

New report preparation system for endoscopic procedures using speech recognition technology ▶



Authors

Toshitatsu Takao¹, Ryo Masumura², Sumitaka Sakauchi², Yoshiko Ohara¹, Elif Bilgic³, Eiji Umegaki¹, Hiromu Kutsumi⁴, Takeshi Azuma¹

Institutions

- 1 Division of Gastroenterology, Department of Internal Medicine, Kobe University Graduate School of Medicine, Kobe, Japan
- 2 NTT Media Intelligence Laboratories, NTT Corporation, Yokosuka, Japan
- 3 Steinberg-Bernstein Centre for Minimally Invasive Surgery and Innovation, McGill University Health Centre, Montreal, Quebec, Canada
- 4 Center for Clinical Research and Advanced Medicine Establishment Shiga University of Medical Science, Shiga, Japan

submitted 7.9.2017

accepted after revision 3.1.2018

Bibliography

DOI <https://doi.org/10.1055/a-0579-6494> |
Endoscopy International Open 2018; 06: E676–E687
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 2364-3722

Corresponding author

T. Takao, MD, PhD, Division of Gastroenterology, Department of Internal Medicine, Kobe University, Graduate School of Medicine, 7-5-2 Kusunoki-cho, Chuo-ku, Kobe, Hyogo, Japan
Fax: +81-78-382-6309
t.takao1234@gmail.com

ABSTRACT

Background and study aims We developed a new reporting system based on structured data entry, which selectively extracts only endoscopic findings from endoscopists' oral statements and automatically inputs them into appropriate columns in real time during endoscopic procedures.

Methods We compared the time for endoscopic procedures and report preparation (ER time) by using an esophagogastroduodenoscopy simulator in three groups: one preparing reports using a mouse after endoscopic procedures (CE group); a second group preparing reports by using voice alone during endoscopic procedures (SR group); and the final group preparing reports by operating the system with a foot switch and inputting findings using voice during endoscopic procedures (SR+FS group). For the SR and SR+FS groups, we identified the recognition rates of the speech recognition system.

Results Mean ER times for cases with three findings each were 162, 130 and 119 seconds in the CE, SR and SR+FS groups, respectively. The mean ER times for cases with six findings each were 220, 144 and 128 seconds, respectively. The times in the SR and SR+FS groups were significantly shorter than that in the CE group ($P<0.017$). The recognition rate of the SR group for cases with three findings each was 98.4%, and 97.6% in the same group for cases with six findings each. The rates in the SR+FS group were 95.2% and 98.4%, respectively.

Conclusion Our reporting system was demonstrated to allow an endoscopist to efficiently complete the report in real time during endoscopic procedures.

Introduction

It is important in medical services to diagnose and treat patients, however, making accurate records is also of importance. Amid various styles of endoscopic reports using different entry methods (mouse and keyboard, or speech) and different entry structures (free text entry or structured data entry), free text entry has the advantage of providing endoscopists leeway to use whatever expressions they choose when inputting their findings, and yet this could often result in incomplete data entry, as well as making reports difficult to search and to extract

data at a later date [1–3] (▶ **Fig. 1** columns 1 and 2). This is why structured data entry is recommended by the European Society of Gastrointestinal Endoscopy (ESGE) [2]. The advantages of structured data entry include a lower occurrence of incomplete records and easier access to the data afterwards for searching and extracting [2–5]. Therefore, the electronic reporting system, which is the current mainstream, requires endoscopists to input endoscopic findings into structured reports using a mouse and keyboard [1, 6–8] (See upper right of ▶ **Fig. 1** highlighted in orange). However, experience has shown that several

Entry structure Entry method	(A) Free text entry	(B) Structured data entry
(I) Mouse and Keyboard	<p>①</p> <ul style="list-style-type: none"> ● Allows free description. → Suitable for describing findings in detail as well as preparing referral letters ● Not suitable for data retrieval after registering findings. ● Easy to cause incomplete data entry since the entries vary. 	<p>②</p> <p>Compared with ①,</p> <ul style="list-style-type: none"> ● Easy to retrieve data after registering findings since they are registered in a template. ● Hard to cause incomplete data entry since the entries are standardized. <p>However, it has the following disadvantages when using a mouse and keyboard;</p> <ul style="list-style-type: none"> ● Need to spend extra time just for report preparation after endoscopic procedures since both cannot be conducted simultaneously. ● Findings could be forgotten before report preparation. ● Need to input findings one by one from higher level columns in order. ● May have to search several columns to locate a particular finding to input.
(II) Speech		<p>③</p> <p>Maintains the same advantages as ②, with the following additional advantages and disadvantages.</p> <ul style="list-style-type: none"> ● Expected to shorten the overall endoscopy time since this allows report preparation during endoscopic procedures. ● Do not have to worry about forgetting findings. ● Enables input of all findings at the same time, not one by one. ● Need to examine if speech is properly recognized during endoscopic procedures, how patients could be affected by preparing reports during endoscopic procedures, and others.

► Fig. 1 Advantages and disadvantages of each entry method and entry structure and issues solved.

issues exist with current structured endoscopic reports. First, because endoscopists cannot input their findings during endoscopic procedures, they have to spend extra time afterwards solely on completing the reports. They sometimes find it difficult to recall the exact details of lesions after endoscopic procedures, especially when several lesions are discovered. Also, it can take a long time to locate the intended findings, with difficulty finding which column could contain them. Report preparation, out of all endoscopic procedure-related tasks, is laborious, which limits the number of procedures that can be completed [9]. Hence, by enhancing the efficiency of report preparation tasks, it is expected that productivity in the endoscopy suite will improve and endoscopists will be able to spend more time on other beneficial tasks [1, 3, 10].

