Improving Multilingual Collaboration by Displaying How Non-native Speakers Use Automated Transcripts and Bilingual Dictionaries

Ge Gao[1,2], Naomi Yamashita[1], Ari Hautasaari[1], Susan R. Fussell[2]

[1]NTT Communication Science Labs
2-4 Hikaridai, Seika-cho, Soraku-gun,
Kyoto, Japan
naomiy@acm.org,
ari.hautasaari@lab.ntt.co.jp

[2]Department of Communication
Cornell University
Ithaca NY 14850 USA
[gg365, sfussell] @cornell.edu

ABSTRACT

Conversational grounding, or establishing mutual knowledge that messages have been understood as intended, can be difficult to achieve when some conversational participants are using a non-native language. These difficulties in grounding can be challenging for native speakers to detect. In this paper, we examine the value of signaling potential grounding problems to native speakers (NS) by displaying how non-native speakers (NNS) use automated transcripts and bilingual dictionaries. We conducted a laboratory experiment in which NS and NNS of English collaborated via audio conferencing on a map navigation task. Triads of one NS guider, one NS follower, and one NNS follower performed the task using one of three awareness displays: (a) a no awareness display that showed only the automated transcripts, (b) a general awareness display that showed whether each follower was reading the automated transcripts and/or translating a word; or (c) a detailed awareness display that showed which line of the transcripts a follower was reading and/or which words he/she was translating. NS guiders and NNS followers collaborated most successfully with the detailed awareness display, while NS guiders and NS followers performed equally across conditions. Our findings suggest several ways to improve systems to support multilingual collaboration.

Author Keywords

Awareness Display; Collaboration; Multilingual; Automated Speech Recognition (ASR); Translation.

ACM Classification Keywords

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

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea

ACM 978-1-4503-3145-6/15/04…$15.00.
http://dx.doi.org/10.1145/2702123.2702498

INTRODUCTION

Globalization makes collaboration in modern organizations increasingly multinational and virtual [16]. As people with different national backgrounds and geographical locations work together, a multilingual context arises in which collaborators speak "a cocktail of languages [18]".

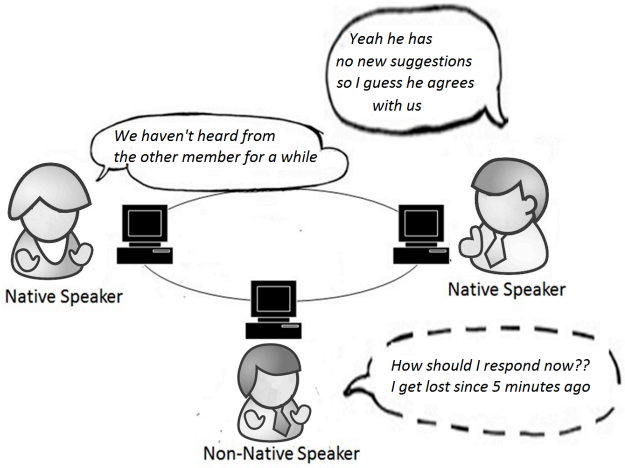

Figure 1. Grounding is hard to achieve during multilingual communication: Non-native speakers sometimes cannot express the problem they encountered, and native speakers may not realize anything is wrong.

In most multilingual collaborations, people use a common language, such as English [12]. Due to differences in fluency, multilingual teams may encounter communication problems that rarely happen within monolingual teams [21][25][27]. For example, non-native speakers (NNS) of English may need extra time to process information and generate speech, but there are few cues for the native speakers (NS) about the reason for the delay [21]. As a result, a NS may wrongly conclude that more information is required or fail to provide the information necessary to make the message clear. Such problems can be especially severe during multiparty audio conferencing due to the lack of visual cues [14][32]. NNS report that audio conversations are hard to follow when many people are

