Chen, W. et al. (Eds.) (2017). Proceedings of the 25th International Conference on Computers in Education.
New Zealand: Asia-Pacific Society for Computers in Education

Usability and Learning Effect Evaluations of an Electrical Note-Taking Support System with Speech Processing Technologies

Hiromitsu NISHIZAKI[a*] & Yosuke NARITA[b]
[a] *Integrated Graduate School of Research, University of Yamanashi, Japan*
[b] *Integrated Graduate School of Medicine, Engineering, and Agricultural Sciences, University of Yamanashi, Japan*
*hnishi@yamanashi.ac.jp

Abstract: This paper describes the evaluation of an electronic note-taking support system that works on a tablet/laptop computer in a real classroom environment. The distinguishing characteristic of the proposed electronic note-taking support system is the incorporation of a few state-of-the-art speech processing technologies. One is deep-neural-network-based automatic speech recognition for transcribing talks. Such transcriptions are very useful for not only note-taking but also recovering missing words. The other speech processing technology is keyword search in the recorded speech. We had already developed the system previously and evaluated it. In this study, we improved the previous support system from the viewpoints of usability and visibility of the system user interface. The improved system was evaluated in the subjective experiment in which listeners (students) who attended the real lectures. In the experiment, not only the usability of the system but also the learning effect for knowledge acquisition were evaluated. The experimental results showed that the improved note-taking support system improved usability and visibility compared to the previous system and the traditional note-taking method. Moreover, the students claimed that the system helped them acquire knowledge about the lectures they attended.

Keywords: Learning effect, note-taking support system, user interface, usability evaluation

1. Introduction

In our daily life, we have many opportunities to take notes such as in classroom lectures and meetings. When we take notes, we usually use a paper notebook with a pen (or pencil) or an electronic device such as a tablet or a laptop personal computer (PC). However, often, we miss words (contents) uttered by a speaker.

Nowadays, most lecturers make slides for lectures using presentation software such as Microsoft PowerPoint®. Comparing a lecture with presentation software to a lecture with a blackboard, most lecturers using presentation software talk and turn slides faster than those not using presentation software. Therefore, most listeners (lecture attendees) may at times miss parts of a talk by a lecturer using presentation software. Moreover, some listeners squiggle characters on a paper notebook. These listeners may not be able to understand the characters, some of which could be messy.

Various types of note-taking support systems and applications (Zyto, et al., 2012), (Livescribe, Inc., 2017), (Google Inc., 2017), (Echo360, 2017) have been developed for solving these problems. They provide listening support and can be used to take notes efficiently.

For example, "NB" (Zyto, et al., 2012) provides a sharing function among listeners who take the same class. The focus of this application is on the sharing function; therefore, it is unsuitable for taking notes at a lecture in real-time. "Livescribe 3 SmartPens" (Livescribe, Inc., 2017) is a commercial note-taking tool that consists of a special pen and a special notebook. This tool records an entire talk by a lecturer; therefore, the user can listen to the talk when reviewing the notes taken using the tool. The "Google Keep" (Google Inc., 2017) application allows users to take notes on portable devices. Users can easily insert a photo in a note and superimpose text onto it. In addition, voices can be recorded and stored in the application. Finally, “Echo360” is an enterprise system for classroom lectures that can record the video and the speech of a lecturer and distribute the recorded content to students. Students can review the lecture video and communicate with the lecturer through the system whenever they like. However, Echo360 does not have any note-taking function.

Although there are a few systems and applications that allow users to review lectures and perform and note-taking tasks, none of these applications/systems have the following characteristics from the viewpoint of classroom use:

- Ease of note-taking
- High visibility of note contents
- Recovering missing words

Therefore, Ota, Nishizaki, and Sekiguchi (2011) developed a note-taking support system for recording classroom lectures. Listeners (note takers) can efficiently use captured images of slides projected onto a screen or figures/characters written on a blackboard for note-taking. In addition, they can also use transcriptions, which are generated by an automatic speech recognition (ASR) system, for note-taking. These characteristics the support system enable lecture listeners to make visible note content without missing any of the lecturer's talk because they can concentrate on the lecture. The system has a function that shows the ASR result on a graphical user interface. Therefore, listeners can not only focus on the lecturer's talk but also watch ‘‘*visible words*’’ by the lecturer. In addition, the visible words can be used to create a memo by touching the displayed words.

