Investigating the Impact of Automated Transcripts on Non-native Speakers' Listening Comprehension

Xun Cao[1,2], Naomi Yamashita[2], and Toru Ishida[1]

[1]Kyoto University, Kyoto, Japan
xun@ai.soc.i.kyoto-u.ac.jp
ishida@i.kyoto-u.ac.jp

[2]NTT Communication Science Laboratories, Kyoto, Japan
naomiy@acm.org

ABSTRACT

Real-time transcripts generated by automatic speech recognition (ASR) technologies hold potential to facilitate non-native speakers' (NNSs) listening comprehension. While introducing another modality (i.e., ASR transcripts) to NNSs provides supplemental information to understand speech, it also runs the risk of overwhelming them with excessive information. The aim of this paper is to understand the advantages and disadvantages of presenting ASR transcripts to NNSs and to study how such transcripts affect listening experiences. To explore these issues, we conducted a laboratory experiment with 20 NNSs who engaged in two listening tasks in different conditions: audio only and audio+ASR transcripts. In each condition, the participants described the comprehension problems they encountered while listening. From the analysis, we found that ASR transcripts helped NNSs solve certain problems (e.g., "do not recognize words they know"), but imperfect ASR transcripts (e.g., errors and no punctuation) sometimes confused them and even generated new problems. Furthermore, post-task interviews and gaze analysis of the participants revealed that NNSs did not have enough time to fully exploit the transcripts. For example, NNSs had difficulty shifting between multimodal contents. Based on our findings, we discuss the implications for designing better multimodal interfaces for NNSs.

Categories and Subject Descriptors

• **Human-centered computing→Natural language interfaces.**

Keywords

Listening comprehension problems; automatic speech recognition (ASR) transcripts; non-native speakers (NNSs); eye gaze

1. INTRODUCTION

Non-native speakers (NNSs) often have difficulty comprehending the speech of native speakers (NSs) [2, 9]. They particularly face comprehension difficulties in real-time settings such as joining audio conferences (as a listener) and listening to live radio broadcasts or lectures where they cannot repeat the audio or listen at their own pace. When NNSs miss some parts of the speech or cannot understand certain words, they cannot exert the time and processing power to timely recover from such problems because they are already overwhelmed by processing continuous streams of speech while listening [16, 19]. As a result, they often get left behind and sometimes even miss the key points of meetings/lectures.

Previous research showed that real-time transcripts generated by automatic speech recognition (ASR) technologies can help NNSs improve their listening comprehension when their accuracy and delay fall within a reasonable range [14, 20]. If such a technology was installed into portable devices such as smartphones, tablets, or laptops, NNSs could view the automatically generated transcripts on the screen while they listened to the speech.

One advantage of providing ASR transcripts to NNSs is that they provide supplemental information to recover from certain problems. For example, NNSs can read transcripts when they miss some parts of the speech or confirm their listening comprehension. However, ASR transcripts often place a burden on NNSs [4, 20]. Since they are already burdened by processing audio information (i.e., NS speech), providing them with textual information (i.e., ASR transcripts) might further overwhelm them with excessive information. Following ASR transcripts while listening may increase their burden and even trigger new problems.

The goal of our research is to provide a design guideline for presenting ASR transcripts to NNSs to effectively support their listening comprehension. To reach our goal, we need to understand the advantages and disadvantages of presenting ASR transcripts to NNSs in further details, and see how they affect their listening experiences. More specifically, a) How do NNSs use ASR transcripts while listening to native speech? b) What types of listening comprehension problems can be solved by reading ASR transcripts? When NNSs fail to solve problems by reading them, what are the factors of failure? c) Do ASR transcripts place an extra burden on NNSs when they fail to solve their listening comprehension problems?

To answer the above research questions, we conducted a laboratory experiment with 20 NNSs who engaged in two listening tasks in different conditions: audio only and audio+ASR transcripts. During the task, the participants pressed a button whenever they

ICMI '16, November 12 - 16, 2016, Tokyo, Japan
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4556-9/16/11 $15.00
DOI: http://dx.doi.org/10.1145/2993148.2993161

encountered anything about which they were not clear or did not understand (i.e., a comprehension problem). Next they explained the kinds of problems they faced and how long they persisted. To better understand how the NNSs used the ASR transcripts, under the audio+ASR transcript condition, we recorded their eye movements using an eye-tracker. Note that this paper focuses on the listening comprehension problems faced by NNSs during their cognitive processing of speech input because such problems are the most common and fundamental obstacles faced by any NNS and lead to cognitive overload [6, 2].