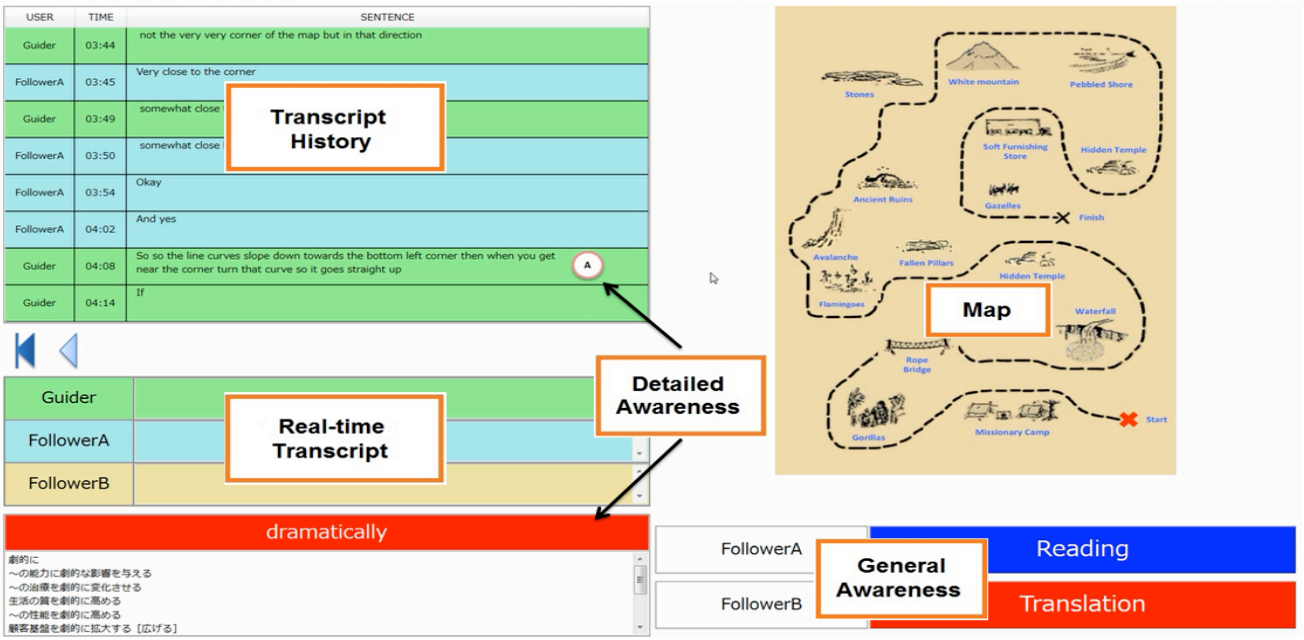

Figure 2. The task interface for NS guider in the detailed awareness display condition. The *general awareness display* (bottom right) signaling whether each follower was reading the automated transcripts and/or translating a word was available to NS guiders in both detailed and general awareness conditions; the *detailed awareness display* showing which line of the transcripts a follower was reading and/or which words he/she was translating was available only in the detailed awareness condition. Followers' interfaces showed the map without the prescribed route, the transcript history and the real-time transcript.

talking, but NSs may have little awareness of NNS's difficulties in understanding [32].

To address these problems, tools such as automated speech recognition (ASR) and translation have been adopted to support multilingual collaboration [14][19][29][32]. This line of work shows the potential of automated transcripts and translation in helping NNS follow group conversations, but overlooks the fact that collaboration is by nature a "joint activity" [7]. Previous work indicates that NS are motivated to make their utterances clearer for NNS addressees [3][14], but little research has been done to help NS assess how well their messages have been understood by NNS partners.

In this paper, we propose a low cost, automated way to signal potential grounding problems in multilingual conversations: providing awareness as to potential points of confusion. Previous studies [14][19][29][32] have shown that when NNS cannot easily express the problems they are experiencing, they try to solve them by checking automated transcripts and/or bilingual dictionaries. These simple behaviors can provide fellow team members with useful cues about points of confusion.

We test this idea in a laboratory study in which triads of two NS and one NNS performed a map navigation task via audio conferencing. The map guiders (all NS of English) were given one of three awareness displays: (a) a *no awareness display* in which the automated transcripts were present but there were no cues about what followers were doing with them; (b) a *general awareness display* that showed whether a follower was looking at the automated transcripts or using the bilingual dictionary; and (c) a *detailed awareness display* that showed where a follower was looking in the transcripts and/or what word he or she was looking up in the bilingual dictionary. We predicted that the more detailed display would facilitate grounding between the NS guider and the NNS follower but would have no effect on grounding between the NS guider and the NS follower.

As predicted, grounding with NNS followers was better when NS guiders had a detailed awareness display. Type of display had no effect on grounding with the NS follower. Guiders with the detailed awareness display were more sensitive to problems encountered by NNS followers, they adjusted their speech to address these problems, and task performance was significantly better than with the other two types of displays. The results add to our understanding of remote multilingual collaboration and suggest new tools to support these collaborations.

RELATED WORK AND HYPOTHESES

In this section, we first review literature on problems of grounding during multilingual collaboration. We then describe how techniques from the literature on awareness displays might be adapted to the context of multilingual communication. We describe how displaying two specific behaviors of followers – use of automated transcripts and looking up words in a bilingual dictionary – might be useful to display to NS guiders and outline our hypotheses.

Problems of Grounding in Multilingual Collaborations

Grounding, or establishing that a message has been understood as intended, is important for successful interpersonal communication [7]. Whether or not a message has been properly grounded is often detected on the basis of cues or evidence received from conversational partners [8]. Partners provide both positive evidence that grounding has occurred (e.g., related next turns, back-channel responses such as "uh huh", or nodding) and negative evidence suggesting grounding has not occurred (e.g., questions or requests for confirmation). When a listener provides little positive or negative evidence of grounding, speakers may have difficulty assessing whether common ground has been established [7].

Differences in fluency may create challenges for the grounding process. For example, research has shown that NNS give fewer responses because of the heavy cognition load of processing a second language [27][28]. In multiparty conversation in particular, NNS may have serious difficulties following others' speech while simultaneously generating their own messages [32]. Further, studies of turn-taking in multilingual groups show that NNS tend to avoid interrupting talk between NSs [25]. Finally, NNS may provide confusing feedback to NS, because they use discourse markers and/or gestures in different ways. For example, NNS use back-channel responses and nodding to express acknowledgement much less often than do NS [13][21][31].

Using Awareness Display to Support Grounding

Much HCI and CSCW research has examined the usefulness of awareness displays for communication and collaboration. This work has shown that awareness of others' activities facilitates a wide range of tasks, including remote assistance (e.g., [2][15][20]), collaborative writing (e.g., [4][11]), collaborative software development (e.g., [5]), collaborative gaming (e.g., [6][24]), and other types of teamwork (e.g., [9][23]). For example, Gergle and colleagues [15] showed that when instruction-givers in a collaborative puzzle task were given a display showing the actions the instruction-follower was taking, they were able to use this awareness information to ground their messages more effectively. They could tell when instructions were misunderstood (i.e., the follower had selected the wrong piece or placed it in the wrong location) or not understood at all (i.e., the follower had not selected a piece) and craft a message to address the problem.

