Do Automated Transcripts Help Non-Native Speakers Catch Up on Missed Conversation in Audio Conferences?

Ari Hautasaari
NTT Communication Science Labs
2-4 Hikaridai, Seika-cho,
Soraku-gun, Kyoto, Japan
ari.hautasaari@lab.ntt.co.jp

Naomi Yamashita
NTT Communication Science Labs
2-4 Hikaridai, Seika-cho,
Soraku-gun, Kyoto, Japan
naomiy@acm.org

ABSTRACT
Previous work has suggested that speeded up playback of recorded audio works well for native speakers (NS) to catch up on conversation they missed in real-time audio conferences. However, this might not be the case for non-native speakers (NNS) who normally have lower listening ability in their second language. In this study, we explore whether automated speech recognition (ASR) technology can aid NNS when combined with speeded up audio playback. We conducted a laboratory experiment in which 18 NS and 18 NNS listened to a pre-recorded audio conference with three English native speakers, during which they were briefly interrupted and missed parts of the ongoing conversation. They then caught up to the conversation with speeded up audio only (1.6x) and speeded up audio with ASR transcripts. Although ASR transcripts did not improve their comprehension of the conversational content when catching up, transcripts allowed NNS to shift their focus between the two modalities depending on their ability to follow second language speech in different audio speeds. The findings inform future development of ASR tools to support multilingual group communication.

Author Keywords
Automated speech recognition; real-time transcripts; multilingual communication; catching up;

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

INTRODUCTION
Technological advancements continue to facilitate multimodal communication over great distances. With the combination of cost-effectiveness and convenience, audio conferencing allows timely decision-making across borders and time zones, and has become one of the most common communication tools used by individuals, organizations and global enterprises.

While audio conferences are common platforms to hold meetings between non-collocated participants, there are numerous coordination issues involved. For some, it may not always be possible to participate in the meeting from the beginning. Others may have to attend to urgent tasks during the meeting, such as answering a phone call, and thus miss parts of the ongoing conversation [16]. Having others reconstruct these missed parts after rejoining the meeting can be disruptive for the whole group.

Technical solutions for catching up on missed parts of the audio conference without disrupting the ongoing conversation include using audio recordings and exploiting automated speech recognition (ASR) technology to present the main points of the meeting as gists [21]. Audio recordings have also been combined with ASR to provide a speeded up playback of the missed parts combined with text transcripts of the spoken dialogue to allow participants to catch up to the real-time conversation [9, 10]. For native English speakers, the speeded up audio seemed sufficient to recover from missed information [10].

However, we suspect that this would not be the case for non-native speakers (NNS). Indeed, even in normal speed, NNS face unique difficulties in audio conferences that are rarely found between native speakers (NS) [11]. Besides NNS being unable to reach the fluency level of NS in their shared language, NNS often find it challenging to follow other's speech under imperfect audio conditions [7, 24].

An increasing number of studies have focused on alleviating the difficulties that NNS face during real-time audio conferences with ASR technology (e.g., [8]). According to previous research, text transcripts may support NNS comprehension in pre-recorded meetings when the transcripts are provided with reasonable accuracy and delay [18, 19, 20, 24]. Given such positive effects of text transcripts on NNS comprehension of second language speech, we speculate that ASR technology might help NNS catch up on the missed parts of the conversation during an audio conference.

CABS'14, August 21–22, 2014, Kyoto, Japan.

http://dx.doi.org/10.1145/2631488.2631495

Our goal is to examine how ASR transcripts may support NNS during audio conferences. As a first step, we focused on the effects of real ASR transcripts on NNS comprehension of second language conversation. We created a catch up interface, which allows participants to briefly leave the audio conference and catch up to the missed parts of the conversation upon rejoining with speeded up audio and ASR transcripts. In the remainder of this paper we describe a laboratory study, where 18 NS and 18 NNS joined a conversation with three native English speakers as passive listeners. We manipulated the accessibility of the ASR transcripts, where the participants listened and caught up to missed parts of an audio conference with speeded up audio (1.6x) or with speeded up audio plus ASR transcripts. Lastly, we will present our results and discuss our findings in relation to previous works, and draw design implications for future development of ASR applications for multilingual audio conferencing.

