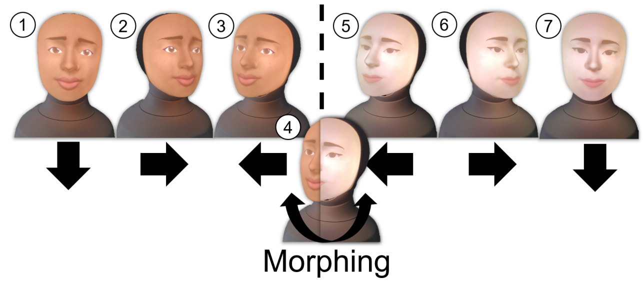

Figure 2: Interaction flow of *MorphHat*. In the beginning, the robot is in a neutral state (1). It then turns toward the active speaker and listens (2). *MorphHat* delivers the translated message to the listener (3). After the message is delivered, the robot morphs its face to the other user's customized face (4). The process is repeated in reverse, with the robot adapting its face to the current speaker (5-7).

was automatically generated through Montreal Forced Aligner[2], supported by Furhat Robotics. It creates .pho files for forced alignment, delivering superior lip-sync quality compared to automatic phoneme recognition. Furhat actively displayed social cues such as attentive gaze, head nods, and eyebrow movements while listening, and facial gestures during speech. These behaviors were complemented by reactive gestures, such as eyebrow raises on speech prominence, as well as continuous micro-expressions (e.g., blinking, subtle facial movements, and small gaze shifts) to make the robot appear more "alive." No cues were shown during our face morphing to avoid unnatural transitions. All social cues were system-driven and not manually triggered. The Furhat Robotics documentation provides detailed information about these features[3].

3.2 Implementation

The system is built on the Furhat Robotics platform[4] (serial number F479), which features a projected face on a 3D mask for expressive interaction. It supports multimodal communication via facial animation, gaze control, and motorized head movements. Both conditions, *MorphHat* and the neutral face used in the study, were implemented in the same way. The only difference between conditions was the change in facial design. While the system leverages the Furhat SDK as a base framework, our implementation required integration work to enable fluent language switching and therefore a multilingual interaction. We designed a custom protocol to orchestrate multiple REST API calls and synchronize robot behaviors with user-triggered events.

Multilingual Interpretation Pipeline. The interpretation pipeline combines modular components for real-time multilingual interaction. Dialogue management is implemented using the Furhat SDK, which operates on a state-machine architecture to coordinate conversations through defined states, transitions, and events. Audio input is transcribed using Azure Speech-to-Text and translated via

[2]Montreal Forced Aligner aligns audio with text using a pronunciation dictionary, see https://montreal-forced-aligner.readthedocs.io/en/v3.3.0/user_guide/index.html.
[3]Furhat Robotics documentation, see: https://docs.furhat.io
[4]Furhat Robotics platform, see https://www.furhatrobotics.com

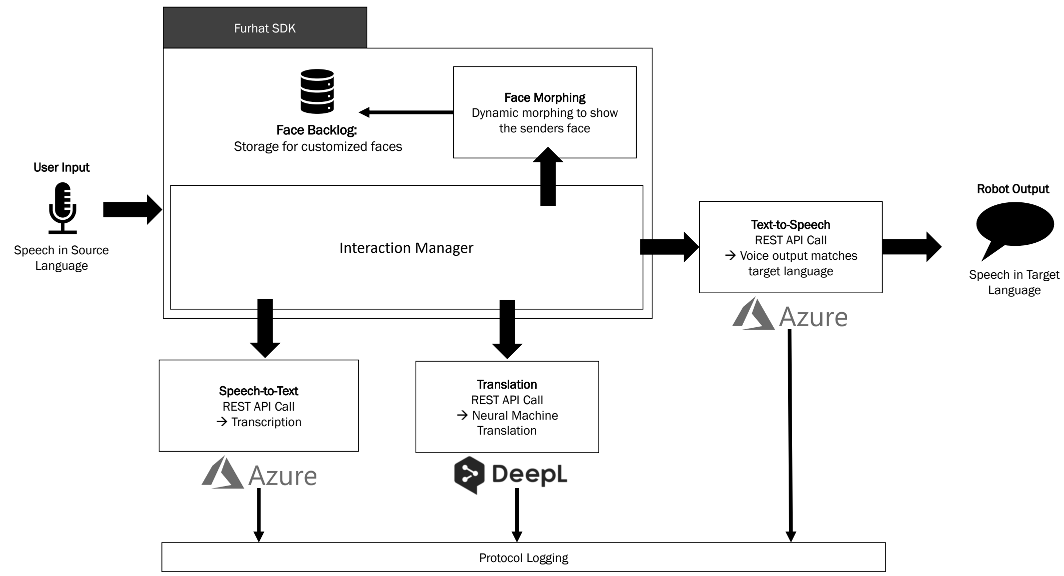

Figure 3: Pipeline architecture for *MorphHat*, built on the Furhat SDK in Kotlin. User speech is processed via Microsoft Azure Speech-to-Text, translated through DeepL API Pro, and converted back to speech using Microsoft Azure Text-to-Speech before the robot vocalizes the output.

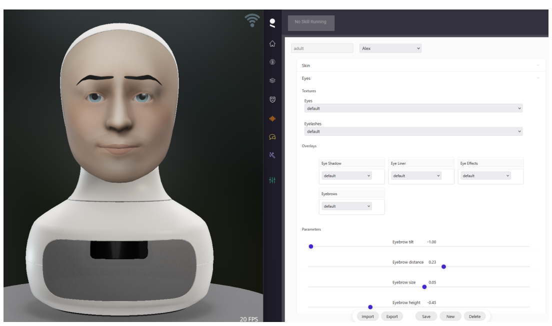

Figure 4: Furhat SDK Face Editor with split-screen view: customization sliders on the right and real-time changes rendering on the left. The participants started with the displayed default character (model: Alex) for their customization.

DeepL API Pro[5]. Unlike large language models, which are more likely to generate unpredictable outputs, neural machine translation systems like DeepL are optimized for accurate word-by-word translation [48, 58]. Finally, Azure Text-to-Speech synthesizes the translated text, which is delivered by the robot with synchronized gaze and orientation. The system currently supports six languages: German, English, French, Spanish, Turkish, and Chinese-Mandarin. These were selected based on local user demographics at our institution and validation availability. While the experimental setup used these predefined languages, the architecture is scalable and can support other languages for bidirectional translations. Speech input relies on a trigger mechanism, i.e., participants used a button to open the microphone, rather than real-time automatic speaker diarization because the latter proved unreliable in our pilot testing. All sessions were logged, including the heard sentence, translated output, and corresponding timestamps to identify error sources (e.g., unclear speech vs. translation inaccuracies).

Facial Customization. The Furhat SDK (version 2.8.3) includes a built-in face editor that enables face customization. Users can select a base character (e.g., skin tone, sex) and then refine its facial design. Participants were also able to choose the voice gender (female or male) for their character, as voice gender can influence how voices are perceived in interaction [42]. Customizable aspects of the face included a wide range of features, with options often supplemented by color wheels and adjustable sliders:

- **Skin:** skin tone selection via color wheel and skin texture and overlays (e.g., freckles, 80 options), plus cheekbone and nose positioning using 5 continuous sliders;
- **Eyes:** eye styles (93), eyelashes (8), eye makeup (26), eyebrows (72), and eye positioning via 10 continuous sliders;
- **Mouth:** teeth (6), lips (39), and mouth positioning using 7 continuous sliders;
- **Hair and Other:** piercings (15), tattoos (18), facial hair (43), and miscellaneous makeup (15).

All modifications are rendered in real-time within the Furhat environment, allowing immediate visual feedback (see Figure 4). In addition to using the face editor, we integrated the face morphing into the program, ensuring that the users facial avatars were synchronized with the conversation and added the morphing logic into the programming code.

[5]DeepL, see https://www.deepl.com

4 Lab Study

We conducted a within-subjects lab study to examine how *MorphHat* influences multilingual dyadic collaboration. The goal of the study was not only to understand the effects between the robot and the human, but also how *MorphHat* might shape the relationship between collaborators. The study was approved by the university's ethics and data protection committee. Participants were informed that participation was voluntary and that they could resign at any time without penalty. In the study, we asked them to solve two collaborative tasks using two conditions:

(1) **Neutral Face**: In the baseline condition, the robot used a neutral-looking default SDK face (model: Titan) and maintained a consistent identity throughout the interaction.
(2) ***MorphHat***: Participants customized their own interpreter. Upon activation, the robot displayed the speaker's customized face as described previously.

