Input Support System for Medical Records Created Using a Voice Memo Recorded by a Mobile Device

K. Kurumizawa and H. Nishizaki
Department of Education, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Kofu-shi, Yamanashi 400-8511, Japan
Email: {kurumizawa, nisizaki}@alps-lab.org

K. Nishizaki and H. Ikegami
Kugawa Hospital for Orthopaedic Surgery, 10-26 Shinden-cho, Kofu, Yamanashi 400-0066, Japan
Email: ari.kanae@snow.plala.or.jp

Abstract—**This paper describes a development of an input support system for medical records using voice memos recorded by a mobile device. The goal of this research is to increase the efficiency of creating medical records. The proposed prototype system enables medical staff to record treatment memos using a mobile device soon after completing patient treatment and to edit a voice memo transcribed by an automatic speech recognition (ASR) system. This system also has ASR error correction technology and a user friendly interface for editing. We evaluated the prototype system by assessing the voice quality recorded by a mobile device using an ASR performance measure, measuring the time required to perform editing, and administering a questionnaire to provide subjective assessments. The experimental results showed that the proposed system was useful and that it may result in more efficient creation of medical records.**

Index Terms—**automatic speech recognition, error correction interface, input support system for medical records, mobile device, voice memo**

I. Introduction

Numerous electronic medical record systems and ordering systems have been developed in recent years. Most hospitals in Japan have begun preparing for the installation of these electronic systems. Computerizing patient data in hospitals has many advantages. For example, medical care becomes more efficient as the use of paper decreases. In addition, computerized patient data can be shared by multiple hospitals, which increases convenience for both hospitals and patients.

However, the Japanese Ministry of Health, Labor and Welfare has said that as of 2012, only 30.3% of all hospitals in Japan have introduced an ordering system. Electronic medical records have been introduced in just 18.7% of Japanese hospitals. Most small hospitals do not utilize information technology (IT) systems for management of patient data.

IT systems are not widely used in for three reasons: cost, lack of user friendliness, and constraint of medical treatment. Hospital medical records systems are very expensive. For example, the initial cost of an electronic medical system for a relatively small hospital (with approximately 100 beds) would be more than one million U.S. dollars. Furthermore, the additional cost that the medical staffs in a hospital learn how to use the system is also required. Therefore, it is difficult for such hospitals to introduce and use the system because of the cost-effectiveness.

In addition, user interfaces for such systems are complex; hospital staffs who are unfamiliar with computers find it difficult to use. Complexity and unfamiliarity cause critical operational errors [1]. Therefore, it is necessary to develop a user friendly and low-cost hospital IT management system.

Finally, for some treatments, particularly orthopedics (rehabilitation), patients are not treated in close proximity to a medical record system, usually located in a fixed position. Therefore, medical staff in such departments cannot input a medical record as they treat the patient.

The goal of this research is to, therefore, facilitate efficient creation of medical records. To achieve this, we have developed a prototype medical record input system for rehabilitation treatment staff, i.e., physical therapist (PT). This study introduces a prototype system that allows a PT to record treatment memos using a mobile device and easily edits the transcriptions of the voice memos using automatic speech recognition (ASR) system on a desktop or laptop personal computer (PC) after completing their daily medical treatments duties. We also have developed a user friendly interface for editing the transcriptions.

Using ASR technology to record medical information has advantages [2]. However, ASR errors must also be considered. Devine, *et al*. [3] claimed that using an ASR system to record medical information would be feasible if ASR technology was improved. Recently, a deep neural network-based acoustic modeling has been developed for

Manuscript received August 1, 2014; revised October 23, 2014.

doi: 10.12720/ijsps.3.2.128-133

an ASR system [4]. This model can significantly improve ASR performance. Therefore, ASR technologies are expected to be increasingly useful for voice recording of medical records. Existing medical record systems, such as AmiVoiceR Ex Clinic/Hospital [5] developed by Advanced Media, Inc. and Amazing Charts [6] developed by Amazing Charts, LLC, use ASR technology. Klann, *et al*. [7] attempted to develop a medical record system that can automatically generate a medical certificate on the basis of medical words or phrases extracted from the ASR transcription of conversational speech between a doctor and a patient. However, they used an ASR system located at a fixed position and did not consider ASR errors or user friendliness. Dragon Medical 360 Mobile Recorder [8] developed by Nuance Communications, Inc. provides a voice recording application that works on mobile devices. Keskinen, *et al*. [9] also introduced a mobile dictation application for healthcare purposes using ASR. Both these devices are similar to the proposed system; however, neither takes ASR errors into consideration.