The support system has other useful functions to support note-takers. For example, a spoken term detection (STD) technology (Akiba et al., 2011), a spoken keyword search/extraction technique, can find the occurrence positions of a query term (keyword) in a recorded lecture talk. This STD enables us to cue a recorded talk easily when a listener wants to review the note content. The efficiency of each useful function implemented in the support system has been already evaluated in previous works (Ota, Nishizaki and Sekiguchi, 2011), (Yonekura et al., 2013).

However, the previous evaluation results showed that the support system was not very handle-able, in other words, it is hard for users to operate the system, and the user interface of the system was not very good compared to the traditional note-taking method of using a pen and paper.

Our note-taking system provides four useful functions, namely, capturing images of projected slides or hand-written characters on a blackboard, ASR of a lecturer’s speech, hand-written and one-touch memos using visible words generated by ASR, and keyword search. These functions are unique to our note-taking system. Moreover, no studies have investigated the effects of note-taking support systems and applications such as the one described above on learning. Therefore, in this study, we experimentally demonstrate the effectiveness of learning using such applications by asking a few students to use our note-taking system in a classroom.

We improve the user-interface of the previous note-taking support system (Yonekura et al., 2013) and re-evaluate the improved system in a real classroom (the previous system was not evaluated in a real classroom but in a simulated classroom). In addition, we investigate the effect of the proposed note-taking system on learning.

The evaluation results obtained in the real environment (real classroom) showed that the latest support system improved usability and visibility compared to the previous system and the traditional note-taking method. Moreover, the evaluation results show that our system helped users acquire more knowledge about the lecture compared to a traditional note-taking method.

2. Note-taking Support System

2.1 System Outline

Figures 1 and 2 show the processing flow of our note-taking support system and a screenshot of the note-taking device, respectively. The newest version of the system, which has been upgraded from previously reported versions (Ota, Nishizaki and Sekiguchi, 2011), (Yonekura et al., 2013) by following the result of previous subjective experiments, was used in this study. As shown in Figure 1, a

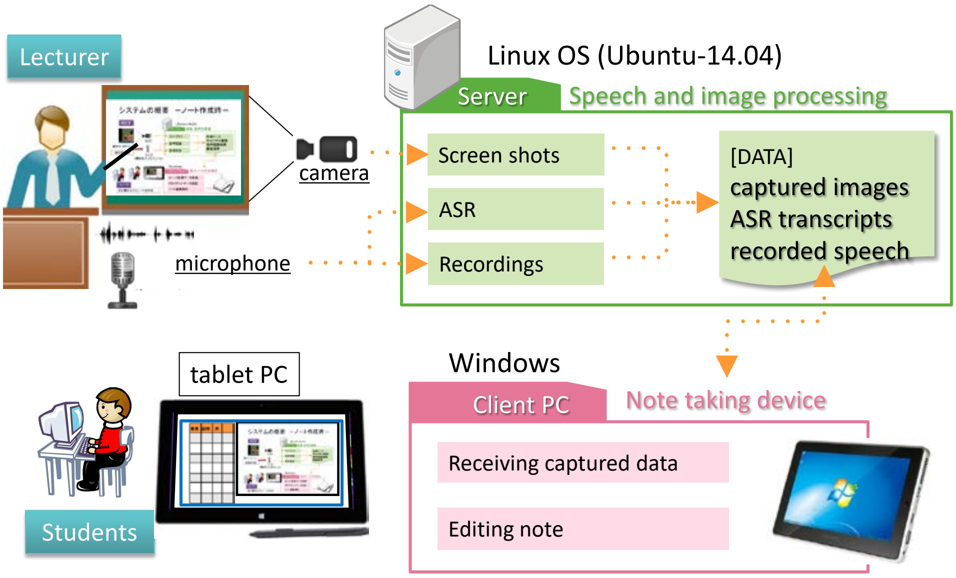

Figure 1. Outline of electronic note-taking support system.

