

RESEARCH



# Conversational Fluency and Attitudes Towards Robot Pilots in Telepresence Robot-Mediated Interactions

Jean E. Fox Tree\*<sup>1</sup>, Susan C. Herring<sup>2</sup>, Allison Nguyen<sup>1</sup>, Steve Whittaker<sup>1</sup>, Rob Martin<sup>1</sup> & Leila Takayama<sup>1</sup>

\*<sup>1</sup>Psychology Department, University of California Santa Cruz, Social Sciences 2 Room 277, Santa Cruz, CA 95064, USA (E-mail: foxtree@ucsc.edu); <sup>2</sup>Department of Information and Library Science, Indiana University, 700 N. Woodlawn Ave., Bloomington, IN 47408, USA

Accepted: 2 June 2023

**Abstract.** In a controlled lab experiment, we compared how in-person and robot-mediated communicative settings affected attitudes towards communicators and discourse phenomena related to conversational negotiation. We used a mock interview within-participants experiment design where each participant (mock interviewee) experienced both types of communication with the same experimenter (mock interviewer). Despite communicating with the same person, participants found the in-person interviewer to be more likable, more capable, more intelligent, more polite, more in control, and less awkward than the same person using a telepresence robot. Behaviorally, we did not detect differences in participants' productions of discourse phenomena (*likes*, *you knows*, *ums*, *uhs*), laughter, or gaze. We also tested the role of communicative expectations on attitudes towards communications. We primed participants to expect that they would be talking to a person via telepresence, a “disabled” robot-person combination using telepresence, or a person in person (between-participants). We did not find differences arising from people's expectations of the communication.

**Keywords:** Mobile telepresence, Attitudes, Discourse markers, Fillers, Laughter, Gaze

Telepresence robot-mediated communication is human–human communication in which at least one party is telepresent via, and remotely controlling, a robot. Although telepresence robots have existed since 1998 (Paulos and Canny, 1998), it has only become feasible to deploy such robots in real-world contexts since high

bandwidth wireless networks became pervasive. Accordingly, a growing number of telepresence robots have become commercially available. Telepresence robots have been field tested with caregiving of homebound elderly people (Fiorini et al., 2020), education of children who could not be in the classroom (Newhart and Warschauer, 2016; Weibel et al., 2020), education of children who live at a distance from the instructor (Kwon et al., 2010), education of children who required special education (Fischer et al., 2019), geocaching between a person in nature and a person indoors (Heshmat et al., 2018), bringing couples in long distance relationships closer together (Yang and Neustaedter, 2018), and shopping with a person in the store and a person elsewhere (Yang et al., 2018). While not as ubiquitous as videoconferencing, telepresence robot-mediated communication has great potential because it is richer in presence, affording human operators the sense of being at the remote location (Weibel et al., 2020; Yang et al., 2017). The human operator can also enjoy the benefits of embodiment with enhanced navigational control, allowing active exploration of the remote environment.

We conducted a controlled experiment to compare how people communicated in an in-person versus a robot-mediated communicative setting and how such communication depended on their communicative expectations. The in-person versus robot-mediated manipulation was within participants: Each participant experienced both conditions. The communicative expectations varied across participants: Some participants expected to talk to a person, others expected to talk to a telepresence robot, and others expected to talk to a person using a telepresence robot who was “disabled” due to the robot’s physical limitations. We did not find differences regarding people’s expectations of the communication. We also did not find differences in communication style. But we did find differences in how people felt about their addressee based on communication modality. People felt more positively about the same person when they were in person versus when they were using the telepresence robot.

## 1 Technology-mediated interaction

There has been relatively little study of conversational processes and outcomes in robot-mediated communication (Herring, 2016); however, this topic has been well studied for video-mediated conversations (Chapanis, 1975; Chapanis et al., 1972; O’Conaill et al., 1993; Short et al., 1976; Sellen, 1995; Whittaker, 1995; Whittaker and O’Conaill, 1997). Comparisons between videoconferencing and face-to-face conversations show relatively few differences in outcomes, for example, as regards learning (Storck and Sproull, 1995), negotiation (Sellen, 1995; Short et al., 1976; Morley and Stephenson, 1970), and in object co-construction tasks (Chapanis, 1975; Reid, 1977). Nevertheless, studies of remote learning suggest that social relationships are affected by video mediation. In a hybrid learning setting where class interactions were a mix of face-to-face and video-mediated,

## Conversational Fluency and Attitudes Towards Robot Pilots...

participants were more positive about classmates they had interacted with face-to-face compared with those they had only met over video (Storck and Sproull, 1995). Although there is little difference in the content of video-mediated versus face-to-face conversations for a variety of tasks (Short et al., 1976; Rutter, 1984; Morley and Stephenson, 1970), in video mediation, participants may attend more to remote participants' verbal communication than to their non-verbal behaviors because body movements are less visible in video-mediated communication (Storck and Sproull, 1995).

Moreover, many studies show differences in conversational processes that make video conversation seem less interactive than face-to-face. For example, video conversations have fewer backchannels (Fox Tree et al., 2021; O'Conaill et al., 1993; Whittaker, 1995) and interruptions (Sellen, 1995). Furthermore, transitions between video speakers are more formal, with greater use of formal handovers (O'Conaill et al., 1993) and more pausing between turns (Hollingsworth, 2022). In a study of narrative structure, an overarching conclusion was that in-person interactions increased story-telling, but telepresence interactions increased story-acting (Fox Tree et al., 2021). Together these findings suggest that video compromises grounding processes (Clark and Brennan, 1991) that allow listeners to show incremental understanding and offer feedback. Video also makes it harder for listeners to gain the conversational floor to clarify or elaborate on what a speaker is saying.

The technical limitations of video help explain these differences. Network limitations can introduce speech lags, which disrupt the flow of conversation and can have major impacts on basic grounding processes, for example by reducing backchannels (Cohen, 1982; Krauss and Bricker, 1967; O'Conaill et al., 1993; Sellen, 1995). Furthermore, gestures and eye gaze, which are critical turn-taking cues, are harder to interpret over video (Argyle, 1990; Beattie, 1978; Monk and Gale, 2002). Gaze misalignment is a pervasive problem in video conferencing systems due to the disparity between the locations of the subject and the camera. This makes mutual eye contact difficult to achieve, as users tend to look at the image of their interlocutor on the screen rather than at the camera (Kuster et al., 2012). Finally, emotional expressions are important when building social relationships, but these, too, can be hard to read over video due to screen resolution, poor internet connections, and other problems (Bruce, 1995; Whittaker and O'Conaill, 1997). All of these studies have examined conversational behaviors in settings where cameras are fixed. In robot-mediated communication, in contrast, telepresence robot pilots are able to navigate through their environment, allowing them more control over where they direct their camera.

An earlier report on how people communicate using telepresence robots supported the idea that using telepresence robots increases psychological distance. Psychological distance is how mentally close communicators feel to other

communicators, and it is related to the concepts of immediacy and social presence (Fox Tree et al., 2021). The researchers observed changed story elements (more abstracts of the stories in person), differently manipulated objects (more object manipulation in telepresence), and differences in backchannels (more backchannels in person), which they argued were a result of increased psychological distance when using telepresence robots (Fox Tree et al., 2021).

In another study, increased psychological distance and changes in discourse patterns were observed when participants communicated with a static humanoid robot versus a robot whose head and lips moved (Tanaka et al., 2014). Participants agreed more strongly with the statement that they felt like they were in the same room with their addressee when the robot's head and lips moved like the (unseen) speaker the robot was emulating than when the robot did not move. Participants also produced more silent pauses when speaking with a robot whose head and lips moved compared to a two-dimensional avatar whose head and lips moved. The researchers proposed that the increase in pauses was caused by increased "tension" in the moving-robot condition (Tanaka et al., 2014, p. 108). While not involving movement through space, these results support the proposal that movement affects psychological distance, which in turn influences how people feel about their interactions and how they produce discourse phenomena.

In the current study, we tested how use of a telepresence robot affected participants' attitudes towards the human communicating through the robot, as well as participants' production of discourse phenomena. We predicted more positive attitudes for in-person interactions over telepresence interactions. The discourse phenomena assessed were discourse markers (such as *you know*), fillers (such as *um*), laughter, and gaze. We predicted that some phenomena would be more likely in person, but that others would not. We also assessed how participants' metaphorical representation of the person they would be interacting with affected their attitudes and discourse phenomena. These representations were primed to be: a person (not using a telepresence robot), a telepresence robot, or a person using a telepresence robot where the robot-person combination was "disabled" due to the limitations of the device (e.g., unable to open doors). We now turn to discussion of the attitudes assessed, the discourse phenomena assessed, and the participants' expectations of their interactions prompted by metaphorical primes on the door of the testing room.

