International Journal of Social Robotics
https://doi.org/10.1007/s12369-024-01187-1

Design and evaluation of a robot telemedicine system for initial medical examination with UK and Thai doctors

Chatchai Chirapornchai[1] · Paul Bremner[1] · Manuel Giuliani[2] · Faatihah Niyi-Odumosu[3]

Accepted: 16 October 2024

Abstract

This study presents the evaluation of a robot telemedicine system used for initial diagnosis of patient with UK and Thai doctors. In a prior study, a set of user requirements for a robot telemedicine system were outlined based on a focus group with medical professionals and an online survey with the public. Based on these recommendations we have developed a robot telemedicine system that consists of a telepresence interface with examination functionality for doctors to control a robot that is present with patients. The system is designed for initial diagnosis of medical conditions, and Bell's palsy diagnosis is chosen as an example use case. This paper presents the design and usability evaluation of this robot telemedicine system. The usability study was conducted using a thinking-aloud protocol, and a semi-structured interview was conducted after using the system. Study results were analysed using framework analysis. Additionally, participants were asked to give subjective ratings of the system. The study was conducted cross-culturally with doctors from the UK and Thailand, allowing us to gain a deeper understanding of system usability with groups that differ in their medical practice and user needs. Study participants rated the system with 72.19 points on the System Usability Scale, which indicates acceptable usability when interpreted semantically. The system also received a presence score of 0.691, normalized between 0 (least) and 1 (highest). Based on the framework analysis the prior set of user requirements has been refined and developed into a set of design recommendations for use in the development of future robot telemedicine systems. Additionally, study results indicate differences in medical practice between UK and Thai doctors which could be relevant to adjust the robot telemedicine system to local users.

Keywords Robot telemedicine system · Medical examination · Usability study · Thinking-aloud · Framework analysis

1 Introduction and Related Work

The maldistribution of healthcare workers is a global challenge, affecting various regions, such as England and Wales [1], Southern Africa [2], and Southeast Asia [3]. According to the Global Burden of Disease statistics [4], to achieve high levels of universal health coverage (UHC), considerable expansion of the world's health workforce is necessary.

Additionally, while half of the world's population lives in rural areas, they receive medical care from less than a quarter of the available physicians [5]. In UK, the uneven distribution of doctors has been a consistent issue since the establishment of the National Health Service (NHS) [1], despite the continuous rise in the number of doctors over the past few years. In England, during October 2022, there were about 1,700 patients for every doctor in average [6]. This number varied greatly among GP practices, with some serving thousands and others serving hundreds. However, the maldistribution of doctors is unlikely to be solved in the short term due to

Paul Bremner and Manuel Giuliani have contributed equally to this work.

✉ Chatchai Chirapornchai
chatchai2.chirapornchai@live.uwe.ac.uk

Paul Bremner
paul2.bremner@uwe.ac.uk

Manuel Giuliani
manuel.giuliani@hs-kempten.de

Faatihah Niyi-Odumosu
faatihah.niyi-odumosu@uwe.ac.uk

1 Bristol Robotics Laboratory, University of the West of England, Bristol, UK

2 Faculty of Electrical Engineering, Kempten University of Applied Sciences, Kempten, Germany

3 School of Applied Sciences, University of the West of England, Bristol, UK

the structural and demographic characteristics of the underserved area. In addition, UK has a major concern when their trained doctors leave their country to work in other countries such as New Zealand due to a more preferable lifestyle there [7].

This uneven distribution is even more critical in Thailand, a developing country, because of (1) the emigration of medical staff from developing to developed countries [8], and (2) the domestic migration from rural regions to urban centres as it provides better living standards, higher incomes, more social recognition and greater job satisfaction [8] and there is an increasing demand for health care services in urban private hospitals and medical tourism [9]. Health centres in sub-districts of Thailand, also known as Tambon (sub-district) health promotion hospitals (THPH), are typically staffed by 3–5 nurses and paramedics without the presence of doctors due to the shortage of medical doctors in rural areas [10]. However, they have to cover approximately 5000 patients on average [11, 12]. THPHs are the first point of contact for most patients in rural areas as they can travel to these hospitals more easily than to provincial hospitals. Health centres focus on health promotion and disease prevention while providing basic diagnosis for acute and stable chronic diseases [12]. However, due to less medical equipment and a lack of medical doctors, some patients may be referred to provincial hospitals, receiving a delay treatment. In addition, in the case of a pandemic such as COVID-19, medical services were disrupted due to the transportation lockdowns [13].

To ease the problem, one possible solution is to use telemedicine systems to make medical expertise available in rural areas where healthcare facilities are inadequate. Telemedicine systems allow patients attending a THPH to consult with doctors from the provincial hospitals at a distance, making healthcare easier to reach while maintaining patients' satisfaction. It has been shown that remote consultations increase accessibility to health services, and patients spend less time waiting for consultation [14, 15]. Fatehi et al. [16] showed that patients with diabetes who were seen remotely by endocrinologists via video conferencing were generally satisfied with remote consultation, despite the inability to perform physical examination. Anderson et al. [17] reported that remote consultations in primary care is generally well-accepted by patients and can significantly reduce the need for face-to-face appointments. Nevertheless, it can still be further improved by educating patients on the remote consultation process. Moreover, telemedicine systems facilitate social distancing, which is necessary in the case of COVID-19 pandemic [18]. Gilbert et al. [19] found that remote consultations are largely acceptable among patients (90/100) and clinicians (78/100) in response to COVID-19. Similarly, Petrazzuoli et al. [20] also revealed that remote consultation is well appreciated by both healthcare professionals and patients, where telephone consultation is the most common approach. In addition, remote consultation can be useful when patients require secondary-care consultation after a referral from doctors or follow-up consultation for previous treatments [21].

Beyond remote consultations, several existing telemedicine systems utilise robots to perform different types of remote medical examinations. There are systems that can carry out minimal invasive surgery remotely [22, 23], perform ultrasound examination [14, 24], ophthalmology examination [25, 26], neurological examination [27–32], ear and oropharyngeal examinations [33], and physical examinations, including palpation [34], and examinations involving tools, including an otoscope, a stethoscope, and an ultrasound probe [35, 36]. Furthermore, telemedicine has the potential to effectively decrease the likelihood of severe long-term outcomes for patients in intensive care units (ICU). The ERIC (Enhanced Recovery after Intensive Care) project[1] has been launched in 2017 in Berlin, Germany [37, 38] to deal with the risk of secondary damage for patients e.g. post-traumatic stress disorder [39], severe muscle weakness [40] after discharge from the intensive care unit. This damage could restrict patients' quality of life and negatively impact treatment outcomes. With the use of telemedicine for daily telemedical visits, it can reduce the risk of long-term consequential damage for intensive care patients, and therefore, has been recommended by the German State for wide introduction in all parts of Germany [41]. With the promising result, it highlighted the importance of patient care in the clinical scenario. However, to the best of our knowledge, there is a paucity of research on the use of social interaction to increase the quality of patient care before and after the doctor connects to the system for the medical examination. Although this is not the focus of this paper, it is the reason why our robot telemedicine system is proposed to automatically initiate patient interactions and routine diagnostic tasks to maintain patient care and reduce the workload of medically trained personnel.

1.1 Use Case Description

The works cited above show that soon telemedicine robots might be used more widely. It is crucial to consider how these robots can maintain patient care, ensure patient safety, and provide essential information or guidance. We aim to develop our robot telemedicine system to reduce the workload of medically trained personnel in health centres and to decrease the time it takes for patients to receive treatment. The scenario of how our robot telemedicine system is envisioned to be used is summarized as follows:

[1] https://www.eric-projekt.net/eric-projekt/

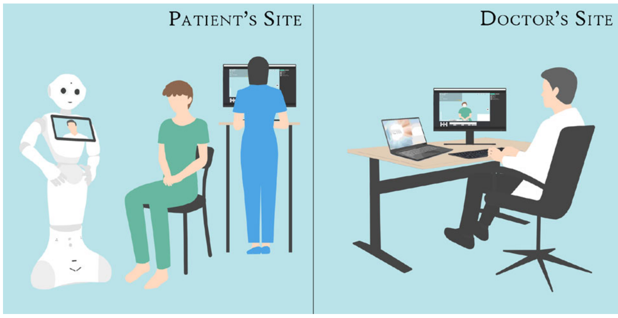

Fig. 1 Robot telemedicine system: patient's site (left) and doctor's site (right). Patient's site refers to the health centre where doctors are unavailable and that a social robot and an assistant help the patient connect to a doctor in an urban centre

1. The patient walks into the health centre in a rural area and is greeted by a medically trained member of staff (who we will call the "assistant" in the following lines).
2. The assistant brings the patient to the robot telemedicine system, which is located in one of the clinic rooms.
3. The assistant sets up the robot telemedicine system, explains to the patient what it can do, and leaves the room.
4. The robot telemedicine system socially interacts with the patient while establishing a remote connection to a doctor in an urban centre. This interaction may include a greeting and obtaining patient medical information. Please note that the investigation on how the robot should behave when the doctor is absent is not in the scope of this paper and will be future work.
5. Once connected to the robot telemedicine system, the doctor executes a remote examination in which the system demonstrates diagnostic movements and is used as part of diagnostic tests (See Sect. 2.7.5). It also provides video conferencing and additional medical information via the telepresence interface. This enables the doctor to properly diagnose the patient and to decide about their required care.
6. Once the doctor disconnects, the robot telemedicine system continues the social interaction with the patient. The interaction may include a reminder of what the doctor has said, scheduling of follow-up appointments, and accompanying the patient back to the assistant. Again, this is outside the scope of this paper.
7. The assistant takes the final steps, e. g. handing out medication, and the patient leaves the health centre.