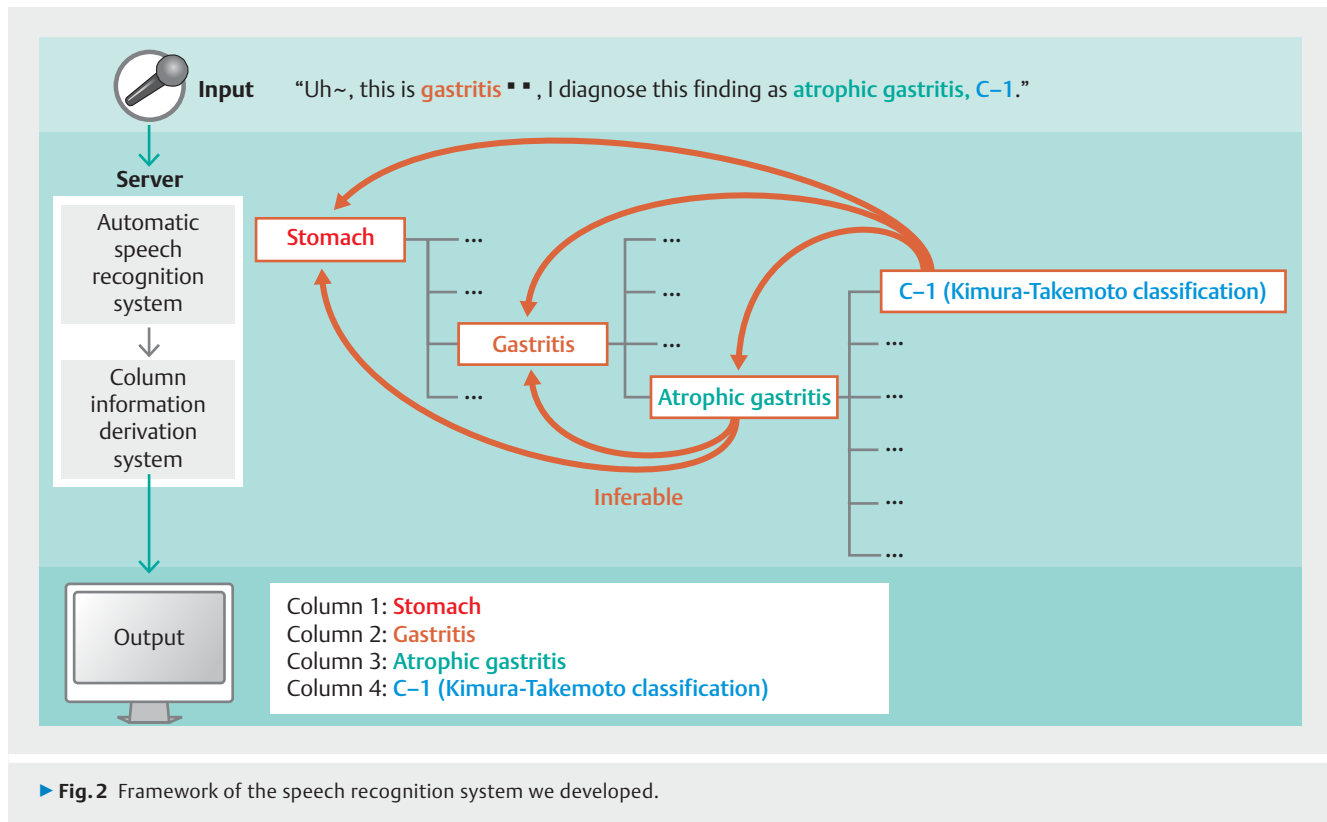
Therefore, we have developed a report preparation system based on structured data entry that selectively extracts only endoscopic findings from endoscopists' oral statements and automatically inputs them into appropriate columns in real

time during endoscopic procedures that occupy both of an endoscopist's hands.

The purpose of this study was to demonstrate that the time spent on endoscopic procedures and report preparation could be shortened by using our newly-developed endoscopic report preparation system.

Methods

This study compared the length of time to complete both the endoscopic procedure and report preparation (ER time) by using an esophagogastroduodenoscopy (EGD) simulator (KOKEN Co., Ltd., Tokyo, Japan) in the following three groups: one with a conventional entry system (CE group); a second group with a speech recognition system (SR group); a the third group with a speech recognition system and a foot switch (SR + FS group). In addition, we also clarified the recognition rate of the speech recognition system for the SR and SR + FS groups.



New speech recognition system

For the SR and SR + FS groups, we developed an endoscopic reporting system utilizing VoiceRex (NTT TechnoCross Corporation, Tokyo, Japan) that extracts only endoscopic findings from endoscopists' spontaneous oral statements and automatically inputs the findings into appropriate columns of the structured reports. The findings do not have to be input into the system in specific order from the first to the last column, unlike in the existing structured data entry system (See lower right highlighted in green of ► **Fig. 1**).

► **Fig. 2** shows the new speech recognition system's framework. A significant feature of this framework is to enable the speech recognition system to directly refer to endoscopic terminologies incorporated in the findings columns. The terminologies are usually configured in the tree structure shown in ► **Fig. 2**. By having the speech recognition system directly refer to this information, we achieved reliable data entry into the columns. This system is also able to infer terms to be input into higher-level columns by recognizing terms entered in a lower level column.

CE group

In the CE group, the reports were prepared by selecting appropriate findings from the pull-down menu, using a mouse after the endoscopic procedures (► **Video 1**). The definition of ER time for the CE group was determined as the total time spent on an endoscopic procedure as well as on report preparation. The time spent on each endoscopic procedure was measured by the endoscopist using a stopwatch, timed from the moment

when the endoscope passed through the mouth of the EGD simulator until it was pulled out from the mouth, after finishing observations of the esophagus, stomach, and duodenum. Furthermore, time spent on report preparation by using a mouse was automatically measured by the reporting system, from the moment when a column was first clicked until the endoscopist selected the last finding and finished data input.



