"Let Me Ask Them to Clarify If You Don't Want To"—A Clarification Agent for Nonnative Speakers

Wen Duan

Dept. of Communication

Cornell University

Ithaca, NY 14853, USA

wd238@cornell.edu

Naomi Yamashita

NTT Communication Science

Labs

2-4 Hikaridai, Seika-cho,

Soraku-gun, Kyoto, Japan

naomiy@acm.org

Sun Young Hwang

Dept. of Communication

Cornell University

Ithaca, NY 14853, USA

sh2284@cornell.edu

Susan Fussell

Dept. of Communication

And Dept. of Information Science

Cornell University

Ithaca, NY 14853, USA

sfussell@cornell.edu

Abstract

When non-native English speakers (NNS) encounter messages they do not understand, they are often reluctant to ask native speakers (NS) for clarification. In this paper, we explored whether a conversation agent that asks clarification questions would increase NNS' willingness to ask questions. We compared two agents: one that asked for clarification about specific message elements and one that asked general clarification questions. NNS and NS rated how disruptive the agent was, the quality of the conversation, and whether they would feel embarrassed to ask their own questions. NNS found both types of agent less disruptive than NS did, but both found the specific agent more disruptive than the generic agent. NS rated the conversations higher in quality than NNS, but there was no effect of agent condition. We discuss potential of using conversational agents to boost NNS's confidence in conversation.

Author Keywords

Multilingual collaboration; conversational agents, computer-mediated communication

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

Introduction

Global collaboration often uses English as a common language, which can create challenges for people who are non-native speakers (NNS). However, research suggests that NNS are reluctant to ask for clarification when they do not understand a message, in part due to face concerns. Previous studies have shown that NNS could benefit from NS's awareness of NNS's problems in language processing [2,3] and from an increase in the length of silence gaps between NS's speech [6].

In theory, if NNS asked for clarification when needed, NS would pause to provide clarification, leading to improved conversational grounding [1]. However, NNS are unable or reluctant to raise clarification questions for at least two reasons [4]: First, it can be difficult to formulate a question in a nonnative language while keeping up with the ongoing conversation. Second, NNS may be concerned about a loss of face from revealing lack of understanding.

To address these problems, we devised and tested a conversation agent for video conferencing systems that generated clarification requests (CRs) at points where a NS's utterance could be potentially confusing for NNS. By asking questions on behalf of NNS, the agent was intended to eliminate the face threat of needing more information and the cognitive load of formulating a question while at the same time slowing down the conversation.

However, there may be trade-offs between the level of detail elicited by a CR and the disruptiveness of the request to the ongoing conversation. To explore these tradeoffs, we compared two types of agent: one that made specific CRs (e.g., "what do you mean by *engagement?*") and one that made generic CRs (e.g., "what do you mean?"). Specific CRs often elicit shorter and more specific responses than do generic CRs [5]. In an experimental study, we compared the effect of the two types of agent on the multiparty collaboration and conversation between NS and NNS. Based on previous literature, we posed the following hypothesis and research questions:

H1: An agent that asks generic CRs will be viewed as more disruptive than an agent that asks specific CRs.

RQ1: Does the type of agent CR influence perceptions of the quality of conversation? Does this differ for NS and NNS?

RQ2: Does the type of agent CR influence feelings of embarrassment to ask their own questions? Does this differ for NS and NNS?

Method

Using a within subjects experimental design, we asked 14 triads of 2 NS and 1 NNS to collaborate via video conferencing on a series of survival tasks, either with an agent that asked specific CRs or with one that asked generic CRs. The order of task and agent conditions were counterbalanced.

Participants

A total of 42 participants were recruited through a third-party personnel recruiting company. Among them

28 (9 female) were native English speakers who grew up and received their education in an English-speaking country. The mean age for native speakers was 45.14 (SD=12.13). The remaining 14 participants (13 female) were native speakers of Japanese to whom English was their non-native language. Their mean age was 45.21 (SD=12.40). NNS were able to communicate in English at low to moderate fluency (listening: M=4.21, SD=1.19; speaking: M=3.71, SD=0.99 on a scale from 1=not fluent at all to 7=very fluent).

