Effects of Public vs. Private Automated Transcripts on Multiparty Communication between Native and Non-Native English Speakers

Ge Gao[1,2], Naomi Yamashita[1], Ari Hautasaari[1], Andy Echenique[1,3], 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

[3]Department of Informatics
University of California, Irvine
Irvine CA 92697 USA
echeniqa@uci.edu

ABSTRACT

Real-time transcripts generated by automated speech recognition (ASR) technologies have the potential to facilitate communication between native speakers (NS) and non-native speakers (NNS). Previous studies of ASR have focused on how transcripts aid NNS speech comprehension. In this study, we examine whether transcripts benefit multiparty real-time conversation between NS and NNS. We hypothesized that ASR transcripts would be more beneficial when the transcripts were publicly shared by all group members as opposed to when they were seen only by the NNS. To test our hypothesis, we conducted a lab experiment in which 14 groups of native and non-native speakers engaged in a story-telling task. Half of the groups received *private transcripts* that were available only to the NNS; the other half received *publicly shared* transcripts that were available to all group members. NS spoke more clearly, and both NS and NNS rated the quality of communication higher, when transcripts were publicly shared. These findings inform the design of future tools to support multilingual group communication.

Author Keywords

Automated speech recognition; real-time transcripts; multilingual communication

ACM Classification Keywords

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

INTRODUCTION

One of the biggest challenges faced by global organizations is how to bring people with different native languages together to work on common problems [6, 22]. Multilingual teams often use English as a common language (*lingua franca*) to communicate between team members [9]. However, since non-native speakers (NNS) often don't reach the fluency level of native speakers (NS) of the common language, they often encounter interactional problems that are rarely found in communication between NSs [12]. For example, NNS often have difficulties following audio conferencing conversations in their second language [7, 28].

Automated speech recognition (ASR) technologies have the potential to alleviate some of the difficulties NNS face when using a common language. Indeed, a small but increasing number of studies have shown that real-time transcripts of an ongoing speech generated by ASR improved NNS's comprehension when provided with reasonable accuracy and time delay [e.g.,18, 19, 23]. In general, however, these studies have looked at non-interactive communication. For example, in Pan et al's [18, 19] and Shimogori et al's [23] studies, NNS engaged in listening comprehension tasks with pre-recorded English speech, conversations and/or lectures given by NS. The ASR transcripts were generated in advance, in order to control the accuracy and latency of the transcripts. In addition, participants didn't need to generate responses on the basis of the speech transcripts. It is thus unclear whether ASR will facilitate conversational interaction.

In this paper, we examine whether and how the use of automated transcripts affects real-time group communication between NS and NNS. We examine two types of transcripts: private transcripts that are shown only to the NNS and publicly shared transcripts that can be viewed by both NS and NNS. Unlike private transcripts, public transcripts can affect how NS speak because they can see how their words are transcribed. We also provided the capability for volunteer editing of transcripts in order to explore how people might use the transcripts in active and flexible ways. These three features of real time interaction, public vs. private transcripts, and optional editing create a new space for studying the effect of automated transcripts on communication between NS and NNS.

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.

http://dx.doi.org/10.1145/2556288.2557303

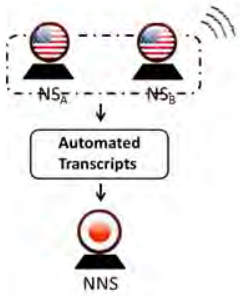
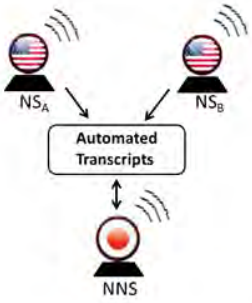
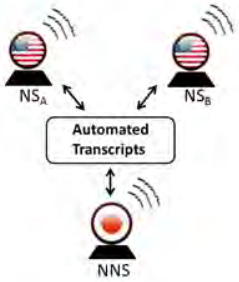
	Previous Studies	Private Transcripts Condition	Public Transcripts Condition
Experimental set up			
Nature of Communication	Pre-record audio speech	Real time conversation	Real time conversation
Accessibility of the automated transcripts	NNS only	NNS only	Everyone

Table 1. Experimental set-up, nature of communication, and accessibility of automated transcripts in previous studies and in the two conditions of the current study. The diagram on the left represents previous studies that examined how transcripts helped NNS achieve listening comprehension on English conversation between NSs, the diagram in the middle represents the private transcripts condition of this study in which only NNS could access to the transcripts, and the diagram on the right represents the public transcripts condition of this study in which both NNS and NSs could access to the transcripts.

The remainder of this paper presents a laboratory study that examined whether and how automated transcripts influence group communication. Fourteen triads of participants, each consisting of two monolingual native English speakers and one native Japanese speaker, who also spoke English as a second language, performed a story-telling task. We manipulated the accessibility of transcripts between groups: In the *private* condition automated transcripts were provided only to the NNS whereas in the *publicly shared* condition automated transcripts were given to all group members. In both conditions, those participants who could see the transcripts could also edit them. Participants assessed their workload and the quality of the story they created after the task. The clarity of speech and number of edits were also calculated based on the system recording.