## 2 Attitudes

How people think about robots may play an important role in conversational interaction. The way people talk to non-human agents is not the same as how they talk to people. For example, some people adopt an "imperious language style" in communicating with digital assistants (Bonfert et al., 2018, p. 96). Thinking in

## Conversational Fluency and Attitudes Towards Robot Pilots...

more human-like terms about a robot can lead to more human-like social expectations of, and behavior towards, the robot (Lee and Takayama, 2011; Takayama and Go, 2012). Based on field notes and interviews with workers in technology-focused companies where telepresence robots were used, Takayama and Go (2012) identified different metaphors that people used for interacting with and talking to the telepresence systems: as nonhuman-like (sub-categorized as: communication medium, robot, and object) or as human-like (sub-categorized as: person and person with disabilities). Remote users operating the robot were defined as *pilots*, and individuals physically co-located with the robot were defined as *local users*. Local users who held a human-like metaphorical model of the robot were more likely to exhibit polite social behaviors toward the robot (e.g., asking the pilot to adjust the volume of the robot audio) and show a similar level of respect for personal space toward the robot as they would toward a human. Local users with the metaphorical model of disabled human would sometimes go out of their way to help the robot (e.g., by talking extra loudly so that the pilot could hear, writing in large letters on a white board so that the pilot could read it, and slowing their pace to walk the robot through the office). In contrast, local users who held a nonhuman-like metaphorical model of the robot were more likely to breach social norms of polite behavior and personal space (e.g., pressing buttons on the robot to adjust its volume directly). In cases where the pilot and local users held differing metaphorical models for the robot, conflicts sometimes occurred (e.g., a local user turning off the robot mid-conversation as if they were hanging up a phone).

In order to approximate users' metaphorical understandings of telepresence robots as identified by Takayama and Go (2012), we explicitly primed participants to interact with a robot, a person, or a "disabled" robot-person combination – that is, a person using a telepresence robot whose mobility was limited due to the limitations of the robotic system. We predicted that the metaphorical prime would affect attitudes, with more positive attitudes when participants were primed to interact with a human being, disabled or not, rather than a machine.

The attitudes we assessed have been explored in previous work primarily with non-telepresence robots (Hoffman et al., 2020; Mirnig et al., 2017; Niemelä et al., 2017; Torrey et al., 2013; Ullman et al., 2014), although politeness has been assessed with telepresence robots (Takayama and Go, 2012). To broaden our understanding of responses to telepresence robots, we assessed: (1) likableness (Hoffman et al., 2020; Huang et al., 2017; Mirnig et al., 2017; Torrey et al., 2013), (2) awkwardness (Huang et al., 2017), (3) capableness (Hoffman et al., 2020; Niemelä et al., 2017), (4) intelligence (Mirnig et al., 2017; Ullman et al., 2014), (5) intimidating-ness (Huang et al., 2017; Niemelä et al., 2017), (6) politeness (Niemelä et al., 2017; Takayama and Go, 2012; Torrey et al., 2013), and (7) in-control-ness (Torrey et al., 2013).

### 3 Discourse phenomena

Discourse phenomena can differ across telepresence and in-person settings. For example, while backchannels (words like *mhm* and *really* spoken by an addressee listening to a floor-holder's turn) were generally similar across telepresence and in-person settings, more *yeahs* were used in person (Fox Tree et al., 2021), aligning with prior observations that people use more social chat in person in comparison to over the phone (Short et al., 1976). In the current study, we investigated additional discourse phenomena, including discourse markers (*like* and *you know*), fillers (*um* and *uh*), laughter, and gaze. Following the prior findings for *yeahs*, we anticipate more of these discourse phenomena in in-person communication compared to telepresence communication.

#### 3.1 Discourse markers

Discourse markers are used in conversation to indicate discourse structure and provide sign-posts to conversational participants about how to interpret talk (Fox Tree, 2010, 2015; Haselow, 2019). Two discourse markers in particular are associated with providing information about how to interpret conversational contributions: *you know* and *like* (Haselow, 2019). They have been called *tailored markers* because they are tailored to the particular addressees engaged in the conversation (Fox Tree, 2015). The discourse marker uses of *like* and *you know* are common in dialogue. Two examples of discourse marker uses of *like* are “this guy came up to me and like tried to run in front of me” (Liu et al., 2016, p. 3160) and “it was like really empty” (Fox Tree, 2006, p. 731). Two examples of discourse marker uses of *you know* is “Everybody wakes up and goes straight to the bathroom, you know, putting on all their make up and everything” (Fox Tree and Tomlinson, 2008, p. 102) and “it’s my my favorite car but you know they’re not they’re not great cars” (Fox Tree, 2001, p. 734).

*Like* is a marker of loose expression of language (Andersen, 1998) – “a precise marker of imprecision” (Fox Tree, 2006, p. 729). Experimental tests demonstrate that *like* is not the same as hedges (Liu and Fox Tree, 2012), and that *like* is functional, not sprinkled in to indicate informal language (Fox Tree, 2006). *Like* is pragmatically useful in interviews, where “*like* is used to focus on salient information, qualify contributions, and introduce examples” (Fuller, 2003, p. 370). *Likes* are also used by interviewers trying to sound less formal (Fuller, 2003). People report adjusting their use of *like* for their addressees, using it more with friends (Fox Tree, 2007). The argument has been made that to use *likes* properly, conversational participants need to know something about each other (Liu and Fox Tree, 2012). Our expectation was therefore that if people feel more able to interpret each other's conversational contributions in person, they should use more *likes* with each other in person than in telepresent settings.

## Conversational Fluency and Attitudes Towards Robot Pilots...

*You know* is used as an invitation to the hearer to fill out the speaker's meaning (Fox Tree and Schrock, 2002). It decreases social distance (Stubbe and Holmes, 1995), and, indeed, people report using *you know* more with friends (Fox Tree, 2007). It has been argued that *you know* requires less tailoring than *like*; Fox Tree (2015) found a bigger difference between written and spoken *like* use than written and spoken *you know* use. We therefore anticipated larger differences in *like* use than *you know* use across telepresent and in-person settings.

Together, we hypothesized that *likes* and *you knows* would occur more often in in-person communication than telepresence communication because of the decreased psychological distance in in-person communication (Fox Tree et al., 2021).

### 3.2 Fillers

Unlike discourse markers which are used in the process of conversational negotiation, *fillers* (the words *uh* and *um*) are associated with speech processing difficulty. Two examples of uses of fillers are “so right where you found that um painting” and “from Walnut you should uh make a left on Cedar” (both examples are from Liu et al., 2016, p. 3162). Fillers indicate upcoming delays in communication, which are often indicated by silent pauses or more fillers (Clark and Fox Tree, 2002). They can be elongated to indicate delay as well (Clark and Fox Tree, 2002), as has been observed with other words (Fox Tree and Clark, 1997). Listeners use fillers to assist in comprehending upcoming speech (Fox Tree, 2001), including making judgments about why the speaker needs to delay, such as when they are uncomfortable with a topic (Fox Tree, 2002) or that they are lying (Fox Tree, 2002; Hosman and Wright, 1987). But delays occur across conversational settings – they would be expected in both telepresent communication and in-person communication. Consequently, we hypothesized that *ums* and *uhs* will not differ across settings.

### 3.3 Laughter

Laughter in conversation accomplishes complex interactional goals. Far beyond being a response to humor, laughter is a response to others (Provine, 1993; Provine and Fischer, 1989). In a week-long diary study, laughter was 30 times more likely to occur with others than when alone (Provine and Fischer, 1989), supporting the argument that laughter is a sign of rapport and playfulness (Provine, 1993). At the same time, in a study of communication across pairs in a variety of settings, laughter was more likely to occur in response to one's own speech than another's speech (Adelswärd, 1989). The settings assessed included job interviews, professional conversations, and simulated negotiations.

Despite the higher rate of laughing at one's own speech, laughing together is important to conversational success. The laughter produced in a dyadic setting can be mutual across two conversational participants or unilateral, where only one of the two participants laughs. Adelswärd (1989) found that job interviews with more mutual laughter compared to unilateral laughter were more likely to lead to job offers. Another finding related to mutual laughter was that in post-trial interviews with defendants accused of fraud, defendants produced more unilateral laughter and initiated more mutual laughter than the interviewers (Adelswärd, 1989). The simulated negotiations were of two types: seeking agreement, or seeking to win. While there was more laughter in the conflict condition and more unilateral laughter across both conditions, the proportion of unilateral laughter was lower in the agreement condition (Adelswärd, 1989). That is, seeking consensus led to more mutual laughter.