Through an exploratory analysis of the experiment data, we found that NNSs adopted different strategies when using the ASR transcripts; some followed the transcripts throughout the listening; some only checked them when necessary. Although the ASR transcripts did seem useful for NNSs to some extent, post-task interviews and gaze analysis of the participants revealed that NNSs did not have enough time or cognitive resources to fully exploit the transcripts. For example, they had difficulty concentrating on listening/reading or shifting between multimodal contents. We also found that the ASR transcripts helped the NNSs solve certain problems (e.g., "do not recognize words they know"), but imperfect ASR transcripts (e.g., errors and no punctuation) sometimes confused the NNSs and even generated new problems. Furthermore, even though NNSs tried to solve certain problems by reading the transcripts (e.g., the words they could not understand), the problems were not necessarily solved, rather their burden was increased.

In the remainder of this paper, we first review previous studies and discuss the framework of our study. We then describe our study that investigated how ASR transcripts impact the listening comprehension of NNSs and conclude with a discussion of the implications of our findings for designing better multimodal interfaces for NNSs.

2. BACKGROUND

2.1 Real-time Listening Comprehension Problems of NNSs

NNSs often face comprehension difficulties when listening to the speech of NSs. In the field of second language learning, researchers have examined the difficulties/problems faced by NNSs from different perspectives. Rubin extensively reviewed the research on second language listening comprehension and attributed the factors that affect listening comprehension into five characteristics: text characteristics (e.g., speech rate), interlocutor characteristics, task characteristics (e.g., task type), listener characteristics (e.g., language proficiency level, memory), and process characteristics (e.g., listening strategies) [15]. While most previous studies explored the factors that influence second language listening (as represented by Rubin's work) [15, 2], Goh classified the listening comprehension problems faced by NNSs into ten categories from a cognitive perspective (Table 1). In her study, 40 non-native students wrote weekly diaries and explained the listening comprehension problems they faced during lectures [6]. Building on Goh's work, Cao et al.'s recent work identified two new problems which tentatively confused the NNSs during listening: "confused about unexpected word appearance" and "unsure about the meaning of words." [3]

Table 1. Listening comprehension problems identified in Goh's work [6]

	Problems
1.	Do not recognize words they know
2.	Unable to form a mental representation from words heard
3.	Cannot chunk streams of speech
4.	Neglect the next part when thinking about meaning
5.	Do not understand subsequent parts of input because of earlier problems
6.	Concentrate too hard or unable to concentrate
7.	Understand words but not the intended message
8.	Confused about the key ideas in the message
9.	Miss the beginning of texts
10.	Quickly forget what is heard

2.2 Technologies to Improve NNSs' Listening Comprehension

Previous studies have shown that real-time transcripts generated by ASR technologies hold the potential to facilitate the listening comprehension of NNSs. ASR transcripts provide textual information that can complement audio speech and improve the comprehension of NNSs [8, 14, 20]. Pan et al. investigated how the quality of ASR transcripts impacts comprehension and subjective evaluations. They found that a 20% word-error-rate (WER) was the most likely critical point for transcripts to be acceptable, and at a 10% WER, comprehension performance significantly improved compared to a no-transcript condition [14].

Yao et al. compared the NNS comprehension performance among three conditions (no-transcript, perfect transcripts with a 2-second delay, and transcripts with a 10% WER and a 2-second delay). The comprehension performance in the latter two conditions was significantly better than that in the no-transcript condition [20].

Despite the positive effects of introducing ASR transcripts, previous research also reported that ASR transcripts burden NNSs who sometimes get overwhelmed when they simultaneously listen to speech and read transcripts that contain errors and delays [4, 20]. In addition, errors and delays negatively impacted how NNSs perceived the value of the ASR transcripts [14, 20].

Overall, the previous studies identified the usefulness of ASR transcripts for supporting NNS listening comprehension and the risk of placing an extra burden on NNSs. However, we still lack a detailed understanding of how NNSs benefit from ASR transcripts (e.g., what types of listening comprehension problems could be solved) and what are the difficulties of using them (e.g., the factors that hinder them from solving their problems).

3. CURRENT STUDY

Previous research has shown that ASR transcripts have the potential to support NNS listening comprehension in real time. However, these results were obtained by asking NNSs comprehension questions and/or from questionnaires; little previous work has scrutinized how NNSs actually use ASR transcripts while they are listening. We believe such knowledge is important for improving the presentation method of ASR transcripts so that they can support NNSs more effectively. Thus, we pose the following research questions:

RQ1: How do NNSs use ASR transcripts while listening to native speech? Is there a common pattern with which they use/read the transcripts, or are there different patterns?

According to Goh, NNSs encounter various types of comprehension problems when they listen to native speech. Among them, we expect that such problems as "do not recognize words they know" can be solved using ASR transcripts, but not such problems as "lack of vocabulary." In addition, since NNSs are often overburdened by processing speech input, ASR transcripts might not always help solve their problems. Transcript errors and delays may exacerbate the situation and even generate new problems. Therefore, we pose the following research questions:

RQ2: What types of listening comprehension problems can be solved by reading ASR transcripts? When NNSs fail to solve problems by reading them, what are the factors of failure?