Both the conditions and the tasks were counterbalanced, and turn-taking and translation behaviors remained constant to isolate the effects of *MorphHat*. We used a mixed-methods approach combining standardized questionnaires and qualitative interviews to assess key metrics of collaboration: trust, rapport and social presence. Due to our small sample size, qualitative data served as the primary source for interpreting.

4.1 Task

Participants completed two collaborative decision-making tasks designed to encourage verbal interaction and engagement with the humanoid robot interpreter. The tasks were selected based on their established use in collaboration research and their suitability for multilingual interactions without requiring domain-specific knowledge. In the "Survival on the Moon" task [1], participants engaged in a collaborative decision-making exercise. The task involved selecting 10 out of 15 items considered essential for survival following a spaceship crash. To avoid repetition, a second scenario, "Survival at Sea" [51], was used. In this task, participants imagined being stranded on a life raft after a shipwreck and again selected 10 out of 15 items to survive. To ensure ease of translation across languages and mutual understanding, simplified item descriptions were provided for both tasks. As both scenarios involve negotiation and prioritization, they are well-suited for studying collaboration and naturally encourage participants to engage in discussion. Full details of the tasks are provided in the supplementary materials.

4.2 Participants

36 participants (20 male, 16 female; M = 24.22 years, SD = 3.7) from diverse academic backgrounds were recruited via flyers, mailing lists, and an institutional study pool. Each dyad included one native German speaker and one non-German speaker fluent in Chinese Mandarin, Turkish, Spanish, or French. Participants were paired to ensure no mutual language understanding, e.g., German speakers did not understand their partner's language, and vice versa. All participants had at least an upper-intermediate level of English proficiency (equivalent to B2 on the standardized CEFR scale[6]), which ensured they could comprehend the study instructions, fill-in the questionnaires and participate in the post-study interview. The overall language distribution was as follows: English (36), German (18), Chinese Mandarin (6), Turkish (6), Spanish (5), and French (1). Language proficiency was self-reported. Participants were also asked about their use of translation systems, including frequency, preferred tools, and typical purposes. Usage varied, with 20 participants reporting daily use, 10 weekly, and a few using translation tools monthly (3) or rarely (3). The most commonly used systems were Google Translate (26), ChatGPT (25), and DeepL (21). Participants primarily used these tools for text-based tasks such as academic reading (29), personal communication (21), writing assignments (16), and lecture translation (15).

[6] B2 refers to the upper-intermediate level of the Common European Framework of Reference for Languages (CEFR), indicating the ability to understand complex texts and communicate effectively in English. See Council of Europe CEFR Levels.

4.3 Experimental Protocol

Upon arrival, participants were briefed and informed that *MorphHat* would act as their personal interpreter during the study, representing them and serving as their voice in another language. They were instructed to customize the robot's face to create a companion that would represent them using the Furhat environment, adjusting features such as skin tone, facial structure, and select voice gender. Participants performed this task on separate laptops without seeing one another. After a warm-up task introducing a fictional character, each dyad completed two 15-minute collaborative tasks: one with a neutral face and one with *MorphHat*. Both the condition order (Neutral Face vs. *MorphHat*) and task type ("Survival on the Moon" vs. "Survival at Sea") were counterbalanced across dyads (e.g., Moon–Neutral → Sea–*MorphHat*; Moon–*MorphHat* → Sea–Neutral; Sea–Neutral → Moon–*MorphHat*; Sea–*MorphHat* → Moon–Neutral). After each task, participants completed the rapport and trust questionnaires during a 10-minute break, and an additional final questionnaire about their overall experience was administered after both conditions. One participant per dyad was randomly selected for a post-study semi-structured interview. The study lasted approximately 90 minutes, with compensation ranging from €20 to €25 depending on interview participation. The full experimental protocol can be seen in Figure 5.

4.4 Apparatus

The study took place in a controlled lab environment. We used the *MorphHat* system, but instead of capturing voice input through the built-in microphone of Furhat, we used a RØDE NT-USB directional microphone to improve audio quality. Turn-taking was triggered by participants using a Logitech wireless presenter to avoid technical problems with speaker diarization due to the variety of different languages [46]. The robot was positioned centrally on a table between two participants, forming a triangular interaction setup commonly used in collaborative interpreting scenarios [50, 64]. Both conditions in our study used the same multilingual interpretation pipeline described in section 3.2.

4.5 Data Collection

We used a mixed-methods approach combining validated scales for trust, rapport, and qualitative interviews to better capture participants' experiences with *MorphHat*. Participants completed questionnaires after experiencing each condition and one final questionnaire

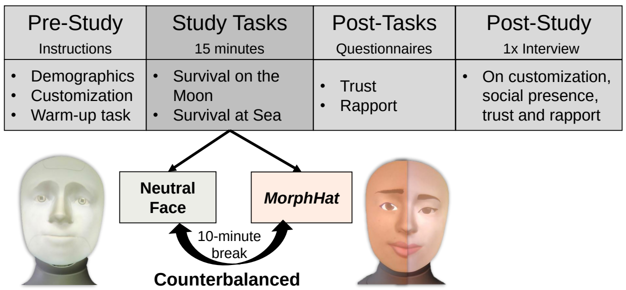

Figure 5: Overview of the experimental procedure: participants first completed a demographics survey and face customization, followed by a warm-up task. They then performed two collaborative tasks under different conditions, each followed by a questionnaire. Finally, participants completed a post-study questionnaire and took part in an interview to provide qualitative insights.

in the end. Qualitative data was collected through semi-structured interviews with randomly selected participants (one from each dyad). All questionnaires and interviews were conducted in English. Interview topics included general impressions of each condition, face customization, social presence with inspiration from Biocca et al. [6], Harms and Biocca [27], and perceptions of trust and rapport. Trust and rapport were measured with validated items, varying in format (5-, 7-, and 8-point scales), and reverse-coded items to preserve consistency from the original questionnaires:

- **Rapport Toward Human Conversation Partner:** Based on rapport components from Tickle-Degnen and Rosenthal [60] and further developed by Bernieri et al. [5] (8-point scale).
- **Trust Toward Human Conversation Partner:** Adapted from trust dimensions by McCroskey and Young [43], using 7-point scale for bipolar adjective format (e.g., caring–uncaring).
- **Trust Toward Robot:** Parallel adaptation of the human-trust items (7-point bipolar scale).
- **Rapport Toward Robot:** Connection-Coordination Rapport (CCR) Scale from Lin et al. [36] (5-point Likert-Scale).

4.6 Data Analysis

We performed a mixed-methods analysis, placing emphasis on qualitative data (N=18, P1–P18) due to the exploratory nature of the study. For practical reasons, we interviewed only one participant per dyad, as the study was conducted by a single experimenter. Additionally, we wanted to explore certain dimensions, e.g. the perception of their partner's competence, goodwill, and trustworthiness. Interviewing them alone allowed for more honest responses, which wouldn't have been possible with the partner present. Interviews were recorded, transcribed using an internal system offered by our university based on OpenAI Whisper, and analyzed using reflexive thematic analysis following the six-phase framework of Braun and Clarke [7]. Our coding combined a deductive frame (e.g., Customization, Social Presence, Trust and Rapport) with inductive codes that emerged during analysis (e.g., Trust Towards System evolved from Trust). Theme development was iterative. Initial codes were clustered into candidate themes and refined through memoing. All supporting materials are provided in the supplementary material, including a coding frame overview, simplified and detailed theme maps, exemplar excerpts (including disconfirming cases), a short reflexivity statement, and reflexivity notes with analytic memos.

