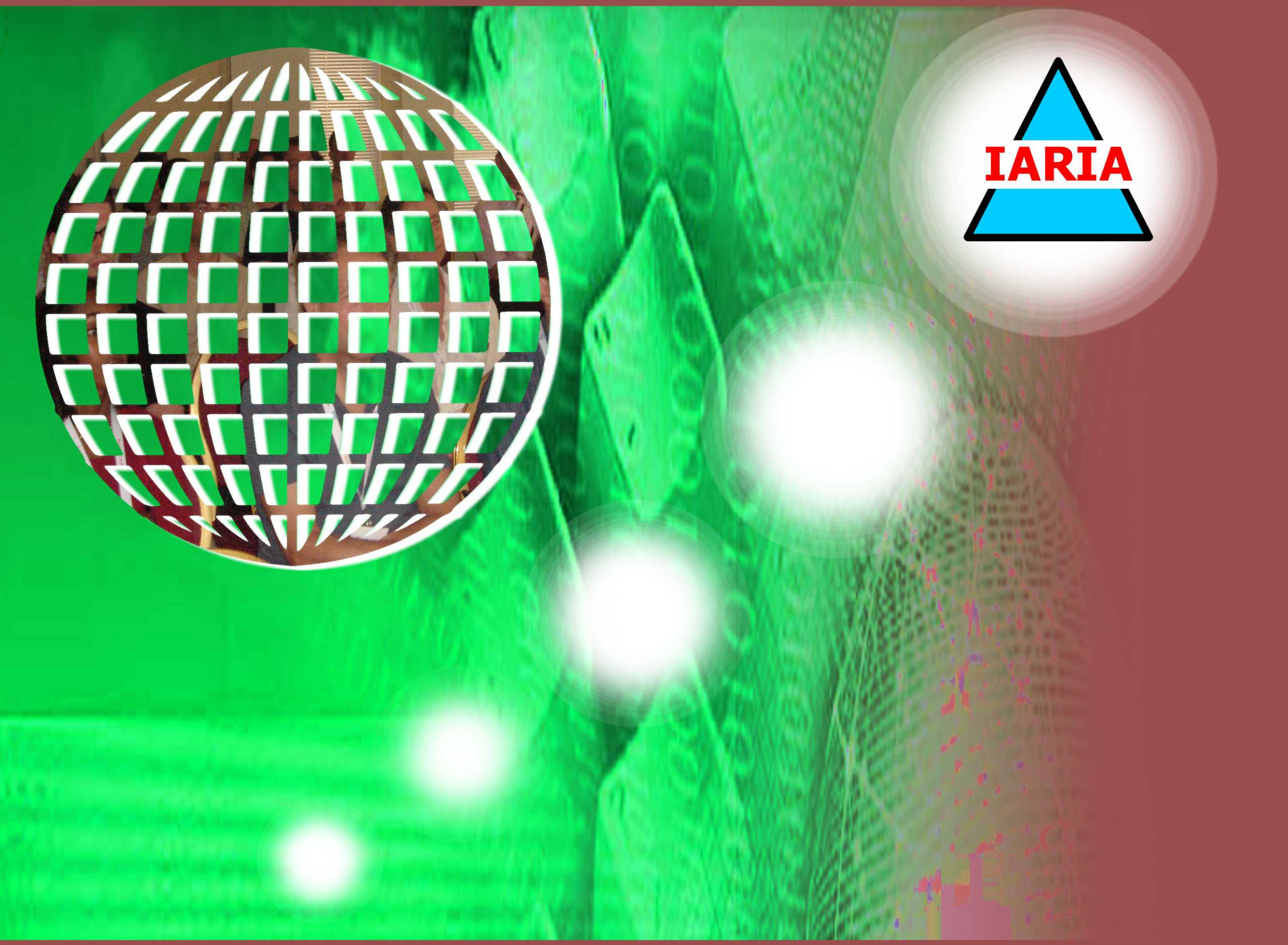
International Journal on Advances in Life Sciences



2019 vol. 11 nr. 3&4

The *International Journal on Advances in Life Sciences* is published by IARIA.
ISSN: 1942-2660
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Life Sciences
Volume 11, Number 3 & 4, 2019

Bee Bee Chua, University of Technology, Sydney, Australia
Udi Davidovich, Amsterdam Health Service - GGD Amsterdam, The Netherlands
Maria do Carmo Barros de Melo, Telehealth Center, School of Medicine - Universidade Federal de Minas Gerais (Federal University of Minas Gerais), Brazil
Kari Dyb, Norwegian Centre for Integrated Care and Telemedicine / University Hospital of North Norway | University of Tromsø, Norway
Juergen Eils, DKFZ, German
Anne G. Ekeland, Norwegian Centre for Integrated Care and Telemedicine / University Hospital of North Norway | University of Tromsø, Norway
El-Sayed M. El-Horbaty, Ain Shams University, Egypt
Ivan Evgeniev, TU Sofia, Bulgaria
Karla Felix Navarro, University of Technology, Sydney, Australia
Joseph Finkelstein, The Johns Hopkins Medical Institutions, USA
Stanley M. Finkelstein, University of Minnesota - Minneapolis, USA
Adam M. Gadomski, Università degli Studi di Roma La Sapienza, Italy
Ivan Ganchev, University of Limerick, Ireland / University of Plovdiv “Paisii Hilendarski”, Bulgaria
Jerekias Gandure, University of Botswana, Botswana
Xiaohong Wang Gao, Middlesex University - London, UK
Josean Garrués-Irurzun, University of Granada, Spain
Hassan Ghazal, Moroccan Society for Telemedicine and eHealth, Morocco
Piero Giacomelli, SPAC SPA -Arzignano (Vicenza), Italia
Alejandro Giorgetti, University of Verona, Italy
Anthony Glascock, Drexel University, USA
Wojciech Glinkowski, Polish Telemedicine Society / Center of Excellence "TeleOrto", Poland
Francisco J. Grajales III, eHealth Strategy Office / University of British Columbia, Canada
Conceição Granja, Conceição Granja, University Hospital of North Norway / Norwegian Centre for Integrated Care and Telemedicine, Norway
William I. Grosky, University of Michigan-Dearborn, USA
Richard Gunstone, Bournemouth University, UK
Amir Hajjam-El-Hassani, University of Technology of Belfort-Montbéliard, France
Lynne Hall, University of Sunderland, UK
Päivi Hämäläinen, National Institute for Health and Welfare, Finland
Anja Henner, Oulu University of Applied Sciences, Finland
Marike Hettinga, Windesheim University of Applied Sciences, Netherlands
Stefan Hey, Karlsruhe Institute of Technology (KIT) , Germany
Dragan Ivetic, University of Novi Sad, Serbia
Sundaresan Jayaraman, Georgia Institute of Technology - Atlanta, USA
Malina Jordanova, Space Research & Technology Institute, Bulgarian Academy of Sciences, Bulgaria
Attila Kertesz-Farkas, University of Washington, USA
Hassan Khachfe, Lebanese International University, Lebanon
Valentinas Klevas, Kaunas University of Technology / Lithuaniain Energy Institute, Lithuania
Anant R Koppar, PET Research Center / KTwo technology Solutions, India
Bernd Krämer, FernUniversität in Hagen, Germany
Ramesh Krishnamurthy, Health Systems and Innovation Cluster, World Health Organization - Geneva, Switzerland
Roger Mailler, University of Tulsa, USA
Dirk Malzahn, OrgaTech GmbH / Hamburg Open University, Germany
Salah H. Mandil, eStrategies & eHealth for WHO and ITU - Geneva, Switzerland
Herwig Mannaert, University of Antwerp, Belgium
Agostino Marengo, University of Bari, Italy
Igor V. Maslov, EvoCo, Inc., Japan
Ali Masoudi-Nejad, University of Tehran , Iran
Cezary Mazurek, Poznan Supercomputing and Networking Center, Poland
Teresa Meneu, Univ. Politécnica de Valencia, Spain

International Journal on Advances in Life Sciences
Volume 11, Numbers 3 & 4, 2019

CONTENTS

Facial-Expression Training Application for Medical Doctors

Akemi Oe, Yoshitoshi Murata
Graduate School of Software and Information Science
Iwate Prefectural University
Iwate, Japan
e-mails: akemioe@gmail.com, y-murata@iwate-pu.ac.jp

***Abstract*— Establishing trust between a patient and a doctor depends as much on their relationship as on the doctor's medical abilities. One of the important factors in building and maintaining a relationship is whether the doctor produces facial expressions appropriate for the patient's condition. We developed a facial-expression training application for medical students and tested its effectiveness as a training tool. In this application, the Microsoft Cognitive Services Emotion API is used to analyze the patient's facial expression. Prior to developing this application, we studied what kind of facial expressions were appropriate for a doctor greeting or treating an adult patient. We focused on four situations. Three of them are greeting situations when a doctor begins a patient interview in the general ward of a hospital. The fourth is when a doctor examines a patient by placing a stethoscope to a patient's chest, i.e., auscultation. We also identified appropriate facial expressions for doctors treating pediatric patients. To verify the training application's effectiveness, the facial expressions before and after learning with the application were evaluated by potential patients or mothers of potential pediatric patients. The results demonstrated the application's utility.**

Keywords-doctor-patient interaction; facial expression; nonverbal communication; facial-expression training application.

I. INTRODUCTION

We study facial expressions appropriate for young medical doctors and develop facial-expression training applications for medical students, one of which was presented at eTELEMED 2019 [1]. We have also identified appropriate facial expressions for when a doctor greets a patient at the beginning of a medical interview in the general ward of a hospital [2].

Patient satisfaction is an important component of medical care [3]. Improving patient satisfaction enhances trust and the relationship between patient and doctor, which leads to stronger adherence to prescribed protocol, such as taking medicine, and to enhanced therapeutic effects [4][5]. Many studies and reviews have shown that the main determinant of patient satisfaction is the patient-doctor relationship [6]–[10] and that patient satisfaction is higher when the patient communicates with a doctor who has strong nonverbal-communication skills [11][12]. However, young inexperienced doctors and medical students often have trouble producing appropriate facial expressions when greeting a patient. The first author of this paper, a lecturer on medical communication, often hears young doctors complaining that, though they intend to smile, patients say that they seem angry.

In response to these complaints, we developed a facial-expression training application that physicians can use for independent study. In this application, the Microsoft Cognitive Services Emotion API is used to analyze the doctor's facial expression. This application uses the results of our study on acceptable expressions for doctors as model doctor expressions [2]. These model expressions are standards to compare with those of learners using the application. To verify the effectiveness of the application, learners' facial expressions before and after learning with the application were evaluated by potential patients. The evaluation results confirmed the utility of the application.

After introducing related work in Section II, acceptable expressions required for doctors are explained in Section III. The system of the facial-expression training application is introduced in Section IV. We discuss the results of our experiments in Section V. Section VI gives our conclusion, and Section VII considers future work.

II. RELATED WORK

This study focuses on facial expressions, we review research on nonverbal communication and research related to learning facial expressions.

A. Nonverbal Communication

Medical interviews have traditionally focused on gathering relevant information from patients [13]. In contemporary medicine, the focus has expanded to building a trusting relationship, sharing decision-making, responding to the patient's emotional state, and supporting actions related to the patient's condition and treatment; this requires the doctor to have a wide range of communication skills [14]. These skills include "looking at a patient not as a case but as a human being" [15] and "building and maintaining a good relationship between doctor and patient" [16]. It has been shown that such skills have a greater effect on patient satisfaction than the doctor's medical skills, the medicine prescribed, the information provided, the questions asked, and the advice and instructions given. In particular, a patient's satisfaction is positively related to the doctor being warm [15]–[17], empathic [15][17]–[19], friendly [17], and giving the impression of being human [18].

Nonverbal communication is a means of communicating these emotional aspects of oneself. Patient satisfaction is higher when the doctor has a strong ability to express his or her emotions and to read the emotions of others through nonverbal cues such as facial expressions, gaze, posture, and tone of voice [10][20]. In short, a doctor's nonverbal communication is an important aspect of patient care.

B. Learning Facial Expressions

Natural, unconscious facial expressions can be seen in humans from infancy [21], but eventually conscious expressions appear. Conscious expressions might also be thought of as "false" expressions. These expressions are skills that form the basis of more complicated expressive behaviors, such as emotional expressions performed in accordance with rules in communication situations [22].

Facial expressions as expressions of emotion are thought to be founded in an understanding of emotions, and there are many developmental studies about understanding emotions. However, studies about learning facial expressions are mostly focused on social-skills training for children with developmental disabilities, and there are not many studies for adults.

In one study, participants were required to express other people's facial expressions by recalling emotions and showing photos of faces with emotions, the showing photos effect was shown about "anger" [23].

In research to develop a smile-training system with the goal of training participants to produce celebrity-like smiles (a facial expression recognized as attractive) [24], researchers created a smile-fitted deformation considering the characteristics of a celebrity-like smile in participants' expressionless faces. As a result of training, participants were able to express a smile highly likely to be recognized as attractive by others.

From the above, it is clear that presenting appropriate facial-expression images is effective for facial-expression training as both a method of expressing emotions and a learning method.

III. ACCEPTABLE FACIAL EXPRESSIONS

The most acceptable facial expressions for doctors when they greet adult patients at the beginning of a medical interview and when auscultating adult patients in the general ward of a hospital were identified. Because the requirements when treating children are quite different from those when treating adults, the most appropriate facial expressions for three types of pediatric patients were also identified.

A. Acceptable Facial Expression for Adult Patients

A previous study [2] has revealed what are considered acceptable facial expressions for a doctor in accordance with the patient's condition. One way to analyze appropriate facial expressions for doctors is to define the facial expressions of experienced doctors. However, their facial expressions are not always deemed appropriate. Also, because many young physicians have trouble presenting appropriate facial expressions when greeting a patient, we chose to find facial expressions that would be acceptable for most patients, including potential patients, from facial expressions that medical students think are suitable. We chose greeting patients in the general ward of a hospital as the target situation.

The procedure was as follows. The participant role-playing the patient portrayed three conditions, as shown in Figure 1: a patient who feels physically healthy (a "bright patient"), one whose physical condition is unknown (an "expressionless patient"), and one who feels badly and is suffering pain (a "patient in pain"). We photographed these role-played conditions. Then, a plurality of medical students greeted the three photographs with the facial expressions that each student deemed appropriate, and we recorded video of them doing so. Although evaluation by actual patients is best, it would have been difficult to request their participation. We asked generally healthy adults who been hospitalized in the past or would be in the future to evaluate the videos. To make the subjective evaluation more effective, we first had 16 people view and evaluate each video recording and removed the ones, in which the student's facial expression was judged to be unacceptable. We then had 31 other people view and evaluate the remaining recordings (Tables I and II).

Comparing an evaluation by adults and analytical results from a computer of the medical student's greeting movies revealed what were considered the acceptable facial expressions in accordance with the patient's condition.

Bright Expressionless in Pain

Figure 1. Conditions portrayed by role-playing adult patient.

The acceptable facial expressions when a young doctor greets an adult patient who is hospitalized in a general ward are as follows;

- For patients who feel physically healthy, the most acceptable facial expression is Figure 2: "continuous happiness" (expressed more as a laugh rather than simply a smile).
- For patients without a facial expression, the most acceptable facial expression is Figure 3: initially "happiness" (expressed as a smile) and then "neutrality" (expressionlessness).
- For patients in bad physical condition suffering pain, the most acceptable facial expression is Figure 4: "neutrality" with a little "sadness" or "surprise."

The cells in Figures 2–4 corresponding to 0 or more and less than 0.2 are shown in blue, 0.2 or more and less than 0.4 in green, 0.4 or more and less than 0.6 in yellow, 0.6 or more and less than 0.8 in orange, and 0.8 or more in red. The total for all emotions is 1, and the value for neutrality is obtained by subtracting the total value for the seven emotions from 1.

TABLE I. RESULTS OF FIRST SUBJECTIVE EVALUATION

	Video-ID	F51	F67	F65	M54	F59	F58	F57	M48	F55	F50	F20	F27	F24	M34	M34	M32
Bright	B-1	1			2			1	2	1	1	2	2	1	1	2	1
	B-2	1	2	2	2	1	1	1	2	1	1	2	1	1	2	2	1
	B-3	1			2			2	2	1	1	2	2	1	2	1	1
	B-4	1	2	1	2	1	1	2	1	2	2	1	2	2	2	2	1
	B-5							1	1	2	1	2	2	1	2	2	2
	B-6							2	2	2	1	2	2	2	2	2	1
	B-7	1			4			3	1	1	1	2	3	2	2	2	1
	B-8	1	2	1	2	1	2	3	2	2	3	2	2	2	4	4	2
	B-9	1	1	2	4	3	2	2	4	2	4	2	2	3	3	2	3
	B-10	2			2			2	4	1	2	2	3	3	4	3	4
Expressionless	E-1	1			2			3	1	2	1	3	3	1	2	1	2
	E-2	1			2			2	2	3	1	2	3	2	2	2	2
	E-3							2	3	2	2	2	2	2	2	2	1
	E-4							2	3	2	2	2	2	2	2	2	1
	E-5	1			2			3	2	1	1	3	4	2	2	2	2
	E-6	1	2	2	2	4	2	2	2	4	2	2	2	2	2	2	1
	E-7	1	2	3	2	1	2	1	4	3	4	3	1	3	3	2	2
	E-8	2	2	1	2	3	1	2	2	2	3	3	2	3	3	3	3
	E-9	2	3	2	3	4	3	2	1	3	2	3	2	3	3	4	1
	E-10	2			2			3	4	1	2	3	3	3	4	2	4
in Pain	P-1							2	2	1	2	2	2	3	2	2	1
	P-2	1			2			2	3	1	1	3	1	3	2	3	2
	P-3							1	1	1	2	2	2	3	3	3	2
	P-4	2			2			1	1	2	2	2	4	2	2	2	3
	P-5	2	2	2	2	2	2	2	3	3	2	1	3	4	2	2	4
	P-6	2	2	2	2	2	2	2	3	3	2	1	3	4	2	2	4
	P-7	2	2	2	4	2	3	2	2	3	2	1	2	4	4	3	3
	P-8	2			2			3	2	3	2	3	3	3	4	2	2
	P-9	1	2	3	4	2	3	4	3	3	3	4	2	5	4	4	1
	P-10	2			4			4	4	2	2	4	3	3	2	3	4

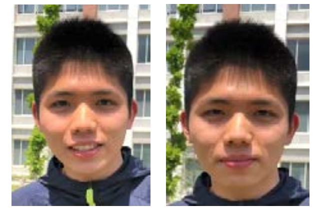
F=female M=male

TABLE II. RESULTS OF SECOND SUBJECTIVE EVALUATION

Video-ID		Score					avg.
		1	2	3	4	5	
Bright	B-2	9	17	4	1	0	1.9
	B-4	13	8	9	1	0	1.9
	B-3	7	10	12	2	0	2.3
	B-5	2	15	13	1	0	2.4
	B-6	5	8	16	2	0	2.5
	B-1	3	9	16	3	0	2.6
Expressionless	E-1	4	10	14	3	0	2.5
	E-3	5	5	19	2	0	2.6
	E-4	4	6	18	3	0	2.6
	E-8	3	5	22	0	1	2.7
	E-2	1	6	13	11	0	3.1
in Pain	P-4	10	8	9	4	0	2.2
	P-3	5	8	16	2	0	2.5
	P-1	3	9	13	5	1	2.7
	P-8	2	10	8	10	1	2.9
	P-2	0	4	14	11	2	3.4

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:00.7	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:01.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:02.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:01.7	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:02.8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.3	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.9	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 2. Acceptable facial expression for adult patients who feel physically healthy.



Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.34
00:00.7	0.71	0.00	0.00	0.00	0.00	0.00	0.00	0.29
00:01.2	0.79	0.00	0.01	0.00	0.00	0.00	0.00	0.20
00:01.6	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.06
00:02.2	0.35	0.00	0.01	0.00	0.00	0.00	0.00	0.63
00:02.7	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.07
00:03.2	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.76
00:03.8	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.70
00:04.0	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.59

Figure 3. Acceptable facial expression for adult patients without a facial expression.

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.7	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.93
00:01.1	0.00	0.00	0.00	0.00	0.00	0.06	0.02	0.92
00:01.6	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.96
00:02.2	0.00	0.00	0.00	0.00	0.00	0.08	0.01	0.92
00:02.6	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.66
00:00.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
00:03.2	0.00	0.01	0.02	0.00	0.00	0.13	0.00	0.84
00:03.6	0.00	0.00	0.01	0.00	0.00	0.05	0.00	0.94
00:03.0	0.00	0.00	0.01	0.00	0.00	0.06	0.00	0.92

Figure 4. Acceptable facial expression for adult patients in bad physical condition suffering pain.

To identify facial expressions appropriate for a doctor performing auscultation, videos were created and evaluated using the following procedure;

Step 1: Take photograph of patient being auscultated (Figure 5).

Step 2: Record videos of two medical doctors producing facial expressions they thought appropriate for the patient (Figure 6).

Step 3: Analyze videos using computer-aided facial-expression-emotion analysis system.

Step 4: Have ten potential patients evaluate doctors' facial expressions on 3-point scale (1: appropriate, 2: neutral, 3: inappropriate).

As shown in Tables III and IV, the results of the facial-expression-emotion analysis are mostly "neutral." As shown in Table V, the potential patients feel that the facial expressions in Figure 6 are appropriate. The potential patients feel that the facial expressions in the two videos are mainly "neutral," meaning that they feel that the doctor is performing the auscultation in a serious manner.

Figure 5. Role-playing patient being auscultated.

(1) Video A-1 (2) Video A-2

Figure 6. Medical students performing auscultation.

TABLE III. COMPUTER ANALYSIS RESULTS FOR VIDEO A-1.

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:00.6	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.96
00:01.1	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.97
00:01.6	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.93
00:02.1	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.98
00:02.6	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.98
00:03.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:03.6	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.99
00:04.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:04.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:05.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:05.6	0.12	0.01	0.01	0.00	0.00	0.04	0.00	0.81
00:05.6	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.98
00:06.2	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.97
00:06.7	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:07.3	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.98
00:07.8	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.93

TABLE IV. COMPUTER ANALYSIS RESULTS FOR VIDEO A-2.

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.00	0.03	0.00	0.00	0.00	0.04	0.00	0.92
00:00.6	0.00	0.03	0.01	0.00	0.00	0.01	0.00	0.94
00:01.1	0.00	0.03	0.02	0.00	0.00	0.01	0.00	0.94
00:01.6	0.00	0.03	0.01	0.00	0.00	0.01	0.00	0.94
00:02.1	0.00	0.03	0.01	0.00	0.00	0.02	0.00	0.94
00:02.6	0.00	0.06	0.01	0.00	0.00	0.02	0.00	0.91
00:03.2	0.00	0.07	0.02	0.00	0.00	0.01	0.00	0.90
00:03.7	0.00	0.03	0.03	0.00	0.00	0.04	0.00	0.90
00:04.2	0.00	0.01	0.05	0.00	0.00	0.05	0.00	0.89
00:04.7	0.00	0.01	0.04	0.00	0.00	0.03	0.00	0.91
00:05.2	0.00	0.02	0.02	0.00	0.00	0.03	0.00	0.91
00:05.7	0.00	0.02	0.03	0.00	0.00	0.02	0.00	0.92
00:06.2	0.00	0.02	0.02	0.00	0.00	0.03	0.00	0.93
00:06.7	0.00	0.01	0.03	0.00	0.00	0.02	0.00	0.93
00:07.3	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.96
00:07.7	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.96
00:08.3	0.00	0.00	0.01	0.00	0.00	0.05	0.00	0.94

TABLE V. RESULTS OF SUBJECTIVE EVALUATION.

Video-ID	Score			avg.
	1	2	3	
A-1	10	0	0	1.0
A-2	9	1	0	1.1

B. *Acceptable Facial Expression for Pediatric Patients*

The expressions appropriate for pediatric patients likely differ from ones appropriate for adult patients. We were able to get cooperation from seven doctors in this part of our study (Table VI). We used our computer-aided facial-expression-emotion analysis system to analyze the facial expressions they produced when shown photographs of pediatric patients in three different conditions.

To identify facial expressions appropriate for greeting hospitalized pediatric patients in the same three conditions described above ("bright patient," "expressionless patient," and "patient in pain"), we used the following procedure;

Step 1: Take photograph of a role-playing pediatric patient (8-year-old) as she produced expressions for each of the conditions (Figure 7) .

Step 2: Record video of the seven doctors as they produced facial expressions they thought appropriate for each photograph.

Step 3: Analyzed the emotion shown by their expressions by using computer-aided facial-expression-emotion analysis.

Step 4: Have evaluation from eleven mothers who has children.

Although the evaluation by children is best, it would have been difficult to request them. We asked mothers who have children.

The results of computer analysis were used to divide the facial expressions of seven pediatricians into three groups.

- Group 1 (Pediatricians 1, 2, 4): for bright pediatric patients who feel physically healthy—"continuous happiness" (expressed with a bright smile). For pediatric patients without a facial expression—"neutral" (expressionlessness). For pediatric patients in bad physical condition suffering pain—"neutral" with a little "surprise" (expressed with a pained expression).

TABLE VI. PROFILES OF PEDIATRICIANS.

	Age	Gender	Number of Years of Experience
Pediatrician 1	30	F	2
Pediatrician 2	33	F	5
Pediatrician 3	35	M	6
Pediatrician 4	39	F	11
Pediatrician 5	40	M	15
Pediatrician 6	55	M	29
Pediatrician 7	64	M	39

Bright Expressionless in Pain

Figure 7. Conditions portrayed by role-playing pediatric patient.

- Group 2 (Pediatricians 3, 5): for bright pediatric patients—"neutral and happiness" (expressed with a natural smile). For pediatric patients without a facial expression—"neutral" (expressionlessness). For pediatric patients in bad physical condition suffering pain—initially "neutral" (expressionlessness) and then "happiness" (expressed with a smile).
- Group 3 (Pediatricians 6, 7); for bright pediatric patients—"happiness" (expressed with a natural smile). For pediatric patients without a facial expression—almost "happiness" (expressed with a smile). For pediatric patients in bad physical condition suffering pain—initially "neutral" (expressionlessness) and then "happiness" (expressed with an encouraging smile).

