



Original Article

Corresponding Author

In Ho Han

<https://orcid.org/0000-0001-7193-6533>

Department of Neurosurgery, Pusan National University Hospital, 179 Gudeok-ro, Seo-gu, Busan 49241, Korea
Email: farlateral@hanmail.net

Co-corresponding Author

Min Gyu Kim

<https://orcid.org/0000-0003-0744-8812>

Human-Robot Interaction Center, Korea Institute of Robotics and Technology Convergence, 39 Jigok-ro, Nam-gu, Pohang 37666, Korea
Email: mingyukim@kiro.re.kr

*Kyoung Hyup Nam and Da Young Kim contributed equally to this study as co-first authors.

Received: October 26, 2021

Revised: December 23, 2021

Accepted: March 2, 2022



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2022 by the Korean Spinal Neurosurgery Society

INTRODUCTION

With the advent of the Fourth Industrial Revolution, efforts to apply artificial intelligence (AI) and machine learning in the medical field are actively underway.^{1,2} In particular, imaging diagnosis, disease diagnosis, and prediction using clinical data and genomic Big data are medical fields of AI that currently receive the most attention.^{3,4} AI technologies associated with natural language processing (NLU) are also being used in healthcare.⁵ Conversational AI is an application of NLU and refers to

Conversational Artificial Intelligence for Spinal Pain Questionnaire: Validation and User Satisfaction

Kyoung Hyup Nam^{1,*}, Da Young Kim^{2,*}, Dong Hwan Kim¹, Jung Hwan Lee¹, Jae Il Lee¹, Mi Jeong Kim¹, Joo Young Park¹, Jae Hyun Hwang¹, Sang Seok Yun³, Byung Kwan Choi¹, Min Gyu Kim², In Ho Han¹

¹Department of Neurosurgery, Medical Research Institute, Pusan National University Hospital, Busan, Korea

²Human-Robot Interaction Center, Korea Institute of Robotics and Technology Convergence, Pohang, Korea

³Division of Mechanical Convergence Engineering, Silla University, Busan, Korea

Objective: The purpose of our study is to develop a spoken dialogue system (SDS) for pain questionnaire in patients with spinal disease. We evaluate user satisfaction and validated the performance accuracy of the SDS in medical staff and patients.

Methods: The SDS was developed to investigate pain and related psychological issues in patients with spinal diseases based on the pain questionnaire protocol. We recognized patients' various answers, summarized important information, and documented them. User satisfaction and performance accuracy were evaluated in 30 potential users of SDS, including doctors, nurses, and patients and statistically analyzed.

Results: The overall satisfaction score of 30 patients was 5.5 ± 1.4 out of 7 points. Satisfaction scores were 5.3 ± 0.8 for doctors, 6.0 ± 0.6 for nurses, and 5.3 ± 0.5 for patients. In terms of performance accuracy, the number of repetitions of the same question was 13, 16, and 33 (13.5%, 16.8%, and 34.7%) for doctors, nurses, and patients, respectively. The number of errors in the summarized comment by the SDS was 5, 0, and 11 (5.2%, 0.0%, and 11.6%), respectively. The number of summarization omissions was 7, 5, and 7 (7.3%, 5.3%, and 7.4%), respectively.

Conclusion: This is the first study in which voice-based conversational artificial intelligence (AI) was developed for a spinal pain questionnaire and validated by medical staff and patients. The conversational AI showed favorable results in terms of user satisfaction and performance accuracy. Conversational AI can be useful for the diagnosis and remote monitoring of various patients as well as for pain questionnaires in the future.

Keywords: Conversational artificial intelligence, Pain questionnaire, Spoken dialogue system, Natural language process, Chatbot, Spine

AI technology that can talk to people, including chatbots or virtual agents.⁶

Unlike the written text-based chatbot, a computer system that can communicate by voice is called a spoken dialog system (SDS).⁷ Unlike the command and control speech system that simply answers requests and cannot maintain the conversation continuously, a SDS can maintain the continuity of the conversation over long periods of time. SDSs are already being applied in everyday life through in-home AI speakers, such as Amazon Alexa (Amazon, Seattle, WA, USA).⁸ Moreover, conversational AI

is being applied in various medical fields, such as patient education, medical appointments, and voice-based electronic medical record (EMR) creation.^{7,9} Recent attempts have been made to collect medical data, such as patient-reported outcomes, health status checks and tracking, and remote home monitoring, using conversational AI.^{7,10}

In assessing patient with spinal disease, doctor-patient dialogue about pain is the first step in diagnosis, and a pain questionnaire is the most important tool during follow-up after treatment or spine surgery. The purpose of our study was to develop a SDS for a pain questionnaire for patients with spinal diseases. We aimed to evaluate user satisfaction and validate the performance accuracy of the system in medical staff and patients. This study is a preliminary study for the development of an interactive medical robot. Based on the results of this study, a follow-up study on robot-based interactive questionnaire is planned.