In this study, we hypothesized that people would be better able to use laughter in the in-person communicative setting than the telepresent setting. We predicted this because people experience less psychological distance in person (Fox Tree et al., 2021).

### 3.4 Gaze

Where we look has a large effect on how we experience conversations. We rely on gaze to facilitate turn transitions (Duncan, 1972; Novick et al., 1996), to disambiguate reference to objects in the environment (Hanna and Brennan, 2007), to check understanding of what was said (Monk and Gale, 2002), and to seek information on how someone is reacting to us (Argyle and Dean, 1965). Moreover, the way we gaze reflects communicative difficulty. Novick et al. (1996) compared two types of gaze patterns; one was the *mutual-break* pattern, where “as one conversant completes an utterance, he or she looks toward the other. Gaze is momentarily mutual, after which the other conversant breaks mutual gaze and begins to speak” (p. 1889). The other pattern was the *mutual-hold* pattern, where “the turn recipient begins speaking without immediately looking away” (p. 1889). Mutual-hold was used when conversational participants had more difficulty communicating (Novick et al., 1996). In our study, we assessed average gaze duration across settings. Based on prior work, we predicted more gaze in the telepresent communication, which we predicted would be more difficult for participants than in-person communication.

## 4 Hypotheses

Telepresence robot-mediated interaction is typically evaluated in comparison to in-person interaction. We therefore tested how people (1) evaluated a telepresence robot interviewer and (2) behaved with a telepresence robot interviewer as compared to an in-person interviewer in a within-participants study design. The



## Conversational Fluency and Attitudes Towards Robot Pilots...

setting was a mock job interview where participants were primed in advance to expect to participate with either a human, a robot, or a human piloting a robot with physical limitations. These primes were intended to approximate the conceptual metaphors for telepresence robots identified by Takayama and Go (2012).

Based on previous research on video-mediated communication (e.g., Storck and Sproull, 1995), we expect participants to be more positive about the in-person interviewer. Based on previous research on metaphorical communication primes (Takayama and Go, 2012), we expected participants to be more positive when they expected a human interviewer. Based on prior work on laughter and discourse markers, we predicted people would produce more mutual laughter, *likes*, and *you knows* with the in-person interviewer because of the use of these elements in the presence of others or with friends or to decrease social distance (Fox Tree, 2007; Fuller, 2003; Provine and Fischer, 1989; Stubbe and Holmes, 1995). Another way to think about this prediction is that telepresence communication increases psychological distance (Fox Tree et al., 2021), leading to less mutual laughter and fewer *likes* and *you knows* with telepresence. While proportionally more unilateral laughter was found in a conflict setting (Adelswärd, 1989), in our study we did not incorporate conflict. We predicted more unilateral and mutual laughter in person. Further, because fillers are used to indicate upcoming delay rather than to decrease social distance (Clark and Fox Tree, 2002), we did not predict differences in filler use across settings. Finally, we predict that participants will gaze more at the robot interviewer than the in-person interviewer. Gaze and eye movements convey important interactional cues (Argyle, 1990; Argyle and Dean, 1965; Duncan, 1972; Novick et al., 1996), but the interviewer's eyes are less visible through the robot's small screen than in person, so we expect that participants will gaze more at the robot interviewer in an attempt to overcome this perceptual limitation. The hypotheses are summarized in Table 1.

## 5 The telepresence robot interview study

We tested the role of setting (in-person, telepresence) and metaphorical prime (robot, person, “disabled” robot-person combination) on attitudes towards the interviewer and the production of discourse, laughter, and gaze phenomena.

### 5.1 Method

Participants participated in a mock job interview that involved multiple activities.

**Participants** Fifty-four people participated in this study, including 53 undergraduate students from a West Coast research university in the United States and 1 participant not affiliated with

Table 1. Hypotheses

Condition	Prediction	Assessed By
In-person versus telepresence	More positive attitudes in person	Ratings of: (1) likableness (2) awkwardness (3) capableness (4) intelligence (5) intimidating-ness (6) politeness (7) in-control-ness
Metaphorical prime (robot, person, "disabled" robot-person combination)	More positive attitudes towards people	Ratings of: (1) likableness (2) awkwardness (3) capableness (4) intelligence (5) intimidating-ness (6) politeness (7) in-control-ness
In-person versus telepresence	More tailored markers in person	Frequency of: (1) <i>likes</i> (2) <i>you knows</i>
In-person versus telepresence	No differences in filler rate across settings	Frequency of: <i>ums</i> and <i>uhs</i>
In-person versus telepresence	More mutual laughter in person	Frequency of instances of mutual laughter
In-person versus telepresence	More unilateral laughter in person	Frequency of instances of unilateral laughter
In-person versus telepresence	More gazing at interviewer in telepresence	Percentage of time participants spent gazing at the interviewer

## Conversational Fluency and Attitudes Towards Robot Pilots...

the university. The undergraduates received course credit for participation. Two participants declined to be filmed, resulting in 52 participants for the behavioral measures.

**Design** The experiment was a 3 (metaphor prime: robot/person/“disabled” robot-person combination) × 2 (interviewer modality: telepresence/in-person) design. The metaphor prime was a between-subject variable and the interviewer modality was a within-subject variable.

Before entering the experiment room, participants were primed with three different metaphors for the interviewer. These conditions were selected based on a subset of the five categories defined by Takayama and Go (2012). The conditions were: (1) the *robot* condition, (2) the *person* condition, and (3) the “*disabled*” *robot-person combination* condition. Participants in each condition received a different version of instructions; all versions contained the same information but featured different wording and a different image representing the interviewer.

In the *robot* condition, participants were primed to think of the robot interviewer as an object. Instructions included a photo of the robot with a non-human smiley face (captioned: “The interviewer, a Beam+robot”) and used phrasing appropriate for an inanimate object (e.g., “you will be greeted by *the robot*... *it* will ask you... answer *its* questions”). In the *person* condition, participants were primed to think of the robot interviewer as an extension of the human operating it. Instructions included a photo of the human interviewer (e.g., captioned: “The interviewer, Robert”) and used phrasing appropriate for a person (e.g., “you will be greeted by *the interviewer*... *he* will ask you... answer *his* questions”). In the “*disabled*” *robot-person combination* condition, participants were also primed to think of the robot as an extension of the human interviewer operating it, but with an additional suggestion that the interviewer has limited physical capacity. Instructions included a photo of the robot with the face of the human operator superimposed on it (e.g., captioned: “The interviewer, Robert, using the robot”) and used the same phrasing as the human condition, with the following additional instruction (“Please be aware that Robert has limited physical capability while piloting the robot, and he may require assistance maneuvering or manipulating objects”).

The main task in the study was an interview with two phases. One interview phase was conducted using the Beam+ in robot-mediated interaction, and the other phase was in person. Twenty-eight interviews were conducted in the robot-mediated interaction first, and 26 were conducted in person first. The interviews analyzed in the present study were conducted in two sets approximately one year apart by two male, native English-speaking students. Both interviewers received training and practice in using the Beam+ robot and the interview protocol prior to conducting their first interviews.

**Procedure** Participants were invited to the lab to participate in a mock job interview. Upon arriving at the lab, participants encountered a poster mounted on the closed lab door which

provided instructions explaining the study procedure and included an image of the interviewer (either a picture of the Beam+, a head shot of the interviewer, or a picture of the Beam+ robot with an image of the interviewer in the screen). Participants were instructed to enter the lab after fully reading the instructions.

Upon entering the lab, participants encountered either the robot interviewer (the Beam+ robot piloted by the interviewer) or the interviewer in person. The order first encountered, robot or human, was counterbalanced. We used a Suitable Technologies Beam+ telepresence robot. It stands approximately 4.4ft (1.35 m) tall and features a 10 inch LCD screen mounted on a long neck attached to a motorized base approximately 14×12 inches (0.36x.3 m) in size. The form factor of the Beam+ is roughly equivalent to a tall seated adult or a short standing adult. The system also features two cameras (one facing downward to assist the pilot in maneuvering and avoiding obstacles), speakers, and a microphone array. The system was controlled remotely using Beam software running on a MacBook Air and connected to the Beam+ over a local WiFi network. This software provides the remote pilot with a simultaneous view of images from both the front-facing and down-facing cameras, and allows the robot to be controlled by keyboard or mouse/trackpad input. During this study we disabled the picture-in-picture view on the Beam+ display so that participants would not see a view of themselves while interacting with the system.

