Catching Up in Audio Conferences: Highlighting Keywords in ASR Transcripts for Non-Native Speakers

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

CABS'14, Aug 21-22 2014, Kyoto, Japan
ACM 978-1-4503-2557-8/14/08.
http://dx.doi.org/10.1145/2631488.2634064

Abstract
Previous works suggest that non-native speakers (NNS) may benefit from viewing textual transcripts of spoken dialogue generated by automated speech recognition (ASR) technology during audio conferences. However, viewing ASR transcripts while listening to the ongoing conversation may impose a higher cognitive load on NNS, especially in adverse audio conditions. We examined how automatically highlighted keywords in real-time ASR transcripts might benefit NNS when catching up on missed parts of an audio conference by reviewing a speeded up (1.6x) audio playback of the missed conversation.

Author Keywords
Automated speech recognition; real-time transcripts; keyword detection; multilingual communication

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

Introduction
Audio conferencing is among the most common communication tools used by individuals and organizations on a global scale. Multilingual teams often

rely on English as a common language to communicate between non-collocated team members. However, as non-native speakers (NNS) rarely reach the fluency level of native English speakers (NS) they may often experience difficulties when following second language conversations in audio conferences [7].

Besides challenges related to language diversity, there are numerous coordination issues involved in holding audio conferences between non-collocated team members. For example, some participants may have to attend to urgent tasks during a meeting, such as answering a phone call, causing them to miss parts of the conversation. Technical solutions for catching up on these missed parts without disrupting the ongoing conversation include using audio recordings and textual transcripts of the conversation generated by automated speech recognition (ASR) technology [2, 3, 6]. In this paper, we examine how automatic keyword highlighting in real-time ASR transcripts might benefit NNS when catching up on missed parts of a multiparty audio conference by reviewing a speeded up audio playback.

Current Study

Previous studies suggest that speeded up audio is sufficient for NS to catch up on missed parts of the conversation during an audio conference [3]. Textual transcripts, on the other hand, may improve NNS comprehension of spoken dialogue in their non-native language [5]. However, following textual transcripts and second language conversation at a faster speed likely increases NNS's cognitive load [1, 4, 7], which may in turn have a negative impact on their comprehension of the spoken dialogue. Drawing from these previous works, we ask:

RQ: Do NNS benefit from viewing highlighted keywords in real-time ASR transcripts when catching up on missed parts of a conversation with speeded up audio during a multiparty audio conference in their non-native language?

Method

We conducted a laboratory experiment, where 18 native English speakers and 18 Japanese non-native English speakers participated in a simulated audio conference as passive listeners. During the audio conference, they were briefly distracted and missed parts of the ongoing conversation. They then had to catch up on the missed conversation using speeded up (1.6x) audio playback and real-time ASR transcripts. We manipulated the accessibility of automatically highlighted keywords when catching up on these missed parts (audio and ASR transcripts vs. audio and ASR transcripts with highlighted keywords).

The participants listened to three 3-minute conversation clips between three native English speakers (2 female) discussing a solution to a survival scenario in both conditions. The ASR transcripts of these conversations were generated by a speech recognition software called Dragon Naturally Speaking[1]. The word error rate (WER) was comparable to the reported WER in previous research with a similar setting and equipment at 23% [1]. Keywords in the ASR transcripts were automatically extracted using a keyword extraction software called GENSEN-Web[2]. Each extracted keyword was highlighted with a red font in a

[1]http://www.nuance.com/naturallyspeaking/pdf/wp_DNS_Field_Reporting.pdf.

[2]http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb_eng.html

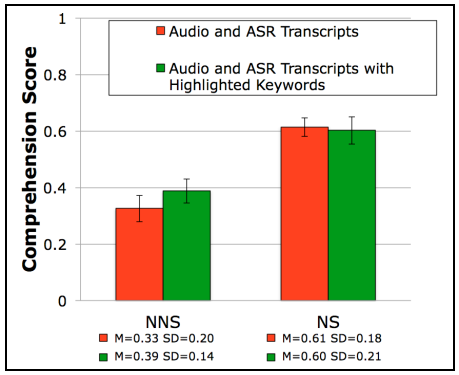

Figure 1. Mean speech comprehension score by keyword accessibility condition when catching up on missed conversation for NNS and NS (error bars represent standard error of the mean).

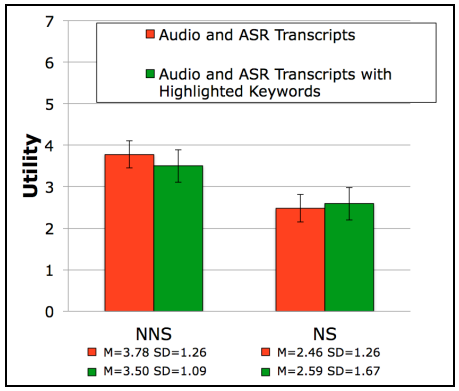

Figure 2. Perceived utility of ASR transcripts by keyword accessibility condition when catching up on missed conversation for NNS and NS (error bars represent standard error of the mean).

real-time transcript tracking interface displayed on a laptop computer assigned to each participant (see [2] for details).

