Koemiru: ICT Tool for Special Needs Schools

Yuichi Muto, Kenta Tetsuzaki, Takako Sato, Masayuki Sugizaki, and Yumiko Matsuura

Abstract

This article introduces Koemiru, our ICT (information and communications technology) system developed for special needs schools. Koemiru, which literally means "see the voice" in Japanese, uses speech recognition technology to support hearing-impaired elementary school pupils. Our system converts utterances into text and displays them on an interactive whiteboard and portable game terminals. Validation experiments were conducted in Okinawa and Tottori to identify the strengths and weaknesses of Koemiru.

1. Introduction

1.1 Challenges in using ICT in the education field

The NTT Group is implementing the Education Square × ICT project with the aim of leveraging information and communications technology (ICT) to develop new learning methods. Through this project, we found that there was a compelling need to use ICT to support the teachers and pupils of special needs schools, but that existing solutions were unable to meet their requirements.

1.2 Our solution for special needs schools

We surveyed special needs schools and consulted with NTT CLARUTY Corporation and found there was a desire for ICT to be applied to meet the fundamental needs of their pupils. These needs include a *visible voice* for hearing-impaired pupils, a *hearing character* for sight-impaired pupils, and *easier conversation* for developmentally disabled pupils. Our market research activities clarified the importance of using ICT technologies for conversation support and for providing information for people with disabilities by alternative methods. We also noticed that pupils wanted to use devices that were already popular at home and in their daily life at school as well. NTT has been researching free conversation speech recognition technology [1]–[3] and has already developed a speech auto-answer system that is used over the tele-phone. In addition, because schools for the deaf are much quieter than conventional schools, their environment is appropriate for the use of speech recognition technology. Accordingly, we decided to create a conversation support system for hearing-impaired pupils by applying speech recognition technology (**Fig. 1**). We focused on the three key goals of visualization of the teacher's voice, utterance training, and conversation support in everyday life.

2. Koemiru

2.1 Overview

We developed Koemiru (see the voice) to support hearing-impaired pupils in special needs schools [4]. The system consists of various servers and terminals. A speech recognition server and administration server are cloud services (**Fig. 2**). The teachers use a personal computer (PC) or smartphone as a terminal, the pupils use Nintendo DS portable game players as terminals, and the classroom has an interactive whiteboard. We chose the game terminal for three reasons. First of all, it is highly popular with pupils, and they can use it easily. Second, it is a familiar device in the community and does not look out of place when the pupils use it outside school. Third, it is very durable and easy to reboot.

2.2 Main functions

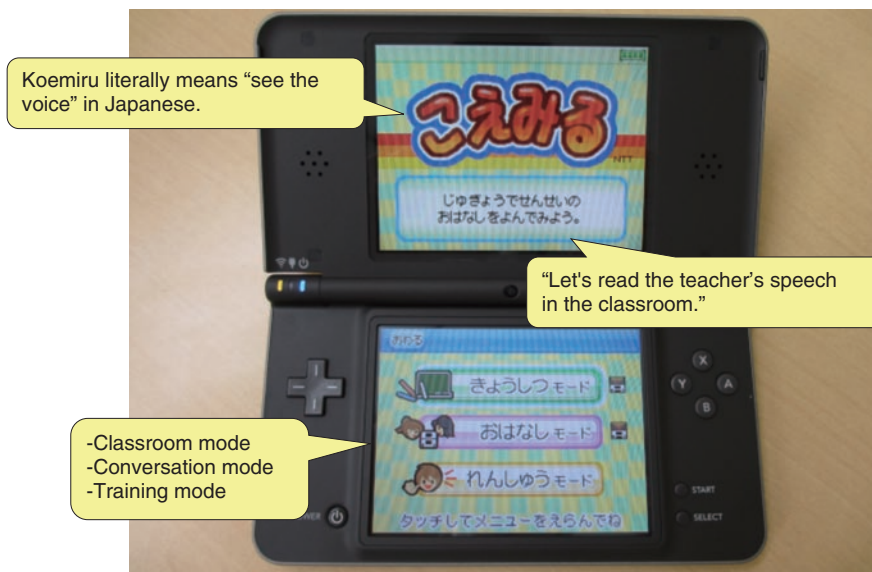Koemiru has three functions: classroom mode,

