Improving Non-Native Speakers' Participation with an Automatic Agent in Multilingual Groups

XIAOYAN LI[1], Cornell University, USA & NTT, Japan
NAOMI YAMASHITA, NTT, Japan
WEN DUAN, Cornell University, USA
YOSHINARI SHIRAI, NTT, Japan
SUSAN R. FUSSELL, Cornell University, USA

Non-native speakers (NNS) often face challenges gaining the speaking floor in conversations with native speakers (NS) of a common language. To help NNS to contribute more, we developed a conversational agent that opens up the speaking floor either automatically, after NS have taken a certain number of consecutive speaking turns, or manually, upon NNS request. We compared these automatic and manual agents to a control condition in a laboratory study in which one NNS collaborated with two NS using English as a common language. Participants (N=48) communicated over video conferencing from separate locations in a research institution to collaborate on three survival tasks. Based on data gathered from the experiments, the automatic agent encouraged NNS to participate more, which previous studies had attempted but failed to achieve. Excerpts from group discussions further showed the crucial role of the automatic agent on NNS participation. Interview results suggested that while NNS appreciated the automatic agent's help to participation, NS perceived the agent's interruption as unfair because they thought all members were speaking equally, which was not the case. The mismatch in their perceptions further emphasizes the need to intervene, and we provide design implications based on the results.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)** → Empirical studies in HCI; • **Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods** → Laboratory experiments

Additional Key Words and Phrases: Multilingual Communication; Participation; Conversational floor management; Conversation agent

ACM Reference format:
Xiaoyan Li, Naomi Yamashita, Wen Duan, Yoshinari Shirai, and Susan R. Fussell. 2022. Improving Non-Native Speakers' Participation with an Automatic Agent in Multilingual Groups. *Proc. ACM Hum.-Comput. Interact,* 7, GROUP, Article 12 (January 2023), 28 pages, https://doi.org/10.1145/3567562.

1 INTRODUCTION

Multilingual workplaces, facilitated by an increasingly interconnected world, thrive based on the premise of communication transcending national, cultural and linguistic borders. With the aim of building an effective channel for communication, a common language such as English is usually endorsed within these workplaces. However, the use of a common language can unintentionally create communication issues between native speakers (NS) and non-native speakers (NNS) of that

2573-0142/2023/01– Article #12 $15.00

https://doi.org/10.1145/3567562

language [41][65][94]. Less proficient speakers often fail to actively participate in conversations due to limited language resources, losing communication opportunities that can lead to valuable social resources [1][64]. In previous studies, NNS expressed concern about being left behind during fast paced exchanges using the common language [41][94]. Instead of fighting for the conversational floor, NNS reported feeling lost due to high cognitive load and problems finding open gaps to join the conversation [29][88][89]. Studies also reported that NS were not fully aware of the significant language barriers faced by NNS and how they affected NNS' communication behaviors (e.g., lower level of participation) [42]. At the same time, lack of effective information sharing between NS and NNS can induce negative emotions and harm organizational development, which builds its success upon expertise from heterogeneous backgrounds [69][70][76].

In response to these problems, researchers have explored providing language support for NNS via a variety of technologies, including machine translation (e.g., [30][85][87]), automated transcripts (e.g., [12][28]) and keyword highlighting (e.g., [16][27]), most of which have focused on improving NNS comprehension rather than motivating them to speak up. In addition, these tools have the potential to raise NS' awareness of the impact that language difficulties have on NNS' communication behaviors. However, even after NS are more aware of the differences in language abilities between themselves and their NNS peers, they were not given clear instructions on how to adjust their communication behaviors accordingly [19][42][82], or provided motivations to do so [21].

To further address these issues, we propose a direct, real-time conversational agent that offers instructions on expected communication behavior adjustment for both NS and NNS discussing over videoconferencing. We focus on floor management, as NNS frequently report difficulty in finding gaps in conversations that would allow them to participate [40][71][94], partially because it is challenging for NNS to project possible "transition-relevant" places at the end of turns [43][72]. More specifically, we developed an automatic agent to monitor the communication process and open up the conversational floor when a preset level of differences in the number of turns taken by NS and NNS is detected (i.e., when two NS take 6 consecutive turns while NNS speak 0 turns). The agent opens the floor by interrupting the conversation and asking someone else to speak. However, if the agent interrupts NS when the discussion is heated or when NNS are not ready to talk, it might elicit negative emotions from NS or impose pressure on NNS instead of facilitating dynamic participation. Therefore, we also developed a manual agent, where NNS simply press a button to enable the agent to ask the same question, without NS being aware who activated the agent. We hypothesize that the agents will help NNS to participate more, reduce group members' workload and improve their perception of collaboration quality.

We explored the effects of the agent on group members' behaviors and perception of the communication experience by conducting a within-subject experiment with 16 groups of three (one NNS and two NS). The participants collaborated on three survival tasks using English as the common language under three conditions: control, automatic agent and manual agent. Based on data gathered from the experiments, the automatic agent encouraged NNS to speak up more: NNS had a significantly higher proportion of words spoken in the automatic agent condition compared to the control condition, which previous studies had attempted but failed to achieve (e.g., [12][28][30][85][87][19][42][82]). The excerpts in Table 1 showed the crucial role of the automatic agent on NNS participation. NNS was invited by NS to speak, shared information that was essential to group performance, and asked important questions about every member's current status. Interview results suggested that while NNS appreciated the automatic agent's help to participation, NS perceived the agent's interruption as unfair because they thought all members were speaking equally, which was

not the case. The mismatch in their perceptions further emphasizes the need to intervene, and we provide design implications based on the results.

2 RELATED WORK AND HYPOTHESES

In this section, we first review literature on challenges faced by multilingual teams and tools developed to address these challenges; we then discuss turn-taking in conversations, followed by an introduction to the current study and hypotheses grounded in literature.

2.1 Challenges of Multilingual Teamwork

Communication tools and social media have enabled people to interact fluidly across national, cultural and linguistic boundaries in ways that would have been difficult if not impossible in the past. In virtual organizations, teams of people from across the globe now work together on common problems, each bringing their own perspective and expertise (e.g., [1] [13][58][60][75]).

Despite the rapid increase of global teams, however, language boundaries remain one of the biggest issues that inhibit their potential. Previous studies highlighted the impact of disparity in linguistic resources on group participation [34][63][83], which, if not properly addressed, could cause asymmetrical information sharing, create language-based subgroup clustering and prevent groups from achieving their potential [1][46][64]. Compared to NS, who generally have sufficient linguistic resources to actively participate in group discussions [22][23][96], NNS often reported having difficulties keeping pace with fast-paced conversations [21][30][41][71]. In addition, Takano et al. [97] reported that producing speech in a second language required substantial cognitive resources, reducing the resources available for thinking about a task. Therefore, NNS can have trouble gaining the conversational floor and clearly expressing their perspectives even if they overcome difficulties in understanding the fast-paced conversations [29][97].

Thus, there may be value to designing tools that oversee the communication process and maximize chances for NNS to take the conversational floor when participation disparities between NS and NNS are too large. Janssens et al. [46] highlighted the need for counterbalancing power disparities caused by cultural and linguistic differences (e.g., by setting rules for clarification and speaking speed) to facilitate more equal contribution to information sharing and group decision-making. However, it is difficult for NS to maintain awareness of the language and power disparities and stick to the set rules, especially when the discussions are task-oriented and time-sensitive. Studies also report that NS tend to overestimate NNS language abilities and underestimate how much language difficulties affect NNS communication behaviors (e.g., lower level of participation) [42]. Thus, technology that could raise NS awareness of NNS' language barriers and give clear instructions on their expected behavior adjustment could be useful. Although completely equal participation is not always possible or even beneficial, balancing perspective sharing when it is too asymmetrical due to linguistic disparities can be advantageous for group performance and organization development [69][70].

2.2 Tools to Support Multilingual Conversation

To address the linguistic challenges of multilingual teams, researchers have developed and tested a variety of tools to aid NNS comprehension, including machine translation [16][30][85][87], automated transcripts [12][28], and keyword highlighting [16][27]. Some of these tools increased NS awareness of differences in language proficiency within multilingual groups. For instance, Gao et al. [28] reported that publicly shared real-time transcripts in a common language improved the perceived quality of communication, in part because NS started speaking more clearly.

Studies have also emphasized differences in communication behaviors by providing visualizations of key elements in group discussions, e.g., number of words spoken, affective features, etc. [41][54].

However, while previous studies explored NS' voluntary behaviors after their awareness of linguistic differences was raised, they typically do not find corresponding behavioral adjustment, i.e., awareness does not necessarily result in the intended communication outcomes. For example, Diamant et al. [19] employed a feedback system that presented the different expression patterns of NNS and NS during communication, only to find that NS did not change their conversational behaviors as a result of the feedback system. Similarly, He et al. [42] found that sharing graphs of NS and NNS communication behaviors (number of words spoken, time smiling, etc.) resulted in little change in subsequent behaviors of NS. Interviews of NS conducted by Duan et al. [21] suggest two reasons why NS fail to adjust: they are not clear how to use the provided information and/or do not feel the need to adjust their behaviors in response to that information. In addition, NNS, along with difficulties in understanding and expressing opinions, report feeling anxious about their language abilities (e.g., consciousness to grammatical mistakes) and thus reluctant to speak up [28][29]. Building on previous work, we explore the effects of direct conversational intervention during group discussions, aimed at achieving a more balanced participation among all members in multilingual groups.

2.3 Turn-Taking in Conversations

Opportunities to speak in conversations are highly dynamic and strategically managed based on moment-by-moment linguistic, behavioral and social cues, e.g., syntax, intonation, gap, pre-turn hearable inbreath, gesturing, eye gaze, and expertise [22][18][24][34][63][96]. A fundamental mechanism guiding speaking opportunities is the organization and distribution of turns. There have been several different lines of work investigating the mechanisms of turn-taking in conversations. For example, Duncan [18] proposed that turn-taking is negotiated through signals (e.g., intonation, body gesture, completion of a grammatical clause, etc.) that people use to indicate their state related to the speaking turn. Three rules were proposed in this model: 1) a speaker could indicate a "turn-yielding" signal (e.g., a clause completion), and an "auditor" could take the next turn in response to such a signal; 2) the auditor could use a backchannel cue to indicate that they do not wish to take the next turn; 3) also, the current speaker could use "suppression of speaking-turn" claims to show that they wish to continue speaking. However, simultaneous turns could happen if the auditor tries to take the next turn without the current speaker giving a turn-yielding signal, or if the current speaker continues to speak after displaying a yielding signal with the auditor attempting to take the next turn. This model takes a psychological perspective and emphasizes that smooth organization of conversations requires collaboration between the speaker and auditor [26][25]: if cues are properly used and responded to, neither of the above situations that lead to simultaneous turns would happen. However, a limitation of this model is the lack of precise descriptions of the pertinent signals/cues [5][17][35]. It is also not clear if people from different cultures use and respond to these cues in the same way [27][28].