MATERIALS AND METHODS

1. Development of the Pain Questionnaire Protocol

First, a pain questionnaire protocol for a SDS was developed by dividing the preoperative and postoperative pain questionnaires to assess the outcomes of patients undergoing spine surgery. The pain questionnaire consisted of questions to reflect the actual conversation between the medical staff and the patient. The items were created based on questions that medical

staff usually ask during rounds of inpatients. The protocol included questions about the location, type, influencing factors, intensity, time of onset, and duration of pain. In addition, questions about the patient’s psychological state, such as questions regarding mood, anxiety, and sleep quality, were included as indirect indicators of pain. Postoperative question items were replaced with question items about pain at the surgical site. Furthermore, a question about whether the patient’s preoperative pain had improved or not was added. Questions about psychological status were the same as the preoperative questions. Each question was structured in a closed question format so that the pain questionnaire system could easily process the patients’ responses. The developed pain questionnaire protocol is shown in Table 1.

2. Collection of Dialogue Dataset for the SDS

To build a database of patients’ various expressions for NLU, real doctor-patient dialogue sets were collected. The study was approved by the Institutional Review Board (IRB No. 1905-023-079). Informed consent was obtained from all patients. A total of 1,314 dialogue sets were collected from 100 hospitalized patients who underwent spinal surgery between September 2019 and August 2021. One dialogue set was defined as one question and one answer. The age range was 22–82 years (mean, 62.6 years), and 47 patients were male. There were 48 spinal stenosis, 13 herniated disc herniation, 13 spinal infection, 11 spinal

Table 1. Questionnaire of the spoken dialogue system

Category	Description	Preoperative situation	Postoperative situation
Pain	Location	Where is the most painful area right now? If there are multiple areas, please tell them briefly in the order of the most pain.	Where do you feel most uncomfortable after surgery? If there are multiple parts, please tell them briefly in the order of discomfort.
	Type	How does the pain feel? Please express it like numbness or aching.	Has the pain that was very painful before the operation improved?
	Influence factor	1. What time of the day do you have the most pain? 2. What posture hurts the most?	1. Is there any pain at the surgical site? 2. What posture hurts the most?
	Intensity	Please rate how severe the pain is on a scale of VAS 0–10.	Please rate how severe the pain is on a scale of VAS 0–10.
	Time and duration	1. Since when have you had pain? 2. When did the pain get worse?	Does the pain at the surgical site last all day?
Psychologic state	Mood	How are you feeling right now? Please tell me between good, average, and bad.	How are you feeling right now? Please tell me between good, average, and bad.
	Anxiety	Are you currently worried or anxious?	Are you currently worried or anxious?
	Quality	1. How many hours did you sleep? 2. Did you sleep well without waking up?	1. How many hours did you sleep? 2. Did you sleep well without waking up?

VAS, visual analogue scale.

tumor, 8 spinal deformity, 4 spine trauma, and 3 myelopathy cases.

Three doctors asked inpatients questions naturally following the pain questionnaire protocol during the rounds, and the conversations were recorded using a voice recorder. The preoperative pain questionnaire was used the day before surgery, and the postoperative pain questionnaire was used between 3 and 7 days after surgery. The recordings were documented in the format of text and stored in a database for NLU. Additionally, the virtual conversations of the researchers were also collected, and 2,000 dialogue sets were used for the database.

3. Development of the SDS

The SDS was structured as shown in Fig. 1. The patient's response voice was entered into the speech recognition module and converted into text data. The text data was the input value of the NLU module. The NLU module played a role in understanding users' intentions by analyzing the intents, name entity recognition, and keywords in the user's answers. The output value of the NLU module was again entered into the dialog management module, which managed the flow of conversation between the user and SDS. It searched the database for information to be given to the user and outputted the content necessary for system utterance. The system utterance output was automatically generated in the format of text data through a natural language generation (NLG) module, which was again keyed

into a speech synthesis module. The speech synthesis module finally completed system utterance generation by outputting the result in a voice format that the user can understand. The pain questionnaire SDS was developed using Python 3.8 for Windows 10. IBM Watson Text-to-Speech (IBM, Armonk, NY, USA) was used for utterance of the SDS. NLU was performed using IBM Watson Assistant and KoNLPy to understand the patient's intent from the text data.¹¹ The utterance was performed using the Google Cloud Speech-to-Text module (Google, Mountain View, CA, USA).

After analyzing the dialogue datasets obtained from the patients, patients' intents that express the character of pain and psychological state were classified into 95 in the intents column of IBM Watson Assistant. A total of 1,229 expression examples were registered in the user example of the intent column. A total of 770 examples for timing, duration, and influence factors were registered in the name entity column.