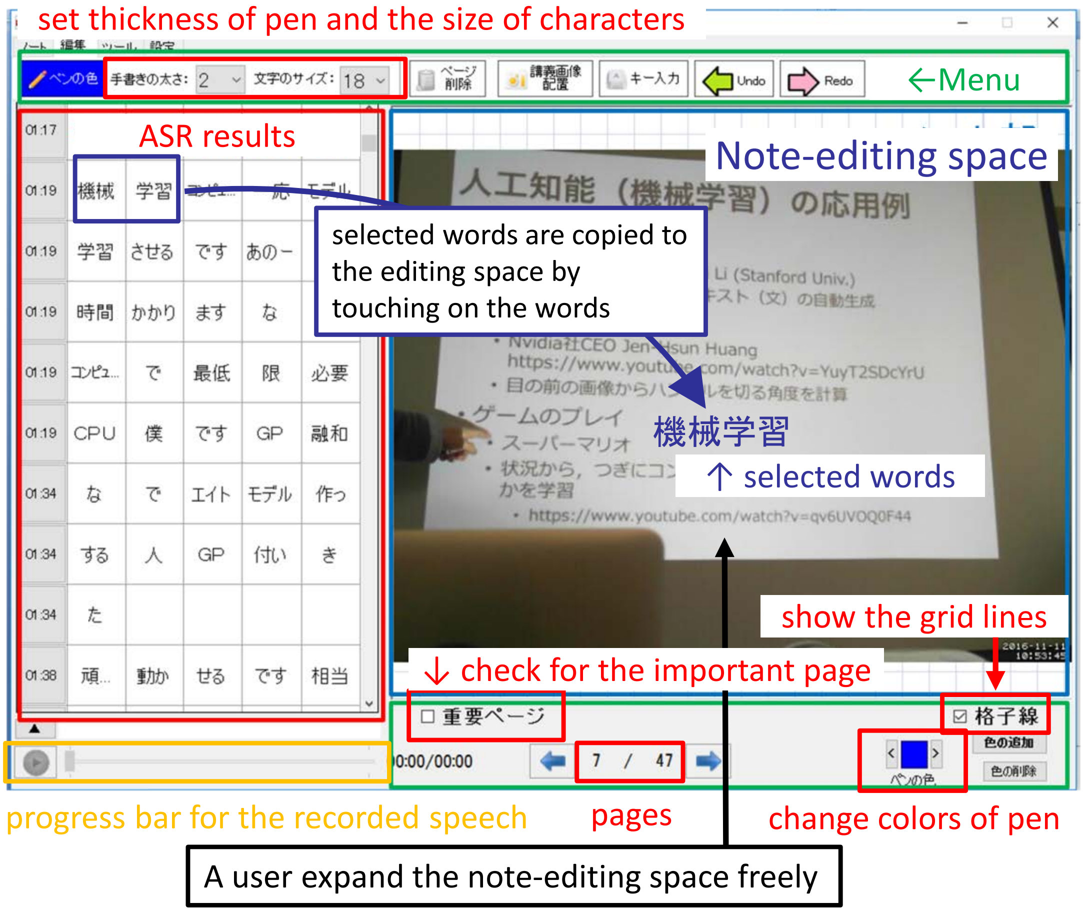

Figure 2. Graphical user interface of electronic note-taking support system.

talk by a lecturer is recorded automatically by a server, which also captures screenshots of the projector screen (or a blackboard) at intervals of 3 s. After the speech is stored on the server, it is transcribed by an

ASR system. The transcription is then transferred to a note-taking device. At this time, the speech is also converted to a PTN-formed index (Natori et al., 2013) for STD by using the ten ASR systems.

The word sequences generated by the ASR system are filtered by rules on the note-taking device and are displayed on the note-taking device. A user can record a note by simply touching or tracing words on the screen. Therefore, handwriting and keyboard operations are unnecessary. The entire speech is recorded on the server, and the locations of recognized words in the speech are identified using word-to-speech alignment information generated by ASR. After note-taking is completed, the recorded speech is transferred to the device. Therefore, users can easily play the speech back from any chosen point. It is not necessary to listen to the entire speech when checking a note.

2.2 Speech Interface

Our note-taking support system has a speech interface as well. This speech interface captures speech, records notes, and performs speech recognition. The recognized words are displayed on the screen. Users can record a note by simply touching or tracing relevant words with their finger. Therefore, this reduces the burden on the user of recording notes while listening to a speech.