Sack et al. [72] proposed a conversational analysis model capturing two turn-allocational techniques: 1) current speaker selecting the next and 2) next speaker self-selecting themselves. When the first technique is employed, the selected speaker has the exclusive right to take the next speaking turn. However, when no one is selected, the current speaker may continue to speak until another speaker self-selects themselves. The transition of turns is based on complicated factors embedded in the situation. Sacks et al. [72] proposed that transfer of speakership happens at "transition-relevant" places, marked by "possible" completion of a syntactic unit, such as a phrase, clause or full sentence. To self-select as the next speaker, one should start speaking at the earliest possible "transition-relevance" place. However, variation in the types of units employed before possible transition-relevance places makes it difficult to accurately time the start of the next turn, making it common for brief overlaps during transition of turns. What is more, Ford et al. [24] argued that syntax alone cannot offer the best projection for completion of turns, e.g., completion of a syntactic utterance could always

be extended by the speaker and must be evaluated in relation with previous context, and therefore should be accompanied by intonation and pragmatics. Even when syntactic, intonational and pragmatic content of utterances are combined to project completion of speakers' utterances, only half of the projected transition-relevance points were reported to actually have a change of speakers [24].

An alternative perspective considers speaking turns in a content-free fashion and defines them as "contiguous intervals" of speech spoken by the same person [26][28], also called "talk spurts" [66]. Shriberg et al. [90] further defined these intervals as "speech regions uninterrupted by pauses longer than 500ms". A turn starts with either the first word in a conversation, or the first word breaking the pause after the previous turn [91]. This framework builds on these simple rules of turn-taking with the aim of providing a straightforward strategy for automatic processing of large amounts of speech data [90]. Due to the complexities embedded in projecting possible turn transition places in the psychological and conversational analysis frameworks, we decided to design the agent based on this last framework, as the emphasis of the paper is to explore the impact of an agent intervention on participation in conversations of multilingual groups. Therefore, we made the decision of having the agent intervene the NS speaker that took the 6th consecutive speaking turn after the first 0.5s acoustic silence was detected.

We also draw from the Sacks et al. [72] framework. When turn taking violations occur, communicators usually employ repair mechanisms by exhibiting verbal or non-verbal signs while interrupting others (e.g., excuse me, leaning forward) [72]; we designed the agent to show signs of interruption by blinking and waving hand when interrupting others. What is more, we decided to use three-member groups considering the nature of turn-taking proposed by this framework. In two-member groups, any non-speaker (e.g., a NNS speaker) may give up opportunities to self-select as next speakers but still be guaranteed to become the next speaker at some point [72]. When there are three or more members, speakers would have to compete for the next turn [95], which suggests that NNS speakers in multilingual groups may be at disadvantage. Previous work have reported how discussions with NS as majority and NNS as minority could be easily dominated by NS, as it is easier for NS to take the next turn following each other's speech, creating long streams of talk where it is difficult for NNS to find a gap to join [21][30][41]. In this study, we decided to have two NS and one NNS in the three-member groups because it is the smallest unit that could replicate this situation.

2.4 The Current Study

In this study, we developed two versions of a conversational agent that opens up the floor during communication in multilingual groups, creating opportunities for NNS to speak: an automatic agent and a manual agent. Both agents interrupted the group discussions by saying "That's Interesting. I was wondering, can we hear from someone else?" The interruption provides implicit instructions on expected NS and NNS behavior adjustment: it reminds NS to pass the conversational floor and encourages NNS to speak more. In combination, these two adjustments have the potential to lead to more balanced group participation.

In the *automatic agent condition,* the agent monitored and interrupted the group discussion when the two NS take six consecutive speaking turns (see 0 for details on rules of interruption). The agent works independently and thus serves as a separate identity from the participants. The interruption rule is based on the assumption that NNS would likely have something to say when they have not talked for a long time. However, the automatic agent can interrupt the conversation at unnecessary places when NNS have nothing to say. Furthermore, the agent interruption could come off as annoying to NS, if they are interrupted at a timing when the discussion is heated and/or when someone is in the middle of making an important point.

In the *manual agent condition,* NNS could activate the agent to interrupt the discussion by pressing a button without being noticed by the NS. Thus, the manual agent is only activated when NNS have something to say but want the agent's help gaining the floor. For example, they might have lost the opportunity to participate in a fast-paced discussion while struggling to find the right timing to interrupt [89], or they might hesitate to interrupt the conversation to ask a clarification question caused by their limited language ability [21]. Therefore, the manual agent works as an assistant for NNS, helping them pause the discussions without having to worry about interrupting at the right moment (as the manual agent will automatically interrupt after NS' next utterance when activated). By making use of the manual agent, we anticipate that NNS could gain extra time in preparing their speech, reduce situations where they fail to find a gap to speak during fast-paced conversations, and relieve them from having to search for the right timing to interrupt. However, the manual agent demands active actions from the NNS, i.e., they need to make a decision and actually press the button to enable the agent. As mentioned earlier, NNS might not be willing to speak up due to a lower confidence of their language proficiency and therefore might not press the button. Therefore, we expect that there will be trade offs between the two agents. In terms of NNS participation, we hypothesize that,

H1a: NNS will participate more in the automatic agent condition than in the control condition.
H1b: NNS will participate more in the manual agent condition than in the control condition.

As discussed earlier, NNS reported having difficulty finding gaps to interrupt and express their opinions, even when they feel their perspective is missing in the group discussion [89][97]. We expect that both agents will help to free NNS from the burden of having to find available gaps to join the conversation. They will then have more bandwidth to focus on understanding their partners and expressing themselves, to better collaborate with other members. Therefore, we expect that the agents will reduce the workload experienced by NNS. Increased NNS ability to collaborate also influences the entire team. Previous study reported that a higher level of participation rate leads to more information sharing [20][77]. By improving NNS' ability to participate and voice their opinions, critical and diverse information could be more effectively shared within the group, which is essential for making informed and creative group decisions [20][76][78][81]. Therefore, we believe that both NS and NNS will have a higher perception of collaboration quality in the agent conditions. At the same time, we expect NS to rate their workload lower when NNS share more perspectives and contribute more to the group discussions, potentially reducing NS burden in completing the group tasks. We hypothesize that,

H2: Both NS and NNS will rate their workload lower in the agent conditions than in the control condition.
H3: Both NS and NNS will rate their collaboration quality higher in the agent conditions than in the control condition.

As both agents, regardless of their type, will remind NS to pass the conversational floor to NNS, we think NNS may perceive their NS peers as more accommodating in the agent conditions. Accommodation here refers to the efforts that communicators make to adjust their behaviors to accommodate other communicators (e.g., by speaking slower or providing gaps to speak) [82]. As NS will likely adjust their behaviors (give up conversational floors) to accommodate NNS as a result of the agent's interruptions, we hypothesize that,

H4: NNS will rate NS as more accommodating in the two agent conditions than in the control condition.

Finally, given the trade offs of the two agents, we are interested in finding out how the two agents will be perceived by the participants. As NS are compelled to pass the conversational floor to their NNS counterparts, we are interested in learning how NS and NNS perspectives on the two agents might differ.

RQ1: How will NNS and NS perceive the agents?

3 METHOD

3.1 Overview

In this study, we employed a within-subjects experimental design to explore the effects of an automatic and a manual agent that opened up the conversational floor on participation and group collaboration in multilingual groups. Each of the groups included two native English speakers and one native Japanese speaker who collaborated on three survival tasks in English. Each group performed one task in each of three conditions: control, automatic agent and manual agent. The sequence of the three survival tasks and three conditions were counterbalanced. We used chat logs generated from IBM real-time speech recognition software to measure participation, and post-task surveys and interviews to investigate the agent's impact on participants' communication experience and participants' perception of the agent.

3.2 Participants

We recruited 48 participants, divided into 16 groups of three (one non-native speaker and two native speakers of English). We recruited 32 native English speakers (13 female, 18 male and 1 gender non-conforming) through a third-party personnel company, and 16 Japanese speakers (2 female, 14 male) by posting an advertisement on a bulletin board in a large Japanese university. The native English speakers were raised and educated in an English-speaking country, including the USA, UK, Australia, etc. Among the 32 native English speakers, 11 had lived in Japan for 3-10 years and 19 had lived in Japan for more than 10 years. Most (13 of the 16) Japanese speakers had lived in an English-speaking country for less than 1 year. Overall, Japanese participants reported their proficiency in English speaking as lower than medium (M=3.13 on Likert scales from 1 (not fluent at all) to 7 (very fluent)). The mean age of the participants was 37.4 (M=44.4 for NS; M=23.3 for NNS). Participants were randomly assigned to one of the 16 groups. They were compensated 10K Japanese yen (around $92 USD) for participating in the study.

3.3 Materials

Survival Tasks. Each group completed three survival tasks and one practice task. The survival tasks were modified versions of the desert, ocean and arctic survival tasks [42][51]; the practice task was a modified version of the lunar survival task. The purpose of these survival tasks is to rank items (e.g., a magnetic compass, a small ax, newspapers, etc.) salvaged from a crashed plane/spaceflight or a sinking yacht based on their importance to survival in different scenarios. Different from the original tasks, which included 15 salvaged items, we only listed six items for each task and required participants to choose three out of the six items and rank them based on their importance to survival (for the practice task, participants chose two out of five items). We made the list of items shorter to accommodate the limited discussion time.

Surveys. Before the experiment started, participants filled out an online demographic survey that collected information on gender, age, profession, years lived in Japan/English speaking countries, etc.

After completing each of the survival tasks, they were asked to complete an online post-task survey that included questions on workload, collaboration quality, and perception of the agent.

Interviews. After all tasks were completed, we interviewed each participant separately on their communication experience and perception of the agent, including the timing and content of the agent's interjections and how it shaped their conversational experiences. Sample questions included: "How do you compare your experience across the three discussions?", "How do you think the agent affected the conversations?", "Was there any change in behavior after the agent interrupted? Why or why not?".

Consent form. All participants read and signed a consent form before taking part in the study. The consent form explained that participation in the study was voluntary, described the purpose and duration of the study, and explained the confidentiality of the data. This research was reviewed and approved by an institutional review board.

3.4 Equipment and Software

Interface. For all the group discussions, participants communicated over video conferencing. As shown in Figure 1, for the automatic and manual agent conditions, the agent's image was also present in the video conferencing interface, in addition to the three participants. The agent had the appearance of a female cartoon avatar and an artificial female voice since previous studies found that users usually prefer female voices in social robots [11][15]. The agent showed signs of interruption by blinking and waving hand when interrupting others, because studies show that when turn taking violations occur, communicators usually employ repair mechanisms by exhibiting verbal or non-verbal signs while interrupting others (e.g., excuse me, leaning forward) [72]. In the control condition, the agent was absent from the group discussions.

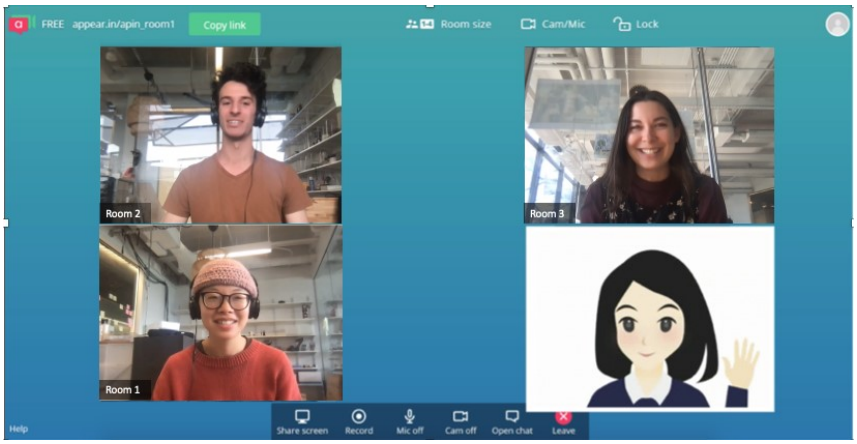

Figure 1: Video Conferencing Interface in Automatic and Manual Agent Conditions

Mechanism for Agent Interruption. IBM Watson ASR (automatic speech recognition) was embedded in the automatic agent to detect the number of turns taken by each group member [21][40]. Participants used different microphones for voice input, which enabled the system to distinguish between speakers. The system detected the start of a turn based on audio input of a specific speaker and the end of a turn based on audio input of another speaker [24][45]. If a speaker's speech was five words or less, it was not recognized as a turn, because short utterances of this type are often used to

provide backchannel responses such as agreeing, showing interest, or exhibiting understanding without breaking the current turn [24][86][93]. Our system was unable to distinguish these backchannels from very short turns (e.g., asking a short question). Therefore, the system only considered a turn to be finished when another speaker took the conversation floor and spoke more than five words.