The facial expressions of Pediatrician 4 (Tables VII–IX), Pediatrician 5 (Tables X–XII), and Pediatrician 6 (Tables XIII–XV) were selected as representative ones for the corresponding groups.

TABLE VII. GROUP 1 (PEDIATRICIAN 4): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "BRIGHT PATIENT"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.71	0.00	0.00	0.00	0.00	0.01	0.00	0.28
00:02.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:01.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:04.2	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.07
00:03.2	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.04

TABLE VIII. GROUP 1 (PEDIATRICIAN 4): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "EXPRESSIONLESS PATIENT"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.91
00:01.2	0.15	0.00	0.00	0.00	0.00	0.00	0.01	0.83
00:02.2	0.92	0.00	0.00	0.00	0.00	0.00	0.01	0.07
00:03.2	0.16	0.00	0.00	0.00	0.00	0.00	0.28	0.56
00:04.3	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.96

TABLE IX. GROUP 1 (PEDIATRICIAN 4): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "PATIENT IN PAIN"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:01.2	0.49	0.00	0.00	0.00	0.00	0.00	0.00	0.51
00:02.2	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.97
00:03.2	0.02	0.00	0.00	0.00	0.00	0.01	0.14	0.82
00:00.1	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.99
00:04.2	0.17	0.00	0.00	0.00	0.00	0.00	0.30	0.52
00:05.2	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.96

TABLE X. GROUP 2 (PEDIATRICIAN 5): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "BRIGHT PATIENT"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.5	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.98
00:00.9	0.06	0.00	0.01	0.00	0.00	0.01	0.31	0.61
00:01.5	0.03	0.01	0.02	0.01	0.00	0.02	0.05	0.85
00:02.1	0.22	0.00	0.05	0.01	0.00	0.02	0.01	0.69
00:02.6	0.30	0.00	0.02	0.00	0.00	0.01	0.00	0.67
00:00.0	0.00	0.00	0.01	0.02	0.00	0.19	0.01	0.78
00:03.1	0.30	0.00	0.01	0.00	0.00	0.00	0.01	0.68
00:03.6	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.50

TABLE XI. GROUP 2 (PEDIATRICIAN 5): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "EXPRESSIONLESS PATIENT"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.05	0.00	0.01	0.00	0.00	0.01	0.00	0.93
00:00.6	0.07	0.00	0.01	0.00	0.00	0.01	0.00	0.92
00:01.1	0.06	0.00	0.01	0.01	0.00	0.01	0.01	0.90
00:01.6	0.01	0.00	0.01	0.00	0.00	0.04	0.01	0.92
00:02.1	0.02	0.00	0.01	0.00	0.00	0.02	0.01	0.93
00:02.7	0.01	0.00	0.01	0.00	0.00	0.04	0.01	0.93
00:03.3	0.02	0.01	0.01	0.02	0.00	0.03	0.01	0.89
00:03.7	0.00	0.01	0.03	0.03	0.00	0.12	0.00	0.81
00:04.2	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.98

TABLE XII. GROUP 2 (PEDIATRICIAN 5): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "PATIENT IN PAIN"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:00.6	0.03	0.01	0.01	0.01	0.00	0.01	0.13	0.80
00:01.1	0.29	0.00	0.04	0.00	0.00	0.00	0.00	0.66
00:01.6	0.01	0.01	0.07	0.03	0.00	0.06	0.03	0.79
00:02.2	0.02	0.00	0.01	0.01	0.00	0.01	0.02	0.94
00:02.7	0.03	0.00	0.01	0.01	0.00	0.06	0.00	0.88
00:03.3	0.25	0.00	0.01	0.00	0.00	0.00	0.01	0.73
00:03.8	0.07	0.00	0.03	0.00	0.00	0.04	0.00	0.86
00:04.3	0.21	0.00	0.01	0.00	0.00	0.00	0.00	0.76
00:04.8	0.10	0.00	0.04	0.01	0.00	0.03	0.00	0.82
00:05.4	0.83	0.00	0.01	0.00	0.00	0.08	0.00	0.09
00:06.0	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.07
00:06.5	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.02

TABLE XIII. GROUP 3 (PEDIATRICIAN 6): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "BRIGHT PATIENT"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.4	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.55
00:01.1	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.20
00:01.8	0.77	0.00	0.00	0.00	0.00	0.00	0.00	0.23
00:02.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:04.5	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:05.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE XIV. GROUP 3 (PEDIATRICIAN 6): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "EXPRESSIONLESS PATIENT"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.4	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.68
00:01.1	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.40
00:01.8	0.96	0.00	0.01	0.00	0.00	0.00	0.00	0.03
00:02.5	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.03
00:03.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:04.5	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE XV. GROUP 3 (PEDIATRICIAN 6): COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "PATIENT IN PAIN"

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.4	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.69
00:01.1	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.81
00:01.7	0.65	0.00	0.00	0.00	0.00	0.00	0.00	0.35
00:02.4	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.14
00:03.0	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.08
00:03.7	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.02
00:04.3	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

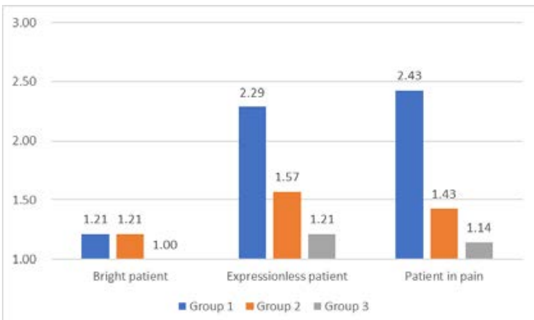


Figure 8. Average score for each condition.

The appropriateness of those expressions was evaluated by 14 mothers who had children between the ages of 2 and 12. They did this by watching videos of the expressions and scoring them on a scale (1: appropriate, 2: neutral, 3: inappropriate).

As shown in Figure 8, the representative facial expressions for the Group 3 pediatricians were evaluated overall as the most appropriate for pediatric patients. As shown in Table VI, the doctor who made the expressions had 29 years of experience. The expressions are shown in Figures 9–11.

For bright pediatric patients, the average score was good for all three groups. The most appropriate facial expression was the same as that for adult patients, a bright smile expressing "happiness." For pediatric patients without a facial expression, a smile held longer than for adults was the most appropriate facial expression. For pediatric patients in bad physical condition suffering pain, an initially "neutral" expression and then an encouraging smile expressing "happiness" was the most appropriate facial expression.

The biggest difference in the appropriate expression between adult and pediatric patients was for patients in bad physical condition suffering pain. As expressed by two of the mothers, if a child is feeling pain and the pediatrician has a pained expression, the child will tend to cry. Another factor is that children are more likely to be scared by something than adults. It is also possible that an expression of encouragement from the pediatrician reassures the child.

We asked an experienced pediatrician (Pediatrician 7) to describe the biggest difference between responding to adults and responding to children. He replied that adults are more often told the truth when the news is bad while it is important to give children hope of being cured. Thus the appropriate facial expression for a "patient in pain" differs greatly between adults and children.

Figure 9. Acceptable facial expression for pediatric patients who feel physically healthy.

Figure 10. Acceptable facial expression for pediatric patients without a facial expression.

Figure 11. Acceptable facial expression for pediatric patients in bad physical condition suffering pain.

IV. FACIAL-EXPRESSION TRAINING APPLICATION SYSTEM

The system of the facial-expression training application is introduced in this section.

A. *Requirements and Design Concept*

Learners study in the following order so that the appropriate facial expressions can be learned efficiently and repeatedly. When starting, to enable learners to notice what their facial expressions are before training, the system has them greet with expressions that they think are suitable without any specific instructions.

Step 1: A learner chooses one of the model patients.

Step 2: The learner greets the model patient with a facial expression that the learner thinks appropriate, and this greeting is recorded.

Step 3: A video of the model doctor greeting the patient with the appropriate facial expression and analysis data are displayed, and important points to notice about the appropriate facial expressions are introduced.

Step 4: The video recorded in Step 2 and its analysis data are displayed.

Step 5: The learner repeats from Steps 2 to 4 until satisfied.

Step 6: If the learner is satisfied, the learner selects a patient with a different condition and returns to Step 1.

In the developed system, in addition to being able to check the facial expressions the learner performed by recording them as a video, the learner can also check the quality of his or her expression on a frame-by-frame basis. Each learner can save images as learning data and facial-expression analysis results by applying security processing because these data and analysis results are important as research data.

B. System Configuration

We developed and used a system that quantitatively analyzes changes in facial expression. It is based on the Cognitive Services Emotion API [25] provided by Microsoft's Azure cloud service, and a facial-expression-emotion detection system for video images. Microsoft Emotion API corresponds to multiple races. Our facial-expression-emotion analysis system calculates the ratio for seven emotions (happiness, anger, contempt, disgust, fear, sadness, and surprise) reflected in the input video image and for neutrality. The total for all emotions is 1, and the value for neutrality is obtained by subtracting the total value for the seven emotions from 1. The configuration of this application is as shown in Figure 12. A video camera is controlled by the Open CV [26]. Recoded video data are converted the Motion-JPEG, and send to the Cognitive Services Emotion API.

First, the learner selects a patient to be a training partner in Figure 13. After clicking the recording button, the learner talks to the model patient. When the greeting and consultation is over, the learner clicks the stop button followed by the next button in Figure 14.

The emotional values of the model doctor's facial expression are displayed. By selecting a timeline, the corresponding facial expression is displayed in Figure 15.

The emotional values of the learner's facial expression are displayed. By selecting a timeline, the corresponding facial expression is displayed in Figure 16.

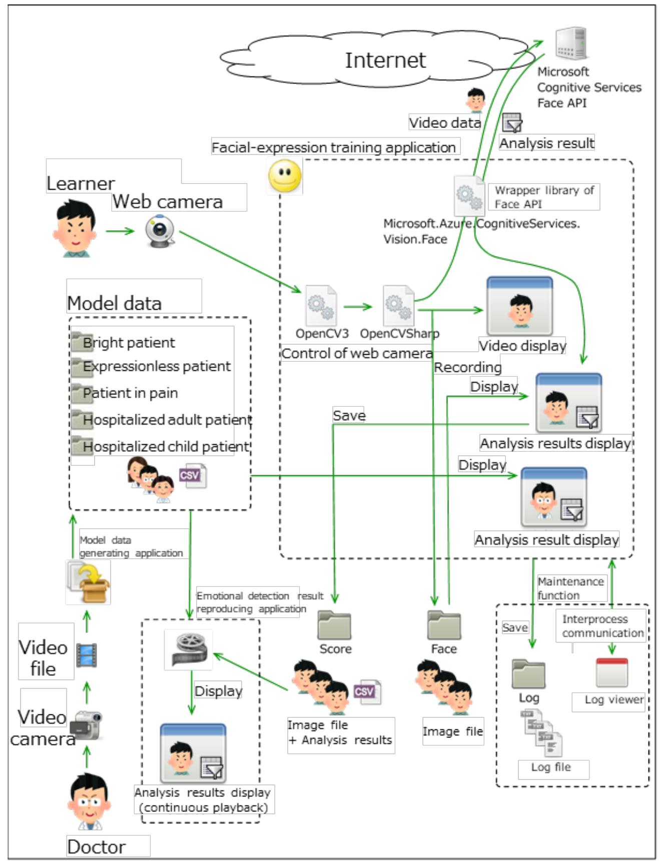


Figure 12. Configuration of facial-expression training application.

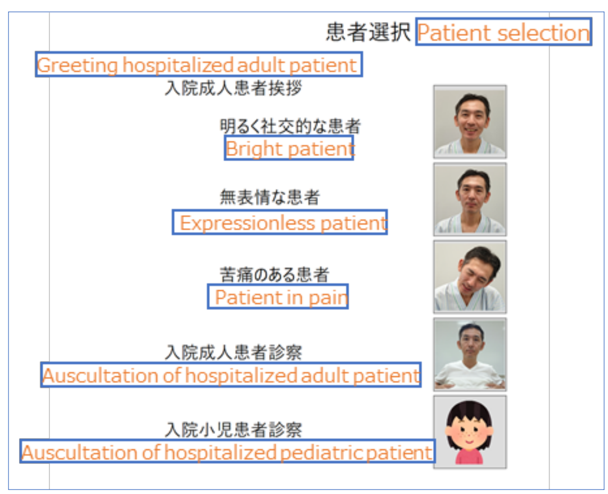


Figure 13. Step 1: Patient-selection screen.

Figure 14. Step 2: Learner greets and examines patient while screen is recorded.

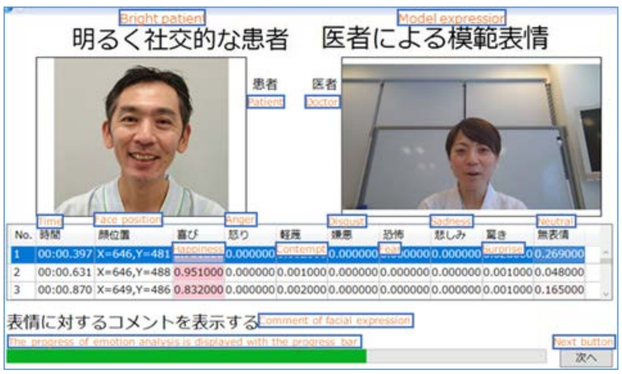


Figure 15. Step 3: Model doctor's appropriate facial expression and analysis results.

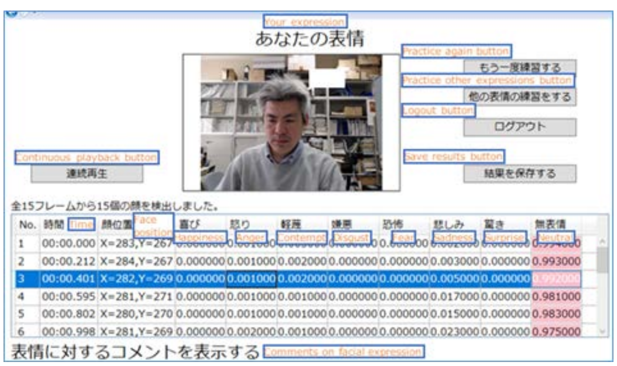


Figure 16. Step 4: Analysis of image taken in Step 2.

C. *Model Doctors*

Videos of model doctors were prepared with the following procedure. In a previous study [2], the acceptable doctors' expressions for the patients were measured and selected. The model doctors (one male and one female) reproduced the acceptable facial expressions, and their reproductions were recorded as videos. Seventy-nine people evaluated these videos of Figures 17–21 and confirmed that they were appropriate, so we adopted them as the model doctor videos.

(1) Woman (2) Man

Figure 17. Model doctor's appropriate facial expression for bright patiets.

(1) Woman (start) (2) Woman (half-way)
(3) Man (start) (4) Man (half-way)

Figure 18. Model doctor's appropriate facial expression for expressionless patiets.

(1) Woman (2) Man

Figure 19. Model doctor's appropriate facial expression for patiets in pain.

(1) Female "doctor" (2) Male "doctor"

Figure 20. Appropriate facial expressions for auscultation.

(1) Female "doctor" (2) Male "doctor"

Figure 21. Appropriate facial expressions for Pediatric patient.

V. EVALUATION

In order to examine the effect of our facial-expression training application, the facial expressions of six participants before and after practice were evaluated.

A. *Experiment*

Six participants including three medical students played doctors and greeted patients with facial expressions that they thought appropriate for each condition of Figure 13. The application recorded them (before training), the participants repeated the facial-expressions training so that they could approximate the analysis results of the model doctors. After practice, we recorded their facial expressions again (after training). Although evaluation by actual patients is best, it would have been difficult to request actual patients' participation. Hence, we asked 30 general healthy adults who had been hospitalized in the past or would be in the future. We showed the video recordings (before and after training) to 10 men and 20 women (average age 41.0 years) without sound. We asked them to judge whether the doctor's facial expression was appropriate for the condition on a 3-point scale (1: appropriate, 2: neutral, 3: not appropriate). We asked them to also comment on anything they felt or noticed. We showed the recordings without sound because we wanted them to focus on the appropriate facial expressions in medical communication conditions, and emotion is easier to read from speech than from facial expressions.

B. *Effects of Training*

The graph of Figures 22–25 showing the result of the rating before and after the training is as follows.

In order to examine the effects of training with the application, we conducted 2 (using the application: before and after) × 6 (doctors) two-way analysis of variance with the evaluation results as the dependent under four conditions. The results showed a significant difference at the 1% level under all conditions (in order from condition 1, $F(59, 295) = 3.19$, $P<.01$; $F(59, 295) = 4.51$, $P<.01$; $F(59, 295) = 4.52$, $P<.01$; $F(59, 177) = 5.26$, $P<.01$).

We consider the effectiveness of this application to have been confirmed because the evaluation scores improved with use of the application under all conditions. We also noticed the following two points during the experiment. The first point is that the evaluation for "patients in pain" was different for each person. Most persons regarded the doctor's serious expression as empathy and evaluated that as adequate. However, a few persons felt that the doctor's expression caused unease and worry, and that person evaluated the expression as not adequate. Therefore, the model expressions for patients in pain may not be limited to one type. The second point is that the facial expressions practiced in the application are unnatural for a few persons. By training in the application, people playing the role of doctor created expressions close to those of the model doctor's videos, and accordingly the evaluation score rose. However, when a facial expression that the participants learned was far from facial expressions that they always do, a few evaluators felt that the expression looked like "a pretended expression" or "artificial expression."

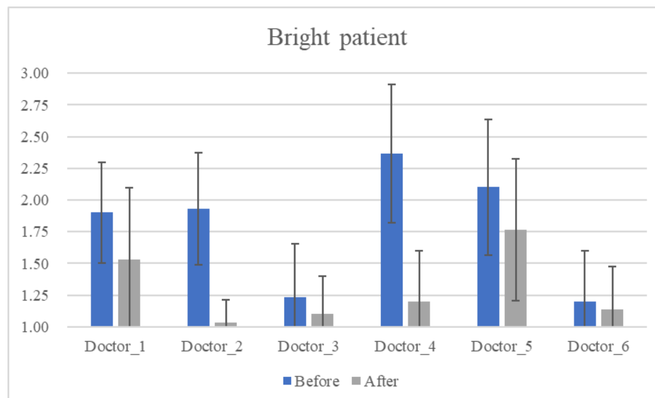


Figure 22. Average and standard deviation of each doctor's evaluation score for bright patients.

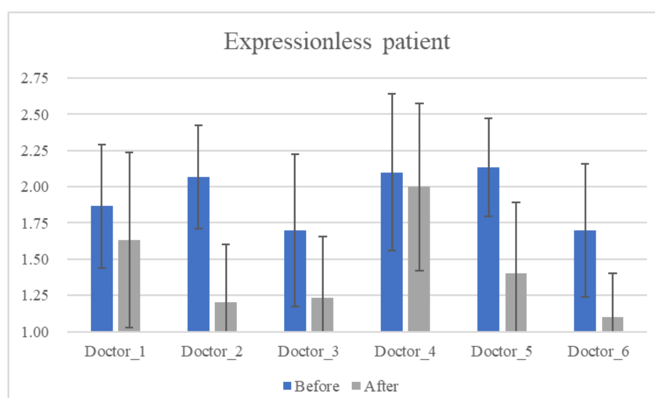


Figure 23. Average and standard deviation of each doctor's evaluation score for expressionless patients.

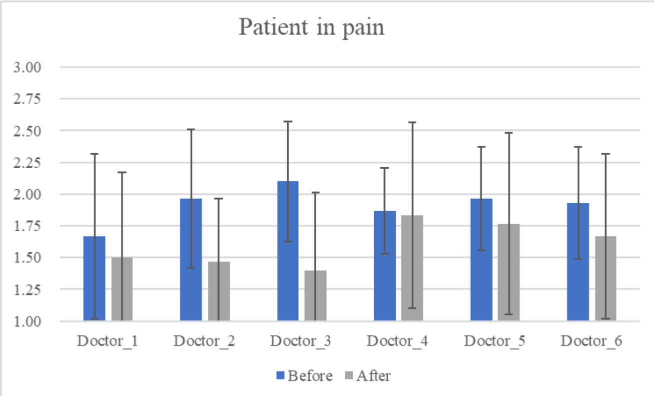


Figure 24. Average and standard deviation of each doctor's evaluation score for patients in pain.

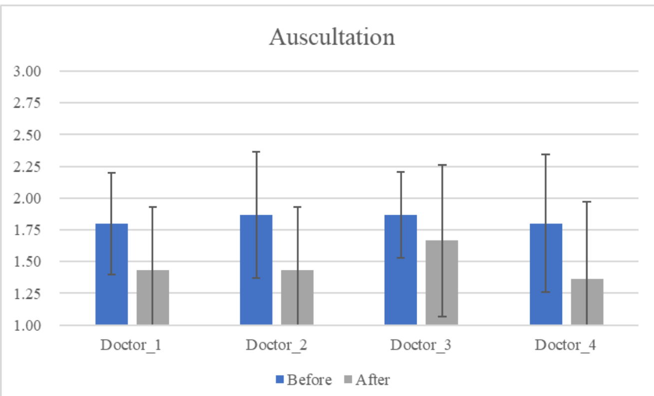


Figure 25. Average and standard deviation of each doctor's evaluation score for auscultation.

VI. CONCLUSION

Our quantitative analysis of medical student facial expressions when greeting adult patients and pediatric patients at the beginning of a medical interview in the general ward of a hospital revealed acceptable facial expressions. For adult patients who feel physically healthy, the most acceptable facial expression is "continuous happiness" (expressed more as a laugh than simply as a smile). For pediatric patients who feel physically healthy, the most acceptable facial expression is "happiness" (a natural smile). For adult patients without a facial expression, the most acceptable facial expression is initially "happiness" (expressed as a smile) and then "neutral" (without expression). On the other hand, for pediatric patients without a facial expression, the most acceptable facial expression is almost "happiness" (expressed as a longer smile). For adult patients suffering pain, the most acceptable facial expression is "neutral" with a little "sadness" or "surprise." On the other hand, for pediatric patients suffering pain, the most acceptable facial expression is initially "neutral" (expressionlessness) and then "happiness" (expressed as an encouraging smile).

During auscultation, the most acceptable facial expression is always "neutral."

We developed a facial-expression training application that physicians can use for independent study. Six participants practiced using this application to express appropriate facial expressions. To verify the effectiveness of this application, the facial expressions that participants made before and after training with it were evaluated by 30 people. On the basis of the results, we consider the usefulness of the facial-expression training application to be verified. We think most learners could express their facial expression naturally by practicing repeatedly.

In the future, we will improve the application so that the on-screen instructions will guide the learner's expressions more properly.

Given the importance of smiles when people interact with each other in many countries or cultures, smiles on "bright patients" in this study are expected to be widely effective.

REFERENCES

[1] A. Oe and Y. Murata, "Quantitative Analysis of Facial-Expression Training Application for Medical Doctors," IARIA, proceeding of eTELEMED 2019, pp. 1–6, 2019.

[2] A. Oe and Y. Murata, "Identification of Appropriate Facial Expressions for Medical Doctors" IARIA, International Journal on Advances in Life Sciences, vol 11 no 1 & 2, pp. 44-55, 2019.

[3] A. Donabedian, The Definition of Quality and Approaches to Its Assessment. Ann Arbor, MI: Health Administration Press, 1980.

[4] S. H. Kaplan, S. Greenfield, and W. E. Ware, Jr, "Assessing the effects of physician-patient interactions on the outcomes of chronic disease," Med. Care, vol. 27, pp. S110–S127, March 1989.

[5] L. M. Noble, A. Kubacki, J. Martin, and M. Lloyd, "The effect of professional skills training on patient-centredness and confidence in communicating with patients," Med. Educ., vol. 41, pp. 432–440, May 2007.

[6] J. A. Hall, D. L. Roter, and N. R. Katz, "Meta-analysis of correlates of provider behavior in medical encounters," Med. Care, vol. 26, pp. 657–675, July 1988.

[7] J. A. Hall, J. A. Harrigan, and R. Rosenthal, "Nonverbal behavior in clinician-patient interaction," Appl. Prev. Psychol., vol. 4, pp. 21–37, Winter 1995.

[8] L. M. Ong, J. C. de Haes, A. M. Hoos, and F. B. Lammes, "Doctor-patient communication: A review of the literature," Soc. Sci. Med., vol. 40, pp. 903–918, April 1995.

[9] T. L. Thompson, "Interpersonal communication and health care," in Handbook of Interpersonal Communication, 2nd ed., M. L. Knapp and G. R. Miller, Eds. Thousand Oaks, CA: Sage, pp. 696–725, 1994.

[10] S. Williams, J., Weinman, J, and J. Dale, "Doctor-patient communication and patient satisfaction: A review," Family Practice, vol. 15, pp. 480–492, October 1998.

[11] M. R. DiMatteo, A. Taranta, H. S. Friedman, and L. M. Prince, "Predicting patient satisfaction from physicians' nonverbal communication skills," Med. Care, vol. 18, pp. 376–387, April 1980.

[12] M. R. DiMatteo, R. D. Hays, and L. M. Prince, "Relationship of physicians' nonverbal communication skill to patient satisfaction, appointment noncompliance, and physician workload," Health Psychol., vol. 5, pp. 581–594, 1986.

[13] J. Bird, S. A. and Cohen-Cole, "The three-function model of the medical interview. An educational device," Advances in Psychosomatic Medicine., vol. 20, pp. 65–88, 1990.

[14] H. de Haes and J. Bensing, "Endpoints in medical communication research, proposing a framework of functions and outcomes," Patient Educ. Couns., vol. 74, pp. 287–294, March 2009.

[15] Z. Ben-Sira, "Affective and instrumental components in the physician-patient relationship: An additional dimension of interaction theory," J. Health Soc. Behav., vol. 21, pp. 170–180, June 1980.

[16] M. K. Buller and D. B. Buller, "Physicians' communication style and patient satisfaction," J. Health Soc. Behav., vol. 28, pp. 375–388, December 1987.

[17] B. M. Korsch and V. F. Negrete, "Doctor-patient communication," In Physician-patient Communication: Readings and Recommendations, G. Henderson, Ed. Springfield, IL: Charles C. Thomas, pp. 29–40, 1981.

[18] S. J. Williams and M. Calnan, "Convergence and divergence: Assessing criteria of consumer satisfaction across general practice, dental and hospital care settings," Soc. Sci. Med., vol. 33, pp. 707–716, 1991.