However, this system depends on speech recognition. If the words the user wants to associate a note with are not recognized correctly, the note cannot be recorded. Speech recognition errors are unavoidable. Therefore, it is undesirable to completely trust the ASR technology.

The key concept of the system is to use ASR as an accessory function. The ASR system is designed to help users record notes; however, it must disturb their work as little as possible. If the words that the user wants to associate a note with are not recognized correctly, the user does not need to touch or trace the words on the screen. Rather, the user can record a note using a keyboard or an electronic pen. This concept of using ASR as an accessory function differs from other systems with speech interfaces. The usability of those system depends entirely on the output of the ASR system, and the system usability worsens if ASR performs poorly.

Speech is recorded simultaneously as ASR operates. Therefore, users can listen to the speech many times if needed. Each annotated word has information that describes where the word is located in the speech. Therefore, system users can easily play the speech starting from any specified word.

2.3 Keyboard and Handwritten Input

As described in Section 2.2, a note-taker can record a note by using a (hardware or software) keyboard or an electronic pen in addition to ASR. If words are transcribed faultily by ASR, the user may input words by using a keyboard or by writing.

A note-taker can draw graphics, such as underlines, on the screen. They can also write words and circle finger-touched or handwritten words for emphasis.

In this study, all handwritten and keyboard-input words are called objects. If an object is drawn on the system screen, it is correlated with the relative time from the beginning of the recorded speech. Therefore, if a user touches an object, they can listen to the speech starting from the object-specified time.

2.4 Recorded Note Review

A user can see a recorded note while listening to the recorded speech. As mentioned in Sections 2.2 and 2.3, a system user can also play speech starting from the time specified by the location of the word or object.

In addition to the notes recorded by the user, all recognized words from the speech are stored in the system. This enables the user to search for a specified word from recorded speech. When the user enters a search term in the search window, the locations of the word are identified if it is recognized correctly.

However, users cannot search for words transcribed incorrectly. To avoid this, we have incorporated the STD technique (Natori, et al., 2013), which is sufficiently robust to speech recognition errors.

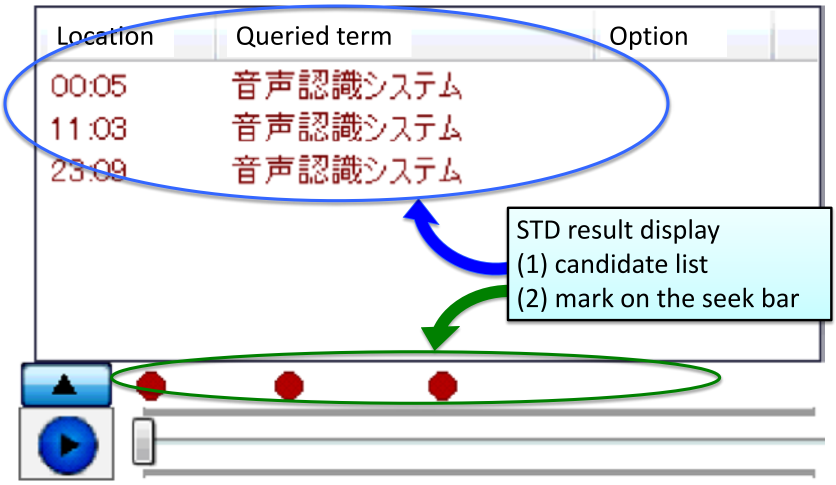

Figure 3. User interface that shows keyword search (STD) results.

2.5 Keyword Search Interface

When a user searches a queried term using the STD engine, the search results are shown on the speech playback seek bar. Figure 3 shows an example of the interface for outputting STD search results. The search results are shown in list form. Furthermore, the expected locations of the queried term, as determined by the STD engine, are indicated on the seek bar (small circle).