In the automatic agent condition, when five consecutive turns were taken by the two NS, the system decided whether or not to interrupt the next speaker based on the following mechanism: if the sixth turn was taken by the NNS, it did not interrupt; if it was taken by one of the two NS, then it interrupted the speaker after the next utterance, defined as a 0.5 second acoustic silence gap within a person's speech [21][40]. However, the agent did not interrupt the same NS more than three times, to prevent situations where the same NS are interrupted by the agent too frequently. In the manual agent condition, the NNS could press a button to enable the agent to interrupt the conversation when it detects the end of the next utterance.

3.5 Procedure

Upon arrival, participants were escorted into separate locations where they could not see each other. After signing consent forms and completing the demographic survey, they were asked to put on the headsets and listen to instructions given by the experimenter. Each group collaborated on a practice task (the lunar survival task) before working on the three main survival tasks (the desert, ocean and arctic survival tasks), to get familiarized with each other, the task procedures and the video conferencing system. Each of the survival tasks included three steps. First, group members were asked to take off their headsets and work independently for 5 minutes; after having written down their independent rankings, they had 15 minutes to discuss within the group and generate a group ranking; finally, they again wrote down their individual rankings. During group discussions, they were encouraged to persuade each other through good reasoning, professional knowledge or personal experience. After each group discussion, they had approximately 15 minutes to complete an online survey. At the end of each experiment, we interviewed the participants about their communication experiences and perceptions of the agent. All participants were debriefed after the study. Each experiment lasted about 2.5 hours.

Before the automatic and manual agent condition, the experimenter introduced the agent to all participants and pointed out that the agent might interrupt their discussions with some requests. NS were not informed about the difference between the automatic and the manual agent, to preserve NNS desires for saving face. However, NNS were told they could press a button to enable the manual agent; they practiced pressing the button before the manual agent condition, so that they were aware what would happen when they press the button and how the manual agent would interrupt the conversations.

All instructions and task materials were given in English; however, the second author (who is a native Japanese speaker) explained the task procedures and materials for the Japanese participants in Japanese before the experiment started. We provided Japanese translations for questions/words in the English surveys that might be difficult for Japanese speakers to understand. They also had the chance to ask any language-related questions in Japanese while they were doing the independent ranking and completing the surveys. The post-task interviews of Japanese speakers were also conducted in Japanese.

3.6 Measures

We collected three types of data:1) data generated from real-time speech recognition, i.e., number of words each participant spoke, based on which we calculated their proportion of words, 2) participants' ratings collected from the surveys and 3) participants' insights from the interviews (see 4.3 for details on interviews).

3.6.1 Participation. We measured both the number of turns and the number of words of all participants generated from IBM Watson automatic speech recognition software. Participants used different microphones for voice input, which enabled the system to distinguish between different speakers. Since our goal is to answer whether NNS could speak more in the agent conditions, we think proportion of turns and proportion of words could complement each other in capturing what we are exploring. i.e., how much someone actually spoke up in the group. Therefore, we calculated the proportion of turns and proportion of words as measurements of participation:

Proportion of turns = individual speaker's numbers of turns/group total number of turns
Proportion of words = individual speaker's numbers of words/group total number of words

3.6.2 Surveys. We collected participants' subjective ratings on workload, quality of collaboration, NNS perception of NS accommodation and perception of the agent from the surveys.

Workload. We used the NASA Task Load Index to measure participants' perceived workload (mental demand, temporal demand, effort, frustration, performance) on 7-point Likert scales (1=low, 7=high) [39]. The last item "performance" did not load on the same factor with the rest and was removed from the scale. The remaining four items formed a reliable scale (Cronbach's $\alpha = .82$). They were averaged to create a measure of "Workload".

Quality of Collaboration. We measured participants' subjective perception of collaboration quality with the following questions on 7-point Likert scales (1= strongly disagree, 7= strongly agree): "We could fully understand each other's thoughts", "Generally, I'm satisfied with our collaboration on this task", "I like the final group ranking", "I felt everyone's thoughts were incorporated into the final group ranking", and "Everyone had a chance to fully express their thoughts during the discussion"[16]. The questions formed a reliable scale (Cronbach's $\alpha = .90$) and were averaged to create a measure of "Quality of Collaboration".

Perception of NS Accommodation. We created a measure of NNS perception of NS accommodating behaviors with five questions. They loaded on two factors and were divided into two reliable subscales: "NS accommodating (Cronbach's $\alpha = .87$)" and "NS encouraging (Cronbach's $\alpha = .70$)". All questions were measured on 7-point Likert scales (1= strongly disagree, 7= strongly agree). "NS accommodating" included three questions: "My partners tried to speak clearly enough for me to understand", "My partners tried to speak slowly enough for me to understand", and "My partners tried to use easier words so that I can understand the conversation better". "NS encouraging" had two questions: "My partners were trying to encourage me to join the conversations" and "My partners dominated the conversation and did not give me a chance to join the conversation" (reverse coded).

Perception of Agent. Participants rated their perception of the agent on 7-point Likert scales (1= strongly disagree, 7= strongly agree). The questions included: "The agent's behavior was appropriate", "The agent's requests made sense", "The timing of the agent's requests was appropriate", "The agent interrupted my thoughts at times", "The agent helped everyone to participate in the conversation better", "I was annoyed by the agent at times", etc. They formed a reliable scale (Cronbach's $\alpha = .86$) and were averaged to create a measure of "Perception of Agent" (some questions were reverse coded to get the overall score on "perception of agent", so that 1 means the most negative and 7 the most positive perception).

4 RESULTS

Across 16 groups, the average discussion time for each of the three conditions was 14.93 mins (SD=2.23). Overall, the automatic agent interrupted 3-5 times (Mean=3.75, SD=.12) in each discussion. Among a total number of 59 agent interruptions, NNS took the floor 25 times, either voluntarily or after being reminded by the native speakers. NNS spoke at least once after agent interruptions in 14 of the 16 groups. Unexpectedly, the manual agent was only activated 3 times across 16 groups. That is, only 3 out of 16 NNS pressed the button once in the manual agent condition, which we explored further in the interviews. In this section, we present findings regarding our hypotheses and research questions. We first examine whether the agent led to behavior changes, measured by NNS level of participation (4.1). Following that, we report participants' perception of the communication experience (4.2), including workload, collaboration quality, NS accommodation and perception of the agent. Finally, we present insights from our interviews (4.3).

4.1 Effects of the Agent on NNS Participation

Hypotheses 1 predicted that NNS' level of participation (measured by proportion of turns and proportion of words) would be higher in the agent conditions than in the control condition. To test these hypotheses, we ran mixed model ANOVAs with proportion of turns and proportion of words as the dependent variables, and condition and trial order as fixed factors. Preliminary analyses showed no effect of task ($F [2, 29.54]=.33$, $p=.72$; $F [2, 29.54]=1.46$, $p=.25$), so we excluded it from the model. NNS English proficiency did not affect participation ($F [1, 14.62]=2.70$, $p=.12$; $F [1, 14.95]=1.36$, $p=.26$), so we also excluded it from the analysis.

The results showed no significant main effect of condition on proportion of turns ($F [2, 27.47] = 1.19$, $p = .32$), no main effect of trial order ($F [2, 23.98]=.75$, $p=.49$), and no interaction between condition and trial order ($F [4, 34.19] = .31$, $p=.87$). Overall, NNS proportion of turns was highest in the automatic agent condition ($M = .17$, $SD = .07$) and lowest in the control condition ($M = .14$, $SD = .08$), with proportion of turns in the manual agent condition in between ($M = .16$, $SD = .05$). Pairwise comparisons showed there was no significant difference in NNS' proportion of turns between the automatic agent and the control condition ($p= .15$), or between the manual agent and the control condition ($p=.27$). There was also no significant difference in NNS' proportion of turns between the automatic and the manual agent condition ($p=.72$).

The results showed a significant main effect of condition on proportion of words ($F [2, 28.23] = 5.35$, $p = .01$), no main effect of trial order ($F [2, 24.63]=.14$, $p=.87$), and no interaction between condition and trial order ($F [4, 32.67] = 1.08$, $p=.38$). Overall, NNS proportion of words was highest in the automatic agent condition ($M = .17$, $SD = .07$) and lowest in the control condition ($M = .12$, $SD = .06$), with proportion of words in the manual agent condition in between ($M = .14$, $SD = .05$) (see Figure 2). Pairwise comparisons showed that H1a was supported: NNS' proportion of words in the automatic condition was significantly higher than the control condition ($p=.003$), but there was no significant difference in NNS' proportion of words between the manual agent and the control condition ($p=.11$), showing that H1b was not supported. There was also no significant difference in NNS' proportion of words between the automatic and the manual agent condition ($p=.11$).

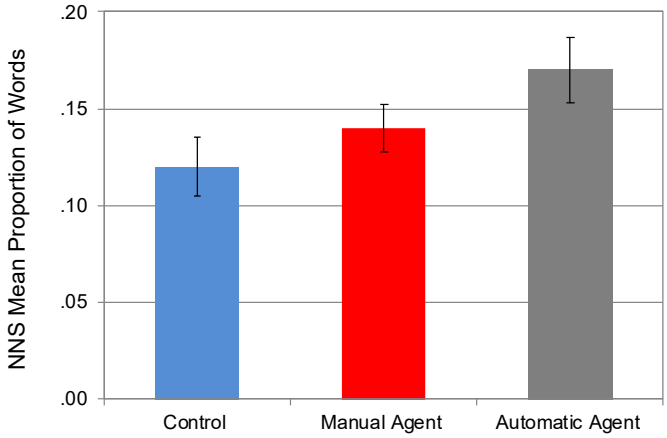

Figure 2: NNS' Mean Proportion of Words in the Control, Manual Agent, and Automatic Agent Conditions. Error bars represent the standard errors of the mean.

4.2 Perception of the Communication Experience

To test whether or not the agent affected participants' perception of the communication experience, we conducted mixed model ANOVAs with workload and collaboration quality as dependent variables and speaker type (NS or NNS), condition (control, automatic agent or manual agent condition), trial order (1,2,3) and task (desert, ocean or arctic survival task) as fixed factors.