In the case of multilingual conversations, awareness of the NNS' cognitive processes – as evidenced in his/her use of automated transcripts of spoken speech into text and/or bilingual dictionaries – could facilitate grounding by showing the speaker when and what kinds of clarifications are needed. For example, in our study, we ask triads to perform a map task. One NS serves as the "guider" and provides directions; one NS and one NNS serve as "followers" and try to draw the route on their individual maps. When a follower spends longer than usual reading the transcript, this may provide evidence of confusion. When a follower is using the bilingual dictionary, this can provide evidence that some words have not been understood. In this way, the display provides evidence about comprehension without the need for explicit verbal feedback.

Awareness displays can vary in their specificity, ranging from a general impression of what someone is doing to revealing their specific actions. With automated transcripts, for example, we could show general information that the follower is reading the transcripts or specific information about which line he or she is reading. Similarly, we could show that a follower is using the bilingual dictionary or we could reveal which word he or she is looking up.

There are tradeoffs between general and specific displays along a number of dimensions including privacy, cognitive load, and informativeness. For task coordination, several studies have found that abstract displays are better than specific displays (e.g., [9][24]). At the same time, the work on shared visual spaces [15][20] suggests that detailed information about what a partner is doing facilitates grounding. In our study, we expect that detailed displays will be superior to general displays because they pinpoint the nature of the follower's difficulties in understanding messages. Thus, we hypothesize that:

H1a. NS guiders will be able to ground their messages more effectively with NNS followers when they have a detailed awareness display vs. no awareness display.

H1b. NS guiders will be able to ground their messages more effectively with NNS followers when they have a detailed awareness display vs. a general awareness display.

H1c. NS guiders will be able to ground their messages more effectively with NNS followers when they have a general awareness display vs. no awareness display.

In our task, we examine triads consisting of one NS guider, one NS follower, and one NNS follower. While it seems unlikely that native speakers would need to use the automated transcripts or bilingual dictionary, the presence of the display may nonetheless influence the grounding process. Thus we ask:

RQ1: How does type of awareness display affect conversational grounding between NS guiders and NS followers?

Task Performance

Problems in grounding messages in multilingual teams can lead to problems in collaboration and performance. In Li's [22] study, for example, English NS pairs and Chinese-English pairs used English to finish a medical information exchange task. The NS pairs were more successful than the multilingual pairs at exchanging information, in part because Chinese listeners didn't provide feedback when they did not understand their partners. A number of other

studies have found worse task performance in multilingual vs. monolingual teams using a variety of tasks (e.g., [10][26][30]). Thus, by improving the grounding process, we anticipate that awareness displays will improve task performance.

H2a. NNS followers will perform better on the map task when NS guiders have a detailed awareness display vs. no awareness display.

H2b. NNS followers will perform better on the map task when NS guiders have a detailed awareness display vs. a general awareness display.

H2c. NNS followers will perform better on the map task when NS guiders have a general awareness display vs. no awareness display.

In addition, we ask:

RQ2: How does type of awareness display affect NS followers' task performance?

Effects of Awareness Displays on the Guider's Cognitive Load

Both the general and the specific awareness displays add information to the guider's interface that is not present in the no awareness display condition. This might increase the guider's perceived workload because there is more for them to attend to, but it might also reduce workload by making it easier to figure out how to ground messages. Therefore we ask,

RQ3: How does type of awareness display affect NS guiders' self-reported cognitive workload?

METHOD

Overview

We conducted a laboratory experiment in which triads consisting of one NS guider, one NS follower and one NNS follower collaborated on a map task using audio conferencing. We manipulated the type of display provided to the guider using a between-groups design. In the *no awareness display condition*, the NS guider saw the automated transcripts but had no cues about what followers were doing with them. In the *general awareness display condition*, the NS guider saw whether each follower was reading the transcripts and/or translating words. In the *detailed awareness display condition*, the NS guider saw which line each follower was reading in the transcript and/or which English word he or she was translating into Japanese. We measured followers' ratings of the quality of the guiders' instructions, how successfully followers completed the map task, and how much cognitive work load the guiders experienced.

Participants

A total of 57 individuals participated in this study. Of these, 38 (6 female) were monolingual native English speakers who currently live in Japan but grew up in English speaking countries and received education in English. Their mean age was 40.57 years (SD = 10.69).

The rest of the participants (N = 19) were native Japanese speakers (6 female) who grew up in Japan and received education in Japanese. Their mean age was 36.16 years (SD = 9.57). They spoke English as a second language but were not fluent (M = 3.58, SD = .88 on a scale of 1-7).

Participants were randomly assigned to triads consisting of a NS guider, an NS follower, and an NNS follower. Triads were then randomly assigned to display condition (5 no awareness display, 7 general awareness display, 7 detailed awareness display).

Materials

Task. We modified two HCRC map navigation tasks [1] for use in this study. One task was used to introduce participants to the technology; the other was used for the experiment itself. Each HCRC map task consists of a guider map and a follower map. The guider's map shows a prescribed route around a set of landmarks. The original follower's map shows some but not all of the landmarks and has no route on it. We modified the original follower's map to create a second follower map matched in the number and type of differences from the guider map, so that each follower would need to independently ground messages with the guider. The labels of landmarks on one of the follower maps were translated into Japanese by a native Japanese speaker.

Surveys. Participants completed an online pre-experiment questionnaire that collected their demographic information (age, gender, country of birth, native language). This survey also asked participants to rate their English fluency on a 7-point scale (1= not fluent at all, 7= very fluent).