Figure 1 shows a graphical representation of the system used in the proposed scenario. On the doctor's site (right), the doctor uses a computer with a graphical user interface to examine the patients using different functionalities on the robot (See Sect. 2.7). On the patient's site (left), a social robot is used to display the video call with the doctor on a tablet mounted on the robot. The additional depicted person on the patient's site is the assistant responsible for the technical setup. A social robot has been used for our system due to planned future work that will involve social interaction with the robot to act as an interaction partner for the patient before and after the doctor connects to the system.

1.2 Aims and Objectives

A key feature of our work not present in the existing literature is the use of a user-centred design (UCD) process that involves the end user throughout system design and development [42]. UCD includes three iterative phases: requirement analysis, system design and development, and user-centred evaluation. The aim of this process is to take into account feedback from future end users throughout system design and implementation.

For the presented robot telemedicine system, our prior work developed user requirements for three involved user groups, doctors, patients, and assistants [43], as part of the requirement analysis (Sect. 2.1). Based upon these requirements, this paper presents the system design of our robot telemedicine system (Sect. 2), and the results of a usability study that involved Thai and UK doctors for user-centred evaluation (Sect. 3). Bell's palsy diagnosis is chosen as an example use case as it allows the evaluation of AR diagnostic information which can be obtained through image processing with existing pieces of diagnostic technology. In the future, AR diagnostic information could be generalised to other conditions where the diagnostic information is difficult to perceive over just video conferencing.

The aims of the study are to evaluate system interface design, usefulness, performance, flexibility and documentation capabilities from the doctors' perspective. In addition, the study results give insights into cross-cultural differences between UK and Thai doctors when using a robot telemedicine system. This allows us to evaluate the utility of the system in both cultures to understand how the system should be focused and developed in the future for various working environments (Sect. 5.1). Finally, results were used to refine the system design recommendations (Sect. 5).

Please note that this work does not collect feedback from the other two user groups, patients and assistants. This will be the subject of upcoming user studies.

2 System Design

This section explains how the interface of the robot telemedicine system containing different functionalities is designed based on the requirement analysis from our prior work.

2.1 User Requirements

Our prior work in [43] presented the first step of the UCD process—requirement analysis—where we established user requirements for the robot telemedicine system. To do so, a focus group with Thai medical staff and an online survey with potential patients has been conducted virtually. By using framework analysis [44], user requirements of three user groups–doctors (D), patients (P), and assistants (A)—were identified as listed below.

D1 Doctors want to communicate with the patient in real time. A chat system is also needed in case of a poor internet connection.
D2 Doctors need the medical information[2] to be displayed on desktop augmented reality (AR) during the entire examination.
D3 Doctors want the assistant to navigate the robot to a safe place, monitor the robot and other equipment, introduce the patient to doctors and vice versa, and take over other supportive tasks for doctors.
D4 To create a good first impression, doctors need robots to interact with the patient in a socially acceptable way.
D5 Doctors need a display on the robot to have the patient follow the examination or to log data.
P1 Patients need the doctor to be present all the time, and therefore, real-time communication is necessary.
P2 Patients need the assistant to be nearby all the time.
P3 Patients need a secure data connection to secure their privacy.
A1 Assistants need an interface to navigate the robot.
A2 Assistants need system functionality to transfer additional data to doctors, e.g., olfactory information.

2.2 Design Requirements

To formulate the design requirements (**R**) below, we combined related user requirements **D1** and **P1** into **R1**, and **D3** and **A1** into **R4**, split requirement **D1** into two parts **R1** and **R2**, and omitted requirements (**P2** and **D4**). We omitted **P2**, because it is not related to technical system development. **D4** is also omitted as it involves social interaction with the robot which is outside the scope of this paper and will be addressed in future work. The rest of design requirements **R3**, **R5**, **R6**, **R7** are derived from **D2**, **D5**, **P3** and **A2**, respectively. This led to the following set of design requirements for the telemedicine system:

R1 The system should have video conferencing capabilities for real-time communication.
R2 The system should have a chat system in case of poor internet connection.
R3 Augmented reality (AR) should be used as a UI paradigm for providing and displaying medical information to the doctor.
R4 The system should provide a navigation interface for doctors and assistants to move the robot.
R5 A screen on the robot is required for the doctor to display information to the patient, and for the patient to see the doctor.
R6 A secure data connection between doctors and patients is required to protect the patient's privacy.
R7 The system should provide a feature for assistants to transfer additional data exceeding the system's capabilities to doctors, e.g. olfactory information.

2.3 Robotics Platform

Pepper[3], a humanoid robot, has been used as the platform for our telemedicine system. Pepper has been used in medical research before, e.g., [45, 46]. Its form and function is suited for social interaction as described in the use case description in Sect. 1.1. The robot has a tablet mounted on the upper body that can be used to display information (**R5**). Pepper's sensors and its control software also enhance safe human-robot interaction.

2.4 Telecommunication Setup

Figure 2 shows an overview of the telecommunication setup for the robot telemedicine system. The system requires two computers, one at the doctor's site and one at the patient's site. The doctor's computer connects to the platform computer using TeamViewer. The platform computer runs Unity and the UI. It connects to the Pepper robot using ROS, due to the robot's software using Python and Unity using C# as programming languages. An MS Teams session between the doctor and patient is established between the doctor's computer and the Android tablet mounted on the robot. This telecommunication setup allowed the Thai and UK doctors to use their personal computers to connect to the telemedicine system in the usability evaluation (Sect. 3).

MS Teams has been chosen as a video conferencing tool (**R1**) with a built-in chat function (**R2**). To run MS Teams on Pepper, the robot's tablet had to be replaced by a 10.1-inch Samsung Galaxy Tab A (2016 version), running Android version Oreo 8.1.0. To obtain a high video resolution, the robot has furthermore been fitted with a GoPro Hero 8 camera

[2] This requirement has been updated by being slightly modified to suit broader circumstances, specifically from *vital signs* to *medical information*

[3] aldebaran.com/en/pepper/

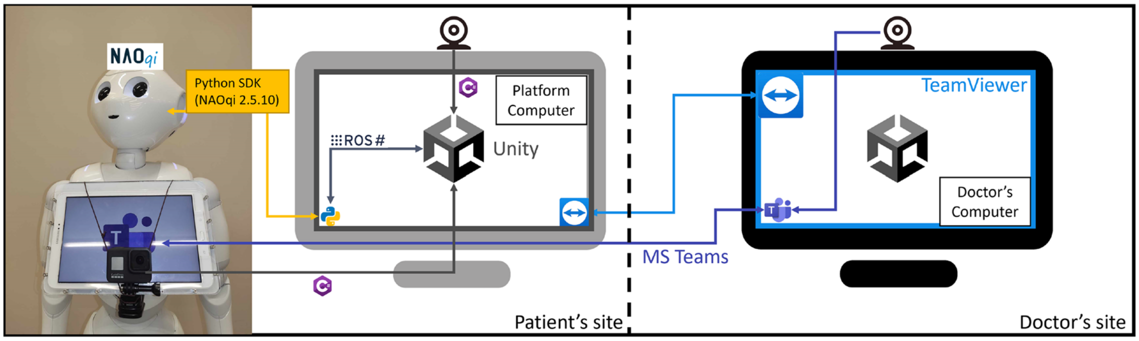

Fig. 2 A brief overview of the telecommunication setup inside the robot telemedicine system

that supports wide-angle video. Figure 2 shows on the left how the tablet and camera were mounted on the robot.

2.5 Robot Teleoperation User Interface

A user interface (UI) has been developed in Unity[4]. It enables doctors and assistants to access all functionality of the telemedicine system. Figure 3 shows a screenshot of the UI. The UI's main window shows the face of the patient overlaid with additional medical information (**R3**). On the right panel, doctors can click the button to start the examination. When the examination starts, the medical examination functions appear in a dropdown menu (Sect. 2.7). Additionally, this panel is used to enable/disable the assistant's camera, the robot's navigation control, patient information, a facial palsy detection feature (Sect. 2.7.1) and put the robot to rest. On the bottom panel of the UI, doctors can use button arrows to navigate the robot (**R4**) and to control the text-to-speech functionality of the robot. In addition, doctors can capture photos. The camera zoom can be adjusted by using a slider on the right of the display or a mouse scroll. The camera can be panned left/right or up/down by two sliders on the bottom-left of the display (or mouse right-click and drag).

2.6 Patient Information Management

This function is used by doctors for recording and displaying patient information. Additionally, assistants can use this feature to transfer additional data, e.g., olfactory information (**R7**). Patient information is displayed on the top-left in the UI (Figure 3). It can be temporarily disabled by clicking the *'Patient Information'* toggle on the right panel of the screen. New patient information can be added by clicking the '+' button on the bottom. Once clicked, doctors can put down the patient's name, age and medical history in the provided text fields. After recording information of a new patient, their name will appear in the dropdown box next to the '+' button. The *'Edit'* button can be used to change patient information. Doctors can also remove a patient record by clicking the '-' button.