On the other hand, our system considers portability, user friendliness, and ASR errors. The system is optimized for medical staffs like PTs who treat patients in a variety of places.

Our input support system has three characteristics. The first characteristic is that a user can record a treatment memo by voice using a mobile device soon after finishing treating a patient. The recorded memo is transferred to an ASR system and file server and automatically transcribed by two types of ASR systems.

The second characteristic is the use of an automatic ASR error correction framework. We have developed the ASR error correction framework [10] using outputs from two ASR systems. The third characteristic is that the support system has a user friendly interface when creating a written treatment record from the transcription of the recorded voice memo with ASR error correction.

We built the prototype system and evaluated it from the following aspects: the voice quality recorded by a mobile device (we used a third generation iPod touch developed by Apple, Inc.), duration required to create a written accurate record from a voice memo, and system usability (subjective assessment). The experimental results indicated that use of a mobile device was not significant; there was negligible reduction of ASR performance due to noise. The ASR error correction and edit interface resulted in faster creation of a written record depending on the subject (the person creating the record). Finally, analysis of a questionnaire administered to assess the system subjectively also indicated that the support system was effective for creating medical treatment records.

II. System Overview

Fig. 1 shows the workflow for the input support system in the rehabilitation department of a hospital. The main features of the workflow are to use a mobile device for recording a voice memo related to the treatment and for editing the transcription of the voice memo.

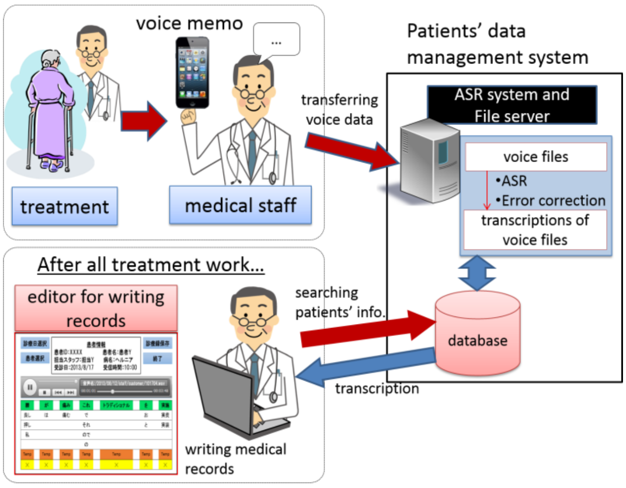


Figure 1. Workflow: Medical treatment, recording a memo, writing a record.

In the rehabilitation department, PTs do not perform medical treatments sitting at a desk equipped with a medical record system. In addition, PTs treat patients continuously without taking a break. Therefore, they do not have time to input medical records to an electronic medical record system. They write medical records for the patients they treat during the day in their spare time or after completing their treatment duties.

To reduce medical records writing burden, we focused on a mobile device. In this study, we used an iPod touch, which is a very small device equipment with a microphone. It is light enough to be carried while performing medical treatment duties. PTs can easily record a voice memo soon after finishing a patient's treatment.

A recorded voice memo is transferred to a file server that stores voice memos and an ASR system that converts speech signals to text. The transcribed voice memos are associated with the voice files and can be saved in the server. However, ASR technologies are not perfect. It is impossible to prevent ASR errors, even in state-of-the-art ASR systems. But medical records should be very accurate. Consequently, a PT has to correct the ASR errors, which can be cumbersome for a PT who is not skilled at operating a computer (in particular keyboard operations) and makes the task of creating medical records less efficient.

Therefore, we have developed an ASR framework with an automatic error correction function and a treatment record editor with a user friendly interface. The ASR error correction framework can reduce the number of ASR errors. A system user can create a medical record easily without spending too much time performing mouse or touch screen operations. A user friendly environment that includes an intuitive interface and reduces the number of keyboard operations is advantageous for a person who is unfamiliar with computer operations.

A. *Voice Memo Application*

The graphical user interface for the voice memo application on an iPod touch (iOS) is shown in Fig. 2. The screen displays entry fields for medical staffs (PTs) and patient names, recording status, and several function buttons. After completing a medical treatment, a PT pushes the "recording button" and records a treatment memo, which generally takes only several tens of seconds. After recording and checking the content (if necessary), the voice data is transferred to the ASR system and file server by pushing the "voice transfer button."