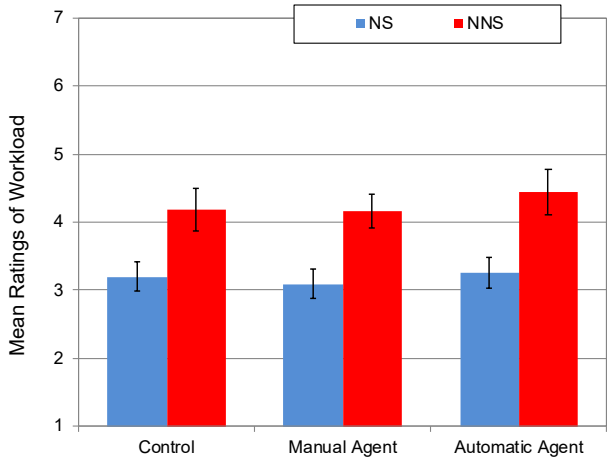

Figure 3: Mean ratings of workload on a scale of 1 (low) to 7 (high) by speaker type and agent condition. Error bars represent the standard errors of the mean.

4.2.1 Workload. Hypothesis 2 predicted that both NS and NNS workload will be lower in the two agent conditions than in the control condition. The results showed that speaker type ($F [1, 39.49] = 10.98$, $p = .002$) had a significant effect on workload, where NNS ($M=4.27$, $SD= 1.18$) had a higher workload than NS ($M=3.18$, $SD=1.22$). However, there was no significant main effect of condition ($F [2, 69.73] = .26$, $p = .77$), and no interaction of speaker type and condition ($F [2, 69.73] = .56$, $p = .57$), indicating that H2 was not supported. We also found a significant main effect of task ($F [2, 69.73] = 4.11$, $p = .02$). Pairwise comparisons showed that workload in the ocean survival task was significantly

lower than the arctic survival task (p=.004) and the desert survival task (p=.06). There was a borderline interaction of speaker type & trial order & task (F [4, 83.68] = 2.17, p = .08) and borderline interaction of condition & task (F [4, 79.73] = 2.06, p = .10).

4.2.2 Collaboration Quality. Hypothesis 3 predicted that both NNS and NS will have a higher perception of the collaboration quality in the two agent conditions. Results (see Figure 4) indicated there was a significant effect of speaker type (F [1, 42.07] = 4.89, p = .03), such that NS (M=6.13, SD=.74) had a higher perception of collaboration quality than NNS (M=5.64, SD=.95). We also found a significant main effect of condition (F [2, 72.98] =3.22, p=.05), but no significant interaction of speaker type & condition (F [2, 72.98]=1.39, p=.26). Overall, participants rated the collaboration quality highest in the automatic agent condition (M=6.06, SD=.73) and lowest in the manual agent condition (M=5.82, SD=.93), with the control condition rated slightly less than the automatic condition (M=6.01, SD=.86). Pairwise comparisons showed that collaboration quality in the automatic agent condition was significantly higher than the manual agent condition (p=.03), but there was no difference between the automatic agent condition and the control condition (p=.44) or between the manual agent condition and the control condition (p=.15). This means that neither the automatic nor the manual agent increased participants' perception of collaboration quality; therefore, H3 was not supported.

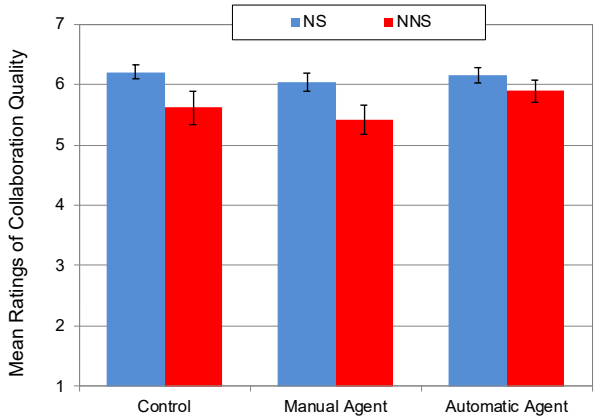

Figure 4: Mean ratings of collaboration quality on a scale of 1 (low) to 7 (high) by speaker type and agent condition. Error bars represent the standard errors of the mean.

4.2.3 NS Accommodation. Hypothesis 4 predicted that NNS will perceive NS as more accommodating in the two agent conditions. After conducting mixed model ANOVAs with NS accommodation (including NS accommodating and NS encouraging) as dependent variables, and condition, trial order and task as fixed factors, we found no significant main effect of condition on NS accommodating (F [2, 15.65] = .77, p = .48) or NS encouraging (F [2, 14.21] = 1.41, p = .28). Therefore, H4 was also not supported. This is an interesting phenomenon that we will return to in the discussion, as NNS mentioned in the interviews that NS often invited them to speak when the agent interrupted.

4.2.4 Perception of the Agent. RQ1 asked how NS and NNS will perceive the agent. Since only 3 out of the 16 groups pressed the button to enable the agent in the manual agent condition, we focus on results from the automatic agent condition here. More specifically, we present descriptive statistics of NS and NNS perceptions of the automatic agent and compare them to see how their perceptions differ.

Overall, NNS mean perceptions of the agent (M=3.81, SD=1.19) are higher than NS perceptions (M=3.06, SD=1.07). NNS had more positive perceptions than NS of agent interruption and the

annoyance caused by the agent. To explain, NNS perception was more positive on "The agent interrupted my thoughts at times" (M=4.75, SD=1.63 for NS; M=3.44, SD=1.90 for NNS)" and "I was annoyed by the agent at times" (M=4.41, SD=1.93 for NS; M=3.13, SD=1.63 for NNS) compared with NS perception. However, both NS and NNS both rated agent timing negatively: "The timing of the agent's requests was appropriate" (M=3.19, SD=1.89 for NS; M=3.06, SD=1.34 for NNS). This means that both NS and NNS thought the automatic agent interrupted with inappropriate timing.

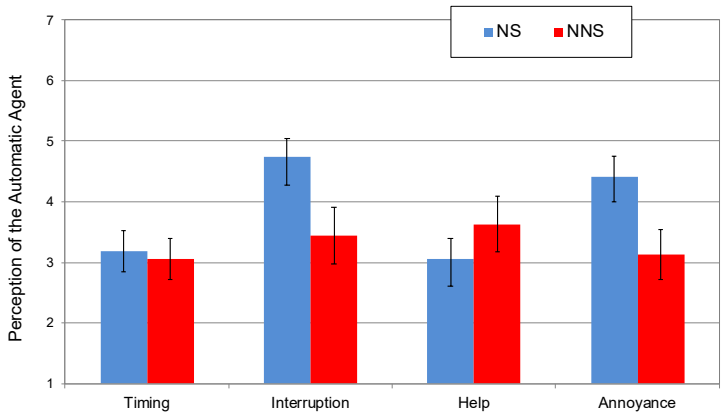

Figure 5: NS and NNS' mean ratings of the Automatic Agent on a scale of 1 (low) to 7 (high). Error bars represent the standard errors of the mean.

4.3 Insights from Interviews

During interviews with participants, we asked questions about their communication experience, with a specific focus on their perception of the automatic and manual agents. Each interview lasted for about half an hour. The interviews were video recorded, transcribed, and the Japanese interviews were then translated by a professional translation company. The transcripts were then coded iteratively using a grounded theory approach [32]. First, we used ATLAS.ti [99] to create initial codes, assigning them to significant instances and quotations. Three researchers engaged in this process independently before comparing the codes together. Following that, we identified emerging concepts and categories, critically examined and continuously rearranged them to explore shared patterns and establish boundaries of higher-level themes and identify relationships between the themes. To get a more in-depth understanding of the themes, we refer back to excerpts of the group discussions based on interesting quotations from the interviews, and present short transcripts to show the agent at work during group discussion.

In this section, we will first report participants' reactions to the automatic agent, by discussing three reoccurring themes from the interviews. During this process, we also present excerpts of conversations happening right before and after two agent interruptions (Table 1), to provide more contextual information of the reoccurring themes. Following that, we briefly discuss NNS' perception of the manual agent.

4.3.1 Insights about the Automatic Agent. In this part, we report three reoccurring themes from the interviews: 1) NNS perceived the automatic agent as helpful in opening up conversational floors and motivating them to contribute to the group discussions; 2) NS perceived the discussion as already "equal" and saw the agent as unfair and unnecessary, with NNS offering different perceptions; 3) and

both NS and NNS showed concerns about the agent's timing, which was sometimes inappropriate, although a minority of NNS presented an alternative argument. We present two excerpts showing the agent at work in one of the group discussions during this process (Table 1).

The Agent Helped NNS to Speak. NNS (10 out of 16) reported that the automatic agent motivated them to speak, either autonomously or after being invited by NS group members. We describe their perceptions in detail below.

NNS reported that the automatic agent motivated them to speak more, simply by reminding them that they weren't speaking enough.

> I remember that once, when the agent broke in and said that it wanted to hear someone else's opinion, we discussed who was not speaking in the discussion. Apparently, I wasn't saying anything and was encouraged to say something. In this situation, I think it helped me speak up (Participant 4C, Japanese).

NNS reflected on how the agent made it easier for them to talk by helping them to stop worrying about grammar mistakes, when they are put in a situation where they had to speak up. As previous studies have shown, NNS sometimes do not speak up due to lack of confidence in their English and concerns about making grammatical mistakes [57]; the comments below showed that when NNS "suddenly have to talk", they might be able to "stop worrying about" using "correct English".

> So, to some degree, before I start talking, I worry about whether or not what I'm going to say is correct English, but if I suddenly have to talk, then I stop worrying about that sort of thing, and I'm able to focus on how much I can get across to the others, so it's like, it would be better to just be directed to talk (Participant 13C, Japanese).

NNS participants frequently emphasized how their NS partners often invited them to speak after the agent's interruption (Table 1, line 3 and line 17 also showed situations like this).

> During the discussion, I was mainly a listener of the conversation. The agent reset the flow of the discussion. And then [my NS partners] asked me "What do you think?" and gave me the chance to speak. In that sense it became easier to talk, because it gave me the opportunity to speak. I was grateful (Participant 7C, Japanese).

NNS appreciated when the agent "cut off" the conversations and gave them opportunities to catch up and ask questions they were "unsure about".

> Yeah. It's like things are cut off temporarily. But if I can't understand the conversation, it's helpful for me if things are cut off (Participant 12C, Japanese).

> The other 2 people turned to listen to me when the agent entered, I felt I should say something… The discussion went by quickly, I mean how fast the others debated. … (After the agent interrupted) I confirmed with them what they chose for their 1st and 2nd choice… Because the discussion stopped, I was able to ask about things I was unsure about (Participant 16C, Japanese).

In the quote above, participant NNS_16C talked about two specific situations: 1) he felt he "should say something" when the automatic agent interrupted, and 2) he was able to ask about things he was "unsure about", i.e., what NS "chose for their 1st and 2nd choice". Table 1 shows two excerpts from the group discussion that participant 16C was referring to. More specifically, it shows parts right before and after the first and the third agent interruption (labeled as Int1 and Int3 respectively).

Excerpt 1

1 16B: I put transistor radio as second. A map and a transistor radio so that you can assume the transistor radio's working (clearing throat).
2 Int1 Agent: That's interesting. I was wondering, can we hear from someone else?
3 16A: Maybe C, do you want to speak?
4 16C: Um… I chose, um… number one is cosmetic case.
5 16A: Oh, why is that?
6 16B: (h)… That's interesting.
7 16C: This is because, we need a mirror.
8 16A: Why?
9 16B: Why?
10 16C: We can… um (thinking sound)… We can send message to people near town.
(Talk continues, including the 2nd agent interruption)

Excerpt 2

11 16A: I put the number three as sea salt because it's gonna be terribly hot and you might need to have some salt in your body. I'm from India so I know handling fifty degrees C is not easy.
12 16B: Yes.
13 16A: And in that point I wouldn't choose any cosmetic case.
14 Int3 Agent: That's interesting. I was wondering, can we hear from someone else?
15 16 B: Okay this is difficult. She keeps interrupting.
16 16A: Okay, what do you think?
17 16B: So I totally agree with you because I had sea salt on it as well because your body dehydrates and you lose salt and you can actually faint if you don't get the right salt in your body, but I wasn't sure if I should put it number two or number three… cuz I put the map as number one. Okay, C, what do you think?
18 16C: Sorry…Um… please more slowly... Um... please speak more slowly, I can't understand.
19 16B: Oh…sorry.
20 16C: Um… checking... um... what's your ranking? What is the number one?