[4] unity.com/

2.7 Medical Examination Functionality

One important design requirement for the robot telemedicine system is to display medical information during examination using AR (**R3**). In order to demonstrate functionality for this, Bell's palsy was chosen as an example application in which AR can aid remote diagnosis by displaying diagnostic information that is challenging to perceive over a video system in remote neurological examinations.

The aim here is to use this application as a proof of concept so that its usability can be analysed, with the intention that the findings will apply to future applications of AR diagnostic information for other conditions. Bell's palsy is an acute facial nerve paralysis, usually affects one side of the face, in which the cause of Bell's palsy is unknown. It is the most common cause of acute facial nerve paralysis. To diagnose Bell's palsy, a diagnosis of exclusion has to be implemented to distinguish whether patients have suffered from Bell's palsy or from other types of facial weakness, for example, a stroke. A stroke occurs due to the blockage of a blood vessel in the brain, causing an inability to move one side of the body in which the symptoms usually become most severe in a matter of seconds to minutes. A diagnosis of exclusion is important because mistaking Bell's palsy for a stroke can lead to wrong or delayed treatment, wasted healthcare resources, and significant emotional stress for patients and their families [47].

The following paragraphs explain the medical background for each of the examinations featured on our robot telemedicine system together with details how each technique was implemented.

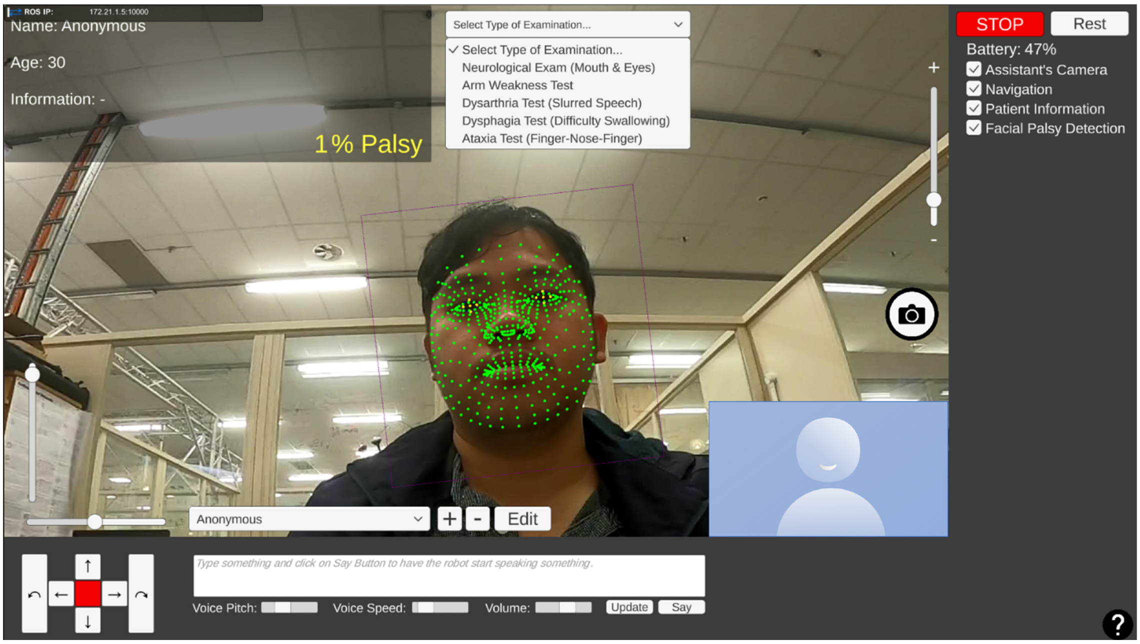

Fig. 3 Robot teleoperation user interface designed in Unity

2.7.1 Neurological examination

The purpose of neurological examination is to determine if the facial weakness of a patient is caused by a peripheral (Bell's palsy) or central lesion (stroke). In general practice, doctors observe for facial asymmetry by paying attention to the nasolabial fold and corners of the mouth and have the patient smile. Then, they ask patients to raise both eyebrows or wrinkle their forehead, close their eyes tightly, and purse their lips. Mouth weakness will be present in both central and peripheral facial palsies, while asymmetry in eyelid closure and forehead wrinkles is a sign of peripheral facial palsy.

In this research, it is assumed that a healthy face shows high symmetry, regardless of the expression. If there's any asymmetry, it is presumed to have some degree of facial paralysis. An effective existing methodology [48] has been utilised to train a classifier to determine facial paralysis. To train the classifier, images were sourced from the YouTube Facial Palsy database [49], representing patients with facial weakness, and the Extended Cohn-Kanade database [50] for healthy people. Images were preprocessed for classification by extraction of facial landmarks using Google MediaPipe[5], which is displayed using AR as a green face mesh (shown in Figure 3). The processed image data was then sent to the classifier to train and extract meaningful information from both the entire face (yellow text in Figure 3) and specific regions at the mouth and eyes (not visible in the UI). Note that the system has not yet implemented facial landmarks around the forehead, this will be investigated in future work. Assessing facial palsy regions from the entire face can help doctors identify whether patients have facial weakness, whereas evaluating specific regions can differentiate the cause of facial weakness. While eye palsy is a sign of peripheral facial nerve palsy, mouth palsy can be found in both peripheral and central lesions.

2.7.2 Arm Weakness Test

This test aims to assess a patient's arm strength, in order to detect abnormal arm weakness. In practice, doctors assess patient's arm muscle strength by asking patients to lift their arms and compare to a scale by the Medical Research Council that grades muscle strength on a scale of 0 to 5. The scale ranges from Grade 0 ("no muscle contraction") to Grade 5 ("normal muscle power"). In general, if the patient's arm muscle strength is less than Grade 3 ("movement against gravity but not resistance"), this is an indicator of a stroke, as the territory involved includes more than just the facial nerve; otherwise, it is most likely Bell's palsy. Therefore, the purpose of this test is for patients to effortlessly follow the instructions by looking at the robot's posture.

Doctors can start the test to initiate the robot's posture of lifting arms, observe how patients lift their arms and later end the robot's posture manually. At the same time, the robot verbally prompts patients to stop lifting their arms.

[5] developers.google.com/mediapipe

2.7.3 Dysphagia (Difficulty Swallowing) Test

Dysphagia is a group of symptoms where a person has difficulty eating, chewing, swallowing and transporting food from the mouth to the stomach. It can be painful and in some cases, swallowing is impossible.

In the robot telemedicine system, doctors can instruct the robot to verbally prompt patients to drink water while observing any signals e.g. coughing. Patients with those symptoms present are suspected to have a stroke, as dysphagia does not show in Bell's palsy.

2.7.4 Dysarthria (Slurred Speech) Test

A Dysarthria test intends to determine whether patients have difficulty speaking due to muscle weakness. It is often caused by damage in the central or peripheral nervous system. The damage makes it difficult to control the tongue or larynx (voice box), causing patients to slur words. Doctors can test patients by asking them to speak certain words or sentences.

In our robot telemedicine system, doctors can use a function to generate words for the robot to speak and ask patients to repeat after. Examples of the pre-defined words are "tip-top", "fifty-fifty", etc, which can be optionally repeated again if needed.

2.7.5 Ataxia (Incoordination) Test

Ataxia can be caused by damage to the brainstem or cerebellum, resulting in difficulty with walking and balance, hand and eye coordination, etc. In this examination, we aim to find out if patients have a lack of coordination between visual inputs and hand movements by implementing a finger-to-nose test. In practice, a finger-to-nose test is performed by asking patients to touch their nose with their forefinger and touch the doctor's finger. After that, the doctor moves their finger to other positions, and patients repeat the process.

In the robot telemedicine system, instead of using the doctor's finger, doctors can move the robot arms to several pre-defined locations and ask patients to repeatedly touch their nose and the robot's finger. With the wide-angle video from a GoPro camera, doctors can see the robot arms and observe how patients perform the finger-to-nose test aided by the robot.

3 System Evaluation with Doctors

This section describes the main objective of this paper which is a usability evaluation of the robot telemedicine system with doctors in a remote medical examination. The study was carried out in accordance with the guidelines and regulations of, and was fully approved by, the University of the West of England Research Ethics Committee (reference number FET.2122.132). Additionally, a decision tool by the UK National Health Service (NHS) Research Ethics Committee (REC)[6] was used to establish that NHS REC approval is not needed for this usability study.

3.1 Methodology

This evaluation aims to uncover usability issues in the robot telemedicine system, improve UI design, and refine the previously specified design requirements. Thus, *usability testing* has been chosen as the paradigm for this user evaluation. Considering the nature of the system, remote usability testing has been implemented. It differs from traditional usability testing only in the fact that the experimenter and participants did not meet in person but over MS Teams. During the test, participants were asked to use the thinking-aloud methodology [51]. Thinking-aloud requires participants to actively voice their thoughts while performing tasks with a new user interface. In the presented usability test, participants were encouraged to talk about what they saw, their thoughts and feelings as they examined different aspects of the system and performed their tasks. This allowed the experimenter to record participants' feedback and hear first-hand about any usability problems.

In the experiment, the participants interacted with a co-author in the same manner that they would do with a real patient via the robot. The only difference here is that they interacted with a healthy person rather than a patient. However, during the task, we also put an image of a patient with facial weakness in front of the co-author for participants to see how the system responded to a patient with facial weakness. For this study, no assistant was used as they are principally involved in setting the patient up to use the system and this was not needed here.