Our results showed that publicly shared transcripts affected people's communication behaviors and the quality of the group conversations. NS spoke more clearly (e.g., at a slower pace with greater articulation) in the publicly shared transcripts condition than in the private transcripts condition. Also, they sometimes manually corrected the errors in the shared automated transcripts. Both NS and NNS participants rated the quality of group communication significantly higher in the publicly shared transcripts condition than in the private transcripts condition. Insights gained from this study provide better understanding of multiparty group communication between NS and NNS and several design implications for future ASR-based communication tools.

BACKGROUND AND RELATED WORK

In this section, we first review previous work on how NS and NNS communicate and how automated transcripts facilitate NNS's comprehension of second language speech. We then describe how automated transcripts might influence communication in groups containing both NS and NNS. Finally, we present the main hypotheses and research questions of the current study.

Difficulties in Communication between NS and NNS

Previous work has shown that although using a common language makes it possible for people with different native languages to communicate, issues of language fluency may decrease the efficiency of group communication [e.g., 13, 27, 29]. Requiring everyone to use a common language also brings certain difficulties to NNS. For example, processing a second language can increase the cognitive load for the NNS [24]. As a consequence, NNS need more time to understand the NS's message as well as organize their own expressions [14]. These processes may be even harder for NNS under compromised communication situations, such as audio conferencing with unclear pronunciations and/or extraneous noises [15, 16].

The problems faced by NNS may be alleviated when NS coordinate and adjust their speaking behavior to the NNS [2, 3], for example by speaking more slowly or enunciating more clearly. The behavioral changes of NS can improve NNS's message understanding and subsequent ability to contribute to the dialogue, thereby benefitting the overall quality of group communication.

Publicly Shared Transcripts and Group Communication

Automated transcripts used in multilingual group communication are designed to help NNS overcome limitations in oral speech comprehension. Previous studies [18, 19, 23, 28] indicated that information given by textual transcripts and audio speech can complement each other, which can improve NNS's comprehension. However, it's worth noting that transcripts used in previous work were shown to NNS only in mono-directional communication scenarios. Transcripts were generated beforehand and controlled in their accuracy and amount of delay. The results of these studies are useful for understanding the role of ASR technologies in such settings as formal presentations or television shows, but their applicability to real-time interactive dialogue is less clear. Real-time dialogue provides challenges for ASR: Transcripts can rarely be generated with perfect accuracy and some time delay is often involved. As noted by Pan [18], tracking transcripts with a certain amount of errors could be distracting and difficult for NNS. The problems created by inaccurate and delayed transcripts may make these less useful to NNS in interactive dialogue than they are in non-interactive settings.

One advantage of interactive settings, however, is that in real-time dialogue the transcripts could be provided publically, to both the NNS and NS members of a group. For the NNS, the transcripts can aid in comprehension. For the NS, the transcripts may help them discover sources of miscommunication and change the way they speak to avoid similar errors in later generated transcripts. Based on this rationale, we hypothesized that:

H1. NS will speak more clearly when transcripts are publicly shared by everyone rather than shown only to NNS.

These possible interactions between NS and the automated transcripts can benefit NNS and lead to better quality group communication. As previous studies have suggested [e.g., 11, 20, 21], the quality of communication between NS and NNS tends to improve as the clarity of NS's speech goes up. Thus, we posed two related hypotheses:

H2. Both NS and NNS will experience better quality of group communication when transcripts are publicly shared by NS and NNS rather than shown only to NS.

H3. NNS's experience of the quality of group communication will be positively correlated with NS's speech clarity.

Given that NNS see the transcripts in both conditions, and they are also dealing with fluency issues (which increase the baseline cognitive load), it is unclear whether the publicly shared transcripts condition will affect their speech clarity. However, the process of receiving edits from partners may still prove somewhat beneficial. Therefore, we posed the following research question:

RQ1. Does the speech clarity of NNS vary between private and publicly shared transcripts conditions?

Publicly Shared Transcripts and Workload

Giving real-time transcripts during the conversation increases the potential workload for communicators. For both NS and NNS, there are multiple types of information, through multiple channels, that they may need to process at the same time, including audio speech from others, oral speech from oneself, and visual transcripts from everyone. However, since communicators have flexibility in allocating their cognitive resources when multitasking, it's hard to predict how NS and NNS's workload will vary between the publicly shared and private transcripts conditions. Thus, we posed the following research question:

RQ2. Does the workload of NS and NNS vary between private and publicly shared transcripts conditions?

Publicly Shared Transcripts and Voluntary Editing

Finally, we also wondered whether and how NS and NNS participants would use the optional function of transcript editing to facilitate their group communication. We therefore posed the following research question:

RQ3. Do NS and NNS correct errors in the transcripts when they have access to them? For NNS, does this editing behavior differ between private and publicly shared transcripts?

METHOD

Overview

We conducted a laboratory experiment with a single factor (accessibility of the transcripts: private vs. publicly shared) between subjects design. Fourteen groups participated in a story-telling task. Each group consisted of 2 monolingual native English speakers and 1 Japanese/English bilingual native Japanese speaker. All participants were required to use English as a common language and work together to create a coherent story. The automated transcripts were provided to all groups but in different ways. For groups in the *private transcripts* condition, only the NNS member of the group could see the transcripts (middle column of Table 1). For groups in the *publicly shared transcripts* condition, both the NNS and NS group members could see the transcripts (right column of Table 1). Participants were not required to edit the transcripts, but they could edit them if they wanted. When the story-telling task was complete, participants answered a post-experiment survey about their mental workload and the quality of their group communication.