BACKGROUND AND RELATED WORK

In this section, we first review previous studies on technological approaches to allow NS to catch up on missed parts of an audio conference. We then review some issues NNS normally face in audio conferencing, and discuss whether the catch up approaches are also effective for NNS. Finally, we introduce recent studies on how ASR transcripts facilitate NNS comprehension, and discuss the possibilities of ASR transcripts helping NNS catch up on missed parts of an audio conference.

Catching Up with Speeded Up Audio

Existing catch up technology allows participants in an audio conference to review any parts of the conversation they missed due to distractions, such as answering an urgent phone call. The technological development, for one, has focused on allowing users to review any missed parts of the ongoing conversation without disrupting the meeting. To achieve this, previous works have exploited audio recordings and language processing technologies to summarize the main points of the missed conversation as gists [21].

However, gisting techniques omit information from the catch up recordings, which might be important for users to fully comprehend the content of the missed conversation. Solutions that include full conversation history play back the missed parts of the audio speech in a higher speed until the point where it catches up to the ongoing conversation (e.g., [9, 10]). The speeded up audio is usually set between 1.4x [5] and 2.0x speed, which is the upper end of what NS can understand [22]. Previous studies have shown that 1.6x speed audio is a reasonable compromise between speed and understandability for NS to catch up on the missed conversation during audio conferences [9, 10].

Previous works have also experimented with using ASR transcripts to help NS catch up on missed parts of an audio conference. However, speeded up audio seemed sufficient, and NS preferred not to view imperfect ASR transcripts while listening to the speeded up audio [9, 10]. Catching up to missed parts of a conversation with speeded up audio alone, however, may be more challenging for NNS considering the difficulties they already face during audio conferences.

NNS Difficulties in Audio Conferences with NS

Previous research has shown that NNS face unique difficulties in audio conferences. Let alone issues with language fluency [13, 23, 25], NNS are further impaired when communicating in their non-native language due to higher cognitive load, and the increased time to process NS utterances [14]. In compromised communication situations, such as audio conferencing with extraneous noise and unclear pronunciation or articulation, these processes may be even harder for NNS [15, 17].

If NS are able and willing to coordinate and adjust their speech to the NNS ability by speaking more slowly or articulating and enunciating more clearly, some of the problems faced by NNS during audio conferences might be alleviated [1, 2, 3, 4, 8]. However, especially in groups where the majority of people are native speakers, this may fall short of expectations [1], also because conversation can move forward only between NS participants (without NNS participation). In order to aid NNS minorities during audio conferences, it is important to consider supporting mechanisms that do not rely on NS ability to ameliorate the blight that NNS face.

ASR Transcripts to Support NNS

As suggested in previous research, ASR technology has potential to alleviate the difficulties NNS face in audio conferences. For example, ASR transcripts provide textual information about the spoken dialogue during audio conferences, which may complement the audio speech and improve NNS comprehension [18, 19, 20, 24]. Previous studies have demonstrated how ASR technologies can improve NNS comprehension during formal presentations and when following television programs in their second language. In a more interactive setting, where ASR transcripts are used in a live multilingual meeting, providing feedback for NS on how their speech is transcribed during real-time audio conferences may facilitate NS adaptation to the technology and improve the accuracy of the ASR transcripts [8].

Altogether, previous works suggest that NNS might benefit even from imperfect transcripts during audio conferences. We wonder whether NNS would still benefit from ASR technology in adverse situations, such as when catching up to missed parts of a conversation with speeded up audio.

CURRENT STUDY

In the current study, we analyze the effects of ASR transcripts on NNS comprehension when catching up on

missed parts of a conversation during a multiparty audio conference in their second language. For our experiment tasks, we use a pre-recorded audio conference corpus in order to control for the variance between speakers.

Research Questions

Previous studies have discussed how text transcripts may improve NNS comprehension of spoken dialogue in their second language [18, 19, 20, 24]. Expanding on these previous works, our first research question asks whether *real* ASR transcripts also improve NNS comprehension during multiparty audio conferences in their second language.

RQ1: Do NNS benefit from viewing real ASR transcripts when listening to a live multiparty audio conference in their second language?

Previous studies suggest that speeded up audio is sufficient for NS to catch up on missed parts of the conversation during audio conferences [10]. However, following second language conversation at faster speed likely imposes a higher cognitive load on NNS [14, 15, 17], affecting their comprehension of the spoken dialogue negatively. Meanwhile, text transcripts seem to help NNS comprehension in their second language. Thus, it is difficult to predict how real ASR transcripts might affect NNS comprehension when combined with speeded up audio for catching up on missed conversation. Hence, our second research question asks:

RQ2: Do NNS benefit from viewing real ASR transcripts when catching up to missed parts of a conversation with speeded up audio during a multiparty audio conference in their second language?