Fig. 2 shows the conversation flow of the pain questionnaire SDS for implementation of the questionnaire protocol in Table 1. The SDS starts by entering a unique identification number (UID) that anonymizes the patient's personal information and stores it in the virtual EMR. When the UID is entered, the SDS checks whether the UID exists in the database. If the UID exists, the SDS starts asking questions after repeating the previous questionnaire's summary. The SDS checks whether the answers to the 10 questions were obtained from the patient during the

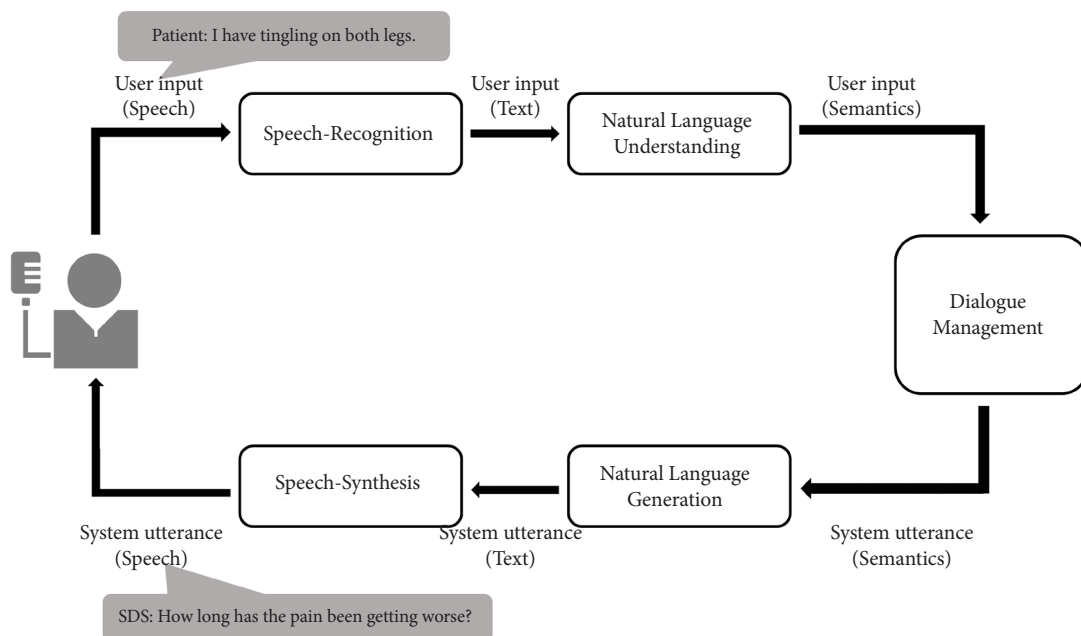


Fig. 1. Architecture of the spoken dialogue system (SDS).

