Internal Medicine
The Japanese Society of Internal Medicine

doi: 10.2169/internalmedicine.9592-22
Intern Med 62: 153-157, 2023
http://internmed.jp

【 ORIGINAL ARTICLE 】

Speech Recognition System Generates Highly Accurate Endoscopic Reports in Clinical Practice

Hiroshi Takayama[1], Toshitatsu Takao[1], Ryo Masumura[2], Yoshikazu Yamaguchi[3], Ryo Yonezawa[4], Hiroya Sakaguchi[1], Yoshinori Morita[5], Takashi Toyonaga[1], Kazutaka Izumiyama[6] and Yuzo Kodama[1]

Abstract:

Objective Endoscopic reports are conventionally written at the end of each procedure, and the endoscopist must complete the report from memory. To make endoscopic reporting more efficient, we developed a new speech recognition (SR) system that generates highly accurate endoscopic reports based on structured data entry. We conducted a pilot study to examine the performance of this SR system in an actual endoscopy setting with various types of background noise.

Methods In this prospective observational pilot study, participants who underwent upper endoscopy with our SR system were included. The primary outcome was the correct recognition rate of the system. We compared the findings generated by the SR system with the findings in the handwritten report prepared by the endoscopist. The initial correct recognition rate, number of revisions, finding registration time, and endoscopy time were also analyzed.

Results Upper endoscopy was performed in 34 patients, generating 128 findings of 22 disease names. The correct recognition rate was 100%, and the median number of revisions was 0. The median finding registration time was 2.57 [interquartile range (IQR), 2.33-2.92] seconds, and the median endoscopy time was 234 (IQR, 194-227) seconds.

Conclusion The SR system demonstrated high recognition accuracy in the clinical setting. The finding registration time was extremely short.

Key words: endoscopy report, speech recognition, recognition rate, upper endoscopy, voice recognition

(Intern Med 62: 153-157, 2023)
(DOI: 10.2169/internalmedicine.9592-22)

Introduction

In recent years, advances in speech recognition (SR) technology have been remarkable, with voice commands being used for many routine activities related to smartphones, the Internet and even housework. Although not common in the medical field, SR has been reported in radiology, emergency medicine, pathology, nursing, endocrinology, dentistry, and psychiatry. The recognition rate has also been reported, with the highest rate being 96% in radiology (1-4). In contrast, only a few studies have described the use of SR in endoscopic reports, and none have accurately evaluated the recognition rate (5-8).

Gastrointestinal endoscopy findings are conventionally handwritten or typewritten at the conclusion of the endoscopic procedure. As such, the endoscopist is required to memorize the findings during the procedure, as the report can only be prepared after endoscopy has concluded. However, this practice poses problems, as the endoscopist may

[1]Division of Gastroenterology, Department of Internal Medicine, Kobe University Graduate School of Medicine, Japan, [2]NTT Computer and Data Science Laboratories, NTT Corporation, Japan, [3]NTT Human Informatics Laboratories, NTT Corporation, Japan, [4]System Solution Headquarters, Fujifilm Medical IT Solutions Co., Ltd, Japan, [5]Department of Gastroenterology, Kobe University Hospital International Clinical Cancer Research Center, Japan and [6]Hyogo Prefecture Health Promotion Association, Japan

Received: February 14, 2022; Accepted: May 6, 2022; Advance Publication by J-STAGE: June 21, 2022

Correspondence to Dr. Toshitatsu Takao, ttakao@med.kobe-u.ac.jp

miss documenting important details of the lesions. Furthermore, there is an increased risk of bacterial infection associated with the doctor touching the keyboard, mouse, and finding sheets (9). Completing handwritten or typewritten reports is also time consuming, which limits the number of procedures that can be performed daily.

To address these issues, we developed a new SR system for endoscopic procedures that was based on structured data entry (10). Our SR system selectively extracts endoscopic findings from spoken conversation and automatically inputs these into the appropriate columns of a structured report. We tested this system in a preliminary study with an endoscopic simulator, which demonstrated that our SR system is 98.4% accurate in recognizing the relevant terminology. It also significantly shortens the overall examination time (10). However, our preliminary study with an endoscopic simulator was conducted in a quiet environment, whereas clinical endoscopy is performed in a space with ambient noise, such as the voices of the patient or nurse, suction sound of the endoscope, and noise from the endoscope washer.