► **Video 1** Report preparation in the CE group.

SR group

In the SR group, by contrast, findings were input during endoscopic procedures for preparation of the reports, using not only oral statements but also voice triggers such as “Start”, “Register” and “Delete” (► **Video 2**). Therefore, for the SR group, the definition of ER time was simply determined as the length of time measured from the moment when the endoscope passed through the mouth of the EGD simulator until it was pulled out from the mouth, after finishing observations of the esophagus, stomach and duodenum, as the endoscopist was able to input and check the findings during each endoscopic procedure. The endoscopist suspended endoscope operation and focused his attention on inputting the findings by speech during endoscopic procedures in the SR group and SR + FS groups. The reason for this was that we considered it difficult to input findings by speech and check if the contents were correct while operating an endoscope in clinical practice.

For the SR group, the endoscopist confirmed the findings by pronouncing “Register” when it was accurate. When it was wrong, the entire row of columns containing the inaccurate finding was deleted by saying “Delete” and the endoscopist pronounced the finding once again. Also, when the system did not react to the endoscopist’s voice, the finding was repeated.

SR + FS group

In the SR + FS group, oral statements were used only for inputting the findings during endoscopic procedures and the USB foot switch (DN-PCACC3UFSWITCH, Dospara Co., Ltd., Tokyo, Japan) was utilized to replace the voice triggers (► **Video 3**). We used the same definition of ER time for the SR + FS group as that for the SR group because the endoscopist was able to input and check the findings during each endoscopic procedure.

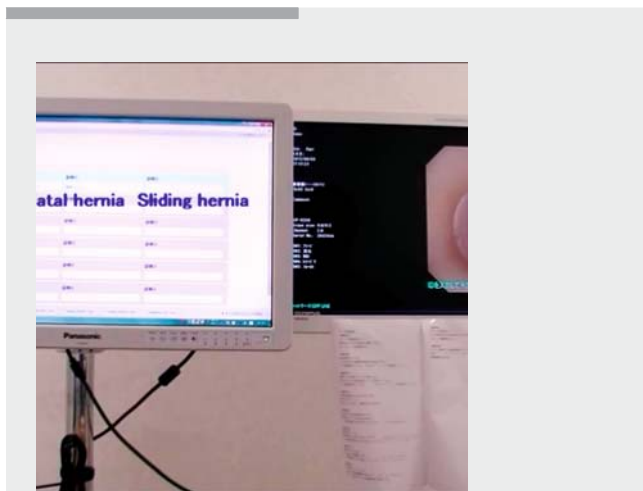
The reporting system started accepting input of findings by speech when the endoscopist pressed the right-hand pedal of the foot switch once. If the finding had been input correctly, the endoscopist pressed the same pedal again to register the finding. If the finding had been input incorrectly, the endoscopist pressed the left-hand pedal of the foot switch to delete the row of findings currently being entered. When the system did not react to the endoscopist’s voice, the finding was spoken once more, in the same way as for the SR group.

Recognition rate and no reaction

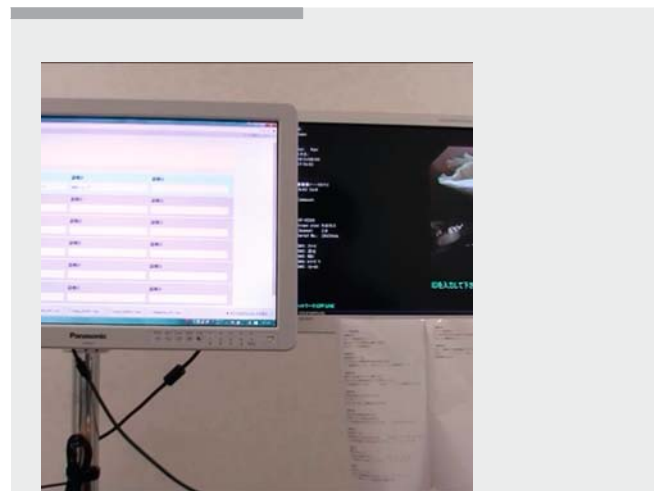
We defined “recognition rate” and “no reaction” as described below.

Each blank space in the reporting system for all three groups is called a column and the system was structured to have names of organs in the “location” column, disease name groups in the “diagnosis 1” column, disease names in the “diagnosis 2” column, and detailed classifications of the diseases in the “diagnosis 3” column (► **Fig. 3**). Whenever the speech recognition system correctly input one complete row of the endoscopist’s dictated findings into the correct columns in one go, without need of correction, this was defined as “accurate.” Four columns placed side by side are called a row of columns, and the “recognition rate” was defined as the number of rows with correct findings divided by the total number of rows. Whenever a finding was not appropriately input in the intended columns, this was defined as an “error.”