For our quantitative statistical analysis, we began by inspecting the raw dataset (36 participants, 72 rows) for outliers. Given the small sample size and the absence of severe outliers, we continued with the original data. Descriptive statistics (mean, median, standard deviation) were computed to summarize distribution. To ensure internal consistency of the multi-item measures, we calculated Cronbach's α (see Appendix A). We fitted linear mixed-effects models for our dependent variables in R[7] using `lme4` [4] and `lmerTest` [33], accounting for dyadic random effects and fixed effects (see Appendix A). We adopted an exploratory approach to specifying fixed effects. The primary fixed effect was the experimental condition (Neutral Face vs. *MorphHat*); additional fixed effects, such as participants' language and prior machine translation usage, were also considered. Our dependent variables focused on the human: Rapport Self-Rated, Rapport Interaction-Based, Trust Competence, Trust Trustworthiness, Trust Goodwill, and the translation technology (i.e., the robot): Rapport Connection, Rapport Coordination, Trust Competence, Trust Trustworthiness and Trust Goodwill. We fitted linear mixed-effect models for each dependent variable with a fixed effect *Condition* (Neutral Face, *MorphHat*). Dyad ID was modeled as a random effect to account for variability between pairs, treating dyads as a random sample representative of the population. We iteratively refined models including and excluding fixed and random effects, but more complex models did not reveal statistical relevant insights or did not reach convergence. We reported R^2_{marginal} and $R^2_{\text{conditional}}$ to quantify the explain variance by our models. Robust inference was applied through non-parametric bootstrap (number of samples = 1000) to mitigate assumption violations. The complete analysis scripts can be found in the supplementary materials.

4.7 Results

In this section, we present our findings combining both quantitative measures and qualitative interviews. All 36 participants successfully completed the two collaborative tasks, with each task lasting about 15 minutes. Beforehand, they were able to customize a face for *MorphHat* within 10 minutes (see Figure 6). Internal consistency was assessed using Cronbach's alpha for all questionnaire scales (see Table 1 in Appendix). Reliability was acceptable to excellent across all scales, with values ranging from $\alpha = .730$ to $\alpha = .943$. Across measures, marginal R^2 values were very low ($R^2_m \leq .20$), while conditional R^2 values were substantially higher (up to $R^2_c = .76$).

***MorphHat* as self-representation.** 27 participants designed *MorphHat* to resemble themselves, often incorporating specific personal features. As one participant explained, *"She is going to speak*

[7]R, a free software environment for statistical computing and graphics, see https://www.R-project.org

Figure 6: Customized robot faces created by all 36 participants. Most designs reflect the participants' own facial features, while a few outliers explored more imaginative styles.

with the robot and I'm going to speak with her robot, so it should resemble us." (P4). Others emphasized distinct facial traits: *"I chose blue eyes and braces because I'm wearing them."* (P12), and *"I considered my eye color, because I feel like that's a very distinct feature in every face."* (P14). Even small details like birthmarks were included: *"It was a fun detail to see the birthmarks in the customization, I thought, let's choose those and make the robot more like me."* (P14). While most aimed for personal resemblance, a few leaned toward idealized versions: *"That's what I would like to look like."* (P15), or adapted *MorphHat* to personal beauty standards: *"I try to make it more beautiful in my beauty standard."* (P5). Cultural differences were also noted: *"I found two very different beauty standards. When seeing the robot faces... a European face and an Asian beauty standard face."* (P5). Interestingly, a subset of faces deviated markedly from this pattern, creating what we term "outlier faces." These included stylized characters with drawn facial features (3), heavy makeup (2), or unnatural skin tones (2), bearing little visual resemblance to the participants themselves. One participant explained their choice: *"She is already seeing me, so I want her to see something fun... something I actually designed."* (P16). Two participants chose to stick with one of the default faces (see Figure 4). For one participant, this decision stemmed from frustration with the available customization options: *"I wasn't satisfied with the options, so I just took this one."* (P6). They also viewed *MorphHat* primarily as a functional tool, placing little emphasis on its visual design: *"I didn't really pay that much details to the design overall, more to the functional aspect."* (P5) We further explored the unconventional and creative customization ideas participants used for *MorphHat*. One participant reflected on how their design choices might influence the perception of their collaborator, stating: *"I thought a blue face would be fun for her and make her trust my robot more."* (P16). In contrast, another participant interacted with a human partner whose avatar featured an unnatural skin tone. When asked why they chose not to pursue similarly bold customization options, the participant explained: *"Because I didn't feel connected with those [customization options]."* (P12).

MorphHat's effects on trust between collaborators. There were only small differences in the quantitative trust ratings between collaborators (see Figure 7). Since lower scores indicate more positive evaluations, participants rated their human partner slightly more favorably with *MorphHat* in competence ($M = 1.68$, $SD = 0.67$), trustworthiness ($M = 1.58$, $SD = 0.63$), and goodwill ($M = 1.57$, $SD = 0.66$), compared to the neutral face (competence: $M = 1.75$, $SD = 0.62$; trustworthiness: $M = 1.61$, $SD = 0.61$; goodwill: $M = 1.68$, $SD = 0.67$). Cohen's d values indicated very small effects (≤ 0.2). Furthermore, the linear mixed-effects models revealed no statistically significant differences for the *MorphHat* condition in competence ($p = .490$), trustworthiness ($p = .783$), or goodwill ($p = .405$). Bootstrap confirmed these findings: for competence, the estimated effect was -0.07 with a 95% CI of [-0.22, 0.07]; for trustworthiness, -0.03 with a 95% CI of [-0.20, 0.11]; and for goodwill, -0.11 with a 95% CI of [-0.21, 0.01], indicating that all intervals include zero. Nevertheless, the qualitative results provided more insights into how *MorphHat* influenced the interaction between the collaborators. For example, P1 stated: *"I had a more personal connection, more humor, and more feelings in the conversation."* corresponding to the dimension of warmth on the trust scale. *MorphHat* was perceived as more emotionally engaging: *"The neutral face was cold and unemotional, while the customized face felt like a normal talk between two people."* (P17). Trustworthiness, which encompasses reliability and transparency, was subtly reflected in how participants related to *MorphHat*. It was described as adding more "intimacy to the interaction" (P16). Through *MorphHat*, participants felt they gained a clearer impression of their partner's personality. This was expressed in statements such as: *"The avatar he chose is really different from me... it's funny... a good way to know your partner better."* (P7) and *"I have like a preview of her and her personality."* (P16). These responses suggest that *MorphHat* contributed to a stronger personal connection.

MorphHat's effects on rapport between collaborators. Participants appeared to customize *MorphHat* with the intention of building a connection towards their collaborator. This intention corresponded to small descriptive differences in rapport ratings (see Figure 7). A difference between conditions emerged in the interaction-based rapport ratings, where participants assessed the quality of the interaction itself. Here, *MorphHat* achieved a mean of *6.56* (SD = 0.74), compared to *6.42* (SD = 0.63) in the neutral face condition, indicating a small trend with Cohen's d = 0.268, which represents a small effect. In the self-rated rapport category, scores showed a slightly stronger trend: the neutral face condition yielded a mean of *6.04* (SD = 0.83), while the *MorphHat* condition resulted in a mean of *6.28* (SD = 0.81). Cohen's d for this comparison was 0.325, suggesting a small-to-moderate effect. However, further analyses revealed no statistically significant differences. Linear mixed-effects models showed no significant effect for the *MorphHat* condition in self-rated rapport ($p = .178$) or interaction-based rapport ($p = .271$). Bootstrap confirmed these findings: for self-rated rapport, the estimated effect was 0.24 with a 95% CI of [-0.05, 0.52], and for interaction-based rapport, the estimated effect was 0.14 with a 95% CI of [-0.11, 0.36]. While the descriptive trends in the quantitative data were minimal, the interview responses provided stronger evidence of rapport toward participants' human conversation partners when using *MorphHat*. Participants frequently described feeling a stronger emotional connection, which is a key indicator of rapport. For example, P7 shared, *"With the customized face, I felt more connected with my partner. In the neutral face condition, I just thought about finishing the task."* Others emphasized that the human-like face enhanced their sense of connection: *"The human face made*