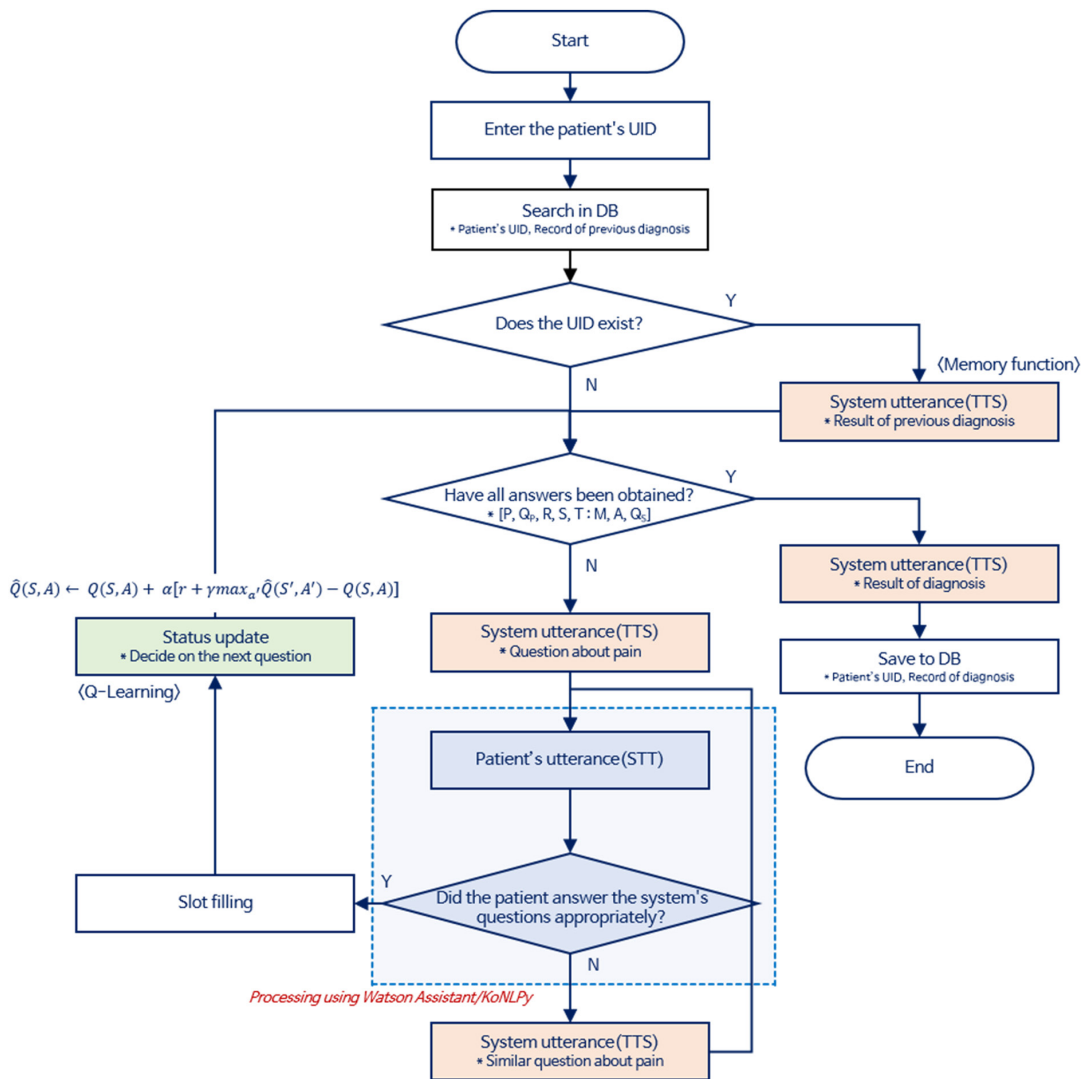


Fig. 2. Flow chart of the spoken dialogue system. UID, unique identification number; DB, database; STT, speech-to-text; TTS, text-to-speech.

Date : 2021-08-18 17:51:51
Position : ['wrist']
Quality : ['tingling']
R_factor1 : ['always']
R_factor2 : ['when using a computer']
Severity : 6
Timing2 : 1 week
Mood : good
Anxiety : no
Sleep1 : 7
Sleep2 : good

Fig. 3. Electric medical record linkage with the spoken dialogue system.

questionnaire. If proper information was not obtained, the SDS asks the question until proper information is obtained. When the patient's answer is not recognized, the SDS utters similar questions without repeating the same question. When the patient's answer is properly recognized, the Q-learning status is updated to determine the next question, and the SDS checks whether all answers are obtained. When all answers are obtained, the SDS utters the summarized result and finishes the questionnaire after saving the results in the form of text in the virtual EMR. An example of the questionnaire results that were transmitted to the virtual EMR is shown in Fig. 3. Supplementary video clip 1 is actual conversation video between SDS and a participant.

4. Validation and User Satisfaction for the SDS

User satisfaction and performance accuracy of the developed pain questionnaire SDS were evaluated. Validation of the SDS was performed for 3 user groups: doctors, nurses, and patients. The participants volunteered to be recruited. The study was approved by the Institutional Review Board of Pusan National University Hospital (IRB No. 2012-010-097). Informed consents were obtained from all patients. Ten participants were included in each group. The participants were pretrained to engage in routine conversations rather than simple short-answer conversations. The participants were provided with basic information about the purpose of the study and SDS, and we helped the participants adapt to the conversation with the SDS. The mean ages of the doctors, nurses, and patients were 35.3 years (range, 25–47 years), 31.2 years (range, 21–58 years), and 64.0 years (range, 48–82 years), respectively. The male-to-female ratios in doctors, nurses, and patients were 9:1, 10:0, and 5:5, respectively. Validation of the SDS was performed in a sitting position on the bed of an inpatient ward. The SDS was mounted on a laptop notebook and placed on a bed table. When the start button was pressed, a conversation was initiated automatically. The SDS first asked a question about pain and recognized the answer, and it followed up with further questions. After the last question and answer, the SDS uttered the summarized result to the users and ended the program immediately after the test, the participants completed a user satisfaction questionnaire about

SDS. The questionnaire consisted of 10 items, including the accuracy of the SDS’s voice, the degree of similarity to human conversation, and overall satisfaction, and followed the 7-point Likert scale.⁷

5. Statistical Analysis

To verify the performance accuracy of the SDS, the recognition error of the patient’s answer, summary error, the causes of the errors, and summary omission of the summarized comment were analyzed. User satisfaction and accuracy between the participant groups were statistically analyzed using 1-way analysis of variance and *post hoc* Tukey honestly significant difference analysis. A p-value of < 0.05 was considered statistically significant.