When NNSs encounter a comprehension problem, they might try to solve it by reading transcripts. However, due to transcript errors and delay, the problem solving process might not be successful and could place an extra burden on NNSs. We investigate how NNSs are burdened during such a process.

RQ3: Do ASR transcripts place an extra burden on NNSs when they fail to solve their listening comprehension problems?

4. METHOD

4.1 Overview

We conducted a laboratory experiment with 20 NNSs who engaged in two listening tasks in different conditions:

- Without-transcript: only audio was presented
- With-transcript: both audio and ASR transcripts were presented

In each condition during the listening task, the participants pressed a button to indicate when they heard confusing language or did not understand something: comprehension problems. Pressing a button marked specific places in the lecture transcripts, which were visited later to explain the details of the problems. We used this "pressing a button" method because it has low-overhead [11]. In addition, it guarantees that we can record the problems faced by NNSs in real time and simultaneously keep the task close to actual listening experiences [3].

The experiment used a within-subject design. Its conditions were counterbalanced across subjects to minimize the order effects. To understand how ASR transcripts are used during listening, NNSs' eye movements were recorded using an eye-tracker under the with-transcript condition (Figure 1) [18].

4.2 Participants

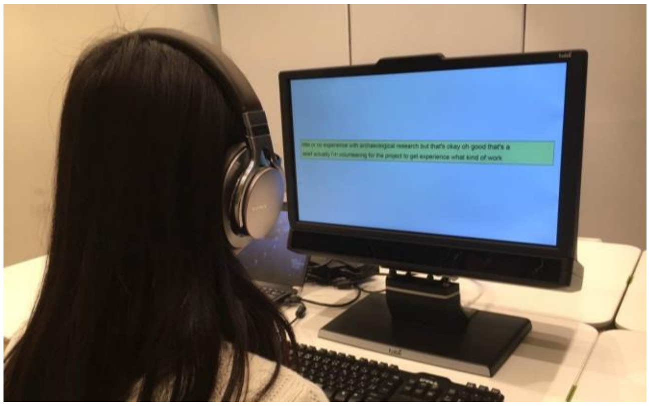
Twenty non-native English speakers participated in our study: ten females and ten males. Their mean age was 25.9 (SD = 2.41). All spoke Chinese as their first language. Their Test of English for International Communication (TOEIC) scores ranged from 690 to 950 (M = 823, SD = 95.05).

4.3 Materials

Four audio clips from the Test of English as a Foreign Language (TOEFL) exam were chosen as task materials. Two clips were conversations and the other two were lectures, both from academic settings. The length of the clips varied from two to five minutes. The average number of words spoken per utterance was about 14 words. Two clips (one conversation and one lecture) were randomly chosen for each experiment condition. Real-time transcripts of each audio clip were generated by Google speech recognition API. The word error rate (WER) of the ASR transcripts was about 10% on average.

4.4 Apparatus

Eye tracking was performed using the Tobii TX300 eye-tracker, which is composed of an eye-tracker unit and a 23", 1920x1080 widescreen monitor. The eye-tracker collects gaze data at 300 Hz and allows large head movements. The gaze data were logged by Tobii Studio. Before starting the tasks, we performed a 9-point calibration of the eye-tracker for each participant using Tobii Studio.



little or no experience with archaeological research but that's okay oh good that's a
[illegible] actually I'm volunteering for the project to get experience what kind of work

Figure 1. Participant in front of screen-based eye-tracker: with-transcript condition.

4.5 Procedure

Step 1 (real-time listening). The participants listened to the audio and pressed a button whenever they encountered a comprehension problem.

Step 2 (retrospective listening). The participants listened to the same audio again. This time using timestamps logged by the software, the computer automatically stopped at the places where they pressed the button during Step 1. At this point, the participants briefly explained what kind of problem they faced and how long it persisted. This step helped them re-experience the first step and recall their comprehension problems.

Under the with-transcript condition, participant eye movements were shown on top of the ASR transcripts. The participants were asked to explain their eye movements. They were also asked the following questions: Did you try to solve your problems using the ASR transcripts? Did the ASR transcripts help? If so, how, and if not, why not?

Step 3 (interview). The participants were handed perfect transcripts of the audio clip on a sheet of paper with markings that indicated their comprehension problems. Based on the marked-up transcripts, they explained the problems they faced during the listening task. This step was designed to get more detailed information about the comprehension problems mentioned in Step 2.

Under the with-transcript condition, they were also asked about their strategies for using the transcripts.

5. RESULTS

Our results are presented as follows. First, we describe how the NNSs used the ASR transcripts as well as the difficulties they faced. Then we report the types of listening comprehension problems that were generally solved by viewing the ASR transcripts. Finally, we describe how the NNSs were burdened when they failed to solve their comprehension problems using ASR transcripts.