Participants completed a second online questionnaire at the end of the study that was customized to their experimental role. The survey included a manipulation check to confirm which type of awareness display was given to the guider in each group. The follower survey also included four questions asking about the quality of the guider's instructions (e.g., "The guider provided appropriate assistance based on my needs"). Responses were on a 7-point scale (1 = strongly disagree, 7 = strongly agree). The guider survey included four questions adapted from the NASA Task Load Index [17]. Participants rated their mental demand, temporal demand, effort, and frustration during the task on a scale of 1 = low to 7 = high.

Interview. We developed an open-ended interview protocol for NS guiders. We asked them whether they ever adjusted the way they gave instructions based on the follower's response. If they said yes, we followed up with questions about how and why they made these adjustments.

Software and Equipment

Speech recognition tool. Participants' speech was automatically transcribed into text using Dragon Naturally

Speaking (DNS). Participants went through a training session before the study started. The time delay required to generate transcripts was between 1-3 seconds.

Eye tracker. A Gazepoint3 (GP3) Eye Tracker was used to trace followers' eye movements in the automatic transcripts. The accuracy of the GP3 is between 0.5 to 1 degree of visual angle, with a 60 Hz update rate. It can capture 25 cm (horizontal) × 11cm (vertical) of eye movement with ±15 cm range of depth movement. The eye-tracker was positioned on the side of the screen and calibrated for each participant. When a follower was reading the transcript, the information was delivered to the awareness display in less than 0.5 sec.

English-Japanese bilingual dictionary. An electronic bilingual dictionary based on Eijiro Translation was set within the interface for NNS followers, allowing them to translate English to Japanese by clicking on a word in the transcript. The information was delivered to the awareness display with a delay of less than 0.5 sec.

Task interface for the guider. The interface for the guider included 4 main components (Figure 2): the map, the real-time transcript, the transcript history, and the awareness display. The *map* (top right) area showed the landmarks and route that the NS guider had to describe to two followers during the navigation task. The *real-time transcript* (middle left) showed the current utterance being generated by DNS, with a 1-3 second delay. This transcript was replaced by a new transcript as new utterances were spoken. The *transcript history* (top left) showed the full conversation history.

The *awareness display* area (bottom right) varied as a function of awareness display condition. In the *no awareness display* condition, no information about what the follower was reading or translating was shown on the guider's interface. In the *general awareness display* condition, the display showed whether each follower was reading the transcript and/or using the bilingual dictionary. A blue light came on when the eye-tracker recognized that a follower was reading the transcript. A red light came on when the translation module recognized that a follower was translating a word.

In the *detailed awareness display* condition, the display gave further information on how each follower was using transcript (top left) and/or bilingual dictionary (bottom left). Once the eye-tracker recognized that a follower was reading the transcript, the follower's icon showed up on the transcript to indicate which line he/she was reading. Once the translation module on the follower's side recognized that he/she was translating a word, the word showed up in red at the bottom left of the interface.

Task interface for the follower. The interface for the followers included three components: the map with no route information on it, the real-time transcript, and the transcript history. No awareness display was provided to the followers.

Equipment. Participants were seated at Sony Vaio laptops with 1.7 GHz CPU and 4GB memory, equipped with three 27 inch monitors. Participants wore headsets with a microphone to communicate with each other as well as receive instructions from the experimenter.

Procedure

Participants were assigned into triads consisting of a NS guider, an NS follower and an NNS follower. They then went through a 20-minute training session that included speech training for DNS and calibration on the eye tracker.

After the training and calibration, the experimenter introduced the study and explained the task and the interface assigned for their condition (no awareness vs. general awareness vs. detailed awareness). They then completed a 10-minute practice task using one of the HCRC maps to familiarize them with the interface. The NS follower and the NNS follower then collaborated with the same guider at the same time on a second HCRC map task. During the task, they used the same awareness interface they were assigned for the practice task and in the same roles (guider vs. follower).

After completing the map task, participants completed the post task survey and guiders were interviewed about how they coordinated their messages with the needs of the NNS follower. Interviews were transcribed by a native English speaker and imported into NVivo10 for analysis. We then identified interview segments focusing on whether and how guiders tailored their assistance to NNS followers for use in our qualitative analysis.

MEASURES

We collected three types of measures: a manipulation check to ensure guiders were aware of which awareness display they had used, survey items measuring quality of the guider's instructions and cognitive workload, and objective task performance.

Manipulation check. Participants' perception of the type of display they used was assessed by a multiple-choice question asking them to identify all information available on their interface (automated transcripts, automated translation, map, awareness display that shows whether a follower was reading transcripts and/or translating, and awareness display that shows detailed content of what each follower was reading and/or translating).

Quality of guider's instructions. The four questions about the quality of the guider's instructions formed a reliable scale (Cronbach's $\alpha = .84$) and were averaged to create a measure of each follower's perception of the quality of the guider's instructions.

Route accuracy. We scored the route each follower drew using the system developed by Diamant and colleagues [10].