Table 1 Excerpts from Group 16's Discussion in the Automatic Agent Condition

In Excerpt 1 presented in Table 1, the first agent interruption (line 2) happened after NS_16B's remarks; after the interruption, NS_16A invited NNS_16C to share his opinions (line 3). For this particular case, the agent interrupted approximately 2 minutes after the discussion started, during which NS_16B spoke the most and NNS_16C spoke the least (16B spoke 83 words, 16A 56 words, 16C 3 words). After being invited to speak, NNS_16C explained why he chose cosmetic case as the most important item for survival (line 10) (which was actually the correct answer for this task).

Excerpt 2 showed the third agent interruption; this time, the agent interrupted NS_16A, and NS_16B invited NNS_16C to speak right after she finished making her point (agreeing with NS_16A that sea salt is important for survival, line 17). NNS_16C used this opportunity to 1) ask his NS partners to "speak more slowly" because he "can't understand" (line 18), and 2) check what items his partners chose in their individual rankings (line 20). In the interviews, he perceived that this was made possible by the agent interrupting the conversation so that "the discussion stopped".

In this group discussion, the first agent interruption made it possible for NNS_16C to share an important piece of information, i.e., the mirror in the cosmetic case was crucial for their survival. While the cosmetic case, which was ranked as the most important item by the NNS_16C (and also ranked the 1st by experts), wasn't chosen as one of the three items on the final group ranking list, the group decision was made with the availability of that important information: that the mirror could help to "send message to people near town". Although NNS_16C did not provide as crucial information after the third agent interruption as he did after the first agent interruption, it was a crucial point for him personally. He was able to 1) share his status (couldn't understand the conversation and hoped that his NS partners could slow down), and 2) get information on others' status (what NS chose for their individual rankings). Although the interruptions led to frustrations (discussed in the next theme below) of both NS_16A and NS_16B, they helped NNS to join the discussion, which was beneficial for both the group (key information for group task was provided) and the NNS personally.

Fairness of the Automatic Agent. In the example shown in Table 1, the interrupted NS speakers did not have as positive perceptions of the automatic agent as NNS. For example, after the third agent interruption, NS_16B commented "Ok, this is difficult. She keeps interrupting" (line 15) and went on to make her point. Although she invited 16C to talk right after completing her point (line 17), she expressed frustrations both in the group discussion and in the interviews. In the interviews, she commented that the agent wasn't necessary because they already tried to give everyone opportunities to speak.

> I think we were all getting frustrated in the first task on the electronic agent who was saying: let the other person speak. I thought we all did a good job in trying to give each other a chance to speak. I couldn't understand the point of this electronic person butting in saying that the other person should speak (Participant 16B, native English speaker).

The other native speaker NS_16A in the same group talked about how the first agent interruption in the discussion made them more conscious of their behaviors, e.g., if someone was speaking too much. However, after later interruptions, she perceived that the agent did not interrupt based on their behaviors, because they "were equally participating".

> (After) the first interruption, we were slightly cautious if somebody was speaking more, but when the second interruption came, we realized that it was not related to us, but it was just a continuous interruption, because we all were equally participating (Participant 16A, native English speaker).

The perception that the participation was already equal and they, as native speakers, did not dominate the conversations, was also shared among other NS. However, quantitative results showed participation wasn't equal (NNS proportion of words was .12, .14 and .17 in the control, manual and automatic agent condition respectively). NS interpreted the agent's interruption as an accusation that they were "the bad guy" that had spoken too much and therefore needed to stop talking. Due to this interpretation, some felt the agent's interruption was "unfair" because they thought the interrupted speaker wasn't speaking too much. For example, NS thought all members "were sharing enough", and expressed it was "unfair" when the agent accused them of "dominating the conversation".

> I think we were sharing enough. But that's my opinion. It's not based on any numbers that she would be monitoring (Participant 5A, native English speaker).

> When I finished the sentence, she (the agent)'s like: that's interesting, can we hear from someone else? Well, I felt it was a little unfair. I wasn't like dominating the conversation. I was trying to give space to everybody to talk, or like C (the NNS in the group) was talking

> a little less obviously, probably because of language barriers. I was trying to encourage her, you know, I'm not the bad guy (Participant 12A, native English speaker).

The non-native speaker in the same group with NS_12A), however, expressed that when the agent interrupted, her NS partners invited her to speak, so that she felt she had to say something. Reflecting back on the discussion, she appreciated being able to contribute to the group discussion.

> (When the agent interrupted) it felt a bit like (it interrupted) because I didn't talk much. They (NS) really respected my opinion when I did (speak). It felt like, ah, they're really taking me into consideration (Participant 12C, Japanese).

Being able to communicate ideas to other group members and contribute more to the final group decision, was also an important factor why participant 7C liked the automatic agent.

> With the automatic agent, the chances for me to communicate my ideas to others increased even if I was still preparing my ideas. Moreover, thanks to the agent, my ideas would be reflected in the overall flow of the discussion. In this aspect, I think the agent helped a lot (Participant 7C, Japanese).

Timing of the Automatic Agent. Both NS and NNS talked about the timing of the agent's interruptions. Participants described that the agent sometimes interrupted in the middle of sentences, and seemed to be random, which disrupted the flow of their conversations.

> I thought that probably there was a time limit on how much someone could speak before the agent broke in, but generally the agent interrupted at random in mid-sentence (Participant 4C, Japanese).

Perceiving the agent's timing as problematic also caused the NS participants to express annoyance towards the agent's interruption and question the agent's credibility, some even ignored the agent's requests as a result. Participant 5B, for example, described the agent as annoying because he was interrupted when he "just started" speaking.

> Well, the agent … seemed to interrupt us at random, which was a little bit annoying. Especially cuz the first time she spoke ever was like I was sort of waiting for my turn… She was like, can we hear from someone else? I was like, I just started! ... I guess it was nice to have a function to remind you, oh, make sure someone else can talk now, but her interjections seemed to be random (Participant 5B, native English speaker).

Participant 11A saw the agent as rude because the agent interrupted in the middle of his sentences; as a result, he chose to ignore the agent's requests and "just continued" speaking. Previous studies also showed that when interruptions are seen as annoying, participants tend to become less compliant [7].

> So I think she needs to know when to interrupt at the right moment. She was interrupting in the middle of the sentence, which is rude... And I just continued, I just told her, could I please finish and continue to speak? Yeah, that was not appropriate for her. If she interrupted at the correct time, it would be fine (Participant 11A, native English speaker).

The timing of the automatic agent was one of the major issues brought up by the participants. We discuss more about its limitations and design implications in section 5. However, an alternative point was argued by some of the NNS participants, i.e., the agent's timing to interrupt wasn't an issue for NNS that are in urgent need of gaps to catch up with conversations; by temporarily stopping the conversations, it was helpful for them, as non-native speakers, to reorganize their thoughts and participate in discussions with two native speakers.

> I thought the timing was pretty good. Specifically, the agent made it possible for me—a Japanese speaker—to speak up when I couldn't during the discussions with two native English speakers. By having the agent interject with good timing when I didn't say anything, it divided up the session into several parts (Participant 3C, Japanese).

> It made it easier to talk, or rather, the conversation stopped temporarily, so then we would reorganize our thoughts, and then it was like we restarted from there, so it actually made some parts easier to talk (Participant 13C, Japanese).

However, some NNS also reported that the agent sometimes interrupted at times when they might not have anything to say, which put extra pressure on them when they were already trying their best to follow the conversations:

> When it interrupted, (the agent) often drew attention. While I was doing all I could to speak in poor phrases in English, it was like a new obstacle came to our attention, and I thought, 'what were we discussing? And I forgot what I wanted to say (Participant 5C, Japanese).

Although NS were critical about the agent's timing, some thought the agent still made sense and had the potential to provide opportunities for everyone to participate in the conversation, if the agent could find more appropriate timing. Participant 8A, for example, pointed out that having a "digital agent" could be helpful because people are less likely to have their feelings hurt. Participant 16A, while showed frustrations towards the agent in the interview, acknowledged that an agent like this could help "everyone" to "participate", if it could find the right time to interrupt.

> I think it could be very helpful because unfortunately, some people don't know that they're talking too much and it's kind of hard to tell them that, so with a digital agent, you don't have to worry about hurting their feelings or, you know, say this: stop talking. Yeah, so we don't have to feel like it's anything personal… It can be just because this person talked for this length of time or they said this many words. It's calculated to interrupt at that time (participant 8A, native English speaker).

> If it was on the right time, then it would have been much helpful for the conversation, so that everybody can participate. Because usually in a group discussion, a few people are more excited to talk and a few people are nervous to talk. So this kind of interruption can cause equal participation (participant 16A, native English speaker).

4.3.2 Perceptions of the Manual Agent. To our surprise, NNS did not like or use the manual agent as much as we predicted; instead, they preferred the automatic agent. As mentioned previously, only 3 NNS in the 16 groups pressed the button to enable the agent in the manual agent condition. Since NS were not aware of the difference between the automatic and manual agent, we focus on NNS perceptions of the manual agent here.

NNS participants reported that they had difficulty finding the right time to press the button in the fast-paced conversations. Although they were told that the agent would interrupt after NS's next utterance (i.e., when the system detects an acoustic gap), they still tried to find the best timing to press the button (and most of them failed to find one).

> I really wanted to press the button, but the fact is that it was difficult to find a good timing. There was ample discussion during the third task, and because there were rapid transitions from one speaker to the next, it made it difficult to find the right timing to press the button (Participant 3C, Japanese).

In addition to challenges in finding the right timing, some NNS just did not want to interrupt "the flow of conversations"; even when they had something important to say, they just "couldn't press the button" to interrupt "a lively discussion".

> It's difficult. It's like interrupting the flow of conversation, so I thought maybe it's better not to press it (Participant 12C, Japanese).

> On several occasions, I really wanted to ask a question, but the other members were having a very lively discussion. … I saw holes in what they were saying, and I mostly just wanted to point out the hole and ask about their ideas. But again, the other two guys were carrying on such a lively discussion, so I actually couldn't press the button (Participant 3C, Japanese).

However, participant 4C reported that having the option to stop the conversation by simply pressing a button provided him some "assurance" that if he wanted to speak, he could do it without the agent. Quantitative results showed he spoke the most proportion of words in the manual agent condition (.15) (.08 in the control condition and .14 in the automatic agent condition).

> I could decide whether to use the agent or not, which offered some assurance in speaking. I didn't actually use the agent, but knowing that it was available gave me the courage to speak out. This made me feel more effortless in speaking in the discussion (Participant 4C, Japanese).