5.1 How NNSs Used the ASR Transcripts

RQ1 asked how the NNSs used the ASR transcripts. To answer this question, we analyzed the post-task interviews and the gaze movement data of our participants. We found that they adopted different strategies when using the ASR transcripts. We further identified why they adopted different strategies with the transcripts.

Our analysis identified different NNS strategies for using the ASR transcripts; some participants generally followed the transcripts (Figure 2), while others only looked at them when needed (Figure 3).

For the NNSs who generally followed the transcripts, they either followed them while listening or gave up listening and concentrated on reading them. For the former group, the ASR transcripts seemed to increase their confidence in what they were hearing. For example, one NNS commented:

While listening, I read the transcripts to check if what I heard was correct. I felt relieved. (NNS 7)

The latter group seemed to have difficulty acquiring information from both the listening and reading channels. One NNS reported:

At first, I wanted to listen and I also wanted to read. I felt dizzy and couldn't catch up with the speech, so I gave up listening and focused on reading the transcripts. (NNS 1)

Figure 2 shows the gaze plot of one participant who followed the transcripts while listening. Even though she thought the transcripts were helpful, she complained that they caused an extra burden. The yellow rectangle represents the "transcript area," and the orange dots indicate her eye gaze locations. The size of the dots indicates the fixation duration.

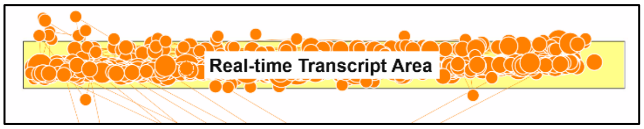


Figure 2. Gaze plot of NNS who followed transcripts.

Some participants only checked the transcripts when necessary, for example, when they encountered a problem or wanted to confirm what they had heard. One participant explained why he adopted such a strategy:

I felt the transcripts were a little distracting. So I focused on listening. If I encountered something I didn't understand, I read the transcripts. After reading, I went back to the listening mode. (NNS 5)

Figure 3 shows the gaze plot of one such participant. While the gaze plots in Figure 2 are centered around the transcript area, the gaze plots in Figure 3 are scattered below the transcript area and only occasionally jump into it.

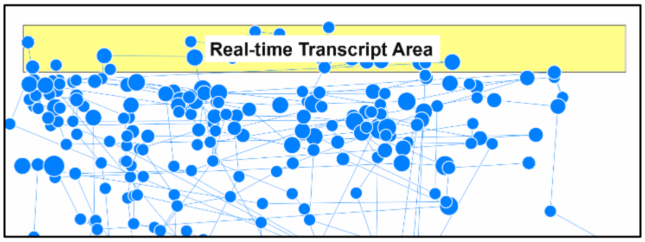


Figure 3. Gaze plot of NNS who did not follow transcripts.

As seen from Figure 3, since these participants mainly focused on listening, their gaze was scattered below the transcript area. When they encountered a problem, their gaze jumped to the transcript area to solve it. However, finding the right place was time-consuming and required some effort.

Sometimes I didn't know where the word I had a problem with was on the screen. I needed to search for it, and that was time-consuming. (NNS 5)

5.2 Listening Comprehension Problems Generally Solved by ASR Transcripts

RQ2 asked the following two questions: a) What types of listening comprehension problems can be solved by reading ASR transcripts? b) When NNSs fail to solve problems by reading them, what are the factors of failure? To answer these questions, we first identified the listening comprehension problems faced by NNSs in each condition and investigated the types of problems that significantly decreased when ASR transcripts were provided.

To identify each type of listening comprehension problem faced by participants during the listening task, we first transcribed the interview data and classified each problem based on the problem categories suggested by two previous works [6, 3]. We used them as a base because they also focus on the listening comprehension problems of NNSs that occur during their cognitive processing of speech input. Note that we added a new category "lack of vocabulary" to the previous categories [6, 3] because it can be solved by adding a dictionary function to the ASR transcripts [5]. All the interview data were coded independently by two coders, and all discrepancies were discussed until an agreement was reached.

We counted the number of times problems occurred based on the markups (times they pressed the button). In a few cases when participants described two problems for one markup, we counted it as two.

Table 2 shows the sample excerpts extracted from our interviews and the percentage of the occurrences of each problem (i.e., number of times each problem occurred/total number of occurrences). Item 14 is problems caused by ASR errors.

Figure 4 shows the distribution of the listening comprehension problems faced by NNSs under the without- and with-transcript conditions. Under the without-transcript condition, 372 problem occurrences were identified; under the with-transcript condition, 267 were identified, including ten problems caused by ASR errors.