METHOD

Overview

We conducted a laboratory experiment with a single factor (transcript accessibility: audio only vs. audio with ASR transcripts) within subjects design. 18 native English speakers and 18 Japanese non-native English speakers participated in a simulated audio conference. The audio conference was pre-recorded, and included three native English speakers discussing a solution to a survival scenario, where they deliberated on the importance of a list of items for survival [12]. The discussions were divided in to 3-minute clips that were used as individual tasks.

To simulate a brief interruption during an audio conference, the participants were asked to leave the conversation for 30 seconds during each task, causing them to miss part of the discussion. In order to catch up with the ongoing conversation, the participants reviewed the missed parts in two conditions: (1) 1.6x speed audio only, or (2) 1.6x speed audio with ASR transcripts. In this experiment, all participants used the catch up functionality in both conditions. We chose the 1.6x speed based on previous works that considered it the best fitting compromise between speed and understandability for NS when catching up to missed conversation [9, 10].

In total, each participant engaged in 6 tasks (3 tasks in each condition). After each task, the participants answered questions about the content of the conversation. The tasks were designed as uniform, and the conditions and tasks were counterbalanced. An open-ended interview about the participants' experiences during the experiment was conducted after they finished all tasks in both conditions.

Participants

We recruited a total of 36 participants for this study. 18 participants (4 female) were native English speakers who resided in Japan at the time of the study, but grew up in English-speaking countries where they received their primary education (from the age of 6 to 18, elementary school to high school). Their mean age was 40.61 (SD = 10.08).

The other half of the participants (N=18) were bilingual native Japanese speakers (14 female) who grew up in Japan and received their primary education in Japanese. Their mean age was 37.78 (SD = 11.21). None of the Japanese participants had lived in English-speaking countries for more than 2 years (M = 0.84, SD = 0.77). We required a minimum score of 700 (M = 850.29, SD = 68.04) in the TOEIC English proficiency test (Test of English for International Communication), which indicated that they were proficient but not fluent in their second language. They did not have extensive experience in communicating in English outside a classroom (M = 2.61, SD = 1.20 on a 7-point scale ranging from 1 = never to 7 = very often).

Experiment Dataset

ASR transcript and audio recording. We hired three native English speakers (2 female) to generate a realistic audio conference recording for the purposes of this study. The contributors were told that they would be taking part in an audio conferencing experiment using ASR tools. They discussed three survival scenarios using an audio conferencing tool combined with real-time ASR transcripts of the contributors' spoken dialogue.

The ASR transcripts used in this study were generated by a speech recognition software called Dragon Naturally Speaking (DNS) [6]. DNS recognizes the speaker's speech and transcribes it into English text. According to previous research, transcripts with a word error rate (WER) below 20% and a delay no more than 2 seconds can be beneficial for NNS [20, 24].