RESULTS

1. User Satisfaction

The results of the user satisfaction survey are shown in Table 2. The overall satisfaction score of 30 participants, consisting of doctors, nurses, and patients, was measured with an average of 5.5 ± 1.4 points out of 7 points. The satisfaction score was 5.3 ± 0.8 for doctors, 6.0 ± 0.6 for nurses, and 5.3 ± 0.5 for patients. The nurse group showed a higher level of satisfaction, but there was no statistically significant difference in user satisfaction between the groups (p = 0.136). The average score for each item

Table 2. Survey results of the spoken dialogue system

Question items	Doctor	Nurse	Patient	p-value
Q1. I could understand SDS’s words well.	6.6 ± 0.699	6.8 ± 0.422	5.7 ± 1.059	0.008*
Q2. The volume, speed, and sound quality of the SDS were adequate.	6.6 ± 0.699	6.7 ± 0.675	6.0 ± 1.155	0.171
Q3. SDS asked the proper questions.	5.4 ± 1.506	6.0 ± 0.816	5.2 ± 1.549	0.390
Q4. SDS gave an appropriate response.	4.9 ± 1.370	5.3 ± 1.418	5.4 ± 1.075	0.664
Q5. In conversation with SDS, I was able to fully express what I wanted to say.	4.6 ± 1.647	5.5 ± 1.179	4.5 ± 1.269	0.222
Q6. SDS seems to understand well what I’m saying.	4.7 ± 1.337	5.6 ± 1.265	4.7 ± 1.252	0.214
Q7. Conversations with SDS were not much different from conversations with people.	4.3 ± 1.337	5.4 ± 1.075	5.1 ± 1.370	0.153
Q8. I think positively about assisting my medical care through the SDS conversation.	5.5 ± 1.179	6.1 ± 0.568	5.0 ± 1.491	0.118
Q9. There was no objection to the conversation with SDS.	5.3 ± 1.494	6.5 ± 0.527	6.1 ± 0.994	0.057
Q10. The conversation with SDS was overall satisfactory.	5.3 ± 1.160	6.2 ± 0.919	5.4 ± 1.265	0.165
Mean	5.3 ± 0.777	6.0 ± 0.547	5.3 ± 0.522	0.136

Values are presented as mean ± standard deviation. SDS, spoken dialogue system.

*One-way analysis of variance, *post hoc* Tukey honestly significant difference analysis was performed, the group that made the difference.

was relatively high at > 6 points for the items Q1, Q2, and Q9, which reflect the clarity and sound quality of the SDS's voice and positivity in conversation. The items, Q5, Q6, and Q7, which

indicate the degree of similarity to conversations with real people, showed relatively low satisfaction. In terms of satisfaction among the participant groups for each item, the patients showed a statistically low score for item Q1 ($p=0.008$). Other items showed no statistical differences between the groups.

Table 3. Summary of errors in the spoken dialogue system

Participant	Total no. of question items	Total no. of questions	Recognition error	Summary error	Omission error
Doctor	96	109	13 (13.5)	5 (5.2)	7 (7.3)
Nurse	95	111	16 (16.8)	0 (0.0)	5 (5.3)
Patient	95	128	33 (34.7)	11 (11.6)	7 (7.4)
p-value			0.063	0.001*	0.857

Values are presented as number (%).

*One-way analysis of variance, *post hoc* Tukey honestly significant difference analysis was performed, the group that made the difference.

2. Performance Accuracy

The SDS asked 95 and 96 questions per group, respectively. The number of repeated questions asked by the SDS because it did not recognize the participant's answers was 13, 16, and 33 (13.5%, 16.8%, and 34.7%) in doctors, nurses, and patients, respectively. The difference in the number of repeated questions was not statistically significant among the 3 groups ($p=0.063$). However, the SDS did not recognize the answers of the patient group and tended to increase the repetition of questions. After

Table 4. Example of an actual conversation between the spoken dialogue system and a patient*