To compare how the ASR transcripts changed the distribution of the problem occurrences, we first counted the problem occurrences of each participant. Next, we conducted a paired t-test (two-tailed) to see whether the average number of problem occurrences per minute changed between the two conditions. Results showed that the NNSs faced significantly fewer problems in the with-transcript conditions for three types of problems: "do not recognize words they know" ($p = 0.000$), "cannot chunk streams of speech" ($p = 0.005$), and "confused about unexpected word appearance" ($p = 0.034$).

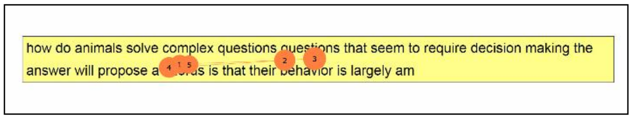
One common element to these problems is that they occur in the early stage of speech comprehension. In other words, they all occur during the cognitive processing phases of perception in language comprehension, which deals with the encoding of acoustic messages [1]. ASR transcripts benefit NNSs during such perceptual processing by transforming acoustic information into textual information.

Table 2. Example and percentage of listening comprehension problems faced by NNSs

Problem	Example interview excerpt	Without-transcript (%)	With-transcript (%)
1. Lack of vocabulary	I didn't know this word: "archaeology." I think it's a vocabulary problem. (NNS 2)	30.6%	45.3%
2. Do not recognize words they know	"Tackle" I knew, but I couldn't recognize it. If I had read it, I would've understood it. (NNS 1)	20.7%	3.4%
3. Unable to form a mental representation from words heard	I knew all of the words. But when combining them, I didn't understand them. (NNS 7)	15.9%	17.2%
4. Cannot chunk streams of speech	I couldn't catch "Joyce in a book called Dubliners." I couldn't divide that chunk into separate words. The words linked together. (NNS 6)	11.0%	7.1%
5. Understand words but not the intended message	Even though I knew the literal meaning, I couldn't understand it in this context. (NNS 19)	3.8%	3.7%
6. Concentrate too hard or unable to concentrate	The whole lecture was too long. At the end, I just couldn't concentrate. (NNS 9)	3.8%	2.6%
7. Neglect the next part when thinking about meaning	I was still thinking about the meaning of "beavers," and so I missed the subsequent words. (NNS 13)	3.2%	4.5%
8. Confused about unexpected word appearance	They were talking about "birds." Then suddenly "mouse" came out. I got confused. (NNS 9)	3.2%	1.1%
9. Unsure about the meaning of words	"Credit" could mean academic "credit" or financial related "credit." I wasn't sure. (NNS 7)	3.0%	2.6%
10. Do not understand subsequent parts of input because of earlier problems	I couldn't understand the meaning of "forage". Due to that, I was unable to understand the subsequent parts. (NNS 13)	2.7%	4.9%
11. Confused about the key ideas in the message	The lecturer explained and explained. I could understand the literal meaning. But I was confused about the key ideas. I didn't know what she wanted to say. (NNS 1)	1.1%	2.6%
12. Quickly forget what is heard	When the lecturer started talking about "another critical issue," I wondered what was the previous issue? But I'd already forgotten what it was. (NNS 3)	0.5%	1.1%
13. Miss the beginning of texts	The audio came too abruptly, and I missed the beginning. (NNS 16)	0.5%	0.0%
14. Confusion caused by ASR errors	I felt what I had heard was "mainly because," but the transcripts show "maybe cuz." The error hindered my understanding. (NNS 2)	0.0%	3.7%

Although most of the three types of problems were solved by showing the ASR transcripts, in some cases they weren't. To identify why, we analyzed the explanations of the NNSs to the interview question, "why didn't the ASR transcripts help you solve your problem?" and attributed three main factors that hindered the NNSs from solving them: ASR transcript errors, lack of time to identify the relevant parts of the transcripts or to consider the meaning of the transcripts, and confusion caused by no punctuation of the transcripts.

Figure 5 shows the gaze plot of a participant who failed to solve his comprehension problem due to ASR errors. In this example, the participant couldn't chunk "of course" from the speech, so he checked the transcripts. However, the transcripts showed an error: "a chorus," which increased his confusion.



how do animals solve complex questions questions that seem to require decision making the answer will propose a chorus is that their behavior is largely am

Figure 5. Gaze plot of NNS who failed to solve his problem due to ASR error.