3.2 Study Procedure

Usability testing sessions took approximately 2 hours per participant. The following user study procedure was employed: **(1)** Participants were sent a participant information sheet and research privacy notice before the session. Participants were also asked to download and install TeamViewer and MS Teams onto their computers. **(2)** Upon connection on MS Teams, the experimenter introduced himself to the participant. Informal conversation occurred to place the participants at ease and to explain the objective of the study. Participants were allowed to ask further questions before giving written consent. **(3)** Participants were given the link to an online questionnaire delivered on Qualtrics[7]. The question-

[6] hra-decisiontools.org.uk/ethics/

[7] questionnaires template can be assessed by at this link.

naire contained the participant information sheet, a consent form, and collected demographic information. Demographic information taken was age, gender, and years of experience as a medical practitioner. Participants were also asked to rate their experience with remote medical examination technology on a 1-to-10 scale. **(4)** Participants were asked to download and install TeamViewer, and to create an account if they had not done it before the session. Then, the experimenter gave instructions to the participants to establish a connection to the platform computer through TeamViewer. It was also tested whether doctors can access the UI on the platform computer and hear the robot's voice. **(5)** Participants were systematically guided through each feature of the robot telemedicine system, i.e., navigation and camera view, patient data management, and medical examination functionality as explained in Sect. 2. **(6)** As the final step before usability testing, participants were instructed on how to use the thinking-aloud method, demonstrated in a short video tutorial. **(7)** Participants were asked to perform a series of system tasks in the following order:

- Navigate the robot and adjust the camera zoom and field of view to see the patient.
- Capture images of the patient and record new patient data.
- Edit the patient data. For Thai GPs, record the data in Thai.
- Use all types of medical examination functions in a self-chosen order. While using the functions, doctors were asked to toggle facial palsy detection and use a text-to-speech module if needed.
- Rest the robot for 10 seconds and wake the robot up.

These tasks were selected in order to simulate the real scenario of a remote consultation. In case of a technical error with the system, the experimenter did explained the error to the participants, recorded their feedback, and noted down the issue for future development. **(8)** After finishing the usability study, participants completed a post-study questionnaire and were asked a set of questions in a semi-structured interview (Sect. 3.4). **(9)** During a study debriefing, participants were allowed to ask additional questions. **(10)** After the study, participants were given either a £15 Amazon (UK) or Starbucks (Thailand) voucher as a token of appreciation.

3.3 Participant Recruitment

The user study was undertaken with GPs from Thailand and the UK. Additionally, a UK-based neurologist consultant was recruited. Thai GPs were recruited from Banphaeo General Hospital by social media. UK GPs were recruited from Derbyshire Healthcare NHS Foundation Trust and NHS Bolton Clinical Commissioning Group (CCG) via email. The UK neurologist was recruited from Coventry and Warwickshire Partnership trusts. In total, four Thai GPs, four UK GPs, and one UK neurologist consultant (male = 7, female = 2) took part in the study. Due to the inherent difficulty of getting hold of medical personnel who work in shifts and have busy schedules, we were unable to recruit additional participants. However, according to [52] during usability testing about 80% of usability problems can be covered within five participants. Thus, we argue that the number of participants is sufficient for our study.

The UK neurologist consultant had to be excluded from results analysis due to his workplace policy that does not allow any software installation, Teamviewer to be specific. However, the experimenter gave a verbal walkthrough of the robot telemedicine system to the consultant and recorded his comments. These will be used to support the discussion in Sect. 5.

3.4 Measures

Questionnaires were used to evaluate the subjective experience of the participants during the test. The participants were asked to complete the questionnaires immediately at the end of the experimental tasks, in the order described below.

System Usability Scale The system usability scale (SUS) is a standardized metric, developed by Brooke [53], to provide a reliable tool for measuring the usability of a system. It consists of 10 items rated on a 5-point Likert scale.

Presence Questionnaire The presence questionnaire (PQ) (v3.0), developed by Witmer and Singer [54], is a commonly used questionnaire in VR and AR research [55]. PQ is robust and reliable to analyze the sense of presence [56]. It was revised by [57] and consists of 24 items rated on 7-point Likert scale. PQ is divided into seven domains: realism, possibility to act, quality of interface, possibility to examine, self-evaluation of performance, haptics, and sounds. Since the robot telemedicine system does not involve haptics, only 22 items in six domains were used for this evaluation.

Semi-structured Interview At the end of the usability test participants were asked a set of questions about their opinions on all system features and how they can be improved. They were also asked to explain the strengths, weaknesses and usefulness of the overall system, and how the system could be used in their work.

3.5 Framework Analysis

In addition to the experimenters' first-hand experience, all usability test sessions were recorded using MS Teams for further analysis. Transcriptions of these recordings from when participants performed the thinking-aloud method and from the answers given to the semi-structured interview were analysed using framework analysis [44]. Framework analysis has been widely applied in health research as it is a systematic and flexible way to analyze qualitative data [58]. The five main steps of framework analysis and how they have been applied in this work are described in this section.

Familiarisation In this step, one member of the team familiarised himself with the data by listening to the recorded sessions twice. Then, we manually transcribed participant speech that contained their comments related to their use of the system and its performance and omitted irrelevant speech from the transcription.

Framework Identification After data familiarisation, five categories were selected as the main themes to form a working analytical framework, consisting of *interface design (I)*, *usefulness (U)*, *system performance (P)*, *flexibility (F)* and *help and documentation (H)*. These themes are chosen based on the commonly talked about themes found in the session recordings.

Indexing In the indexing step, each statement made by participants is allocated to one of the five main themes. In this user study, a spreadsheet was used to record a theme for each sentence using the theme abbreviation as a code and a '+' or '−' symbol to indicate if the sentence was a positive or negative comment respectively.

Charting and Summarizing Once the themes have been identified and indexed, the data is rearranged according to system features, so that the number of comments relating to each feature can be counted. That is to say that when multiple participants provide comments on a particular system feature, they are added to a count of comments related to that feature.

Mapping and Interpretation In the final step, the count for each data/comment is mapped onto each word code/index (See Table 4).

4 Results

4.1 Participants

Thai doctors all reported their age bracket as 18–35, UK doctors were all in the 36–55 age bracket. Participants rated their familiarity with remote examination technology relatively low at 3.75 ($SD = 2.44$) on average (out of 10). Table 1 reports detailed demographics.

4.2 Questionnaires Data

In the post-study questionnaires (bottom of Table 1), the overall system received a SUS score of 72.19 which is considered as acceptable [59]. Thai doctors gave lower SUS scores than UK doctors. For the presence scores, the system received a PQ score of 0.691, normalized between 0 (least) and 1 (highest sense of presence). UK doctors gave lower presence scores than Thai doctors. Among the presence subscales (Table 2), participants rated the interface quality least and possibility to examine most favorably compared to the remainder of the PQ items.

4.3 Framework Analysis

Table 4 shows a summary of positive and negative comments made by participants for each system feature, ordered by the main themes discussed in Table 3. In total, there are 254 comments, sorted into 91 positive and 163 negative comments. These comments address 117 unique usability issues, some of which were mentioned more than once. The feature that received the most negative comments is *Patient Data Management* (38), whereas *Camera View and Navigation* received the most positive feedback (22). Among the five criteria, *Performance* received the most comments, both positive (53) and negative (71).

4.4 Participant Quotes

In order to discuss user feedback on each system feature, we have summarised comments made and where appropriate provided quotes from study participants. Quotes by UK doctors U1 to U4 are given verbatim, and quotes by Thai doctors T1 to T4 have been directly translated into English.

4.4.1 Overall UI, Appearance and Speech

For the overall UI of the system, six doctors found the system to be easy to use, despite their non-IT background, due to the straightforward and self-explanatory interface. The users found the robot's speech accent easy to understand and human-like, and the appearance of the robot friendly.

[T2]: *The interface is straightforward and the movement of the robot seemed natural to me. The voice is also human-like and the face of Pepper is friendly.* (Q1)

[U2]: *I think it was easy enough for me. It's a bit complicated in the beginning, during installing TeamViewer. But when I got the hang of it, I was able to figure out what to use to control the robot, which was easy.* (Q2)

However, at the same time, one doctor had a different opinion about the robot's appearance.

[U4]: *The appearance of the robot seemed a bit too scary, may be because I'm not get used to seeing any kinds of robot [sic]. The white colour is fine, but the structure makes me scared [sic].* (Q3)

For the negative comments, one doctor discovered the failure of the interface to display certain Thai words properly. In addition, many participants thought the text-to-speech

Table 1 Demographics and the corresponding scores from each participant

	Thai T1	T2	T3	T4	Total	UK U1	U2	U3	U4	Total	Total
Age	18–35	18–35	18–35	18–35		36–55	36–55	36–55	36–55		
Gender	Male	Male	Male	Male		Male	Female	Male	Female		
Years	1–2	3–5	1–2	2–3		< 1	5–10	10–25	2–3		
RT	3	5	5	4	4.25	0	0	6	7	3.25	***3.75***
SAF	0	4	4	4	3.25	6	2	0	7	3.75	3.375
RPM	6	4	4	0	3.5	0	0	0	0	0	1.75
SUS score	77.5	57.5	47.5	80	65.63	90	70	80	75	78.75	***72.19***
PQ score	0.682	0.753	0.636	0.747	0.705	0.747	0.604	0.695	0.662	0.677	***0.691***

T1 to T4 represents four Thai GPs and U1 to U4 represents four UK GPs
RT, SAF, RPM = Experience with Real-time, Store-and-forward, and Remote Patient Monitoring

Table 2 Means of presence scores and their subscales

Subscales	Thai		UK		Total	
	Mean	N. mean	Mean	N. mean	Mean	N. mean
Realism (R)	34.5	0.704	33	0.673	33.75	0.689
Possibility to act (PA)	20.25	0.723	19.5	0.696	19.88	0.71
Interface quality (IQ)	14.75	0.702	11.75	0.560	13.25	0.631
Possibility to examine (PE)	14.5	0.69	15.5	0.738	15	0.714
Self-evaluation of performance (SP)	9.5	0.679	9.75	0.696	9.63	0.688
Audio fidelity (AF)	15	0.714	14.75	0.702	14.88	0.708
Total PQ score	108.5	0.705	104.25	0.677	106.38	0.691

*N. mean = Normalized mean

function unnecessary to use as they could speak to patients directly.