Figure 2. Voice recording interface (iOS application).

In addition to voice recording, the iOS application has a treatment booking management function. PTs can check their treatment schedule and also book a new treatment or revise a patient's treatment schedule.

B. ASR Error Correction

We previously developed an ASR error correction method based on a confusion network combination (CNC) [10]. The goal of the CNC-based error correction method is to reduce the time required to correct ASR errors.

Correction time can be reduced by displaying multiple word candidates on the edit screen. ASR errors can be corrected by a mouse or touch operation and do not require keyboard operations. However, too many word candidates may confuse a user. Therefore, suitable candidate words must be selected.

The CNC-based method uses two types of ASR systems. Using multiple ASR systems, such as Recognizer Output Voting Error Reduction (ROVER) [11] and CNC [12], has been reported to improve speech recognition performance. The two ASR systems produce two types of transcriptions, each of which uses a different word set. Our method combines the confusion network output from the two ASR systems depending on posterior probabilities attached to each word.

The CNC method is based on a combination of confusion networks derived by the two ASR systems. In the support system, we commonly used Julius [13], an open source large vocabulary continuous speech recognition engine, as a decoder in the two ASR systems. A language model is also commonly used in the systems.

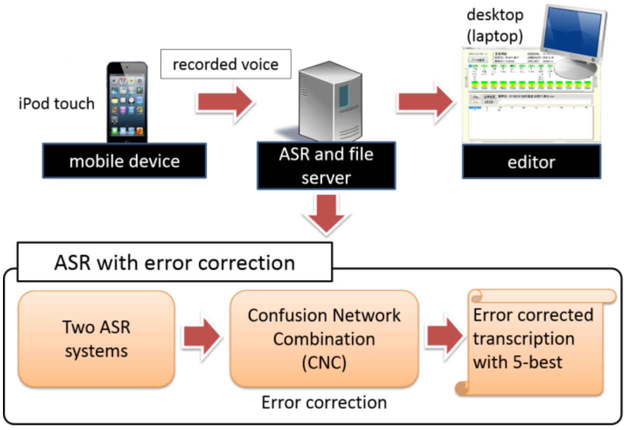


Figure 3. Outline of an error correction framework.

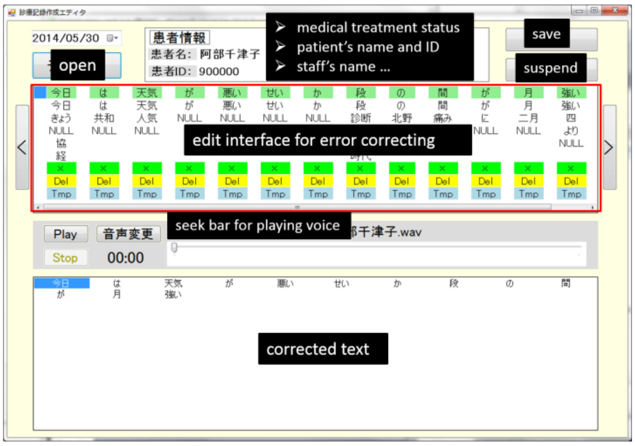


Figure 4. Screen shot of the editing interface.

ASR systems differ according to the type of acoustic models used. We prepared two types of acoustic models: a triphone-based hidden Markov model (HMM) and a syllable-based HMM.

The Julius decoder can recognize an input utterance and can output a confusion network-formed transcription. This method comprises the following steps.

C. Editing Interface for Writing Medical Records

Fig. 3 shows an outline of the error correction framework using the voice memo transcription edit interface with the CNC-based method to reduce ASR errors. First, a voice memo is recognized by the two ASR systems. Next, the CNC-based method combines the confusion networks derived from the two ASR systems. The transcription created by the method is displayed on the editing system's graphical user interface. A user can find ASR errors and correct them by replacing incorrect words with the correct ones.

A screenshot of our editing interface is shown in Fig. 4. This interface is composed of two parts: edit interface for error correcting of transcription by the ASR systems and the representation of the editing result. Users can listen to a voice memo using a seek bar while editing a transcription. Users can also cue the recording using a transcribed word, i.e., users can start playback from a particular point.