[19] R. C. Wasserman, T. S. Inui, R. D. Barriatua, W. B. Carter, and P. Lippincott, "Pediatric clinicians' support for parents makes a difference: An outcome-based analysis of clinician-parent interaction," Pediatrics, vol. 74, pp. 1047–1053, December 1984.

[20] H. S. Friedman, M. R. DiMatteo, and A. Taranta, "A study of the relationship between individual differences in nonverbal expressiveness and factors of personality and social interaction," J. Res. Pers., vol. 14, pp. 351–364, September 1980.

[21] L. A. Camras and M. M, Shuster, "Current emotion research in developmental psychology," Emotion review, 5, pp. 321-329, 2013.

[22] M. Lewis, M. W. Sullivan, and A. Vasen, "Making faces: Age and emotion differences in the posing of emotional expression," Developmental Psychology, 23, pp. 690-697, 1987.

[23] N. Takahashi and I. Daibo, "The facial expressions by emotion-instruction and photograph-instruction, and the effect of presence of another person in the facial expressions of anger and sadness," Japanese Journal of Interpersonal and Social Psychology, 3, pp. 65-72, 2003.

[24] F. Kameyama and Y. Muraoka, "An Interactive Design System of an attractive smile," J IEICE technical report. Human information processing, 95(554), pp. 15-22, 1996.

[25] Microsoft Emotion API, https://azure.microsoft.com/en-us/services/cognitive-services/emotion/, [retrieved: November 2019].

[26] Open CV, https://opencv.org/, [retrieved: November 2019].

Modelling Employee Resilience Using Wearables and Apps: A Conceptual Framework and Research Design

Herman de Vries, [1,2,3,7], Wim Kamphuis[2,8], Hilbrand Oldenhuis[1,9], Cees van der Schans[3,4,5,10] & Robbert Sanderman[3,6,11]
[1] Hanze University of Applied Sciences, Professorship Personalised Digital Health, Groningen, The Netherlands
[2] TNO, Department of Human Behaviour & Organisational Innovation, Soesterberg, The Netherlands
[3] Department of Health Psychology, University Medical Center Groningen, Groningen, The Netherlands
[4] Department of Rehabilitation Medicine, University Medical Center Groningen, Groningen, The Netherlands
[5] Hanze University of Applied Sciences, research group Healthy Ageing Allied Health Care and Nursing, Groningen, The Netherlands
[6] Deptartment of Psychology, Health and Technology, University of Twente, Enschede, The Netherlands
E-mail: [7] h.j.de.vries@pl.hanze.nl; [8] wim.kamphuis@tno.nl; [9] h.k.e.oldenhuis@pl.hanze.nl;
[10] c.p.van.der.schans@pl.hanze.nl; [11] r.sanderman@umcg.nl

***Abstract* – Occupational stress can cause health problems, productivity loss or absenteeism. Resilience interventions that help employees positively adapt to adversity can help prevent the negative consequences of occupational stress. Due to advances in sensor technology and smartphone applications, relatively unobtrusive self-monitoring of resilience-related outcomes is possible. With models that can recognize intra-individual changes in these outcomes and relate them to causal factors within the employee's context, an automated resilience intervention that gives personalized, just-in-time feedback can be developed. This paper presents the conceptual framework and methods behind the WearMe project, which aims to develop such models. A cyclical conceptual framework based on existing theories of stress and resilience is presented as the basis for the WearMe project. The operationalization of the concepts and the daily measurement cycle are described, including the use of wearable sensor technology (e.g., sleep tracking and heart rate variability measurements) and Ecological Momentary Assessment (mobile app). Analyses target the development of within-subject (n=1) and between-subjects models and include repeated measures correlation, multilevel modelling, time series analysis and Bayesian network statistics. Future work will focus on further developing these models and eventually explore the effectiveness of the envisioned personalized resilience system.**

Keywords – Occupational Stress; Personalized eHealth; Sensors; Wearables; Virtual Coaching.

I. INTRODUCTION

The *Wearables and app-based resilience Modelling in employees (WearMe)* project focuses on the mental resilience of employees with a stressful occupation [1]. Occupational stress can cause health problems, such as musculoskeletal disease, cardiovascular disease, depression and burnout [2]. Consequently, it can also lead to financial burdens due to treatment costs, productivity loss and absenteeism [3]. The cumulative wear and tear on bodily systems caused by stress is particularly detrimental for health and well-being [4]; this so-called 'allostatic load' increases the brain's sensitivity to appraise stimuli as threats and reduces resources to cope, which can result in a loss spiral [5].

Resilience can be defined as the process of positively adapting to adverse events [6]. It entails the use of individual (e.g., personality) and contextual (e.g., social support) resources to cope with adversity [7]. By utilizing these resources, resilient individuals are able to recover from the negative impact of stress relatively quickly and thus decrease their risk of negative long-term consequences.

Companies and institutions may offer resilience interventions to their employees to promote their health and employability and prevent stress-related problems. These interventions often target a broad population which unfortunately disregards the variability between employees. More personalized approaches might monitor for early signs of stress-related outcomes, link these to causal factors in the employee's own context, and provide personalized advice to sustain relevant resources that may prevent the aforementioned loss spiral. Due to advances in sensor technology and smartphone applications, relatively unobtrusive self-monitoring of changes in resilience related outcomes is increasingly possible [8]. While these advances open up the possibility of personalized monitoring in resilience interventions, models are needed to recognize intra-individual changes in these outcomes and relate these to causal factors and future consequences; this would allow for the opportunity to create automated resilience interventions that give personalized, just-in-time feedback, for employees to utilize in workplace applications.

In this paper, we present the conceptual framework and the study protocol of the ongoing WearMe project. After introducing the rationale behind the WearMe project in Section I, Section II describes a cyclical conceptual framework that is based on existing theories on stress and resilience. This framework represents the concepts and interrelations between concepts that we predict are necessary to model employee resilience. In Section III, we elaborate on how these concepts are operationalized in the WearMe Project, including the use of consumer-available wearables and an Ecological Momentary Assessment (EMA) app. Afterwards, we describe in Section IV the methods of the first WearMe study. Finally, Section V discusses possible directions for future work that can help develop predictive employee resilience models and personalized interventions.

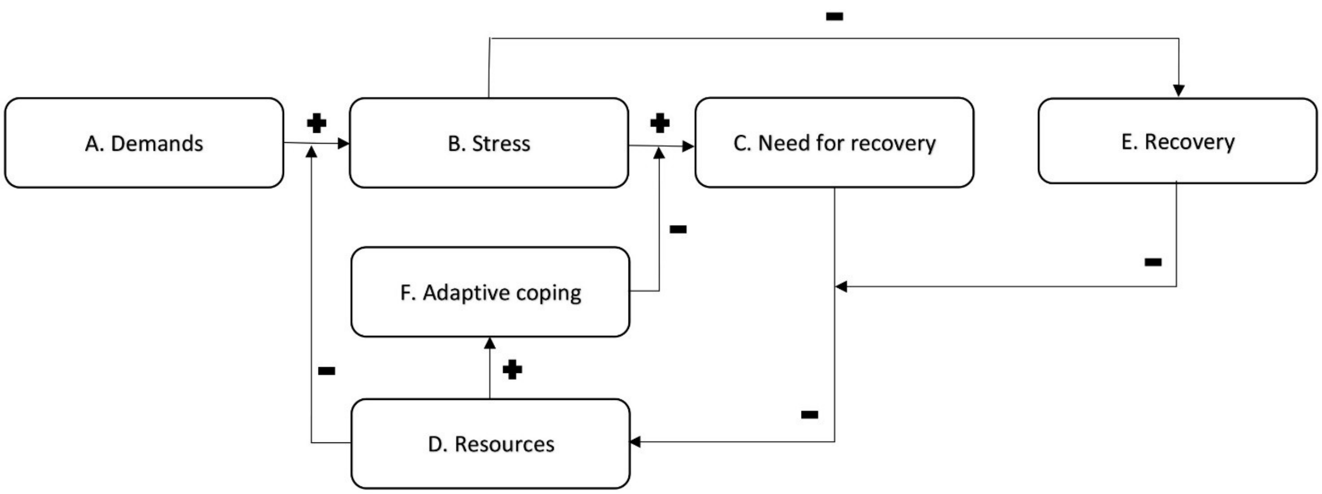


Figure 1. Conceptual framework for the WearMe study.

II. CONCEPTUAL FRAMEWORK

The conceptual framework of the WearMe project is presented in Figure 1. It illustrates our hypotheses on how the accumulation of the negative consequences of stress has a cyclical nature and how it can contribute to a loss spiral. This framework is based on the *Transactional Model of Stress and Coping* [9], the *Job Demands-Resources Model of Burnout* [10], the *Effort-Recovery Model* [11] and the *Conservation of Resources Theory* [5].

Stress accumulates when (job) *demands,* such as time pressure or physical workload, are appraised as threats due to inefficient available *resources* to *adaptively cope* with them [9]. Afterwards, an individual's *need for recovery*, characterized by feelings of exhaustion and reduced vigor to undertake new activities, depends on the individual's ability to utilize the available resources to adaptively cope with the demands [9][10]. A high need for recovery (i.e., little vigor to undertake activities), has a negative impact on an individual's resources to appraise and cope with new demands – unless there is sufficient *recovery* to alleviate this effect [11]. Aside from causing a perceived need for recovery, stress can also decrease sleep quality [12] and psychological detachment [13], which are aspects of *recovery* [14].

This framework's cyclical nature is supported by the Conservation of Resources theory [5], which states that initial loss of resources increases one's vulnerability to stress. Since additional resources are necessary to battle stress, this may lead to a depletion of resources or a loss spiral.

III. OPERATIONALIZATION

Based on the conceptual framework described above, we developed a measurement cycle to operationalize concepts using consumer-available wearables and an EMA smartphone application. All concepts are measured daily except adaptive coping—due to its highly context-specific nature which makes it difficult to quantify. In this section, we will first briefly present our daily measurement cycle. Following this, we will describe each concept and its operationalization.

The presented conceptual framework is not bounded by a specific timeframe. However, since the WearMe study particularly aims to investigate day-to-day and multi-day trends, we operationalized the concepts in a daily measurement cycle (Figure 2). For the daily measures, the WearMe study protocol utilizes: (1) a wrist-worn tracker for unobtrusive, continuous measurements throughout the day and night, (2) a Bluetooth chest strap and a smartphone application for a physiological measurement taken upon awakening and (3) a smartphone application for EMA questionnaires taken upon awakening and before bedtime.

A. Demands

Demands refer to the physical, social or organizational aspects that require sustained physical or mental effort and are therefore associated with certain physiological costs [15]. Participants' perceived daily demands are scored with the evening EMA questionnaire and is based on the self-composed diary question *"How demanding was your day?"*;

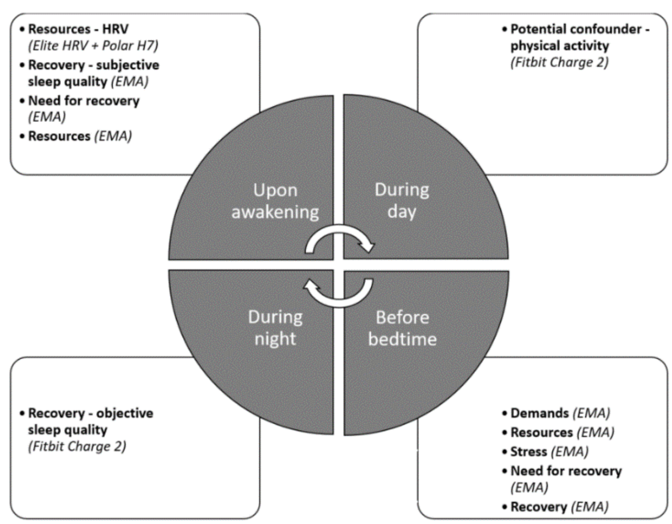


Figure 2. Measurement cycle of the WearMe study.

this is scored on an 11-point Numeric Rating Scale (NRS) that ranges from 0 (*"Not at all"*) to 10 (*"Extremely"*).

B. Stress

Participants' perceived total daily stress is scored in the evening EMA questionnaire with a validated single-item scale [16]: *"How much stress did you perceive today?"*. The question was rephrased to be applicable for daily use and the NRS that ranged from 1 (*"No stress"*) to 6 (*"Extreme stress"*) was adjusted to range from 0-10 for consistency.

C. Need for recovery

Need for recovery can be defined as a conscious emotional state and is connected with a temporal reluctance to continue with the present demands or to accept new demands; it is related to the depletion of resources following effort to meet certain demands [17]. The concept is characterized by a combination of perceiving high fatigue, as well as low vigor to undertake new activities. Participants' perceived fatigue is questioned in both the morning and evening EMA questionnaires to allow the calculation of within-day changes, while mental exhaustion is only measured during the evening. For fatigue, a validated single-item scale (*"How fatigued do you currently feel?"*) is used [18]. Item 3 of the Need For Recovery Scale is used to inquire mental exhaustion [19]: *"I felt mentally exhausted as a result of my activities"*. All items are scored on an 11-point NRS ranging from 0 (*"Not at all"* for fatigue and *"Strongly disagree"* for exhaustion) to 10 (*"Extremely"* for fatigue and *"Strongly agree"* for exhaustion).

D. Resources

According to the Job Demands-Resources model, job resources refer to physical, psychological, social or organizational aspects of a job that: (1) are functional in achieving work goals, (2) reduce job demands and the associated physiological and psychological costs and (3) stimulate personal growth, learning and development [10]. The resources in our conceptual framework can be seen as personal resources that enable an individual to better deal with stress. These resources include vigor, fitness, general self-efficacy (GSE), happiness, work engagement, and heart rate variability (HRV). Items for vigor, fitness, general self-efficacy (GSE) and happiness are included in both the morning and evening EMA questionnaires, and are all scored on an 11-point NRS ranging from 0 (*"Not at all"*) to 10 (*"Extremely"*). Below, the measured resources are described in more detail.

Vigor can be characterized by high levels of energy and mental resilience, the willingness to invest effort in one's work and persistence even in the face of difficulties [20]. Having high perceived vigor can therefore be seen as an individual resource during the appraisal of and coping with high demands. The item for vigor (measured in the morning and the evening) is based on an item of the vigor subscale of the Utrecht Work Engagement Scale (UWES) and rephrased for daily use in a neutral setting (*"Do you feel like undertaking activities?"*) [21]. Additionally, one item from the dedication subscale of the UWES is only included in the evening EMA questionnaire (*"Today, my activities were full of meaning and purpose."*) [21].

Fitness is also an individual resource for the appraisal of and coping with high demands; it is scored with a self-composed item that is similarly phrased to the fatigue item: *"How fit do you currently feel?"*. The item on fitness is included due to its more physical characteristics in comparison to the other items.

GSE is the belief in one's competence to tackle novel tasks and cope with adversity in a broad range of stressful or challenging encounters [22]. High GSE is associated with high optimism, self-regulation and self-esteem, and low depression and anxiety [22]; it can therefore be seen as an individual resource that is addressed during the appraisal of a stressor. The EMA item for GSE is based on the item with the highest factor loading (item 6) of the Generalized Self-Efficacy Scale and is rephrased for daily use: *"Do you feel capable of solving problems today?"*. During the evening, *"today"* is replaced with *"tomorrow"*.

Happiness is a state of well-being and contentment, characterized by frequent positive affect, high life satisfaction and infrequent negative affect [23]. Happiness has an inverse correlation with stress [24] and contributes to the psychological capital (resources) that may be key in better understanding the variation in perceived symptoms of stress [25]. Positive emotions like happiness can also predict increases in (trait) resilience and life satisfaction [26]. Participants' perceived happiness is scored using a validated single-item scale (*"Do you feel happy?"*) [27].

HRV refers to the variation in the inter-beat-intervals between heartbeats and is considered a proxy for autonomous nervous system functioning [28]. While HRV mostly serves as a parameter that illustrates physiological changes during acute stress, the resting HRV can remain decreased during and after acute stress [15][16]. In addition, having a lower resting HRV has been associated with increased sensitivity for stress [31], decreased emotion-regulation [32], decreased physical performance [33] and an increased risk of long-term physical or mental health problems [34]. In the WearMe study, resting HRV is therefore considered to be a potential indicator for the accumulation of stress, as well as an individual resource used in the appraisal of and coping with upcoming demands. Participants measure their resting HRV in the morning after waking up and before getting out of bed for 2 minutes in a supine position using the Elite HRV smartphone application [35] and a Polar H7 chest strap [36]. This aligns with existing standards that suggest a duration of 1-5 minutes under consistent circumstances with as little influence of circadian rhythms, meals, smoking, posture changes and significant mental or physical exertion [36][37]. We chose not to apply guided breathing, as respiratory rate influences HRV [38][39], and we intend to measure the natural resting state of the participant. The exported inter-beat-interval data are analysed using Kubios Premium software, version 3.1.0 [41]. Our analyses will focus on a time-domain outcome called Root Mean Square of the Successive Differences (RMSSD).

E. Recovery

Recovery refers to the recuperation from potential load effects after the exposure to certain demands [11]. The concept of recovery consists of two components that are known to limit the spillover of a perceived need for recovery from the previous day to the next day: (1) sleep and (2) being able to psychologically detach from work during leisure time [42]. Since stress is known to have a negative effect on sleep quality [12] and psychological detachment [13], deteriorated sleep and psychological detachment are also considered to be potential indicators for the accumulation of the negative consequences of stress. Sleep deprivation contributes to the accumulation of allostatic load [42][43], but also attenuates the relationship between negative affect experienced at work and negative affect in the next morning [45]. Sleep is therefore an important component in the recovery from (work-related) stress and helps limit the potential loss of resources.

Detachment is measured with an item from the psychological detachment subscale of the Recovery Experience Questionnaire that had the highest average correlation to the other three included subscale questions [14]: *"During my off-job time, I distanced myself from my work"*. Additionally, the perceived availability of time to recover throughout the day is measured based on an item used in a prior study [17]: *"Today I had enough time to relax and recover from work"*. Both items are included in the evening EMA questionnaire and scored on an 11-point NRS ranging from 0 (*"Strongly disagree"*) to 10 (*"Strongly agree"*).

The Fitbit Charge 2 wrist-worn tracker is used to objectively measure the total sleep time and sleep efficiency. Additionally, the subjective sleep quality is measured in the morning EMA questionnaire with a validated single-item [47]: *"How was the quality of your sleep?"* and is scored on an 11-point NRS ranging from 0 (*"Worst possible sleep"*) to 10 (*"Best possible sleep"*).

F. Other

In order to account for potentially confounding effects and explain relevant variance, two other variables are included in the daily measures: (1) alcohol intake and (2) physical activity. Alcohol intake is associated with a lower resting HRV [48], but is sometimes also used as a strategy to cope with increased stress [49]. Alcohol intake is therefore measured during the morning EMA questionnaire by asking for the number of alcoholic beverages that the participant consumed during the previous day. While the absolute amount of alcohol in different types of beverages may deviate, asking for the number of alcoholic beverages consumed is both convenient for daily inquiry and consistent with the widely used AUDIT-C questionnaire [50]. Finally, physical activity (steps, sedentary minutes, minutes of moderate-to-vigorous physical activity) is measured throughout the day using the Fitbit Charge 2 [51]. Physical activity levels are associated with decreased stress reactivity [52], a higher resting HRV [53] and improved sleep [54]; therefore, physical is a potential confounder.

IV. PRESENT STUDY

The first WearMe study aims to test the usability of the described measurement protocol, as well as to gather a first wave of data to be able to test the hypothesized relations in the conceptual model. Additionally, the development of both intra-individual and population models will be explored. The study protocol was approved by the ethical committee of the Hanze University of Applied Sciences Groningen (heac.2018.008).

A. Population

For the first WearMe study, students who are starting their first full-time internship for Social Work and Applied Psychology are invited to participate. We anticipate this population to be at risk of experiencing stress due to the potentially stressful nature of these disciplines and the fact that these are the first full-time internships in the participants' curriculum. The students need to own an Android or iOS smartphone in order to participate. For recruitment, a message is placed on the school's digital learning environment and the students who are scheduled for their first internships receive an e-mail. Participation in the study is voluntary. In order to facilitate recruitment and optimize adherence during participation, participants who collect at least 80% valid data points are rewarded with a €25 gift voucher. Additionally, participants who collect enough data to create intra-individual models receive individual feedback. Since this first WearMe study is exploring a new topic, it was impossible to perform an accurate power calculation based on the considered data-analysis methods (paragraph IV.C). Due to the availability of materials, a maximum of 15 participants can be simultaneously recruited. Therefore, the recruitment and data-collection processes are divided over two waves. The first recruitment wave started in September 2018, whereas the second waive started in September 2019.

B. Data collection

The total data collection period is 15 weeks, targeting a maximum of 105 full days of data per participant. The operationalization of the conceptual model and items included in the EMA questionnaires are described in Section III. The participants use a Polar H7 Bluetooth chest strap in combination with the Elite HRV smartphone application to measure their resting HRV upon awakening and used a Fitbit Charge 2 wrist-worn tracker to continuously measure their physical activity and sleep. In order to collect the subjective EMA questionnaire data, TNO's self-developed *"How am I?"* smartphone application is used. Participants are instructed to fill in their morning EMA questionnaire (7 items) after measuring their resting HRV and fill in their evening EMA questionnaire (12 items) before going to bed. The morning questionnaire is available between 06:00 and 15:00 and the evening questionnaire is available between 21:00 and 06:00 in order to offer participants a broad window to fill in the questionnaires (e.g., when potentially staying up late or sleeping in during weekends). Additionally, participants receive smartphone notifications as reminders at 06:00 for the morning questionnaires and at 21:00 for the evening questionnaires. Where available, validated Dutch versions of

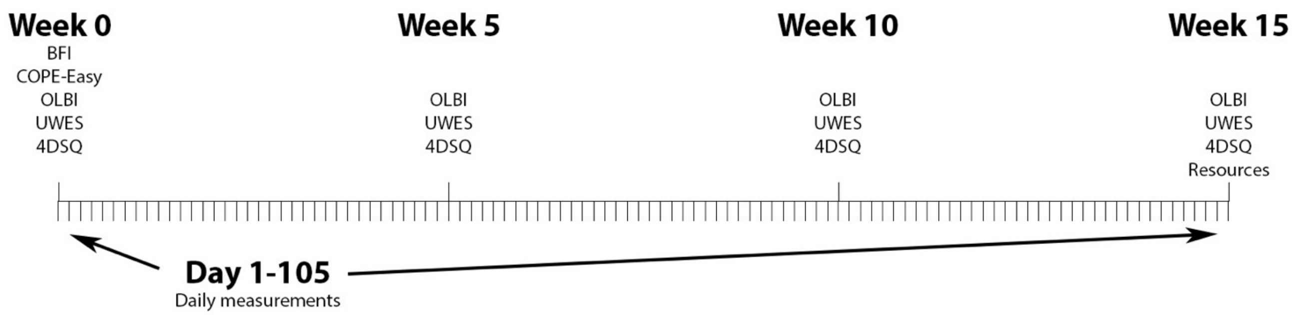


Figure 3. The WearMe study timeline.

the questionnaires described in Section III are used. Items based on questionnaires that were only available in English were translated into Dutch. For validation of these items, backwards translation by a native English speaker was performed. No differences that significantly changed the meaning of the items were found during this process.

The daily measurements described in Section II consisted of concepts that can vary on a day-to-day basis. However, some of the concepts of the conceptual framework included aspects that are more trait-like (e.g., personality traits as potential resources or preferred coping strategies) or could be expected to vary over a longer timeframe (e.g., burnout, depression). Therefore, several full questionnaires are administered to benefit the development of population models using between-subject analyses: a questionnaire on personality traits (the *Big Five Inventory*; BFI) [55], coping strategies (the *COPE-Easy*) [56], burnout (the *Oldenburg Burnout Inventory*; OLBI) [57], work engagement (the *Utrecht Work Engagement Scale*; UWES) [20] and symptoms of somatization, distress, depression and anxiety (the *Four-Dimensional Symptom Questionnaire*; 4DSQ) [58]. The questionnaires on burnout, work engagement and symptoms of somatization, distress, depression and anxiety are also administered after 5, 10 and 15 weeks. Finally, after 15 weeks, participants fill out a resources questionnaire to retrospectively assess the perceived personal and environmental resources throughout the internships, since participants are not able to accurately assess the environmental resources prior to or at the beginning of their internship. This resources questionnaire was inspired by resources questionnaires that were developed for other domain-specific work environments [58][59] and adjusted to better align with the participants' internship contexts. Additionally, the distributed questionnaires consisted of items that were derived from existing validated questionnaires such as the *Life Orientation Test* [61], the *Connor Davidson Resilience Scale* [62] and the *Dispositional Resilience Scale* [63]. Figure 3 illustrates the timeline for the measurements in the first WearMe study.

C. Data analysis

Several approaches to data-analysis will be explored. First, the hypotheses formulated in the conceptual framework that were introduced in Section II will be tested using within-day relations and, if possible, on multi-day trends. The repeated measures correlation technique as described by Bakdash and Marusich [64] will be used to analyze the correlation between two variables while taking into account that data points are repeated measures within participants. Random intercept, fixed slopes multilevel modelling will be applied when two or more variables within a specific concept or potential confounders are included to predict the variance within a single dependent variable. Both methods allow the scores between participants to differ (random intercepts), but explore a fixed effect between the two variables (fixed slopes). We anticipate that there will be insufficient data available to explore whether the effect between the included variables differ between participants (random slopes).

Second, we will explore the development of intra-individual (n=1) models for within-day and, if possible, multi-day trends using the data of the participants with the highest adherence. Aside from the aforementioned techniques, the use of time series analysis techniques and Bayesian statistics will be considered for the multi-day trend analyses.

Finally, the data of the full questionnaires will be used to explore (1) if trends in relevant daily outcomes like sleep, resting HRV and the presence of resources and need for recovery can be predicted based on personality traits or preferred coping strategies measured at baseline, (2) if these trends are also predictive for changes in burnout, work engagement and symptoms questionnaires and (3) if there is an association between the daily measured state-related variables (e.g., individual resources and perceived stress) and the trait-variables measured at baseline (the personality traits and preferred coping strategies).