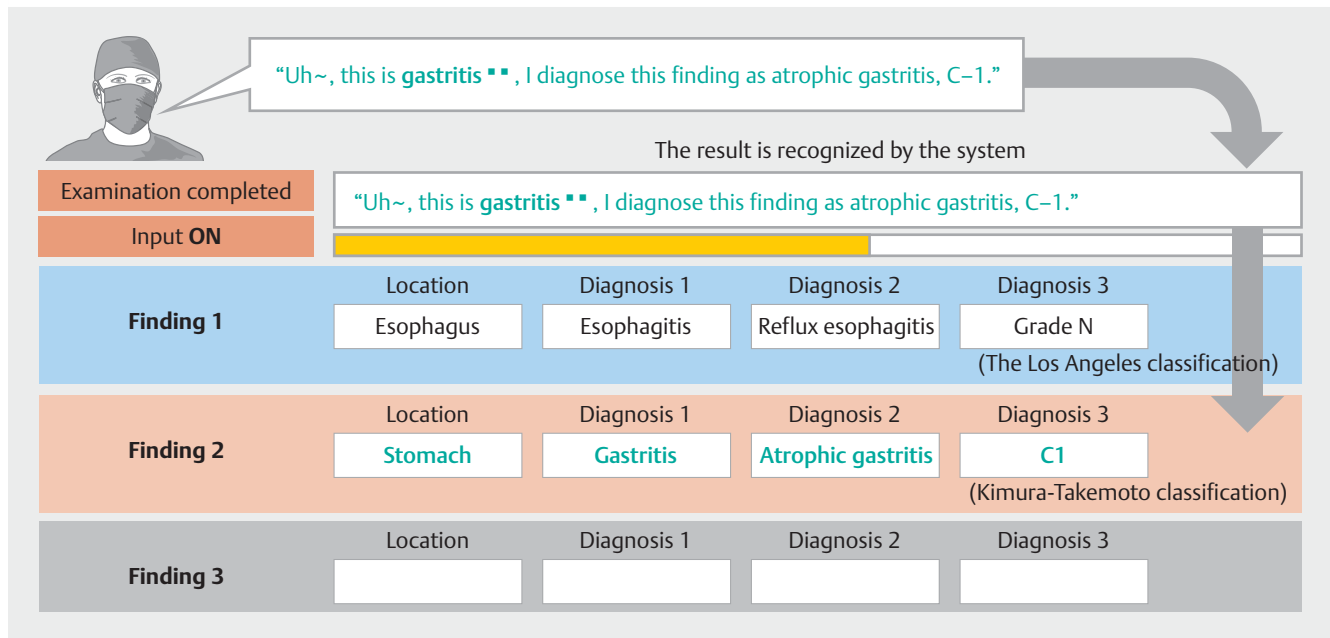
In addition to “accurate” and “error,” it was also possible that the speech recognition system would not react to the endoscopist’s voice when dictating a finding, and this was defined as “no reaction.” As such, the frequency of “no reaction” was studied in two different situations: when an attempt was made to dictate findings in the form of a sentence and when control words such as “start” and “register” were attempted to operate the system by voice.



► **Video 2** Report preparation in the SR group.



► **Video 3** Report preparation in the SR + FS group.



► **Fig. 3** The report preparation system based on structured data entry created for this study. The top column on the screen shows how the system has recognized an endoscopist's naturally-spoken sentence. Next, the system classifies which of the recognized terms are endoscopic terminologies and enters each one of them into the appropriate columns.

Lists of findings to input in the report were prepared before the experiment, and the same lists were used for all three groups (► **Appendix 1**, ► **Appendix 2**). Although Appendix 1 and 2 are written in English, the endoscopist in this study used the lists of findings written in Japanese and vocalized them in Japanese. To assess how the number of findings that were input per case could affect ER time and recognition rate, the lists included 21 cases with three findings per case, and another 21 cases with six findings per case. The vocalized findings were sentences naturally spoken in day-to-day clinical practice, including filler words. It is internationally desirable to use the Minimal Standard Terminology for endoscopic terminologies in a reporting system, however, the Gastroenterological Endoscopy Glossary (Japan Gastroenterological Endoscopy Society) is applied in this system because that is commonly used in Japan. Before the experiment, white marks were placed in nine locations around the EGD simulator. It was agreed that the white marks would be touched in a predetermined order with the tip of biopsy forceps sticking out from the endoscope tip so as to not arbitrarily prolong or shorten the time spent on the endoscopic procedure. The endoscope used for this study was GIF-H260 (Olympus Corporation, Tokyo, Japan), and displays were set as shown in ► **Video 2** and ► **Video 3**. One experienced endoscopist was in charge of all endoscopic procedures as well as report preparation so as to eliminate any variation in results due to different operators. This study was conducted after the endoscopist had become familiar with the simulated procedure and did not take into account a learning curve for the endoscopic procedure. Even so, considering the possibility of the learning curve, we made an effort to reduce this effect by taking one case at a time from each of the CE, the SR, and the

SR + FS groups and changing the order per case, instead of carrying out the procedures one group at a time. A microphone (Savi GO WG100/B wireless headset system, Plantronics, Inc., Santa Cruz, California, United States) was worn on one ear as a speech-input device during endoscopic procedures.

There was no conflict of interest to declare.

Statistics

Our preliminary investigation confirmed that the ER time for the SR and the SR + FS groups was significantly shorter than for the CE group. Accordingly, when comparing the ER time of each group, we expected a large effect size, to which J. Cohen's proposal can be applied [11]. For that reason, to have a statistically significant difference with a certain two-sided significance level ($\alpha=0.05$) and statistical power (80%), we needed 21 cases by referring to the report by J. Cohen [11]. A paired *t*-test was applied to compare the ER time among the three groups using EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan) to study whether there would be a significant difference between them [12]. Because this was a multiple comparison among the three groups, the Bonferroni method was used to adjust the significance level.

► **Table 1** Study results regarding ER time.