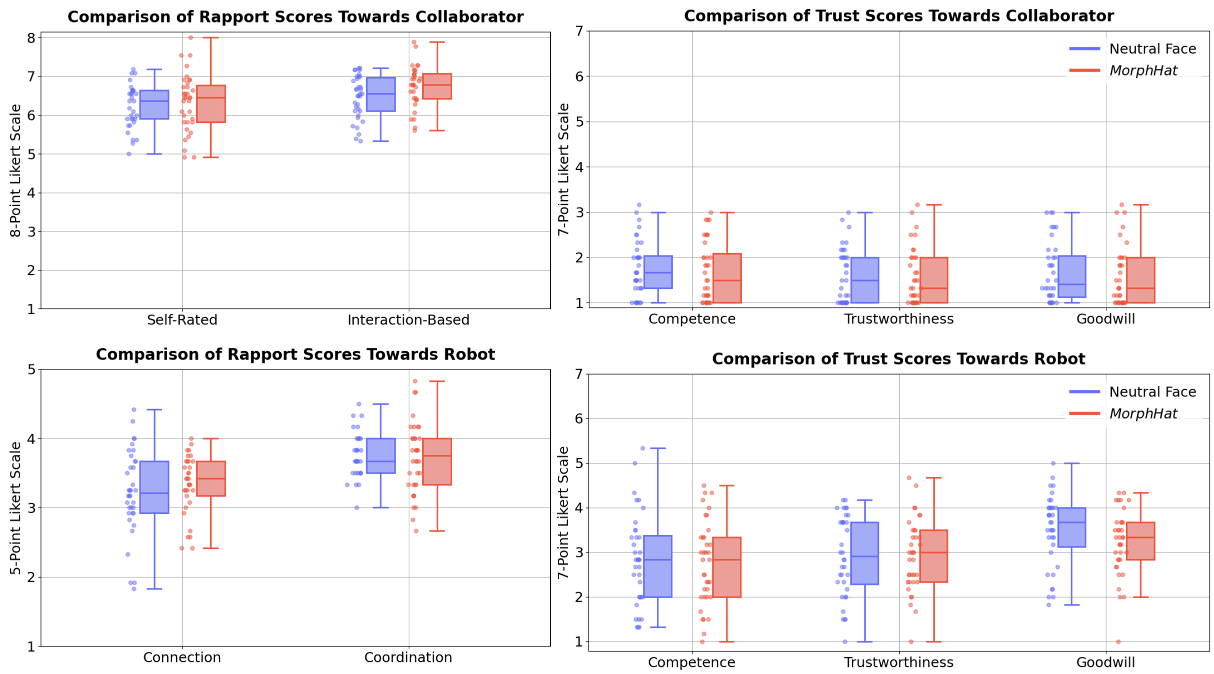

Figure 7: Boxplots of rapport and trust scores towards the robot and collaborator. Top row: rapport and trust scores towards the human partner. Bottom row: rapport and trust scores towards the robot partner. *Note: Lower scores on the trust scale represent higher levels of trust.*

us feel closer" (P18), *"The visible facial movements made me feel more connected"* (P4), and *"I felt more connected to my conversation partner because of the faces"* (P1). These statements suggest that *MorphHat* helped foster interpersonal relationships. P16 elaborated, *"Her character gave more personality of herself, it was like a personality card. (...) It seems very approachable, cute, and nice, like a sweet personality."* Rapport was also characterized by mutual understanding and shared emotional resonance, as P15 explained, *"The connection between the participants was stronger with the customized face. ... I really have more connectors."* Participants described feeling more relaxed and comfortable, with P4 adding, *"It was not so stiff because the avatar was very funny (...) it made the conversation more loosened up."* Communication was perceived as more natural and human-like: participants remarked that it felt more like talking to *"an actual person"* (P14). P18 added that *MorphHat* *"makes the conversation more real."*. However, *MorphHat* also introduced a potential distraction. P7 remarked that the neutral face was *"more boring,"* which helped them concentrate more on their human conversation partner. P12 explained, *"With the neutral face, I was more focused on my conversation partner, and the robot was more like a tool that I used. But with the customized face, I looked the robot in the eye when it was speaking to me."* Other participants expressed similar views. For instance, P16 noted: *"I looked more at the customized face rather than the neutral one. For the customized face, I was actually analyzing its face, its color, and its features (more)."* (P16). Likewise, P17 commented on the neutral face: *"I didn't really look at the robot at all and tried to focus more on my partner."*

MorphHat's effects on trust in the robot. *MorphHat* received slightly improved trust ratings in the aggregated data (see Figure 7). In the *Competence* category, it showed marginally more favorable ratings (*M = 2.75*, *SD = 0.94*) compared to the neutral face condition (*M = 2.83*, *SD = 1.03*), reflecting higher perceived trust given that lower scores indicate greater trust. Similarly, in *Goodwill*, ratings were lower for *MorphHat* (*M = 3.22*, *SD = 0.74*) than for the neutral face (*M = 3.50*, *SD = 0.81*), again indicating a slightly more positive trend. This difference is reflected in Cohen's $d = -0.363$, suggesting a small-to-moderate effect. In contrast, the dimension of *Trustworthiness* showed no difference between conditions (*MorphHat*: *M = 2.93*, *SD = 0.81*; neutral face: *M = 2.92*, *SD = 0.91*), indicating similar levels of trust in both systems, with Cohen's d close to zero. However, none of these differences reached statistical significance in the linear mixed-effects model (*Competence*: $p = .731$; *Trustworthiness*: $p = .968$; *Goodwill*: $p = .128$). Bootstrap likewise revealed no effects, with only the *Trust Competence* dimension calculated for the robot condition (estimate = -0.08, 95% CI [-0.33, 0.11]), as other dimensions could not be computed due to model singularity. While the quantitative results revealed no substantial effects, qualitative data suggested that participants perceived trust in the robot's translation capabilities as important. Participants expressed uncertainty and a sense of being *"out of control"* (P17) when relying on the neutral face output. One participant shared: *"I looked at her face to gauge her reaction… maybe the robot said complete nonsense."* (P8). Addressing the perceived intelligence of *MorphHat*, P9 noted: *"The customized face was better at putting more context in* *the translation than the neutral face.".* Others attributed fewer errors to it: *"With the neutral face, I thought the robot made a translation mistake. With the customized face, I blamed the speaker more."* (P12). Some also felt better understood: *"I kind of feel like the customized face understands me better, also emotion-wise."* (P7). This suggests that *MorphHat* was seen as more knowledgeable and context-aware. Participants described it as more human-like, which increased their trust. As P17 put it: *"It's just that I trust a human person more than a robot."* Participants also seemed to have trusted *MorphHat* more because it helped them form impressions of their partner: *"I trust the customized face more because it gave me a better first impression of the person I was going to talk to."* (P16). *MorphHat* was often described as *"not just a tool"* (P12), but rather as *"a personality card"* (P16) or even *"an alter ego"* (P12), which enhanced trust. Regarding the neutral face, it was perceived as more distant and impersonal: *"Just doing its job."* (P9), *"A third party that doesn't belong to either of us."* (P12), and *"The neutral face looks more like a robot, it feels so artificial. So I don't trust the robot 100 percent."* (P10). These impressions were linked to reduced emotional connection and lower trust. In contrast, two participants reported trusting robots and machines more than humans, which extended to the neutral face. P5 explained: *"It's a robot and it cannot make any errors. (...) I trust a robot more than a human."* Similarly, P9 remarked: *"I can see the machine behind it more and I trust the machine to do its job right."*

MorphHat's effects on rapport towards the robot. To better understand how participants felt toward the Robot, we examined the rapport they experienced during the interaction. Descriptively, participants reported slightly higher connection with *MorphHat* (see Figure 7). In the *Connection* category, *MorphHat* received slightly higher ratings ($M = 3.29$, $SD = 0.61$) compared to the neutral face condition ($M = 3.22$, $SD = 0.62$). Similarly, the *Coordination* dimension showed a small advantage for *MorphHat* ($M = 3.72$, $SD = 0.52$) over the neutral face ($M = 3.62$, $SD = 0.66$). Cohen's d values again indicated very small effects (≤ 0.2). Statistical analysis using linear mixed-effects models revealed no significant differences between conditions (*Connection*: $p = .631$; *Coordination*: $p = .486$). Bootstrap analysis could not compute *Rapport Connection* and *Rapport Coordination* due to singularity. Focusing on the interview data, participants in the category of *Connection* described feelings of warmth and empathy. For example, participants perceived *MorphHat* as more emotionally expressive, as illustrated by statements such as: *"The eyes of the customized face transported a lot of the emotions."* (P1) and *"The face was more familiar... like talking to someone that is close to me."* (P7). P6 also stated that he *"felt more attached"* to the things he creates. Emotional connection was further expressed through feelings of being understood and seen, as participants noted: *"I felt more seen."* (P12) and *"It understands me more."* (P13). Participants expressed excitement about interacting with the robot, which further contributed to the sense of connection: *"Something is special about the customized face and makes me want to talk to it."* (P7). Turning to *Coordination*, the morphing behavior of *MorphHat* supported turn-taking. As P17 explained: *"I was confused about why, in the neutral face condition, the face didn't change. So I was wondering: when should I speak? It didn't feel like there was a natural switching."* However, not all experiences with face switching were positive. P4 described initial confusion about the frequent changes: *"In the beginning of the first conversation, it swapped faces the whole time. And this was a little bit weird, because you talk to someone who switches faces."* Additionally, rapport was associated with experiencing a fluid conversation. As P10 summarized: *"I feel a smoother interaction with the customized face."* Participants also reported feeling more comfortable: *"My own robot makes it more familiar and I feel more relaxed."* (P13). It is important to note that conditions and tasks were counterbalanced, but participants observed that the second interaction felt *"less awkward"* (P4). This pattern may reflect increased familiarity with the system over time.