We conducted a pilot study to determine the performance of our SR system in terms of maintaining the same level of accuracy in clinical endoscopic settings with various degrees of ambient noise.

Materials and Methods

Study design

This study was a prospective observational pilot study conducted in collaboration with Kobe University Hospital and the Hyogo Prefectural Health Promotion Association.

The study protocol was approved by the Institutional Review Board of Kobe University Hospital (no. B200249, approved on November 16, 2020) and performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. We obtained written informed consent from all of the participants.

Study population

Patients who underwent upper endoscopy as part of a routine health checkup at the Hyogo Prefectural Health Foundation Health Examination Center were included in this study. Patients who had strong anxiety during a previous endoscopic examination were excluded.

SR system

We jointly developed the original Japanese SR system for reporting endoscopic findings equipped with VoiceRex® (Nippon Telegraph and Telephone Corporation, Tokyo, Japan) in collaboration with Fujifilm Medical IT Solutions, NTT TechnoCross Corporation, and Nippon Telegraph and Telephone Corporation. The original SR system was used in a preliminary study with an endoscopic simulator (10). In the present study, we upgraded the VoiceRex® system by introducing the latest technology (11, 12). We also upgraded the text-processing method and vocabulary database of the system. The SR system automatically extracts endoscopic findings from spoken conversation and inputs these data into the appropriate columns in structured reports. We opted for structured data entry because free text entry would likely be incomplete and difficult to search (13-19).

The SR system processed keywords through a hierarchical structure that comprised organs, disease names, and remarks. When an endoscopic finding was identified, the system selected the appropriate terminology from the built-in dictionary linked with the Japan Endoscopy Database and inserted it into the appropriate column. When a lower-level finding in the column tree was recognized, the corresponding terminology was automatically entered into the upper column. For example, when “Grade M” was identified from spoken conversation, the system entered “esophagus,” “reflux esophagitis,” and “Grade M” in the organ, disease name, and remarks columns, respectively. Another feature of this system was the ability to recognize multiple expressions for a single finding. There were 196 types of findings and 446 ways of expressing them. For example, when we registered the findings of gastric cancer, we were able to register it by saying *igan* in Japanese or using an abbreviation for gastric cancer, such as ”MK” in German. Therefore, malignant findings were able to be registered by voice alone, even during endoscopy for conscious patients. As there were multiple verbal expressions for a finding, we prepared multiple patterns to ensure that the findings were registered with a high recognition rate.

The SR system is a software program that is installed onto a personal computer. A wireless microphone transmitted all spoken conversation to the computer. The command “Start examination” launched the report registration screen, and the program automatically added the relevant data to the report throughout the procedure. All words spoken by the endoscopist were displayed in the “All detected words” section on the bottom of the screen. The SR system only recognized the findings included in the built-in dictionary, and the recognized findings were first displayed in the “Pre-registration findings” section on the upper part of the screen. If wrong data were added at any time point, we were able to revise the data by rephrasing our findings in this stage. When the proper findings were displayed, the endoscopist used the command “Register” to finalize the findings in the center of the screen as “Registration findings.” The endoscopist was able to directly finalize the registration by uttering the word “Register” immediately after the findings had been uttered. The command “Complete examination” closed the report registration screen at the end of the procedure (Figure, Supplementary material 1).

Endoscopic procedures

A single endoscope (GIF-PQ260; Olympus Corporation, Tokyo, Japan) and endoscopic system (EVIS LUCERA ELITE; Olympus Corporation) were used to perform endoscopy. No sedation was used during endoscopy for any par-