		CE group			SR group	SR + FS group
		Endoscopy	Report preparation	Time to complete simulated endoscopy procedure-related tasks	Time to complete simulated endoscopy procedure-related tasks	Time to complete simulated endoscopy procedure-related tasks
Three findings						
Mean, sec		116	46	162	130	119
SD, sec		6	13	15	13	10
Median, sec		116	42	161	129	120
Range, sec		105 – 129	29 – 82	136 – 198	113 – 159	98 – 135
Recognition rate					98.4% (62/63)	95.2% (60/63)
No-reaction rate	When inputting findings				6.3% (4/63)	6.3% (4/63)
	When giving voice commands				1.6% (2/126)	
Six findings						
Mean, sec		117	104	220	144	128
SD, sec		7	20	24	14	17
Median, sec		116	104	218	141	130
Range, sec		106 – 133	77 – 147	189 – 267	123 – 168	103 – 158
Recognition rate					97.6% (123/126)	98.4% (124/126)
No-reaction rate	When inputting findings				0.8% (1/126)	6.3% (8/126)
	When giving voice commands				4.0% (10/252)	
SD, significant deviation.						

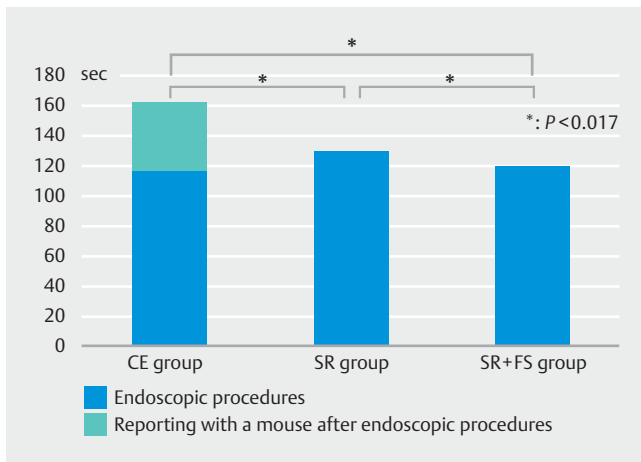
Results

Results of this study regarding ER time appear in ► **Table 1**, ► **Fig. 4** and ► **Fig. 5**. The mean ER time (\pm SD) for cases with three findings each was 162 ± 15 sec for the CE group, 130 ± 13 sec for the SR group, and 119 ± 10 sec for the SR+FS group, which shows that the time for the SR and the SR+FS groups was significantly less than that for the CE group ($P < 0.017$) (► **Table 1**, ► **Fig. 4**). In the same way, the mean ER time (\pm SD) for cases with six findings each was 220 ± 24 sec, 144 ± 14 sec and 128 ± 17 sec for the CE, the SR and the SR+FS group, respectively, showing that the time for the SR and the SR+FS groups was significantly less than that for the CE group ($P < 0.017$) (► **Table 1**, ► **Fig. 4**).

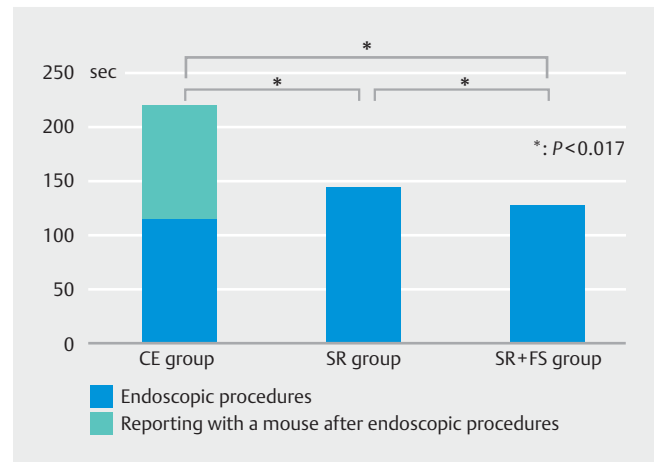
In addition, the recognition rate for the SR Group was 98.4% (62/63 findings) for cases with three findings each, and 97.6% (123/126 findings) for cases with six findings each. The recognition rate for the SR+FS group, by the same token, was 95.2% (60/63 findings) for cases with three findings each, and 98.4% (124/126 findings) for cases with six findings (► **Table 1**). The

findings wrongly recognized by the speech recognition system are listed in ► **Table 2**.

On top of that, we calculated the frequency for which the speech recognition system did not react at all when input of endoscopic findings was attempted. This occurred in 4 out of 63 findings (6.3%) with the SR group for cases with three findings each, and in 1 out of 126 findings (0.8%) for cases with six findings each. At the same time, when control words were spoken to control the system by voice, the system did not react for the SR group 2 times out of 126 voice commands (1.6%) for cases with three findings each, and 10 times out of 252 voice commands (4.0%) for cases with six findings. For the SR+FS group, the speech recognition system did not react in 4 out of 63 findings (6.3%) input for cases with three findings each, and in 8 out of 126 findings (6.3%) input for cases with six findings each (► **Table 1**).



► **Fig. 4** The comparison of the mean ER time for cases with three findings each.



► **Fig. 5** Comparison of the mean ER time for cases with six findings each.

► **Table 2** Performance of the speech recognition system regarding findings.

Speech recognition group	Number of findings	Findings wrongly recognized by speech recognition system
SR	3	There is no abnormal finding in the stomach.
SR	6	There is a gastric telangiectasia.
SR	6	In the stomach, there are some telangiectasias.
SR	6	I'll add SSBE in the report as well.
SR + FS	3	There is no abnormal finding in the stomach.
SR + FS	3	The stomach has Mallory-Weiss syndrome, doesn't it?
SR + FS	3	For the disease name, this finding is diagnosed as gastric lipoma.
SR + FS	6	A part of the stomach is narrowing.
SR + FS	6	This is duodenal Brunner's gland hyperplasia.

Discussion

We newly developed the endoscopic reporting system based on structured data entry using speech recognition technologies. This study demonstrated the possibility of reducing the total time spent on endoscopy-related tasks in clinical practice by using this system.