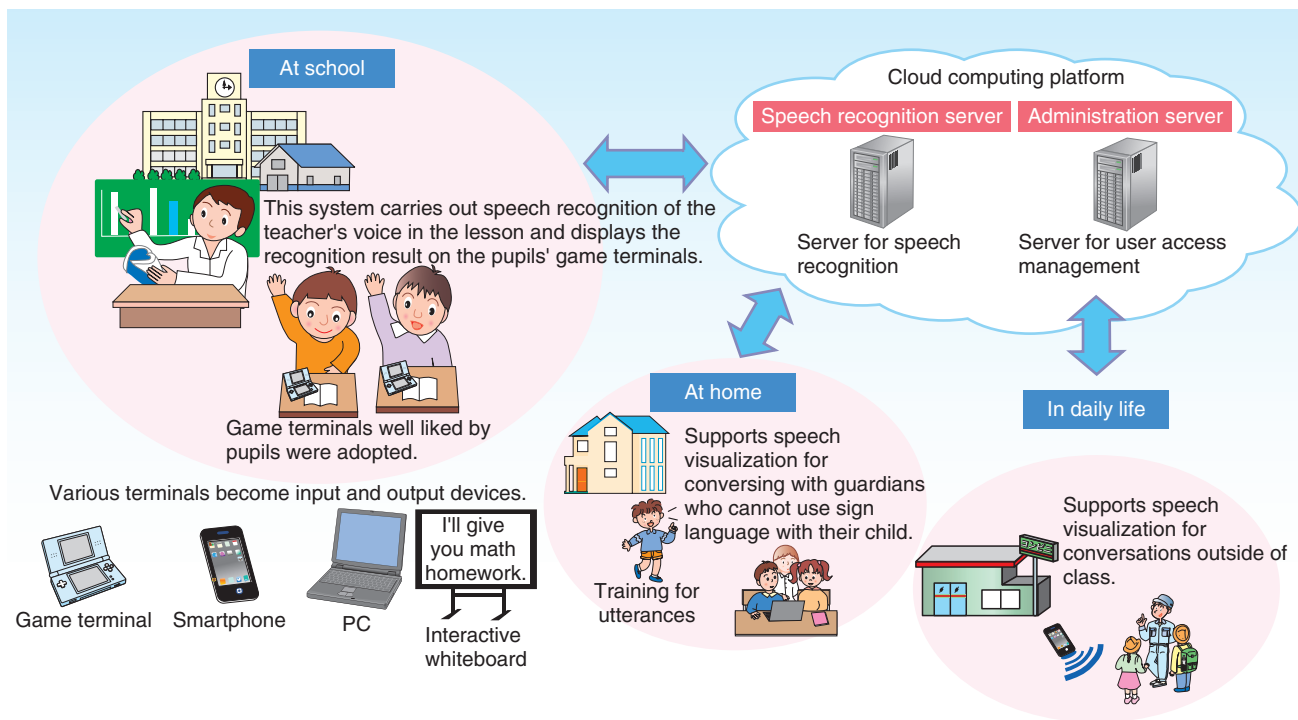Fig.1.   Welcome screen of Koemiru on a game terminal.

Fig. 2.   Overview of Koemiru.

training mode, and conversation mode.

Classroom mode is used in school lessons. When the teacher speaks into the wireless microphone, his or her voice is sent to the speech recognition server on the cloud computing platform and is recognized and output as text. The text output is displayed on the
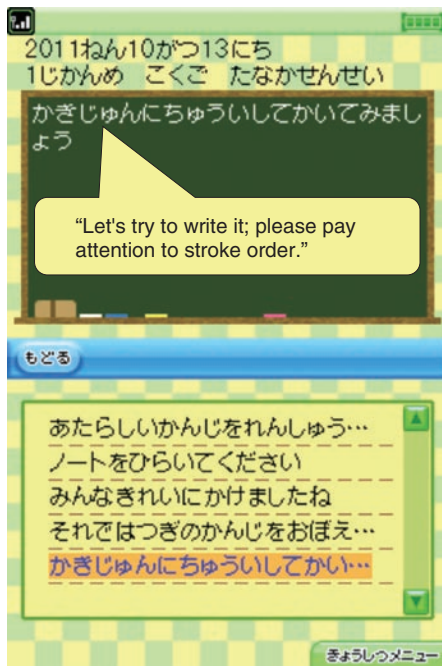
Fig. 3.   Example of classroom mode on a game terminal.