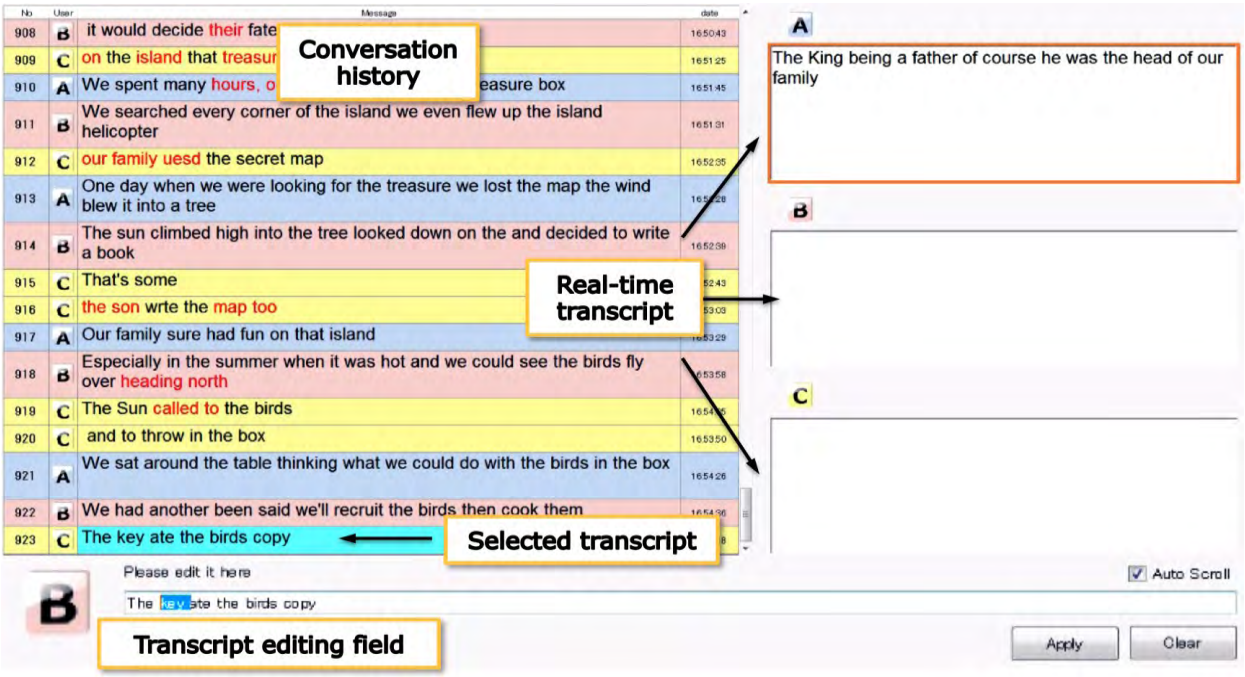


Figure 1. The transcripts tracking interface for the NNS in the private transcripts condition and for all participants in the publicly shared transcripts condition (in the current snapshot, the conversation history shows transcripts of all speech given by A, B and C; the real-time transcripts shows the on-going speech of A, but this transcripts will disappear from the real-time box and go to the conversation history box when A's speech is finished; the selected transcript shows the transcripts of C's speech which is edited by B in the transcript editing field; the edited parts in C's transcripts will be marked in red later).

Participants

The study involved a total of 42 participants. Among them, 28 of the participants (4 female) were native monolingual English speakers who currently live in Japan but grew up in English speaking countries and received education in English. Their mean age was 42.85 years (SD = 11.68). They reported having little previous experience using ASR (M = 3.29, SD = 1.90 on a 7-point scale ranging from 1 = never to 7 = very often). They were all experienced in intercultural communication using English as a common language (M = 6.50, SD = 1.04 on a 7-point scale ranging from 1 = never to 7 = very often).

The rest of the participants (N = 14) were bilingual native-Japanese speakers (14 female) who grew up in Japan and received education in Japanese. Their mean age was 39.29 years (SD = 10.93). None of these participants had lived in English speaking countries for more than 2 years. Their English proficiency level was relatively high within the Japanese population but they did not identify themselves as fully fluent (M = 3.71, SD = 0.72 on a 7-point Likert scale; 1 = not fluent at all, 7 = very fluent). They reported having little previous experience using ASR (M = 1.86, SD = 1.10) and conducting multilingual communication using English as a common language (M = 2.71, SD = 1.20).

Software and Equipment

Speech recognition tool. Transcripts used in this study were generated by ASR technology. According to previous research [19, 28], ASR generated transcripts can benefit multilingual communication when the word error rate (WER) of the transcripts is below 20% and their time delay is less than 4 seconds.

Dragon Naturally Speaking (DNS) [8] was used to recognize speakers' speech and transcribe it into English text. DNS is one of the most popular speech recognition technologies used worldwide. The optimal WER of DNS is below 10% [8]. To get the most out of DNS, participants need to go through a training session before the formal speech recognition starts. During the training session, participants are required to read aloud some English materials provided by DNS. By doing so, DNS learns the way participants speak and automatically adjusts its recognition results to accommodate the participants' speech. The time delay required to generate transcripts is between 1-3 seconds [8].