2.6 Other functions

Our note-taking system has the following other useful functions:

- Overhead-view of all pages
  A user can display a thumbnail list of all pages. Therefore, they can smoothly move to the target page when they want to review a note.
- Marking important pages (bookmark)
  A user can bookmark favorite or important pages in a note.
- Register and emphasize important words from an ASR transcription
  A user can register important (or key) words into the system in advance. If a lecturer utters these words, the system alerts the user and the words are emphasized with bold font and displayed in the ASR result area.
- Change color of pen easily by touching a "<" or ">" bottom (see Figure 2). The color of the pen is set to blue by default, and the only other registered color is red. Of course, a user can register other colors freely. It is well-known that blue and red colors are brain-stimulating; therefore, these colors enhance the effectiveness of learning (Mehta and Zhu 2009).

3. Improvement over Previous Note-taking System

We evaluated the previous note-taking system (Yonekura et al., 2013). The evaluation results showed that the previous system reduced users' note-taking burden compared to the traditional (hand-written) method of note-taking. However, it did not fare well in terms of operability and the visibility. Therefore, we had to improve the previous system. A few test subjects indicated the following points of improvement:

- Expansion of note-taking space
- Show grid lines
- Ability to change color of characters easily

- Create a new page when a user captures an image
- Improve ASR performance

We have refined the previous note-taking system based on the above suggestions, and we will compare the usability of the revised version of the system to the previous one.

4. System Evaluation

We performed a subjective experiment to evaluate the usability of the newest note-taking support system and the visibility of the user interface on the system in a real classroom environment.

In the experiment, we test the usability of the system and its effect on learning in lectures by comparing three note-taking methods: the newest support system, the previous system, and the traditional note-taking method (pen and paper).

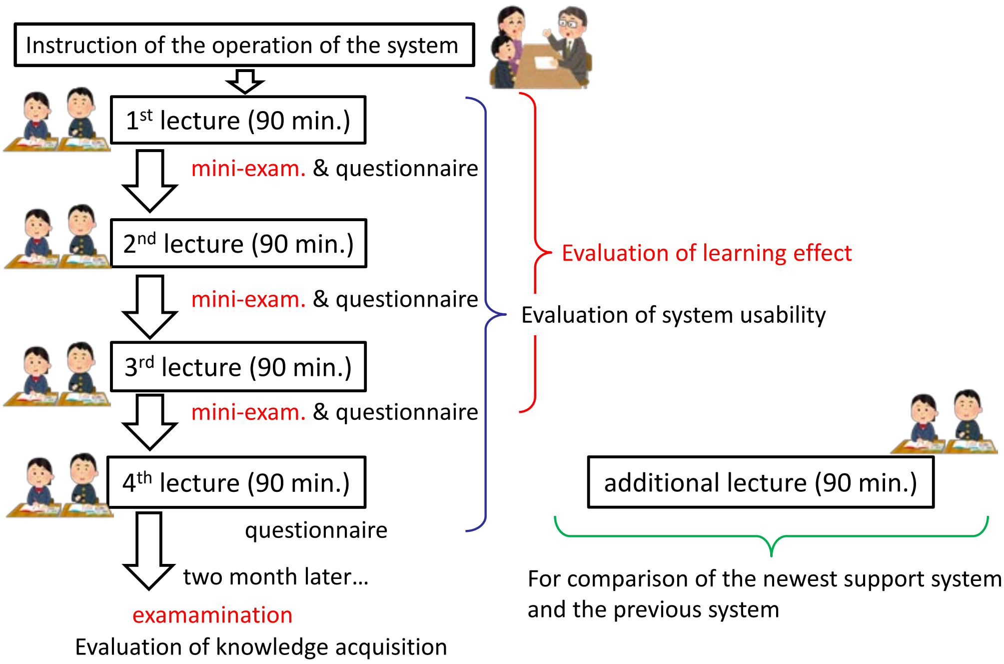

Figure 4. System evaluation design in subjective experiment.

Table 1: Experimental conditions for note-taking evaluation.