interactive whiteboard and on the portable game terminals of authorized pupils. Japanese uses both kanji and kana (hiragana and katakana). Because the pupils can read only kana, teachers can choose either kana alone or both kanji and kana when outputting the recognition results (**Fig. 3**). This selection is managed by the administration server; the recognition results are displayed only to authorized teachers and pupils. Additionally, the teacher can correct any mistakes in the recognition results via this server.

When hearing-impaired pupils perform voice training, they access the training mode on the Nintendo DS, as shown in **Fig. 4**. If a pupil utters something, the recognition results are displayed on the DS. Voice-volume training is supported by the visualized volume bar.

When pupils are at home or outside of school, they can use the conversation mode to communicate with other people. They can pose questions on the Nintendo DS using stored questions or by entering the text themselves. When the respondents speak into the Nintendo DS to reply to the hearing-impaired pupils,
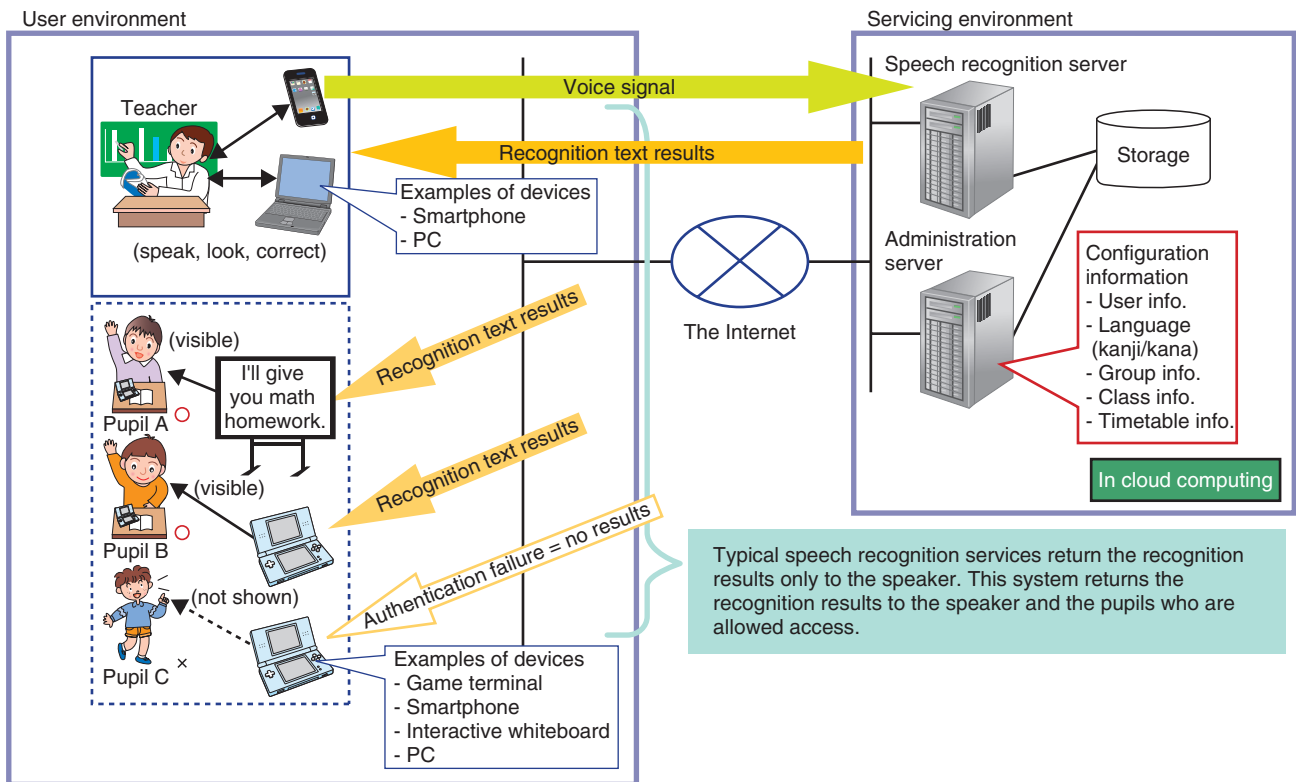

Fig. 4.   System configuration.

the recognized text appears on the display of the Nintendo DS.