A study on use of endoscopic reporting utilizing speech recognition systems has been previously published, however, it employed free text entry, which is no longer recommended, and its insufficient recognition rates required extra time for corrections [9]. Furthermore, endoscopic reporting systems using well-known speech recognition technologies, such as Siri for Apple, have not been available for sale so far, and currently there exist no reports on studies using such systems for endoscopic reporting [10, 13].

As such, it is conceivable that the data on this endoscopic reporting system using our newly-developed speech recognition technology will be valuable preliminary data when a speech re-

cognition system is going to be introduced in clinical practice in the future.

ER time

As shown in ► **Table 1**, ► **Fig. 4** and ► **Fig. 5**, the ER time for the SR and the SR + FS groups was significantly shorter than for the CE group. The CE group took longer to prepare reports for cases with six findings than for cases with three findings. This was simply because cases with six findings had more information to input. Moreover, the report preparation time in this group varied per case. A possible explanation for this could be that report preparation time was prolonged due to difficulty locating the findings to be input in the appropriate columns.

For the SR group, there was a delay before the system reacted to speech, not only when inputting findings but also when given voice commands such as "Start" and "Register." This could prolong the length of time spent on the endoscopic procedure itself. As such, we introduced a foot switch in the SR + FS group, which replaced the voice-activated commands. This

group did not have issues where the system would not react, or had a delay before reacting to voice-activated commands each time, yielding a shorter ER time when compared with the SR group. It is conceivable that use of a foot switch could serve in clinical practice, however, there are some points that need to be considered, including that its location should always be fixed and that endoscopists should not be required to look away from the screen.

Speech recognition performance

In this study, the system attained recognition rates of 95% or higher by having the speech recognition system specialize in endoscopic terminologies, in addition to the superior performance of VoiceRex. The number of findings input that contained errors requiring correction was 9 out of 378 findings in total for the SR and the SR+FS groups. Out of these nine findings, “There is no abnormal finding in the stomach” and “telangiectasia” were both recognized wrongly twice (► **Table 2**).

In terms of “no reaction” from the system, voice-activated operations by vocalizing control words in the SR group were required a total of 378 times, with a total of 12 findings (3.2%) resulting in no reaction. Meanwhile, the total number of findings that needed to be input by vocalizing sentences for the SR and the SR + FS groups was 378, with 17 findings (4.5%) resulting in no reaction. Generally speaking, the speech recognition system has lower recognition for short words and words starting with plosives. Therefore, we estimated prior to this study that the system was more likely to be unresponsive to voice commands compared to inputting findings, however, it was not the case this time. This study did not pursue the causes of errors and no reaction. However, prolonged examination time for repeating voice input due to errors or lack of reaction creates disadvantages not only for patients but also for endoscopists. Consequently, further investigation into speech that results in errors and no reaction is warranted to identify the causes and the solutions.

Future issues

Going forward, we are planning to introduce this system in clinical practice, although there are still many issues to solve as stated below. The possible agenda includes: 1) issues about the speech recognition system (measures to deal with surrounding environment such as noise, multilingualization to accommodate not only Japanese but also other languages such as English, an enhancement in speech recognition speed, upgrade of voice-activated operation function); 2) issues other than the speech recognition system (introduction of a system to minimize eye movement of endoscopists); 3) issues when dealing with patients (a possibility of extended time for endoscopy, distraction of attention from endoscopic images owing to the necessity to check the findings input in a separate monitor, less understanding of patient status, a response to lesions that require sensitivity, such as cancer, when communicating the information to patients); and other factors. Moreover, new technologies could cause new types of human errors not listed above [13–15]. Thus, consideration should be given not to cre-

ate disadvantages to patients upon the introduction of this speech recognition system in clinical practice.

Limitations

The limitations of this study included the use of an EGD simulator. Unlike endoscopic procedures in clinical practice, using a simulator requires neither detailed observation of mucosal patterns nor shooting still -images. Usually, the 5 to 8 minutes per patient is spent on endoscopic screening in Japan, although it differs depending on the endoscopist. However, for the above-stated reasons the length of time spent on endoscopic procedures in this study using a simulator was shortened. Nonetheless, we used a simulator in this study as it was necessary to standardize the shape and condition of the stomach when comparing the length of time spent on each group’s endoscopic procedures, otherwise it could affect the outcome. Another reason to note is that use of this reporting system on patients at a stage when its operation had not yet been fully ensured might have caused disadvantages to patients.

In addition, prepared findings could have affected the results. For the CE group, the reports were prepared using a mouse while reading printed materials showing the prepared findings to be entered. However, in clinical practice, as a larger number of lesions and findings are identified, it becomes more difficult to remember all of them precisely. It is thus necessary to examine whether clinical practice will bear the same results to this study going forward.

The speech recognition ability of the system also was evaluated by a single endoscopist, which should be included as a limitation of this study. Speech recognition ability may vary depending on the endoscopist’s age, dialect, voice volume and articulation, as well as individual speaking speeds. Hence, it is necessary that the system be evaluated by multiple endoscopists in the future.

Conclusions

The current study has demonstrated that time spent on endoscopic procedures and report preparation could be shortened by using our newly-developed endoscopic report preparation system.