Prior to each interview that began with the robot interviewer, the Beam+ robot was positioned next to the table, facing the door that participants would use to enter the room. When the interview began in person, the interviewer was seated in a chair in the same location. The interviewer greeted participants, welcomed them to the lab, directed them to read and complete a consent form, and verified that they had fully read the instructions, thereby ensuring they received the priming condition. A second copy of the instructions was placed on the table next to the consent form in case any participant had not fully read the instructions on their own. After giving consent, the participants were invited to sit down at the table across from the interviewer. If the interviewer was using the Beam+, the interviewer moved the Beam+ robot to a position at the table approximating a comfortable seated conversation. If the interviewer was in person, he seated himself in a chair located in the same place. The participant sat in a single open chair was placed on the opposite side of the table next to a collection of office supplies (several sheets of paper, a stapler, a box of staples, a small digital timer, a whiteboard eraser, three whiteboard pens, and two ballpoint pens). These objects were selected as items which could be used as props during the interview and would not seem out of place in an office setting. The chair(s) and office supplies were arranged in the same position before each interview.

During one half of the interview, the participants interacted with the Beam+ robot piloted remotely by their interviewer (the *robot interviewer*). During the other half of the interview, they were interviewed by the same interviewer in person (the *human interviewer*). After completing the interview, participants

## Conversational Fluency and Attitudes Towards Robot Pilots...

filled out a short online survey asking them about their experience during the interview and their subjective rating of the robot interviewer and the in-person interviewer.

Audio and video recordings were captured using a GoPro Hero4 video camera mounted on a tripod located next to the table opposite the study participants. The GoPro was positioned so that both the interviewer and participant were visible in the recording. Secondary recordings were also captured using Screencastify (screen capture software) running on the MacBook Air to record the robot-interviewer portion of the interview, and using a MeCam Classic camera with a lanyard mount (worn around the interviewer's neck) to record the human-interviewer portion of the interview.

**Interview questions** The first half of the interview opened with a series of warm-up questions, such as: “How are you doing today?,” “Can you tell me a little bit about your previous work experience?,” and “What kind of job would you like to have in the future?” We refer to this as the *conversational* portion of the interview. Next, the interviewer asked a series of questions framed as creative thinking questions; we refer to this as the *formal* portion of the interview. Some questions were designed to be answered entirely verbally (e.g., “Why is the earth round?” or “If you were a box of cereal, what would you be and why?”). These were based on a blog post discussing the use of creative interviewing questions (Greenberg, 2015). The remaining questions required interaction with the objects on the table (e.g., “Using the items on the table in front of you, please act out a scene from a movie, show or book. Spend about a minute or two on this and give as much detail as you can.” or “Using the items on the table, please arrange them to represent a map of a place you have lived...”).

The interviewer asked a series of 10 questions in the following order: three verbal, two interactive, three verbal, two interactive.

After these questions, the interviewer stated that the first portion of the interview was over. In interviews where the first half was conducted via the Beam+, the interviewer explained that they would return the robot to its charging station and come to the room in person to continue the interview. The interviewer then piloted the robot to the side door of the room, at which point they turned the robot to face the participant and asked the participant to open the door so that the interviewer could exit the room. The interviewer parked the robot in its charging station in the adjacent room, disconnected from the robot, stopped screen recording, activated the wearable MeCam camera, and joined the participant in the other room. In interviews where the first half was conducted by the human interviewer, the interviewer explained that the robot was charged and ready to conduct the second half of the interview (after stating earlier that it needed to charge), left the

room by the side door, connected to the robot, and piloted the robot around to the door that participants used to enter the room. At that point the interviewer asked the participant to open the door so that they could enter the room. The interviewer then piloted the robot to approximately the same position across the table from the participant that the human interviewer previously occupied.

The second half of the interview was designed to match the structure and content of the first half of the interview, with the exception that the conversational portion followed the formal questions to better fit the structure of an interview and maintain a more natural conversational flow. Upon entering the room, the interviewer thanked the participant for waiting, confirmed that the participant was ready to continue, and then began the second formal portion of the interview. These questions were matched as closely as possible to the previous questions and contained the same proportion of verbal and interactive questions. Ten questions were asked in the same order as before, three verbal, two interactive, three verbal, two interactive. The interviewer concluded with a second conversational portion of the interview. We attempted to match question content and duration to the warm-up questions at the beginning of the interview (e.g., “If this had been a real job interview, how do you think you did?” and “What did you think of the questions that we asked you?”).

**Dependent measures** There were two sets of dependent measures: (1) a post-study survey of attitudes and (2) an assessment of discourse phenomena in transcripts of the interviews, including discourse markers, fillers, laughter, and gaze.

The attitude questions assessed the participants’ view of the telepresent interviewer and the in-person interviewer. There were two sets of seven identical statements, with the statements making claims about *the robot interviewer* or *the human interviewer*, and participants saw both sets. The seven statements probed how likable, awkward, capable, intelligent, intimidating, polite, and in control the participants thought the interviewer was. In the set about the robot interviewer, participants saw statements of the form “The robot interviewer was polite,” and in the set about the human interviewer, participants saw “The human interviewer was polite.” Participants were asked to respond on a scale of 1 to 5, with 1 being *strongly disagree* and 5 being *strongly agree*.

The interviews were transcribed by trained research assistants using a modified and simplified version of the Jeffersonian system used in conversation analysis research (Jefferson, 2004). Discourse markers, fillers, laughter, and gaze were hand-coded in a subset of the transcribed interviews. For laughter, research assistants counted the number of times a participant laughed during the interview. Each instance was coded as being produced by only the participant (unilateral laughter), or by the participant and the interviewer in immediately adjacent turns, including when their laughter overlapped (mutual laughter). The process of gaze

## Conversational Fluency and Attitudes Towards Robot Pilots...

assessment was laborious. It involved reviewing the video recordings and indicating in the transcripts whenever the participant looked at the robot, along with the duration of the gaze. In this study, we report the average time spent gazing throughout the entire interview per interviewee. In all, 50% of the interviews were coded for discourse markers, fillers, laughter, and gaze.

### 5.2 Results

Results are presented for attitudes and discourse phenomena.

**Attitudes** There was no effect of order of condition on attitudes,  $F(35, 230)=0.876, p=0.672$ .

To investigate the role of prime (robot interviewer, human interviewer, “disabled” human interviewer) on participant attitudes toward the interviewer, we conducted a MANOVA. There was no effect of prime on any of the attitude questions,  $F(14, 92)=1.31, p=0.220$ .

To investigate the role of interviewer setting (telepresence, in-person) on attitudes, we conducted a MANOVA and found that interviewer setting has a statistically significant effect on attitudes,  $F(7, 47)=7.124, p<0.001$ . These were followed by univariate tests with Bonferroni corrections to see which attitudes were affected. Participants rated the robot interviewer as more awkward than the in-person interviewer. They rated the in-person interviewer as more likable, more capable, more intelligent, more polite, and more in control than the robot interviewer. See Table 2 for results of the attitude assessments.

**Table 2.** Attitude assessment results.

	Mean (SD) Telepresence	Mean (SD) In-Person	Statistics
AWKWARD The human / robot interviewer was awkward	2.26 (1.08)	1.81 (0.99)	$F(1,53)=9.217, p=0.004$
LIKABLE The human / robot interviewer was likable	4.20 (0.83)	4.72 (0.53)	$F(1,53)=24.44, p<0.001$
CAPABLE The human / robot interviewer was capable	4.13 (0.99)	4.78 (0.42)	$F(1,53)=25.96, p<0.001$
INTELLIGENT The human / robot interviewer was intelligent	4.20 (0.77)	4.56 (0.63)	$F(1,53)=17.44, p<0.001$
POLITE The human / robot interviewer was polite	4.11 (0.63)	4.80 (0.41)	$F(1,53)=6.08, p=0.017$
IN CONTROL The human / robot interviewer was in control	3.74 (1.03)	4.13 (1.08)	$F(1,53)=4.99, p=0.03$
INTIMIDATING The human / robot interviewer was intimidating	2.07 (1.24)	2.17 (1.19)	$F(1,53)=0.321, p=0.574$

**Table 3.** Discourse phenomena results.

	Mean (SD) Telepresence	Mean (SD) In-Person	Test statistics
<i>like</i>	0.03 (0.03)	0.04 (0.04)	$t(25) = -2.76, p = 0.06$
<i>you know</i>	0.005 (0.01)	0.007 (0.01)	$t(25) = -1.91, p = 0.4$
fillers	0.07 (0.06)	0.08 (0.08)	$t(25) = -1.71, p = 0.6$
mutual laughter	0.004 (0.004)	0.006 (0.008)	$t(25) = -1.80, p = 0.50$
unilateral laughter	0.01 (0.01)	0.02 (0.01)	$t(25) = -1.57, p = 0.76$
gaze	0.28 (0.09)	0.27 (0.09)	$t(25) = 0.53, p = 0.1$

For measures besides gaze, Mean(SD) represents the number of instances of the phenomenon divided by the total number of words. For gaze, Mean(SD) represents the percentage of time the participant spent gazing at the interviewer divided by the length of the interview

These results are consistent with previous researchers' findings that people tend to be rated as less intelligent, and generally less positively, when communicating via audio or video conferencing technologies as compared to face-to-face communication (Short et al., 1976; Whittaker and O'Conaill, 1997).