MorphHat's effects on social presence. Participants rated *MorphHat* as more positive than the neutral face: *more present* (23 vs. 4), *attention-capturing* (30 vs. 2), *mood-influencing* (16 vs. 3), *emotionally responsive* (15 vs. 1), and *expressive* (18 vs. 3). Participants described *MorphHat* as *"more alive"* (P2), *"more present"* (P10), and *"more like a real person"* (P9). Subtle facial movements, like eyebrow movements, were cited as contributing to this effect: *"The eyebrows moved, the corners of the mouth shifted...it looked just more like a live face"* (P14). In contrast, the neutral face was seen as static and less engaging: *"When I look at a face with real skin color, it's more like I'm talking to a human interface"* (P9), even though they were both implemented in the same way. Participants reported that their attention sometimes shifted toward *MorphHat*, as it appeared more engaging. For example, P12 noted: *"In the neutral face condition, I was more focused on my conversation partner... But in the customized face condition, I looked the robot in the eye when it was speaking to me."* However, this shift occasionally reduced focus on the human partner. P17 explained: *"I ignored my partner, my human partner a bit more, because I had more of an interaction with the (customized) robot. And with the neutral face, I focused more on my actual partner."* Similarly, P15 remarked: *"I think the neutral face was more eye-catching that I more looked at it."*

5 Discussion

5.1 *MorphHat* for Multilingual Collaboration

Customization and Representation. Most participants designed *MorphHat* to resemble themselves, often adding unique facial details. Participants' descriptions indicate that customization played an important role in how they related to the robot, though our study cannot make claims about trust mechanisms. Some participants described the robot as more socially present when it visually represented them or their partner. While some participants associated their customized face with emotional resonance, these interpretations did not show measurable effects. Interestingly, some participants preferred avatars diverging from their own appearance, revealing tension between visual similarity and embodied identification, as noted by Javier et al. [31]. Our observations align with Luria et al. [40], who found that co-embodiment can create seamless collaborative experiences, though it may introduce challenges in high-attention tasks. This mirrors findings from avatar-based systems, where customization fosters identification and engagement [9]. While avatar customization is common in virtual environments, face customization in physical robots is less explored due to hardware constraints. Our qualitative findings suggest that insights from avatar customization may be applicable to physical robots, although further research is needed. Future robot interpreters could leverage customization to enhance self-representation and emotional bonds, particularly in multilingual collaboration, where shared identity may help bridge language gaps.

Designing Translation Systems for Trust and Rapport. Our qualitative results suggest that some participants perceived the customized faces as more expressive compared to the neutral face condition, even though trust and rapport scores did not differ significantly. These impressions are consistent with prior work showing that customization influences user perception, including feelings of inclusion [9]. While homophily theory predicts that similarity fosters trust and social presence [55], we observed that similarity, enabled through customization and the ability to create one's own representation, also promoted involvement. It is important to emphasize that these experiences were not reflected in the quantitative data, which showed no measurable differences in trust or rapport, likely due to the modest sample size (N=36). Campagna and Rehm [10] state that trust is a dynamic cognitive construct that evolves over time. This may explain why deeper interpersonal trust did not emerge in our study in quantitative data, given the short duration of interaction (15 minutes per condition) and the novelty of the system. Trust, whether between humans or robotic systems, typically develops through repeated interactions [70], which were limited in this experimental setting. Longer-term studies may help determine whether robot interpreters can be seen as relational partners rather than tools. Additionally, *MorphHat* may not have fully leveraged its potential to enhance rapport and trust-related dimensions due to limited self-resembling cues. Beyond facial customization, additional customization, such as voice adaptation and cultural cues, could strengthen psychological ownership, as culture-adaptive communication strongly influences trust [67].

Creating Social Environments. Our qualitative data suggested that *MorphHat* contributed to an enhanced sense of social presence. Participants rated it as more present, expressive, and emotionally responsive than the neutral face. This effect appears to stem from two mechanisms: *co-embodiment*, where *MorphHat* visually represented both participants, and *shared representation*, which reinforced its role as an interpreter. Systems, like Socibot [22], visually presented static characters before, but rarely match the current conversational partner or give the user the chance to influence the character [22]. Moreover, face morphing approaches have been limited to one person rather than dynamically representing current speakers in conversation [71]. Thus, *MorphHat* extends these approaches by enabling real-time representation during ongoing interaction. Furthermore, research on physical embodiment indicates that embodied agents significantly increase perceived social presence and positively influence evaluations of the interaction [34]. These accounts suggest that embodiment and visual cues may influence how people experience a humanoid robot interpreter. Future studies should compare shared embodiment (one morphing robot) with individualized embodiment (two robots representing each participant) to examine which approach better supports trust and rapport. Such comparisons are important given that some participants found morphing helpful for turn-taking, while others experienced it as distracting. Rapid face changes could evoke uncanny reactions, especially in fast-paced dialogue where users have limited time to process shifting facial identities. This could introduce them to cognitive load, potentially competing with the actual collaboration task. Beyond facial morphing, future work could explore how additional cues such as culturally adapted expressions shape inclusion, as gestures convey cultural meaning and influence communication in interpretation contexts [13, 54].

5.2 Lost in Translation

While large language models offer promising contextual translation capabilities, they can introduce unintended content [48, 58]. To ensure consistency, we used neural machine translation with word-by-word translation, which sometimes resulted in rigid or emotionally flat output. This provides an opportunity for future work, which could explore hybrid approaches that balance contextual knowledge and fluency in translation. Furthermore, we did not have the resources to validate all translation transcripts due to language diversity. But we took random samples of every language and gave them to native speakers to validate translation accuracy. While their feedback was generally positive, occasional mistranslations may still have occurred and could have introduced noise into the trust and rapport measures. All participants (regardless of the language) were able to understand the translations and did not report any major problems. To this day, translation accuracy remains a challenge, with occasional errors and technical issues disrupting the flow and causing confusion. Some participants were irritated by the system's premature speech cut-offs due to the automatic detection of the sentence's end, suggesting the need for a more flexible input-closing mechanism. These interruptions and the need to manually activate voice input may have disrupted conversational flow and therefore rapport formation. Together with the short interaction duration, this may have limited quantitative differences across conditions. Additionally, language levels were self-reported, which may have introduced variability in participants' reliance on the system versus their own language skills. Moreover, we interviewed one participant per dyad to encourage more open reflections that might not surface in joint interviews. As a result, the qualitative data reflect the perspective of one participant per dyad rather than both collaborators. Some participants reported attributing translation errors differently across conditions, blaming themselves rather than the system when using *MorphHat*, raising the question of how visual resemblance might shift responsibility for automated translation. This aligns with prior work showing that users judge human-mediated translation as more reliable and are more likely to attribute errors to the system when they know it is automated [19]. If psychological ownership increases identification with the robot, users may also become less critical of its output, shifting accountability. Future research should examine how representation influences users' interpretations of system reliability.

5.3 Virtual and Physical Embodiments

While *MorphHat* was generally preferred by participants, we fully acknowledge that not everyone has access to such a technology yet. This raises the question, how fully virtual embodiments integrated into communication platforms may benefit from our concept, as they can offer greater scalability compared to physical robots. These digital agents can still convey presence and connection, particularly when designed with customizable features [31, 66]. Future research should explore how virtual avatars compare to physical language interpreters in multilingual collaboration. Although virtual systems may currently lack the tactile presence, they present a promising direction for accessible and cost-effective mediated communication. We decided against comparing the *MorphHat* system to a virtual version because many studies have explored the benefits of a physical embodiment [56, 65]. Nevertheless, the use of physical robots should be critically evaluated. Since some participants found the morphing unfamiliar, highlighting a potential downside that future studies should examine more directly. These questions emphasize the need for nuanced comparisons and longitudinal studies to better understand the psychological and practical implications of both virtual and physical embodiments. Another key limitation is the absence of a true baseline such as a human interpreter or standard voice-to-voice translation system. We chose a neutral face as a within-concept baseline to isolate the effect of facial morphing rather than embodiment itself. An additional third condition (e.g., a mobile translation app) would have strengthened the generalizability of our findings. However, this was not feasible given the logistical difficulty of recruiting 36 participants (18 dyads) with diverse language backgrounds for an on-site study. Future work should compare embodied and non-embodied translators to understand trade-offs, for example, in cognitive load.