3. Demonstration experiment

3.1 Overview

We conducted trials to evaluate the effectiveness of Koemiru. These trials were held from the end of January to early March 2012 in a fifth-grade class at Tottori Prefectural School for the Deaf, Himawari campus, and in a first-grade class at Okinawa Prefectural School for the Deaf. We prepared broadband connections by establishing Wi-Fi links from each terminal to a Wi-Fi router in the classroom, which was linked via optical fiber to the Internet. We optimized the acoustic model of the speech recognition server by using speech samples recorded by the actual teacher in the classroom. We also added the
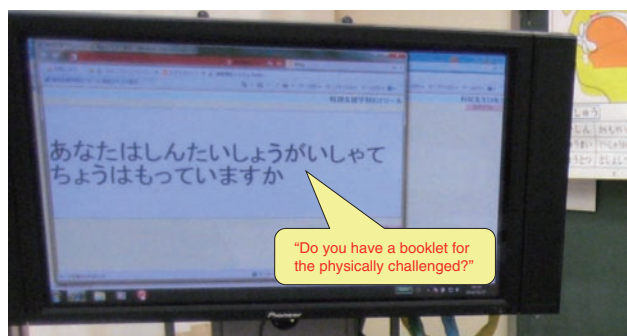

Fig. 5. Recognized speech displayed on interactive whiteboard.

words in the textbooks used in the classroom to the speech recognition dictionary, so the server could accurately recognize the speech of the teacher. The teachers requested that all recognition results be output in hiragana because they wanted to teach the pupils how to read kanji (**Fig. 5**). We also conducted free-form interviews with the teachers using a prepared questionnaire to determine service acceptability.

3.2 Results

The results of the trials (**Fig. 6**) are summarized as follows.
(1) This system is effective in teaching new words to pupils, since physical objects can be imagined from the words.
(2) This system stores all the speech recognition results of the teacher. Thus, teachers and pupils can review what was said at the end of the class.
(3) All devices are very user-friendly for both teachers and pupils. In particular, the Nintendo DS, one of the most popular portable game terminals, was well received by the pupils because no special instruction is needed to use it. Moreover, the pupils do not need to hold it when using it; they can simply put it on their desks.
(4) This system provided a very easy-to-use way for the teacher to edit the recognized text. The teachers used this function often, even though recognition accuracy was extremely high, to improve the content of their messages.

The results also highlighted the following challenges.

Good points

This system is effective in teaching new words to pupils: objects imagined from the words.

Teachers and pupils can review what was spoken at the end of the class.

All devices for teachers and pupils are very user-friendly.

The teachers liked the correction function and wanted to use it even when recognition accuracy was high.

Challenges

Response time should be shorter for short speech.

Teachers would like to be able to modify the recognition result and voice input operation on the interactive whiteboard rather than going to a PC.

Teachers should be able to add unrecognized words to the dictionary.

It is impractical for teachers to use smartphones. They must have their hands free to use sign language for pupils, so they cannot hold a smartphone at the same time.

Fig. 6. Experimental results.

- The response time should be reduced to improve the service acceptability, particularly with short sentences. The teachers found that with short sentences, the current version did not start to return the text until the teacher had finished speaking the sentence, and they felt the waiting time was annoying.
- Output correction required that the teacher access his or her PC. However, because the teacher is often in front of the classroom's blackboard or whiteboard, it is desirable to enable output correction at the interactive whiteboard.
- It is preferable for teachers to be able to add words to the dictionary by themselves.
- The teachers had difficulty in using the smartphone as the interface. Because they must be able to use sign language for the pupils, they cannot simultaneously hold the phones.

4. Future plans

We developed the Koemiru system based on speech recognition technology to support conversations with and among pupils with hearing disabilities. The results of trials indicated that Koemiru is effective in providing support to pupils in special needs schools. We plan to work on eliminating the aforementioned problems so that we can introduce this system into the market at the earliest possible date.