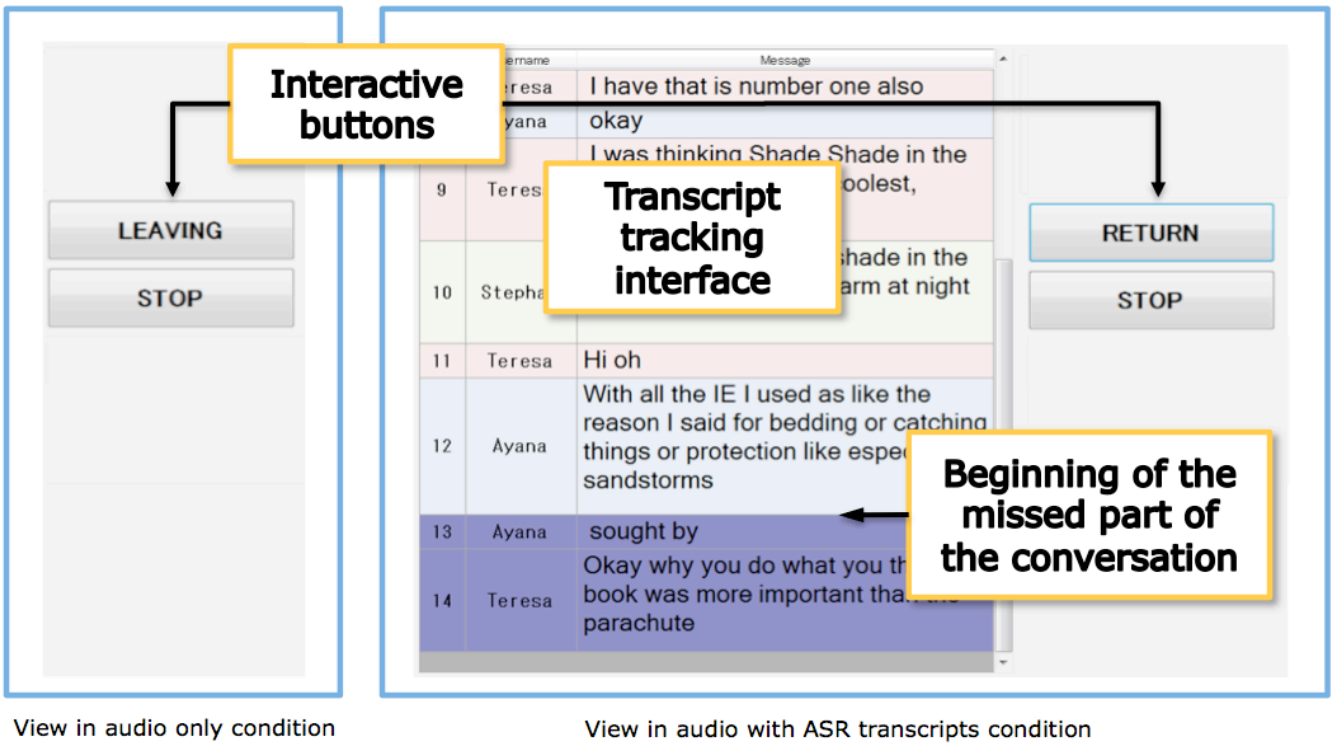

Figure 1. Catch up and real-time transcript tracking interface in audio only and audio with ASR transcripts conditions.

The contributors concluded a training session with DNS before the formal speech recognition started in order to familiarize them with the ASR software and adjust the recognition results to accommodate for their speech and articulation. WER calculated by comparing random samples of transliterated audio excerpts to ASR results was 23%, which is comparable to the reported WER in previous research with a similar setting and equipment [7].

We recorded all audio and transcripts generated during the discussions between the contributors. We then extracted 3-minute clips from each discussion, and used each clip as one task in the experiment.

Software and Equipment

Catch up interface. The catch up interface consisted of two components: interactive buttons for leaving and returning to the meeting, and a real-time transcript tracking interface. In the audio only condition, only the interactive buttons were visible to the participants (Figure 1). In the audio with ASR transcripts condition, the transcript tracking interface was shown on the left side of the screen (Figure 1, right). The transcript tracking interface displayed the ASR transcripts generated by DNS with a 1 to 3 seconds time delay as they appeared during the real-time conversation. The conversation history shows the full transcripts of the conversation. Participants could drag the scroll bar to see who said what in what order during the audio conference.

After the participants pressed the "Leaving" button on the interface, the audio was muted and the interactive button changed to display "Return" (Figure 1). When the participants pressed the "Return" button, the audio continued from the time stamp when they pressed the "Leaving" button and was played back in 1.6x speed until it caught up with the ongoing conversation. In the audio with ASR transcripts condition, the missed parts of the conversation were highlighted with a purple background in the transcript tracking interface (Figure 1, right). After the participants pressed the "Return" button, the highlighted part of the transcripts followed the speeded up audio in real time until the audio caught up with the ongoing conversation.

Tasks and Procedures

The participants were directed to a room and assigned to a laptop computer equipped with an external mouse to manipulate the catch up interface. Before the experiment began, the participants were asked to sign a consent form and fill out a demographic survey on a web browser.

After all participants finished the demographic survey, the experiment organizers explained the task instructions in English for the native speakers (NS) and in Japanese for the non-native speakers (NNS). In the experiment, the participants were asked to imagine that they are part of a multiparty audio conference as passive listeners. They were told that their task was to listen to a conversation in English and try to catch the conversational content the best they could. We allowed note-taking during this experiment.