5 DISCUSSION

In this paper, we examined the effects of an agent that opened up opportunities for NNS to speak during a conversation with NS on NNS levels of participation in the conversation, as well as NS and NNS perceptions of the agent and the communication experience. Overall, the results showed that the automatic agent significantly increased NNS' level of participation without increasing their workload. However, neither NS nor NNS perceived the collaboration quality as higher in the agent conditions. We discuss these results in detail below.

5.1 Effects of the Agents on Participation

The quantitative results showed that the automatic agent successfully encouraged NNS to participate more, which previous studies had attempted but failed to achieve, either through improving NNS comprehension (e.g., [12][28][30][85][87]) or getting NS to accommodate to NNS language needs (e.g., [19][42][82]). This study shows that there is promise in providing direct instructions on expected behaviors of NS and NNS in conversations that use a common language. Although NNS proportion of turns did not change across conditions, they spoke significantly more proportion of words in the automatic agent condition. We think the reason might be that NNS used each turn to explain a piece of information in more detail as an effort to participate more, because the automatic agent reminded them that they weren't speaking enough (as discussed in the interviews). This uncovers a second level of the automatic agent's potential effect on NNS that is beyond simply providing the 3-5 turns opened up by making its request. This means that NNS might have tried to provide more information each time they speak up, as opposed to speaking up by taking more turns.

Drawing from the interviews, the majority of NNS thought the automatic agent made it easier for them to join the discussion, by "dividing up" the discussions and providing a pause for them to reflect on the fast-paced conversations. They were also less nervous about making grammatical mistakes when they were "suddenly" asked to speak and thus focused more on getting their ideas across. The excerpts in Table 1 showed the crucial role of the automatic agent on NNS participation. NNS was

invited by NS to speak when the agent interrupted; by speaking up, NNS was able to share information that was essential to group performance, and ask important questions sharing their own status and checking on other members' status. These communicative possibilities were only made possible by having the automatic agent temporarily pause the conversation and ask NS to pass the conversational floor.

Contrary to our expectations, only 3 out of 16 NNS activated the manual agent. NNS participants reported that they had difficulty finding the right time to press the button, although they were told that the agent would automatically interrupt after NS's next utterance. They also felt reluctant to break the flow of others' conversation and discussed possible social signals the agent might send (e.g., that they had something more important to say than the current speaker). These indicate traits of high-context cultures [37], where the Japanese speakers attend to their surroundings and how their individual behaviors might affect others. Although the manual agent could serve as a reminder to NNS that they are expected to join the conversation when they have something to say, they did not speak significantly more in the manual agent condition. Possibly, this was because NS did not pass the speaking floor to NNS as they did in the automatic agent condition (e.g., by inviting NNS to talk), because the agent was not actually activated in the manual condition for most groups. This further highlights the crucial role of NS in affecting how much NNS can participate in multilingual conversations. Also, NNS' unwillingness to activate the manual agent is contradictory to the findings of work on persuasive technologies (e.g. [2][25][26]) designed for more equal group participation. Tools such as ScoringTalk [2] and TurnTable [25] showed that simply by providing visual feedback on members' participation status helped groups to achieve more balanced participation. However, these studies primarily used participants who are native speakers; our study sheds light on the literature by showing an alternative possibility when similar strategies are tested with NNS participants.

5.2 Perceptions of Collaboration Quality

The interviews provided insights into why the automatic agent did not increase participants' perceptions of collaboration quality. NS perceived the agent's interruption as unfair and unnecessary because they thought the discussion was fair and all members were speaking equally; however, this was not the case (NNS proportion of words was .12, .14 and .17 respectively in the control, manual and automatic agent condition). Drawing from this insight, future work could monitor group members' participation and share feedback [21] to the interrupted speaker, to help them understand why the agent interrupted. As we showed in 4.3, while NS was annoyed by the agent's interruption, NNS in the same group often appreciated the agent interrupted and gave them a gap to catch up with or contribute to the conversations. Interestingly, even when the automatic agent provided more NNS speaking opportunities by reducing NS participation, it did not alter NNS perceptions of how accommodating their NS partners were. We suspect that this is due to the fact that the existence of these speaking opportunities was not attributed to the NS but to the agent itself. We elaborate on this point in 5.3.2.

We designed the automatic agent based on a simple framework that defined turns as intervals of speech uninterrupted by pauses longer than 0.5s, which sometimes led to the agent's requests overlapping with the interrupted NS speaker's speech [32][33]. As Suchman [79] pointed out, one of the problems of human machine communication is the inability of the machine to effectively employ repair mechanisms when signs of communication problems occur. Although it is also challenging for human communicators to accurately time the start of the next turn, making it common for brief overlaps during transition of turns, these overlaps are often resolved quickly through repair mechanisms [72]; however, since the automatic agent wasn't designed to attend to and respond to the

problems of real-world communication, some of the turn taking violations and overlaps did not get resolved as they would have in natural conversations. The asymmetry in resources available to the agent and the NS participants was a crucial reason for most of the communication problems that emerged in the study. Possible solutions include making the agent smarter and more contextualized (5.3.1) and making it more clear to the participants that the agent has limitations [1].

5.3 Design Implications

We used this study to explore communicative possibilities [1] that are yet unrealized in multilingual groups. We showed it is possible that even with a simple design, an agent that directly intervenes discussions and asks NS to pass conversational floors to NNS can achieve more balanced participation in multilingual group decision making discussions. What is more, the majority of NNS appreciated being able to have a break in the fast-past conversations and an opportunity to re-organize their thoughts, regardless of the agent's timing. Many emphasized their appreciation of being able to contribute more to the group discussion. However, the negative emotions (e.g., annoyance from the NS side) it created within the participants led to some limitations of the study, including unchanged perceptions of workload and collaboration quality. We provide suggestions for the design of future tools to improve conversational dynamics between NS and NNS speakers in multilingual groups.

5.3.1 Future Agent Designs. In our study, the agent's interruption sometimes overlaps with the current speaker's next utterance. Although overlapping talks are a natural result of turn-transitioning and are often resolved quickly in natural conversations (e.g., one party stops talking) [72][74], the agent's timing issues and lack of attendance to context led to annoyance from the NS Participants. Simple first steps for future designs include: the agent could make sounds (e.g., "hmm", "ah") in addition to interactive gestures (e.g., hand waving) to send signals [14][18] so that others know that the agent was about to say something.

In this study, NS complained about being interrupted when they thought the discussion was fair and all members were speaking equally, which was not the case (NNS proportion of words was .12, .14 and .17 respectively in the control, manual and automatic agent condition). Future agents could monitor group members' participation and share feedback about the speaking time or number of words spoken [21] to the interrupted speaker, to help NS understand why the agent interrupted. Also, human communicators often adapt to interventions by adjusting their "pre-existing practices" [1][8][9]. In our study, some NS participants mentioned that they tried to give NNS chances to talk after the first agent interruption made them aware of the imbalance of participation, and the agent failed to acknowledge those efforts because it wasn't made aware of the changes. Future agents could interrupt with continuously updated rules based on participants' behavioural changes, and potentially interrupt only the dominant speakers [98]. Also, the design for the timing of interruption could attend to the dynamics of the conversation and avoid interrupting when the discussion is at a crucial point (e.g., when a final decision is being made) [10].

This study also sheds light on other methods of intervention in conversations. While having a direct intervention that opens up conversational floors helps to balance participation, agents with other functions could also have potential to change communication dynamics in multilingual groups. For example, as suggested by previous studies [21] and interviewees in this study, the agent could interrupt by asking NS to summarize previous points, or asking clarification questions when it detects difficult words, which could help to "divide up" the conversations and provide a gap for NNS to reorganize their thoughts before participating in the discussion.

5.3.2 Attributions of NS Accommodation and Politeness of Interruptions. We found it interesting that the quantitative data showed NNS did not perceive NS to be more accommodating in the automatic agent condition, despite the fact that they were able to speak more and that NNS interviewees talked

about NS inviting them to speak when the agent interrupted. We suspect this is because NNS attributed the conversational floor opening to the agent instead of NS.

An alternative option could be, instead of publicly saying "Can we hear from someone else?", the agent privately sends requests to NS, such that NS are not directly interrupted. Results from the interviews showed that NS often perceived participations as equal, although that was not actually the case based on quantitative findings. Therefore, we believe it is also important to show NS information about each member's participation level, in addition to sending them requests to opening up floors. In this way, NS would be reminded to make informed decisions, but they can decide on appropriate times to open the conversational floor, e.g., they could finish up their points before letting someone else talk, as participant 16B did in Table 1, line 17. More importantly, the agent could be seen as more polite and empathetic, and previous studies showed using empathy in interruptions (e.g., asking users if it was a good time to interrupt) could promote more positive user experiences [7][59], reducing negative emotions from NS' side. Finally, NNS could attribute NS as more accommodating to their language abilities and communication styles, which has the potential to encourage positive emotions and group cohesion.

However, it is crucial to emphasize here, that the solution presented above needs to consider potential drawbacks, as previous studies providing feedbacks to NS reported that NS often trusted their own judgement instead of system recommendations and failed to make behavioural adjustment [19][21][42]. Alternative designs could employ agents that interrupt automatically but show signs of empathy, e.g., attending to the participants' frustration, stress or asking them if they were interrupted at the right time [49][59], or asking NS to give up conversational floors privately, and only openly making such requests if NS refuse to do that.

5.4 Limitations and Future Directions

Our study has a number of limitations that we plan to address in future work. First, we only studied conversations between native Japanese and native English speakers in Japan, collaborating in English as a common language; there was also a large age difference between NS and NNS participants and a gender difference within NNS participants. While the within-subjects design could help mitigate some of these issues, results may still differ for speakers of other languages, other age groups, gender combinations and/or from other cultural backgrounds. In this paper, we mainly focused on participants' language ability in analyzing our results and less on their cultural background, partially because the NS participants are from different cultures, and partially because NNS highlighted the importance of language barriers on their participation behaviors in previous studies [16][21][41]. However, language and culture are interconnected [57][94]; future studies should investigate the relationship between participation, language and culture. For example, if NNS participants are from individualist cultures and NS are from collectivist cultures, how would results differ? Personality traits of group members also affects a number of factors in group collaboration, such as compliance to rules [6][36], team dynamics [68], as well as preferred conversation topics [84]; for example, introverts showed more engagement in problem-centered topics, whereas extraverts were more active in a wider scope of topics [84]. Future studies could further explore if personality traits [48][61] could affect NNS participation and NS compliance to the agent's requests.

Second, we used three-member groups with 2NS and 1NS to recreate the smallest unit with NS as majority in a group, and to explore situations where speakers need to compete for the next speaking turn, as opposed to pairs where any non-speaker could be guaranteed the next turn at a certain point (see more in 2.3). However, in groups with more than three members, there is more complexity in both the turn-taking mechanisms and group dynamics; for example, when groups have more than one NNS, they might engage in code switching (where they switch back and forth between the common

language and their native language) and helping behaviors (e.g., those with higher common language proficiency may help those less fluent)[16][31]. Future studies are needed to explore members' behaviors in larger groups, which are also common in multilingual communication.

Third, as noted earlier, we designed our agent in a particular way, to interrupt based on six NS speaking turns in a row to provide an opportunity for NNS team members to contribute. The agent also made very direct requests as opposed to more subtle ones because they often failed to lead to behavioral changes in previous studies (e.g., [12][28][30][85][87][19][42][82]). As the interviews indicate, this design approach created some limitations in conversational interaction. Future work might explore the value of other design strategies for these agents, including those mentioned in the Design Implications section (5.3). Finally, we used a laboratory study with a realistic but not entirely natural team activity, the survival tasks. Additional work is needed to determine whether our findings generalize to other team configurations and tasks.