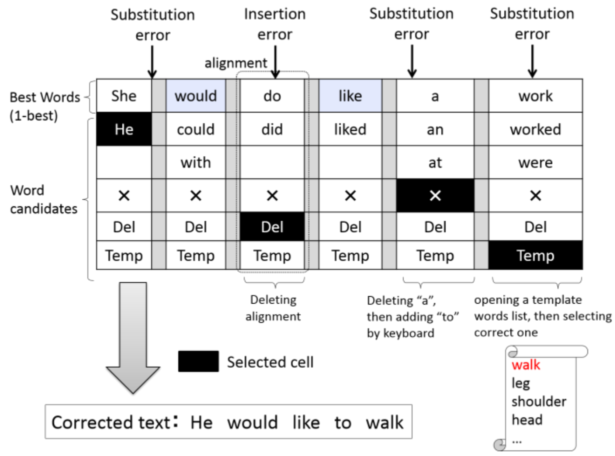


Figure 5. Scaled-Up illustration of the error correction interface.

Fig. 5 shows a scaled-up illustration of the error correction interface. The word sequence on the row labeled "Best Words" is the best recognition hypothesis (1-best) by the CNC-based method. The "Word candidates" rows present candidates for error correction; each candidate is aligned to each 1-best word. The interface can display a maximum of 5-best candidates for each alignment because too many candidates may degrade correction efficiency. Each alignment has "X," "Del," and "Temp," buttons. If a user touches (or clicks) on a candidate cell, the 1-best word corresponding to the touched (or clicked) cell is replaced by the word in the 1-best word cell. A 1-best word can be deleted from the "Best Words" line by touching (or clicking) the "X" button. Users can insert a new word to replace the deleted word. These works are done when a user corrects substitution errors. If candidate insertion errors occur, users can delete the alignment (all candidates and buttons) by pushing the "Del" button.

By repeating these actions, users can obtain an error-free voice memo transcription. However, if the correct word is not in any of the cells, users have to use a keyboard to input the correct word. Users can activate a word input form by selecting the blank space between alignments at the position where the work should be inserted. The word is added when input is complete. Keyboard operations increase work time. Therefore, to reduce time required to correct a transcription, it is important to display as many candidates as possible.

The "Temp" button is used to open a template words list, which includes words frequently used in medical records. Medical staff can register these words prior to undertaking a transcription/correction exercise.

III. Evaluation

We performed three types of evaluation: ASR performance of voice memos recorded by two types of microphones, the automatic error correction performance, and a subjective evaluation of editing a transcribed voice memo.

We evaluated the voice quality by comparing the microphone in a mobile device with a handheld microphone. We also evaluated the improvement in ASR performance achieved by the CNC-based method. In addition, we evaluated the time required for editing a voice memo with and without the proposed editing system. To determine the time required to edit without the editing system, subjects edited a voice memo using standard Windows OS text editor software and also manually created a record using pen and paper.

A. Experimental Setup

Seven subjects who were familiar with computer-based keyboard operations participated in the experiment. In a subjective experiment, we compared the voice quality recorded by a mobile device (third generation iPod touch) and that recorded by a handheld microphone directly connected to the ASR system and file server using an ASR performance measure.

The experimental procedure was as follows:

1) A subject records a voice memo using a mobile device and handheld microphone. (This procedure was used to investigate the extent to which the microphone in a mobile device degrades voice quality.)
2) The CNC-based error correction method is performed when the system receives the two types of confusion networks.
3) The transcription with correction candidates is displayed on the editing system interface.
4) The subject corrects the errors in the transcription with and without the editing system.

To assess ASR performance, each subject spoke three sentences related to medical care and one additional sentence, with a duration of approximately 20s, and hand wrote the final copy of the transcription produced by the ASR system with the CNC-based error correction. We measured the time required to complete the manual transcription of this additional sentence.

Both types of acoustic models come from the Japan Newspaper Article Speech corpus, a word trigram-based language model with vocabulary of 20,000 words, provided by the Acoustic Society of Japan [14].

B. Experimental Results and Discussion

Table I shows the ASR performance of 1-best and 5-best ASR suggested candidates for voice memos recorded by two types of devices and shows a comparison of recordings with and without CNC-based error correction.

TABLE I. ASR Performances (Word-Based Accuracy [%]) of Voice Memos Recorded by Two Types of Devices and the Effectiveness of the CNC-Based ASR Error Correction.