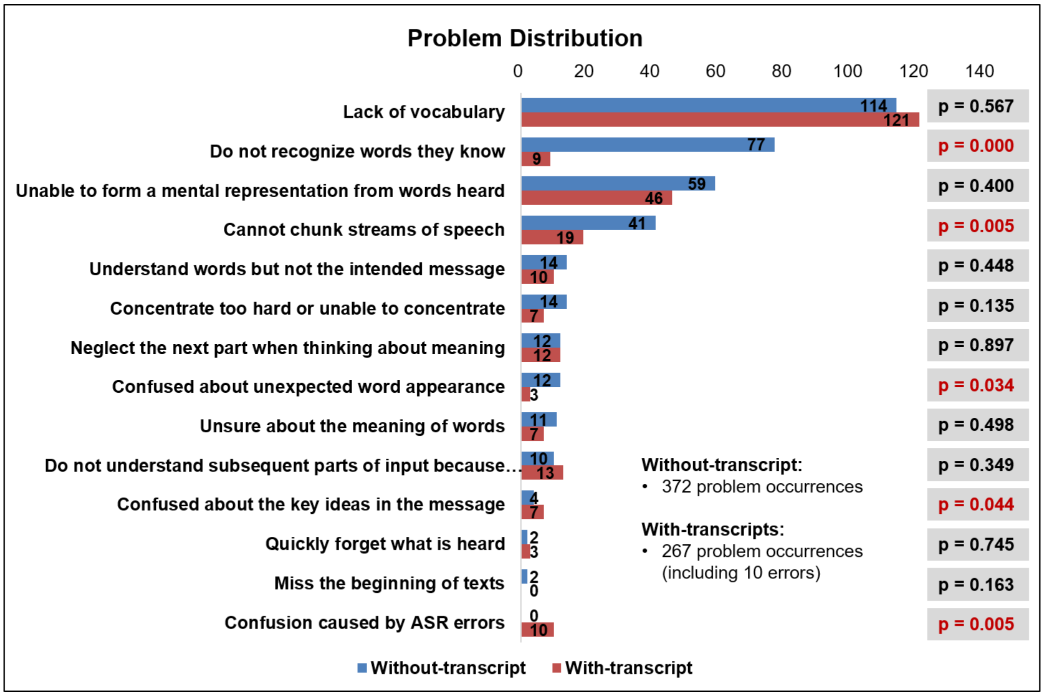


Figure 4. Distribution of listening comprehension problems faced by NNSs under without- and with-transcript conditions.

Table 3 summarizes the three factors and shows some extracts from the interviews. Since these factors hindered our participants from solving their problems, removing the influence of them would improve NNS comprehension.

Table 3. Factors that hindered NNSs from solving their problems

Factor	Example interview excerpt	Percentage (%)
ASR errors	I couldn't understand, so I checked the transcripts. After seeing errors in them, I became even more confused. (NNS 1)	61.3%
Lack of time	The sentence (I had a problem with) was a bit too long. Although I checked the transcripts, I didn't have enough time to think. (NNS 10)	25.8%
No punctuation	There was no period between "yet" and "the" in the transcripts. I thought they belonged to one sentence, but actually they belonged to two sentences, so I didn't understand. (NNS 4)	6.5%
Others		6.5%

5.3 NNSs' Workload for Using ASR Transcripts

RQ3 asked whether ASR transcripts placed an extra burden on NNSs when they failed to solve their listening comprehension problems. Note that we focus on the problems reported by the NNSs, which means that we discounted the problems which were solved by reading the ASR transcripts.

We consider "response time" one rough measure for NNS workload. Response time is the time taken to press a button when a NNS recognized a listening comprehension problem. The longer it takes to react (i.e., press the button), the heavier is the burden. We calculated the response time by counting the number of words spoken from where the problems started to where the NNSs pressed the button.

Figure 6 shows the average response time of each listening comprehension problem under the without- and with-transcript conditions. The response times of four types of listening comprehension problems significantly increased (t-test, two-tailed): "lack of vocabulary ($p = 0.000$)," "do not recognize words they know ($p = 0.023$)," "unable to form a mental representation from words heard ($p = 0.000$)," and "cannot chunk streams of speech ($p = 0.004$)".

This result suggests that even though NNSs tried to solve certain problems by reading the transcripts (e.g., the words they could not understand), these problems were not necessarily solved, rather their burden was increased. For example, although the "lack of vocabulary" and "unable to form a mental representation from words heard" problems tend to be unsolvable by reading the transcripts, NNSs seemed to read the transcripts to check whether they are actually unsolvable; this did not help them resolve the problem but only increased their workload.