**Discourse phenomena** To investigate the role of interviewer setting (telepresence, in-person) on discourse phenomena, we conducted pairwise comparisons with Bonferroni corrections. We did not find significant differences for likes, you knows, fillers, unilateral laughter, mutual laughter, or gaze. See Table 3 for behavioral results.

### 5.3 Discussion

Only attitudes varied depending on conversational setting. Even when communicating with the same interlocutor, participants felt more positively about them in the in-person setting. We did not find significant differences in behavioral phenomena (discourse markers, fillers, laughter, and gaze). We also did not find evidence that attitudes differed depending on the way people were primed to think about their interlocutor (as a robot, a person, or a "disabled" robot-person combination). One possibility is that our primes (a poster that participants read on the door of the experiment room) were not strong enough to produce a detectable difference.

## 6 General discussion

The COVID-19 pandemic made everyone acutely aware of the ways that technology influences communication, including both the advantages and the disadvantages of telepresence versus in-person communication. For example, many people the world over learned how to work remotely via Zoom. This type of telepresence communication involves face-forward head-and-shoulders images. An



## Conversational Fluency and Attitudes Towards Robot Pilots...

advantage of this is that it directs attention to communicators' faces which contain a lot of information, such as mouth opening to indicate a desire to speak (Krause and Kawamoto, 2019, 2021), raising eyebrows to indicate prosodic structure (Krahmer and Swerts, 2007), and head movements that indicate listener comprehension (Li, 1999) or affiliation (Stivers, 2008). But video conferencing also has disadvantages, like restricting movement. Movement has been shown to be useful for indicating topic shifts and turn exchanges (Cassell et al., 2001). Also, videoconferencing fatigue can result from an overemphasis on work at the expense of sociality (Bergmann et al., 2022).

Telepresence robots provide advantages in comparison to stationary remote communication. They have the potential to provide a greater sense of presence, and they allow a remote communicator to physically move around a space like an in-person communicator would. Even slight movements of a telepresence robot can be a form of body language for initiating and ending conversations (Neustaedter et al., 2016). Most telepresence robots have wide-angle cameras and some can swivel screens (Nichols, 2022), both elements that are missing in Zoom telepresence. One recent model exploits maps of the area to improve navigation (Nichols, 2022), which frees up a communicator's energy to focus on interactions instead of driving the robot. Studies of telepresence robots in the workplace have found that they promote casual interaction and can build social connections among geographically distributed team members (Lee and Takayama, 2011). Industry experts predict that the Covid-19 pandemic could drive greater demand for telepresence robots (Nichols, 2022), especially as remote work becomes more widespread and workers are reluctant to return to in-person workplaces (Goldberg, 2022).

Yet despite efforts to more closely model the in-person experience, telepresence communication still falls short, as the findings of this study show. Robot-mediated communication was assessed as less socially desirable than face-to-face communication. Importantly, we found these results even though the participants were communicating with the same addressee in the same session – each participant experienced both types of communication with the same interviewer. Because of our within-participants design, we can conclude that what we observed was a product of the communicative medium, not the communicator.

We took a close look at how people communicated across telepresence and in-person communication, as well as testing attitudes towards these communicative modalities. We anticipated that discourse phenomena that are hallmarks of casual conversations – words like *um*, *like*, and *you know*, as well as laughter and gaze patterns – might occur more frequently in face-to-face communication as opposed to robot-mediated videochat. We did not find evidence of such differences in usage, however. We note that these data were collected before the COVID-19 pandemic. Increased familiarity with videochat communication since the pandemic could affect discourse phenomena usage. For example, people

might speak more naturally with robots in our post-pandemic world, much like in the early days of texting when people with more experience texted more like they spoke (Fox Tree et al., 2011). Alternatively, increased expertise could induce different effects; for example, participants might know that their gaze patterns are not properly transmitted through a videochat camera and therefore might adjust their behavior, such as gazing directly into the camera instead of at their addressee's virtual face (O'Conaill et al., 1993). Closer analysis of discourse phenomena might reveal information about how people use technology that is not evident from measurements of how people feel about technology.

This study has many theoretical and practical implications. Theoretically, the study provides new knowledge about the relationship between feelings about interlocutors who use different communicative modalities and indicators of communicative effectiveness, such as the production of discourse phenomena. We found that feelings about the interlocutor can be affected by modality even though communicative phenomena were not. Practically, the study provides knowledge about how communicative modalities can change attitudes *towards the same person*. This has many implications. For example, it highlights the importance of using the same modality for all interviewees during the hiring process. If some are interviewed by Zoom and others in person, the people interviewed by Zoom may be at a disadvantage. Likewise, as hybrid work becomes the norm, teams constituted of a mix of remote and co-present workers may experience interpersonal attitudinal differences that reflect the use of mediated communication.

## 6.1 Future work

The design of telepresence robots could improve in ways that would improve social interaction and conversational flow in the future. For example, autonomous navigation could reduce social awkwardness associated with bumping into objects (Desai et al., 2011). To help interlocutors establish eye contact, screens with cameras embedded in the center (Kristofferson et al., 2013) and other techniques involving semi-reflective screens have been proposed (Ishii and Kobayashi, 1992). Remotely-controllable arms could make robots more socially desirable by enabling them to gesture, shake hands, and hug. These robots would also be less “disabled” and dependent on assistance from others (Herring, 2016).

It is also possible that as people gain experience with telepresence robots, they may change how they feel about them and how they use them (Fox Tree et al., 2011; Lei et al., 2022; Oviedo and Fox Tree, 2021). Only a handful of study participants commented on interacting previously with a robot; 93% rated their experience with robots as *none* (82%) or *a little* (11%). Future researchers could examine whether more experience with telepresence robots results in people's interactions more closely resembling their in-person communications. Researchers could also study how people think about and behave when interacting with novice robot pilots, a situation that is likely to occur in real-world contexts where

## Conversational Fluency and Attitudes Towards Robot Pilots...

telepresence robots are available for public use, such as to attend conferences, visit museums, or go on campus tours (e.g., Neustaedter et al., 2016, 2018).

A related direction for future research concerns the effects of telepresence robotics on discourse in naturalistic (non-experimental) settings. So far, such data have been hard to come by due to privacy concerns and the challenge of getting informed consent from people who might happen to interact with one's research robot "in the wild," such as in a museum or at a conference reception (Neustaedter et al., 2016, 2018). Authentic, unplanned interactions with a telepresence robot raise many questions about discourse use (Herring, 2016). For example, absent priming, how do others refer to the robot—as "you," "s/he," or "it"—and what factors condition variation in reference? Lee and Takayama (2011) found that people in a workplace setting who thought of the robot as a machine were more likely to refer to "it"; does this depend on the robot pilot's activity and discourse behaviors? To what extent do local persons and telepresence robot pilots accommodate to each other stylistically? Do interlocutors' social status and gender influence this and other features of participant alignment, such as informality and use of pronouns that signal group identity and grounding?

Future researchers might also study how settings affect communicative effectiveness. For example, researchers might test how settings affect conversational balance or grounding. Static videoconferencing has been shown to be unbalanced, with isolated remote participants contributing fewer turns and less content (O'Conaill et al., 1993). Other researchers have found that conversational participants who are on the same social footing strive to rebalance conversations after periods where one participant contributed an outsized share of the dialogue, and success at rebalancing was related to positive feelings about the conversation (Guydish et al., 2021; Guydish and Fox Tree, 2022). How people feel about conversations has also been related to successful grounding (Guydish and Fox Tree, 2021). One reason people find in-person communication more comfortable than telepresence communication may be because they are better able to balance their conversations in person.