Recording device	N-best	w/o correction	w/ correction
iPod touch	1-best	66.6	71.5
	5-best	69.8	78.1
Handheld mic.	1-best	74.9	77.1
	5-best	77.6	81.9

The ASR performance without error correction is from the transcription by the single ASR system with the triphone-based HMM as acoustic model. The values are word accuracy rates averaged among 21 voices (seven subjects x three sentences). Evidently, the correction method is very useful for improving the ASR performance on both the 1-best and 5-best candidates. In particular, the correction drastically improved the performance of the 5-best candidates for the memo recorded by the iPod touch.

There are significant differences between the ASR performance of the 1-best candidates for iPod touch and handheld microphone recordings. Although the ASR performance for the 5-best candidates with iPod touch recordings is not as good as that with handheld microphone recordings, the different is not significant. The iPod touch microphone degrades voice quality slightly due to noise. We did not take denoising technologies into account.

TABLE II. RECORD CREATION EXECUTION TIMES ([M: SS]) AND ASR PERFORMANCE ([%]).

Subject	System	Keyboard	Handwriting	ASR perf.
#1	3:39	2:21	3:03	78.0
#2	1:43	1:40	2:23	77.1
.#3	1:47	1:37	3:33	73.5
#4	2:41	1:34	1:50	81.5
#5	1:19	2:43	2:25	96.1
#6	2:28	1:56	2:05	73.5
#7	1:23	1:57	2:07	85.7
Ave.	2:09	1:58	2:30	80.8

Table II shows the execution time for manual (keyboard and handwriting) creation of a voice memo record and ASR performance for each voice memo. "System" indicates the time using the editing system, "Keyboard" indicates the time using a standard Windows OS text-editor, and "Handwriting" indicates the time required to manually create a record. As shown in Table II, five subjects created a record most quickly using a keyboard. However, for Subjects #2 and #3, there are no significant executive time differences between using the system and the keyboard. The fastest execution time was recorded by Subject #5 using the editing system. And Subject #7 also finished writing a record quickly. The creation time is not much different from Subject #5. This is probably because ASR performance for Subject #5 and #7 was very high (96.1% and 85.7%, respectively). Therefore, improving ASR performance is certain to reduce the time required for performing editorial work.

TABLE III. QUESTIONNAIRE RESULTS (AVERAGE FOR ALL SUBJECTS).

Questionnaire items	Rating
Was the transcription work easy?	4.1
Usability of the system	3.9
Which is better the system or the keyboard?	4.2

We administered a questionnaire to investigate the usability of the editing system. The questionnaire was administered to the subject who participated in the experiments described in previous sections. Questions and results are presented in Table III. The subjects used a 5-point scale (1, 2, 3, 4, 5) for each questionnaire item. For the first and second questions, "5" was the highest rank and "1" was the lowest rank. The final item in the questionnaire was a competitive question comparing simplicity of use between the proposed editing system and keyboard input. "5" means that a subject liked the system. This result indicates that our system's usability and ease of use are very high. Furthermore, we asked all subjects an additional question: "which is the best editing way to write a record?" All subjects answered that the editing system was the best.

From the evaluations of voice quality and ASR performance, as well as subjective assessment, it is evident that the editing system's user interface was good. However, ASR performance is very important to reduce editing time. Therefore, we should take effective measures to address ASR performance improvement.

IV. CONCLUSION

This paper introduced an input support system for creating medical records using a voice memo recorded by a mobile device. A mobile device is useful for creating a voice memo immediately after completing a medical treatment. We developed a voice recording application that runs on an iOS device and a user interface to correct errors in ASR transcriptions.

We investigated the quality of voices recorded by the mobile device and evaluated the support system in a subjective experiment. The voice quality recorded by the iPod touch was slightly lower than that recorded using a handheld microphone on the hypotheses of up to 5-best ASR candidates in which some errors were corrected. The editing interface received good evaluations from all subjects in the subjective experiment. However, improving ASR performance would reduce workload.

In future, we intend to develop a language model training tool and create an ASR dictionary to improve ASR performance, specifically for voice recordings containing medical terms. In addition, we also wish to improve the record editing user interface.

ACKNOWLEDGMENT

This study was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) Grant Number 26282049.

REFERENCES

[1] R. Koppel, *et al.*, "Role of computerized physician order entry systems in facilitating medication errors," *The Journal of the American Medical Association*, vol. 293, no. 10, pp. 1197-1203, 2005.