[T1]: *The speed and volume of the robot are already good. The speech that the robot used during instructions is good but the text-to-speech feature is unnecessary. However, it might be good to have in case we have sets of pre-defined sentences and want to use it when we are tired speaking.* (Q4)

In contrast, one UK participant thought this function has the potential to be used as a translation service, e.g., for UK GPs to examine patients who are non-native English speakers.

[U2]: *This function can be helpful because sometimes you may speak to somebody but they might not hear me very well due to a long-distance connection, but by typing it there, the robot says it directly to the patient, especially if there is a way for the robot saying what I have typed in their own language.* (Q5)

4.4.2 Camera View and Navigation

Five participants found it difficult to navigate the robot through the 2D camera. One of the reasons reported for this is because the setup of the camera was too low. It was also reported that the lack of depth information causes difficulties in distance estimation.

[T3]: *Does this bump? I feel like it is quite difficult to estimate the distance between the patient and the robot. I'm not sure if there should be some distance indication signs.* (Q6)

[T2] *At the moment, I think the robot and patient are not on the same level. I feel like I'm looking from a low angle. The robot is supposed to be at the same height as the patient. Let's get a little bit closer. I also feel that I could not estimate the distance well, like there isn't any map showing the current position of the robot or telling how many metres are there between the distance.* (Q7)

Six out of eight doctors experienced video feed delays during navigation. This could result in a low score of PQ subscale - Interface Quality (IQ). Nevertheless, the robot's slow moving speed allowed participants to adapt to the delay.

[T4]: *Slide, stop. Is it in the right position? Hmm, it is not quite following. I feel some slight delays.* (Q8)

Table 3 Definitions of the criteria used for system evaluation

Criteria	Descriptions
Interface design	Comments about the distribution of visual elements e.g. Layout & Positioning, Shape & Size, Colour, etc. to achieve users' goals
Usefulness	Comments about the need of the elements
Performance	Comments about the system's aspects that allow users to act and explore in a natural manner and not restrict normal physical actions
Flexibility	Comments about sufficiency in features and shortcuts to do certain actions to facilitate ease of use
Documentation	Comments about guidelines for helping users achieve certain action

[U4]: *I think I notice some slight delays but it can still be adaptable as the moving speed is slow.* (Q9)

However, three doctors stated that it would be better if there were an option to adjust the robot speed to travel larger distances faster.

[T1]: *I think the speed is okay, not too fast or too slow. Maybe because the room size is not that big, so I didn't feel the robot is moving too slowly. However, if the distance is larger, it would be better to have an option to adjust the speed.* (Q10)

Participants reported that the use of colouring the navigation buttons while the robot is moving is a useful feature.

[T3]: *It is quite obvious in UI that when pressing the button, the colour changes to green, indicating that the robot is moving.* (Q11)

Participants were satisfied with the high resolution of the GoPro camera, resulting in high score of PQ subscale - PE.

[U1]: *The resolution is very good and clear for me. I could easily do an assessment remotely in this way.* (Q12)

[U3]: *Yeah, I think the resolution is brilliant.* (Q13)

However, the GoPro camera's wide-angle lens causes a distorted video image. Thus, one Thai participant suggested having options to switch between lens modes so that an undistorted view can be used during robot navigation, and a wide-angle view during the examination.

4.4.3 Patient Data Management

All participants commented that the patient data management function requires more fields to record the demographics and medical information of the patients as the current system can only record patients' names, ages, and general information. In real scenarios, all information such as chief complaint, present illness, allergy, etc. and the result obtained from the examination should be recorded separately.

[U1]: *Usually when you work, you need to be quite efficient with time. So it has to be easy to read by separating the patients' information into different categories. And also, the examination findings, is it gonna [sic] be integrated into the notes automatically or do I need to type? In the latter case, it should be recorded separately as well. So, that needs to go somewhere.* (Q14)

In addition, all Thai doctors prefer to record patient information in the main UI window rather than adding and editing information using the provided buttons as they found the task to be time-consuming.

[T4]: *About the patient recording system, it's like I have to press a number of buttons to show the window and type to record, right? It may be a bit too challenging as I have to do two things at a time. It would be easier for me to work with patient information in the main UI window.* (Q15)

One participant suggested using scoring systems to capture patient conditions. Another suggested having a search function and commonly searched-for keywords for medication and examination types.

4.4.4 Neurological Examination

All participants reported that facial palsy detection has high accuracy in the detection of mouth palsy, but low accuracy in the detection of eye palsy. Specifically, it was commented that the number indicating detection accuracy would change too frequently.

[U3]: *I think the fact that you can get the percentage of palsy eyes and mouth is brilliant. It's just the number of palsy eyes is not stable, compared to palsy mouth.* (Q16)

Participants commented that the correct value for eye palsy detection was only visible when the patient looked straight at the camera without moving their eyes and face. To eliminate false positives, the asymmetry feature should not include the landmarks at the eyeballs since the extraocular muscles (EOM), muscles for rolling eyeballs are not affected by Bell's palsy.

Table 4 Summary of positive and negative comments for each feature in the robot telemedicine system, clusters of comments highlighted in bold font

	Overall UI, appearance & speech		Camera view & navigation		Patient data management		Neurological examination		Arm weakness test		Dysphagia test		Dysarthria test		Ataxia test		Total	
	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
Interface design	***10***	4	2	2	2	6	0	2	0	1	0	0	0	0	0	1	14	16
Usefulness	3	7	1	2	0	0	4	2	2	0	0	9	5	3	2	0	17	23
Performance	8	8	***25***	***24***	0	9	6	***11***	4	***14***	0	0	7	0	3	5	***53***	***71***
Flexibility	0	3	4	8	0	***22***	0	5	0	3	0	0	0	1	2	4	6	***46***
Docum entation	1	2	0	1	0	1	0	0	0	0	0	0	0	0	0	3	1	7
Total	22	24	32	27	2	38	10	20	6	18	0	9	12	4	7	13	91	163

*N. mean = Normalized mean

[T1]: *I was curious about something. What happens if you glance sideways without moving your head, does the percentage of eye palsy increase? Normally, the number should not go up since the ability to roll the eyes is caused by EOM, and it has nothing to do with facial palsy.* (Q17)

It is recommended that the robot should also be able to detect asymmetry in forehead wrinkles.

[U2]: *Right now, the system detects asymmetry in the eyeballs and that makes the accuracy goes [sic] wrong. In fact, what can distinguish stroke and facial palsy is forehead muscle movement. If patients cannot wrinkle their forehead, that is not a stroke, but it is facial nerve palsy.* (Q18)

4.4.5 Arm Weakness Test

Five participants faced a problem with the robot not properly lifting its arms when executing the arm weakness assessment function. The reason for the malfunction was the built-in collision avoidance function of the robot detecting the experimenter's hands as obstacles.

In terms of usability, the arm weakness test was partially useful because the functionality of the robot only allowed participants to assess the arm muscle strength up to Grade 3 (against gravity). To assess higher grades, the participants recommended involving the assistant. One mentioned solution was to hand out weights to the patient when administering the test.

[U3]: *I think this test is alright but I think there is room for improvement. If the patient has the muscle power 4 out of 5 [sic], this may not detect it [sic] because the patient is not fully weak. Having the patient copy the movements will not be enough. Maybe if the assistant can help me assess in a higher grade by asking the patient to lift a dumbbell.* (Q19)

In addition, all Thai doctors prefer to have a timer to help them count time while lifting their arms.

[T4]: *About the robot's dialogue, the 10-second interval is exactly what I use during the weakness test, but do I have to count the time by myself? I thought there would be a timer showing up. It would be better to have a timer for me to refer to.* (Q20)

4.4.6 Dysphagia Test

All participants found the dysphagia feature to be impractical and useless as they can ask patients during history taking whether they find it difficult to swallow food or liquids, and notice any coughing while swallowing.

[T3]: *I think dysphagia can only be found by history taking. It is very difficult to find out by examining, only when dysphagia happens in the upper parts of the body, which is not always true.* (Q21)

One UK doctor mentioned that most stroke patients do not experience any difficulty swallowing, only those whose conditions are severe have trouble swallowing.

[U2]: *Difficulty in swallowing can be one of the symptoms of stroke, but usually that happens only in a massive stroke.* (Q22)

In addition, one Thai doctor commented that in normal practice, they do not ask patients to swallow anything as it is unsafe to force patients with severe dysphagia to swallow food or liquids.