6 CONCLUSIONS

We examined the effects of a conversational agent that opens up the speaking floor for NNS during multilingual conversations in a controlled laboratory study. We compared an automatic agent and a manually controlled agent to a control condition with no agent. We found that NNS participated significantly more in the automatic agent condition than in the control condition. Excerpts from group discussions and interview results highlighted the crucial role of the automatic agent on NNS participation. However, there was a mismatch in NNS and NS perceptions of the agent, where NNS appreciated being able to contribute more after the agent's interruption, while the majority of NS perceived the agent's interruption as unfair. We provided design suggestions for improving future agents to support multilingual communication.

ACKNOWLEDGMENTS

We thank all the participants for their time and patience. We also thank NTT software developer Kobayashi-san for his excellent problem-solving skills.

REFERENCES

[1] Mark Aakhus and Sally Jackson. 2005. Technology, interaction, and design. *Handbook of Language and Social Interaction*, 411-436.

[2] Hiroyuki Adachi, Myojin Seiko and Shimada Nobutaka Shimada. 2015. ScoringTalk: a tablet system scoring and visualizing conversation for balancing of participation. *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications* (Kobe Japan, Nov. 2015), 1–5.

[3] Jo Angouri. 2013. The multilingual reality of the multinational workplace: Language policy and language use. *Journal of Multilingual and Multicultural Development* 34, 6: 564-581. https://doi.org/10.1080/01434632.2013.807273

[4] Alberto Alesina and Eliana La Ferrara. 1999. Participation in Heterogeneous Communities. *The Quarterly Journal of Economics*, 115(3), 847-904. https://doi.org/10.1162/003355300554935

[5] Geoffrey W. Beattie. 1981. The regulation of speaker turns in face-to-face conversation: Some implications for conversation in sound-only communication channels. *Semiotica* 34, 55–70.

[6] Laurent Bègue, Jean-Léon Beauvois, Didier Courbet, Dominique Oberlé, Johan Lepage, and Aaron A. Duke. 2015. Personality Predicts Obedience in a M ilgram Paradigm. *Journal of Personality, 83*(3), 299-306.

[7] Timothy Bickmore, Daniel Mauer, Francisco Crespo, and Thomas Brown. 2007. Persuasion, task interruption and health regimen adherence. In *International conference on persuasive technology* (pp. 1-11). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-77006-0_1

[8] Geoffrey C. Bowker and Susan Leigh Star. 1999. *Sorting things out: Classification and its consequences*. MIT Press: Cambridge, MA.

[9] John Seely Brown and Paul Duguid. 2000. *The social life of information.* Cambridge, MA: Harvard Business School Press.

[10] Cynthia Breazeal and Brian Scassellati. 1999. A Context-Dependent Attention System for a Social Robot. In *Proceedings of 16th International Joint Conference on Artificial Intelligence,* vol. 2, pp. 1146-1151. https://dl.acm.org/doi/10.5555/1624312.1624382

[11] Julia Cambre and Chinmay Kulkarni. 2019. One Voice Fits All?: Social Implications and Research Challenges of

Designing Voices for Smart Devices. In *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 1–19. https://doi.org/10.1145/3359325

[12] Xun Cao, Naomi Yamashita and Toru Ishida. 2018. Effects of Automated Transcripts on Non-native Speakers' Listening Comprehension. *IEICE Transactions on Information and Systems*, Vol.E101-D, No.03, 2018. https://doi.org/10.1587/transinf.2017EDP7255

[13] Erran Carmel (1999). Global Software Teams: Collaborating Across Borders and Time Zones. Prentice-Hall, Upper Saddle River, NJ.

[14] Justine Cassell, Matthew Stone and Hao Yan. 2000. Coordination and context-dependence in the generation of embodied conversation. In INLG'2000 *Proceedings of the First International Conference on Natural Language Generation*. Mitzpe Ramon, Israel. https://dl.acm.org/doi/10.3115/1118253.1118277

[15] Rebecca Cherng-Shiow Chang, Hsi-Peng Lu, and Peishan Yang. 2018. Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan. *Computers in Human Behavior* 84 (July 2018), 194–210. https://doi.org/10.1016/j.chb.2018.02.025

[16] Mei-Ling Chen, Naomi Yamashita and Hao-Chuan Wang. 2018. Beyond Lingua Franca: System-Facilitated Language Switching Diversifies Participation in Multiparty Multilingual Communication. In *Proceedings of the ACM on Human-Computer Interaction,* Vol. 2, CSCW, Article 34 (November 2018). https://doi.org/10.1145/3274303

[17] Anne Cutler and Mark Pearson. 1986. On the analysis of prosodic turn-taking cues. In *Intonation in Discourse*. College-Hill, San Diego, CA, pp. 139–156.

[18] Starkey Duncan. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology,* 23(2), p.283.

[19] Ilana Diamant, Brian Lim, Andy Echenique, Gilly Leshed, and Susan R. Fussell. 2009. Supporting intercultural collaboration with dynamic feedback systems: preliminary evidence from a creative design task. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09).* ACM, New York, NY, USA, 3997-4002. https://doi.org/10.1145/1520340.1520607

[20] Joan Morris DiMicco, Anna Pandolfo, and Walter Bender. 2004. Influencing group participation with a shared display. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW '04).* ACM, New York, NY, USA, 614-623. DOI=http://dx.doi.org/10.1145/1031607.1031713

[21] Wen Duan, Naomi Yamashita and Susan R. Fussell. 2019. Increasing Native Speakers' Awareness of the Need to Slow Down in Multilingual Conversations Using a Real-Time Speech Speedometer. In *Proceedings of the ACM on Human-Computer Interaction,* Vol. 3, No. CSCW, Article 171 November 2019. ACM, New York, NY, USA. 25 pages. https://doi.org/10.1145/3359273

[22] Vitaly J. Dubrovsky, Sara Kiesler, and Beheruz N. Sethna. 1991. The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. *Human-Computer Interaction,* 6, 119-146. https://doi.org/10.1207/s15327051hci0602_2

[23] Norah E. Dunbar and Judee K. Burgoon. 2005. Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships* 22, 2 (2005), 207–233. https://doi.org/10.1177/0265407505050944

[24] Cecilia E. Ford and Sandra A. Thompson. 1996. Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. Schegloff, & S. Thompson (Eds.), *Interaction and Grammar* (Studies in Interactional Sociolinguistics, pp. 134-184). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620874.003

[25] Heather A. Faucett, Matthew L. Lee, and Scott Carter. 2017. I Should Listen More: Real-time Sensing and Feedback of Non-Verbal Communication in Video Telehealth. *Proceedings of the ACM on Human-Computer Interaction.* 1, CSCW (Dec. 2017), 1–19. doi:https://doi.org/10.1145/3134679

[26] Fu Carolyn, Kritika Dhanda, Marc Exposito Gomez, Haeyoung Kim, and Yan Zhang. 2017. TurnTable: Towards More Equivalent Turn-Taking. *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction* (New York, NY, USA, 2017), 609–615.

[27] Ge Gao, Hao-Chuan Wang, Dan Cosley and Susan R. Fussell. 2013. Same translation but different experience: The effects of highlighting on machine-translated conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13).* ACM, New York, NY, USA, 449–458. https://doi.org/10.1145/2470654.2470719

[28] Ge Gao, Naomi Yamashita, Ari MJ Hautasaari, Andy Echenique, and Susan R. Fussell. 2014. Effects of Public vs. Private Automated Transcripts on Multiparty Communication Between Native and Non-native English Speakers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14).* ACM, New York, NY, USA, 843–852. https://doi.org/10.1145/2556288.2557303

[29] Ge Gao, Naomi Yamashita, Ari 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 SIGCHI Conference on Human Factors in Computing Systems (CHI '15).* ACM, New York, NY, USA, 3463–3472. https://doi.org/10.1145/2702123.2702498

[30] 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 (CSCW' 15).* ACM, New York, NY, USA, 852–863. https://doi.org/10.1145/2675133.2675197

[31] Ge Gao, and Susan R. Fussell. 2017. A kaleidoscope of languages: When and how non-Native English speakers shift between English and their Native language during multilingual teamwork. In *Proceedings of the 2017 CHI Conference*

on Human Factors in Computing Systems. https://doi.org/ 10.1145/3025453.3025839
[32] Barney G Glaser and Anselm L. Strauss. 1967. The discovery of grounded theory: Strategies for qualitative research. Chicago, IL: Aldine Publishing Co.
[33] Erving Goffman. 1967. *Interaction Ritual*. Garden City, N.Y.: Doubleday Anchor.
[34] Charles Goodwin and Marjorie H. Goodwin. 2004. Participation, in A. Duranti (ed.) *A Companion to Linguistic Anthropology*, pp. 222–43. Oxford: Basil Blackwell.
[35] Agustín Gravano and Julia Hirschberg. 2011. Turn-taking cues in task-oriented dialogue. *Computer Speech & Language, 25*(3), pp.601-634.
[36] Gisli H. Gudjonsson, Jon Fridrik Sigurdsson, Olafur O. Bragason, Emil Einarsson, and Eva B. Valdimarsdottir. 2004. Compliance and personality: The vulnerability of the unstable introvert. *European Journal of Personality, 18*(5), 435-443.
[37] Edward Twitchell Hall. 1976. *Beyond culture*, New York: Anchor Press–Doubleday.
[38] Jay Hall and W. H. Watson. 1970. The effects of normative intervention on group decision-making performance. *Human Relations, 23*, 299-317. doi.org/10.1177/001872677002300404
[39] Sandra G. Hart and Lowell E. Staveland. 1998. Development of NASA-TLX: Results of empirical and theoretical research. *Advances in Psychology*, 52, 139-183. https://doi.org/10.1016/S0166-4115(08)62386-9
[40] Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38, 4: 555–568. https://doi.org/10.1016/j.wocn.2010.08.002
[41] Helen A. He and Elaine M. Huang. 2014. A qualitative study of workplace intercultural communication tensions in dyadic face-to-face and computer-mediated interactions. In *Proceedings of the 2014 conference on Designing interactive systems (DIS' 14)*. ACM, New York, NY, USA, 415–424. https://doi.org/10.1145/2598510.2598594
[42] Helen A. He, Naomi Yamashita, Ari Hautasaari, Xun Cao, and Elaine M. Huang. 2017. Why Did They Do That? Exploring Attribution Mismatches Between Native and Non-Native Speakers Using Videoconferencing. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW' 17)*. ACM, New York, NY, USA, 297–309. https://doi.org/ 10.1145/2998181.2998205
[43] Naoko Morita. 2004. Negotiating Participation and Identity in Second Language Academic Communities. *TESOL Quarterly* 38 (4): 573–603. https://doi.org/10.2307/3588281
[44] Pamela J. Hinds, Tsedal B. Neeley, and Catherine D. Cramton. 2014. Language as a lightning rod: Power contests, emotion regulation, and subgroup dynamics in global teams. *Journal of International Business Studies* 45, 5: 536-561. https://doi.org/10.1057/jibs.2013.62
[45] Hanneke Houtkoop and Harrie Mazeland. 1985. Turns and discourse units in everyday conversation. *Journal of Pragmatics* 9:595-619. https://doi.org/10.1016/0378-2166(85)90055-4
[46] Maddy Janssens and Jeanne M. Brett. 2006. Cultural Intelligence in Global Teams: A Fusion Model of Collaboration. *Group & Organization Management*, 31(1), 124–153. https://doi.org/10.1177/1059601105275268
[47] Gail Jefferson. 1988. Preliminary notes on a possible metric which provides for a standard maximum silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective*. Clevedon, UK: Multilingual Matters.
[48] Oliver P. John, and Sanjay Srivastava. 1999. The Big-Five trait taxonomy: History, measurement, and theoretical perspectives.
[49] Jonathan Klein, Youngme Moon, and Rosalind W. Picard. 2002. This computer responds to user frustration: Theory, design, and results. *Interacting with computers, 14*(2), 119-140. https://doi.org/10.1016/S0953-5438(01)00053-4
[50] Daan Van Knippenberg. 2000. Work motivation and performance: A social identity perspective. *Applied psychology, 49*(3), 357-371. https://doi.org/10.1111/1464-0597.00020
[51] JC Lafferty, Patrick M Eady, and J Elmers. 1974. The desert survival problem. *Experimental Learning Methods* (1974).
[52] Kornel Laskowski, 2010. Modeling norms of turn-taking in multi-party conversation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 999-1008).
[53] Gene H. Lerner. 2003. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society* 32, 177–201. https://doi.org/10.1017/S004740450332202X
[54] Gilly Leshed, Dan Cosley, Jeffrey T. Hancock, and Geri Gay. 2010. Visualizing language use in team conversations: Designing through theory, experiments, and iterations. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 4567-4582. https://doi.org/10.1145/1753846.1754195
[55] Han Z. Li, Young-ok Yum, Robin Yates, Laura Aguilera, Ying Mao, and Yue Zheng. 2005. Interruption and Involvement in Discourse: Can Intercultural Interlocutors be Trained? *Journal of Intercultural Communication Research* 34, 4: 233–254.
[56] Huisi (Jessica) Li, Connie Y. Yuan, Natalya N. Bazarova, and Bradford S. Bell. 2018. Talk and Let Talk: The Effects of Language Proficiency on Speaking Up and Competence Perceptions in Multinational Teams. *Group & Organization Management* 44, no. 5 (October 2019): 953–89. https://doi.org/10.1177/1059601118756734
[57] Xiaoshi Li, and Xuerui Jia. 2006. Why don't you speak up?: East Asian students' participation patterns in American and Chinese ESL classrooms. *Intercultural Communication Studies*, 15(1), 192.
[58] Jessica Lipnack and Jeffrey Stamp (1997). Virtual Teams: Reaching Across Space, Time and Organizations with Technology. NY: John Wiley & Sons.
[59] Karen K. Liu, and Rosalind W. Picard. 2005. Embedded empathy in continuous, interactive health assessment. In *CHI Workshop on HCI Challenges in Health Assessment* (Vol. 1, No. 2, p. 3).
[60] Yadong Luo and Oded Shenkar. 2006. The multinational corporation as a multilingual community: Language and organization in a global context. *Journal of International Business Studies, 37*, 321-339.