Material and equipment
As is shown in Figure 1, we implemented the agent feature into the three-way videoconferencing interface by creating a fourth "participant" whose image was presented the same way as participants were, except that instead of a camera-captured motion picture, the image of the agent was an unanimated female cartoon avatar matched with a female human voice.

We used the confidence score returned for each utterance by the speech recognition system as an indicator of potential confusion. When the NS's utterance received a score lower than a threshold, the agent would ask the speaker to clarify. The threshold for each speaker was determined during the training session by the utterance receiving the 10th lowest confidence score. To avoid frequent interruption, we limited the number of interruptions per discussion session to 4 times. After the agent interrupted one speaker, it would not do so again for 2 minutes.
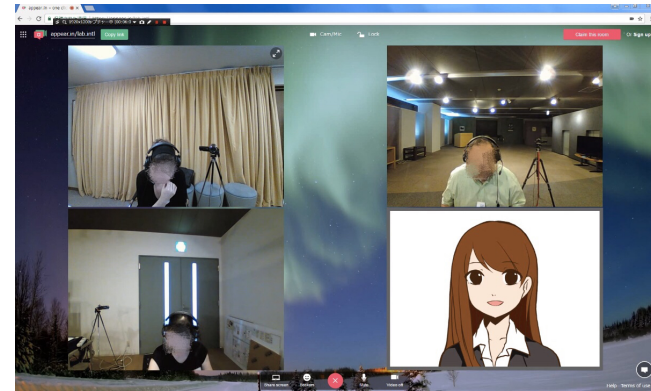

Figure 1: Interface with CR agent

We preprogramed the agent's CRs based on types of conversation breakdowns [1], as is shown in Table 1. In specific CR condition, once the agent was activated, she asked one of the three questions in Column 2, randomly picked by the system (replacing xx with the portion of the message that had received a confidence score lower than threshold). In the generic CR condition, the agent asked one of the three questions in Column 3 without repeating the problematic element that activated the agent in the first place.

Source of problem	Specific CRs	Generic CRs
acoustic	Did you say xx?	What did you say?
semantic/ pragmatic	What do you mean by xx?	What do you mean?
	Could you explain xx.	Could you explain?

Table 1: Agent's clarification questions

Task and procedure
Each experiment lasted about 2 hours. After completing a demographic survey, participants went through a training session in which they completed a desert survival task. They first individually ranked a list of items in the order of their importance for their survival in the desert, then, they discussed the rankings with their partners in order to come up with a group agreement. The speech recognition confidence scores for training discussion were used for determining individual speaker's interrupting thresholds. After the training session, they completed two other survival tasks (lunar and ocean) in the same fashion, except they were told that a Japanese agent would join their discussion and ask for clarification when "she" thinks is needed. They completed a post-task survey after each session.

Measures
Agent disruptiveness was measured using a three-item, 7-point scale (Cronbach's α=.80) asking whether the agent "was annoying", "interrupted abruptly" and "interrupted too many times". Conversational quality was measured by a three-item, 7-point scale (Cronbach's α=.76) asking about the "efficiency", "friendliness" and "success" of the discussion perceived by participants. We also asked about their feeling of embarrassment about asking CR by using a single-item on 7-point scale (for all measures: 1=strongly disagree, 7=strongly agree).