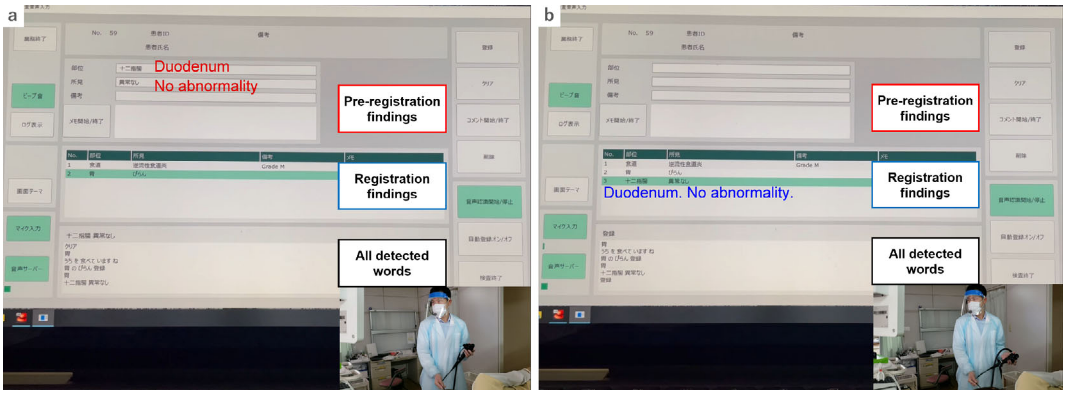

Figure. Speech recognition (SR) system workflow. All words spoken by the endoscopist were displayed in the "All detected words" section at the bottom of the screen. The recognized findings were first displayed in the "Pre-registration findings" section at the upper part of the screen (a). When the proper findings were displayed, the endoscopist used the command "Register" to finalize the findings in the center of the screen as "Registration findings" (b).

ticipants. All endoscopic procedures were performed by a single experienced endoscopist in a clinical setting that included ambient noise from the endoscopic machines and washers. Upper gastrointestinal endoscopy was performed with the personal computer with the SR system positioned next to the endoscopic image screen. The program was run throughout each procedure, and a handwritten report was also completed by the endoscopist at the conclusion of each endoscopy; this procedure was performed in a blinded manner. The reports from the SR system and endoscopist were compared, and the accuracy of the automated report was also analyzed. We also analyzed all audio files recorded using the SR system.

Outcomes

The primary outcome was the correct recognition rate of the SR system. Correct recognition was defined as the ratio of correct identification of a set of data, which included the organs, disease names, and remarks, corresponding to the findings in the handwritten report.

The secondary outcomes included the initial correct recognition rate, number of revisions, finding registration time, and endoscopy time. The initial correct recognition rate was defined as the number of rows with correct findings registered on the first instance divided by the number of rows finally registered in the system. The number of revisions was defined as the number of times rephrasing was required prior to registration of the findings. The finding registration time was defined as the time from when the endoscopic findings were first spoken to when a set of findings was registered in the system. The endoscopy time was defined as the time between when "start examination" and "complete examination" were spoken. The finding registration time and endoscopy time were retrieved from the information recorded in the SR system.

Sample size calculation

Our previous study with the SR software program and an upper gastrointestinal endoscopy simulator demonstrated a correct recognition rate of 98.4% (10). The role of SR software programs in other disciplines has been examined, and the highest correct recognition rate of 96% was documented in radiologic reports (1-4). We expected a lower rate in this study than our previous study due to ambient noise during the procedure. We hypothesized that the rate in this study would be 96%, which was the highest rate reported previously. We set the width of the 95% confidence interval at 4%, so that the correct recognition rate would be between 92% and 100%. The required sample size based on this setting value was found to be 93. We estimated that at least three findings could be obtained in each patient. As such, 31 patients were required to obtain 93 findings. We expected a 10% dropout rate, so the final required sample size was 34 participants.

Statistical analyses

The number of revisions, finding registration time, and endoscopy time were presented as medians with interquartile ranges (IQRs). The statistical program "R", version 4.0.4 (R Foundation, Vienna, Austria) was used for analysis. EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan) was utilized for all statistical analyses and sample size calculations (20).

Results

Upper endoscopy was performed in 34 participants. We evaluated 128 findings in 22 disease names. The disease names and number of findings are listed in Table 1. The number of findings correctly recognized was 128. All of these were registered correctly on the first instance without revisions. Thus, the correct recognition rate and initial correct recognition rate were both 100%. The number of revisions was 0.

The median finding registration time was 2.57 (IQR, 2.33-2.92) seconds, whereas the median endoscopy time was 234 (IQR, 194-227) seconds (Table 2).