Lecture name	Fundamentals of machine learning
Number of lectures	Five (only one lecture was used for comparison of the newest system and the previous system)
Duration time	90 minutes for all lectures
Number of subjects	12 (undergraduate students who majoring in Mechatronics engineering)
ASR system	Julius rev. 4.3.2 (Lee & Kawahara, 2009)
Acoustic model	Triphone-based Deep Neural Network (DNN)-Hidden Markov Model (HMM) (Hinton et al., 2016)
Language model	Word-trigram trained from Corpus of Spontaneous Japanese (Maekawa, 2013)

4.1 *Experimental Setup*

Figure 4 shows the outline of the subjective experiment. In addition, Table 1 lists the conditions of the note-taking evaluation. The subjects were 12 undergraduate students majoring in Mechatronics engineering. All students were not accustomed to operating a computer, but they used their smartphones regularly. All students attended the four classes (lectures) on fundamentals of machine learning, and each lecture lasted 90 min. After the lecture (except for the fourth lecture), all students took a mini-examination related to the lecture and answered a questionnaire on system usage. They were allowed to review note contents (generated by the system or hand-written) during the examination. The scores of the mini-examinations were a good indicator of the effect of the system on learning during the lectures. The questionnaire results will be used to assess system usability.

In these four lectures, all subjects used the newest (refined) note-taking support system and the traditional method. For a comparison of the newest and the previous systems, we scheduled an additional lecture in which the subjects evaluated the previous system.

We instructed (for a few minutes) the all students about the use of the note-taking system before the lectures. However, all students did not practice system operation for long. Instead, we provided a manual in which system operating instructions were explained to all students. They were free to read the manual whenever they liked during the lectures.

Table 2 summarizes the number of students using the various note-taking methods. As shown in Table 2, three or four students per lecture (except for the 5th lecture) used the newest system. All students evaluated the newest system once. For assessment of the improvement in the user interface, we compared the newest note-taking system and the previous version. Five students used the previous system in the fifth lecture. All five students had used the newest system beforehand.

Table 2: Number of students per lecture.

	Newest system	Paper pen	Previous system
1st lecture	4	10	N/A
2nd lecture	3	5	N/A
3rd lecture	4	7	N/A
4th lecture	4	N/A	N/A
5th lecture	N/A	N/A	5

Table 3: Objective evaluation results.

	Newest system	Paper pen	Previous system
How many times missed lecturer's utterances [times]	**2.9**	4.8	3.0
Time rate of listening lecturer's talk in concentration [%]	**73.1 %**	55.7 %	66.6 %
Ease of reviewing note [10-step evaluation]	**9.4**	5.0	8.7
Usability of systems	**8.4**	N/A	7.0
Visibility of user interface	**8.4**	N/A	6.7

4.2 Experimental Results

4.2.1 Objective Evaluation

Table 3 summarizes the objective results of the questionnaires in which we assessed five items as follows:

- How many times did each student miss the lecturer's utterances (average of all students)?
- Time rate of listening lecturer's talk in terms of concentration (subjective)
- Ease of reviewing notes
- System usability (user-friendliness)
- Visibility of graphical user interface (how easy was it to catch the ASR-based caption and how easy was it to review recorded notes?)

As can be inferred from Table 3, the newest system bested the paper-based (traditional) note-taking method on all evaluation items. Because electronic systems can record and transcribe the entire lecture by means of the ASR system, the students who used the systems were able to recover missing words. In addition, the students could capture images of the slides projected on the screen easily; therefore, they concentrated on listening to the lecture.

By contrast, in the comparison of the newest system with the previous one, the new system bested the previous one on all evaluation items. This result shows that the previous system was improved successfully. The upgrades in terms of the five points described in Section 3 were very effective for improving the usability and the visibility of the user interface of the note-taking system.

4.2.2 Evaluation on Learning Effect

Table 4 shows the accuracy rates of the mini-examinations conducted soon after the end of each lecture. As can be inferred from Table 4, the students who used the note-taking system obtained better results in each examination. From Table 3, when the students used the electronic note-taking system, the time rate of concentrating on listening to the lecture was increased drastically compared to that with the hand-written method. In other words, the students who used the system could concentrate on the lecture, resulting in improvements in their examination scores. In addition, the ASR-based captions allowed the students to easily create a memo on a captured image. They could write words or phrases in the note-editing space by using ASR-based captions. This let the students focus on the talk. From the results, the electronic note-taking system helped the subjects study the topic of the lecture more effectively compared to the traditional method of taking notes using pen and paper.
Table 5 also shows the accuracy rates of mini-examinations conducted two months after the end of the fourth lecture. All questions were the same as the ones in Table 4. Different from the examination in Table 4, the students were not allowed to review the notes made by themselves. They answered the questions from their memory only. As may well be expected, the accuracy rates in all examinations degraded compared to the results in Table 4. However, the subjects who used the system scored significantly higher than the other subjects in two examinations (1st and 2nd examinations). However, in the 3rd lecture, the average of the examination scores did not change compared to that in Table 4 (from 66.7 to 66.7). This is because the students who did not use the system studied deep learning programming after finishing the lectures, while the questions in the 3rd examination were related to autoencoders and recurrent neural networks. In other words, the subjects gained knowledge about the 3rd examination from elsewhere. This is a special case. The results in Table 5 indicate that our system was very useful for acquiring knowledge about lectures, even when considering the special case.