Results

Agent Disruptiveness
A 2 (agent condition: generic or specific) x 2 (language: NNS vs. NS) mixed models ANOVA using subjects nested within groups as a random effect

showed a significant main effect of agent condition ($F_{[1, 40]}$ = 6.23, p = .02). Participants in the specific condition rated the agent as more disruptive (M = 3.81, SE = .24) than participants in the generic condition (M = 3.35, SE = .24). There was also a borderline significant effect of language ($F_{[1, 40]}$ = 2.88, p = .10). Native speakers rated the agent as more disruptive (M = 3.96, SE = .26) than non-native speakers (M = 3.20, SE = .36). There was no interaction between agent condition and language ($F_{[1, 40]}$ = .56, p = .46). See Figure 2.
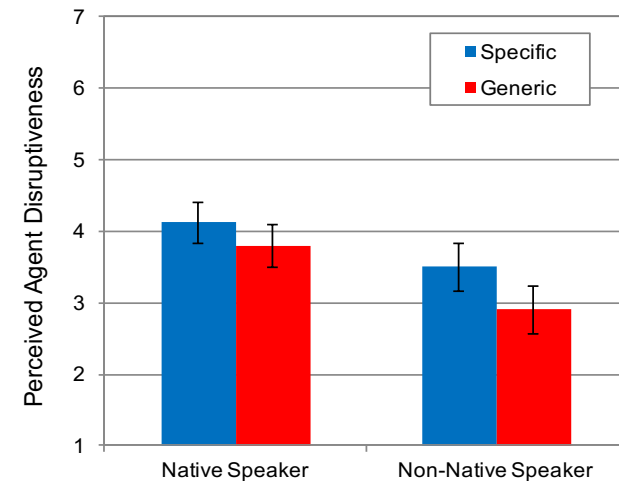

Figure 2: Mean agent disruptiveness as a function of type of agent CR and language (error bars represent standard errors of the mean).

Conversational quality
A 2 (agent condition) x 2 (language) mixed models ANOVA of the form outlined above showed no main effect of agent condition ($F_{[1, 40]}$ = .18, p = .68); for the specific agent, M = 6.35, SE = .10, for the generic

agent M = 6.39, SE = .10). However, there was a significant effect of language (F [1, 40] = 10.39, p = .003). Native speakers rated the conversation higher in quality (M = 6.64, SE = .10) than non-native speakers (M = 6.10, SE = .14). There was no significant interaction between agent condition and language (F [1, 40] = 1.03, p = .32). See Figure 3.
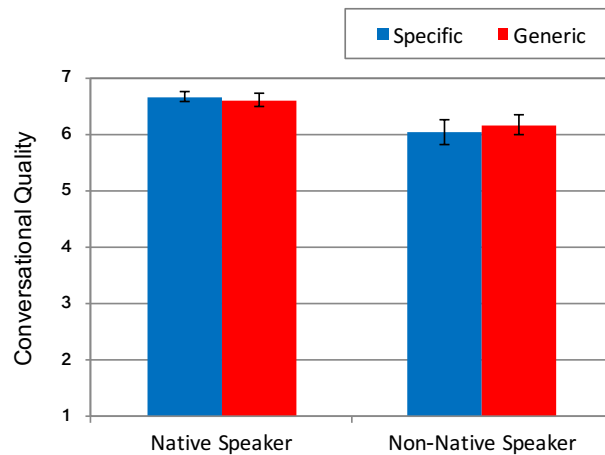

Figure 3: Mean conversational quality as a function of type of agent CR and language (error bars represent standard errors of the mean).