V. CONCLUSION AND FUTURE WORK

This article presented the conceptual framework for the WearMe project and a detailed description of the operationalization of these concepts in the first (ongoing) WearMe study. Data collected with a wrist-worn wearable tracker, a Bluetooth chest-strap and a smartphone EMA questionnaire app on a daily will be used to explore if the hypotheses that are presented in the conceptual framework are indeed supported.

When the results affirm that tracking sleep and resting HRV with the use of consumer wearables is feasible and can be useful in resilience modelling, the current models will be expanded. Future studies will therefore focus on the development of predictive models that allow early detection of stress-related symptoms. In addition, expanding the current model by using additional consumer-available wearables or apps that can unobtrusively collect potentially relevant data (e.g., GPS location, calendar events) may be explored. When our conceptual framework is validated, a more inductive approach to data-analysis may also be explored (e.g., using machine learning) to increase the explained variance of the individual models. If successful, these models can be implemented in applications that create personalized feedback on how to cope with demands or limit the loss of relevant resources, which may help employees optimize their resilience.

Furthermore, it is likely that the development of within-subject models requires a long period of data collection. This means that in the envisioned automated resilience system, an individual will have to collect data for a relatively long period before receiving personalized feedback. The creation of a classification algorithm and the identification of subgroups with similar outcome trajectories using between-subject analyses of baseline and first-week data in a larger sample might allow for the development of a system that combines both methods [65]. In such a system, participants could receive semi-personalized feedback early on based on their subgroup classification and receive fully personalized feedback when enough within-subject data are available. Such a method would be a compromise between deductive methods that test assumptions based on existing knowledge and inductive methods that allow specific intra-individual predictors to be included in even more personalized feedback.

ACKNOWLEDGMENTS

The authors thank Dr. Heather Young (TNO) for performing the backwards translations of the EMA items and Tamar Schaap (TNO) for proofreading the article.

REFERENCES

[1] H. de Vries, W. Kamphuis, H. Oldenhuis, C. van der Schans, and R. Sanderman, ‘Wearable and App-based Resilience Modelling in Employees (WearMe)’, presented at the The Eleventh International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED), Athens, 2019.

[2] T. W. Colligan and E. M. Higgins, ‘Workplace stress: Etiology and consequences’, *J. Workplace Behav. Health*, vol. 21, no. 2, pp. 89–97, 2006.

[3] S. Béjean and H. Sultan-Taïeb, ‘Modeling the economic burden of diseases imputable to stress at work’, *Eur. J. Health Econ.*, vol. 6, no. 1, pp. 16–23, 2005.

[4] B. S. McEwen, ‘Stress, adaptation, and disease: Allostasis and allostatic load’, *Ann. N. Y. Acad. Sci.*, vol. 840, no. 1, pp. 33–44, 1998.

[5] S. E. Hobfoll, ‘The Influence of Culture, Community, and the Nested-Self in the Stress Process: Advancing Conservation of Resources Theory’, *Appl. Psychol.*, vol. 50, no. 3, pp. 337–421, 2001.

[6] T. D. Cosco, A. Kaushal, R. Hardy, M. Richards, D. Kuh, and M. Stafford, ‘Operationalising resilience in longitudinal studies: a systematic review of methodological approaches.’, *J. Epidemiol. Community Health*, vol. 71, no. 1, pp. 98–104, Jan. 2017.

[7] T. W. Britt, W. Shen, R. R. Sinclair, M. R. Grossman, and D. M. Klieger, ‘How Much Do We Really Know About Employee Resilience?’, *Ind. Organ. Psychol.*, vol. 9, no. 02, pp. 378–404, 2016.

[8] R. L. Drury, ‘Wearable biosensor systems and resilience: a perfect storm in health care?’, *Front. Psychol.*, vol. 5, p. 853, 2014.

[9] R. S. Lazarus and S. Folkman, ‘Transactional theory and research on emotions and coping’, *Eur. J. Personal.*, vol. 1, no. 3, pp. 141–169, 1987.

[10] A. B. Bakker and E. Demerouti, ‘The Job Demands-Resources model: state of the art’, *J. Manag. Psychol.*, vol. 22, no. 3, pp. 309–328, 2007.

[11] M. Van Veldhoven, ‘Need for recovery after work: An overview of construct, measurement and research’, in *Occupational health psychology: European perspectives on research, education and practice*, vol. 3, de J. Houdmont and S. Leka, Eds. Nottingham, 2008, pp. 1–25.

[12] E.-J. Kim and J. E. Dimsdale, ‘The effect of psychosocial stress on sleep: a review of polysomnographic evidence.’, *Behav. Sleep. Med.*, vol. 5, no. 4, pp. 256–278, 2007.

[13] S. Sonnentag, I. Kuttler, and C. Fritz, ‘Job stressors, emotional exhaustion, and need for recovery: A multi-source study on the benefits of psychological detachment’, *J. Vocat. Behav.*, vol. 76, no. 3, pp. 355–365, 2010.

[14] S. Sonnentag and C. Fritz, ‘The Recovery Experience Questionnaire: development and validation of a measure for assessing recuperation and unwinding from work.’, *J. Occup. Health Psychol.*, vol. 12, no. 3, p. 204, 2007.

[15] E. Demerouti, F. Nachreiner, A. B. Baker, and W. B. Schaufeli, ‘The Job Demand-Resources Model of Burnout’, *J. Appl. Psychol.*, vol. 86, no. 3, pp. 499–512, 2001.

[16] A. J. Littman, E. White, J. A. Satia, D. J. Bowen, and A. R. Kristal, ‘Reliability and validity of 2 single-item measures of psychosocial stress’, *Epidemiology*, pp. 398–403, 2006.

[17] S. Sonnentag and F. R. H. Zijlstra, ‘Job characteristics and off-job activities as predictors of need for recovery, well-being, and fatigue.’, *J. Appl. Psychol.*, vol. 91, no. 2, p. 330, 2006.

[18] M. L. M. Van Hooff, S. A. E. Geurts, M. A. J. Kompier, and T. W. Taris, ‘“How fatigued do you currently feel?” Convergent and discriminant validity of a single-item fatigue measure’, *J. Occup. Health*, vol. 49, no. 3, pp. 224–234, 2007.

[19] M. Van Veldhoven and S. Broersen, ‘Measurement quality and validity of the “need for recovery scale”’, *Occup. Environ. Med.*, vol. 60, no. suppl 1, pp. i3–i9, 2003.

[20] W. B. Schaufeli, A. B. Bakker, and M. Salanova, ‘The measurement of work engagement with a short questionnaire: A cross-national study’, *Educ. Psychol. Meas.*, vol. 66, no. 4, pp. 701–716, 2006.

[21] W. B. Schaufeli, M. Salanova, V. González-Romá, and A. B. Bakker, ‘The measurement of engagement and burnout: A two sample confirmatory factor analytic approach’, *J. Happiness Stud.*, vol. 3, no. 1, pp. 71–92, 2002.

[22] A. Luszczynska, B. Gutiérrez-Doña, and R. Schwarzer, ‘General self-efficacy in various domains of human functioning: Evidence from five countries’, *Int. J. Psychol.*, vol. 40, no. 2, pp. 80–89, 2005.

[23] S. Lyubomirsky, K. M. Sheldon, and D. Schkade, ‘Pursuing happiness: the architecture of sustainable change.’, *Rev. Gen. Psychol.*, vol. 9, no. 2, p. 111, 2005.

[24] H. H. Schiffrin and S. K. Nelson, ‘Stressed and happy? Investigating the relationship between happiness and perceived stress’, *J. Happiness Stud.*, vol. 11, no. 1, pp. 33–39, 2010.

[25] J. B. Avey, F. Luthans, and S. M. Jensen, ‘Psychological

capital: A positive resource for combating employee stress and turnover', *Hum. Resour. Manage.*, vol. 48, no. 5, pp. 677–693, 2009.

[26] M. A. Cohn, B. L. Fredrickson, S. L. Brown, J. A. Mikels, and A. M. Conway, 'Happiness unpacked: positive emotions increase life satisfaction by building resilience.', *Emotion*, vol. 9, no. 3, p. 361, 2009.

[27] A. M. Abdel-Khalek, 'Measuring happiness with a single-item scale', *Soc. Behav. Personal. Int. J.*, vol. 34, no. 2, pp. 139–150, 2006.

[28] J. F. Thayer, F. Ahs, M. Fredrikson, J. J. Sollers III, and T. D. Wager, 'A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health.', *Neurosci. Biobehav. Rev.*, vol. 36, no. 2, pp. 747–756, Feb. 2012.

[29] M. Hall *et al.*, 'Acute stress affects heart rate variability during sleep.', *Psychosom. Med.*, vol. 66, no. 1, pp. 56–62, Feb. 2004.

[30] E. Hynynen, N. Konttinen, U. Kinnunen, H. Kyrolainen, and H. Rusko, 'The incidence of stress symptoms and heart rate variability during sleep and orthostatic test.', *Eur. J. Appl. Physiol.*, vol. 111, no. 5, pp. 733–741, May 2011.

[31] G. Park, J. J. Van Bavel, M. W. Vasey, and J. F. Thayer, 'Cardiac vagal tone predicts inhibited attention to fearful faces.', *Emot. Wash. DC*, vol. 12, no. 6, pp. 1292–1302, Dec. 2012.

[32] D. P. Williams, C. Cash, C. Rankin, A. Bernardi, J. Koenig, and J. F. Thayer, 'Resting heart rate variability predicts self-reported difficulties in emotion regulation: a focus on different facets of emotion regulation.', *Front. Psychol.*, vol. 6, p. 261, 2015.

[33] S. Jimenez Morgan and J. A. Molina Mora, 'Effect of Heart Rate Variability Biofeedback on Sport Performance, a Systematic Review.', *Appl. Psychophysiol. Biofeedback*, vol. 42, no. 3, pp. 235–245, Sep. 2017.

[34] D. Liao, M. Carnethon, G. W. Evans, W. E. Cascio, and G. Heiss, 'Lower heart rate variability is associated with the development of coronary heart disease in individuals with diabetes: the atherosclerosis risk in communities (ARIC) study', *Diabetes*, vol. 51, no. 12, pp. 3524–3531, 2002.

[35] A. S. Perrotta, A. T. Jeklin, B. A. Hives, L. E. Meanwell, and D. E. R. Warburton, 'Validity of the Elite HRV Smartphone Application for Examining Heart Rate Variability in a Field-Based Setting', *J. Strength Cond. Res.*, vol. 31, no. 8, pp. 2296–2302, 2017.

[36] D. J. Plews, B. Scott, M. Altini, M. Wood, A. E. Kilding, and P. B. Laursen, 'Comparison of heart-rate-variability recording with smartphone photoplethysmography, Polar H7 chest strap, and electrocardiography', *Int. J. Sports Physiol. Perform.*, vol. 12, no. 10, pp. 1324–1328, 2017.

[37] D. S. Quintana, G. A. Alvares, and J. A. J. Heathers, 'Guidelines for Reporting Articles on Psychiatry and Heart rate variability (GRAPH): recommendations to advance research communication', *Transl. Psychiatry*, vol. 6, p. e803, May 2016.

[38] M. Malik *et al.*, 'Heart rate variability: Standards of measurement, physiological interpretation, and clinical use', *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.

[39] J. W. Denver, S. F. Reed, and S. W. Porges, 'Methodological issues in the quantification of respiratory sinus arrhythmia', *Biol. Psychol.*, vol. 74, no. 2, pp. 286–294, 2007.

[40] D. S. Quintana and J. A. J. Heathers, 'Considerations in the assessment of heart rate variability in biobehavioral research', *Frontiers in Psychology*, vol. 5. p. 805, 2014.

[41] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, and P. A. Karjalainen, 'Kubios HRV--heart rate variability analysis software.', *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 210–220, 2014.

[42] S. Sonnentag and C. Binnewies, 'Daily affect spillover from work to home: Detachment from work and sleep as moderators', *J. Vocat. Behav.*, vol. 83, no. 2, pp. 198–208, 2013.

[43] B. S. McEwen and I. N. Karatsoreos, 'Sleep deprivation and circadian disruption: stress, allostasis, and allostatic load', *Sleep Med. Clin.*, vol. 10, no. 1, pp. 1–10, 2015.

[44] B. S. McEwen, 'Sleep deprivation as a neurobiologic and physiologic stressor: allostasis and allostatic load', *Metabolism*, vol. 55, pp. S20–S23, 2006.

[45] S. Sonnentag and C. Binnewies, 'Daily affect spillover from work to home: Detachment from work and sleep as moderators', *J. Vocat. Behav.*, vol. 83, no. 2, pp. 198–208, 2013.

[46] M. de Zambotti, A. Goldstone, S. Claudatos, I. M. Colrain, and F. C. Baker, 'A validation study of Fitbit Charge 2^{TM} compared with polysomnography in adults', *Chronobiol. Int.*, vol. 35, no. 4, pp. 465–476, 2018.

[47] J. C. Cappelleri, A. G. Bushmakin, A. M. McDermott, A. B. Sadosky, C. D. Petrie, and S. Martin, 'Psychometric properties of a single-item scale to assess sleep quality among individuals with fibromyalgia', *Health Qual. Life Outcomes*, vol. 7, no. 1, p. 54, 2009.

[48] D. S. Quintana, A. J. Guastella, I. S. McGregor, I. B. Hickie, and A. H. Kemp, 'Moderate alcohol intake is related to increased heart rate variability in young adults: Implications for health and well-being', *Psychophysiology*, vol. 50, no. 12, pp. 1202–1208, 2013.

[49] M. L. Cooper, M. Russell, and W. H. George, 'Coping, expectancies, and alcohol abuse: A test of social learning formulations.', *J. Abnorm. Psychol.*, vol. 97, no. 2, p. 218, 1988.

[50] K. Bush, D. R. Kivlahan, M. B. McDonell, S. D. Fihn, and K. A. Bradley, 'The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking', *Arch. Intern. Med.*, vol. 158, no. 16, pp. 1789–1795, 1998.

[51] I. A. Figueroa, N. D. Lucio, J. L. Gamez Jr, V. E. Salazar, and M. D. Funk, 'Validity of Daily Physical Activity Measurements of Fitbit Charge 2', in *International Journal of Exercise Science: Conference Proceedings*, 2018, vol. 2, no. 10, p. 27.

[52] K. R. Fox, 'The influence of physical activity on mental well-being', *Public Health Nutr.*, vol. 2, no. 3a, pp. 411–418, 1999.

[53] K. L. Rennie, H. Hemingway, M. Kumari, E. Brunner, M. Malik, and M. Marmot, 'Effects of moderate and vigorous physical activity on heart rate variability in a British study of civil servants', *Am. J. Epidemiol.*, vol. 158, no. 2, pp. 135–143, 2003.

[54] D. L. Sherrill, K. Kotchou, and S. F. Quan, 'Association of physical activity and human sleep disorders', *Arch. Intern. Med.*, vol. 158, no. 17, pp. 1894–1898, 1998.

[55] J. J. A. Denissen, R. Geenen, M. A. G. Van Aken, S. D. Gosling, and J. Potter, 'Development and validation of a Dutch translation of the Big Five Inventory (BFI)', *J. Pers. Assess.*, vol. 90, no. 2, pp. 152–157, 2008.

[56] W. C. Kleijn, G. L. Van Heck, and A. Van Waning, 'Ervaringen met een Nederlandse bewerking van de COPE copingvragenlijst: De COPE-Easy', *Gedrag Gezondh.*, vol. 28, pp. 213–226, 2000.

[57] J. R. B. Halbesleben and E. Demerouti, 'The construct validity of an alternative measure of burnout: Investigating the English translation of the Oldenburg Burnout Inventory', *Work Stress*,

vol. 19, no. 3, pp. 208–220, 2005.

[58] B. Terluin *et al.*, 'The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization', *Bmc Psychiatry*, vol. 6, no. 1, p. 34, 2006.

[59] R. (TNO) Delahaij, O. (TNO) Binsch, and W. (TNO) Kamphuis, 'Weerbaarheidsmonitor voor de politie', TNO, Soesterberg, M10280, 2012.

[60] R. (TNO) Delahaij, W. (TNO) Kamphuis, O. (TNO) Binsch, and W. (TNO) Venrooij, 'Ontwikkeling Militaire Resilience Monitor', TNO, Soesterberg, R11652, 2015.

[61] M. F. Scheier, C. S. Carver, and M. W. Bridges, 'Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test.', *J. Pers. Soc. Psychol.*, vol. 67, no. 6, p. 1063, 1994.

[62] K. M. Connor and J. R. Davidson, 'Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC)', *Depress. Anxiety*, vol. 18, no. 2, pp. 76–82, 2003.

[63] P. T. Bartone, R. J. Ursano, K. M. Wright, and L. H. Ingraham, 'The impact of a military air disaster on the health of assistance workers', *J. Nerv. Ment. Dis.*, vol. 177, no. 6, pp. 317–328, 1989.

[64] J. Z. Bakdash and L. R. Marusich, 'Repeated Measures Correlation', *Front. Psychol.*, vol. 8, pp. 456–456, Apr. 2017.

[65] G. Spanakis, G. Weiss, B. Boh, L. Lemmens, and A. Roefs, 'Machine learning techniques in eating behavior e-coaching: Balancing between generalization and personalization', *Pers. Ubiquitous Comput.*, vol. 21, no. 4, pp. 645–659, 2017.

Human Pose Estimation, Anthropomorphism and Gamification in the Promotion of Physical Activity Among Breast Cancer Survivors

João P. Monteiro
Carolina T. Lopes
Nuno C. Duarte

INESC TEC / Faculdade de Engenharia
Universidade do Porto
Rua Dr. Roberto Frias s/n
4200-465 Porto, Portugal
jpsm@ieee.org
ee11101@fe.up.pt
up201502854@fe.up.pt

André T. Magalhães

Serviço de Cirurgia Geral
Centro Hospitalar de São João
Alameda Prof. Hernâni Monteiro
4200-319 Porto, Portugal
amag1976@gmail.com

Hélder P. Oliveira

INESC TEC / Faculdade de Ciências
Universidade do Porto
Rua Campo Alegre 1021/1055
4169-007 Porto, Portugal
helder.f.oliveira@inesctec.pt

Abstract—**As breast cancer survivors are living longer, the adverse effects impacting quality of life resulting from the cancer treatment are more frequent. Concurrently, an emergence of new technologies has led to major changes in society with hypothesized potential to automate and assist in otherwise time-consuming tasks while promoting engagement. In this context, technology from the video games industry has been used with emphasis in the recovery and follow-up stages to evaluate and motivate the patient after treatment. The present work aims to evaluate a set of graphical user interfaces that make use of data acquired with a colour and depth sensor to monitor and provide real-time feedback to a given user. Design guidelines from serious games are explored within the context of developing a system aid for physical follow-up care in the form of a set of exercises selected by the medical community. The proposed interfaces were evaluated in a clinical setting with a group of breast cancer survivors and the expressed preferences collected. Results are discussed under the light of future developments of anthropomorphization and gamification as means to promote engagement.**

Keywords*–*Computer vision; Computerized monitoring; Computer aided instruction; User interfaces; Medical services.

I. INTRODUCTION

As previously observed [1], while contributing for improved overall survivorship, contemporary breast cancer treatment techniques may result in several impairments in women's upper-body function and, consequently, contribute to a decreased quality of life [2] that could potentially be monitored through the application of readily available low cost computer vision based sensing approaches.

On the subject of the adverse effects resulting from the cancer treatment, upper body morbidity (e.g., decreased range of motion, muscle strength, pain and lymphedema) can be recognized to be among the most prevalent side effects. Regarding lymphedema alone, a swelling condition resulting from lymphatic ablation commonly associated with breast cancer treatment, it has been estimated that over 1 million Breast Cancer Survivors (BCS) in the United States of America (USA) and 10 million women worldwide may meet the criteria for breast cancer-related variant of the condition [2]. Persistent postsurgical pain is also an increasingly documented problem, negatively impacting quality of life and affecting approximately 20% of new chronic pain patients. The reported incidence of persistent post mastectomy pain (PPMP) ranges from 25-60%, in an estimated total of around 2.5 million survivors in the United States. Among breast cancer patients, PPMP is rated as the most troubling symptom, leading to disability and psychological distress, and is notably resistant to management. While surgical factors, including more extensive surgery (total or partial mastectomy), axillary lymph node dissection and reconstruction have been suggested to serve as important risk factors for chronic pain, several studies do not support this association [3]. Adjuvant treatment, such as radiation, chemotherapy, and hormone therapy, has also been associated with persistent pain. Among demographic factors, younger age correlates with increased persistent pain incidence in some studies but not others. Pre-existing pain is also more frequent in those who go on to develop PPMP [4]. The most commonly cited theory for post mastectomy pain syndrome is the removal of the intercostobrachial nerves that run through the axillary region into the arm which provokes chronic post-operative pain in breast cancer patients, followed by chemotherapy and radiation therapy. The treatment involves physical therapy, topical agents, anticonvulsants, antidepressants, antiarrhythmic, nerve block and scar desensitization injections with dilute local anaesthesia and steroids [5].

While the assessment of the oncological outcome of the cancer treatment can be easily objectively quantified by disease-free and overall survival rates, the same does not hold for functional aspects closely related to quality of life within the target population. Assessment of BCS symptoms and health-related quality of life outcomes are usually made using Patient Reported Outcome (PRO) questionnaires that quantify significant outcome variables from the patient's perspective [2]. A prospective surveillance model for BCS has been proposed, highlighting the importance of monitoring for functional and physical impairment commonly associated with treatment [6]. Despite available methods for monitoring and assessing, an integrated approach able to achieve early detection, promote risk-reduction and self-management, while engaging the user in an adequate follow-up strategy, is still considered missing [7].

In Section II, an outline of topics related to the application of typical elements of game playing is presented in order to contextualize the proposed methodology, that is presented in Section III. The paper continues with a discussion of results in Section IV regarding key questions relating to the application of strategies of anthropomorphization and gamification as means to promote engagement to particular physical activities within the context of patient empowerment systems. Lastly, a discussion on future developments is presented in Section V.

II. Related Work

Engaging patients in their healthcare can be recognized as a paramount topic that has evolved through time also as a reflection of specific technological and societal contexts [8]. In this sense, growing trends of the quantified-self movement, personal health records tools dissemination and interactive video games that combine physical exercise with game-play and have a primary purpose other than entertainment present themselves as currently active research lines.

A. Gamification

Physical activity promotion programmes tested in patients with disabilities and impairment problems demonstrate that patients' functionality can improve with an intensive training split that is contextualised and oriented as a pursuit in the achievement of a well defined goal. However, this task division is prone to present a major set-back, which is the lack of interest of the patient in performing repetitive tasks [11].

On the other hand, it is possible to note that a game, overall, aims to offer the player a challenge of a physical or/and mental nature that can be completed using a set of rules, being able to install feelings of amusement or entertainment in the participant while returning feedback in a form of grades or scores, while possibly unlocking new challenges based on the feedback received. Video games have the same goals, only a computer is used as an intermediary [12].

The concept of serious games is one that is hard to define, but it usually refers to games used for training, advertising, simulation or education. A particular example of such a gamified approach, commonly referred to as exergaming or exergames, can be described as a type of video game, or multimedia interaction that requires the player to physically move in order to play [13]. With the evolution of video game acceptance by the general public, serious games have begun to surge, spreading into healthcare where they can eventually provide a more personalized experience to users, improving not just physical, but also mental aspects of care. This surge, and the evolution of visual computing, seems to enable the development of personalized home systems, which could objectively evaluate the patient's state, while motivating for continued physical activity [14]. Specifically for rehabilitation, game prototypes have been tested for specific circumstances, in particular scenarios, such as upper limb rehabilitation [15].

The usage of games as a rehabilitation tool is a rather young topic especially taking into consideration that these usually tend to depend on virtual/augmented reality and low-cost effective equipment that has only begun being available relatively recently, with early examples including applications of devices such as the Playstation EyeToy dating back to 2003. A selection of reference works related to rehabilitation games is briefly reviewed and presented in the following list:

- Esfahlani et al. [16], evaluated a system aiming to monitor kinematic activity of upper and lower limbs by using a group of capture devices (Xbox Kinect, Mya armband and Rudder Pedal), to create a game in which levels are proposed taking into account current and expected abilities of the user.
- Caurin et al. [17], tested a dynamic difficulty adaptation game, using an adapted version of the Pong game playable with a wrist rehabilitation system as an input mechanic for patients with motor deficiencies.
- Barzilay, et al. [18], proposed the usage of neural networks paired with a virtual reality platform, composed of a Vicon^{TM} motion capture system and a wireless Aurion^{TM} surface EMG ZeroWire, to create a neuromotor training system for upper-limb rehabilitation that proposes exercises based on the feedback from patient's initial usage as well as therapist input.
- Darzi et al. [19], studied a system that regulates the difficulty of a game by analysing the physiological responses of the users by measuring respiration and electromyography signals from the posterior deltoid.
- Ma et al. [20], assessed the use of a Microsoft Kinect in terms of quantification of maximum range for hand movement, peak velocity and mean velocity, through its integration in a rehabilitation game, validating it by comparing it to a Vicon^{TM} motion capture system.

From these latter mentioned studies we can observe that multiple capture devices can be used, although some of them represent not just custom made solutions but also high-priced solutions, with the Microsoft Kinect being identified as a relative accurate device at a relative low-cost price.

On the other hand, it is also possible to find multiple examples of studies more focused on the usability of such games and how the patients reacted to them and elements that should be taken into consideration when developing such games. Some examples are listed bellow:

- Alankus et al. [21], created multiple games that used two Nintendo Wii remotes attached to the user's arm in order to detect elbow and arm movement. The study mainly contributed by studying a set of game design elements proposed to be considered when attempting to create rehabiliation games.
- Seo et al. [22], measured stroke rehabilitation patients' expectations for virtual rehabilitation games before these engaging in three different games. After the gaming experience the patients are again asked to answer a survey in which they evaluate the games. The games were developed using the Microsoft Kinect and P5 Glove MIDI as capture devices.
- Burke et al. [23], conducted a study in which multiple games were tested, using different capture devices, such as the P5 Glove MIDI, Nintendo Wii remote or off-the-shelf webcams. The study focused on the usability of these games on able-bodied users before conducting it on stroke rehabilitation patients. This secondary study [24] eventually occurred in which the webcam games were tested at home by the stroke rehabilitation patients, proving to be successful both in usability and playability, with potential to be deployed for home usage.