Transcripts tracking interface. Transcripts generated by ASR were transferred into an interface we designed for this study (see Figure 1). In the private condition, only the NNS participant in each group could see this interface with full transcripts of their group conversation. In the publicly shared transcripts condition, however, this transcripts tracking interface was accessible to both NS and NNS.

The transcripts tracking interface included 3 main components: the real-time transcript, a conversation history, and a transcript editing field. In the private transcripts

condition, only the NNS in each group could see the transcripts tracking interface. NS participants knew their NNS group members could see the transcripts, but the screen they faced was blank. In the publicly shared transcripts condition, both NNS and NS could see the same transcripts tracking interface shown on the computer screen.

The *real-time transcripts* component (top right side of figure1) showed transcripts generated by DNS-11 with a 1-3 seconds delay. Participants could track "who is speaking what" in the streaming mode. The transcript of one speech will disappear and be replaced by the next transcript when new speech comes out.

The *conversation history* (top left side of Figure 1) shows the full transcripts of the group conversation. Transcripts that disappear from the real-time field will go to this history field. Participants can drag the scroll bar and see "who spoke what at which time" during the whole conversation.

The *transcript editing field* (bottom side of Figure 1) provides an optional function that allows participants to edit any part of the transcript. Editing on the transcripts was allowed but not mandatory. The edited transcripts, if any, were again shared by everyone. If participants notice an error in the transcripts and want to edit it, they can use the mouse to select the sentence. The selected sentence will go to the editing field on the bottom; meanwhile, the original sentence in the conversation history will turn blue to indicate that the sentence is being edited. After the editing is done, participants can press enter or click the *apply* button on the bottom right to send the edited sentence back to the conversation history and share it with the other participants. Words and/or sentences that have been edited in the transcripts will be marked in red automatically.

Equipment. All participants were seated at Sony Vaio laptops with 1.8 GHz CPU, 8GB memory and 15.5 inch monitors in a separated room. They wore headsets with a microphone during the study to communicate with each other as well as receive instructions from the experimenter.

Task and Procedures

Participants were assigned into groups consisting of two native English speakers (NS) and one native Japanese speaker (NNS). Before the experiment began, participants went through the speech training in DNS, so that the speech recognition software could recognize their voice with optimal accuracy. The duration of this training varied between 10-20 minutes, depending on the vocal volume, articulation, and accent of each individual speaker.

After the training session, participants were presented with a list of keywords randomly selected from the "1000 most frequent words" provided by the British National Corpus [17]. For the main part of the experiment, participants were asked to participate in a story-telling task in which they built a coherent story together using the keywords they received. To build a coherent story, participants needed to understand others' story lines and also make their own story lines clear to the others. In this task, to make both NS and NNS speak equally, the three participants in a same group were required to speak in turns, following the order of "A→ B → C → A → …". On their own turns, participants were required to use at least one keyword (out of three) in their own list, and create a sentence based on previous story lines given by other group members.

A five-minute practice task was conducted before the main task to familiarize participants with the task and the experiment system. The main task lasted for 10 minutes. Keywords used in the main task were identical for all the groups (see Table 2).

As noted earlier, we manipulated the accessibly of transcripts during the experiment. All 14 groups were randomly assigned to either the private transcripts condition or the publicly shared transcripts condition. After completing the main task, each participant rated his/her workload and the quality of their group communication during the experiment. After the entire session, we conducted open-ended interviews about participants' experiences during the session. Interviews were conducted in each participant's native language.

Speaker A	Speaker B	Speaker C
Tree	Wheel	Map
Family	Island	King
Table	Fly	Box

Table 2. Keywords given to each speaker in the main task

MEASURES

We collected two types of measures: objective measures of participants' communication behavior that were reflected in the transcripts given by the experiment system, and subjective measures of communication experience that were self-reported by participants.

Objective Communication Behavior

Speech clarity. The clarity of articulations was measured by calculating the Word Error Rate (WER) of the original transcripts. WER is calculated by comparing random samples of transliterated audio excerpts to corresponding automatic transcripts. It takes into account the number of substitutions, deletions and insertions needed to match the reference sentence to the ASR result. For ideal transcripts with no errors, the WER will be 0; as the accuracy of the transcripts goes down, the WER value rises. The WER of transcripts reflects the overall clarity of articulation based on a mixture of a set of speaker characteristics, including tone, speaking speed and accent [18]. The lower the WER, the clearer the articulation.

Transcript editing. Editing of the transcripts could be recognized by red marks in the final transcripts. We calculated the number of edits done by NNS and NS during

the task. All edits observed and analyzed were corrections of transcripts errors rather than extra notes.

Subjective Communication Experience

Workload. Participants' subjective workload during the task was measured using three 7-point Likert scales adapted from the Task Load Index (TLX [10]) ("How much mental and perceptual activity you felt was required", "How much time pressure you felt due to the rate or pace at which the tasks or task elements occurred", and "How much mental and physical effort you had to make to accomplish your level of performance," 1 = low, 7 = high). The questions formed a reliable scale (Cronbach's α = .83) and were averaged to create a measure of workload.