## 6.2 Conclusion

Interviewers using mobile telepresence communication were considered more awkward, less likable, less capable, less intelligent, less polite, and less in control. Behaviorally, we did not detect differences in participants' productions of discourse markers (*likes* and *you knows*), fillers (*ums* and *uhs*), laughter, or gaze. We did not observe differences in the way people were primed to think about their addressee (as a robot, a person, or a "disabled" robot-person combination) on the attitudes they held about their addressee, but our primes may not have been strong enough. These are many avenues for future exploration, including analyzing how conversational participants produce other discourse phenomena,

how they balance their conversations, how they ground using different communication technologies, and how their level of experience with mobile telepresence – both as pilots and as local users – affects their communication and attitudes. Future experimenters could explore other methods of priming individuals before their interactions. Future researchers could also seek to collect and analyze more interactions with telepresence robots in naturalistic settings.

### **Acknowledgements**

We thank the many research assistants who helped us with this project, with a special thank you to Kevin Weatherwax, Isabel Whittaker-Walker, and Madison Chartier. We also thank Yasmin Chowdhury for transcript supervision.

**Authors' contributions** The experiment was designed by R.M., S.W., J.F.T, and S.H. The experiment was run by R.M., S.W., and L.T. The data were coded by S.H. and A.N. The final data analyses were completed by A.N. The main manuscript text was written by J.F.T. with sections written by S.H., A.N., S.W., R.M., and L.T. All authors reviewed the manuscript.

**Funding** Research funds granted to Susan Herring by Indiana University Bloomington were used for this project. Research funds granted to Leila Takayama by the University of California Santa Cruz were used for this project.

**Data availability** To protect confidentiality of participants, recordings are not available.

### **Declarations**

**Ethical approval** This project was approved by the University of California Santa Cruz Institutional Review Board.

**Competing interests** The authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Conversational Fluency and Attitudes Towards Robot Pilots...

### References

- Adelswärd, Viveka (1989). Laughter and dialogue: The social significance of laughter in institutional discourse. *Nordic Journal of Linguistics*, vol. 12, no. 2, pp. 107-136. <https://doi.org/10.1017/S0332586500002018>
- Andersen, Gisle (1998). The pragmatic marker like from a relevance-theoretic perspective. In A. H. Jucker; and Y. Ziv (eds): *Discourse Markers: Descriptions and Theory*, Amsterdam: John Benjamins, pp. 147-70. <https://doi.org/10.1075/pbns.57.09and>
- Argyle, Michael (1990). *Bodily Communication*. Abingdon, Oxfordshire, UK: Routledge.
- Argyle, Michael; and Janet Dean (1965). Eye contact, distance and affiliation. *Sociometry*, vol. 28, no. 3, pp. 289-304. <https://doi.org/10.2307/2786027>
- Beattie, Geoffrey W. (1978). Sequential temporal patterns of speech and gaze in dialogue. *Semiotica*, vol. 23, nos. 1-2, pp. 29-52.
- Bergmann, Rachel; Sean Rintel; Nancy Baym; Advait Sarkar; Damian Borowiec; Priscilla Wong; and Abigail Sellen (2022). Meeting (the) pandemic: Videoconferencing fatigue and evolving tensions of sociality in enterprise video meetings during COVID-19. *Computer Supported Cooperative Work (CSCW)*, pp. 1-37. <https://doi.org/10.1007/s10606-022-09451>
- Bonfert, Michael; Maximilian Spliethöver; Roman Arzaroli; Marvin Lange; Martin Hancil; and Robert Porzel (2018, October). If you ask nicely: A digital assistant rebuking impolite voice commands. *ICMI'18: Proceedings of the 20th ACM International Conference on Multimodal Interaction, Boulder, Colorado, 16 October - 20 October 2018*. New York: ACM Press, pp. 95-102. <https://doi.org/10.1145/3242969.3242995>
- Bruce, Vicki (1995). The role of the face in face-to-face communication: Implications for videotelephony. In S. Emmot (ed): *Information Superhighways: Multimedia Users and Futures*. San Diego: Academic Press, pp. 227-237.
- Cassell, Justine; Yukiko I. Nakano; Timothy W. Bickmore; Candace L. Sidner, and Charles Rich (2001). Non-verbal cues for discourse structure. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France, 6 July - 11 July 2001*, pp. 114-123. <https://doi.org/10.3115/1073012.1073028>
- Chapanis, Alphonse (1975). Interactive human communication. *Scientific American*, vol. 232, no. 3, pp. 36-42.
- Chapanis, Alphonse; Robert B. Ochsman; Robert N. Parrish; and Gerald D. Weeks (1972). Studies in interactive communication: The effects of four communication modes on the behavior of teams during cooperative problem solving. *Human Factors*, vol. 14, no. 6, pp. 487-509. <https://doi.org/10.1177/00187208720140060>
- Clark, Herbert H.; and Susan E. Brennan (1991). Grounding in communication. In L. B. Resnick; J. M. Levine; and S. D. Teasley (eds): *Perspectives on Socially Shared Cognition*. Washington, D.C.: American Psychological Association, pp. 127-149.
- Clark, Herbert H.; and Jean E. Fox Tree (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, vol. 84, pp. 73-111. [https://doi.org/10.1016/S0010-0277\(02\)00017-3](https://doi.org/10.1016/S0010-0277(02)00017-3)
- Cohen, Karen M. (1982). Speaker interaction: Video teleconferences versus face-to-face meetings. In *Proceedings of Teleconferencing and Electronic Communications*. Madison, Wisconsin: University of Wisconsin Press, pp. 189-199.
- Desai, Munjal; Katherine M. Tsui; Holly A. Yanco; and Chris Uhlik (2011). Essential features of telepresence robots. *TEPRA '11: Proceedings of the IEEE International Conference on Technologies for Practical Robot Applications, Woburn, MA: 11 April - 12 April 2011*. IEEE Press, pp. 15-20. <https://doi.org/10.1109/TEPRA.2011.5753474>
- Duncan, Starkey (1972). Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology*, vol. 23, no. 2, pp. 283-292. <https://doi.org/10.1037/h0033031>

- Fiorini, Laura; Gianmaria Mancioppi; Claudia Becchimanzi; Alessandra Sorrentino; Mattia Pistolesi; Francesca Tosi; and Filippo Cavallo (2020). Multidimensional evaluation of telepresence robot: Results from a field trial. *RO-MAN 2020: 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy: 31 August – 4 September 2020*. IEEE Press, pp. 1211–1216. <https://doi.org/10.1109/RO-MAN47096.2020.9223467>
- Fischer, Aaron J.; Bradley S. Bloomfield; Racheal R. Clark; Amelia L. McClellan; and William P. Erchul (2019). Increasing student compliance with teacher instructions using telepresence robot problem-solving teleconsultation. *International Journal of School and Educational Psychology*, vol. 7, no. S1, pp. 158-172. <https://doi.org/10.1080/21683603.2018.1470948>
- Fox Tree, Jean E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory and Cognition*, vol. 29, no. 2, pp. 320-326. <https://doi.org/10.3758/bf03194926>
- Fox Tree, Jean E. (2002). Interpreting pauses and *ums* at turn exchanges. *Discourse Processes*, vol. 34, no. 1, pp. 37-55. [https://doi.org/10.1207/S15326950DP3401\\_2](https://doi.org/10.1207/S15326950DP3401_2)
- Fox Tree, Jean E. (2006). Placing *like* in telling stories. *Discourse Studies*, vol. 8, no. 6, pp. 749-770. <https://doi.org/10.1177/1461445606069287>
- Fox Tree, Jean E. (2007). Folk notions of *um* and *uh*, *you know*, and *like*. *Text and Talk*, vol. 27, no. 3, pp. 297-314. <https://doi.org/10.1515/TEXT.2007.012>
- Fox Tree, Jean E. (2010). Discourse markers across speakers and settings. *Language and Linguistics Compass*, vol. 3, no. 1, pp. 1–13. <https://doi.org/10.1111/j.1749-818X.2010.00195.x>
- Fox Tree, Jean E. (2015). Discourse markers in writing. *Discourse Studies*, vol. 17, no. 1, pp. 64–82. <https://doi.org/10.1177/1461445614557758>
- Fox Tree, Jean E.; and Clark, H. H. (1997). Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, vol. 62, pp. 151-167. [https://doi.org/10.1016/S0010-0277\(96\)00781-0](https://doi.org/10.1016/S0010-0277(96)00781-0)
- Fox Tree, Jean E.; and Josef C. Schrock (2002). Basic meanings of *you know* and *I mean*. *Journal of Pragmatics*, 34, 727-747. [https://doi.org/10.1016/S0378-2166\(02\)00027-9](https://doi.org/10.1016/S0378-2166(02)00027-9)
- Fox Tree, Jean E.; and John M. Tomlinson, Jr. (2008). The rise of *like* in spontaneous quotations. *Discourse Processes*, vol. 45, pp. 85-102. <https://doi.org/10.1080/01638530701739280>
- Fox Tree, Jean E.; Sarah A. Mayer; and Betts, Teresa. E. (2011). Grounding in instant messaging. *Journal of Educational Computing Research*, vol. 45, no. 4, 455-475. <https://doi.org/10.2190/EC.45.4.e>
- Fox Tree, Jean E.; Steve Whittaker; Susan C. Herring; Yasmin Chowdhury; Allison Nguyen; and Leila Takayama (2021). Psychological distance in mobile telepresence. *International Journal of Human-Computer Studies*, vol. 151, p.102629. <https://doi.org/10.1016/j.ijhcs.2021.102629>
- Janet M. Fuller (2003). Use of the discourse marker *like* in interviews. *Journal of Sociolinguistics*, vol.7, no. 3, pp. 365-377. <https://doi.org/10.1111/1467-9481.00229>
- Jefferson, Gail (2004). Glossary of transcript symbols with an introduction. *Conversation Analysis*, pp. 13-31. <https://doi.org/10.1075/pbns.125.02jef>
- Goldberg, Emma (2022, March 10). A two-year, 50-million-person experiment in changing how we work. *New York Times* <https://www.nytimes.com/2022/03/10/business/remote-work-office-life.html>
- Greenberg, Andrew (2015, April 9). Ask these creative interview questions. Retrieved June 04, 2016, from <http://www.recruitingdivision.com/creative-interview-questions>
- Guydish, Andrew J.; and Jean E. Fox Tree (2021). Good conversations: Grounding, convergence, and richness. *New Ideas in Psychology*, vol. 63, no. 1, p. 100877. <https://doi.org/10.1016/j.newideapsych.2021.100877>
- Guydish, Andrew J.; and Jean E. Fox Tree (2022). Reciprocity in instant messaging conversations. *Language and Speech*, vol. 65, no. 2, pp. 404-417. <https://doi.org/10.1177/00238309211025070>
- Guydish, Andrew J.; J. Trevor D'Arcey; and Jean E. Fox Tree (2021). Reciprocity in conversation. *Language and Speech*, vol. 64, no. 4, pp. 859–872. <https://doi.org/10.1177/0023830920972742>