Discussion

We demonstrated that our revised SR system was highly accurate and practical in the clinical setting, despite the presence of ambient noise. All findings were accurately registered at the first instance and did not require revision.

There are a few reports describing the use of SR in endoscopic reports, but none have accurately evaluated their recognition rate (5-8). The SR system we previously developed had a high recognition rate and contributed to shortening of the examination time in a quiet environment with an endoscopic simulator. However, we had not evaluated whether or not the SR system operated accurately during an actual endoscopic examination with ambient noise (10). In addition, our previous SR system involved using a wired microphone.

Table 1. **List of All the Endoscopic Findings Obtained in This Study.**

Organs	Disease names	Remarks	Number of findings
Esophagus	Barrett's esophagus	SSBE	14
Esophagus	No abnormality		10
Esophagus	Hiatal hernia		7
Esophagus	Reflux esophagitis	Grade M	5
Esophagus	Reflux esophagitis	Grade A	5
Esophagus	Papilloma		2
Esophagus	Candidiasis		1
Esophagus	Submucosal tumor		1
Stomach	Erosion		11
Stomach	Fundic gland polyp		8
Stomach	No abnormality		6
Stomach	Atrophic gastritis	O-II	5
Stomach	Atrophic gastritis	C-II	4
Stomach	Erosive gastritis		4
Stomach	Atrophic gastritis	C-III	3
Stomach	Atrophic gastritis	O-I	3
Stomach	Xanthoma		2
Stomach	Hyperplastic polyp		1
Stomach	Telangiectasia		1
Stomach	Ulcer		1
Duodenum	No abnormality		32
Duodenum	Ulcer scar		2

SSBE: short segment Barrett's esophagus

In this present study, the revised SR system with a wireless microphone demonstrated a superior correct recognition rate of 100% despite ambient noise, such as the voice of the patient or a nurse, the suction sound of the endoscope, and the noise from the endoscopy washer. In comparison, our previous study with the original SR system demonstrated a correct recognition rate of 98.4% (10). We expected a lower rate in this study due to ambient noise during the procedure. However, the result was 100%, which was the highest correct recognition rate we initially expected. The high correct recognition rate in this study may be due to the following reasons: we utilized structured data entry, which allowed our SR system to selectively extract endoscopic findings that matched those in the built-in dictionary, so none of the other noises were considered as findings; and we improved the accuracy of the SR system by using the latest deep learning technology, adjusting the word processing method, and enhancing the vocabulary choices in the software dictionary.

The median finding registration time with the revised system was 2.57 (IQR, 2.33-2.92) seconds. This time was extremely short and had little effect on the overall endoscopy time. Our system was able to register findings efficiently because of the high processing power of the software program. In particular, our system was able to input simultaneous findings (location, disease name, remarks) based on a single lower-level finding in the column tree. When a lower-level finding in the column tree was recognized, the corresponding findings were automatically entered into the other columns. The prompt registration of SR findings during endoscopy may shorten the overall examination time by eliminating the time required to create reports at the end of each procedure.

Several limitations associated with the present study warrant mention. First, it was a single-center study conducted by a single endoscopist in Japanese participants. The SR ability of our software may have been influenced by the age, sex, dialect, speaking volume, and speaking speed of the endoscopist. As such, our results may not be generalizable. Second, we only evaluated a limited number of findings in our study. Finally, we did not evaluate the usefulness of the SR system in terms of total time spent on the endo-

Table 2. **Correct Recognition Rate, Initial Correct Recognition Rate, Number of Revises, Finding Registration Time and Endoscopy Time for 128 Findings.**

Number of findings correctly recognized	128 (100%)
Number of findings correctly recognized on the first instance	128 (100%)
Number of revisions	0
Finding registration time (s)	2.57 (2.33-2.92)
Endoscopy time (s)	234 (194-227)

Data represent the number (%) or median (interquartile range) scopic procedure and report preparation, as in the study of Yokota et al. (8).