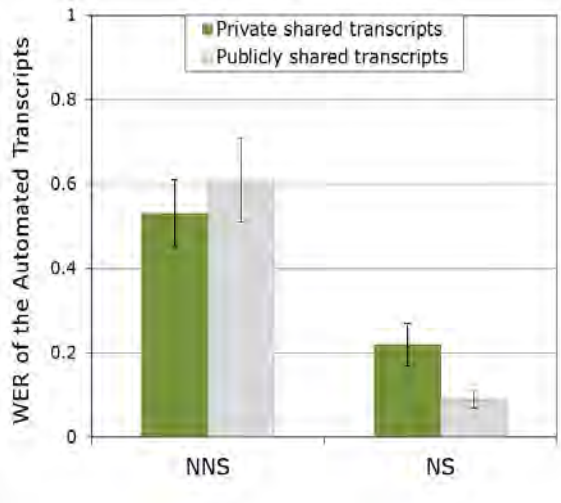


Figure 2. Mean WER of the automated transcripts by accessibility condition for NNS and NS (error bars represent standard errors of the mean).

Quality of group communication. Each participant's perception of the quality of group communication was measured using three 7-point Likert scales ("I felt the story we just built flowed well", "I felt we built upon each other's story line", and "We successfully built up a coherent story", 1 = strongly disagree, 7 = strongly agree). The questions formed a reliable scale (Cronbach's α = .94) and were averaged to create a measure of quality of group communication.

Manipulation Checks

Language proficiency of group members. Participants' perception of the language proficiency of their group members was assessed by a single-choice question. This question asked them to indicate whether they were speaking to two native speakers, two non-native speakers, or one native speaker and one non-native speaker.

Accessibility to automated transcripts Participants' perception of the accessibility of the transcripts was assessed by a single-choice question asking what communication medium each participant was using during the task (audio conferencing only, automated transcripts only, or audio conferencing with automated transcripts).

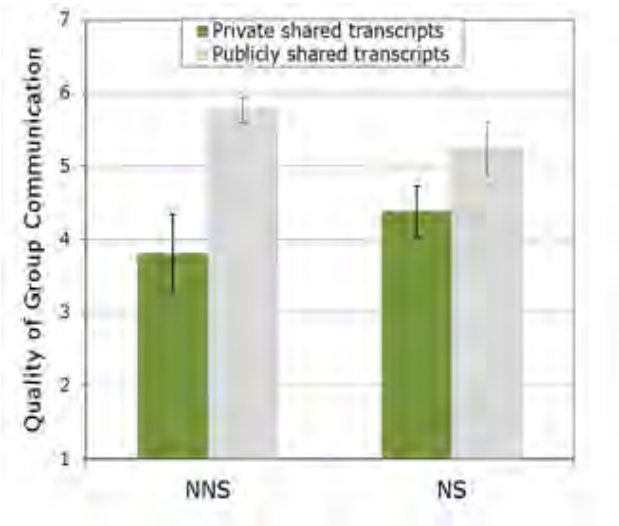


Figure 3. Mean quality of group communication by accessibility condition for NNS and NS (error bars represent standard errors of the mean)

RESULTS

To explore our hypotheses and research questions, we conducted 2 (transcripts accessibility: private vs. publicly shared) × 2 (language background: NNS vs. NS) Mixed Model ANOVAs. Participant was nested within groups. Transcripts accessibility and language proficiency were set as independent fixed variables. The demographic backgrounds (e.g., age and gender), previous experience on using ASR, and previous experience on multilingual communication in English of each participant were set as control variables in all the models. Since the effects of the control variables were generally not significant, we do not discuss them further.

Manipulation Checks

Our manipulation checks on the perception of partners' language proficiency and transcripts accessibility showed that both manipulations were successful. All participants (100%) correctly perceived the language level of their group members. All participants (100%) also correctly perceived their accessibility to the transcripts.

Group Communication

H1-H3 hypothesized that NS will generate clearer speech in the publicly shared rather than private transcripts condition. The quality of group communication will also be improved in the publicly shared condition.

Speech clarity. To explore H1 and RQ1, we conducted a 2 × 2 Mixed Model ANOVA on the WER of transcripts. Lower WER scores indicate that a speaker is speaking more clearly. The results fully supported H1. There was a significant main effect of language background on the WER

(F [1, 27.74] = 11.92, p = .002), which indicated that the WER of NS (M = .16, SE = .03) was significantly lower than the WER of NNS (M = .56, SE = .06; see Figure 2). This main effect was further qualified by an interaction between transcripts accessibility and language background (F [1, 20.74] = 3.51, p = .04). Consistent with H1, NS's WER was significantly lower in the public transcripts condition (M = .09, SE = .02) than in the private transcripts condition (M = .22, SE = .05): F [1, 8.50] = 3.48, p = .04. That is, NS articulated their messages more clearly in the publicly shared transcripts condition than in the private transcripts condition.

With respect to RQ1, NNS's WER didn't show significant change between the publicly shared transcripts (M = .61, SE = .10) and the private transcripts conditions (M = .53, SE = .08): F [1, 11] = 0.41, p = .54. That is, NNS's speaking clarity did not change between the two conditions.

Quality of group communication. H2 addressed how transcripts condition affected perceived quality of group communication. To test H2, we conducted a 2 × 2 Mixed Model ANOVA analysis on the self-reported quality of group communication (see Figure 3). Consistent with H2, there was a significant main effect of transcripts accessibility: F [1, 15.55] = 14.35, p = .002. There was no effect of language proficiency (F [1, 31.92] = 3.52, p = .07) and no interaction effect (F [1, 22.85] = 2.55, p = .12). Both NS and NNS perceived better quality of group communication in the publicly shared transcripts condition (M = 5.41, SE = .25) than in the private transcripts condition (M = 4.19, SE = .29).