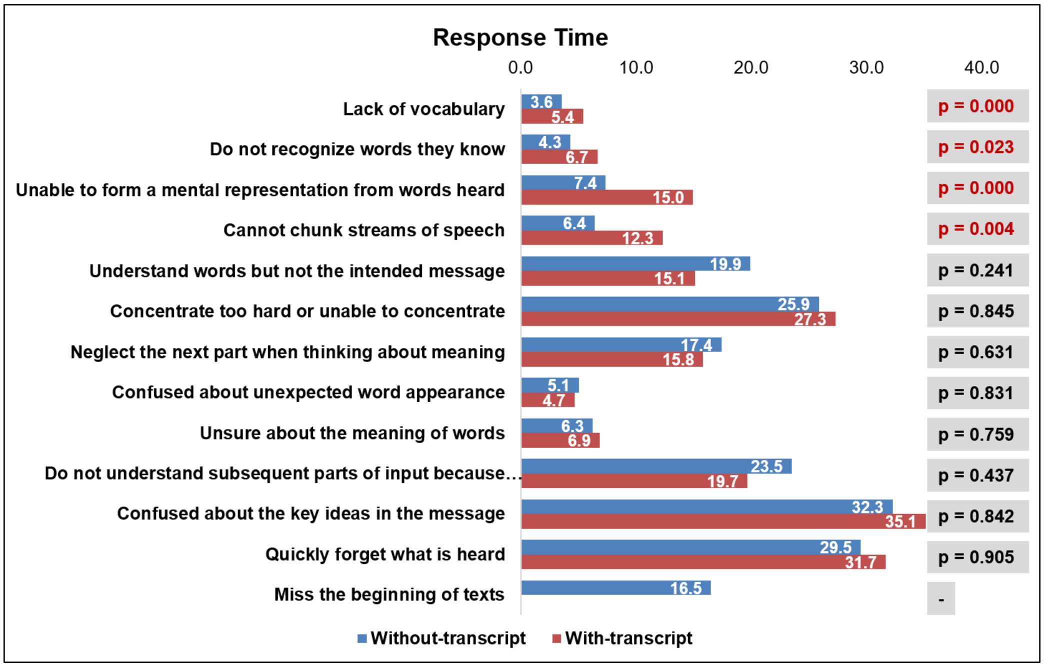


Figure 6. Response times of each listening comprehension problem under without- and with-transcript conditions.

6. DISCUSSION

To help NNSs solve their problems and reduce their burden, we suggest the following design implications.

6.1 Improving Effectiveness of Using ASR Transcripts

Reducing confusion caused by ASR errors by exploiting word recognition confident scores: ASR errors not only hindered NNSs from solving their problems but they also increased confusion and decreased NNSs' confidence in their own listening comprehension. When presenting ASR transcripts, we suggest exploiting the word recognition confident scores [12], which indicate the reliability of the recognition results. The lower the recognition confidence score, the greater is the likelihood of ASR error.

By embedding word recognition confident scores into the presentation of ASR transcripts, we might prevent NNSs from getting confused by ASR errors. For example, words with low confidence scores could be shown in gray and words with high confidence scores could be shown in bold.

Reducing NNSs workload by displaying keywords: We also found that most NNSs had difficulty simultaneously reading the ASR transcripts and listening to native speech. One possible explanation is that they lack sufficient cognitive resources to follow both text and audio, especially when the transcripts include errors and are shown with delays.

Previous studies found that some NNSs benefit more when only keywords are presented as captions rather than entire sentences [7]. This strategy may also be beneficial when presenting ASR transcripts to NNSs because the keywords could help them understand the key points of the conversations/lectures without attracting excessive attention.

Marking places where NNSs encountered problems to reduce search time: In our study, some NNSs concentrated on listening and viewed the transcripts only when they faced problems or wanted to confirm their listening (Figure 3). Although the NNSs seemed to intentionally adopt this viewing method to efficiently use the transcripts, shifting between multimodal contents seemed to place an additional burden on them. Indeed, NNSs had to search through the transcripts to spot the relevant place when they faced some problems. We suggest helping NNSs locate where they had problems in the transcripts. For example, when a NNS encounters a problem and presses the button, the system could automatically mark that place on the transcripts.

6.2 Introducing Other Technologies to Supplement ASR Transcripts

We found that the response times of some types of listening comprehension problems significantly increased. One possible reason is that even though NNSs tried to solve certain problems by reading the transcripts (e.g., the words they could not understand), the problems were not necessarily solved, rather their burden was increased.

For problems that were difficult or impossible to solve by viewing ASR transcripts, we suggest introducing other technologies to supplement ASR transcripts [10]. For example, the system could automatically provide dictionaries and images based on when a button was pressed.

Previous works suggested that eye tracking is not only useful for analyzing user behavior, but it can also be used as an input mechanism and a means of interacting with a program, a game, or some other technology [13, 17]. In our study, we observed some typical eye movements of NNSs when encountering certain problems: (1) fixating on a word or phrase; (2) looking back and forth at words or phrases; (3) shifting from no-transcript to transcript areas. These gaze patterns could be useful for detecting the types of problems experienced by NNSs. If a system could detect them in real time, it may provide a suitable support for NNSs to solve the problems without extra burdens. For example, if a NNS is fixated on a word, the system could automatically provide a dictionary definition or an image of it to support comprehension.