Competing interests

None

References

- [1] Manfredi MA, Chauhan SS, Enestvedt BK et al. Endoscopic electronic medical record systems. *Gastrointest Endosc* 2016; 83: 29–36
- [2] Bretthauer M, Aabakken L, Dekker E et al. Requirements and standards facilitating quality improvement for reporting systems in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2016; 48: 291–294

- [3] 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* 2012; 4: 76–81
- [4] Gouveia-Oliveira A, Raposo VD, Salgado NC et al. Longitudinal comparative study on the influence of computers on reporting of clinical data. *Endoscopy* 1991; 23: 334–337
- [5] Soekhoe JK, Groenen MJ, van Ginneken AM et al. Computerized endoscopic reporting is no more time-consuming than reporting with conventional methods. *Eur J Intern Med* 2007; 18: 321–325
- [6] Aabakken L, Westerheim J, Hofstad B et al. SADE database for endoscopic procedures: aspects of clinical use. *Endoscopy* 1991; 23: 269–271
- [7] ASGE Technology Committee; Conway JD, Adler DG, Diehl DL et al. Endoscopic electronic medical record systems. *Gastrointest Endosc* 2008; 67: 590–594
- [8] Wassef W, Canto M, Birk J. Toward more user-friendly electronic endoscopy information systems: role of accessories. *Gastrointest Endosc Clin N Am* 2004; 14: 755–762
- [9] Massey BT, Geenen JE, Hogan WJ. Evaluation of a voice recognition system for generation of therapeutic ERCP reports. *Gastrointest Endosc* 1991; 37: 617–620
- [10] Johnson M, Lapkin S, Long V et al. A systematic review of speech recognition technology in health care. *BMC Med Inform Decis Mak* 2014; 14: 94
- [11] Cohen J. A power primer. *Psychol Bull* 1992; 112: 155–159
- [12] Kanda Y. Investigation of the freely available easy-to-use software 'EZ' for medical statistics. *Bone Marrow Transplantation* 2013; 48: 452–458
- [13] Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. *J Am Med Inform Assoc* 2016; 23: e169–e179
- [14] Coiera E. Technology, cognition and error. *BMJ Qual Saf* 2015; 24: 417–422
- [15] Coiera E, Aarts J, Kulikowski C. The dangerous decade. *JAMIA* 2012; 19: 2–5

► **Appendix 1** Lists of findings for 21 cases with three findings per case

Case	Findings
1	Nothing particular in the esophagus.
	For the disease name, this finding is diagnosed as erosive gastritis.
	Nothing particular in the duodenum.
2	There is no abnormal finding in the esophagus.
	This is obviously atrophic gastritis, open type 2.
	There is no abnormal finding in the duodenum.
3	Although it's mild, I diagnose this finding as a sliding hernia.
	There are multiple hyperplastic polyps mainly in the gastric body.
	No abnormal finding in the duodenum.
4	I diagnose this reflux esophagitis as grade B.
	There is no abnormal finding in the stomach.
	There is erosive duodenitis in the bulb.
5	This is reflux esophagitis, grade A.
	I see erosive gastritis in the gastric antrum.
	There is extrinsic compression in the duodenal bulb.
6	There is no abnormal finding in the esophagus.
	This atrophic gastritis is evaluated to be closed type 2.
	This duodenal ulcer scar is at S2 stage.
7	SSBE is there.
	This lesion is a gastric ulcer, A1 stage.
	This duodenal ulcer is at A2 stage.
8	I think there is no abnormal lesion in the esophagus.
	I see a gastric ulcer scar with S2 stage.
	I'll diagnose this lesion as duodenal Brunner's gland hyperplasia.
9	Well, nothing particular in the esophagus.
	Billroth I reconstruction has been performed on this stomach.
	Okay, no lesion in the duodenum.
10	Oh! Those are quite big esophageal varices, aren't they?
	It's PHG, isn't it?
	There is also periampullary diverticulum, right?
11	Let me see, no abnormal findings in the esophagus.
	For the disease name, I'll call it hemorrhagic gastritis.
	Okay, no abnormalities in the duodenum.

► **Appendix 1** (Continuation)

Case	Findings
12	There is nothing abnormal in the esophagus.
	This is obviously nodular gastritis.
13	There is no abnormal finding in the duodenum.
	Although it's mild, I diagnose this finding as a sliding hernia.
	Possibly, this lesion is eosinophilic gastroenteritis.
14	There is no abnormal finding in the duodenum.
	I diagnose this reflux esophagitis as grade B.
	This lesion is acute gastric mucosal lesion.
15	There is erosive duodenitis in the bulb.
	This finding is reflux esophagitis, grade A.
	The stomach has Mallory-Weiss syndrome, doesn't it?
16	There is extrinsic compression in the duodenal bulb.
	There is no abnormal finding in the esophagus.
	There is food residue in the stomach.
17	This duodenal ulcer scar is at S2 stage.
	SSBE is there.
	I see some gastric angiectasias.
18	This duodenal ulcer is at A2 stage.
	I think there is no abnormal lesion in the esophagus.
	There is a gastric submucosal tumor.
19	I'll diagnose this lesion as duodenal Brunner's gland hyperplasia.
	Well, nothing particular in the esophagus.
	I see gastric aberrant pancreas in the antrum.
20	Okay, no lesions in the duodenum
	Oh ! There are quite big esophageal varices.
	I suspect that this elevation is gastric GIST.
21	There is also periampullary diverticulum, right?
	Let me see, no abnormal findings in the esophagus.
	For the disease name, this finding is diagnosed as gastric lipoma.
	Okay, no abnormalities in the duodenum.