https://doi.org/10.1057/palgrave.jibs.8400197
[61] Gerald Matthews, Ian J. Deary, & Martha C. Whiteman. 2003. *Personality traits*. Cambridge University Press.
[62] Nicoleta Meslec and Petru L. Curşeu. 2013. Too close or too far hurts: cognitive distance and group cognitive synergy. *Small Group Research, 44*(5), 471-497. http://doi.org/10.1177/1046496413491988
[63] Lorenza Mondada. 2012. The dynamics of embodied participation and language choice in multilingual meetings. *Language in Society, 41*(2), 213-235. Retrieved from http://www.jstor.org/stable/41487754
[64] Raffaella Negretti and Miguel Garcia-Yeste. 2014. 'Lunch Keeps People Apart': The Role of English for Social Interaction in a Multilingual Academic Workplace. *Multilingua, 34*(1), 93-118. https://doi.org/10.1515/multi-2014-1038
[65] Duyen T. Nguyen and Susan R. Fussell. 2013. Effects of Message Content on Cognitive and Affective Processes in Cross-culture and Same-culture Instant Messaging Conversations. In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW' 13)*. ACM, New York, NY, USA. https://doi.org/10.1145/2441776.2441782
[66] A.C. Norwine and O.J. Murphy. 1938. Characteristic time intervals in telephonic conversation. *Bell System Technical Journal, 17*(2), pp.281-291.
[67] Mei-hua Pan, Naomi Yamashita, and Hao-chuan Wang. 2017. Task Rebalancing: Improving Multilingual Communication with Native Speakers - Generated Highlights on Automated Transcripts. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW' 17)*. ACM, New York, NY, USA. https://doi.org/10.1145/2998181.2998304
[68] Rui Prada, Samuel Ma, and Maria Augusta Nunes. 2009. Personality in social group dynamics. In *2009 International Conference on Computational Science and Engineering* (Vol. 4, pp. 607-612). IEEE.
[69] Pamela Rogerson-Revell. 2007. Using English for international business: A European case study. *English for Specific Purposes* 26. 103–120. https://doi.org/10.1016/j.esp.2005.12.004
[70] Pamela Rogerson-Revell. 2008. Participation and performance in international business meetings. *English for Specific Purposes* 24. 401–421. https://doi.org/10.1016/j.esp.2008.02.003
[71] Jonathon Ryan, and Leslie Forrest. 2019. 'No chance to speak': developing a pedagogical response to turn-taking problems. *Innovation in Language Learning and Teaching*, 1-14. https://doi.org/10.1080/17501229.2019.1687709
[72] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn taking for conversation. In *Language* 50: 696-735. https://doi.org/10.2307/412243
[73] Claude Sammut. 2001. Managing Context in a Conversational Agent. *Electronic Articles in Computer and Information Science,* 3(7):1–13.
[74] Emanuel A. Schegloff. 2003. Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29, 1–63. https://doi.org/10.1017/S0047404500001019
[75] Pnina Shachaf. 2008. Cultural diversity and information and communication technology impacts on global virtual teams: An exploratory study. *Information Management*, 45, 131-142. https://doi.org/10.1016/j.im.2007.12.003
[76] Günter K Stahl, Martha L Maznevski, Andreas Voigt, and Karsten Jonsen. "Unraveling the Effects of Cultural Diversity in Teams: A Meta-analysis of Research on Multicultural Work Groups." *Journal of International Business Studies* 41, no. 4 (2010): 690-709. Accessed April 3, 2020. www.jstor.org/stable/40604760
[77] Garold Stasser and Laurie A. Taylor. 1991. Speaking Turns in Face-To-Face Discussions. *Journal of Personality and Social Psychology,* vol 60, 5, 675–684. https://psycnet.apa.org/doi/10.1037/0022-3514.60.5.675
[78] Garold Stasser and William Titus. 1987. Effects of Information Load and Percentage of Shared Information on the Dissemination of Unshared Information during Group Discussion. *Journal of Personality and Social Psychology,* vol 53 (1), 81–93. https://psycnet.apa.org/doi/10.1037/0022-3514.53.1.81
[79] Lucy A. Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge, UK: Cambridge University Press.
[80] Jane F. Stephens. 1987. *Towards a Model of Turn-taking in Conversation.* Ph.D. Dissertation. University of Sheffield, UK.
[81] Dennis D. Stewart and Garold Stasser. 1996. The Sampling of Critical, Unshared Information in Decision-Making Groups: The Role of an Informed Minority. *European Journal of Social Psychology,* vol 28, 95–113. https://psycnet.apa.org/doi/10.1002/(SICI)1099-0992(199801/02)28:1%3C95::AID-EJSP847%3E3.0.CO;2-0
[82] Emma Sweeney and Zhu Hua. 2010. Accommodating toward your audience: Do native speakers of English know how to accommodate their communication strategies toward nonnative speakers of English? *Journal of Business Communication* 47, 4: 477–504. https://doi.org/10.1177/0021943610377308
[83] Hanne Tange & Jakob Lauring. 2009. Language management and social interaction within the multilingual workplace. *Journal of Communication Management, 13*, 218-232.
[84] Avril Thorne. 1987. The press of personality: A study of conversations between introverts and extraverts. *Journal of Personality and Social Psychology, 53*(4), 718.
[85] Hao-Chuan Wang, Susan R. Fussell and Dan Cosley. 2013. Machine translation vs. Common language: Effects on idea exchange in cross-lingual groups. In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW' 13)*. ACM, New York, NY, USA, 935-944. https://doi.org/ 10.1145/2441776.2441882
[86] Nigel Ward and Wataru Tsukahara. 2000. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics,* 23: p. 1177-1207. https://doi.org/10.1016/S0378-2166(99)00109-5
[87] Naomi Yamashita, and Toru Ishida. 2006. Effects of machine translation on collaborative work. In *Proceedings of the 2006 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW' 06)*. ACM, New York, NY, USA, 515–523. https://doi.org/10.1145/1180875.1180955

[88] Naomi Yamashita, Andy Echenique, Toru Ishida, and Ari Hautasaari. 2013. Lost in Transmittance: How Transmission Lag Enhances and Deteriorates Multilingual Collaboration. In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work Social Computing Social Computing (CSCW' 13)*. ACM, New York, NY, USA, 923–934. https://doi.org/10.1145/2441776.2441881
[89] Pamela Rogerson-Revell. 2008. Participation and performance in international business meetings. *English for Specific Purposes* 27, 3: 338–360. https://doi.org/10.1016/j.esp.2008.02.003
[90] Elizabeth Shriberg, Andreas Stolcke, and Don Baron. 2001. Observations on overlap: Findings and implications for automatic processing of multi-party conversation. In *Seventh European Conference on Speech Communication and Technology*.
[91] Karl Weilhammer and Susan Rabold. 2003. Durational aspects in turn taking. In *Proceedings of the International Conference of Phonetic Sciences* (pp. 2145-2148).
[92] John Wiemann and Mark Knapp. 2017. Turn-taking in conversations. *Communication Theory*, pp.226-245.
[93] Victor Yngve. 1970. On getting a word in edgewise. Papers from the *Sixth Regional Meeting [of the] Chicago Linguistic Society, p*age 568.
[94] Chien Wen Yuan, Leslie D. Setlock, Dan Cosley, and Susan R. Fussell. 2013. Understanding informal communication in multilingual contexts. In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW' 13)*. ACM, New York, NY, USA, 909-922. https://doi.org/10.1145/2441776.2441880
[95] Steve Whittaker. 2003. Theories and methods in mediated communication. In *The Handbook of Discourse Processes,* A. C. Graesser,M. A. Gernsbacher, and S. R. Goldman, Eds. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 243–286.
[96] Connie Yuan, Natalya N. Bazarova, Janet Fulk, and Zhi-Xue Zhang. 2013. Recognition of expertise and perceived influence in intercultural group collaboration. *Journal of Communication*, Volume 63, Issue 3, June 2013, Pages 476–497. https://doi.org/10.1111/jcom.12026
[97] Yohtaro Takano and Akiko Noda. 1993. A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology* 24, 4: 445–462. https://doi.org/10.1177/0022022193244005
[98] Ilana Volfin and Israel Cohen. 2013. Dominant speaker identification for multipoint videoconferencing. *Computer Speech Language*, 27(4):895 – 910. https://doi.org/10.1016/j.csl.2012.03.002
[99] ATLAS.ti: The Qualitative Data Analysis Research Software. Retrieved from http://atlasti.com

Received: May 2022, Revised: August 2022, Accepted: September 2022.