Table 4: Mini-examination results soon after end of lectures (accuracy rates [%]).

	w/ System	w/o System
1st lecture	**65.0**	56.4
2nd lecture	**100**	71.2
3rd lecture	**95.8**	66.7
Average	**88.9**	63.7

Table 5: Mini-examination results about two months after end of lectures (accuracy rates [%]).

	w/ System	w/o System
1st lecture	**53.3**	40.0
2nd lecture	**78.6**	28.6
3rd lecture	61.1	**66.7**
Average	**64.3**	45.1

5. Conclusions

This paper described evaluations of an electronic note-taking support system in a real classroom environment. The support system allows users to take notes quickly by using captured images of slides projected on a screen and transcriptions created by the ASR system. Recording and transcription of a lecturer's voice by the ASR system can help a note-taker to not miss the lecturer's talk. Therefore, a note-taker can concentrate on the lecture to a greater extent than they would be able to do with the traditional note-taking method.

The experimental results of usability and visibility showed that our system provides a better user interface for creating or reviewing notes. Our system did not prevent the system users from listening to the lectures. In addition, the results of the examination scores claimed that the use of our system helped users to create memories of lecture contents, and finally, the system facilitated the students in acquiring knowledge.

In this paper, we evaluated the system on five lectures of one topic (machine learning). In future works, we will evaluate the system in other lectures and we will increase the number of students.

Acknowledgements

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) Grant Number 17H01977 and Grant-in-Aid for Scientific Research (C) Grants Number 15K00254.

References

Akiba, T., Nishizaki, H., Aikawa, K., Kawahara, T., & Matsui, T. (2011). Overview of the IR for Spoken Documents Task in NTCIR-9 Workshop, *Proceedings of NTCIR-9 Workshop Meeting,* 223–235.

Echo360 (2017). Ech360: Education's First Video and Active Learning Platform, from https://echo360.com/

Google Inc. (2017). Google Keep, from https://keep.google.com/

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., & Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups, *IEEE Signal Processing Magazine*, 29(6), 82–97.

Lee, A. & Kawahara, T. (2009) Recent Development of Open-Source Speech Recognition Engine Julius, *Proceedings of the 1st Asia-Pacific Signal and Information Processing Association Annual Summit and Conference,* 131–137.

Livescribe, Inc (2017). Livescribe 3 Smartpen, from https://www.livescribe.com/en-us/smartpen/ls3/

Maekawa, K. (2003). Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, 7–12.

Mehta, R. & Zhu, J. R. (2009). Blue or Red? Exploring the Effect of Color on Cognitive Task Performances. *Science Magazine*, 323(5918), 1226–1229.

Natori, S., Furuya, Y., Nishizaki, H., & Sekiguchi, Y. (2013). Spoken Term Detection Using Phoneme Transition Network from Multiple Speech Recognizers' Outputs. *Journal of Information Processing,* 21(2), 176–185.

Ota, K., Nishizaki, H., & Sekiguchi, Y. (2012). Development of Note-Taking Support System with Speech Interface. (2012). *Proceedings of the 4th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 4 pages.

Yonekura, C., Furuya, Y., Nishizaki, H., & Sekiguchi, Y. (2013). Evaluation of the Usefulness of Spoken Term Detection in an Electronic Note-Taking Support System. *Proceedings of the 5th Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 4 pages.

Zyto, S., Karger, D., Ackerman, M., & Mahajan, S. (2012). Successful Classroom Deployment of a Social Document Annotation System, *Proceedings of CHI 2012,* 1883–1892.