► Appendix 2 Lists of findings for 21 cases with six findings per case

Case	Findings
1	Let's see. This reflux esophagitis is at grade A.
	This is a mild sliding hernia.
	You see, there is erosive gastritis.
	I see several fundic gland polyps.
	There is an adenoma in the stomach.
	There is no abnormal finding in the duodenum.
2	Okay, I'd say there is no abnormal finding in the esophagus.
	This atrophic gastritis is categorized into open type 2.
	These are all gastric hyperplastic polyps.
	Gastric xanthoma is also seen.
	There is erosive duodenitis.
	Oh, there is also periampullary diverticulum.
3	Mild sliding hernia is seen there.
	Let me see, this reflux esophagitis is at grade A.
	Oh, SSBE is also there.
	There are some fundic gland polyps.
	This lesion is called duodenal Brunner's gland hyperplasia.
	Oops, this is periampullary diverticulum, isn't it?
4	Cough, cough. I'm sorry. This lesion is reflux esophagitis, grade B.
	This is Candida esophagitis.
	This is a sliding hernia and the severity is moderate.
	I'll diagnose this as erosive gastritis.
	Oh, there is also a submucosal tumor in the stomach.
	Well, this is erosive duodenitis.
5	There is Candida esophagitis.
	There is also solitary esophageal varix.
	This is Mallory-Weiss syndrome.
	Superficial gastritis is also seen.
	Erosive gastritis is there, isn't it?
	There is no abnormal finding in the duodenum.
6	Alright. There is no abnormal finding in the esophagus.
	There is atrophic gastritis, closed type 2.
	This would be fine as intestinal metaplasia.
	Oh, I can see gastric xanthoma as well.
	There is a gastric telangiectasia.
	Oh, there is a duodenal ulcer scar at S2 stage.

► Appendix 2 (Continuation)

Case	Findings
7	The esophagus is being invaded by a primary cancer of another organ.
	The origin is advanced gastric cancer.
	A part of the stomach is narrowing.
	I see gastric carcinoid tumor as well.
	Atrophic gastritis, open type 3, is seen in the background mucosa.
	Nothing particular in the duodenum.
8	There is no lesion in the esophagus.
	I think I'll classify this condition as cascade stomach.
	Wow, there's an acute gastric mucosal lesion.
	I exchanged gastrostomy tube today.
	Well, this is duodenal Brunner's gland hyperplasia.
	Ectopic gastric mucosa is in the duodenum.
9	Reflux esophagitis is seen. Its grade is M.
	A sliding hernia is also seen. It's mild though.
	Also I want to say this is SSBE.
	I suspect that this lesion in the antrum is gastric malignant lymphoma.
	This lesion in the duodenum may well be GIST.
	The neighboring lesion is duodenal Brunner's gland hyperplasia.
10	There is esophageal diverticulum in the middle thoracic esophagus.
	This patient has previously undergone a gastric ESD.
	As previously recognized, GAVE is still present.
	This patient has undergone APC therapy before.
	The protrusion with the recess on top is gastric aberrant pancreas.
	There is also periampullary diverticulum in the duodenum.
11	Reflux esophagitis, grade A, is seen.
	This is a mild sliding hernia.
	Hmm, atrophic gastritis is seen over a broad area.
	Some gastric hyperplastic polyps are seen.
	There is adenoma in the stomach.
	There is no abnormal lesion in the duodenum.

► Appendix 2 (Continuation)	
Case	Findings
12	There is no abnormal finding in the esophagus.
	I diagnose the grade of this atrophic gastritis as open type 2.
	Gastric hyperplastic polyp is also seen.
	This is gastric ulcer scar at S2 stage.
	There is erosive duodenitis.
	There is also periampullary diverticulum.
13	This is a mild sliding hernia.
	This reflux esophagitis is categorized into grade A.
	The finding is slight, but SSBE is recognized.
	There is a lesion suspected as gastric MALT lymphoma.
	There is duodenal Brunner's gland hyperplasia in the bulb.
	Periampullary diverticulum is also seen.
14	I see reflux esophagitis, grade B.
	This is mild Candida esophagitis.
	This sliding hernia is mild.
	I diagnose this finding as superficial gastritis for now.
	There is a protrusion suspected as a submucosal tumor in the stomach.
	Erosions are seen in the duodenum. I diagnose them as erosive duodenitis for the time being.
15	This candida esophagitis is mild, right?
	Solitary esophageal varix in the middle thoracic esophagus.
	Oh my gosh! Mallory-Weiss syndrome is seen.
	This patient has previously undergone a gastric ESD.
	Intestinal metaplasia is seen in the stomach.
	Nothing particular in the duodenum.
16	Okay, there is no abnormal finding in the esophagus.
	Atrophic gastritis, closed type 3, is seen.
	There is bleeding.
	That's Dieulafoy's lesion.
	In the stomach, there are some telangiectasias.
	There is a duodenal ulcer scar, S2 stage.

► Appendix 2 (Continuation)	
Case	Findings
17	Nothing particular in the esophagus.
	There is large advanced cancer.
	A part of the stomach is narrowing.
	Atrophic gastritis, open type 3, is seen in the background mucosa.
	The stomach isn't fully observed yet.
	Nothing particular in the duodenum.
18	There is no abnormal lesion in the esophagus.
	Shortening of lesser curvature is seen.
	I exchanged gastrostomy tube today.
	This is duodenal Brunner's gland hyperplasia.
	There is also a duodenal ulcer scar at S2 stage.
	Extrinsic compression is there in the duodenum.
19	This reflux esophagitis is at grade C.
	A moderate sliding hernia is also seen.
	I'll add SSBE in the report as well.
	There is no lesion in the stomach.
	Is the lesion in the duodenum a case of GIST?
	The neighboring lesion is duodenal Brunner's gland hyperplasia.
20	There is an esophageal granular cell tumor in the middle thoracic esophagus.
	This patient has previously undergone a gastric ESD.
	As previously recognized, DAVE is still present.
	This patient has undergone APC therapy before.
	Atrophic gastritis is also seen.
	There is also periampullary diverticulum in the duodenum.
21	Reflux esophagitis, grade A, is seen.
	This is a mild sliding hernia.
	This stomach has undergone distal Gastrectomy with Roux-en-y reconstruction.
	Some gastric hyperplastic polyps are seen.
	There is an adenoma in the stomach.
	There is no abnormal finding in the duodenum.