[T1]: *I don't think this would be practical. Well, in today's practice. I wouldn't go through all these steps. Normally, I would do history taking to find whether they have dysphagia or not. In fact, I would not even allow them to swallow as it may worsen the pain.* (Q23)

4.4.7 Dysarthria Test

The comments about this feature are split. Three doctors thought the dysarthria examination was not useful as they could tell whether patients had slurred speech by talking to them.

[U1]: *You know what, if the patient talks to you, you will have already known whether this is normal for you, is this how you sound like? since I would have had an interaction with the patient before we start the whole examination, and I might tell 'oh this sounds off'.* (Q24)

The other five doctors thought that the test was useful to specifically confirm a patient's level of slurred speech.

[U4]: *For me, it is a useful test. Although I have already known from the conversation, it is used for confirming the symptoms.* (Q25)

According to [60], patients expected the consultation to include medical examination, tests, diagnosis, treatment, prescriptions, medication, and outcomes. Therefore, two doctors also said that the availability of this feature will increase patients' trust even though the outcome of using or not using this test is similar.

[T4]: *It's alright. This test is still good to have so that it will be more clear to the patient to know that we are doing examination.* (Q26)

4.4.8 Ataxia Test

Two doctors faced technical problems with robot arm movement when using the ataxia function, due to the robot's built-in collision avoidance.

Overall the participants confirmed that the use of the wide-angle lens camera allowed them to observe how patients touch the robot's finger and their own nose, which did lead to a feeling of immersiveness.

[U3]: *Good, good. This is good. I think this feature is the best. I think it is crucial to be able to see patients touching the robot, to see how smooth they can perform the test.* (Q27)

However, one Thai doctor reported that the finger-to-nose test needs to be administered carefully because patients can sometimes be confused about how to perform the test properly,

[T3]: *The robot's dialogue says only that it will perform finger-to-nose test, and ask if the patient is not fully understanding the test. I'm just thinking. It should have quite a lot of description, like what we will do in the beginning. When I examine finger-to-nose test with real patient, the patient is also confused. They make a mistake of using their left finger to touch their nose and their right finger to touch the doctor's finger which in fact should be the same finger. So I think it should be explained what we will do first or having a diagram about how to perform finger-to-nose test inside the room would be helpful as well.* (Q28)

The participant also noted that the distance between the robot and the patient may affect the quality of the test. In a correctly administered test, the patient's arms must be fully stretched.

[T2]: *Here, I have some comments. When doing this test, the patient has to fully stretch their arm. So when navigating the robot, there should be a way to fix the distance between the patient and the robot for the test to be effective.* (Q29)

5 Discussion

This section describes how the findings in Sect. 4 can be used to refine the design requirements (R) presented in Sect. 2.2 into updated design requirements (Ra).

First of all, MS Teams has been chosen as a video conferencing tool (R1) to establish a connection between the doctor's computer and the tablet. This has proved to be successful since all eight participants were able to conduct the remote study without communication problems. Therefore, we conclude that the tablet is sufficient to establish real-time communication (R1a).

Secondly, MS Teams comes with a built-in chat function (R2) to be used in the case of a poor internet connection. A text-to-speech function can be used alternatively in case of a poor internet connection (R2a) to have the robot speak directly to the patient (Q5).

Thirdly, AR is implemented as a UI paradigm for providing and displaying medical information to the doctor (R3). In this study, the use of AR can help doctors to perform neurological examination. Participants have no concern about using AR as they found the UI to be easy to use (Q1 and Q2), and therefore in the future, AR should still be utilised as the main UI paradigm to deal with other clinical conditions that are harder to diagnose using a video feed than they would be in person.

Navigation is also one of the important functionalities for doctors and assistants to move the robot (R4). However, participants found this feature difficult to use due to the lack of depth information (Q6 and Q7) and delay from the robot's poor connectivity (Q8 and Q9). Therefore, the navigation interface should implement depth information to estimate the distance ahead, and the robot's connectivity has to be improved to minimize the delay (R4a).

Before the robot telemedicine system was designed, the design requirement suggested that using a screen on the robot would be enough for the doctor to display information to the patient, and for the patient to see the doctor (R5). However, to enable telecommunication capabilities while maintaining high video resolution with a wide-angle camera view, the robot's tablet has to be replaced by the Android tablet and the GoPro camera is required (Sect. 2.4). In addition, the camera position was too low (Q7). Therefore, both the external screen and camera are required for doctor-patient communication and the frame for holding the camera needs to be re-designed to allow the camera height to be adjusted by the doctor (R5a).

Next, a secure data connection between doctors and patients is required to protect the patient's privacy (R6). In this study, for the implementation of the video conferencing feature, third-party software such as TeamViewer and MS Teams were utilized to minimize the development time. It is important to note that in a real application, all medical communications have to be executed through Healthcare Insurance Portability and Accountability Act (HIPAA) compliant telemedicine platforms. Using third-party software is both illegal and unethical in terms of data storage security. Therefore, it should be noted that these software tools were applied solely for the purpose of usability testing and the telecommunication setup will have to be redesigned.

Lastly, it is suggested that the system should provide a feature for assistants to transfer additional data exceeding the system's capabilities to doctors (R7). In the presented system prototype, TeamViewer is implemented for the doctor's computer to connect to the platform computer. Therefore, assistants can transfer additional data using the same user interface as doctors (R7a).

Overall, participant feedback from the semi-structured interview has been used to refine the design requirements (R) presented in Sect. 2.2 into updated design requirements (Ra) as follows:

R1a A tablet is sufficient to establish real-time communication.

R2a In case of a poor internet connection, text-to-speech can be used in addition to chat functionality.

R4a The navigation interface should implement depth information to estimate the distance ahead, and the robot's connectivity has to be improved to minimize the delay.

R5a Both external screen and camera are required for doctor-patient communication. The frame for holding the camera needs to be re-designed to adjust the camera height.

R7a Assistants can transfer additional data using the same user interface as doctors.

In addition, our data suggests that additional design requirements for the robot telemedicine system are needed to meet user feedback that is not covered by already existing design requirements. First of all, to detect whether patients have suffered from Bell's palsy, there is a concern about eye asymmetry detection that has an unstable accuracy due to false positives from rolling eyeballs. Eye asymmetry detection should only utilise eyelid landmarks, not eyeball landmarks (R8a) (Q17). To enhance neurological examination, asymmetry in forehead wrinkles will assist doctors since it is a sign of peripheral facial palsy which is presented in Bell's palsy (R9a) (Q18). Then, in the arm weakness test and ataxia test, the robot's built-in collision avoidance function detected participants' hands as obstacles (Sect. 4.4.5 and 4.4.8), causing the robot to stop its arm movement halfway. Hence, the compatibility of robot obstacle detection and medical functionality needs to be ensured so that doctors can perform these two tests properly while ensuring the safety of patients (R10a). Again in the arm weakness test, a timer should be added to help them count time while assessing arm weakness (R11a) (Q20) and weights are required to assess the higher grade of arm muscle strength (R12a) (Q19).

For the dysphagia test, all participants found the feature to be impractical and not useful, thus it should be removed (R13a) (See Sect. 4.4.6). This demonstrates the importance of extensive consultation during system design to maximise the utility of implemented features.

Apart from medical examination functionality, all participants suggested that the interface for patient data management needed to be improved to support better patient recording (R14a) (Q14). In addition, Thai doctors requested for the recording system to be in a single window to increase ease of use (R14a) (Q15). Again, to minimise the time, some doctors suggested having a scoring system to evaluate patient conditions (R15a) while others suggested having a search function for medication and examination types (R15a) (Sect. 2.6). Lastly, most doctors found Pepper's accent to be human-like, and the appearance to be friendly (Q1), thus suitable for social interaction (R16a), but its behaviour must be evaluated in the future to see whether it is suitable for clinical scenarios and increases the quality of medical examination for the patient or not.

The requirement extensions can be summarised as follows:

- **R8a** Eyes asymmetry detection should only utilise eyelid landmarks, not eyeball landmarks.
- **R9a** Forehead asymmetry detection would aid in facial palsy detection.
- **R10a** Compatibility of robot obstacle detection and medical functionality needs to be ensured.
- **R11a** A timer should be added for assessing arm weakness.
- **R12a** Weights should be used to assess higher-grade arm strength.
- **R13a** The dysphagia function should be removed as it is not useful.
- **R14a** The interface has to support better patient information recording and should be integrated into a single main window.
- **R15a** It is recommended to use score ratings and search functions to minimize typing.
- **R16a** Pepper's accent and appearance are suitable for social interaction, but its behaviour must be evaluated.

5.1 Cultural Differences: Thai vs UK Doctors

Observing participants during usability testing and analyzing their feedback uncovered that there are major differences in doctors' working environments between Thailand and the UK. Thai GPs are more likely to examine much higher numbers of patients than UK GPs, both in-person and remotely. As a result, they tend to have more experience with remote examination technology (Table 1) despite having fewer years of clinical practice. Comments from the interview show that UK doctors spend more time on how to make patients feel comfortable before conducting a remote examination (**R16a**), whereas the comments from Thai doctors focus on how to improve the interface of the robot telemedicine system so that they spend less time for each patient (**R11a**, **R14a**, **R15a**). This may explain the SUS and PQ results in Table 1 and 2. Thai GPs focus on overall system usability which results in lower SUS scores. UK GPs focus on the feeling of presence when using the system to talk with patients, resulting in lower PQ scores. In summary, the system design for Thai doctors should focus on enhancing system accessibility to reduce diagnostic time, while the one for UK doctors should focus on implementing social interaction to build rapport with patients.