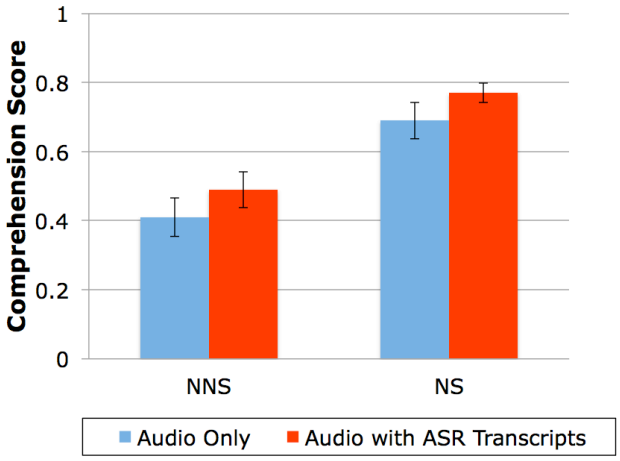

Figure 2. Mean speech comprehension score during a live audio conference in audio only and audio with ASR transcripts conditions for NNS and NS (error bars represent standard error of the mean).

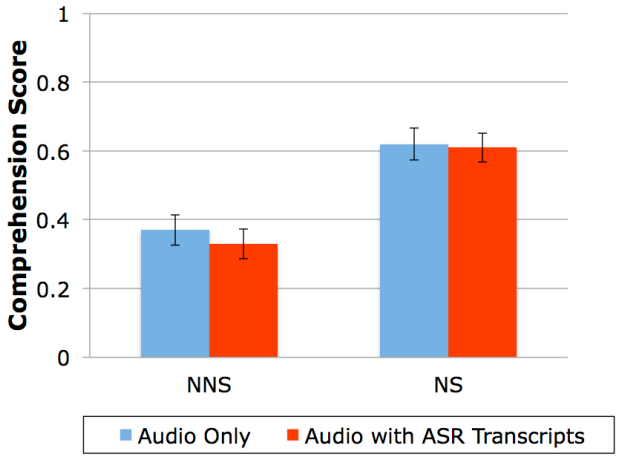

Figure 3. Mean comprehension score when catching up in audio only and audio with ASR transcripts conditions for NNS and NS (error bars represent standard error of the mean).

To simulate a short disturbance during the audio conference, such as an urgent phone call, we asked the participants to solve a short riddle. The riddles were written in English for NS participants and in Japanese for NNS participants on a piece of paper, which was turned face down on a desk next to each participants' laptop. Once the participants attended to this disturbance, they were forced to miss part of the conversation. During each of the experiment tasks, the participants attended to the disturbance (i.e., solving a riddle) once.

All together, the experiment included one practice task in each condition to familiarize the participants with the catch up and transcript tracking interface followed by three actual tasks in each condition. Each task was a 3-minute long snippet of an actual conversation between native English speakers discussing about a survival scenario. After each task, the participants answered a post-task quiz about the content of the conversation they just heard in English. Half of the questions were about the conversation they heard in live audio, and half about the missed part of the conversation they reviewed with the catch up functionality. After the entire experiment session, we conducted open-ended interviews about the participants' experiences. Interviews were conducted in each participant's native language.

Measures

We measured the participants' level of comprehension about the conversational content by administrating a post-task quiz. The participants' score (0-1) in the post-task quiz for reflected their level of comprehension.

RESULTS

Comprehension in Live Audio Conference

In order to answer our *RQ1* regarding the effects of real ASR transcripts on NNS comprehension during a live audio conference, we conducted a 2 (transcript accessibility: audio only vs. audio with ASR transcripts) × 2 (language background: NNS vs. NS) repeated measures ANOVA (Figure 2). There was a marginal main effect for transcript accessibility ($F[1, 34] = 3.17$, $p = .085$), and a significant main effect for language background ($F[1, 34] = 29.78$, $p < .05$). The interaction effect between transcript accessibility and language background was not significant ($F[1, 34] = 1.00$, $p =$ *n.s.*). NNS comprehension score with audio and ASR transcripts ($M = 0.49$, $SD = 0.23$) was higher than with audio only ($M = 0.41$, $SD = 0.24$). Similarly, NS comprehension score was higher with audio and ASR transcripts ($M = 0.77$, $SD = 0.12$) than with audio only ($M = 0.69$, $SD = 0.22$).