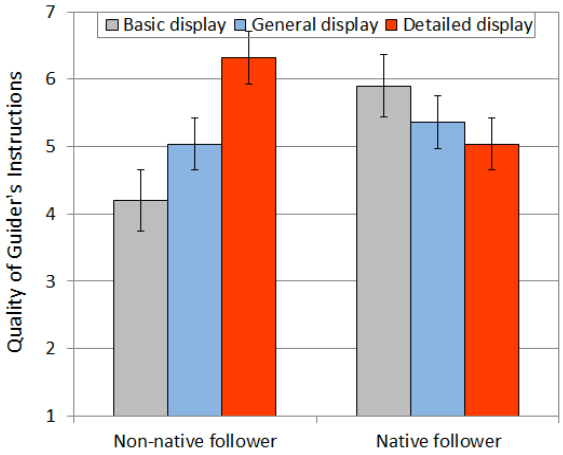

Figure 3. Mean rating of the quality of guider's instructions (on a scale of 1-7) by display type with non-native and native English followers (error bars represent standard errors of the mean)

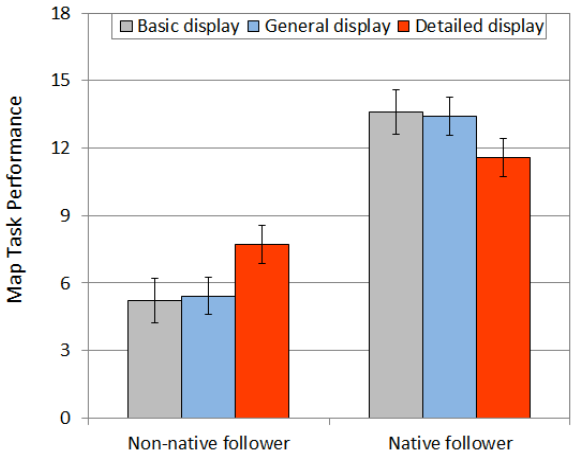

Figure 4. Mean map task performance (on a scale of 0-17) for non-native and native English followers by type of awareness display (error bars represent standard errors of the mean)

Each guider-follower dyad earned one point every time they hit the correct landmark in the right order, but got zero points if they hit a wrong landmark or went to the landmark in the wrong order. Route accuracy was calculated as the total score of hitting correct landmarks.

Guider's workload. The four NASA-TLX questions [17] (mental demand, temporal demand, effort, and frustration) formed a reliable scale (Cronbach's α = .83) and were averaged to create a measure of the guider's workload.

RESULTS

To explore our hypotheses and research questions, we conducted a series of ANOVAS as well as qualitative analysis of the interview data. For quantitative measures collected from the followers, we used 3 (type of the awareness display: basic vs. general vs. detailed) × 2 (language background: NNS vs. NS) Mixed Model ANOVAs, nesting participants within groups. For quantitative measures collected from the guiders, we used a One-Way ANOVA to test the effect of display type. Since demographic variables in all models were not significant, we do not discuss them further.

Manipulation Check

Our manipulation check on the perception of display type showed all participants (100%) correctly perceived the type of display guiders used during the task.

Quality of Guider's Instructions

Our first set of hypotheses is about the quality of guiders' instructions. We hypothesized that NS guiders would be able to ground their messages with NNS followers most effectively when using the detailed awareness display (H1a, H1b) and least effectively when using the no awareness display (H1c). As shown in Figure 3, the results were consistent with H1a and H1b but not H1c.

An ANOVA of the form outlined above showed a significant interaction effect between display type and follower's language background on the rating of the quality of guiders' instructions ($F [2, 32] = 6.29$, $p < .01$; $\omega^2 = 0.23$). The main effects of display type ($p = .47$) and language background ($p = .29$) were not significant.

To test our hypotheses, we ran pairwise comparisons among the three display types separately for NNS followers and NS followers. Consistent with H1a and H1b, NNS followers perceived the guiders' assistance to be better when guiders used the detailed awareness display ($M = 6.32$, $SE = .39$) vs. the general awareness display ($M = 5.04$, $SE = .39$; $F [1, 32] = 5.48$, $p < .05$; Cohen's $d = 3.28$) or the no awareness display ($M = 4.20$, $SE = .46$; $F [1, 32] = 12.43$, $p < .01$; Cohen's $d = 5.39$). However, contrary to H1c, there was no difference between the general and no awareness display conditions ($F [1, 32] = 1.93$, $p = .17$).

RQ1 asked about the effects of the guiders' awareness displays on NS followers. NS followers' ratings of the guiders' instructions did not differ significantly between conditions (for the detailed display, $M = 5.04$, $SE = .39$; for the general awareness display, $M = 5.36$, $SE = .39$; for the no awareness display, $M = 5.90$, $SE = .46$). (For detailed vs. no displays, $F [1, 32] = 2.06$, $p = .16$; for detailed vs. general displays, $F [1, 32] = 0.34$, $p = .56$; for general vs. no displays, $F [1, 32] = 0.81$, $p = .37$.)

Map Task Performance

Our second set of hypotheses addressed task performance. We hypothesized that NNS followers would draw the most accurate paths when guiders had a detailed awareness display (H2a, H2b) and the least accurate paths when guiders had no awareness display (H2c). See Figure 4.

An ANOVA of the form outlined above indicated a significant main effect of the follower's native language on map task performance ($F [1, 32] = 86.33$, $p < .001$; $\omega^2 =$

0.70). NNS followers hit fewer landmarks correctly (M = 6.11, SE = .51) than did NS followers (M = 12.87, SE = .51). The main effect of display type was not significant ($p = .95$), but there was a significant interaction between display type and follower's language background ($F [2, 32] = 4.22$, $p < .05$; $\omega^2 = 0.15$).