5.4 Adapting to Cultural and Emotional Cues

Cross-cultural differences can significantly shape the dynamics of multilingual collaboration, highlighting the need for culturally adaptive systems for interpreting [13]. In this work, we focused primarily on communication and face customization. However, facial expressions can vary across cultures, which means they may be interpreted differently depending on cultural background. This highlights the need for integrating culturally adaptive expressions as a direction for future work. Future research should therefore explore in greater depth how embodiment and translation technologies can mediate cultural norms and foster more inclusive interactions. Particularly given that some participants interpreted emotional expressions differently, suggesting that uniform facial cues may not support all cultural backgrounds equally. A compelling vision for future systems is exemplified by the character *C-3PO* from *Star Wars*, a humanoid robot that not only facilitates multilingual communication but also embodies cultural knowledge, functioning as an intelligent interpreter across diverse societies. Inspired by this concept, future systems should incorporate cultural cues, nonverbal gestures, and contextual awareness to enrich interpersonal communication.

5.5 Beyond Dyadic Collaboration

Our study focuses on collaboration between dyads due to pragmatic reasons, but the *MorphHat* concept is not limited to that. It can support multiple collaborators by allowing each individual to create a customized stand-in. *MorphHat* can directly address the target person by facing them, co-embodying individuals' trustworthy companions. However, whether people can recognize their customized translator in a team collaboration setting and how *MorphHat* can address multiple people simultaneously, remains to be explored. This also offers the potential to go beyond the capabilities of a human interpreter who can typically handle up to triadic conversations [16, 61], offering a supporting tool for more complex multilingual settings. This may also extend beyond the context of collaboration and enable socially rich multilingual communication in everyday situations, such as in social services. Future research should examine how facial morphing scales to more complex conversational settings, such as group interactions, and identify when face switching may become confusing.

6 Conclusion

In this paper we proposed *MorphHat*, a humanoid robot interpreter designed to support multilingual collaboration through facial customization and dynamic face morphing for representing speakers in multilingual collaboration. Acting as a personal representative for participants, *MorphHat* aimed to address trust and representation, which are central to inclusive collaboration. In our within-subjects study with 36 participants, qualitative descriptions suggested that customization shaped feelings of representation and may have indirectly supported trust, despite no significant quantitative differences. Participants' perceptions of *MorphHat* shifted from a neutral tool to a more supportive partner, indicating effects on rapport. Overall, *Morphhat* makes the first steps towards embodied live translation that helps people navigate multilingual settings more comfortably. Future work should compare embodied and non-embodied translation tools and examine how trust and representation develop over longer term use.

Acknowledgments

This work was support by Heidelberger Services AG (hsag).

References

[1] National Aeronautics and Space Administration. 1970. NASA Exercise: Survival on the Moon. https://www.csuchico.edu/anthmuseum/_assets/documents/nasa-exercise-survival-on-the-moon.pdf Accessed August 2025.

[2] S. Ashkenazi, G. Skantze, J. Stuart-Smith, and M.E. Foster. 2025. The Multilingual Student Support Robot. In *Proceedings of the 2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 953–958. doi:10.1109/RO-MAN63969.2025.11217768

[3] S. Ashkenazi, R. Srour-Zreik, G. Skantze, J. Stuart-Smith, and M.E. Foster. 2025. Participatory Design for Human-Robot Interaction with Syrian Refugees and Asylum Seekers. In *Proceedings of the 2025 International Conference on Social Robotics + AI (ICSR+AI)*. Naples, Italy. 10–12 Sep 2025.

[4] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. doi:10.18637/jss.v067.i01

[5] Frank J. Bernieri, John S. Gillis, Janet M. Davis, and Jon E. Grahe. 1996. Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of personality and social psychology* 71, 1 (07 1996), 110. https://www.proquest.com/scholarly-journals/dyad-rapport-accuracy-judgment-across-situations/docview/209802738/se-2 Copyright - Copyright American Psychological Association Jul 1996; Last updated - 2025-11-13; CODEN - JPSPB2.

[6] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence: Teleoperators and Virtual Environments* 12, 5 (2003), 456–480.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[8] Thomas Breideband, Jeffrey Bush, Chelsea Chandler, Michael Chang, Rachel Dickler, Peter Foltz, Ananya Ganesh, Rachel Lieber, William R. Penuel, Jason G. Reitman, John Weatherley, and Sidney D'Mello. 2023. The Community Builder

(CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 376–380. doi:10.1145/3584931.3607498

[9] Benjamin Buck et al. 2023. Avatar Customization, Personality, and the Perception of Work Group Inclusion in Immersive VR. In *CSCW '23 Companion: Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. ACM, New York, NY, USA. doi:10.1145/3584931.3606992

[10] Giulio Campagna and Matthias Rehm. 2025. A Systematic Review of Trust Assessments in Human–Robot Interaction. *J. Hum.-Robot Interact.* 14, 2, Article 30 (Jan. 2025), 35 pages. doi:10.1145/3706123

[11] Mei-Ling Chen, Naomi Yamashita, and Hao-Chuan Wang. 2018. Come Together: Facilitating Collocated Multilingual Group Discussion with a Language Support Tool. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3170427.3188595

[12] Sam Clark. 2025. *'The worst thing': Meloni-Trump interpreter stumbles at White House amid Ukraine war.* https://www.politico.eu/article/worst-thing-meloni-trump-interpreter-stumbles-at-white-house-ukraine-war/ Accessed: August 21, 2025.

[13] Kyle Conway. 2013. Cultural translation. In *Handbook of Translation Studies: Volume 3*. John Benjamins Publishing Company, 21–25.

[14] Yichao Dai, Zhirui Zhang, Yichao Du, Shujie Liu, Lemao Liu, and Tong Xu. 2023. Datastore Distillation for Nearest Neighbor Machine Translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 32 (2023), 807–817.

[15] Sarah Dawkins, Amy Wei Tian, Alexander Newman, and Angela Martin. 2017. Psychological ownership: A review and research agenda. *Journal of organizational behavior* 38, 2 (2017), 163–183.

[16] Bruce T. Downing. 2021. Community/Liaison Interpreting in Healthcare Settings. In *The Routledge Handbook of Translation and Health*, Şebnem Susam-Saraeva and Eva Spišiaková (Eds.). Routledge, Chapter 14. doi:10.4324/9781003167983-14

[17] Kurt Eberle, Johanna Geiß, Mireia Ginestí-Rosell, Bogdan Babych, Anthony Hartley, Reinhard Rapp, Serge Sharoff, and Martin Thomas. 2012. Design of a hybrid high quality machine translation system *(EACL 2012)*. Association for Computational Linguistics, USA, 101–112.

[18] Martin Feick, Terrance Mok, Anthony Tang, Lora Oehlberg, and Ehud Sharlin. 2018. Perspective on and Re-orientation of Physical Proxies in Object-Focused Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173855

[19] Ge Gao, Bin Xu, Dan Cosley, and Susan R. Fussell. 2014. How beliefs about the presence of machine translation impact multilingual collaborations. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1549–1560. doi:10.1145/2531602.2531702

[20] Ge Gao, Bin Xu, David C. Hau, Zheng Yao, Dan Cosley, and Susan R. Fussell. 2015. Two is Better Than One: Improving Multilingual Collaboration by Giving Two Machine Translation Outputs. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 852–863. doi:10.1145/2675133.2675197

[21] Ge Gao, Naomi Yamashita, Ari M.J. Hautasaari, and Susan R. Fussell. 2015. Improving Multilingual Collaboration by Displaying How Non-native Speakers Use Automated Transcripts and Bilingual Dictionaries. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3463–3472. doi:10.1145/2702123.2702498

[22] Aimi Shazwani Ghazali, Jaap Ham, Emilia Barakova, and Panos Markopoulos. 2020. Investigating the effect of social cues on social agency judgement. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (Daegu, Republic of Korea) *(HRI '19)*. IEEE Press, 586–587.