This result answers our *RQ1*, and is consistent with previous work by Pan et al. (2009, 2010) regarding the effects of ASR transcripts on NNS comprehension. Both NNS and NS comprehension marginally improved when viewing ASR transcripts during a live audio conference.

Comprehension When Catching Up

To answer our *RQ2* regarding the effects of real ASR transcripts on NNS comprehension when catching up on missed parts of an audio conference with speeded up audio, we conducted a 2 (transcript accessibility: audio only vs. audio with ASR transcripts) × 2 (language background: NNS vs. NS) repeated measures ANOVA (Figure 3). There was no significant main effect for transcript accessibility ($F[1, 34 = 0.99$, $p =$ *n.s.*), but a significant main effect for language background ($F[1, 34] = 36.61$, $p < .05$). The interaction effect between transcript accessibility and language background was not significant ($F[1, 34] = 1.00$, $p =$ *n.s.*). NNS comprehension score with audio and ASR transcripts ($M = 0.33$, $SD = 0.20$) was slightly lower than with audio only ($M = 0.37$, $SD = 0.19$). NS comprehension score was also slightly lower with audio and ASR transcripts ($M = 0.61$, $SD = 0.18$) than with audio only ($M = 0.62$, $SD = 0.21$).

This result answers our *RQ2*. NNS comprehension of the conversational content did not improve by viewing ASR transcripts when catching up to missed parts of a conversation during a multiparty audio conference compared to speeded up audio only.

DISCUSSION

Overall, our data suggests that NNS do not benefit from using ASR transcripts when catching up on missed parts of a conversation during a multiparty audio conference in their second language. In the next sections, we aim to illuminate our findings in more detail by reflecting on the post-experiment interviews with NS and NNS participants. [1]

Effects of ASR Transcripts on NS Behavior

In the post-experiment interviews, NS participants provided several reasons why they gained little benefit from using ASR transcripts. Firstly, NS participants often referred to the transcripts as being distracting because of the delay.

> Especially the delay of the transcripts is distracting in normal speed. [NS10]

> The transcripts are showing a different place of the discussion. [NS5]

Furthermore, some NS participants noted that the mistakes in the transcripts caused them to be less than helpful for speech comprehension. Some even referred to the mistakes in the transcripts as being entertaining.

Another reason for NS gaining little benefit from ASR transcripts seemed to stem from their familiarity in listening to fast English speech. The participants also mentioned that they would rather listen to the speeded up audio again when catching up to confirm their understanding of the conversational content rather than rely on the ASR transcripts.

> The mistakes in the transcripts are not helpful. I would rather listen to the fast sections again than read the text [if I missed some part of the conversation]. [NS12]

Meanwhile, regardless of the delay and errors in ASR transcripts, some NS participants did seem to find the ASR transcripts useful for catching up on parts of the conversation they missed.

> If I missed some parts, I checked the text transcripts to catch up with the audio. This is for the parts I didn't actually hear for some reason, [but] I was used to the fast speed because I listen to podcasts in 1.8 speed. [NS15]

Effects of ASR Transcripts on NNS Behavior

While some NNS also vocalized the detrimental effects of transcript delay and imperfect accuracy, many NNS expressed the difficulties of focusing on two modalities (audio and ASR transcripts) at the same time. Interestingly, more than half of the NNS participants reported developing some sort of strategy to alleviate the burden of concentrating on the two modalities:

> In normal speed, I found it difficult to listen and read at the same time. So I tried not to look at the transcripts but concentrate on listening. When it came to fast speed, it just became impossible to understand what was said. So I started to read the transcripts. I think it helped a lot. [NNS6]

> In fast speed, I couldn't follow the conversation so I concentrated on reading the transcripts. [NNS4]

While many NNS tried to concentrate on reading the ASR transcripts when catching up to missed conversation, some NNS participants seemed to shift their focus on listening to the speeded up audio only.

> In normal speed, I looked at the transcripts when I missed some words. That was very useful. But in fast speed, I didn't have the time to check the transcripts [even when I missed some parts]. I had to move forward and concentrate on listening [to the speeded up audio only]. [NNS12]

Summary of Findings

Our results suggested that both NNS and NS might benefit from ASR transcripts even with imperfect accuracy and delay during live multiparty audio conferences. Although our results did not reach statistical significance, they indicated a trend that ASR technology might provide additional information about the conversational content for NNS and improve their understanding. Thus, our results complement previous works on the positive effects of text transcripts on NNS comprehension [18, 19, 20, 24] during multiparty audio conferences in their second language [8].