Pairwise comparison between the three display types showed that in support of H2a and H2b, the paths drawn by NNS followers were most accurate when guiders had the detailed awareness display (M = 7.71, SE = .84) vs. the general awareness display (M = 5.43, SE = .84; $F [1, 32] = 3.74$, $p = .06$; Cohen's $d = 2.71$) or the no awareness display (M = 5.20, SE = .99; $F [1, 32] = 3.77$, $p = .06$; Cohen's $d = 2.73$). Contrary to H2c, there was no significant difference between the general and no awareness displays ($F [1, 32] = 0.03$, $p = .86$.)

With respect to RQ2, the accuracy of path drawn by NS followers was equally good when guiders had a detailed awareness display (M = 11.57, SE = .84), general awareness display (M = 13.43, SE = .84), or no awareness display (M = 13.60, SE = .99). (For detailed vs. no displays, $F [1, 32] = 2.45$, $p = .13$; for detailed vs. general displays, $F [1, 32] = 2.47$, $p = .13$; for general vs. no displays, $F [1, 32] = 0.02$, $p = .90$.)

Guider's Workload

RQ3 asked how the workload experienced by NS guiders varied as a function of the type of awareness display provided. A one-way ANOVA of display type on self-reported workload showed no significant effect of display type ($F [2, 18] = 0.77$, $p = .48$). There were no significant differences between how much workload guiders reported when using the detailed awareness display (M = 4.29, SE = .47), general awareness display (M = 4.31, SE = .47), or no awareness display (M = 3.50, SE =. 56).

Insights from Interviews with Guiders

After they completed the task, we conducted open-ended interviews with each guider to ask whether and how they adjusted the way they gave instructions based on the followers' responses.

NS guiders reported three issues about coordination management: difficulties interpreting NNS' status from audio feedback, adjusting guidance based on NNS followers' use of the transcripts, and adjusting guidance based on NNS followers' use of the bilingual dictionary.

The first theme that ran through the interviews of all NS guiders concerned their problems understanding how to ground their messages based on NNS followers' verbal feedback. Guiders reported that it was hard to determine whether NNS followers understood an instruction because NNS followers were less vocal than NS followers.

There wasn't a lot of input from the non-native follower, so I was not sure if he is following or not. I assumed that he was following. [G13, no awareness display]

In the beginning I first thought that maybe the non-native follower understood very well and that is why she was quiet, but then I realized that is not true. Her map too is different [from mine], even if she didn't speak. [G17, detailed awareness display]

The second point that emerged from the interviews was that guiders adjusted their instructions based on the NNS followers' use of the automated transcripts. Guiders with both awareness displays reported that they modified the way they gave instructions upon seeing that a NNS follower was reading transcripts, but the way that they did so depended on the type of awareness display. For guiders with general awareness displays, seeing that a NNS follower was using the transcripts seemed to initiate a verbal rephrasing of the last steps presented.

If he was reading something on the transcript, he couldn't understand something [from the audio]. I would reconfirm [the path just described] … So even if he said ok, I made sure that it is ok. [G18, general awareness display]

When I tried to explain what my map looked like, the native speaker responded quickly. The Japanese speaker needed time to process the information, so I wasn't sure if he understood what I said... or not. So rather than asking "do you understand, do you understand?" I would check the status of the [NNS follower's] reading, and also I was saying different words in case the transcript didn't work correctly. Hoping that he could pick it. [G2, general awareness display]

In contrast, with detailed awareness displays, guiders were able to make adjustments to their speech that took into account the specific utterances that the NNS followers found problematic.

When the non-native follower was reading transcript, I was slowing down and clarifying … [based on] his physical position on the transcript, and if he was behind the native speaker. I think I also probably used more words, for example northwest or top-left, to give [the speech recognition] more choice to transcribe the words [G9, detailed awareness display]

I did try to make use of the reading information, to track, to clarify. [G11, detailed awareness display]

Guiders in both the general and detailed awareness display conditions also reported that they made adjustments to their instructions based on NNS followers' use of the bilingual dictionary. However, the strategies they used differed between display conditions. With the general display, they couldn't pinpoint the exact word that was unclear, but they were sensitive to the possibility that translating would take some time and slowed their speech accordingly:

When the red button is on, I slowed down or used a different word. … if I say a sentence and notice the red color, I just say again in case there is any difficulty. [G1, general awareness display]

When the non-native speaker was translating ... I would slow down a little bit, because I felt that it was time consuming. [G4, general awareness display]

With the detailed display, guiders identified specific words that were problematic for their NNS followers and tried to use alternative words in their future utterances. In some cases such as G10 below, guiders used the absence of translation as a cue that the NNS follower knew a term.

I check the translation to see what words he had hard time with. Like, I tried not to use "diagonal", because I didn't think he knew that word, he translated it. So I just tried to use other words like "curvy". I didn't know if he knew curvy, but he didn't look it up, so maybe he knew. [G10, detailed awareness display]

First time I saw he couldn't understand "vertical" and from that on I tried to use "vertical" less and use north/south lines instead. So I try to change my vocabulary, although I am sure I still used vertical a few times but I tried to use it less. [G12, detailed awareness display]

Finally, we asked guiders whether they attended to the NS followers' use of transcripts and all reported that they did not because they could communicate successfully with the NS follower through audio communication alone.

DISCUSSION

In this study, we provided participants three types of displays that delivered different amounts of awareness information. Our findings indicate that providing NS guiders with detailed awareness displays best supports the grounding between the guiders and their NNS followers. In the rest of this discussion, we connect our study with previous work on awareness and grounding. Specifically, we consider cues used by NS guiders to support grounding with NNS followers, and trade-offs between general and detailed displays.