## Conversational Fluency and Attitudes Towards Robot Pilots...

- Hanna, Joy E.; and Susan E. Brennan (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, vol. 57, no. 4, pp. 596–615. <https://doi.org/10.1016/j.jml.2007.01.008>
- Haselow, Alexander (2019). Discourse marker sequences: Insights into the serial order of communicative tasks in real-time turn production. *Journal of Pragmatics*, vol. 146, pp. 1–18. <https://doi.org/10.1016/j.pragma.2019.04.003>
- Herring, Susan. C. (2016). Robot-mediated communication. In R. A. Scott; M. C. Buchmann; and S. M. Kosslyn (eds): *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*. Hoboken, NJ: John Wiley & Sons, pp. 1–16. <https://doi.org/10.1002/9781118900772.etrds0414>
- Heshmat, Yassamin; Brennan Jones; Xiaoxuan Xiong; Carman Neustaedter; Anthony Tang; Bernhard Riecke; and Lillian Yang (2018). Geocaching with a Beam: Shared outdoor activities through a telepresence robot with 360 degree viewing. *CHI'18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, Canada, 21–26 April, 2018*. New York: ACM Press, pp. 1–13. <https://doi.org/10.1145/3173574.3173933>
- Hoffmann, Laura; Melanie Derksen; and Stefan Kopp (2020). What a pity, pepper! How warmth in robots' language impacts reactions to errors during a collaborative task. *HRI'20: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, Cambridge, UK, 23 March - 26 March 2020*. New York: ACM Press, pp. 245–247. <https://doi.org/10.1145/3371382.3378242>
- Hollingsworth, Kara (2022). *Transition-Relevance Places in Video-Mediated Conversations*. Honors thesis. Baylor University, Texas. <https://hdl.handle.net/2104/11837>
- Hosman, Lawrence A; and Wright II, John W. (1987). The effects of hedges and hesitations on impression formation in a simulated courtroom context. *Western Journal of Speech Communication*, vol. 51, no. 2, pp. 173–188. <https://doi.org/10.1080/10570318709374263>
- Huang, Lixiao; Daniel McDonald; and Douglas Gillan (2017). Exploration of human reactions to a humanoid robot in public STEM education. *HFES 61: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Austin, TX, 9 October - 13 October 2017*. Sage, pp. 1262–1266. <https://doi.org/10.1177/1541931213601796>
- Ishii, Hiroshi; and Minoru Kobayashi (1992). ClearBoard: A seamless medium for shared drawing and conversation with eye contact. *CHI92: Proceedings of ACM CHI Conference on Human Factors in Computing, Monterey, CA, 3 May - 7 May 1992*. New York: ACM Press, pp. 525–532. <https://doi.org/10.1145/142750.142977>
- Krahmer, Emiel; and Marc Swerts (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, vol. 57, no. 3, pp. 396–414. <https://doi.org/10.1016/j.jml.2007.06.005>
- Krause, Peter A.; and Alan H. Kawamoto (2019). Anticipatory mechanisms influence articulation in the form preparation task. *Journal of Experimental Psychology: Human Perception and Performance*, vol. 45, no. 3, pp. 319–335. <https://doi.org/10.1037/xhp0000610>
- Krause, Peter A.; and Alan H. Kawamoto (2021, July). Predicting one's turn with both body and mind: Anticipatory speech postures during dyadic conversation. *Frontiers in Psychology*, vol. 12, p. 2856. <https://doi.org/10.3389/fpsyg.2021.684248>
- Krauss, Robert; and Peter Bricker (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America*, vol. 41, pp. 286–292. <https://doi.org/10.1121/1.1910338>
- Kristofferson, Annica; Silvia Coradeschi; and Amy Loutfi (2013). A review of mobile robotic telepresence. *Advances in Human-Computer Interaction*, vol. 2013, January, article 3. <https://doi.org/10.1155/2013/902316>

- Kuster, Claudia; Tiberiu Popa; Jean-Charles Bazin; Craig Gotsman; and Markus Gross (2012). Gaze correction for home video conferencing. *ACM Transactions on Graphics*, vol. 31, no. 6, pp.1-6. <https://doi.org/10.1145/2366145.2366193>
- Kwon, Oh-Hun; Seong-Yong Koo; Young-Geun Kim; and Dong-Soo Kwon (2010). Telepresence robot system for English tutoring. *ARSO'10: IEEE Workshop on Advanced Robotics and Its Social Impacts, Seoul, S. Korea, 26 October - 28 October 2010*. New Jersey: IEEE Press, pp. 152–155. <https://doi.org/10.1109/ARSO.2010.5679999>.
- Lee, Min Kyung; and Leila Takayama (2011). “now, i have a body”: Uses and social norms for mobile remote presence in the workplace. *CHI'11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, Canada, 7 May - 12 May 2011*. New York: ACM Press, pp. 33–42. <https://doi.org/10.1145/1978942.1978950>.
- Lei, Ming; Ian Clemente; Haixia Liu; and John Bell (2022). The acceptance of telepresence robots in higher education. *International Journal of Social Robotics*, vol. 14, no. 4, pp. 1025–1042. <https://doi.org/10.1007/s12369-021-00837-y>
- Li, Han Z. (1999). Grounding and information communication in intercultural and intracultural dyadic discourse. *Discourse Processes*, vol. 28, no. 3, pp. 195-215. <https://doi.org/10.1080/01638539909545081>
- Liu, Kris; Jean E. Fox Tree; and Marilyn Walker (2016). Coordinating communication in the wild: The Artwalk dialogue corpus of pedestrian navigation and mobile referential communication. *LREC 2016: Proceedings of the International Conference on Language Resources and Evaluation, Portorož, Slovenia, 23 May – 28 May 2016*, pp. 3159–3166.
- Liu, Kris; and Jean E. Fox Tree (2012). Hedges enhance memory but inhibit retelling. *Psychonomic Bulletin & Review*, vol. 19, no. 5, pp. 892-898. <https://doi.org/10.3758/s13423-012-0275-1>
- Mirnig, Nicole; Gerald Stollnberger; Markus Miksch; Susanne Stadler; Manual Giuliani; and Manfred Tscheligi (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, vol. 4. <https://doi.org/10.3389/frobt.2017.00021>.
- Monk, Andrew; and Caroline Gale (2002). A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes*, vol. 33, no. 3, pp. 257-278. [https://doi.org/10.1207/S15326950DP3303\\_4](https://doi.org/10.1207/S15326950DP3303_4)
- Morley, Ian; and Stephenson, Geoffrey M. (1970). Formality in experimental negotiations: A validation study. *British Journal of Psychology*, vol. 61, no. 3, pp. 383- 384. <https://doi.org/10.1111/j.2044-8295.1970.tb01256.x>
- Neustaedter, Carman; Samarth Singhal; Rui Pan; Yasamin Heshmat; Azadeh Forghani; and John Tang (2018). From being there to watching. *ACM Transactions on Computer-Human Interaction*, vol. 25, no. 6, pp. 1–39. <https://doi.org/10.1145/3243213>
- Neustaedter, Carman; Gina Venolia; Jason Procyk; and Daniel Hawkins (2016). To Beam or not to Beam: A study of remote telepresence attendance at an academic conference. *CSCW'16: Proceedings of ACM Conference on Computer Supported Cooperative Work, San Francisco, CA, 27 February - 2 March 2016*. New York: ACM. <https://doi.org/10.1145/2818048.2819922>
- Newhart, Veronica; and Mark Warschauer (2016). Virtual inclusion via telepresence robots in the classroom: An exploratory case study. *The International Journal of Technologies in Learning*, vol. 23, no. 4, pp. 9-25. <https://doi.org/10.18848/2327-0144/CGP/v23i04/9-25>
- Nichols, Greg (2022, March 29). The 5 best telepresence robots: Super-charge remote work. ZDNET. <https://www.zdnet.com/article/best-telepresence-robot>
- Niemelä, Marketta; Anne Arvola; and Lina Aaltonen (2017). Monitoring the acceptance of a social service robot in a shopping mall: First results. *HRI'17: Proceedings of the Companion*