[23] Amit Ghimire, Anova Hou, Ig-Jae Kim, and Dongwook Yoon. 2025. AvatARoid: A Motion-Mapped AR Overlay to Bridge the Embodiment Gap Between Robots and Teleoperators in Robot-Mediated Telepresence. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 906, 26 pages. doi:10.1145/3706598.3713812

[24] Cristina B. Gibson and Jennifer L. Gibbs. 2006. Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative Science Quarterly* 51 (2006), 451–495.

[25] Linda Haffner, Karin Ruschke, Shiva Bidar-Sielaff, Maria-Paz Avery, Bruce Downing, Carola Green, and Cynthia Roat. 2003. *Guide to Interpreter Positioning in Health Care Settings.* Working Paper. National Council on Interpreting in Health Care. https://www.ncihc.org/assets/documents/workingpapers/NCIHC%20Working%20Paper%20-%20Guide%20to%20Interpreter%20Positioning%20in%20Health%20Care%20Settings.pdf Accessed August 2025.

[26] Megan K. Halpern, Ingrid Erickson, Laura Forlano, and Geri K. Gay. 2013. Designing collaboration: comparing cases exploring cultural probes as boundary-negotiating objects. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1093–1102. doi:10.1145/2441776.2441900

[27] Chad Harms and Frank Biocca. 2004. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Unpublished manuscript* (2004).

[28] Xiaoxuan Hei, Chuang Yu, Heng Zhang, and Adriana Tapus. 2024. A bilingual social robot with sign language and natural language. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, New York, NY, USA, 526–529.

[29] Elaine K. Horwitz, Michael B. Horwitz, and Jo Ann Cope. 1986. Foreign language classroom anxiety. *The Modern Language Journal* 70, 2 (1986), 125–132. doi:10.2307/327317

[30] Sheng Hu and Cuiying Zou. 2024. An Investigation of College Students' Use of Online Machine Translation in English Learning. In *Proceedings of the 7th International Conference on Educational Technology Management (ICETM 2024)*. ACM, Changsha, China. doi:10.1145/3711403.3711449

[31] Alex Javier et al. 2024. The Alt-Self: Investigating the Inclusivity of Self-Avatar Representations in Social VR. In *CHI PLAY Companion '24: Extended Abstracts of the 2024 Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, NY, USA. doi:10.1145/3665463.3678783

[32] Monika Johansson, Gabriel Skantze, and Joakim Gustafson. 2013. Head Pose Patterns in Multiparty Human-Robot Team-Building Interactions. In *Social Robotics*, G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, and U. Leonards (Eds.). Springer International Publishing, 351–360. doi:10.1007/978-3-319-02675-6_35

[33] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, 13 (2017), 1–26. doi:10.18637/jss.v082.i13

[34] Kwan Min Lee, Younbo Jung, Jae Kim, and Sang Rok Kim. 2006. Are physically embodied social agents better than disembodied social agents?: the effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies* 64, 10 (2006), 962–973.

[35] Daniel J. Liebling, Michal Lahav, Abigail Evans, Aaron Donsbach, Jess Holbrook, Boris Smus, and Lindsey Boran. 2020. Unmet Needs and Opportunities for Mobile Translation AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376261

[36] Ting-Han Lin, Hannah Dinner, Tsz Long Leung, Bilge Mutlu, J Gregory Trafton, and Sarah Sebo. 2025. Connection-Coordination Rapport (CCR) Scale: A Dual-Factor Scale to Measure Human-Robot Rapport. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 869–879.

[37] K Casey Lion, Chuan Zhou, Paul Fishman, Kirsten Senturia, Allison Cole, Kenneth Sherr, Douglas J Opel, James Stout, Carmen E Hazim, Louise Warren, et al. 2023. A sequential, multiple assignment randomized trial comparing web-based education to mobile video interpreter access for improving provider interpreter use in primary care clinics: the mVOCAL hybrid type 3 study protocol. *Implementation Science* 18, 1 (2023), 8.

[38] Nichola Lubold and Heather Pon-Barry. 2014. Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues. In *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (Istanbul, Turkey) *(MLA '14)*. Association for Computing Machinery, New York, NY, USA, 5–12. doi:10.1145/2666633.2666635

[39] Le Luo, Dongdong Weng, Guo Songrui, Jie Hao, and Ziqi Tu. 2022. Avatar interpreter: improving classroom experiences for deaf and hard-of-hearing people based on augmented reality. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. Association for Computing Machinery (ACM), New York, NY, USA, 1–5.

[40] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 633–644. doi:10.1145/3322276.3322340

[41] Yujie Ma and Yifan Qian. 2023. Methods of Improving Japanese-Chinese Machine Translation System through Machine Learning and Human-Computer Interaction. *ACM Transactions on Asian and Low-Resource Language Information Processing* (2023).

[42] Amama Mahmood and Chien-Ming Huang. 2024. Gender Biases in Error Mitigation by Voice Assistants. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 60 (April 2024), 27 pages. doi:10.1145/3637337

[43] James C. McCroskey and Thomas J. Young. 1981. Ethos and credibility: The construct and its measurement after three decades. *Central States Speech Journal* 32, 1 (1981), 24–34. doi:10.1080/10510978109368075

[44] Kana Misawa, Yoshio Ishiguro, and Jun Rekimoto. 2012. Livemask: A telepresence surrogate system with a face-shaped screen for supporting nonverbal communication. In *Proceedings of the international working conference on advanced visual*