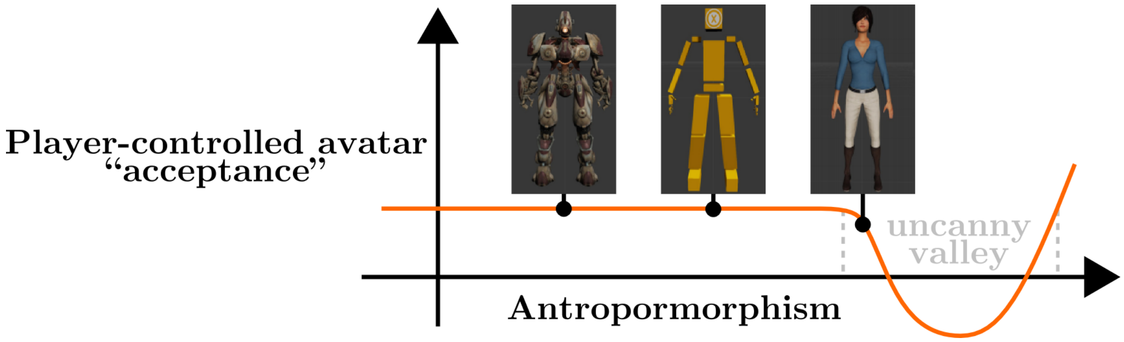


Figure 1. Example assessment of the strength of the IVBO in relation to the degree of anthropomorphism of a user controlled avatar, and the hypothesized uncanny valley effect [9], adapted from [10].

Previous studies seem to suggest that the evaluated systems are not just viable, but appreciated by its users. It is, however, important to note that although those games are graded as serious games, many of the principles being put to use were described for the creation of purely entertainment pastimes. On that perspective, it is possible to highlight that the ability to keep a patient engaged in a time-consuming repetitive activity is usually the biggest hurdle to overcome. Notwithstanding, in the case of [25], it has been noted that games that are created for the specific purpose of rehabilitation make the game effective, but end up lacking qualities traditional games possess, especially entertainment, thus leading to a less motivating experience, and thus resulting in the activity to be less likely to be repeated. The opposite can also happen when games that are designed purely for entertainment purposes are adapted to rehabilitation, in which the user with limited mobility expresses to have great fun with experience, but is unable to complete the level without help from a therapist [26]. These games also tend to focus on the individual recovery, thus lacking social qualities to it. As an additional variable to be taken into consideration, previous studies reinforce the need to create a social rehabilitation game for added motivation, suggesting the participation of a relative, or even multiple patients. An additional note to take from [21] is the consideration for the possibly older audience and the eventual need to hold their focus by the usage of colourful scenes and sound effects.

Overall, the ability of games to easily create fun challenges, be it for an individual or a group, seems to make them a good candidate for a rehabilitation aid tool. This, together with the existence of various input controllers, accessible systems and the possibility of such systems to be taken into the user's home, make it a very appealing candidate to help ease the problem of physical activity promotion. But still, does not yet seem to be settled how to exactly materialize such ideas.

B. Anthropomorphism

Different elements are being included in serious games as strategies to promote improved adherence [27]. Of those, it is possible to highlight virtual representations of the self, through which players are presented to the possibility of assuming the role of a character in the game [28]. On the topic of player controlled game characters, the Illusion of Virtual Body Ownership (IVBO) considers the effect of game players experiencing a sense of artificial body parts to be their own, within the context of an Virtual Reality (VR) setting [10]. Previous research [10] tends to suggest that the IVBO may result from an interaction of both synchronous visual, motor and tactile sensory inputs, as well as pre-existing visual and proprioceptive body representation factors. Another factor is the virtual body realism in terms of visual human resemblance, or anthropomorphism [29]. On a related note, the Uncanny Valley appertains to a theorized relationship between humans and robots [9] (e.g., Figure 1), that hypothesis that it should exist a positive relationship between how human a robot looks, and how comfortable people are with its appearance, up to the moment a robot would get too close to being human in appearance, without being fully human, at which point human reaction would became negative [9].

C. Human Pose Estimation

In vivo measurements of body mechanics have been typically acquired with optical motion capture or inertial sensor-based methods in a laboratorial setting [30]. Most commonly applied techniques employ optical systems that use high-speed cameras to capture the 3D motion of reflective markers that have been placed on anatomically relevant landmarks of the subject's limbs, trunk, and pelvis, being the supportive assumption that these markers' motion represent the movement of the rigid bony segments observed during the movement (Kernozek et al., 2013). It is possible to verify that an increasing number of studies have been focused on monocular mark-less approaches based on visual data. Accordingly, it is possible to recognize that the visual data streams, acquired by cameras, present the benefit of allowing a person to be monitored without the need of additional markers to be employed. From the several types of data that can be captured, RGB and depth, are, as already recognized, two of the most commonly used modalities being used. This approach tends to be much less expensive compared with those specialised opto-electronic apparatuses for acquiring motion data [31], and can also be considered to be used in most natural, everyday life settings. Commercial products include Microsoft's Kinect or Intel's Realsense that also provide an application programming interface (APIs) to acquire said depth data.

Range of motion is an important element to be taken into account when evaluating body mobility [32] that can be challenging for a patient to self report. Despite associated performance compromises, the Microsoft Kinect has been previously evaluated as a tool that could easily be used to monitor such measures without assistance from a trained examiner [33].

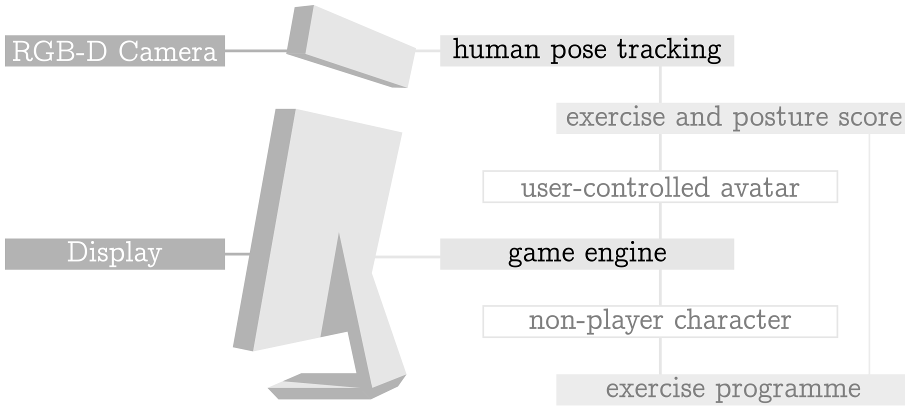


Figure 2. Overall architecture of the proposed system, comprising the use of colour and depth data to monitor, and provide real-time feedback to, a user in the context of a physical activity promotion for breast cancer survivors.

III. Proposed Research Approach

While there are several games that include serious topics, the inclusion of serious game elements is not yet enough to induce learning or real-world action [35]. Overall, despite engagement being considered a valuable resource, research on patient engagement technologies regarding impact on health outcomes has been limited [36]. Given this, this work's main goal is to develop and assess a game to promote an adequate exercise routine for BCS, to be used independently, as a self-management system to support breast cancer survivorship while monitoring one's physical status. The overall architecture of the proposed system is outlined in Figure 2.

We consider the Microsoft Kinect as an easily accessible, Color and Depth (RGB-D) sensor-device that enables to monitor a user's movement and provide feedback through the usage of an avatar, so that the user is aware of the performed movement, aiming at promoting adherence to exercise [37]. Both versions of the Kinect range sensor, i.e., the Kinect^{SL}, which is based on the Structured Light principle, and the Time-of-Flight variant Kinect^{ToF}, were considered [38]. To create the game environment, Unity was selected as the game engine, given its accessibility and widespread use.

In this paper, we pursue the following main topics:

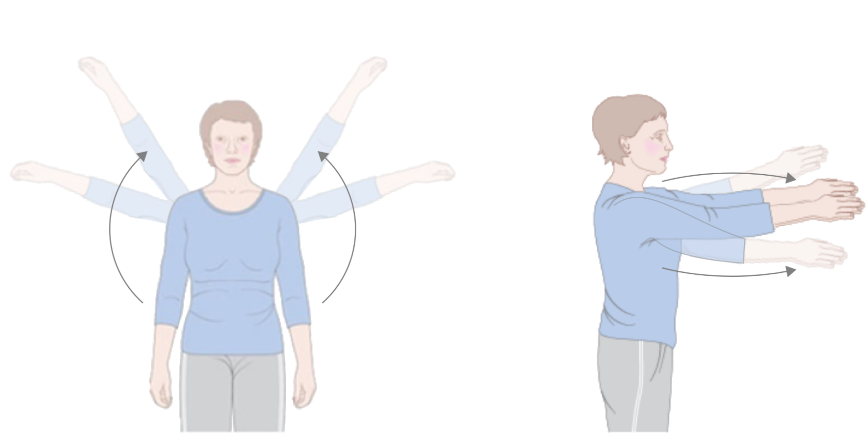
1) anthropomorphism as a strategy to engage, and
2) gamification as a mean to promote physical activity,
3) evaluation of RGB-D based human pose estimation systems for shoulder and elbow angle measurements

about which we present a body of exploratory work, with particular interest on the investigation of expressed preferences of the target user population to evaluate the developed demonstrators in a clinical setting.

A. Exercise programme selection

A standardized exercise programme consisting of shoulder flexion, abduction, and horizontal adduction was selected in accordance to the National Institute for Health and Clinical Excellence (NICE) guidelines [34]. The individual exercises comprised in the programme are illustrated in Figure 3.

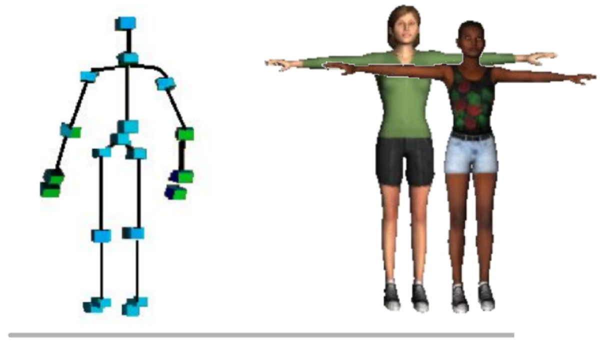
The exercise routine is composed of three sets, each comprising ten repetitions of one of the three exercises included in the programme, and small breaks between sets.



(a) (b) (c)

Figure 3. Illustration of the elements of the exercise programme considered for BCS physical activity promotion intervention, consisting of shoulder flexion (a), abduction (b) and horizontal abduction (c). Adapted from [34].

Levels of user controlled avatar anthropomorphism (while exercising)	
no visual feedback	
skeleton	a_1)
humanoid avatar	a_2)
mirror	



a_1) a_2)

Figure 4. Illustration of the user controlled character: avatar based characters animated with the user's tracked movement with different levels of anthropomorphism a_1) and a_2).

B. Expressed acceptance assessment

Analysis of engagement can be considered valuable in providing insights into game mechanisms that can then be applied to games for learning, or physical activity promotion [39], although not trivial to measure. In order to assess the acceptance of particular contexts of a given physical activity promotion intervention, a criteria set, based on [40], was used as basis for user expressed acceptance assessment. The criteria comprised the following aspects of testing:

- c_1) suitability for the task,
- c_2) information accessibility,
- c_3) continuity correction,
- c_4) visual pleasingness,
- c_5) self-descriptiveness,
- c_6) adequacy of user workload.

Based on that criteria, a questionnaire composed of six questions was formulated in Portuguese, and a five point scale, ranging from strong disagreement (1) to strong agreement (5), considered for range of response options.

C. Study on anthropomorphism

1) Participants and design: Following the approval by the clinical service direction, seventy-two adults (mean age of the cohort was 57.79 $\pm$ 11.16 years, all female) participated in the study. Participants were recruited via personal invitation from surgeon-led follow-up consultations of BCS. Participants were informed that the study was voluntary and part of the development of an aid designed to promote physical activity recommend for BCS within the context of an academic dissertation work. Written informed consent granting permission to the use of the anonymously collected data was obtained from all participants. All participants were fluent in Portuguese and did not get paid for their participation.

2) Procedure and materials: Participants were invited to participate in this study via personal invitation at the end of a follow up consultation at the Breast Center of São João Hospital during the period from the end of October until the beginning of December, 2016. The recruited participants were prompted to use the system in an adjacent room to the consultation room.

The architecture illustrated in Figure 2 was adapted so that it would entail a Non-Player Character (NPC) in the form of a virtual assistant that exemplified the movements to be performed according to the established exercise programme while the user was exercising. The same programme would be repeated four times, considering additional breaks between routines, one for each of the considered levels of the user controlled avatar anthropomorphism (illustrated in Figure 4).

After using the system, each patient was inquired of its satisfaction level of the usage of the system through a questionnaire that required the user to rate each of the tested interfaces according to a five point scale ranging from least preferred (1) to most preferred (5). Each session took approximately 30 minutes, comprising the usage of the system for the proposed exercise programme and the filling of the questionnaire.

3) Results: Each of the four interfaces were evaluated using the aforementioned score in a five point scale after the user completed the exercise programme using all of the proposed interfaces. Table I presents the mean expressed preferences for the user controlled character variations.

Although it seems to not exist an abrupt drop on the collected expressed preference between evaluated interfaces with different levels of user controlled avatar anthropomorphism, both skeleton and humanoid examples seem to be preferred over the alternatives with either no visual feedback, or mirror-based feedback.

TABLE I. AVERAGE AND STANDARD DEVIATION (SD) OF EXPRESSED PREFERENCES OVER THE DIFFERENT USER CONTROLLED CHARACTER TESTED BY SEVENTY-TWO BCS IN A CLINICAL SETTING.

		avatar		
	no visual feedback	skeleton	humanoid	mirror
Average	4.10	4.22	4.22	4.19
SD	0.77	0.88	0.77	0.82

(a) Experimental set-up

(b) Printed pamphlet

Figure 5. a) Acquisition environment for the study on gamification for the tested system comprising a Kinect$^{\text{ToF}}$, laptop and additional screen; b) Printed pamphlet produced at the Breast Center of São João Hospital and distributed to BCS.

D. Study on gamification

1) Participants and design: Sixty-eight adults (mean age of the cohort was 59.09 $\pm$ 10.92 years, all female) participated. The same recruitment method mentioned in Subsection III-C (Study on anthropomorphism) was used. A sub group of 22% of participants (15 out of 68) were randomly assigned to receive printed information resources, in form of a pamphlet produced at the Breast Center of São João Hospital.

2) Procedure and materials: As in the study on anthropomorphism, participants were invited to participate after a surgeon-led follow-up consultation at the Breast Center of São João Hospital. The recruitment took place from the beginning of November until the end of December, 2017. Participants were informed about the study being part of the development of an aid to promote physical activity recommend for BCS, and prompted to use the system, in an adjacent to the consultation room (as illustrated in Figure 5).

The architecture used for the study on anthropomorphism, was considered, including the NPC virtual assistant exemplifying the exercise programme to be performed. To provide real-time feedback of the user's own movement only a human avatar was used. Differently from the previous study, the user controlled avatar was animated with the human pose provided by a Kinect$^{\text{ToF}}$. A novelty introduced by the second Kinect version (through its corresponding SDK and respective tools) is the gesture builder tool.

The gesture builder tool was used to create a database containing the set of considered movements, and to assess the completion of a given movement being performed. After building a library of the selected exercises, this was used to score the performance of the user. The normal scoring of the game attributed 1 point for every 1% of progress in each repetition, and a final score was presented as a percentage of the routine completed (the complete routine corresponds to 3000 points).

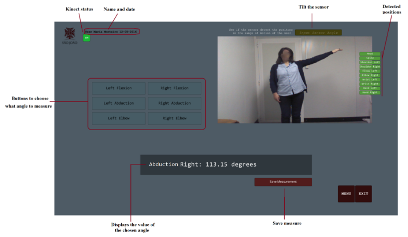
After the usage of the system, each patient was inquired to express level of acceptance that required the user to rate each of the previously identified criteria according to a five point scale ranging from strong disagreement (1) to strong agreement (5). Each session took approximately 10 minutes, which comprised the usage of the interface for the proposed exercise programme by the user and the filling of the questionnaire.

3) Results: Table II presents the mean expressed acceptance for the proposed Gamified Aid for Monitoring Exercise (GAME) with a humanoid player controlled character and an NPC virtual assistant, against an informative printed pamphlet. Of the total cohort of sixty-eight BCS, fifty-three were randomly assigned to use the GAME and fifteen assigned for being shown the printed pamphlet.

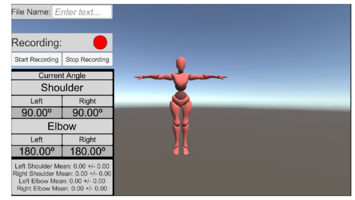
In the context of the evaluation, it seems to exist a stronger agreement, across considered criterion, for the proposed GAME being a preferred medium over printed materials.

TABLE II. AVERAGE AND STANDARD DEVIATION (SD) OF EXPRESSED ACCEPTANCE FOR THE PROPOSED GAME AND A PRINTED PAMPHLET CONTAINING INFORMATION ABOUT THE SELECTED EXERCISE PROGRAMME.

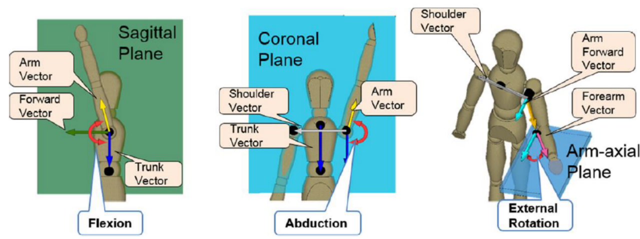
	GAME		pamphlet	
Criteria	Average	SD	Average	SD
c_1)	4.60	0.74	4.00	1.11
c_2)	4.72	0.50	4.08	1.27
c_3)	4.92	0.32	5.00	0.00
c_4)	4.88	0.29	4.60	0.84
c_5)	4.96	0.19	4.60	0.84
c_6)	4.88	0.39	4.00	1.11



(a) Interface for Kinect^{SL} virtual goniometer



(b) Interface for Kinect^{ToF} virtual goniometer



(c) Reference measurments of shoulder and elbow range of motion angles

Figure 6. Snapshots of the developed interfaces for the virtual goniometer for both Kinect^{SL}, (a), and Kinect^{ToF}, (b), and illustration of the measured postural angles, adapted from [41], (c).

E. Study on postural angles from RGB-D

1) Participants and design: Due to time constraints from the clinical setting this study of validation of the Kinect sensors for the task of measuring joint angles was performed in an academic setting, having participated seven adults (mean age of the cohort was 27.6 ± 5.44 years, 5 male, 2 female) recruited via personal invitation from a group of post graduate students from the Faculty of Engineering of Porto with no known shoulder impairments. All participants were fluent in Portuguese and did not get paid for their participation.

2) Procedure and materials: Participants were invited to participate in this study via personal invitation at two moments: December, 2016 and December 2017. Participants were informed that the study was part of the development of an aid designed to promote physical activity recommend for BCS. The recruited participants were prompted to use the system, in an indoor room of the Faculty of Engineering of Porto.

An application for measuring range of shoulder and elbow motion using input data from both Microsoft Kinect RGB-D sensors was developed using the cross-platform game engine Unity. Furthermore, the results are compared against the data acquired with a clinical gold standard goniometer. The goniometer used to register the values of postural angles of the shoulder and elbow has a range between 0 and 270 degrees, with a minimum scale value of 2 degrees.

Each subject was recorded performing a shoulder flexion with the instruction of obtaining a 90 degree angle. For the shoulder abduction and elbow rotation exercises, the same procedure as the shoulder flexion has considered. In the case of shoulder rotation the vectors can be seen as the vector defined by the points consisting of the chest and the hip, against the vector defined by the points consisting of the shoulder and elbow. In the case of elbow rotation, the angle is obtained by the vector defined from the shoulder to the elbow and the vector from the elbow to the wrist. These vectors can be visualized in different planes in Figure 6.

3) Results: Table III presents the mean absolute difference (disagreement) between the readings of a trained annotator using a goniometer, and the computed angles from the three-dimensional location of the anatomical landmarks detected by both versions of Kinect. The results suggest that despite an improved design regarding depth estimation of the Kinect^{ToF} over the Kinect^{SL}, the difference to the goniometer measures is smaller for the older version of the RGB-D camera, that seemed to deal better with the partial occlusion of body parts in the range of view of the acquisition device in the context of this study. Despite constituting a limitation to the present study, it does not seem evident to the authors whether a different result could have been observed would have been possible to have access to BCS to perform this technical validation.

TABLE III. AVERAGE AND STANDARD DEVIATION (SD) OF SHOULDER AND ELBOW RANGE OF MOTION ABSOLUTE ANGLE DIFFERENCE IN DEGREES BETWEEN GONIOMETER MEASURES BY A TRAINED ANNOTATOR AND ESTIMATIONS BASED ON THE HUMAN POSE RECOVERY METHODS PROVIDED WITH BOTH STRUCTURED LIGHT (SL) AND TIME-OF-FLIGHT (TOF) VARIANTS OF KINECT ($|\theta_{KINECT_V} - \theta_{GONIOMETER}|, V = SL, TOF$).

Absolute Disagreement [°]						
	Shoulder abduction		Shoulder flexion		Elbow flexion	
	Kinect^{SL}	Kinect^{ToF}	Kinect^{SL}	Kinect^{ToF}	Kinect^{SL}	Kinect^{ToF}
Average	0.936	3.190	0.532	5.170	1.034	7.080
SD	0.630	2.205	0.421	2.200	0.603	7.305

IV. Conclusion

The present work investigates the impact of providing real-time feedback to BCS within the context of a physical activity promotion intervention. A system comprised of a RGB-D sensor with a processing pipeline to monitor the user, and in that way animate a user controlled avatar, was considered.

In the first exploratory study the effect of different levels of anthropomorphism of the user controlled avatar was investigated. Seventy-two BCS participated in the cohort. The results seem to agree with the hypothesised Uncanny Valley effect, in the sense that a more anthropomorphised representation of the self (a mirror), seems not to be the preferred interface. Although not possible to assess from the presented results, but also supported considering previous research, i.e., [10], subjectively constructed proprioceptive body representations of the self, seems to be an apparently worth considering factor in the context of BCS, with potential impact to adherence to systems using anthropomorphised avatars.

In the second study a gamified approach considering an humanoid avatar and an NPC assistant was evaluated against a printed pamphlet. From a total of sixty-eight participants, a subgroup of 15 was randomly assigned to be shown the pamphlet containing information about appropriate care following breast cancer treatment, including the recommendation to perform simple exercises to be repeated throughout survivorship. The remaining participants played a game where an NPC assistant would exemplify the recommend exercise programme, while a humanoid avatar would replicate the user's movements, and in real-time provide feedback of the exercise being executed. Overall, the collected expressed acceptance suggests that the proposed gamified aid for monitoring exercise seems suitable for the task, informative, visual pleasing, self-descriptive, and providing an adequate workload to the user.

For the third study the accuracy of both versions of the Microsoft Kinect RGB-D capture device in calculating angles of the shoulder and elbow in specific poses and the possibility of using it instead of a tradition instrument the goniometer was evaluated. A control group of 7 healthy adults was considered and two separated sessions were performed for the distinct Structured Light and the Time-of-Flight based depth sensors. The results compared to the goniometer seem to suggest that the Microsoft Kinect can be considered as an auxiliary virtual goniometer in order to facilitate frequent measurement taking, specially if in coordination with individualized monitoring context. It seems also possible to recognize a higher disagreement with the gold standard measures for the Kinect$^{\text{ToF}}$ that can be related with a more frequent interference from detection errors associated with occluded body parts.

Overall it seems possible to recognize that in the context of physical activity promotion interventions target at particular populations monitoring particular measures of physical status trough automatic methods based on human pose estimation is feasible. Moreover, the application of gamified strategies to promote engagement seem to be well perceived by users, even though the mid to long term adhesion is yet to be properly characterized. Furthermore, for the particular context of breast cancer survivors, the hypnotized relation between humans and virtual characters in which users would experience higher levels of comfort when interacting with more anthropomorphic looking avatars, does not seem to hold.

V. Future Work

The considered prospective surveillance model for breast cancer survivors highlights the importance of monitoring for functional and physical impairment commonly associated with breast cancer treatment. Low cost device-based methods have been studied, and its potential to "enable a continuum of time scale from a summary of entire interactions to second-by-second dynamics" continuously highlighted in a myriad of application scenarios [42]. Notwithstanding, from the presented work, various topics seem to still present themselves as pertinent to be explored in future work. Among several, the problem of recovering the spatial pose of the human body during dynamic movements, from mark-less set-ups of acquisition is highlighted bellow.

Computer vision and pattern recognition fields present recent insightful research that is constantly innovating. Even if, at the very front end of development, progress may look a lot like long-lived methodologies as may be the case of variations of perceptron inspired learning approaches [43] or related weights estimation method of back-propagation [44], rediscovered through a new context of parallel computation capabilities and extensive data availability [45]. And despite all that, or even the recent explosion of deep convolutional neural networks (DCNN), *general purpose learning algorithms that improve themselves in provably optimal ways* still seem a distant future [45]. By the same token, human pose estimation still remains with several challenges, especially in the 3D space as reviewed by [46]. One of the challenges arises from the ill-posed nature of the 3D pose estimation task itself, especially from a single monocular image. Similar image projections can be derived from completely different 3D poses. In such cases, self-occlusions result in ambiguities that limit the applicability of existing techniques. Furthermore, recent research primarily focus on frontal views with few occlusions despite the abundance of occlusion and partial-poses in object detection in natural environments.

Besides a trend of deep neural networks-based methods, the existence of prominent publicly made available datasets [47], [48], [49], recurrently used to stablish benchmarks for the task of recovering the tri-dimensionality of the human pose from bi-dimensional visual data, seem to contribute to advances in the respective field. Despite the recent trend of methods outperforming feature learning strategies in a myriad of applications [50], the challenging tasks of establishing a proper learning approach and parameters tuning for a given task [51], as well as, a compromised interpretability [52] of the resulting models, still remain relevant open challenges that gain special importance in clinically related applications. On that regard, recent regulation in the European Union proposes that individuals affected by algorithmic decisions have a right to explanation [53], despite not being completely clear on how a clinician treating a patient who is aided by a machine learning algorithm may be expected to explain decisions that use the patient's data [54]. The need for less opaque ways to explain algorithms outcomes, has also motivated the recent DARPA's Explainable Artificial Intelligence (XAI) program [55].