Measures

We measured the participants' level of comprehension about the conversational content regarding the parts they missed and had to review with the catch up functionality by administrating a post-task quiz. The participants' score (0-1) in the post-task quiz reflected their level of comprehension.

The participants' perception of the utility of ASR transcripts when catching up on missed conversation with speeded up audio was measured using three 7-point Likert scales ("real-time transcripts/highlighted keywords helped me organize my thoughts", "real-time transcripts/highlighted keywords helped me recover from missed information in the conversation", and "real-time transcripts/highlighted keywords helped me follow the flow of the conversation", 1 = strongly disagree, 7 = strongly agree). The questions formed a reliable scale (Cronbach's α = .95) and were averaged to create a measure of utility.

Results

In order to answer our *RQ*, we conducted 2 (keyword accessibility: audio and ASR transcripts vs. audio and ASR transcripts with highlighted keywords) × 2 (language background: NNS vs. NS) repeated measures ANOVAs on the effects of highlighted keywords in real-time ASR transcripts on NNS comprehension and perceived utility when catching up on missed parts of an audio conference.

There was no significant main effect for keyword accessibility ($F[1, 34]=0.30$, $p=n.s.$), but a significant main effect for language background ($F[1, 34]=38.63$, $p<.05$) on the level of comprehension about the speeded up audio content (Figure 1). The interaction effect between keyword accessibility and language background was not significant ($F[1, 34]=0.65$, $p=n.s.$).

Secondly, there was no significant main effect for keyword accessibility ($F[1, 34]=0.03$, $p=n.s.$), but a significant main effect for language background ($F[1, 34]=17.53$, $p<.05$) on the perceived utility of ASR transcripts when catching up with speeded up audio (Figure 2). The interaction effect between keyword accessibility and language background was not significant ($F[1, 34]=0.06$, $p=n.s.$).

These results partly answered our *RQ.* While NNS level of comprehension did not improve significantly when automatically highlighted keywords were available in ASR transcripts, they perceived the utility of ASR transcripts and highlighted keywords higher than NS for catching up on missed parts of an audio conference with speeded up audio.

Discussion

In this section we discuss our results in more detail by reflecting on the post-experiment interviews with NS and NNS participants. [3]

NS participants, despite the ASR transcripts and highlighted keywords including errors, found the

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

NS6	Keywords help find information in the transcripts that I missed in the fast sections.
NS13	I listened to the audio first, and then went through the text with highlights again to see if I missed something.
NS5	The keywords helped me remember them without hearing them. For example, I didn't remember hearing the word 'scissors', but when I filled out the questionnaire, I remembered seeing it as a keyword. I forgot to read the transcripts, so keywords helped me more**.**

Table 1. NS interview quotes.

NNS9	Highlighted keywords gave me some hint what the conversation was about. I could listen to the conversation based on these keywords.
NNS4	I think highlighted words tend to stay in my mind.
NNS16	Highlighted keywords caught my attention, but I didn't think they were particularly useful or important keywords.

Table 2. NNS interview quotes.

keywords useful for checking if they missed any information when using the catch up functionality (Table 1: NS6, NS13). Further, the highlighted keywords worked as visual reminders of the audio content for some NS participants (Table 1: NS5).

NNS were also able to make use of the keywords to confirm their understanding of the ongoing conversation. Moreover, the highlighted keywords offered the NNS a supporting channel to follow the second language conversation when they could not keep up with the speeded up audio alone (Table 2: NNS9). Similarly to NS participants, NNS also vocalized the value of highlighted keywords as visual reminders of the audio content (Table 2: NNS4). However, due to errors in ASR transcripts and automatic highlighting, some NNS participants may have found them distracting (Table 2: NNS16).

While we found no significant increase in NNS comprehension of speeded up audio content when catching up, our results suggest that enriching real-time ASR transcripts with automatically highlighted keywords may benefit both NNS and NS during multiparty audio conferences.

Conclusion

We presented a study, where 18 native English speakers and 18 Japanese non-native English speakers caught up to missed parts of an audio conference with speeded up audio and real-time ASR transcripts. Our results showed that automatically highlighted keywords in the ASR transcripts did not significantly improve NNS comprehension of the conversational content in speeded up audio. However, NNS may benefit from highlighted keywords in ASR transcripts as visual reminders of the conversational content, and as an additional supporting channel for keeping up, especially, with speeded up review of an audio conversation.

References

[1] 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.

[2] Hautasaari, A. and Yamashita, N. Do automated transcripts help non-native speakers catch up on missed conversation in audio conferences? In *Proc. CABS 2014*, ACM (2014), to appear.

[3] 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.

[4] Luisa, M., Lecumberri, G., Cooke, M. and Culter, A. Non-native speech perception in adverse conditions: A Review. *Speech Communication 52* (2010), 864-886.

[5] 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.

[6] 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.

[7] 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.