Cues Used by NS Guiders to Support Grounding

Both ratings of the guiders' instructions and NNS followers' task performance show that type of awareness display affects conversational grounding between NS guiders and NNS followers. Specifically, a detailed awareness display led to better communication and performance than either a general awareness display or no awareness display.

Our three types of displays intentionally provided different types of cues that could be useful to the grounding process. The no awareness display was similar to regular audio conferencing. The general awareness display offered additional cues of whether the follower was reading transcripts and/or translating words. The detailed display further indicated which line of transcripts the follower was reading and/or which words the follower was translating from the bilingual dictionary.

From the interviews, it is clear that guiders believed that the effectiveness of their instruction improved as the richness of the awareness information increased. NS guiders using the no awareness display reported that they tried to guess the NNS follower's level of understanding based on limited audio feedback. Given that the NNS followers often provided little verbal response, guiders may not have been able to accurately detect when to slow down and/or how clarify their instructions.

The general awareness display showed whether the NNS follower was requesting language-related help from the system, and the guiders believed that this information helped them adjust their instruction. However, as can be seen from the followers' ratings of the guiders' instructions, which showed no significant difference between this display condition and the no display condition, their adjustments probably did not meet the needs of the NNS followers. The interviews provide some suggestions why this might be the case. For example, guiders assumed that if a NNS follower was reading the transcript, he or she was reading the last instruction. But this would not address actual problems in grounding if the NNS followers had fallen behind in the conversation. In future work, we plan to examine the accuracy of speakers' assumptions about the grounding process as a function of different types of awareness cues.

The detailed awareness display provided concrete information that was directly linked to the on-going grounding process. This display serves similar functions as the shared visual space examined by Gergle and colleagues [15]. However, it's worth noting that displaying how NNS followers use transcripts and/or a dictionary may bring extra support compared to a display of shared visual space. In our task, for example, giving the guiders a view of the NNS follower's path on the map might also have helped them identify grounding problems, though a shared view alone is less clear as to which part of the message needs clarification (e.g., which words were misunderstood).

Trade-offs between General and Detailed Displays

Our results suggest that at least for audio conversations between native and non-native English speakers, grounding and performance will be better with detailed awareness displays. This finding is consistent with some previous work, such as that of Gergle and colleagues [15]; but contrasts with other work, such as that of Dabbish and Kraut [9]. It is therefore important to examine the trade-off between using general and detailed awareness displays.

The trade-offs between general and detailed awareness displays revolve around two issues: a) what tasks the display was intended to support and b) the richness of information required to accomplish that task. In both this study and Gergle and colleagues' study [15], the central purpose of collaboration is for an information-giver to convey instructions to an information-receiver and ensure

that grounding has occurred. To accomplish such tasks, the information-giver needs cues that are rich enough to infer whether the information-receiver has understood the message, and if not, what kind of grounding problem has occurred. For situations in which grounding is primary, our results suggest detailed awareness displays may be best.

In other settings, the primary goal may be something other than conversational grounding, for example, appropriate timing of interruptions [9]. To accomplish this type of task, it may be more important for the interrupter to know whether the interruptee is busy or not than to know precisely what he or she is doing. Thus, there is little benefit to providing the additional information in a detailed awareness display.

An additional factor to consider is people's privacy preferences. For example, some people may not want others knowing precisely which words they need to look up but be more open them knowing that they are using a dictionary. We plan to explore this issue in our future work.

DESIGN IMPLICATIONS

Our findings suggest several ways that collaboration tools might be enhanced to facilitate the detection, repair, and prevention of grounding problems in multilingual teams.

Displaying NNS's Behaviors for Detecting Grounding Problems

The most direct design implication is that designers of tools to support multilingual collaboration should include detailed awareness displays. Guiders relied on these detailed displays to detect and repair grounding problems, and NNS followers reported that the instructions were better in the detailed display condition than in the general or no awareness display condition. For collaboration systems that have speech recognition and/or bilingual dictionary modules embedded in them already, it should be relatively straightforward to implement detailed awareness displays. For other types of tasks and systems, designers may need to create new mechanisms for detecting and displaying awareness of NNS difficulties in grounding messages.

Helping NS Interpret NNS's Behaviors

Further, the NS guiders' descriptions of how they used the awareness display indicate a design space for facilitating the repair of grounding problems. When the display showed the NNS follower was translating a word, for example, NS guiders reported that they would rephrase the instruction with what they thought was simpler word. However, it might actually be more effective for the NS to keep using the same word, given that NNS already looked up what it meant. In other words, NS's adjustments and NNS's needs may not match.

One way to address this mismatch between the grounding needs of NNS listeners and NS' perceptions of these needs is for the system to help interpret the data collected by awareness tools and generate recommendations to the NS about how to clarify their messages, perhaps on data from the conversation itself or from theories of second language learning. An automated system of this type might also help address possible privacy concerns caused by directly sharing NNS behaviors with their NS partners.

Learning from Different NNS's Behaviors to Prevent Grounding Problems

A related recommendation is for systems to collect general information about potential points of confusion in the grounding process, including specific words or grammatical structures that are difficult for non-native speakers to understand. For example, if a system is used by many different multilingual collaborations, it could collect data on which words are often looked up in the bilingual dictionary and which are not, or which kinds of messages NNS listeners spend a long time looking at in the automated transcripts and which they never look at. Based on this data, system might both provide suggestions to the NS about how to better phrase his or her messages and automatically provide translations to the NNS of words and phrases that are frequently problematic for other NNS.