Correlation between NS's speech clarity and NNS's perceived quality of group communication. To test H3, we calculated the Spearman correlations between the WER of NS's transcripts and NNS's rating on the quality of group communication. Consistent with H3, there was a significant correlation between NS's WER and NNS's ratings of communication quality: r = -0.63, p = .008. These findings indicated that the communication experience of NNS was improved when NS spoke more clearly.

	NNS team member		NS team member	
	NNS transcript	NS transcript	NNS transcript	NS transcript
Private	0	0	N/A	N/A
Publicly shared	19	0	18	60

Table 3. Frequency of voluntary editing initiated by NNS and NS on transcripts by NNS and NS in the private and publicly shared transcripts conditions

Workload

RQ2 asked how the perceived workload of NS and NNS varied as a function of transcripts accessibility. To answer RQ2, we conducted a 2 × 2 Mixed Model ANOVA on the self-rated workload score (see Figure 4). There was a significant main effect of transcripts accessibility on workload: F [1, 14.58] = 5.04, p = .04. There was no effect of language background (F [1, 29.17] = 0.18, p = .67) but a marginal interaction between these two factors (F [1, 23] = 3.83, p = .05). NS's rating of workload remained the same in the publicly shared (M = 3.55, SE = 0.38) and private transcripts conditions (M = 3.71, SE = 0.27): F [1, 10.21] = 0.03, p = .86. NNS's rating of workload was higher in the publicly shared (M = 5.95, SE = 0.39) than the private condition (M = 4.52, SE = 0.37): F [1, 9] = 8.17, p = .02.

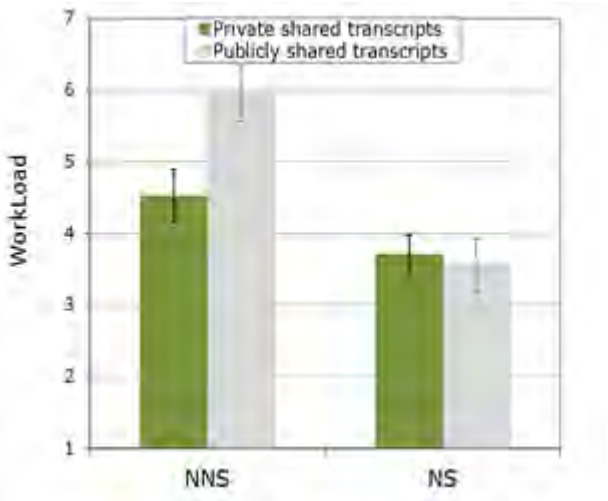


Figure 4. Workload by accessibility condition for NNS and NS (error bars represent standard errors and the mean)

Transcript editing

RQ3 asked whether and how NS and NNS participants used the optional function of editing to facilitate their group communication. To answer RQ3, we calculated the frequency of edits done by each group in each experimental condition (see Table 3).

In the publically shared condition, NS edited both their own transcript errors and the NNS's errors (but not the transcripts of the other NS in their group). For NNS, although they could see the transcripts in both conditions, no editing behavior was observed when the NNS was the only one to see the transcripts. NNS edited their own transcripts only in the publicly shared condition.

DISCUSSION

Overall, our data suggest that sharing automated transcripts between NS and NNS has benefits for communication in multilingual groups. Compared to the private transcripts condition, NS in the publicly shared transcripts condition spoke more clearly. They also voluntary corrected some errors in the shared transcripts. Furthermore, when NS articulated their speech more clearly, NNS gave higher ratings for the quality of the group communication. However, NNS also rated their mental workload as higher

when transcripts were publicly available. We explore each of these findings in more detail in the sections below.

Effects of Shared Transcripts on NS communication

There are several reasons why NS might have articulated their messages more clearly in the shared transcripts condition. One possibility is that the public transcripts gave NS a sense of when and where problems arose in the group communication. This strategy was explicitly mentioned by some of our participants in the post-experiment interview.

> … I tried all the time. Just speak as clearly as possible. I tried to not speak with my usual accent actually. I think the software wouldn't pick my initial way of speaking clearly, so I just speak with a neutral accent. [P24-1, NS]

Other participants explicitly mentioned editing the transcripts to make their meaning clearer for NNS:

> I did the editing a lot to try to help other group members. And also to make sure I'm clear about myself …I was trying to help all the three of us, just trying to make it a group thing and make it clear. I think it's helpful for them. [P23-1, NS]

These benefits to group communication did not come at the cost of greater work load for NS, perhaps because they found multitasking in their native language fairly easy:

> It was easy … I could easily multitask. You know, all I edited was adding some words and punctuations. It was not that taxing. [P26-2, NS]

Interestingly, although there were much more errors in NNS's transcripts, NS corrected only a small percentage of them. NS reported that despite these errors, they could understand NNS's oral messages.

Effects of Shared Transcripts on NNS communication

The effects of public vs. private transcripts had no notable effects on NNS speech clarity, which is not surprising because they received transcripts in both conditions. However, public transcripts did improve NNS's ratings of the quality of the group conversation. We believe that this is directly due to the effects of public transcripts on NS speech clarity, and the strong negative correlation between NS WER scores and NNS ratings of conversational quality supports this view.