Acknowledgments

The authors would like to thank the direction, members, and users of the Breast Center of São João Hospital, that, valuably, supported and participated in the research.

This work is financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia as part of project UID/EEA/50014/2013; and also by FCT within the Ph.D. grant number SFRH/BD/138823/2018.

REFERENCES

[1] J. P. Monteiro, C. T. Lopes, N. C. Duarte, A. Magalhães, and H. P. de Oliveira, "Investigations on the impact of anthropomorphism and gamification on breast cancer survivors' expressed preferences in a physical activity promotion intervention," in Proceedings of the eTELEMED 2019 : The Eleventh International Conference on eHealth, Telemedicine, and Social Medicine. IARIA, 2019, pp. 139–143.

[2] V. M. Boquiren, T. F. Hack, R. L. Thomas, A. Towers, W. B. Kwan, A. Tilley, E. Quinlan, and B. Miedema, "A longitudinal analysis of chronic arm morbidity following breast cancer surgery," Breast Cancer Research and Treatment, vol. 157, no. 3, 2016, pp. 413–425.

[3] R. Gärtner, M.-B. Jensen, J. Nielsen, M. Ewertz, N. Kroman, and H. Kehlet, "Prevalence of and factors associated with persistent pain following breast cancer surgery," Jama, vol. 302, no. 18, 2009, pp. 1985–1992.

[4] K. G. Andersen and H. Kehlet, "Persistent pain after breast cancer treatment: a critical review of risk factors and strategies for prevention," The Journal of Pain, vol. 12, no. 7, 2011, pp. 725–746.

[5] Y. Briskin and T. Odinets, "Improvement of upper limb's condition of women with post mastectomy syndrome with the help of problem–oriented program of physical rehabilitation," Pedagogics, psychology, medical-biological problems of physical training and sports, vol. 19, no. 11, 2015, pp. 20–25.

[6] N. L. Stout, J. M. Binkley, K. H. Schmitz, K. Andrews, S. C. Hayes, K. L. Campbell, M. L. McNeely, P. W. Soballe, A. M. Berger, A. L. Cheville, C. Fabian, L. H. Gerber, S. R. Harris, K. Johansson, A. L. Pusic, R. G. Prosnitz, and R. A. Smith, "A prospective surveillance model for rehabilitation for women with breast cancer," Cancer, vol. 118, no. S8, 2012, pp. 2191–2200.

[7] L. Lai, J. Binkley, V. Jones, S. Kirkpatrick, C. Furbish, P. Stratford, W. Thompson, A. Sidhu, C. Farley, J. Okoli, D. Beech, and S. Gabram, "Implementing the prospective surveillance model (PSM) of rehabilitation for breast cancer patients with 1–year postoperative follow–up, a prospective, observational study," Annals of Surgical Oncology, vol. 23, no. 10, 2016, pp. 3379–3384.

[8] W. Tauxe, "A tumour through time," Nature, vol. 527, no. 7578, 2015, pp. S102–S103.

[9] M. Mori, K. MacDorman, and N. Kageki, "The uncanny valley," IEEE Robotics & Automation Magazine, vol. 19, no. 2, 2012, pp. 98–100.

[10] J. Lugrin, J. Latt, and M. E. Latoschik, "Avatar anthropomorphism and illusion of body ownership in vr," in Proceedings of the 2015 IEEE conference on Virtual Reality (VR), 2015, pp. 229–230.

[11] M. Simon, "Gamification and serious games for personalized health," Studies in Health Technology and Informatics, vol. 177, 2012, pp. 85–96.

[12] F. Laamarti, M. Eid, and A. E. Saddik, "An overview of serious games," International Journal of Computer Games Technology, vol. 2014, 2014, pp. 1–15.

[13] S. Wüest, N. A. Borghese, M. Pirovano, R. Mainetti, R. van de Langenberg, and E. D. de Bruin, "Usability and effects of an exergame–based balance training program," Games for Health Journal, vol. 3, no. 2, 2014, pp. 106–114.

[14] B. Lange, S. Koenig, E. McConnell, C.-Y. Chang, R. Juang, E. Suma, M. Bolas, and A. Rizzo, "Interactive game–based rehabilitation using the microsoft kinect," in Proceedings of the 2012 IEEE Conference in Virtual Reality (VR). IEEE, 2012.

[15] R. Moreira, A. Magalhães, and H. Oliveira, "A kinect–based system for upper–body function assessment in breast cancer patients," Journal of Imaging, vol. 1, no. 1, 2015, pp. 134–155.

[16] S. S. Esfahlani, S. Cirstea, A. Sanaei, and G. Wilson, "An adaptive self–organizing fuzzy logic controller in a serious game for motor impairment rehabilitation," in Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), 2017, pp. 1311–1318.

[17] G. A. Caurin, A. A. Siqueira, K. O. Andrade, R. C. Joaquim, and H. I. Krebs, "Adaptive strategy for multi–user robotic rehabilitation games," in Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 1395–1398.

[18] O. Barzilay and A. Wolf, "Adaptive rehabilitation games," Elsevier Journal of Electromyography and Kinesiology, vol. 23, no. 1, 2013, pp. 182–189.

[19] A. Darzi, M. Goršič, and D. Novak, "Difficulty adaptation in a competitive arm rehabilitation game using real-time control of arm electromyogram and respiration," in Proceedings of the 2017 International Conference on Rehabilitation Robotics (ICORR), 2017, pp. 857–862.

[20] M. Ma, R. Proffitt, and M. Skubic, "Quantitative assessment and validation of a stroke rehabilitation game," in Proceedings of the 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017, pp. 255–257.

[21] G. Alankus, A. Lazar, M. May, and C. Kelleher, "Towards customizable games for stroke rehabilitation," in Proceedings of the Special Interest Group on Computer-Human Interaction (SIGCHI) Conference on Human Factors in Computing Systems, 2010, pp. 2113–2122.

[22] N. J. Seo, J. K. Arun, P. Hur, V. Crocher, B. Motawar, and K. Lakshminarayanan, "Usability evaluation of low-cost virtual reality hand and arm rehabilitation games." Journal of rehabilitation research and development, vol. 53, no. 3, 2016, pp. 321–334.

[23] J. W. Burke, M. McNeill, D. Charles, P. Morrow, J. Crosbie, and S. McDonough, "Serious games for upper limb rehabilitation following stroke," in Proceedings of the 2009 Conference in Games and Virtual Worlds for Serious Applications, 2009, pp. 103–110.

[24] J. W. Burke, M. McNeill, D. K. Charles, P. J. Morrow, J. H. Crosbie, and S. M. McDonough, "Optimising engagement for stroke rehabilitation using serious games," The Visual Computer, vol. 25, no. 12, 2009, p. 1085.

[25] E. Flores, G. Tobon, E. Cavallaro, F. I. Cavallaro, J. C. Perry, and T. Keller, "Improving patient motivation in game development for motor deficit rehabilitation," in Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, ser. ACE '08. ACM, 2008, pp. 381–384.

[26] D. Rand, R. Kizony, and P. Weiss, "Virtual reality rehabilitation for all: Vivid gx versus sony playstation ii eyetoy," in Proceedings of the 5th Intl Conf. Disability, Virtual Reality and Assoc. Tech (ICDVRAT), 2004, vol. 4, pp. 87–94.

[27] D. Thompson, "Designing serious video games for health behavior change: Current status and future directions," Journal of Diabetes Science and Technology, vol. 6, no. 4, 2012, pp. 807–811.

[28] M. Rice, R. Koh, Q. Lui, Q. He, M. Wan, V. Yeo, J. Ng, and W. P. Tan, "Comparing avatar game representation preferences across three age groups," in Proceedings of the 2013 ACM SIGCHI Conference on Human Factors in Computing Systems CHI. ACM Press, 2013.

[29] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik, "The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response," IEEE Trans. on Visualization and Computer Graphics, vol. 24, no. 4, 2018, pp. 1643–1652.

[30] L. Donath, O. Faude, E. Lichtenstein, C. Nüesch, and A. Mündermann, "Validity and reliability of a portable gait analysis system for measuring spatiotemporal gait characteristics: comparison to an instrumented treadmill," Journal of neuroengineering and rehabilitation, vol. 13, no. 1, 2016, p. 6.

[31] A. Patrizi, E. Pennestrì, and P. P. Valentini, "Comparison between low-cost marker–less and high–end marker–based motion capture systems for the computer–aided assessment of working ergonomics," Ergonomics, vol. 59, no. 1, 2016, pp. 155–162.

[32] S. H. Lee, C. Yoon, S. G. Chung, H. C. Kim, Y. Kwak, H. won Park, and K. Kim, "Measurement of shoulder range of motion in patients with adhesive capsulitis using a kinect," PLOS ONE, vol. 10, no. 6, 2015, p. e0129398.

[33] M. Huber, A. L. Seitz, M. Leeser, and D. Sternad, “Validity and reliability of kinect skeleton for measuring shoulder joint angles: a feasibility study,” Physiotherapy, vol. 101, no. 4, 2015, pp. 389–393.

[34] J. Yarnold, “Early and locally advanced breast cancer: Diagnosis and treatment national institute for health and clinical excellence guideline 2009,” Clinical Oncology, vol. 21, no. 3, 2009, pp. 159–160.

[35] K. Starks, “Cognitive behavioral game design: a unified model for designing serious games,” Frontiers in Psychology, vol. 5, 2014, p. 28.

[36] J. E. Prey, J. Woollen, L. Wilcox, A. D. Sackeim, G. Hripcsak, S. Bakken, S. Restaino, S. Feiner, and D. K. Vawdrey, “Patient engagement in the inpatient setting: a systematic review,” Journal of the American Medical Informatics Association, vol. 21, no. 4, 2014, pp. 742–750.

[37] J. Han, L. Shao, D. Xu, and J. Shotton, “Enhanced computer vision with microsoft kinect sensor: A review,” IEEE Transactions on Cybernetics, vol. 43, no. 5, oct 2013, pp. 1318–1334.

[38] H. Sarbolandi, D. Lefloch, and A. Kolb, “Kinect range sensing: Structured–light versus time–of–flight kinect,” Computer Vision and Image Understanding, vol. 139, oct 2015, pp. 1–20.

[39] E. A. Boyle, T. Hainey, T. M. Connolly, G. Gray, J. Earp, M. Ott, T. Lim, M. Ninaus, C. Ribeiro, and J. Pereira, “An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games,” Computers & Education, vol. 94, 2016, pp. 178–192.

[40] K. S. Park and C. H. Lim, “A structured methodology for comparative evaluation of user interface designs using usability criteria and measures,” International Journal of Industrial Ergonomics, vol. 23, no. 5-6, 1999, pp. 379–389.

[41] S. H. Lee, C. Yoon, S. G. Chung, H. C. Kim, Y. Kwak, H.-w. Park, and K. Kim, “Measurement of shoulder range of motion in patients with adhesive capsulitis using a kinect,” PloS one, vol. 10, no. 6, 2015, p. e0129398.

[42] C. Leclère, M. Avril, S. Viaux-Savelon, N. Bodeau, C. Achard, S. Missonnier, M. Keren, R. Feldman, M. Chetouani, and D. Cohen, “Interaction and behaviour imaging: a novel nethod to neasure mother–infant interaction using video 3d reconstruction,” Translational psychiatry, vol. 6, no. 5, 2016, p. e816.

[43] F. Rosenblatt, “Principles of neurodynamics. perceptrons and the theory of brain mechanisms,” Cornell Aeronautical Lab Inc Buffalo NY, Tech. Rep., 1961.

[44] D. E. Rumelhart, G. E. Hinton, R. J. Williams et al., “Learning representations by back–propagating errors,” Nature, vol. 323, no. 3, 1986, pp. 533–536.

[45] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural networks, vol. 61, 2015, pp. 85–117.

[46] N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris, “3d human pose estimation: A review of the literature and analysis of covariates,” Computer Vision and Image Understanding, vol. 152, 2016, pp. 1–20.

[47] L. Sigal, A. O. Balan, and M. J. Black, “Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion,” International journal of computer vision, vol. 87, no. 1-2, 2010, p. 4.

[48] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, “2d human pose estimation: New benchmark and state of the art analysis,” in Proceedings of the IEEE Conference on computer Vision and Pattern Recognition, 2014, pp. 3686–3693.

[49] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments,” IEEE transactions on pattern analysis and machine intelligence, vol. 36, no. 7, 2013, pp. 1325–1339.

[50] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, “Deep patient: An unsupervised representation to predict the future of patients from the electronic health records,” Scientific reports, vol. 6, 2016, p. 26094.

[51] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” Neurocomputing, vol. 187, 2016, pp. 27–48.

[52] B. J. Lengerich, S. Konam, E. P. Xing, S. Rosenthal, and M. Veloso, “Towards visual explanations for convolutional neural networks via input resampling,” arXiv preprint arXiv:1707.09641, 2017.

[53] B. Goodman and S. Flaxman, “European union regulations on algorithmic decision–making and a “right to explanation”,” AI Magazine, vol. 38, no. 3, 2017, pp. 50–57.

[54] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman et al., “Opportunities and obstacles for deep learning in biology and medicine,” Journal of The Royal Society Interface, vol. 15, no. 141, 2018, p. 20170387.

[55] M. T. Ribeiro, S. Singh, and C. Guestrin, ““why should i trust you?”: Explaining the predictions of any classifier,” in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, 2016, pp. 1135–1144.

Evaluation of Joint Range of Motion Measured by Vision Cameras

Oky Dicky Ardiansyah Prima, Yuta Ono, Yoshitoshi Murata, Hisayoshi Ito
Graduate School of Software and Information Science, Iwate Pref. Univ.
Takizawa, Japan
prima@iwate-pu.ac.jp, g231q005@s.iwate-pu.ac.jp, {y-murata, hito}@iwate-pu.ac.jp

Takashi Imabuchi
Office of Regional Collaboration, Iwate Pref. Univ.
Takizawa, Japan
t_ima@ipu-office.iwate-pu.ac.jp

Yukihide Nishimura
Dept. of Rehabilitation Medicine
Iwate Medical University
Morioka, Japan
ynishi@iwate-med.ac.jp

***Abstract*— Joint Range of Motion (ROM) can be measured through a variety of methods including the use of sophisticated devices such as goniometers and non-intrusive three-dimensional (3D) sensor devices such as motion capture systems. The Microsoft Kinect has been proposed as an affordable motion capture device as an alternative to goniometers. However, due to limited measurement range and complex setup, this device cannot be used as a self-measurement during home rehabilitation or flexibility training. With the recent progress in human pose estimation based on computer vision approaches, it has become possible to estimate human joint positions in real time from vision cameras. This study evaluates joint ROM measured by two vision cameras using 3D human pose estimation based on a single camera and a stereo camera. The ROM of major joints, which consist of shoulders, elbows, a hip, and knees was evaluated for 10 users. The stereo camera gives the best results with a small bias to the goniometer compared to the single camera and the Kinect. Vision cameras have advantages on estimating semi-occluded joint locations than the Kinect. The 3D human pose based on a single camera opens up possibilities to build Tele-Rehabilitation (TR).**

Keywords-rehabilitation; computer vision; range of motion; activities of daily living; 3D human pose estimation.

I. INTRODUCTION

Human movement is dependent on the amount of range of motion (ROM), the amount of motion available at a synovial joint. This movement is unique to each joint and is dependent upon the shape of the articular surfaces of bones and the integrity and flexibility of the periarticular soft tissues. ROM can be measured as either active and passive. While the former is measured by the person contracting the muscles around the joint, the latter by an external force pushing on the body around the joint. Passive motion can either limit or perform full joint ROM. This study is an extension of our previous work on the assessment of ROM from body joints estimated by vision cameras [1].

There are close relationships between joint ROM and Activities of Daily Living (ADL) [2][3]. The loss of ROM may occur at all ages due to injuries, diseases, surgery and normal aging, giving a direct effect on posture and movement. Although loss of ROM may not be associated with complete loss of function, people who have impaired ROM need to perform their activities by using compensatory strategies [4]. For example, a patient with impaired shoulder flexion motion may not be able to raise his upper limb but may still be able to conduct most ADL tasks.

Joint ROM can be assessed through a variety of methods including goniometers, inclinometers, photographs, and Motion Capture (MoCap) systems. The double-armed goniometer is the most common device to use, whereas ROM is measured at the end of its full range of movement. To obtain reliable measurements, clinicians are suggested to take repeated measurements. Since the universal goniometer has scale in 5° increments, the measurement fluctuation is usually expected up to ±5°. On the one hand, goniometers may introduce error into the measurement because the positions of the bones and axis points must be estimated, on the other hand, inclinometers are easier because no such alignment is required. Inclinometers have dials that indicate the angle at which the inclinometer is located with respect to the line of gravity. Photographs can be used to measure a certain joint ROM. Since most smartphones currently available are equipped with cameras, the use of smartphones as non-intrusive ROM assessment tools is increasing. DrGoniometer, a photo-based iPhone app, potentially offers an easy tool of ROM measurements [5][6]. It also has an ability storage of all related information to build up historical data for each movement for further analysis. MoCap systems can be categorized into marker-based and markerless system. Marker-based MoCap systems, such as Vicon, can accurately capture human movement. Vicon is often regarded as a standard in motion capture. These systems utilize multiple vision cameras to detect the light reflected by the marker and calculate the three-dimensional (3D) position using triangulation. Markerless MoCap systems use a depth sensor to measure the 3D position of the target within range of the sensor. Microsoft Kinect is widely used as an inexpensive markerless MoCap system that can track human movement and posture in three dimensions. Accuracy assessment of the Kinect against Vicon show that the Kinect is sensitive enough to be used as a portable MoCap system for workplace ergonomic assessments [7]. Other studies have shown that the Kinect performs well for a range of healthcare imaging applications [8][9].

Due to the lack of specialized medical institutions that provide rehabilitation services, there is a growing need for simple methods of ROM self-measurement. Marker-based MoCap systems are superior on handling occlusion with multiple vision cameras fixed in multiple directions to capture

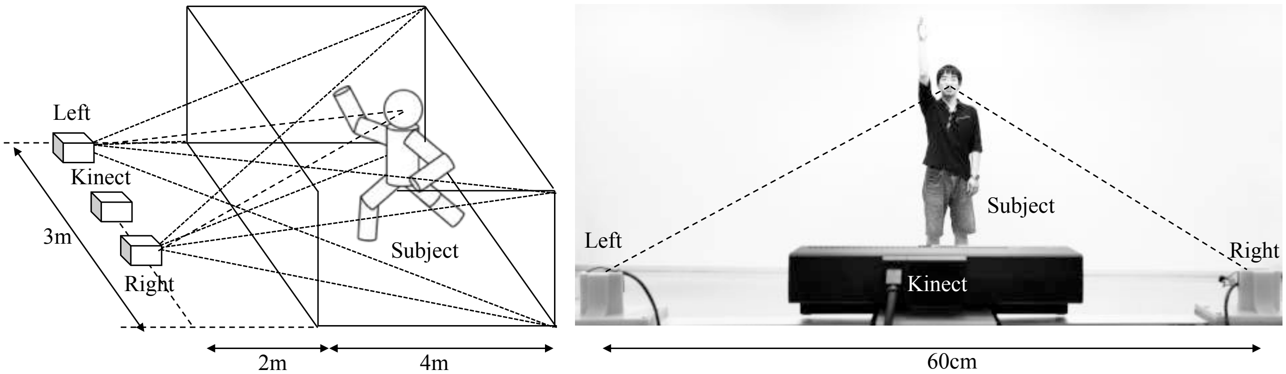


Figure 1. The experimental setup in this study.

the target. However, these systems can retrieve data only in a limited area. In addition, these systems are mostly used indoors because sunlight interferes with infrared cameras used for measurement [10]. These drawbacks also exist with markerless MoCap systems. The complexity to setup MoCap systems has prevented these systems to be used as self-measurements during home-based rehabilitation or flexibility training.

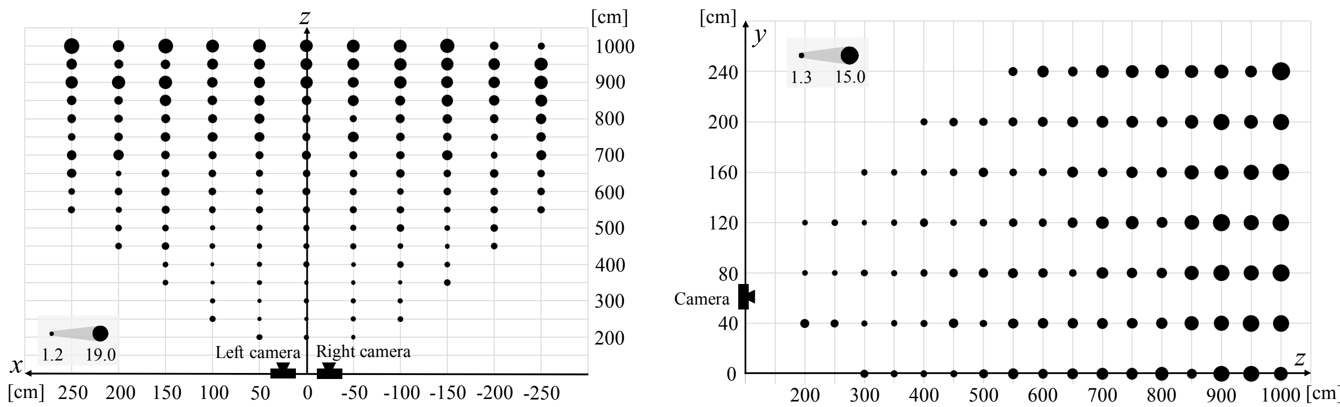
Various rehabilitation programs require a system that is capable to measure joint ROM indoors or outdoors. The basic ADL includes a functional mobility to move from one place to another while performing tasks, such as walking, getting in and out of bed, and getting into and out of a chair. This study seeks to expand the use of vision cameras: a single camera and a stereo camera, as a simple and low-cost tool to promote basic self-care by measuring joint ROM during ADL, home rehabilitation, or flexibility training. Joint ROM is measured according to the method and guidelines for joint range of motion measurement by the Japanese Association of Rehabilitation Medicine and the Japanese Orthopedic Association (JARM & JOA) [11]. ROM of the major joints, consisting of the shoulders, elbows, hips and knees is measured using the vision cameras. Measurement results are compared with the Kinect's to determine whether the vision camera is suitable for practical use of joint ROM quantification.

This paper is organized as follows. Section II describes related work on detection of 3D body joints using vision cameras. Section III describes methods to measure joint ROM using vision cameras and the Kinect based on JARM and JOA guidelines. Section IV evaluates accuracies of the resulted joint ROM obtained from each modality. Finally, Section V summarizes the results and describes the future prospects of this study.

II. Related Work

The task of estimating human pose without using a marker is attracting attention in the field of computer vision research. Many studies have been conducted to enable the use of cameras to detect various joints with complex postures. Toshev et al. (2014) use a regression model using cascade Deep Neural Networks (DNN) to detect 2D joints and associates the corresponding joints throughout the body posture [12]. Newell et al. (2016) proposed Stacked Hourglass Network (SHN) to improve detection by processing a diverse and challenging series of poses using a simple mechanism for initial prediction evaluation [13]. SHN was known to have robust performance against various challenges related to joint detection of multiple people. Both [12] and [13] require a human detection process as a pre-processing to detect joints in the body, and if the pre-processing fails to detect humans, joint detection cannot be performed. The latest method, "OpenPose", uses Part Confidence Maps (PCM) to detect joints and Part Affinity Field (PAF) to associate corresponding joints directly without the human detection process in advance [14]. SHN and OpenPose are available online as open source software for research purposes. Ono et al. (2018) applied OpenPose to the stereo camera and estimated the 3D joint from the corresponding 2D joints using a stereo vision approach [15]. 3D joint measurement based on stereo vision seems promising. As long as stereo cameras are available, patients can create a self-report ROM at home and send reports to the clinician to assess their ability to engage in ADL tasks. Measurement results show that this approach is effective in measuring joint ROM as an alternative to the Kinect. Since the stereo camera only captures visible spectrum, 3D measurements can be performed indoors and outdoors in relatively bright light conditions [1]. The rapid growth of Virtual Reality (VR) has made stereo cameras widely available in the market, making it easier to implement.

Along with the breakthrough in 2D human pose estimation, studies on 3D human pose estimation from a single camera have made significant progress. This estimation requires two steps: joint position estimation in 2D image coordinates and 3D coordinate estimation of each corresponding 2D joint. Many studies have investigated the problem of inferring 3D joints from 2D projections. These studies include traditional 2D to 3D methods that define bone length and estimate 3D joints using binary decision trees [16], or deep net based to estimate 3D joints with DNN. Martinez et al. (2017) proposed a relatively simple deep feedforward DNN [17] via Human3.6M, the largest 3D human pose dataset that includes 3.6 million human poses and corresponding images to estimate 3D joints from 2D projections [18]. Unlike the



(a) Horizontal errors (X-Z plane)

(b) Vertical errors (Y-Z plane)

Figure 2. Distribution of errors measured at ground truth points used in this study.

current MoCap system, this technique is capable of estimating body joints in semi-occluded image regions. Hence, the proposed system is much better at handling occlusion than depth sensor-based systems, such as the Kinect. This work has been made available as open source software, namely “3D-pose-baseline [19].”

Both 2D and 3D human pose estimation described above require a dedicated Graphics Processing Unit (GPU) and result in higher costs to implement. However, this calculation can be efficiently performed using GPU-accelerated cloud services over the Internet. Using services over the Internet, clinicians can measure joint ROM during physical rehabilitation with patients at home. Prima et al. (2019) proposed an IoT-based Tele-Rehabilitation (TR) framework that uses a single camera to observe the joints of a client's body in 3D when performing an ADL task [20]. Measurement results show that although the Kinect gives better results in terms of absolute accuracy, the proposed framework is less sensitive to noise than the Kinect. Further expansion of this application will enable ROM measurement in aquatic therapy and fitness pools.