References

[1] Y. Noda, Y. Yamaguchi, K. Ohtsuki, and A. Imamura, "Development of the VoiceRex Speech Recognition Engine," NTT Technical Journal, Vol. 11, No. 12, pp. 14–17, 1999 (in Japanese).

[2] H. Masataki, D. Shibata, Y. Nakazawa, S. Kobashikawa, A. Ogawa, and K. Ohtsuki, "VoiceRex—Spontaneous Speech Recognition Technology for Contact-center Conversations," NTT Technical Review, Vol. 5, No. 1, pp. 22–27, 2007.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2007 01022.pdf

[3] S. Watanabe, T. Iwata, T. Hori, A. Sako, and Y. Ariki, "Topic Tracking Language Model for Speech Recognition," Computer Speech & Language, Vol. 25, No. 2, pp. 440–461, 2011.

[4] H. Shinohara, "R&D to Create the Future of ICT," NTT Technical Review, Vol. 10, No. 4, 2012.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2012 04fa2.html

Yuichi Muto
Senior Research Engineer, Promotion Project 2, NTT Service Evolution Laboratories.

He received the B.E. and M.S. degrees in information sciences from Tohoku University, Miyagi, in 1998 and 2000, respectively. He joined NTT Service Integration Laboratories as an engineer in 2000 and engaged in development activities in the Communication Traffic & Service Quality Project. From 2004 to 2009, he worked at the R&D center of NTT WEST. He moved back to NTT Service Integration Laboratories in 2009. His research interests include vehicle telematics services. As a result of organizational changes at the end of May 2012, he is now in NTT Service Evolution Laboratories and is involved in promoting and producing NTT R&D products.

Kenta Tetsuzaki
Research Engineer, Promotion Project 2, NTT Service Evolution Laboratories.

He received the B.S. degree in earth science from Hokkaido University in 2006 and the M.S. degree in earth planetary science from the University of Tokyo in 2008. He joined Corporate Marketing Headquarters, NTT WEST, as a system engineer in 2008 and was involved in the construction of large enterprise network systems. He moved to NTT Cyber Solutions Laboratories in 2010, where he developed the 3D-VOD server system for IPTV. As a result of organizational changes in May 2012, he is now in NTT Service Evolution Laboratories and is developing new services with R&D products.

Takako Sato
Research Engineer, Promotion Project 2, NTT Service Evolution Laboratories.

She received the B.S. degree in mathematics from Hirosaki University, Aomori, in 1994. She joined NTT Multimedia Systems Department in 1994 and developed video transmission systems. She was also involved in the development of IP-based broadcasting systems. As a result of organizational changes in May 2012, she is now in NTT Service Evolution Laboratories and is developing new services based on R&D products..

Masayuki Sugizaki
Senior Research Engineer, Supervisor, Promotion Project 2, NTT Service Evolution Laboratories.

He received the B.S. and M.S. degrees in information science from Tokyo University of Science in 1993 and 1995, respectively. He joined NTT Human Interface Laboratories in 1995 and studied web search technologies, text mining, and search log analysis. From 2004 to 2010, he was with NTT Resonant Inc., where he worked as a developer creating various web-based search services such as blog, news, and shopping searches for goo (http://www.goo.ne.jp/). He moved to NTT Service Integration Laboratories in 2010. As a result of organizational changes in May 2012, he is now in NTT Service Evolution Laboratories and is developing new services based on R&D products.

Yumiko Matsuura
Senior Research Engineer, Supervisor, Promotion Project 2, NTT Service Evolution Laboratories.

She received the B.S. and M.S. degrees in computer science from Keio University, Kanagawa, in 1991 and 1993, respectively. She joined NTT Human Interface Laboratories as an engineer in 1993 and engaged in the development of multimedia systems and a content delivery platform. From 2004 to 2006, she was in the R&D Strategy Department and was involved in developing and managing a vision for R&D. She moved to NTT Cyber Solutions Laboratories in 2006. Her research interests include search engine technologies. As a result of organizational changes in May 2012, she is now in NTT Service Evolution Laboratories and is involved in promoting and producing NTT R&D products. Her other activities include J-Win, a nonprofit organization for promoting and accelerating diversity in the workplace. She is also active as an organizer of the alumni network.