NNS's better communication experience in the shared transcripts condition did come with a cost, as evidenced by their significantly higher ratings of workload in the publicly shared vs. private transcripts conditions. This added workload seems to have stemmed at least in part from NNS concerns about how their NS partners would view errors in their transcripts.

> My utterance was transcribed poorly due to my bad pronunciation. I felt that the wrong transcript might lead other NS to confusion so I tried to edit them when I could. Unfortunately, I couldn't edit as much as I wanted. I was fully occupied with other stuff. [P5, NNS (translated into English)]

> … Because the transcripts of my utterance were garbage, I tried various things to improve the quality. I moved the microphone closer to my mouth, but the quality didn't change. I also tried to make my sentence as short as possible so that I do not place burden on the system and other members. . But it didn't help… I got more and more shocked as I noticed that my bad pronunciation cannot be helped. I hope other members are not mad at me. [P11, NNS (translated into English)]

Seeing errors in their own transcripts led NNS to try to correct the transcripts in addition to speaking and listening, thereby creating substantial extra workload.

Publicly Shared Transcripts as a Tool for Conversational Grounding

When NS and NNS interact as a group, transcripts shared publicly with everyone improve the perceived quality of their group communication (as in Figure 3). Our data indicated that groups might have received benefits from the public transcripts in several ways.

First, the transcripts seemed to give them a clearer clue to track the on-going conversation. Unlike purely audio speech, textual transcripts are reviewable. The reviewability of the transcripts seemed to help participants get better sense of the whole conversation.

> I think, when having the transcripts, our conversation was more organized, because you can visually see the text. When you look at transcripts, you get an image. It's like reminding me I should follow this. [P25-1, NS]

In addition, the transcripts facilitated the process of grounding between group members by providing supplementary information through multiple channels. As pointed out by Clark and Brennan [4], grounding sometimes requires communicators to use alternative media to overcome constrains imposed by an original medium. In oral communication, communicators have few cues to track and/or confirm what others are saying at the moment. Shared visual transcripts provide an alternative way to overcome such constraints, which supports the grounding between communicators.

> I understood the outline of each person's speech, but I sometimes missed small parts of others' speech. When I missed some details of what others said, I read the transcripts to compensate for the missed parts. [P8, NNS (translated into English)]

DESIGN IMPLICATIONS

In the above sections, we showed how publicly shared transcripts improve group communication between NS and NNS. The interviews helped us understand why this is the case. Based on our findings, we propose several recommendations for the design of future ASR-based multilingual communication systems.

Improving NNS's Transcripts Accuracy by Using NS's Editing

From the interviews, we found that participants were bothered by the low quality transcripts of NNS's utterances ([P23-1, NS], [P11, NNS], [P5, NNS]). NNS seemed particularly worried if the corrupt transcripts of their own utterance would confuse other group members. Although NS were quite successful in adapting to the ASR technology both in terms of oral adaptation and manual correction, NNS seemed to have difficulties adapting to the technology. NNS's oral adaptation did not seem to improve the accuracy of the transcripts ([P11, NNS], Figure 2), and manual adaptation (editing) was costly in terms of cognitive load. A possible solution would be to simply remove the NNS's transcripts from the interface. Alternatively, a system could make it easier for NS to edit NNS transcripts. Ideally, the edits by NS could be used to train the ASR to better recognize the NNS's speech.

Simplifying Transcript Editing by Tagging Keywords

Although NS's edits appeared beneficial for group communication, NS reported that they sometimes faced difficulties in editing other members' utterances. Particularly when the transcript quality was low, as was the case for many NNS's transcripts, it became difficult for them to remember everything the NNS said ([P23-1, NS]). In such cases, edits were not corrected properly. Sometimes, the edits changed NNS's utterances into something different from what the NNS had actually said. One NNS expressed confusion and seemed upset when she found her utterance being altered in this way. One strategy for preventing erroneous edits might be to limit the amount of edits they can make to the original transcripts. For example, rather than allowing full editing on the whole transcripts, we may set restrictions on editable parts (e.g., keywords) of the transcripts. Another alternative would be to provide audio recordings that could be consulted when memory for the original utterance was low.

Reducing Participants' Workload by Making the Interface More Friendly to NNS

Altogether, the improvement of group communication in the publicly shared condition was accomplished at the cost of extra workload for the NNS. In particular, NNS seemed overloaded with listening, reading and thinking about what to say in the common language. Since previous work has shown that second language message processing and message construction can be mentally taxing [24], using machine translation technology to translate the conversation history into bilingual transcripts may help reduce NNS's workload.

The layout of the ASR interface might also be redesigned to minimize NNS's workload. In our interview data, participants reported that some of the workload came from multitasking on the same interface. For example, the conversation history shifted rapidly as conversation carried on even during someone's edits. A separate space for making corrections or taking notes may help users identify the edited messages faster and keep track of them easier.

LIMITATIONS AND FUTURE DIRECTIONS

There were also several limitations to this study. We examined the effect of shared transcripts by conducting a story-telling task. This task aligns with certain forms of group collaboration in the real world. For example, the general setting of this task resembles an organized distant meeting in which project members with different roles and/or expertise present their ideas one by one with a common goal. Each set of keywords represents a different role or area of expertise. Participants were asked to speak in turns, which is similar to meetings in which the chairman manages turn taking. However, this study leaves open the question of how the shared transcripts would work in other forms of collaboration, such as group work with more flexible interactivity and/or tasks that require minimal verbal-communication.