III. Methods

This study measures the ROM of the major body joints, which consist of shoulders, elbows, hip, and knee joints, using three modalities: a single camera, a stereo camera, and the Microsoft Kinect V2 (hereafter, simply called the Kinect). The resulting joint ROM obtained from each device is verified using the ROM measured with a conventional goniometer. Goniometric measurements were performed in a standardized way [21]. Figure 1 shows the experimental setup for this study. Two cameras at 60cm intervals are stereo-calibrated on images with resolution of 1280×720 pixels. The Kinect is set on the middle of these cameras. The subject performs various joint movements at a distance of 4m from the center line between the two cameras. Data processing is synchronized at 30 frame per second (fps).

A. 3D Joint Measurement Using a Single Camera

The left side camera (Figure 1) is used to capture the subject's movement. For each frame, positions of joints are estimated based on PCM calculated using the OpenPose library [22]. Here, the resulting joint structure is rearranged to match the structure used in SHN. 3D coordinates corresponding to these joints are estimated using the 3D-pose-baseline. A weighted is applied to reduce noise in the resulted 3D joints.

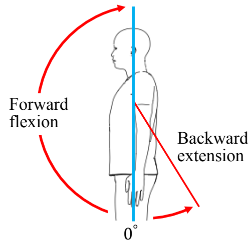
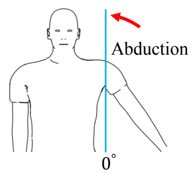
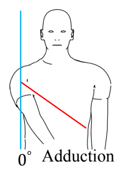
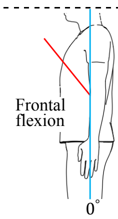
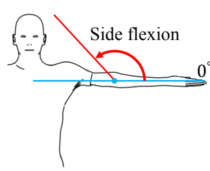
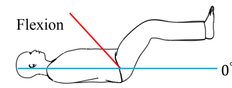
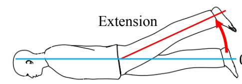
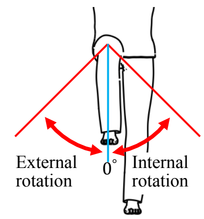
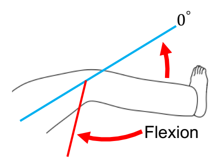
B. 3D Joint Measurement Using a Stereo Camera

The method of Ono et al. (2019) is used to measure 3D joints using a stereo camera [1]. To improve measurement accuracies, the stereo camera is re-calibrated using ground truth points. A total 1,004 ground truth points were regularly placed in an area covered by the stereo camera. This area accounts for 5m×10m. After the re-calibration process, Root Mean Square Error (RMSE) of the ground truth points was measured 8.12cm. This RMSE is considerably acceptable for our study. In the location where the subject performs joint movements (4m from the center line between the two cameras), RMSE was measured 5.07cm. Error distributions are shown in Figure 2.

C. 3D Joint Measurement Using Kinect

The Microsoft Software Development Kit (SDK) for the Kinect is used to access 3D body joints from data taken from the Kinect’s sensor. Here, the resulting 3D joint is not calibrated using ground truth points. The temporal synchronization of the captured data between the Kinect and other cameras was performed using Network Time Protocol (NTP).

TABLE I. ROM MEASUREMENTS OF THE MAJOR BODY JOINTS IN THIS STUDY.

Joint	Motion	ROM	Posture
Shoulder	Forward flexion	180°	
	Backward extension	50°	
	Abduction	180°	
	Adduction	75°	
Elbow	Frontal flexion	145°	
	Side flexion	145°	
Hip	Flexion	125°	
	Extension	15°	
	Internal rotation	45°	
	External rotation	45°	
Knee	Flexion	130°	

D. Data Extraction

ROM measurements of the major body joints are shown as in Table I. Measurements are performed according to JARM and JOA guidelines. For the shoulder joints, ROM of four movements: forward flexion, backward extension, abduction, and adduction are measured. In this measurement, the torso is fixed to the wall so that the spine does not bend back and forth while moving. For the elbow joints, ROM of two movements: frontal flexion and side flexion are measured where the forearm is in the supination position. ROM measurements of hip joints include flexion, extension, internal rotation, and external rotation. The subject lies firmly on his back on a flat surface during hip flexion, while the subject lies firmly in an anatomical position during hip extension. However, in this study, ROM measurements for the internal and external rotation is performed with the subject standing and with the back fixed to the wall because these movement cannot be observed from either vision cameras or the Kinect. The knee joint flexion is performed with the hip joint in flexion. For each movement, angle values between the minimum and maximum angles are measured. The maximum angle does not represent a precise full ROM because external forces such as partner stretching are not involved during the measurement.

Joint angle measurements from data obtained with three modalities: a single camera, a stereo camera and the Kinects, were performed as the relative angle between the longitudinal axis of two adjacent segments. As an example, for elbow joint angles, the adjacent segments are the upper arm and the forearm. Whereas, for knee joint angles, the adjacent segments are the upper and the lower legs.

In this study, two types of measurements are performed: static and continuous. Static measurement measures the

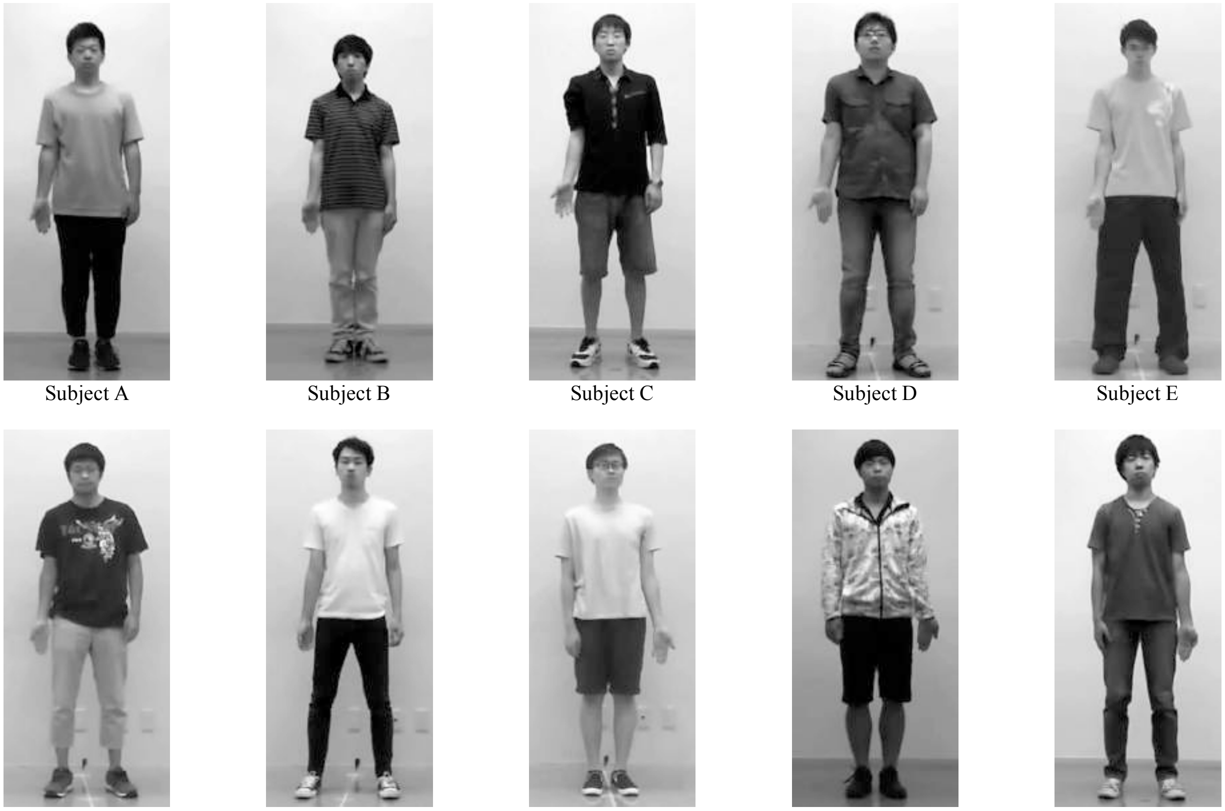


Figure 3. 10 Subjects participated in the experiment.

absolute accuracy of a particular posture. Continuous measurement will reveal how stable the values of angles are measured. Ten healthy male subjects (mean age 21.8 ± 0.92 years) were recruited for the experiments (Figure 3). All participants agreed to participate and signed the consent forms, to allow their data to be used in publications of this research. All subjects were instructed to wear normal clothing to analyze how clothes affects the 3D joints measured by the three modalities. Single cameras and stereo cameras are considered to have a greater impact than the Kinect.

a. Static measurement

Subjects were asked to perform each movement for both left and right joints as shown in Table I. By the time the subject performed a maximum ROM, the examiner aligned the goniometer to the bone and axis position, and another examiner recorded the reading. Subjects were instructed to move each joint to its maximum ability to obtain maximum ROM. The agreement of measurements from the three modalities against the goniometer were evaluate by studying the mean bias and constructing Limits of Agreement (LOA) to determine validity [23][24]. Here, the 95% LOA were defined as the mean bias to ±1.96 Standard Deviation (SD).

b. Continuous measurement

Measurements using three modalities were conducted for 15s from the start position, the position where the goniometer is aligned at 0°, to the end position, the position where full ROM is achieved. During the measurement, a progress bar is displayed on the monitor, so the subject can adjust the movement speed. The resulting measurements were individually fitted using a 4^{th} order polynomial regression model to investigate the stability of each measurement by each device. The RMSE, which is an absolute fit to the model data, was calculated to evaluate the model.

IV. Evaluation of the Resulted Joint ROM Measured by Three Modalities

Motion data was collected for 10 subjects performing 11 motions (Table I). Each motion was performed at the joints of the left and right bodies. Therefore, each user generated 22 motion data for the experiment. For statistical analysis, the goniometer measured the maximum ROM for each motion.

a. Static measurement

Table II shows the mean bias and the 95% LOA of the maximum ROM measurements of joint angles of shoulders, elbows, hip, and knees obtained from the three modalities versus the measurements from the goniometer. Due to 5° scale in goniometer, the measurement fluctuation is usually expected up to ±5°. However, aligning the goniometer correctly with its axis on the joint axis is a difficult task. Here, we considered that a mean bias between ±10° is acceptable for the resulting measurements obtained from the three

TABLE II. MEAN BIAS AND LOA THE MAXIMUM ROM MEASUREMENTS OF JOINT ANGLES FOR SHOULDERS, ELBOWS, HIP, AND KNEE JOINTS ACQUIRED USING THREE MODALITIES AGAINST THE GONIOMETER.

Side	Joint	Motion	Single camera		Stereo camera		Kinect	
			Mean bias	95% LOA	Mean bias	95% LOA	Mean bias	95% LOA
Left	Shoulder	Forward flexion	-25.55°	-36.78 to -14.32°	**3.05°**	-8.55 to 14.66°	**2.79°**	-8.33 to 13.90°
		Backward extension	12.88°	-1.20 to 26.96°	**0.75°**	-22.72 to 24.23°	**7.62°**	-5.75 to 20.98°
		Abduction	-30.65°	-45.45 to -15.85°	**-3.28°**	-14.17 to 7.60°	**-3.78°**	-18.60 to 11.05°
		Adduction	**7.23°**	-7.23 to 21.69°	13.43°	-2.92 to 29.79°	**8.92°**	-5.40 to 23.25°
	Elbow	Frontal flexion	**-5.37°**	-19.98 to 9.24°	16.44°	-6.28 to 39.17°	11.90°	-0.20 to 24.00°
		Side flexion	-29.25°	-47.40 to -11.09°	**-7.02°**	-17.06 to 3.01°	-23.49°	-31.60 to -15.38°
	Hip	Flexion*)	-46.38°	-67.90 to -24.87°	**1.06°**	-7.13 to 9.26°	19.32°	-79.26 to 117.90°
		Extension*)	**-2.25°**	-24.95 to 20.44°	15.32°	-19.42 to 50.06°	78.76°	-63.94 to 221.46°
		Internal rotation	**4.68°**	-14.35 to 23.70°	17.32°	-14.20 to 48.84°	**-3.73°**	-39.02 to 31.55°
		External rotation	**-2.17°**	-15.68 to 11.34°	11.54°	-4.56 to 27.63°	**4.58°**	-6.61 to 15.78°
	Knee	Flexion*)	-57.85°	-81.82 to -33.88°	**-2.35°**	-21.02 to 16.32°	-34.45°	-152.53 to 83.64°
Right	Shoulder	Forward flexion	-27.29°	-44.35 to -10.23°	**5.15°**	-9.23 to 19.53°	**5.04°**	-3.52 to 13.59°
		Backward extension	**3.10°**	-10.28 to 16.48°	**-4.58°**	-15.82 to 6.66°	**2.95°**	-3.36 to 9.25°
		Abduction	-35.02°	-49.37 to -20.66°	**-0.28°**	-12.01 to 11.46°	**-5.27°**	-20.31 to 9.76°
		Adduction	**-0.08°**	-16.38 to 16.22°	10.25°	-4.03 to 24.53°	10.45°	-2.84 to 23.73°
	Elbow	Frontal flexion	**1.04°**	-13.12 to 15.20°	16.81°	-1.89 to 35.51°	16.63°	3.86 to 29.40°
		Side flexion	-23.32°	-36.58 to -10.06°	**-4.79°**	-18.97 to 9.38°	-15.03°	-31.85 to 1.80°
	Hip	Flexion*)	-26.22°	-56.36 to 3.92°	**-1.85°**	-15.08 to 11.38°	-22.00°	-132.40 to 88.39°
		Extension*)	**7.34°**	-8.09 to 22.77°	12.86°	-12.36 to 38.08°	**8.88°**	-19.24 to 37.01°
		Internal rotation	15.30°	-18.70 to 49.30°	14.69°	-18.38 to 47.75°	**-4.34°**	-35.17 to 26.48°
		External rotation	22.86°	-6.37 to 52.09°	14.62°	0.55 to 28.69°	**4.58°**	-9.12 to 18.28°
	Knee	Flexion*)	-66.46°	-91.15 to -41.77°	**0.01°**	-14.75 to 14.78°	-42.97°	-155.55 to 69.61°

*) Measurements were taken with the subject lying on the floor (sleeping posture).

modalities, as shown by bold numbers in Table II. These numbers cover 41% of the measurements obtained with a single camera and 55% with stereo cameras and the Kinect. High mean bias was observed in resulting measurements with a single camera and the Kinect, as shown by underlined numbers. 3D joint measurement using a single camera relies on the 3D human pose dataset used in the 3D-pose-baseline library. Therefore, this library may not provide optimal results when estimating 3D human postures with specific postures such as sleeping postures. Kinects also appear to be insufficient to measure body joints in these postures. Morever, the Kinect suffers from occluded joints which cannot be observed by the its depth sensor. The resulting measurements using the stereo camera indicate relatively better accuracies than those of the single camera and the Kinect. Overall, our measurements show that the 95% LOA for the discrepancy of the three modalities against the goniometer exceeded ±5°, which can be considered as clinically significant. For the stereo camera and the Kinect, this finding is consistent with [1].

ROM measurement results for all joints from each user are presented in Table III. Overall, regardless the subjects, the measurements taken with a single camera significantly show higher mean bias than those taken with the stereo camera and the Kinect, $F(1, 28)= 21.82$, $p<0.01$. There is no trend of results found with specific subjects. Differences in results are more likely depending on the type of motion and posture, as shown in Table II. Here, we consider that there is no significant difference in measurement results depending on clothes. This observation is different from what previous studies suggested [1].

b. Continuous measurement

Table IV shows RMSE for polynomial regression model fitted to the measurement results from the three modalities. Here, the smaller the RMSE, the more stable the measurement. To compare measurement stability among the three modalities, two-way Analysis of Variance (ANOVA)

TABLE III. MEAN BIAS AND LOA OF THE MAXIMUM ROM MEASUREMENTS FOR ALL JOINT ACQUIRED USING THREE MODALITIES AGAINST THE GONIOMETER.

Subject	Single camera		Stereo camera		Kinect	
	Mean bias	95% LOA	Mean bias	95% LOA	Mean bias	95% LOA
A	-10.00°	-72.20 to 52.20°	8.86°	-25.54 to 43.26°	-7.33°	-98.41 to 83.74°
B	-13.69°	-63.99 to 36.61°	4.87°	-16.70 to 26.45°	5.38°	-92.20 to 102.97°
C	-13.11°	-66.24 to 40.01°	8.29°	-16.00 to 32.58°	2.72°	-86.36 to 91.80°
D	-15.89°	-61.31 to 29.52°	3.30°	-14.53 to 21.13°	-3.24°	-42.68 to 36.20°
E	-17.05°	-63.21 to 29.10°	9.32°	-20.47 to 39.11°	-4.99°	-66.11 to 56.13°
F	-13.78°	-67.41 to 39.86°	5.09°	-9.62 to 19.80°	4.96°	-45.23 to 55.14°
G	-10.71°	-57.19 to 35.77°	2.22°	-21.53 to 25.98°	10.22°	-52.06 to 72.51°
H	-10.61°	-63.76 to 42.54°	6.06°	-9.93 to 22.06°	5.03°	-52.00 to 62.06°
I	-15.71°	-59.97 to 28.55°	3.18°	-28.30 to 34.66°	6.11°	-77.21 to 89.44°
J	-17.37°	-71.80 to 37.06°	7.51°	-14.90 to 29.91°	-6.42°	-71.31 to 58.47°

TABLE IV. RMSE FOR MODEL FITTING RESULTS.

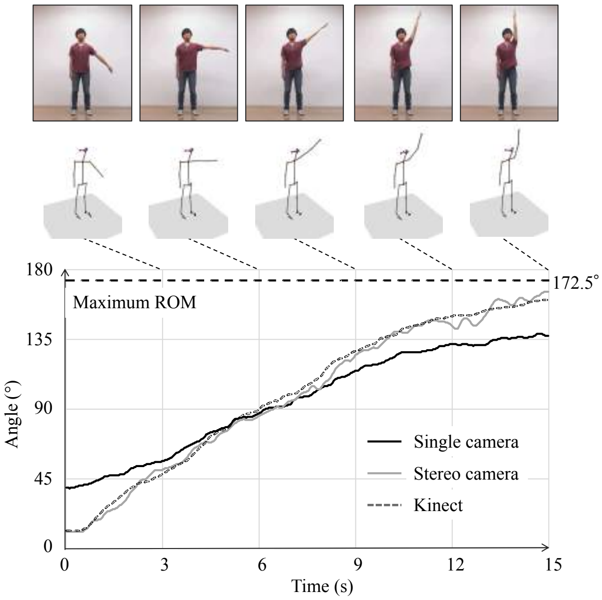
Joint	Motion	Single camera	Stereo camera	Kinect	Fluctuation
Shoulder	Forward Flexion	7.121°	7.624°	15.610°	High
	Backward Extension	0.797°	3.439°	1.131°	Low
	Abduction	2.463°	4.110°	3.384°	Low
	Adduction	3.652°	4.233°	1.993°	Low
Elbow	Frontal Flexion	5.605°	9.734°	5.791°	Moderate
	Side Flexion	2.152°	9.007°	3.918°	Moderate
Hip	Flexion	6.050°	4.129°	8.128°	Moderate
	Extension	1.110°	3.880°	6.501°	Moderate
	External Rotation	3.834°	3.935°	4.576°	Low
	Internal Rotation	2.522°	3.991°	3.050°	Low
Knee	Flexion	8.829°	7.105°	11.019°	High
Mean		4.0124°	5.5625°	5.9184°	
Standard deviation		2.57969°	2.32978°	4.28928°	

for two factors (joint movement and modality) was performed. The main effect of the joint movement was not significant, $F(1, 29) = 1.93$, $p = 0.18$. The main effect of the modality was also not significant, $F(1, 29) = 0.56$, $p = 0.46$. Hence, there are no significant difference in RMSE of measurement data with the three modalities.

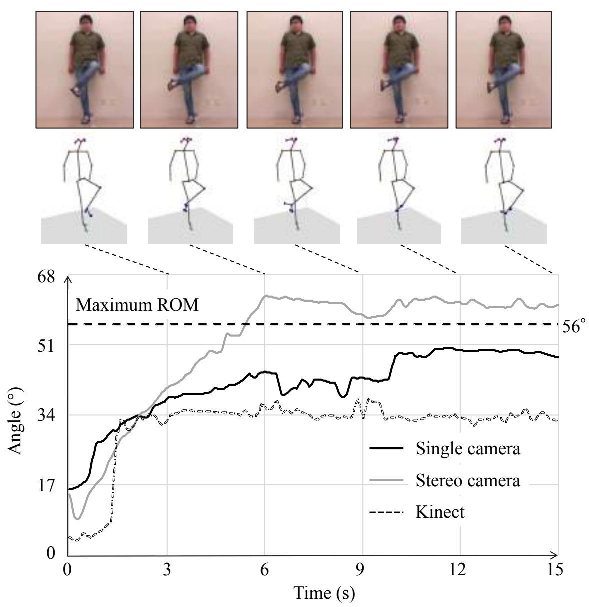
For further analysis, a visual interpretation is conducted to visualize the fluctuation in the resulting measurement during 15s motion. The amount of fluctuations were categorized according to the RMSE values: low, moderate, and high. Figures 4 ~ 6 show samples with low, medium and high fluctuations in the measurement results, respectively. The Dashed straight lines represent maximum ROM meausured by the goniometer. Figure 4 shows two samples of measurement results (shoulder abduction and hip internal rotation) with low fluctuations. Joint angles measured with the three modalities show similar trends. However, the Kinect measures the maximum ROM close to the goniometer. Figure 5 shows two samples of measurement results (elbow frontal flexion and elbow side flexion) with moderate fluctuations. Occlusion at the elbow joint has a slight effect on the Kinect measurements. Figure 6 shows two measurements of shoulder forward flexion with high fluctuations. Occlusion at the elbow joint greatly affects the Kinect measurements. During the experiment, the Kinect fails to measure some postures as indicated by the drops in the curve. Figure 7 shows Kinect's failure to measure hip extension and hip flexion in sleeping posture.

V. CONCLUSION AND FUTURE WORK

In this study, 2D human pose and 3D human pose estimation techniques were used to measure 3D body joints and calculate their ROM for various motions and postures. The former was applied to a stereo camera to estimate 3D joints using triangulation. The latter was solely applied to a single camera to estimate 3D joints by referring to the 3D human pose dataset. Based on our experiments, the stereo camera gives the best results with a small bias to the goniometer compared to the single camera and the Kinect.

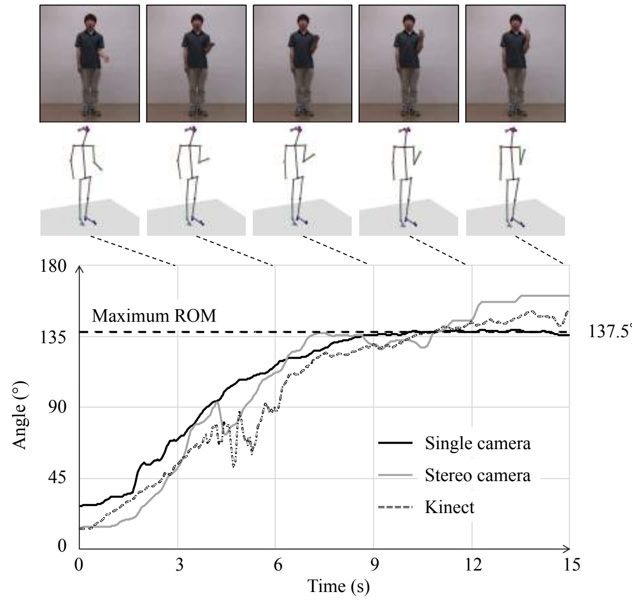


(a) Shoulder abduction

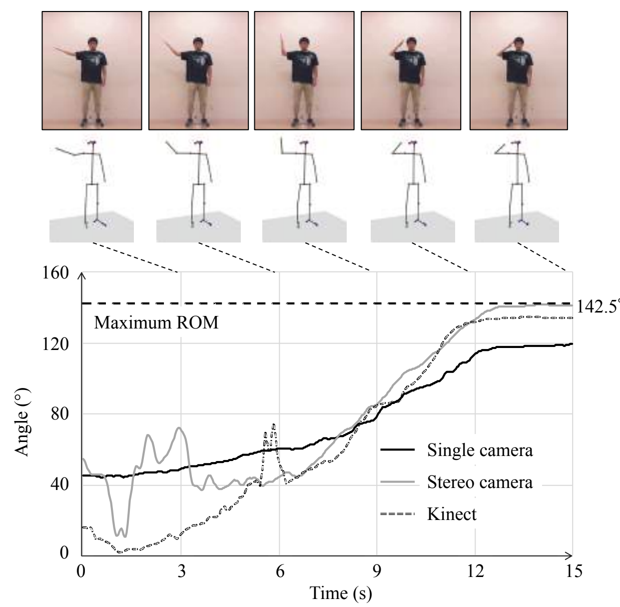


(b) Hip internal rotation

Figure 4. Samples of measurement results with low fluctuations.



(a) Elbow frontal flexion



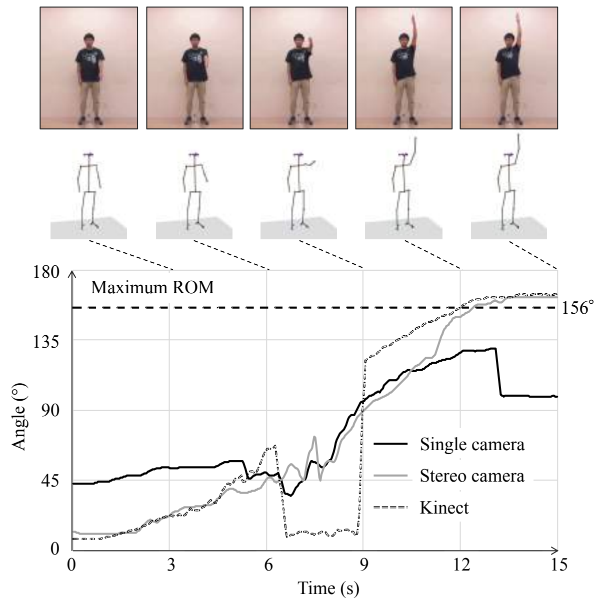
(b) Elbow side flexion

Figure 5. Samples of measurement results with moderate fluctuations.