interfaces. 394–397.
[45] Sandra Mueller, Martin Feick, and Alexander Maedche. 2025. WorldHat: A Humanoid Social Robot Interpreter for Multilingual Dyadic Conversations. In *Mensch und Computer 2025 – Workshopband*. Gesellschaft für Informatik e.V., Chemnitz, Germany.
[46] Prasanth Murali, Ian Steenstra, Hye Sun Yun, Ameneh Shamekhi, and Timothy Bickmore. 2023. Improving Multiparty Interactions with a Robot Using Large Language Models. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 175, 8 pages. doi:10.1145/3544549.3585602
[47] Naoto Nakazato, Shigeo Yoshida, Sho Sakurai, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2014. Smart Face: enhancing creativity during video conferences using real-time facial deformation. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 75–83.
[48] Jianhui Pang, Fanghua Ye, Derek Fai Wong, Dian Yu, Shuming Shi, Zhaopeng Tu, and Longyue Wang. 2025. Salute the classic: Revisiting challenges of machine translation in the age of large language models. *Transactions of the Association for Computational Linguistics* 13 (2025), 73–95.
[49] Franz Pöchhacker. 2006. Research and methodology in healthcare interpreting. *Linguistica Antverpiensia, New Series–Themes in Translation Studies* 5 (2006).
[50] Nike K Pokorn. 2017. "There is always some spatial limitation" Spatial positioning and seating arrangement in healthcare interpreting. *Translation and interpreting studies* 12, 3 (2017), 383–404.
[51] PACEsetters Program. 2016. Teambuilder: LOST AT SEA. https://www.airman.af.mil/Portals/17/002%20All%20Products/003%20PACEsetters/Teambuilder_LOST_AT_SEA.pdf Accessed August 2025.
[52] Lin Qiu and Izak Benbasat. 2010. A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human-Computer Studies* 68, 10 (2010), 669–688. doi:10.1016/j.ijhcs.2010.05.005
[53] Yuqing Ren and Jeffrey Clement. 2024. Augmenting Human Teams with Robots in Knowledge Work Settings: Insights from the Literature. *ACM Transactions on Human-Robot Interaction* 13, 2 (2024), Article 20, 34 pages. doi:10.1145/3649884
[54] Jessica A. Sasser, David S. McConnell, and Janessa A. Smither. 2024. Investigation of Relationships Between Embodiment Perceptions and Perceived Social Presence in Human–Robot Interactions. *International Journal of Social Robotics* 16, 8 (2024), 1735–1750. doi:10.1007/s12369-024-01138-w
[55] Reza Salek Shahrezaie, B. A. Anima, and David Feil-Seifer. 2021. Birds of a Feather Flock Together: A Study of Status Homophily in HRI. In *Social Robotics*, H. Li, S. S. Ge, Y. Wu, A. Wykowska, H. He, X. Liu, D. Li, and J. Perez-Osorio (Eds.). Springer International Publishing, 281–291. doi:10.1007/978-3-030-90525-5_24
[56] Ameneh Shamekhi, Q. Vera Liao, Dakuo Wang, Rachel K. E. Bellamy, and Thomas Erickson. 2018. Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173965
[57] Keir Simmons, Carol E. Lee, Dan De Luce, and Courtney Kube. 2025. *Trump envoy relied on Kremlin interpreter in meetings with Putin to end war in Ukraine*. https://www.nbcnews.com/world/russia/russia-ukraine-war-trump-envoy-witkoff-interpreter-kremlin-rcna205878 Accessed: August 21, 2025.
[58] Jungha Son and Boyoung Kim. 2023. Translation performance from the user's perspective of large language models and neural machine translation systems. *Information* 14, 10 (2023), 574.
[59] Iris Straub. 2016. 'It looks like a human!' The interrelation of social presence, interaction and agency ascription: A case study about the effects of an android robot on social agency ascription. *AI & SOCIETY* 31, 4 (2016), 553–571. doi:10.1007/s00146-015-0632-5
[60] Linda Tickle-Degnen and Robert Rosenthal. 1990. The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry* 1, 4 (1990), 285–293. doi:10.1207/s15327965pli0104_1
[61] Rebecca Tipton. 2019. Community Interpreting. In *Routledge Encyclopedia of Translation Studies* (3rd ed.). Routledge, Chapter 18. doi:10.4324/9781315678627-18
[62] Enrica Troiano, Roman Klinger, and Sebastian Padó. 2020. Lost in back-translation: Emotion preservation in neural machine translation. In *Proceedings of the 28th International Conference on Computational Linguistics*. 4340–4354.
[63] Jean Ventre-Dominey, Gilles Gibert, Marie Bosse-Platière, Alessandro Farnè, Peter Ford Dominey, and Flavia Pavani. 2019. Embodiment into a robot increases its acceptability. *Scientific Reports* 9, 1 (2019), 10083. doi:10.1038/s41598-019-46528-7
[64] Jelena Vranjes and Geert Brône. 2021. Interpreters as laminated speakers: Gaze and gesture as interpersonal deixis in consecutive dialogue interpreting. *Journal of Pragmatics* 181 (2021), 83–99.
[65] Joshua Wainer, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. 2006. The role of physical embodiment in human-robot interaction. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 117–122.
[66] Rebecca Wald, Evelien Heijselaar, and Tibor Bosse. 2021. Make your own: The Potential of Chatbot Customization for the Development of User Trust. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (Utrecht, Netherlands) *(UMAP '21)*. Association for Computing Machinery, New York, NY, USA, 382–387. doi:10.1145/3450614.3463600
[67] Liangyan Wang, Paul-Lou Chia-Yi Rau, Vanessa Evers, Brandon K. Robinson, and Pamela Hinds. 2010. When in Rome: The role of culture & context in adherence to robot recommendations. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 359–366. doi:10.1109/HRI.2010.5453165
[68] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (2016), 38–44. doi:10.1145/2907069
[69] Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka, and Toru Ishida. 2009. Difficulties in establishing common ground in multiparty groups using machine translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 679–688. doi:10.1145/1518701.1518807
[70] Rongbin Yang and Santoso Wibowo. 2022. User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets* 32, 4 (2022), 2053–2077.
[71] Hirokazu Yoshida and Fumihide Tanaka. 2018. Maintaining the Presence of Remote Speakers on Telepresence Robots by Visual Morphing to Reduce Loneliness. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 404–409. doi:10.1109/ROMAN.2018.8525537
[72] Nungduk Yun and Seiji Yamada. 2024. Investigation of factors that influence human presence and robot anthropomorphism in telepresence robot. *IEEE Access* 13 (2024), 4996–5008.
[73] Zihan Zhang, Black Sun, and Pengcheng An. 2025. Breaking Barriers or Building Dependency? Exploring Team-LLM Collaboration in AI-infused Classroom Debate. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 137, 19 pages. doi:10.1145/3706598.3713853
[74] Naim Zierau, Edona Elshan, Camillo Visini, and Andreas Janson. 2020. A Review of the Empirical Literature on Conversational Agents and Future Research Directions. In *Proceedings of the International Conference on Information Systems (ICIS)*. Association for Information Systems, Virtual Conference, n/a. https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/5/

A Data Analysis Numbers

Table 1: Cronbach's Alpha for Rapport and Trust Questionnaires

Questionnaire	Dimension	Cronbach's α
Rapport Towards Human	Self-Rated	.769
	Interaction-Based	.809
	Overall	.869
Trust Towards Human	Competence	.918
	Trustworthiness	.822
	Goodwill	.878
	Overall	.943
Rapport Towards Robot	Connection	.887
	Coordination	.778
	Overall	.887
Trust Towards Robot	Competence	.881
	Trustworthiness	.773
	Goodwill	.730
	Overall	.889

Note: Higher rapport scores indicate stronger perceived rapport. Lower trust scores indicate higher perceived trust due to counterbalanced scales.

Table 2: Descriptive Statistics for Neutral Face and *MorphHat* with Cohen's d

	Median		Mean		Std. Dev.		Cohen's d
	Neutral	*MorphHat*	Neutral	*MorphHat*	Neutral	*MorphHat*	
Rapport Self-Rated	6.135	6.450	6.036	6.276	0.825	0.805	0.325
Rapport Interaction-Based	6.530	6.780	6.420	6.564	0.628	0.738	0.268
Trust Competence Human	1.670	1.500	1.754	1.680	0.621	0.666	-0.164
Trust Trustworthiness Human	1.500	1.330	1.607	1.575	0.611	0.625	-0.066
Trust Goodwill Human	1.415	1.330	1.676	1.569	0.667	0.661	-0.200
Rapport Connection	3.210	3.375	3.223	3.292	0.615	0.608	0.114
Rapport Coordination	3.670	3.750	3.620	3.718	0.658	0.515	0.165
Trust Competence Robot	2.830	2.830	2.833	2.754	1.029	0.943	-0.082
Trust Trustworthiness Robot	2.915	3.000	2.922	2.930	0.910	0.806	0.009
Trust Goodwill Robot	3.670	3.330	3.500	3.218	0.813	0.739	-0.363

Note: Higher rapport scores indicate stronger perceived rapport. Lower trust scores indicate higher perceived trust due to counterbalanced scales.

Table 3: Linear Mixed Model Estimates with Model Fit

Questionnaires	Term	Estimate	Std. Error	t-value	p-value	R_m^2	R_c^2
Rapport Self-Rated	(Intercept)	6.036	0.149	40.591	0.000	0.090	0.594
	MorphHat	0.239	0.175	1.366	0.178		
Rapport Interaction-Based	(Intercept)	6.420	0.126	50.769	0.000	0.082	0.764
	MorphHat	0.144	0.129	1.114	0.271		
Trust Competence Human	(Intercept)	1.754	0.133	13.181	0.000	0.026	0.815
	MorphHat	-0.074	0.108	-0.690	0.493		
Trust Trustworthiness Human	(Intercept)	1.607	0.124	12.917	0.000	0.031	0.647
	MorphHat	-0.032	0.115	-0.277	0.783		
Trust Goodwill Human	(Intercept)	1.676	0.126	13.327	0.000	0.041	0.713
	MorphHat	-0.107	0.127	-0.839	0.405		
Rapport Connection	(Intercept)	3.223	0.102	31.621	0.000	0.009	0.666
	MorphHat	0.069	0.144	0.482	0.631		
Rapport Coordination	(Intercept)	3.620	0.098	36.784	0.000	0.019	0.639
	MorphHat	0.098	0.139	0.700	0.486		
Trust Competence Robot	(Intercept)	2.833	0.170	16.667	0.000	0.174	0.296
	MorphHat	-0.078	0.226	-0.346	0.731		
Trust Trustworthiness Robot	(Intercept)	2.922	0.143	20.389	0.000	0.022	0.656
	MorphHat	0.008	0.203	0.040	0.968		
Trust Goodwill Robot	(Intercept)	3.500	0.130	27.023	0.000	0.051	0.714
	MorphHat	-0.282	0.183	-1.539	0.128		

Note: Higher rapport scores indicate stronger perceived rapport. Lower trust scores indicate higher perceived trust due to counterbalanced scales.