7. CONCLUSIONS

We investigated the impact of ASR transcripts on the listening comprehension of NNSs in our study. Through an exploratory analysis of the experiment data, we found that NNSs adopted different strategies when using the ASR transcripts; some followed them throughout the listening; some only checked them when necessary. Although the ASR transcripts did seem useful for NNSs to some extent, post-task interviews and gaze analysis of the participants revealed that the NNSs did not have enough time or cognitive resources to fully exploit the transcripts. For example, NNSs had difficulty concentrating on listening/reading or shifting between multimodal contents. We also found that the ASR transcripts helped the NNSs solve certain problems (e.g., "do not recognize words they know"), but imperfect ASR transcripts (e.g., errors and no punctuation) sometimes confused the NNSs and even generated new problems. Furthermore, even though NNSs tried to solve certain problems by reading the transcripts (e.g., the words they could not understand), the problems were not necessarily solved, rather their burden was increased. Based on our findings, we suggest implications for designing better multimodal interfaces for NNSs.

8. ACKNOWLEDGMENTS

This research was partially supported by a Grant-in-Aid for Scientific Research (S) (24220002, 2012-2016) from Japan Society for the Promotion of Science (JSPS).

9. REFERENCES

[1] J. R. Anderson. *Cognitive psychology and its implications.* WH Freeman, New York, 1995.

[2] A. Bloomfield, S. C. Wayland, E. Rhoades, A. Blodgett, J. Linck, and S. Ross. What makes listening difficult? factors affecting second language listening comprehension. Technical report, DTIC Document, 2010.

[3] X. Cao, N. Yamashita, and T. Ishida. How non-native speakers perceive listening comprehension problems: Implications for adaptive support technologies. In *International Conference on Collaboration Technologies,* pages 89–104. Springer, 2016.

[4] G. Gao, N. Yamashita, A. M. Hautasaari, A. Echenique, and S. R. Fussell. Effects of public vs. private automated transcripts on multiparty communication between native and non-native English speakers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* pages 843–852. ACM, 2014.

[5] G. Gao, N. Yamashita, A. M. Hautasaari, and S. R. Fussell. Improving multilingual collaboration by displaying how non-native speakers use automated transcripts and bilingual dictionaries. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3463–3472. ACM, 2015.

[6] C. C. Goh. A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1):55–75, 2000.

[7] H. G. Guillory. The effects of keyword captions to authentic French video on learner comprehension. *Calico Journal*, pages 89–108, 1998.

[8] A. Hautasaari and N. Yamashita. Do automated transcripts help non-native speakers catch up on missed conversation in audio conferences? In *Proceedings of the 5th ACM international conference on Collaboration across boundaries: culture, distance & technology*, pages 65–72. ACM, 2014.

[9] E. Hinkel. *Handbook of research in second language teaching and learning*, volume 2. Routledge, 2011.

[10] T. Ishida. *The language grid: Service-oriented collective intelligence for language resource interoperability*. Springer Science & Business Media, 2011.

[11] V. Kalnikaitė, P. Ehlen, and S. Whittaker. Markup as you talk: establishing effective memory cues while still contributing to a meeting. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 349–358. ACM, 2012.

[12] J. C. Lai and J. G. Vergo. Speech recognition confidence level display, Dec. 21 1999. US Patent 6,006,183.

[13] L. Lorigo, M. Haridasan, H. Brynjarsdóttir, L. Xia, T. Joachims, G. Gay, L. Granka, F. Pellacini, and B. Pan. Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59(7):1041–1052, 2008.

[14] Y. Pan, D. Jiang, L. Yao, M. Picheny, and Y. Qin. Effects of automated transcription quality on non-native speakers' comprehension in real-time computer-mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1725–1734. ACM, 2010.

[15] J. Rubin. A review of second language listening comprehension research. *The modern language journal*, 78(2):199–221, 1994.

[16] Y. Takano and A. Noda. A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology*, 24(4):445–462, 1993.

[17] I. Umata, S. Yamamoto and M. Nishida. Effects of language proficiency on eye-gaze in second language conversations: toward supporting second language collaboration. In *Proceedings of the ACM on International Conference on Multimodal Interaction (ICMI)*, pages 413–420, 2013.

[18] P. Winke, S. Gass, and T. Sydorenko. Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, 97(1):254–275, 2013.

[19] N. Yamashita, A. Echenique, T. Ishida, and A. Hautasaari. Lost in transmittance: how transmission lag enhances and deteriorates multilingual collaboration. In *Proceedings of the 2013 conference on Computer Supported Cooperative Work*, pages 923–934. ACM, 2013.

[20] Y L. Yao, Y.-x. Pan, and D.-n. Jiang. Effects of automated transcription delay on non-native speakers' comprehension in real-time computer-mediated communication. In Human-Computer Interaction–INTERACT 2011, pages 207–214. Springer, 2011