One of the explanations for the above differences could be that in the UK patients have to book appointments for non-urgent medical examinations, which gives GPs a predefined amount of time with each patient. In Thailand, patients can walk into the hospital and receive treatment on the same day without booking an appointment. This leaves GPs with less diagnosis time per patient. This difference in the healthcare systems has also been studied. It has been shown that Thailand is likely to have better non-urgent medical services in comparison to other countries, which causes a rapidly growing medical tourism sector [61–63].

However, although the robot telemedicine system implements the diagnosis of exclusion to examine Bell's palsy, both the UK neurologist consultant and all the UK GPs mentioned that it is very unlikely for them to diagnose the patients with stroke remotely since they can call an ambulance. Thus, the UK neurologist consultant advised that the robot telemedicine system should focus on examinations for more other non-emergency symptoms in which we intend to expand the functionalities of the robot in the future. In Thailand in contrast, even though Thai patients can call for emergency medical services (EMS), 1 of 5 patients with serious illness do not use EMS due to being badly informed and not having previous experiences with the service [64]. Moreover, it has been found that most Thai laymen cannot identify stroke symptoms and underestimate the severity of the condition [65, 66]. In addition, some areas in Thailand suffer from a lack of EMS coverage and bad traffic conditions [67]. Thus, some of the Thai GPs mentioned that in rural areas there is a small chance that GPs have to remotely examine patients and refer them to the nearest hospital in case of stroke symptoms. In light of this feedback, we note that while our choice of clinical condition to be examined was atypical from the doctors' point of view, it allowed us to test the principles of system operation.

5.2 Limitations

In terms of a discussion of the limitations of this work, first, the system has not been evaluated by doctors and patients at the same time. This decision has been made due to the tight schedule of the participating doctors. This proved to be a good decision as some study sessions had to be arranged on very short notice, and one study had to be conducted out-of-hours to meet one participant's schedule. Secondly, the study has not been conducted with patients who are suffering from facial paralysis. Due to the current iteration of the system's facial palsy detection not having been tested in a clinical context, the decision has been made not to invite real patients for ethical reasons. Third, the assistant is not involved in this study to investigate whether the doctors can carry on their tasks successfully without the need of an assistant. However, this approach does not reflect the real use case as and future system evaluation will also have to include an investigation of the doctor-assistant connection. Fourth, the number of participants is too small to definitively establish that there are cultural difference between UK and Thai doctors. Note that we did not intend to investigate the cultural difference between UK and Thai doctors from the beginning, and therefore we did not conduct the study with more participants. Nevertheless, our findings suggest that differences in cultures should be taken into account when designing robot telemedicine systems. Moreover, further study is required to investigate the cultural differences of medical practitioners in different countries where telemedicine systems are being considered. Lastly, one limitation is that we are not considering user interface accessibility principles for users with disabilities. Thus, the current study may not cover some usability issues such colour blindness. If GP has a colour blindness, one potential usability problem is the navigation part, where the navigation interface uses the colour as a sign of whether the robot is moving or not.

6 Conclusion and Future Work

In this work, a user-centred design process was implemented to develop a robot telemedicine system for facial paralysis examination. The robot telemedicine system has been evaluated by usability testing conducted with Thai and UK doctors. Results from quantitative and qualitative methods show that there is room for improvement in terms of the usability of the robot telemedicine system. Based on participant feedback, a new set of design recommendations has been developed. It was also found that the views of UK and Thai doctors about the system differ, most likely because of differing working conditions in the UK and Thai healthcare systems. Hence, the specific culture the system will be used in must be considered for its design.

In terms of future work, for the robot telemedicine system to be functional, it is suggested to focus more on other non-urgent clinical conditions. We could add the use of person-tracking technologies to examine participant behaviour, emotional responses, etc.

Secondly, it is important to note that this study only involved evaluation by doctors. To provide a more holistic view of system usability studies involving other stakeholders, e.g., patients and medical assistants will be conducted in the future.

Lastly, a social robot is employed as the robotics platform in this work, since it is planned to implement social behaviour on the robot with the goal to increase the quality of the medical examination for the patient. A future study will be investigated how the robot is perceived when it acts in the role of a receptionist facilitating the interaction with the patient before and after the doctor connects to the system for the medical examination.

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the work presented in this article.

References

1. Hann M, Gravelle H (2004) The maldistribution of general practitioners in England and Wales: 1974–2003. Br J Gen Pract 54(509):894–898
2. Padarath A, Chamberlain C, McCoy D et al (2003) Health personnel in Southern Africa: confronting maldistribution and brain drain. Equinet Discussion Papers, vol 4, no 1, pp 1–41
3. Kanchanachitra C, Lindelow M, Johnston T et al (2011) Human resources for health in southeast Asia: shortages, distributional challenges, and international trade in health services. Lancet 377(9767):769–781. https://doi.org/10.1016/S0140-6736(10)62035-1
4. GBD 2019 Human Resources for Health Collaborators (2022) Measuring the availability of human resources for health and its relationship to universal health coverage for 204 countries and territories from 1990 to 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 399:2129–2154. https://doi.org/10.1016/S0140-6736(22)00532-3
5. World Health Organization (2010) Increasing access to health workers in remote and rural areas through improved retention: global policy recommendations. https://iris.who.int/handle/10665/44369
6. Office for National Statistics (2022) Trends in patient-to-staff numbers at GP practices in England: 2022. [online]. Available from: https://www.ons.gov.uk/releases/trendsinstafftopatientnumbersatgppracticesinengland2022
7. Sharma A, Lambert TW, Goldacre MJ (2012) Why UK-trained doctors leave the UK: cross-sectional survey of doctors in New Zealand. J R Soc Med 105(1):25–34. https://doi.org/10.1258/jrsm.2011.110146
8. Salafsky B, Glasser M, Ha J (2005) Addressing issues of maldistribution of health care workers. Ann Acad Med Singapore 34(8):520–525
9. Wibulpolprasert S, Pachanee CA (2008) Addressing the internal brain drain of medical doctors in Thailand: the story and lesson learned. Glob Soc Policy 8(1):12–15. https://doi.org/10.1177/14680181080080010104
10. Nundy M, Bhatt P (2022) The health system in the kingdom of Thailand: reforms, achievement, and challenges. [pdf]. Available from: https://csep.org/wp-content/uploads/2022/11/Health-System-in-Thailand.pdf