Question of spoken dialogue system	Answer of patient
Hello, it's interview time. I'll start the interview now. What is most uncomfortable symptom right now? If there are multiple parts, please tell them briefly in the order of the most pain.	Yes, My butt hurts. It hurts on both sides of the buttocks on the legs
How does the pain feel? Please express it like numbness or aching.	I get numb and numb. I have tingling when I walk.
How long has the pain been getting worse? Please answer with a period such as 1 day, 1 week, 1 month and 1 year.	No, what I'm doing here is that it hurt before the operation, but after the operation, it seems like yesterday.
I didn't hear well. When did the pain get worse? Please answer with a period such as 1 day, 1 week, 1 month and 1 year.	I can't remember how many years it was before the surgery, and it's been less than a week since I came in after the operation, but it's been 5 days since the operation.
Please rate how severe the pain is on a scale of 0-10.	It seems like the pain is something like 9 or 100 points or something. It hurts when I walk. Buttock. I can't get out of bed.
What time of day do you have severe pain? Please answer the time such as in the morning or before bed.	Mainly, I lay down and went to work, and when the bed goes up, I have to move my ass with it with that. The bed comes down and I can't do it properly with that. I need help.
I didn't hear well. When is the most painful time of the day? Please answer the time such as in the morning or before bed.	Usually, when I lie on the bed and then get up and try to get down from the bed to urinate, this buttock doesn't move properly. That's why I have to keep helping you so I can come down. I was sick. it's time. What, when will you come down?
Please tell me how you are feeling right now? Please answer your feeling between good, average and bad.	It's uncomfortable to sit right now.
Are you currently worried or anxious?	Yes
This is question about sleep quality. Did you sleep well without waking up?	Yes, I sleep well.
How many hours did you sleep last night?	Well, I didn't sleep for 3 days after surgery and I slept well last night.
I didn't hear well. how much time did you sleep yesterday?	How many hours. Well, 9 hours and 7 hours.
[Summarized comment of SDS]	
Yes, I see. Then I will inform you about the results of the questionnaire. Your pain area is your leg, and you have a tingling sensation, getting worse when you wake up. The intensity of the pain is 10, and the pain has gotten worse since about a week ago. And you say you feel bad and have anxiety. You said that you slept 9 hours and the quality of your sleep was good.	

*Translated questions and answers conducted in Korean into English.

the pain questionnaire was completed, the number of errors in the summarized comment was measured to be 5, 0, and 11 (5.2%, 0.0%, and 11.6%) for doctors, nurses, and patients, respectively. In particular, there were no summary error in nurses. There was a statistically significant difference between the groups ($p = 0.001$). The number of summarization omissions was 7, 5, and 7 (7.3%, 5.3%, and 7.4%), respectively, and there was no statistical difference between the groups ($p = 0.857$) (Table 3).

3. A Case of an Actual Conversation Between a Patient and the SDS

Table 4 shows the actual conversation content between the pain questionnaire SDS and a patient who was an 82-year-old woman. The patient reported a pain intensity score of 9 points for 2 or 3 days prior, but the SDS recognized it as a 10-point intensity for 1 week prior. Because the patient described her symptoms in great detail, it was difficult to accurately recognize specific factors, such as the timing and intensity of the pain. For example, when explaining the symptom duration, she did not mention the exact date, saying "It's been less than a week since I came in after the operation, but it's been 5 days since the operation." Even when talking about the score for pain intensity, she did not express it accurately. Even though the patient used a regional Korean dialect, the dialect had no effect on the processing results because the SDS processed the answers centered on the keywords. However, it was confirmed that the patient did not predict the end time of the utterance of the SDS and thus responded during the utterance of the SDS.

DISCUSSION

Conversational AI is increasingly being used in medical health-care field.^{6,9} Conversational AI, such as voice chatbots and voice assistants, can provide primary medical education services that answer common questions based on knowledge databases. For example, if people ask a question about first aid in the case of a fever or insect bites, the SDS can tell the treatment method via voice.¹² Recently, hospitals have been actively introducing a doctor appointment service using chatbots.¹³ Currently, the most actively researched field is document automation through voice recognition.^{7,14} Speech recognition technology can dramatically reduce the time required to write medical records for doctors and nurses by automatically inputting data in the medical records. It has been reported that this technology has reduced the burden and fatigue experienced by doctors and nurses and increased the time spent caring for patients.¹⁵ In addition, conver-

sational AI can be used to automate patient data collection as the SDS used in this study can collect important medical history and patient-reported outcomes. It can also be used for remote home monitoring.^{9,10,15}

The term, voice-based conversational AI, is used interchangeably with chatbot or voice assistant; however, the more specialized term is "spoken dialogue system." The SDS can be defined as a dialog software system that can communicate with people using voice.¹⁶ SDS includes several NLU technologies, such as speech recognition, NLU, and NLG. Dialogue systems can be broadly classified into 4 categories depending on whether the type of dialogue is open or closed dialogue and whether the dialogue system is based on a retrieval or a generative model.¹⁷ A retrieval model-based dialogue system called closed conversation responds to a specific topic with a premade answer. The pain questionnaire SDS is based on a retrieval model that allows a closed conversation. In many dialogue systems, the user initiates the conversation, and the conversation flow is determined by the user requesting information to the dialogue system.¹⁶ The pain questionnaire SDS in our study has a flow of asking and processing information from the patients as the system takes the initiative in conversation.

This SDS was developed for the purpose of being mounted on a medical assistant robot that provides medical services to the inpatients, especially those undergoing spinal surgery since pre- and postoperative pain assessments in these patients are the most important items for diagnosis and treatment follow-up. Therefore, the conversation flow of the SDS actually followed the pre- and postoperative pain assessments for inpatients with spinal diseases. Although the SDS was developed with a focus on inpatients, it can be sufficiently used for first outpatient visits or remote monitoring due to the general content of the conversation.