Feeling of embarrassment

An ANOVA of the form described above showed a borderline significant main effect of agent condition (F [1, 40] = 3.40, p = .07); for the specific agent, M = 1.93, SE = .20, for the generic agent M = 2.23, SE = .20). There was also a significant effect of language (F [1, 40] = 16.80, p < .001). Native speakers reported feeling less embarrassed about asking for clarification (M = 1.34, SE = .21) than non-native speakers (M = 1.34, SE = .21). There was also a significant interaction between agent condition and language (F [1, 40] = 8.58, p = .006). As shown in Figure 4, while NS reported low levels of embarrassment in both conditions, NNS reported higher embarrassment in the generic condition than in the specific condition.
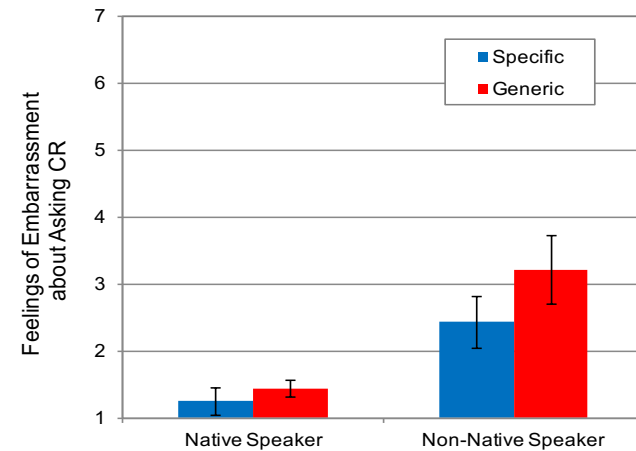

Figure 4: Mean feelings of embarrassment about asking CR as a function of type of agent CR and language (error bars represent standard errors of the mean).

Discussion

Our results show that NS found the agent more disruptive than NNS, which is not surprising given that the agent only interrupted NS. However, the NS did not see the agent as highly disruptive, suggesting that our approach could aid NNS without inconveniencing NS. However, contrary to our hypothesis, both NS and NNS found the specific CR agent more disruptive than generic CR agent. We suspect this was due to a flaw in our interruption mechanism— utterances with low confidence score that were pulled out to clarify on in

the specific CR condition were often transcribed inaccurately and even hard for NS to catch and make sense of. To make the activation of the CR agent more reliable, we are currently modifying our mechanism for detecting the speech elements that are potentially confusing for NNS, by incorporating SLA (second language acquisition) and NLP (natural language processing) resources.

There was no significant difference between agent conditions as to ratings of conversational quality. However, NNS rated conversational quality in both conditions lower than NS did. Language fluency and confidence in using the language might have played a role in NNS' conversation experiences, especially when working with two NS as was the case in our study. In future work, we will compare the two agent conditions with a baseline condition without an agent to see how a CR agent affects NNS's conversation experience.

It is not surprising that NNS felt much more embarrassed to ask their own clarification questions than NS since they used a nonnative language to communicate. It is interesting however, that NNS felt less embarrassed in the specific CR condition. In light of our interview data, we speculate that this could result from the mis-transcribed utterances spoken out by the specific CR agent. NNS reported feeling more relaxed and confident when the agent appeared to be even more incompetent in English (i.e. uttering nonsense words) than NNS perceived themselves to be.

In summary, our study showed that a CR agent can potentially facilitate multiparty multilingual collaboration with minimal disruption to the conversation flow. It also has the potential to encourage NNS to overcome their face concerns and ask clarification for themselves.

Acknowledgements

References
1. Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge, UK.

2. 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 *Proc. CHI 2015*, 3463–3472.

3. Helen Ai He, Naomi Yamashita, Chat Wacharamanotham, Andrea B. Horn, Jenny Schmid, and Elaine M. Huang. 2017. Two Sides to Every Story: Mitigating Intercultural Conflict through Automated Feedback and Shared Self-Reflections in Global Virtual Teams. In *Proc. ACM Hum.-Comput. Interact. 1, CSCW, Article 51*

4. 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, 233–254.

5. Verena Rieser and Johanna D Moore. 2005. Implications for Generating Clarification Requests in Task-oriented Dialogues. In *the 43rd Annual Meeting of the ACL*, 239–246.

6. Naomi Yamashita, Andy Echenique, Toru Ishida, and Ari Hautasaari. 2013. Lost in Transmittance: How Transmission Lag Enhances and Deteriorates Multilingual Collaboration. *ACM Conference on Computer Supported Cooperative Work Social Computing Social Computing*: 923–934.