11. World Health Organization. Regional Office for the Western Pacific (2015) The Kingdom of Thailand health system review, vol 5. WHO Regional Office for the Western Pacific
12. World Health Organization and Alliance for Health Policy and Systems Research (2017) Primary health care systems (primasys): case study from Thailand: abridged version. Technical documents
13. World Health Organization (2020) The impact of the COVID-19 pandemic on noncommunicable disease resources and services: results of a rapid assessment. World Health Organization, Geneva
14. Boman K, Olofsson M, Berggren P et al (2014) Robot-assisted remote echocardiographic examination and teleconsultation: a randomized comparison of time to diagnosis with standard of care referral approach. JACC Cardiovasc Imaging 7(8):799–803. https://doi.org/10.1016/j.jcmg.2014.05.006
15. Pfeil JN, Rados DV, Roman R et al (2023) A telemedicine strategy to reduce waiting lists and time to specialist care: a retrospective cohort study. J Telemed Telecare 29(1):10–17. https://doi.org/10.1177/1357633X20963935
16. Fatehi F, Martin-Khan M, Smith AC et al (2015) Patient satisfaction with video teleconsultation in a virtual diabetes outreach clinic. Diabetes Technol Ther 17(1):43–48. https://doi.org/10.1089/dia.2014.0159
17. Anderson J, Walsh J, Anderson M et al (2021) Patient satisfaction with remote consultations in a primary care setting. Cureus 2019(9):6–8. https://doi.org/10.7759/cureus.17814
18. Rothan HA, Byrareddy SN (2020) The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 109:102433. https://doi.org/10.1016/j.jaut.2020.102433
19. Gilbert AW, Billany JC, Adam R et al (2020) Rapid implementation of virtual clinics due to COVID-19: report and early evaluation of a quality improvement initiative. BMJ Open Qual 9(2):1–8. https://doi.org/10.1136/bmjoq-2020-000985
20. Petrazzuoli F, Gokdemir O, Antonopoulou M et al (2022) Patient consultations during SARS-CoV-2 pandemic: a mixed-method cross-sectional study in 16 European countries. Rural Remote Health 3(1):144–151. https://doi.org/10.22605/RRH7196
21. Buvik A, Bugge E, Knutsen G et al (2019) Patient reported outcomes with remote orthopaedic consultations by telemedicine: a randomised controlled trial. J Telemed Telecare 25(8):451–459. https://doi.org/10.1177/1357633X18783921
22. Preusche C, Ortmaier T, Hirzinger G (2002) Teleoperation concepts in minimal invasive surgery. Control Eng Pract 10(11):1245–1250. https://doi.org/10.1016/S0967-0661(02)00084-9
23. Sung GT, Gill IS (2001) Robotic laparoscopic surgery: a comparison of the da Vinci and Zeus systems. Urology 58(6):893–898. https://doi.org/10.1016/S0090-4295(01)01423-6
24. Arbeille P, Capri A, Ayoub J et al (2007) Use of a robotic arm to perform remote abdominal telesonography. AJR Am J Roentgenol 188(4):317–322. https://doi.org/10.2214/AJR.05.0469
25. Labiris G, Fanariotis M, Christoulakis C et al (2003) Tele-ophthalmology and conventional ophthalmology using a mobile medical unit in remote Greece. J Telemed Telecare 9(5):296–299. https://doi.org/10.1258/135763303769211337
26. Hadziahmetovic M, Nicholas P, Jindal S et al (2019) Evaluation of a remote diagnosis imaging model vs dilated eye examination in referable macular degeneration. JAMA Ophthalmol 137(7):802–808. https://doi.org/10.1001/jamaophthalmol.2019.1203
27. Craig JJ, Mcconville JP, Patterson VH et al (1999) Neurological examination is possible using telemedicine. J Telemed Telecare 5(3):177–181. https://doi.org/10.1258/1357633991933594
28. Handschu R, Littmann R, Reulbach U et al (2003) Telemedicine in emergency evaluation of acute stroke: interrater agreement in remote video examination with a novel multimedia system. Stroke 34(12):2842–2846. https://doi.org/10.1161/01.STR.0000102043.70312.E9
29. Pennington L, Stamp E, Smith J et al (2019) Internet delivery of intensive speech and language therapy for children with cerebral palsy: a pilot randomised controlled trial. BMJ Open 9(1):1–12. https://doi.org/10.1136/bmjopen-2018-024233
30. Khan AJ, Szczepura A, Palmer S et al (2022) Physical therapy for facial nerve paralysis (Bell's palsy): an updated and extended systematic review of the evidence for facial exercise therapy. Clin Rehabil 36(11):1424–1449. https://doi.org/10.1177/02692155221110727
31. Vargas BB, Shepard M, Hentz JG et al (2017) Feasibility and accuracy of teleconcussion for acute evaluation of suspected concussion. Neurology 88(16):1580–1583. https://doi.org/10.1212/WNL.0000000000003841
32. Awadallah M, Janssen F, Körber B et al (2019) Telemedicine in general neurology: interrater reliability of clinical neurological examination via audio-visual telemedicine. Eur Neurol 80(5–6):289–294. https://doi.org/10.1159/000497157
33. Cai Y, Zheng YJ, Gulati A et al (2021) Patient use of low-cost digital videoscopes and smartphones for remote ear and oropharyngeal examinations. JAMA Otolaryngol Head Neck Surg 147(4):336–342. https://doi.org/10.1001/jamaoto.2020.5223
34. Stollnberger G, Moser C, Beck E et al (2014) Robotic systems in health care. In: 2014 7th international conference on human system interactions (HSI). IEEE, pp 276–281. https://doi.org/10.1109/HSI.2014.6860489
35. Carranza K, Day N, Lin L et al (2018) Development of a telepresence robot with integration of a general practice medical consultation module. Int J Eng Technol (UAE) 7(566):203–207
36. Carranza KAR, Day NJB, Lin LMS et al (2019) Akibot: a telepresence robot for medical teleconsultation. In: 2018 IEEE 10th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management, HNICEM 2018, pp 5–8. https://doi.org/10.1109/HNICEM.2018.8666283
37. Adrion C, Weiss B, Paul N et al (2020) Enhanced recovery after intensive care (ERIC): study protocol for a German stepped wedge cluster randomised controlled trial to evaluate the effectiveness of a critical care telehealth program on process quality and functional outcomes. BMJ Open 10(9):e036096. https://doi.org/10.1136/bmjopen-2019-036096
38. Spies CD, Paul N, Adrion C et al (2023) Effectiveness of an intensive care telehealth programme to improve process quality (ERIC): a multicentre stepped wedge cluster randomised controlled trial. Intensive Care Med 49(2):191–204. https://doi.org/10.1007/s00134-022-06949-x
39. Myhren H, Ekeberg Ø, Tøien K et al (2010) Posttraumatic stress, anxiety and depression symptoms in patients during the first year post intensive care unit discharge. Crit Care 14(1):1–10. https://doi.org/10.1186/cc8870
40. Fan E, Cheek F, Chlan L et al (2014) An official American thoracic society clinical practice guideline: The diagnosis of intensive care unit-acquired weakness in adults. Am J Respir Crit Care Med 190(12):1437–1446. https://doi.org/10.1164/rccm.201411-2011ST
41. Charité – University Medicine Berlin (2022) Charité project ERIC: tele-visit for intensive care patients recommended as standard care. [online]. Available from: https://www.charite.de/service/pressemitteilung/artikel/detail/charite_projekt_eric_tele_visite_fuer_intensivpatienten_zur_regelversorgung_empfohlen/
42. Norman DA, Draper SW (1988) User centered system design: new perspectives on human–computer interaction. L. Erlbaum Associates Inc., Mahwah
43. Chirapornchai C, Niyi-Odumosu F, Giuliani M et al (2022) User Requirements for a robot teleoperation system for general medical examination. In: 2022 31st IEEE international conference on robot and human interactive communication (RO-MAN). IEEE, Naples,

Italy, pp 858–863. https://doi.org/10.1109/RO-MAN53752.2022.9900855
44. Ritchie J, Spencer L (2002) Qualitative data analysis for applied policy research. In: Analyzing qualitative data. Routledge, pp 187–208
45. Tanioka R, Locsin R, Yasuhara Y et al (2018) Potential legal issues and care implications during care-prevention gymnastic exercises for the elderly using pepper in long term health care facilities. Intell Control Autom 09(03):85–93. https://doi.org/10.4236/ica.2018.93007
46. Zhou Y (2019) Robot application in the medical area. PhD thesis, Vaasa University of Applied Sciences
47. Induruwa I, Holland N, Gregory R et al (2019) The impact of misdiagnosing Bell's palsy as acute stroke. Clin Med J R Coll Physicians Lond 19(6):494–498. https://doi.org/10.7861/clinmed.2019-0123
48. Parra-Dominguez GS, Garcia-Capulin CH, Sanchez-Yanez RE (2022) Automatic facial palsy diagnosis as a classification problem using regional information extracted from a photograph. Diagnostics. https://doi.org/10.3390/diagnostics12071528
49. Hsu GSJ, Kang JH, Huang WF (2019) Deep hierarchical network with line segment learning for quantitative analysis of facial palsy. IEEE Access 7:4833–4842. https://doi.org/10.1109/ACCESS.2018.2884969
50. Lucey P, Cohn JF, Kanade T et al (2010) The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE computer society conference on computer vision and pattern recognition—workshops, CVPRW 2010, pp 94–101. https://doi.org/10.1109/CVPRW.2010.5543262
51. Lewis C (1982) Using the "thinking aloud" method in cognitive interface design. Research report, IBM TJ Watson Research Center Yorktown Heights, NY
52. Nielsen J, Landauer TK (1993) A mathematical model of the finding of usability problems. ACM Press, New York, pp 206–213. https://doi.org/10.1145/169059.169166
53. Brooke J (1996) SUS: a "quick and dirty'usability. Usability Eval Ind 189(3):189–194
54. Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. Presence Teleoperators Virtual Environ 7:225–240. https://doi.org/10.1162/105474698565686
55. Hein D, Mai C, Hußmann H (2018) The usage of presence measurements in research: a review. In: Proceedings of the international society for presence research annual conference (presence). The International Society for Presence Research Prague, pp 21–22
56. Witmer BG, Jerome CJ, Singer MJ (2005) The factor structure of the presence questionnaire. Presence Teleoperators Virtual Environ 14:298–312. https://doi.org/10.1162/105474605323384654
57. UQO Cyberpsychology Lab (2004) Presence questionnaire. [pdf]. Available from: https://marketinginvolvement.files.wordpress.com/2013/12/pq-presence-questionnaire.pdf
58. Gale NK, Heath G, Cameron E et al (2013) Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 13(1):117. https://doi.org/10.1186/1471-2288-13-117
59. Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 4:114–123. https://doi.org/10.5555/2835587.2835589
60. Kenten C, Bowling A, Lambert N et al (2010) A study of patient expectations in a Norfolk general practice. Health Expect 13(3):273–284. https://doi.org/10.1111/j.1369-7625.2010.00603.x
61. Cohen E (2008) Medical tourism in Thailand. AU-GSB e-journal 1(1):24–37
62. Wong KM, Velasamy P, Arshad TNT (2014) Medical tourism destination SWOT analysis: a case study of Malaysia, Thailand, Singapore and India. In: SHS web of conferences, vol 12, p 01037. https://doi.org/10.1051/shsconf/20141201037
63. Reed CM (2008) Medical tourism. Med Clin N Am 92:1433–1446. https://doi.org/10.1016/j.mcna.2008.08.001
64. National Institute for Emergency Medicine (Thailand) (2017) What causes Thai patients not to think of 1669 during medical emergency. [pdf]. Available from: niems.go.th/1/upload/migrate/file/256011140906534322_UxQ25Uds6UmTRb8Z.pdf
65. Sirisamutr T, Ponsen K, Wachiradilok P (2018) Knowledge, perspective, and reasons of not-calling emergency medical services in Thailand. J Health Syst Res 12
66. Sirisamutr T, Wachiradilok P (2019) Awareness, perspective, and reasons behind patients' rejection of emergency medical services in Thailand. Prehosp Disaster Med 34(s1):s104–s104. https://doi.org/10.1017/S1049023X19002176
67. Sittichanbuncha Y, Prachanukool T, Sarathep P et al (2014) An emergency medical service system in Thailand: providers' perspectives. J Med Assoc Thail 97(10):1016–1021