In the user satisfaction evaluation of the SDS, there was no statistical difference in satisfaction among the 3 groups, but satisfaction of nurses was slightly higher than that of doctors and patients. In the nurse group, there was no summary errors; hence, the overall accuracy was high, and it is presumed that the expectation for the use of the SDS was reflected in the nurse groups with a high actual workload. On the other hand, it seems that doctors showed relatively low satisfaction because the accuracy of the SDS did not meet their expectations as they require a high level of information accuracy. As for the satisfaction of patients, the mean age was relatively older; hence, unfamiliarity with the digital system may have contributed to the low score. In particular, in item Q1, patients showed significantly lower satisfaction

than medical staff; hence, their understanding of the SDS question may have been low. Therefore, the question content and method should be upgraded to be easier to understand for elderly patients. In the performance evaluation of the SDS, recognition errors in the patient group were significantly more in number. The high error rate may be due to the fact that many unstructured speech recognitions occurred because the patient's answer was long, specific, and varied as a routine expression. In addition, the patient's voice tended to be lower in volume and unclear; hence, the recognition error was likely to be high. On the other hand, due to their prior education for natural conversation, doctors and nurses tended to intentionally give clear and simple answers so that the SDS could recognize the answers themselves. There were cases in which the user could not predict the end time of the utterance of the SDS and answered before the end of the question. Therefore, it is necessary to improve the usability by adding system feedback so that patients can predict the end point of the SDS utterance. Finally, when users answered a question with multiple contents, the SDS recognized only one content. For example, when users answered about the location of the pain, they complained of pain in several locations, including the back, buttocks, and legs. However, the SDS only recognized only one of the 3 pain sites. This is because the SDS fills the slot by selecting only one keyword from the user's answer. Therefore, the SDS should be upgraded to recognize these types of answers.

To improve the overall accuracy of the SDS, it is necessary to significantly improve the current voice recognition technology. Despite the rapid development of voice recognition, the rate of its use is still 80% or less, which is not adequate for medical information that requires high accuracy.¹⁸ In addition, it is important to secure the vocabulary and sentences for patients' expressions through the collection of more dialog sets from actual conversations between medical staff and patients. However, because the doctor-patient dialog is protected by the patient's right to privacy, collection of a large number of dialog sets is challenging, unlike general dialog sets that can be easily obtained from the internet.

Until now, commercialized conversational AI for collecting medical information through voice conversations with patients has not been developed. Conversational AI for collection of medical information can reduce the time and effort needed of medical staff by automating the questionnaire during the first outpatient visit in the future. In addition, it is expected that it will be applied in telemedicine and remote patient monitoring, which is receiving increasing interest due to the recent coronavirus

disease 2019 pandemic. In particular, for older patients, collection of patient outcome reports using text-based chatbots or apps are limited due to presbyopia and difficulty in using smart devices. Therefore, it will be more useful if remote monitoring can be performed using conversational AI in elderly patients. If clinical decisions supporting AI and conversational AI are combined in the future, it could be applied to software in medical devices for diagnosis, treatment, and prevention beyond collecting medical information.⁹

SDS can be used for remote pain monitoring of spinal patients through automation of pain questionnaires for spine patients, and shortening of doctor consultation time through automation of initial consultations. In this case, the collection of pain information can be automated through follow-up of the patient before and after surgery, which can help in tracking the patient's prognosis. By frequently performing additional pain questionnaires as well as pain evaluation during rounds by medical staff, pain evaluation will be possible more frequently while reducing the medical staff's work loading.

A limitation of this study is the small number of test subjects; thus, there may be bias in the evaluation of user satisfaction and performance accuracy. Nevertheless, our study reports the first development of conversational AI for a spinal pain questionnaire. Our study can also provide an important starting point and reference for future related research as our findings validate the accuracy and satisfaction of real patients and medical staff. In the future, we hope to improve the SDS and evaluate user satisfaction and performance accuracy in a large sample of patients.

CONCLUSION

This study is the first report in which voice-based conversational AI was developed for a spinal pain questionnaire that was validated by medical staff and patients. Conversational AI showed favorable results in terms of user satisfaction and performance accuracy. If a large amount of dialogue sets between patients and medical staff are collected and voice recognition technology is improved, it is expected that conversational AI can be used for diagnosis and remote monitoring of various patients as well as help in creating pain questionnaires in the near future.

NOTES

Supplementary Material: Supplementary video clip can be found via <https://doi.org/10.14245/ns.2143080.540>.

Conflict of Interest: The authors have nothing to disclose.