However, much like in previous works [9, 10], the NS participants in our study reported that the imperfect transcript quality combined with 1-3 second delay was distracting. Our experiment did not include interactive aspects between the NNS and NS participants during the audio conference. Thus, the behavioral adjustments that NS speakers might adopt in the presence of NNS receivers were not present in our experimental setting [1, 2, 3, 4], which may have also affected the accuracy of the ASR transcripts [8]. While few NS did find the transcripts helpful when they missed some information, our results beg the question whether the detrimental effects of current ASR technology outweigh the positive effects for NS in live audio conferences.

Our results for using speeded up audio in combination with ASR transcripts to catch up on missed parts of the conversation showed no significant difference compared to speeded up audio only for NNS or NS. This result reflects previous research, where NS reported the imperfect

[1] All NNS interview quotes are translated from Japanese by the Authors.

accuracy as a limitation for using only ASR transcripts to catch up on missed conversation [10].

However, NNS adopted an interesting strategy, where they changed their focus from audio to text depending on their ability to follow the speeded up second language speech. When the NNS were unable to adequately comprehend the audio speech in 1.6x speed, they either shifted their focus completely to the ASR transcripts ignoring the speeded up audio [NNS 4, 7] or concentrated on listening to the speeded up audio ignoring the ASR transcripts [NNS12].

This switch between modalities may explain why ASR transcripts did not improve NNS comprehension when catching up. Although some NNS preferred reading the ASR transcripts over listening to the speeded up audio, this does not mean that their reading skill level surpassed their second language listening skills. Some NNS participants were able to compensate their listening skills by reading the ASR transcripts in normal speed [NNS12], which might have helped them improve their comprehension score during the live audio conference.

We wonder whether there is a technical solution to present the ASR transcripts in a way that accommodates the NNS shift in focus between modalities when catching up on missed conversation. One way to accomplish this would be to combine automatic keyword or key phrase extraction methods with ASR technology. As audio conferences are often unstructured, include overlaps between speakers and sudden changes between topics, highlighting the keywords in the transcripts might allow NNS to focus their attention to the key points of the conversation, and better detect sudden topic changes in speeded up audio.

Future Directions

In future studies, we are interested in combining automatic keyword or key phrase extraction with ASR technology to better support NNS when catching up on missed parts of an audio conference with speeded up audio. Secondly, as our experiment did not include interactive aspects between the NNS participants and NS speakers, we are interested in exploring how ASR transcripts combined with keyword extraction might accommodate communication between NNS and NS in multiparty audio conferences. Investigating how speeded up audio and ASR transcripts might help NNS and NS catch up on longer periods of missed conversation would further inform the potential applications of the current technology for audio conferencing systems.

CONCLUSION

We presented a study, where Japanese non-native English speakers (NNS) and native English speakers (NS) participated as passive listeners in an audio conference with three native English speakers. During the audio conference, the participants were briefly distracted and missed parts of the conversation. To catch up on the conversation, the participants used speeded up audio (1.6x) and speeded up audio with transcripts generated by automated speech recognition (ASR) software to review the missed parts of the audio conference.

Our results indicated that while ASR transcripts might improve NNS comprehension in live multiparty audio conferences, NNS did not benefit from viewing the transcripts when catching up to missed parts of the conversation with speeded up audio. However, ASR transcripts allowed the NNS to shift their focus from audio to text depending on their ability to follow the spoken dialogue in different speeds. This provided an alternative channel for NNS to catch up on the missed conversation when they were unable to follow the second language conversation in speeded up audio.

ACKNOWLEDGMENTS

We would like to extend our gratitude to the NTT development team for their technical support, and to the anonymous reviewers for their invaluable comments.