CONCLUSION

We conducted a laboratory experiment in which NS and NNS of English collaborated via audio conferencing on a map navigation task using one of three awareness displays (no awareness vs. general awareness vs. detailed awareness). We found that NS guiders and NNS followers collaborated most successfully using the detailed awareness display. Specifically, NS guiders grounded their messages with NNS followers most effectively when the detailed awareness display was provided. NNS followers also performed best on the map task with the detailed awareness display. In interviews, NS guiders stated that they modified instructions to make grounding easier based on the cues they received about NNS' problems grounding the messages. Our findings suggest several ways that future collaboration tools might better support multilingual teams.

ACKNOWLEDGMENTS

This research was funded in part by National Science Foundation grant #1318899. We thank Leslie Setlock for editorial assistance and the NTT development team for their technical support.

REFERENCES

1. Anderson, A., Bader, M., Bard, E., et al. (1991). The HCRC Map Task Corpus. *Language and Speech, 34,* 351-366.
2. Bardram, J. E., Hansen, T. R., & Soegaard, M. (2006). AwareMedia- A shared interative display supporting social, temporal, and spatial awareness in surgery. *Proc. of CSCW 2006,* 109-118.

3. Beebe, L.M., & Giles, H. (1984). Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language, 46,* 5-32*,*
4. Birnholtz, J., & Ibara, S. (2012). Tracking changes in collaborative writing: Edits, visibility, and group maintenance. *Proc. of CSCW 2012,* 809-818.
5. Carter, J., & Dewan, P. (2010). Are you having difficulty? *Proc. of CSCW 2010,* 211-214.
6. Cheung, V., Chang, Y. L. B., & Scott, S. D. (2012). Communication channels and awareness cues in collocated collaborative time-critical gaming. *Proc. of CSCW 2012,* 569-578.
7. Clark, H.H. (1996). *Using Language*. Cambridge Press.
8. Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley, (Eds.). *Perspectives on Socially Shared Cognition* (pp. 127-149). APA Press.
9. Dabbish, L., & Kraut, R. (2008). Awareness displays and social motivation for coordinating communication. *Information Systems Research, 19,* 221-238.
10. Diamant, E.I., Fussell, S. R., & Lo, F. L. (2008). Where did we turn wrong? Unpacking the effects of culture and technology on attributions of team performance. *Proc. CSCW 2008*, 383-391.
11. Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. *Proc. of CSCW 1992,* 107-114.
12. Feely, A. J., & Harzing, A. W. K. (2003). Language management in multinational companies. *Cross-Cultural Management: An Intl'l Journal, 10,* 27-52.
13. Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics, 28,* 410-439.
14. Gao, G., Yamashita, N., Hautasaari, A., Echenique, A., & Fussell, S. R. (2014). Effects of public vs. private automated transcripts on multiparty communication between native and non-native English speakers. *Proc. of CHI 2014,* 843-852.
15. Gergle, D., Kraut, R. E., & Fussell, S. R. (2013). Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction, 28,* 1-39.
16. Gibson, C. B., & Gibbs, J. L. (2006). Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. *Administrative Science Quarterly, 51,* 451-495.
17. Hart, S. G., & Staveland, L. E. (1998). Development of NASA-TLX: Results of empirical and theoretical research. *Advances in Psychology, 52,* 139-183.
18. Henderson, J.K. (2005). Language diversity in international management teams. *International Studies of Management and Organization, 35,* 66–82.
19. Ishida, T. (2011). *The language grid*. Springer Press.
20. Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human Computer Interaction, 18,* 13-49.
21. Li, H. Z., Yum, Y., Yates, R., Aguilera, L., Mao, Y., & Zheng, Y. (2005). Interruption and involvement in discourse: Can intercultural interlocutors be trained? *J. Intercultural Communication Research, 34,* 233-254.
22. Li, H. Z. (1999). Grounding and information communication in intercultural and intracultural dyadic discourse. *Discourse Processes*, *28*, 195-215.
23. Lim, B. Y., Brdiczka, O., & Bellotti, V. (2010). Show me a good time: Using content to provide activity awareness to collaborators with Activity Spotter. *Proc. of GROUP 2010,* 263-272.
24. Reynolds, L., Birnholtz, J., & Lee, A. (2012). The effect of communication channel and visual awareness display on coordination in online tasks. *Proc. iConference 2012*, 120-128
25. Rogerson-Revell, P. (2008). Participation and performance in international business meetings. *English for Specific Purposes, 27,* 338-360.
26. Setlock, L. D., Fussell, S. R., & Neuwirth, C. (2004). Taking it out of context: Collaborating within and across cultures in face-to-face settings and via instant messaging. *Proc. CSCW 2004,* 604-613.
27. Takano, Y., & Noda, A. (1993). A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology, 24,* 445-462.
28. Tange, H., & Lauring, J. (2009). Language management and social interaction within the multilingual workplace. *J. Communication Management, 13,* 218-232.
29. Wang, H. C., Fussell, S. R., & Cosley, D. (2013). Machine translation vs. common language: Effects on idea exchange in cross-lingual groups. *Proc. of CSCW 2013,* 935-944.
30. Wang, H-C., Fussell, S. R. & Setlock, L. D. (2009). Cultural difference and adaptation of communication styles in computer-mediated group brainstorming. *Proc. CHI 2009,* 669-678.
31. Wong, J. (2000). The token" yeah" in nonnative speaker English conversation. *Research on Language and Social Interaction, 33,* 39-67.
32. Yamashita, N., Echenique, A., Ishita, T., & Hautasaari, A. (2013). Lost in translation: How transmission lag enhances and deteriorates multilingual collaboration. In *Proc. of CSCW 2013*, 923-934.