## Conversational Fluency and Attitudes Towards Robot Pilots...

- of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6 March - 9 March 2017. New York: ACM Press, pp. 225–226. <https://doi.org/10.1145/3029798.3038333>.
- Novick, David; Brian Hansen; and Karen Ward (1996). Coordinating turn-taking with gaze. *ICSLP'96: Proceedings of the International Conference on Spoken Language Processing, Philadelphia, PA, 3 October - 6 October 1996*. New Jersey: IEEE Press, pp. 1888–1891. <https://doi.org/10.1109/ICSLP.1996.608001>
- O'Conaill, Brid; Steve Whittaker; and Sylvia Wilbur (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction*, vol. 8, no. 4, pp. 389–428. [https://doi.org/10.1207/s15327051hci0804\\_4](https://doi.org/10.1207/s15327051hci0804_4)
- Oviedo, Vanessa Y.; and Jean E. Fox Tree (2021). Meeting by text or video-chat: Effects on confidence and performance. *Computers in Human Behavior Reports*, vol. 3, pp. 100054. <https://doi.org/10.1016/j.chbr.2021.100054>
- Paulos, Eric; and John Canny (1998). PRoP: Personal Roving Presence. In *CHI98: Proceedings of the ACM SIGCHI conference on Human Factors in Computing Systems, Los Angeles, CA, 18 April - 23 April 1998*. New York: ACM Press, pp. 296–303. <https://doi.org/10.1145/274644.274686>
- Provine, Robert R. (1993). Laughter punctuates speech: Linguistic, social and gender contexts of laughter. *Ethology*, vol. 95, no. 4, pp. 291–98. <https://doi.org/10.1111/j.1439-0310.1993.tb00478.x>
- Provine, Robert R.; and Kenneth R. Fischer (1989). Laughing, smiling, and talking: Relation to sleeping and social context in humans. *Ethology*, vol. 83, pp. 295–305. <https://doi.org/10.1111/j.1439-0310.1989.tb00536.x>
- Reid, Alex (1977). Comparing the telephone with face-to-face interaction. In I. Pool (ed): *The social impact of the telephone*. Cambridge, MA: MIT Press, pp. 386–414.
- Rutter, Derek R. (1984). *Looking and Seeing: The Role of Visual Communication in Social Interaction*. Chichester, UK: Wiley.
- Sellen, Abigail (1995). Remote conversations: The effects of mediating talk with technology. *Journal of Human Computer Interaction*, vol. 10, no. 4, pp. 401–441. [https://doi.org/10.1207/s15327051hci1004\\_2](https://doi.org/10.1207/s15327051hci1004_2)
- Short, John; Williams, Ederyn; and Bruce Christie (1976). *The Social Psychology of Telecommunications*. London, UK: Wiley.
- Stivers, Tanya (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, vol. 41, no. 1, pp. 31–57. <https://doi.org/10.1080/08351810701691123>
- Storck, John; and Lee Sproull (1995). Through the glass darkly: What do people learn in videoconferences? *Human Communication Research*, vol. 22, no. 2, pp. 197–219. <https://doi.org/10.1111/j.1468-2958.1995.tb00366.x>
- Stubbe, Maria; and Janet Holmes (1995). You know, eh and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication*, vol. 15, no. 1, pp. 63–88. [https://doi.org/10.1016/0271-5309\(94\)00016-6](https://doi.org/10.1016/0271-5309(94)00016-6)
- Takayama, Leila; and Janet Go (2012). Mixing metaphors in mobile remote presence. *CSCW'12: Proceedings of Computer Supported Cooperative Work, Seattle, WA, 11 February - 15 February 2012*. New York: ACM Press, pp. 495–504. <https://doi.org/10.1145/2145204.2145281>
- Tanaka, Kazuaki; Nakanishi, Hideyuki; and Hiroshi Ishiguro (2014). Comparing video, avatar, and robot mediated communication: Pros and cons of embodiment. *CollabTech 2014: Communications in Computer and Information Science*, vol. 460, pp. 96–110. [https://doi.org/10.1007/978-3-662-44651-5\\_9](https://doi.org/10.1007/978-3-662-44651-5_9)

- Torrey, Cristen; Susan R. Fussell; and Sara Kiesler (2013). How a robot should give advice. *HRI 2013: Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction, Tokyo, Japan, 4 March - 6 March 2013*. New York: ACM Press, pp. 275–282. <https://doi.org/10.1109/HRI.2013.6483599>
- Ullman, Daniel; Iolanda Leite; Jonathan Phillips; Julia Kim-Cohen; and Brian Scassellati (2014). Smart human, smarter robot: How cheating affects perceptions of social agency. *CogSci 2014: Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 36, no. 36, pp. 2996–3001.
- Weibel, Mette; Martin Kaj Fridh Nielsen; Martha Krogh Topperzer; Nanna Maria Hammer; Sarah Wagn Møller; Kjeld Schmiegelow; and Hanne Bækgaard Larsen (2020). Back to school with telepresence robot technology: A qualitative pilot study about how telepresence robots help school-aged children and adolescents with cancer to remain socially and academically connected with their school classes during treatment. *Nursing Open*, vol. 7, no 4, pp. 988–997.
- Whittaker, Steve (1995). Rethinking video as a technology for interpersonal communications: Theory and design implications. *International Journal of Human-Computer Studies (IJHC)*, vol. 42, no. 5, pp. 501–529. <https://doi.org/10.1006/ijhc.1995.1022>
- Whittaker, Steve; and O’Conaill, Brid (1997). The role of vision in face-to-face and video-mediated communication. In K. Finn, A Sellen, and S. Wilbur (eds): *Video-Mediated Communication*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, pp. 23 - 49
- Yang, Lillian; and Carman Neustaedter (2018, November). Our house: Living long distance with a telepresence robot. *CSCW’18: Proceedings of the ACM on Human-Computer Interaction*, vol 2, pp. 190:1–190:18. <https://doi.org/10.1145/3274459>
- Yang, Lillian; Carman Neustaedter; and Thecla Schiphorst (2017, May). Communicating through a telepresence robot: A study of long distance relationships. *CHI’17: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, Colorado, 6 May - 11 May 2017*. New York: ACM Press, pp. 3027–3033. <https://doi.org/10.1145/3027063.3053240>
- Yang, Lillian; Jones, Brennan; Neustaedter, Carman; and Singhal, Samarth (2018, November). Shopping over distance through a telepresence robot. *CSCW’18: Proceedings of the ACM Conference on Human-Computer Interaction*, vol. 2, pp. 191:1–191:18. <https://doi.org/10.1145/3274460>