Funding/Support: This study was supported by the Technology Innovation Program (No. 20000515) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Author Contribution: Conceptualization: KHN, DYK, JIL, SSY, BKC, MGK, IHH; **Data curation:** KHN, DYK, DH Kim, JIL, MJK, JYP, JHH, MGK, IHH; **Formal analysis:** KHN, DYK, DH Kim, JH Lee, JIL, MJK, SSY, BKC, MGK, IHH; **Funding acquisition:** MGK, IHH; **Methodology:** KHN, DYK, JH Lee, MJK, JYP, JHH, SSY, BKC, MGK, IHH; **Project administration:** MGK, IHH; **Visualization:** SS Yun, BK Choi, IHH; **Writing - original draft:** KHN, DYK, IHH. **Writing - review & editing:** MGK, IHH.

ORCID

Kyoung Hyup Nam: 0000-0002-3749-4660

Da Young Kim: 0000-0002-3998-3927

Dong Hwan Kim: 0000-0001-8982-7917

Jung Hwan Lee: 0000-0002-1393-7105

Jae Il Lee: 0000-0003-1412-4146

Mi Jeong Kim: 0000-0003-3307-0315

Joo young Park: 0000-0001-5646-0899

Jae Hyun Hwang: 0000-0002-6730-2092

Sang Seok Yun: 0000-0002-0805-0227

Byung Kwan Choi: 0000-0001-5244-4592

Min Gyu Kim: 0000-0003-0744-8812

In Ho Han: 0000-0001-7193-6533

REFERENCES

- Nam KH, Seo I, Kim DH, et al. Machine learning model to predict osteoporotic spine with hounsfield units on lumbar computed tomography. *J Korean Neurosurg Soc* 2019;62:442-9.
- Lee JH, Han IH, Kim DH, et al. Spine computed tomography to magnetic resonance image synthesis using generative adversarial networks: a preliminary study. *J Korean Neurosurg Soc* 2020;63:386-96.
- Tran KA, Kondrashova O, Bradley A, et al. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* 2021;13:152.
- Kim M, Yun J, Cho Y, et al. Deep learning in medical imaging. *Neurospine* 2019;16:657-68.
- Sarker A, Al-Garadi MA, Yang YC, et al. Defining patient-oriented natural language processing: a new paradigm for research and development to facilitate adoption and use by medical experts. *JMIR Med Inform* 2021;9:e18471.
- Tudor Car L, Dhinakaran DA, Kyaw BM, et al. Conversational agents in health care: scoping review and conceptual analysis. *J Med Internet Res* 2020;22:e17158.
- Mairittha T, Mairittha N, Inoue S. Evaluating a spoken dialogue system for recording systems of nursing care. *Sensors (Basel)* 2019;19:3736.
- Shaughnessy AF, Slawson DC, Duggan AP. "Alexa, can you be my family medicine doctor?" The future of family medicine in the coming techno-world. *J Am Board Fam Med* 2021;34:430-4.
- Jadczyk T, Wojakowski W, Tendra M, et al. Artificial intelligence can improve patient management at the time of a pandemic: the role of voice technology. *J Med Internet Res* 2021;23:e22959.
- Piau A, Crissey R, Brechemier D, et al. A smartphone Chatbot application to optimize monitoring of older patients with cancer. *Int J Med Inform* 2019;128:18-23.
- Park L, Cho S. Korean natural language processing in Python. Paper presented at Proceedings of the 26th Annual Conference on Human and Cognitive Language Technology; 2014 Oct 10-11; Chuncheon, Korea. *Human and Language Technology*; 2014.
- García-Queiruga M, Fernández-Oliveira C, Mauríz-Montero MJ, et al. Development of the @Antidotos_bot chatbot tool for poisoning management. *Farm Hosp* 2021;45:180-3.
- Palanica A, Flaschner P, Thommandram A, et al. Physicians' perceptions of Chatbots in health care: cross-sectional web-based survey. *J Med Internet Res* 2019;21:e12887.
- White AA, Lee T, Garrison MM, et al. A randomized trial of voice-generated inpatient progress notes: effects on professional fee billing. *Appl Clin Inform* 2020;11:427-32.
- Joseph J, Moore ZEH, Patton D, et al. The impact of implementing speech recognition technology on the accuracy and efficiency (time to complete) clinical documentation by nurses: a systematic review. *J Clin Nurs* 2020;29:2125-37.
- Jurafsky D, Martin JH. *Speech and language processing*. London: Pearson; 2004.
- Kojouharov S. *Ultimate guide to leveraging NLP & machine learning for your Chatbot [Internet]*. [cited 2016 Sep 18]. Available from: <https://chatbotslife.com/ultimate-guide-to-leveraging-nlp-machine-learning-for-you-chatbot-531ff2d-d870c>.
- Bajorek JP. *Voice recognition still has significant race and gender biases [Internet]*. Brighton (MA): Harvard Business Publishing; c2022 [2019 May 10]. Available from: <https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases>.