CONCLUSION

We presented a study comparing publicly shared transcripts, available to both NS and NNS, with private transcripts available only to NNS. NS speech clarity, and both NS and NNS ratings of the quality of the conversation, were higher when transcripts were publicly shared. However, public transcripts were also associated with higher cognitive load on the part of NNS. The findings suggest ways to enhance ASR technologies to make them better suited to multilingual group communication.

ACKNOWLEDGMENTS

This research was funded in part by National Science Foundation grant #1318899 and #1025425. We thank Leslie Setlock, David Hau, Michael Schramm from Cornell for their assistance. We also thank the NTT development team for their technical support and the anonymous reviewers for their valuable comments.

REFERENCES

1. Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D.,. & Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication*, 49(10), 763-786.

2. Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, 112, 272-284.

3.Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707-729.

4.Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, R. M. Levine, & S. D. Teasley (Eds.). *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA.

5.Crookes, G., Rulon, K. A., & Enright, D. S. (1988). Topic and feedback in native-speaker and non-native-speaker conversation. *TESOL Quarterly*, 22(4), 675-681.

6.Desanctis, G., & Monge, P. (1998). Communication processes for virtual organizations. *Journal of Computer-Mediated Communication*, 3(4), 1-16.

7.Dunkel, P. (1991). Listening in the native and second/foreign language: Toward an integration of research and practice. *Tesol Quarterly*, 25(3), 431-457.

8.Dragon Natually Speaking (DNS): Dragon Solutions Field Report. Full text accessible at: http://www.nuance.com/naturallyspeaking/pdf/wp_DNS_Field_Reporting.pdf.

9.Feely, A.-J., & Harzing, A.W.K. (2003). Language management in multinational companies. *Cross-Cultural Management: An Int'l Journal, 10*, 37–52.

10.Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

11.Krause, J. C., & Braida, L. D. (1995). The effects of speaking rate on the intelligibility of speech for various speaking modes. *The Journal of the Acoustical Society of America*, 98, 2982.

12.Kurhila, S. (2001). Correction in talk between native and non-native speaker. *Journal of Pragmatics*, 33(7), 1083-1110.

13.Li, N., & Rosson, M. B. (2012). At a different tempo: what goes wrong in online cross-cultural group chat? In *Proc. of the 17th ACM International Conference on Supporting Group Work*, 145-154.

14.Li, N., & Rosson, M. B. (2012). Instant annotation: early design experiences in supporting cross-cultural group chat. In *Proc.s of the 30th ACM International Conference on Design of Communication*, 147-156.

15.Luisa, M., Lecumberri, G., Cooke, M., Culter, A. (2010). Nonnative speech perception in adverse conditions: A Review. *Speech Communication*, 52, 2010, 864-886.

16.Nabelek, A.K., Donahue, A.M. (1984). Perception of consonants in reverberation by native and non-native listeners. *Journal of Acoustical Society of America, 75*, 1984, 632-634.

17.One Thousand Most Frequent Words by British National Corpus. Full list accessible at: http://simple.wiktionary.org/wiki/Wiktionary:Most_frequent_1000_words_in_English.

18.Pan, Y., Jiang, D., Picheny, M., & Qin, Y. (2009, April). Effects of real-time transcription on non-native speaker's comprehension in computer-mediated communications. In *Proc. of CHI 2009*, 2353-2356.

19.Pan, Y., Jiang, D., Yao, L., Picheny, M., & Qin, Y. (2010, April). Effects of automated transcription quality on non-native speakers' comprehension in real-time computer-mediated communication. In *Proc. of CHI 2010*, 1725-1734.

20.Picheny, M.A., Durlach, N.I., & Braida, L.D. (1989), Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, 32, 600-603.

21.Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78(2), 199-221.

22.Shachaf, P. (2008). Cultural diversity and information and communication technology impacts on global virtual teams: An exploratory study. *Information & Management*, 45(2), 131-142.

23.Shimogori, N., Ikeda, T., & Tsuboi, S. (2010). Automatically generated captions: will they help non-native speakers communicate in English? In *Proc. of the 3rd International Conference on Intercultural Collaboration*, 79-86.

24.Takano, Y. & Noda, A. (1995). Interlanguage dissimilarity enhances the decline of thinking ability during foreign language processing. *Lang. Learn.*, *45*, 657–681.

25.Van Compernolle, D. (2001). Recognizing speech of goats, wolves, sheep and… non-natives. *Speech Communication*, 35(1), 71-79.

26.Vigil, N., & Oller, J. (1976). Rule fossilization: A tentative model. *Language Learning*, 26, 281-295.

27.Yamashita, N., Echenique, A., Ishida, T., & Hautasaari, A. (2013, February). Lost in transmittance: how transmission lag enhances and deteriorates multilingual collaboration. In *Proc. of CSCW 2013*, 923-934.

28.Yao, L., Pan, Y. X., & Jiang, D. N. (2011). Effects of Automated Transcription Delay on Non-native Speakers' Comprehension in Real-Time Computer-Mediated Communication. In *Human-Computer Interaction–INTERACT 2011*, 207-214.

29.Yuan, C. W., Setlock, L. D., Cosley, D., & Fussell, S. R. (2013). Understanding informal communication in multilingual contexts. In *Proc. of CSCW 2013*, 909-9.