REFERENCES

1. Bradlow, A. R. and Bent, T. Perceptual adaptation to non-native speech. *Cognition 106, 2* (2008), 707-729.
2. Bradlow, A. R. and Bent, T. The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America 112* (2002), 272-284.
3. Bradlow, A. R. and Pisoni, D. B. Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America 106* (1999), 2074-2085.
4. Bradlow, A. R., Torretta, G. M. and Pisoni, D. B. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication 20, 3* (1996), 255-272.
5. Christel, M., Winkler, D., Taylor, R. and Smith, M. Evolving video skims into useful multimedia abstractions. In *Proc. CHI 1998,* ACM (1998), 171-178.
6. Dragon Naturally Speaking (DNS): Dragon Solutions Field Report. Full text accessible at: http://www.nuance.com/naturallyspeaking/pdf/wp_DNS_ Field_Reporting.pdf.
7. Dunkel, P. Listening in the native and second/foreign language: Toward an integration of research and practice. *Tesol Quarterly 25, 3* (1991), 431-457.
8. Gao, G., Yamashita, N., Hautasaari, A., Echenique A. and Fussell, S. Effects of public vs. private automated transcripts on multiparty communication between native and non-native English speakers. In *Proc. CHI 2014*, ACM (2014), 843-852.
9. Inkpen, K., Hegde, R., Junuzovic, S., Brooks, C., Tang, J.C. and Zhang, Z. AIR Conferencing: Accelerated instant replay for in-meeting multimodal review. In *Proc. MM'10,* ACM (2010), 663-666.

10. Junuzovic, S., Inkpen, K., Hegde, R., Zhang, Z., Tang, J. and Brooks, C. What did I miss? In-meeting review using multimodal accelerated instant replay (AIR) conferencing. In *Proc. CHI 2011*, ACM (2011), 513-522.
11. Kurhila, S. Correction in talk between native and non-native speaker. *Journal of Pragmatics 33, 7* (2001), 1083- 1110.
12. Lafferty, J. C., Eady, P. M. and Elmers, J. The desert survival problem. Plymouth, Michigan. Experimental Learning Methods (1974).
13. Li, N. and Rosson, M. B. At a different tempo: What goes wrong in online cross-cultural group chat? In *Proc. GROUP 2012*, ACM (2012), 145-154.
14. Li, N. and Rosson, M. B. Instant annotation: Early design experiences in supporting cross-cultural group chat. In *Proc. SIGDOC 2012*, ACM (2012), 147-156.
15. Luisa, M., Lecumberri, G., Cooke, M. and Culter, A. Non-native speech perception in adverse conditions: A Review. *Speech Communication 52* (2010), 864-886.
16. Meetings in America V: Meeting of the minds. *An MCI® Executive White Paper*, (2003). Full text at: https://e-meetings.verizonbusiness.com/meetingsinamerica/pdf/MIA5.pdf (accessed: 25.6.2014).
17. Nabelek, A.K. and Donahue, A.M. Perception of consonants in reverberation by native and non-native listeners. *Journal of Acoustical Society of America 75* (1984), 632-634.
18. Pan, Y., Jiang, D., Picheny, M. and Qin, Y. Effects of real-time transcription on non-native speaker's comprehension in computer-mediated communications. In *Proc. CHI 2009*, ACM (2009), 2353-2356.
19. Pan, Y., Jiang, D., Yao, L., Picheny, M. and Qin, Y. Effects of automated transcription quality on non-native speakers' comprehension in real-time computer-mediated communication. In *Proc. CHI 2010*, ACM (2010), 1725-1734.
20. Shimogori, N., Ikeda, T. and Tsuboi, S. Automatically generated captions: Will they help non- native speakers communicate in English? In Proc. *ICIC 2010*, ACM (2010), 79-86.
21. Tucker, S., Bergam, O., Ramamoorthy, A. and Whittaker S. Catchup: A useful application of time-travel in meetings. In *Proc. CSCW 2010*, ACM (2010), 99-102.
22. Wildemuth, B., Marchionini, G., Yang, M., Geisler, G., Wilkens, T., Hughes, A. and Gruss, R. How fast is too fast? Evaluating fast forward surrogates for digital video. In *Proc. JCDL 2003*, ACM (2003), 221-230.
23. Yamashita, N., Echenique, A., Ishida, T. and Hautasaari, A. Lost in transmittance: How transmission lag enhances and deteriorates multilingual collaboration. In *Proc. CSCW 2013*, ACM (2013), 923-934.
24. Yao, L., Pan, Y. X. and Jiang, D. N. Effects of automated transcription delay on non-native speakers' comprehension in real-time computer-mediated communication. In *Proc. INTERACT 2011*, ACM (2011), 207-214.
25. Yuan, C. W., Setlock, L. D., Cosley, D. and Fussell, S. R. Understanding informal communication in multilingual contexts. In *Proc. CSCW 2013*, ACM (2013), 909-922.