[2] A. Zafar, J. M. Overhage, and C. J. McDonald, "Continuous speech recognition for clinicians," *Journal of the American Medical Informatics Association*, vol. 6, no. 3, pp. 195-204, 1999.

[3] G. Evermann and P. Woodland, "Comparative evaluation of three continuous speech recognition software packages in the generation of medical reports," *Journal of the American Medical Informatics Association*, vol. 7, no. 5, pp. 462-468, 2000.

[4] G. Hinton, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, 2012.

[5] Advanced Media, Inc. AmiVoiceR Ex clinic/hospital. [Online]. Available: http://www.advanced-media.co.jp/english/business/medical.html

[6] Amazing Charts, LLC. Amazing charts. [Online]. Available: http://amazingcharts.com/

[7] J. G. Klann and P. Szolovits, "An intelligent listening framework for capturing encounter notes from a doctor-patient dialog," *BMC Medical Information and Decision Making 2009*, vol. 9, pp. 1-10, 2009.

[8] Nuance Communications, Inc. Dragon medical 360 mobile recorder. [Online]. Available: http://www.nuance.com/products/dragon-medical-360-mobile-recorder/index.htm.

[9] A. A. Nanavati, N. Rajput, *et al.*, "Mobile dictation with automatic speech recognition for healthcare purposes," in *Proc. 8th Workshop on Speech in Mobile and Pervasive Environments (SiMPE2013)*, 2013, pp. 1-5.

[10] S. Ishimaru, H. Nishizaki, and Y. Sekiguchi, "Effect of confusion network combination on speech recognition system for editing," in *Proc. APSIPA ASC 2011*, 2011, pp. 1-4.

[11] J. G. Fiscus, "A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (ROVER)," in *Proc. IEEE ASRU Workshop*, 1997, pp. 347-352.

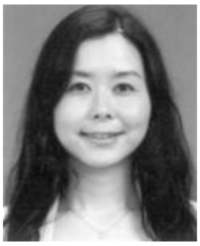
[12] G. Evermann and P. Woodland, "Posterior probability decoding, confidence estimation and system combination," in *Proc. NIST Speech Transcription Workshop*, 2000.

[13] A. Lee, T. Kawahara, and K. Shikano, "Julius - An open source real-time large vocabulary recognition engine," in *Proc. EUROSPEECH 2001*, 2001, pp. 1691-1694.

[14] K. Itou, *et al.*, "The design of the newspaper-based Japanese large vocabulary continuous speech recognition corpus," in *Proc. 5th ICSLP*, 1998, pp. 3261-3264.

Keisuke Kurumizawa was born in 1991. He received his B.E. degree in computer and media sciences from University of Yamanashi in 2013. He is now a master course student at Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi. His research interests include spoken language processing and speech interface. He is a student member of the Acoustical Society of Japan.

Hiromitsu Nishizaki was born in 1975. He received his B.E., M.E., and D.Eng. degrees in information and computer sciences from Toyohashi University of Technology in 1998, 2000, and 2003. He is now an assistant professor in Department of Research, Interdisciplinary Graduate School of Medicine and Engineering at University of Yamanashi. His research interests include spoken language processing, speech interface, and human-computer interaction. He is a member of IEEE, the Institute of Electronics, Information and Communication Engineers (IEICE), Information Processing Society of Japan (IPSJ), and the Acoustical Society of Japan.



Kanae Nishizaki was born in 1975. She graduated from the Faculty of Rehabilitation of Fujita Health University in 1998. She also got her physical therapist certificate in 1998. She worked as a physical therapist at Kamiiida Daiichi General Hospital in Nagoya, Japan from 1998 to 2004. She is now a general manager of the Department of Rehabilitation, Kugawa Hospital for Orthopaedic Surgery. She received her Ph.D of engineering from University of Yamanashi in 2014.

Hitoshi Ikegami was born in 1963. He graduated from the School of Medicine of Tokyo Medical University. He got his medical license in 1989 and he is also a medical doctor. He worked as a medical doctor at the orthopedic surgery of Tokyo Medical University Hospital, then he was an assistant professor at the department of orthopedic surgery of Tokyo Medical University in 1995. He became a director of orthopedic surgery at Ohuna Chuo Hospital in 1997. He is now the director of Kugawa Hospital for Orthopedic Surgery since 2007. He received the Hippocrates award of Tokyo Medical University in 1994, the Taisho award in 1996, and the AAOS New Program Travelling Fellow in 1998.