We intend to conduct a multi-center study in the future to prove that a high recognition rate can be achieved among various endoscopists and findings. Since we were able to achieve an extremely high correct recognition rate in an actual endoscopy setting, which was not evaluated previously, we believe that this system can improve endoscopic efficiency. Our aim is to increase the operational efficiency of the endoscopy workflow and decrease the stress of the endoscopist.

In conclusion, our SR system demonstrated a high correct recognition rate of 100% in the actual clinical setting. The finding registration time was extremely short.

The authors state that they have no Conflict of Interest (COI).

References

1. Blackley SV, Huynh J, Wang L, et al. Speech recognition for clinical documentation from 1990 to 2018: a systematic review. J Am Med Informatics Assoc **26**: 324-338, 2019.
2. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. J Am Med Informatics Assoc **23**: e169-e179, 2016.
3. Johnson M, Lapkin S, Long V, et al. A systematic review of speech recognition technology in health care. BMC Med Inform Decis Mak **14**: 94, 2014.
4. Pezzullo JA, Tung GA, Rogg JM, et al. Voice recognition dictation: radiologist as transcriptionist. J Digit Imaging **21**: 384-389, 2008.
5. Massey BT, Geenen JE, Hogan WJ. Evaluation of a voice recognition system for generation of therapeutic ERCP reports. Gastrointest Endosc **37**: 617-620, 1991.
6. Cass OW. Automated speech technology for gastrointestinal endoscopy reporting and image recording. Proc Annu Symp Comput Appl Med Care 968-969, 1991.
7. Molnar B, Gergely J, Toth G, et al. Development of a speech-based dialogue system for report dictation and machine control in the endoscopic laboratory. Endoscopy **32**: 58-61, 2000.
8. Yokota Y, Iwatsubo T, Takeuchi T, et al. Effects of a novel endoscopic reporting system with voice recognition on the endoscopic procedure time and report preparation time: propensity score matching analysis. J Gastroenterol **57**: 1-9, 2022.
9. Choi ES, Choi JH, Lee JM, et al. Is the environment of the endoscopy unit a reservoir of pathogens? Intest Res **12**: 306, 2014.
10. Takao T, Masumura R, Sakauchi S, et al. New report preparation system for endoscopic procedures using speech recognition technology. Endosc Int Open **06**: E676-E687, 2018.
11. Masumura R, Tanaka T, Ando A, et al. Role play dialogue aware language models based on conditional hierarchical recurrent encoder-decoder. Proc Annu Conf Int Speech Commun Assoc Interspeech **2018**: 1259-1263, 2018.
12. Tanaka T, Masumura R, Masataki H, et al. Neural error corrective language models for automatic speech recognition. Proc Annu Conf Int Speech Commun Assoc Interspeech **2018**: 401-405, 2018.
13. de Lange T, Moum BA, Tholfsen JK, et al. Standardization and quality of endoscopy text reports in ulcerative colitis. Endoscopy **35**: 835-840, 2003.
14. Aabakken L. Quality reporting - finally achievable? Endoscopy **46**: 188-189, 2014.
15. Hoff G, Ottestad PM, Skafløtten SR, et al. Quality assurance as an integrated part of the electronic medical record - a prototype applied for colonoscopy. Scand J Gastroenterol **44**: 1259-1265, 2009.
16. Bretthauer M, Aabakken L, Dekker E, et al. Reporting systems in gastrointestinal endoscopy: requirements and standards facilitating quality improvement: European Society of Gastrointestinal Endoscopy position statement. United Eur Gastroenterol J **4**: 172-176, 2016.
17. Manfredi MA, Chauhan SS, Enestvedt BK, et al.; ASGE Technology Committee. Endoscopic electronic medical record systems. Gastrointest Endosc **83**: 29-36, 2016.
18. Maserat E, Safdari R, Maserat E, et al. Endoscopic electronic record: a new approach for improving management of colorectal cancer prevention. World J Gastrointest Oncol **4**: 76-81, 2012.
19. Gouveia-Oliveira A, Raposo VD, Salgado NC, et al. Longitudinal comparative study on the influence of computers on reporting of clinical data. Endoscopy **23**: 334-337, 1991.
20. Kanda Y. Investigation of the freely available easy-to-use software ‘EZR’ for medical statistics. Bone Marrow Transplantation **48**: 452-458, 2013.