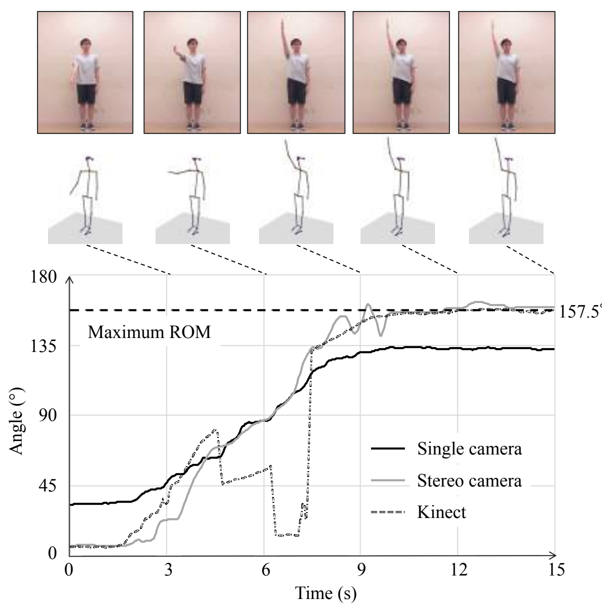
The Kinect surprisingly shows a high bias to goniometer in several measurement cases (Table II).

Occluded joints are a concern for 3D joint measurements. Since our method of calculating 3D joints with either a single camera or a stereo camera is based on the OpenPose library, in many cases, semi-occluded joint locations can be estimated to calculate the 3D locations. In contrast, the Kinect's algorithm does not take special measures against occlusion. Another drawback is that the Kinect needs to detect the posture of the whole body in order to properly detect a particular joint.

The 3D human pose based on a single camera has various applications. People can easily measure their posture everywhere using their own camera. It can be used to build a TR service, allowing clinicians to interact with patients in real-time. By promoting TR, we can expect to reduce the potential time and cost of rehabilitation services, especially for individuals who have economically disadvantaged.

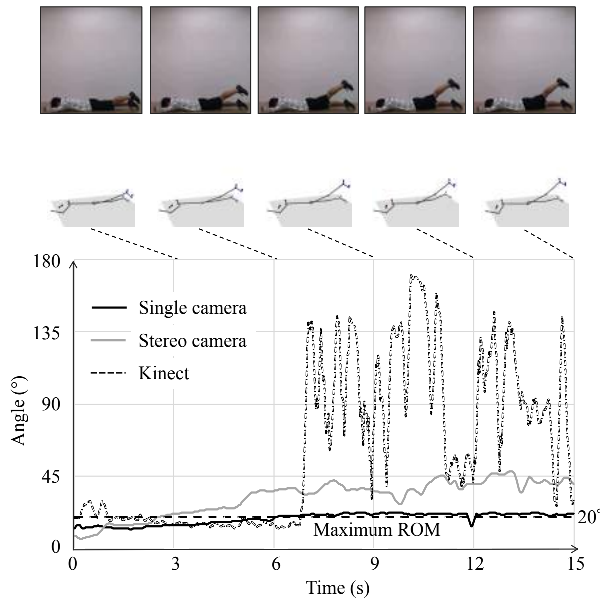


(a) Shoulder forward flexion

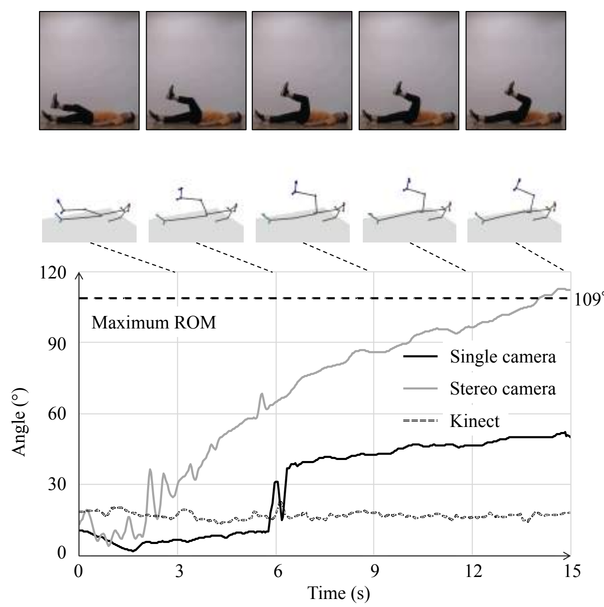


(b) Shoulder forward flexion

Figure 6. Samples of measurement results with high fluctuations.



(a) Hip extension



(b) Hip flexion

Figure 7. Samples of the Kinect's failures.

Our future work includes improving 3D human pose technology to enable body orientation measurement and detection of specific human behavior. We are also working on 3D human poses based on a 360-degree camera that allows measurement of 3D joints in all directions.

ACKNOWLEDGEMENT

We acknowledge the effort from the authors of OpenPose and 3d-pose-baseline to make 3D joint measurements using a single camera possible. This work was supported by the MIC/SCOPE #181602007. We wish to acknowledge the four anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

REFERENCES

[1] Y. Ono, O. D. A. Prima, T. Imabuchi, Y. Murata, H. Ito, and Y. Nishimura, "Assessment of Joint Range of Motion Measured by a Stereo Camera," eTELEMED 2019 : The Eleventh

International Conference on eHealth, Telemedicine, and Social Medicine, IARIA, pp. 23-29, 2019.

[2] A. M. Oosterwijk, M. K. Nieuwenhuis, C. P. van der Schans, and L. J. Mouton, "Shoulder and elbow range of motion for the performance of activities of daily living: A systematic review," *Physiotherapy Theory and Practice*, vol. 34, no. 7, pp. 505-528, 2018.

[3] A. M. Oosterwijk, M. K. Nieuwenhuis, H. J. Schouten, C. P. van der Schans, and L. J. Mouton, "Rating scales for shoulder and elbow range of motion impairment: Call for a functional approach," *PLOS ONE*, vol. 13, no. 8, pp. 1-13, 2018, https://doi.org/10.1371/journal.pone.0200710.

[4] B. P. Pereira, A. Thambyah A, and T. Lee, "Limited forearm motion compensated by thoracohumeral kinematics when performing tasks requiring pronation and supination," *Journal of Applied Biomechanics*, vol.28, pp. 127–138, 2012.

[5] K. Mitchell, S. B. Gutierrez, S. Sutton, S. Morton, and A. Morgenthaler, "Reliability and validity of goniometric iPhone applications for the assessment of active shoulder external rotation," *Physiotherapy Theory and Practice*, vol. 30, no. 7, pp. 521–525, 2014.

[6] DrGoniometer. http://www.drgoniometer.com/ [retrieved: December, 2019]

[7] T. Dutta, "Evaluation of the Kinect sensor for 3D kinematic measurement in the workplace," *Applied Ergonomics*, vol. 43, no. 4, pp. 645–649, 2012, doi:10.1016/j.apergo.2011.09.011.

[8] S. Aleesandro, C. Andrea, M. Matteo, and M. T. Lorenzo, "Kinect V2 Performance Assessment in Daily-Life Gestures: Cohort Study on Healthy Subjects for a Reference Database for Automated Instrumental Evaluations on Neurological Patients," *Applied Bionics and Biometrics*, pp. 1-16, 2018, https://doi.org/10.1155/2017/8567084.

[9] S. H. Lee, C. Yoon, S. G. Chung, H. C. Kim, Y. Kwak, H-w. Park, and K. Kim, "Measurement of Shoulder Range of Motion in Patients with Adhesive Capsulitis Using a Kinect," *PLOS ONE*, vol. 10, no. 6, pp. 1-12, 2015, doi:10.1371/journal.pone.0129398.

[10] J. Spörri, C. Schiefermüller, and E. Müller, "Collecting kin- ematic data on a Ski track with optoelectronic stereophotogram- metry: A methodological study assessing the feasibility of bringing the biomechanics Lab to the field," *PloS ONE*, vol. 11, no. 8, e0161757, 2016.

[11] K. Yonemoto, S. Ishigami, and T. Kondo, "The Method Guidelines for Range of Motion Measurement," The Japanese *Journal of Rehabilitation Medicine*, vol. 32, no. 4, pp. 207–217, 1995. (in Japanese)

[12] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," *Computer Vision and Pattern Recognition* (*CVPR*), 2014 IEEE Conference, pp. 1653–1660, 2014.

[13] A. Newell, K. Yang, and J. Deng, "Stacked Hourglass Networks for Human Pose Estimation," *Computer Vision – ECCV* 2016, pp. 483–499, Amsterdam, Netherlands, 2016.

[14] Z. Cao, T. Simon, S.E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *Computer Vision and Pattern Recognition* (*CVPR*), 2017 IEEE Conference, pp. 7291-7299, 2017.

[15] Y. Ono, O. D. A. Prima, and H. Ito, "3D Human Pose Estimation for Motion Analysis," The 80th Nation Convention of Information Processing Society of Japan, pp. 263-264, 2018. (in Japanese)

[16] H. J. Lee and Z. Chen, "Determination of 3D Human Body Postures from a Single View," *Computer Vision, Graphics and Image Processing*, vol. 30, pp. 148–168, 1985.

[17] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A Simple Yet Effective Baseline for 3d Human Pose Estimation," arXiv preprint arXiv:1705.03098, pp. 1-10, 2017.

[18] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325-1339, 2014.

[19] 3D-pose-baseline. https://github.com/una-dinosauria/3d-pose-baseline. [retrieved: December, 2019]

[20] O. D. A. Prima, T. Imabuchi, Y. Ono, Y. Murata, H. Ito, and Y. Nishimura, "Single Camera 3D Human Pose Estimation for Tele-rehabilitation," eTELEMED 2019 : The Eleventh International Conference on eHealth, Telemedicine, and Social Medicine, IARIA, pp. 13-18, 2019.

[21] N. B. Jain, R. B. Wilcox, J. N. Katz, and L. D. Higgins, "Clinical Examination of the Rotator Cuff," *Physical Medicine and Rehabilitation*, vol. 5, pp. 45–56, 2013.

[22] OpenPose, https://github.com/CMU-Perceptual-Computing-Lab/openpose. [retrieved: December, 2019]

[23] M. E. Huber, A. L. Seitz, M. Leeser, and D. Sternad, "Validity and Reliability of Kinect Skeleton for Measuring Shoulder Joint Angles: a Feasibility Study," *Physiotherapy*, vol. 101, no. 4, pp. 389–393, 2015.

[24] J. M. Bland, and D. G. Altman, "Statistical Method for Assessing Agreement between Two Methods of Clinical Measurement," *Lancet*, vol.327, pp. 307-310, http://dx.doi.org/10.1016/S0140-6736(86)90837-8, 1986.

A Collective Intelligence Framework for Lifestyle Management pro Mental Health Systems

Sakirulai Olufemi Isiaq
Department of Computing
Solent University
Southampton, United Kingdom
femi.isiaq@solent.ac.uk

Drishty D Sobnath
Research, Innovation and Enterprise
Solent University
Southampton, United Kingdom
drishty.sobnath@solent.ac.uk

Abstract— **Health Information Management Systems are becoming a central fixture in healthcare settings, but only a few frameworks exist to provide guidelines for the development of an innovative and sustainable system. This study employs a collective intelligence approach by corroborating knowledge, skills and contributions of various stakeholders to develop a Framework for Lifestyle Management pro Mental Health Management Systems (FLMMHS). A mixed-methods approach was employed and covered in two principal phases namely; document analysis (analysis of existing facts about mental health in the body of knowledge) and empirical analysis (experts' validation using four core parameters namely; efficacy, effectiveness, simplicity and flexibility). FLMMHS' components are apportioned into three core layers namely; Research Design Evaluation (RDE wrapper), Guidelines and Requirements (G&R), and Diagnosis Prevention Alleviation (DPA). While these components are flexibly designed to allow seamless system integration, its comprehensive representation serves as an implementation platform for the development of mental health systems. Although the suitability of FLMMHS for system development is based on the premise of lifestyle management for mental health, successful evaluation following qualitative and quantitative measures by expert judges impresses its aptness for the development mental health management systems.**

Keywords- mental health systems; lifestyle management; collective intelligence; Diagnosis Prevention and Alleviation; framework; standards; Guidelines and Requirements; barriers and facilitators; mental health support; FLMMHS.

I. INTRODUCTION

This work extends our existing research [1], which contends that lifestyle management approach as an effectual management practice for mental health disorders. A subsequent failing of aspects of human health such as intelligence, imagination and thought is considered a mental health disorder [2]. Globally, this health phenomenon is increasingly becoming popular with notable consumption of various aspects of human resources. Currently, a significant proportion of adult's population now suffers a form of mental health disorder with a record of about 26.2% of US adult population suffering a form 'serious' mental illness and 27% in the EU having mental issues [3]. Contemporary, records show that the gap between mental health treatment and its accessibility is increasingly becoming wider, currently estimated between 35% and 50% [4]. This gap is often widened by known factors including; stigma associated with mental health candidates, ineffective therapies, lack of adequate and awareness of mental health resources among other factors. Consequently, a significant proportion of mental health disorder candidates are left undiagnosed or diagnosed with no adequate attention or treatment.

Even more prominent is the traditional approach of dealing with these disorders (i.e., the process of Diagnosis, Prevention and Alleviation - DPA) also contributes to these noted shortcomings thus, aiding this increase perhaps, exponentially. Effectively, common DPA practice involves consultation with healthcare professionals, such as a General Practitioner (GP) or psychiatrist, who utilises standard tools such as, Hospital Anxiety and Depression Scale (HADS) [5], the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [6] and International Classification of Diseases (ICD) criteria [7], Generalized Anxiety Disorder (GAD-7) [8] among other appropriate tools to examine the candidate's situation. Nonetheless, almost half of the world's population lives in countries with fewer than two psychiatrists per a hundred thousand (100,000) people [9], therefore, optimal access to these practitioners is becoming highly impossible. Arguably, occurrences of prominent mental health disorders such as depression, bipolar affective disorder, anxiety, schizophrenia and dementia are rather deteriorative of one's health rather than instantaneous. Hence, on many occasions, appropriate lifestyle management may be better suited to avert the occurrence, or/and perhaps, manage these disorders effectively. Nevertheless, appropriate lifestyle management could be effectively accomplished with utmost acceptance of the life owner.

Mentioned earlier that traditional diagnosis and subsequent treatment of mental disorder conditions require the expertise of skilled medical practitioners who often are not readily available, recorded advances in technology can support bridging this gap. Not only are the advances of technology necessitating its prominences in various aspects of human lives including education, business and health, but the availability, size and power of hardware components and sensors are significantly aiding its presence. Besides, the existence of concepts such as artificial intelligence (Machine Learning, Natural Language Processing and Analytics) combined with the power of internet reinforces technology's importance in different areas of life, particularly health. For instance, in health, smartphone technology, which combines the communication and computation of a handheld device is used to facilitate point of care services using mobile computing [10]. Also, computer-based cognitive behavioural therapy (CBT) has proven to be clinically effective similar to a face-to-face treatment of various mental disorder [4].

Although technology-based approaches of health treatment and management have been characterised by low-cost and ability to reach a larger audience, the lack of traceable standards for many of these tools raises potent questions perhaps, in terms of their efficacy, effectiveness and acceptability. For instance, an online mental health diagnostic tool is prone to contests such as inaccuracies, exaggerations or misrepresentations that may influence accurate diagnosis thus, constituting false or negative effects. This work aims to curtail these aforementioned problems by presenting a research-oriented Framework for Lifestyle Management pro Mental Health Systems, FLMMHS. The framework adopts a collective intelligence approach by utilising the knowledge of literature, various stakeholders, existing systems and other sourced components among others.

The rest of the paper is structured as follows; Section II explores some existing mental health systems and the importance of a standard framework; Section III discusses existing mental health frameworks and Section IV explains the methodology comprising of empirical and document analysis. Section V discusses the barriers and facilitators of implementing mental health systems while Section VI describes the components of FLMMHS and its evaluation. Finally, conclusions and future work were presented in Section VII.

II. Current Mental Health Systems and Framework Importance

Health information systems are becoming a central fixture in the healthcare settings, but only a few standards are currently associated with the implementation and adaptation of these system solutions. The integration of mental health information systems into primary care is increasingly growing popularity as an effective means of treating mental health conditions and depression has been a good case example [11]. While benefits like lower healthcare costs, improved medication adherence, early diagnosis, and better patient/treatment follow-ups have been associated, healthcare system developments rely on suitable infrastructures, effective polices and perhaps cutting-edge technologies. More so, the acceptance of newer technologies for diagnosis, prevention and alleviation by major stakeholders is still very feeble for several reasons.

With technological advancements such as fast internet and 5G network, patients are now able to remotely receive real-time support/treatment for conditions that do not necessarily require the physical presence of physicians. For example, in recent times, the delivery of cognitive behavioural therapies (CBT) over the internet has been proven effective [12] and such internet-based protocols have been widely adopted to date [13]. Besides, numerous behavioural intervention technologies (BIT) - a technological application of behavioural and psychological intervention to address behavioural, cognitive and affective targets – are currently being adopted to treat or support physical and behavioural mental health disorders [14]. More recently, further advancements are being recorded in the areas of mobile hardware and sensor infrastructure. For instance, the traditional method of examining blood alcohol, nicotine and vitamin D levels are being substituted with technological hardware. Such advancements have been further magnified by organisations such as Samsung who recently released a smartwatch device for blood pressure measurements [15]. Additionally, other advancements that have been recently presented include devices such as BACtrack, Digital Health Age and others that have been mentioned in different studies [16] [17] [18] for carrying out vitamin D and other physiological measurements.

The readiness of smartphones and other handheld devices for e-health has severally been explored with no exception to its utilisations for the management of mental health situations. For example, face-to-face therapy presented in [4] adopts this technique for DPA activities. Additionally, momentary and intervention triggered configurable commonly utilised for assisting distressed patients, in remote locations further justifies the importance of technology for the attainment of urgent treatment particularly, in areas such as mental health. In recent times, numerous mobile applications have been available for mental health management (i.e., diagnosis, prevention and alleviation) however, many of these applications are prone to risks including mismanagement, misinterpretation, misdiagnosis or recommendation of unsuitable alleviation techniques. Not only could these risks worsen the situation of vulnerable users, but they could lead to potential health relapse.

Google Play and Apple Store play host platforms for numerous mental health apps analysed by Shelton, Psycom (top 25 mental health apps in 2018) in 2018 [19]. These mental health apps were categorised into general mental health, addiction, anxiety, suicide prevention, depression, bipolar disorder and obsessive-compulsive disorder apps among other categories. In these categories are apps such as Self-help for Anxiety Management (SAM), CBT Thought Record Diary, MoodKit, IMoodJournal and Talkspace Online Therapy among others [19]. Although some of these apps are accessible to users at a cost, others are available for free to improve mood, life-quality and user's mood among other activities. For example, "Depression CBT Self Guide" is a pocket guide that helps users to learn about CBT and how to cope with depression; it also allows users to measure the severity of depression, develop positive thoughts and encourages meditation practice. Other highly rated apps in Google store [20] include "Positive Thinking" or "Operation Reach Out" which provides support via different resources such as hotlines, videos to military personnel and veteran who suffer depression. Also "Moodkit" in Apple Store provides over two hundred mood improvement activities to support distressing thoughts.

Apart from the aforementioned apps, other mental health mobile apps that present users-functionalities such as, diagnosis and progress tracking include Moodtrack diary [21] and depression screening test [22], which tracks activity progression and diagnosis respectively. Pacifica [23] and Relieve depression PRO [24] provide prevention and alleviation functionalities, although without diagnosis or personalisation functionalities. While many of these apps support users via a variety of techniques, there is no known standards or framework typical to these systems. Moreover, more profound information about their implementations or policies adopted for deployment are not generally publicised to the best of our knowledge.

III. Existing Mental Health Frameworks

Several service-oriented frameworks provide guidance and coordination supports for mental health care delivery to enhance patient experience and service quality. Numerously, different health bodies have developed various service-oriented frameworks, which cater for different mental health aspects. For example, the Organisation for Economic Co-operation and Development (OECD) in the United Kingdom concentrates on diagnosis and assessment, access to mental health services, personal well-being and care programmes, and treatment of patients according to defined standards [25] [26].

Occasionally are these frameworks revised to satisfy the ever-evolving stakeholders' requirements. For instance, a recent review of the Department of Health service framework for mental health [25] aided the updates, which include (i) deepening the health and social care services integration; (ii) health and social well-being improvements; (iii) promote evidence-based practices and (iv) multidisciplinary and intersectoral workings enhancement [25]. Although the framework builds on the 2010 version, it offers a more streamlined approach to include service and experience indicators. Also, the values and principles of the revised framework are based on the recommendation of National Institute for Health and Care Excellence (NICE), which focuses on safe and effective care, patient's experience and recovery principles.

Another example of existing frameworks is the Community Mental Health Framework for Adults by the National Health Service [27]. Its implementation breaks down the current barriers as follows: (i) mental health and physical health, (ii) health, social care, voluntary, community, social enterprise organisations and local communities, and (iii) primary and secondary care in order to provide an integrated, and personalised service. Other positives aimed at this framework is enabling candidates with mental health complications: have unhindered access to mental health care, manage their conditions, move to individualised treatment plans, and contribute to the local community. The goal of the framework is similar to the Service-based Framework focusing on local communities' needs.

Besides, in the United States, mental health services follow the Donabedian framework model - a quality assurance-based framework, which considers the organisation and structure of the health care system delivery with the aims of providing better health care outcomes [28] [29]. The National Institute of Mental Health (NIMH)'s Research Domain Criteria (RDoC) framework is not only considered as a diagnostic tool for mental health problems but also serves as a basis of understanding the biological, social, developmental and environmental factors that may affect individual psychological functions. Often, is it used alongside other models such as, the International Classification of Diseases (ICD) [7] or the Diagnostic and Statistical Manual of Mental Disorders (DSM) [30].

While many of these frameworks are commonly considered to play effective roles in mental health service, management and delivery; their deployments for system development have not considered or published to the best knowledge of the authors. For example, technological factors such as interoperability, usability, technology acceptance or scalability are not explicitly reflective in most of these frameworks. Hence, this study bridges the identified gap by combining service and technical oriented requirements (incorporating medical and technical principles) to develop a system deployable framework for system developers. Taking a collective intelligence approach, the framework proposes to serve as a standard for mental health management system development following a lifestyle management approach. The following section illustrates the adopted methodology and the role of collective intelligence in the development and evaluation of FLMMHS.

IV. Methodology

Different study methods may be adopted at various stage of a research life cycle. For this work, a multifaceted mixed-methods approach was adopted, albeit, classified into two main phases namely; 1) documentary analysis and 2) empirical analysis. While the phase of documentary-study involved the analysis of existing body of facts in the knowledgebase, the empirical analysis phase involved qualitative and quantitative evaluation of derived facts from documentary analysis and the developed framework. Therefore, the Framework for Lifestyle Management pro Mental Health System is developed following a concept of collective intelligence i.e., combining the knowledge, skills and collaborative outputs of diverse sources and stakeholders.

A. Document Analysis

Enhanced methodical approach of literature analysis, PRISMA [31] was utilised to collect intended relevant contents. The principles of PRISMA was adopted in four cardinal stages that include identification, screening, eligibility and inclusion. As these cardinal stages were carried out iteratively, varying keywords were employed at different stages to distinctly improve robustness of document inclusion. Key terms including collective intelligence, barriers, facilitators diagnostic, prevention and alleviation approaches about mental health were utilised for extracting facts from reputable platforms. Employed platforms include British Library [32], German National Library of Science and Technology [33], Google scholar [34] among others. Documents such as white papers, journal and conference articles were examined to elicit mental health efforts. These efforts were further analysed in order to classify key mental health concepts and principles. In addition to general mental health efforts and concepts, specific common mental health disorders such as depression, bipolar-disorder, sleeping-disorder, and schizophrenia, among other conditions were singularly and commonly evaluated. Key concepts including definitions, symptoms, impacts, methods of diagnosis, prevention and alleviation were also intensely examined. In additions, standards and policies of health system development, system design methods, guidelines and system requirements were appraised and suitably categorised in line with mental health system development as framework components. Separate from the efforts of the framework development was a further scientific appraisal of the framework's suitability and reliability as explained in the empirical analysis section.

B. Empirical Analysis

To validate framework rigor, multiple empirical methods (qualitative and quantitative) including survey, discussion groups and data triangulation were applied. Also, qualitative and quantitative validity criteria that include face and content validity index were adopted. A thematic survey was utilised primarily to gather regular user's perspectives and their understanding of technology adoption for managing mental health and its disorders. While there are no general exclusion criteria for the survey participants, associated limitations of online survey limit the survey respondents to technological informed candidates with basic knowhow of electronic system operation (i.e., basic understanding of mobile or Web technologies). The outputs of the survey and documentary analysis did not only aid the development process of the framework, further were the utilised empirical methods (secondary survey, discussion group and data analysis) to strengthen its evaluation. Twenty-six expert judges (software engineers and developers of varying level of expertise) completed a validation survey followed by three discussion groups of distinct expert judges (with varying level of expertise). Although the outputs of the survey and discussion groups were later triangulated to justify the validity of the adopted instruments, the reliability of questions was subsequently measured using Cronbach's alpha technique. The following section considers collective intelligence and its significance as a tool for the framework development.

C. Collective Intelligence

The concept of Collective Intelligence embraces the utilisation of skills, knowledge, sources and collaboration of multiple stakeholders to solve a problem. Although the concept has existed for decades, the recent conception of its combination with machine learning principles has been catalytical for the generation of newer and more interesting facts. Nowadays, advances recorded in technological visions, computing powers and machine learning abilities (collecting/analysing data from millions of records over the cloud) has opened up new possibilities of finding solutions to modern-day problems [35]. Interestingly, the health domain has been one of the various domains that benefits from technological sophistications particularly, through collaborative measures between individuals and companies leading to the development of smart algorithms. For example, health organisations now uses machine learning concepts to predict disease and symptom deterioration, hence, preventive measures are taken to reduce hospitalisation and mortality rates. Taking to these benefits, the use of machine learning techniques on collectively gathered facts will not only aid the accumulation of newer knowledge yet, exponentially.

Considerably, this work considers the five factors of collective intelligence (Autonomous Commons, Balance, Focus, Reflexive and Integrate for action) highlighted crucial to solving problems by Mulgan [36] as imperative for the development of effective mental health system. Accordingly, an intricate collection and analyses of existing mental health data was utilised to accomplish a rich background for developing a framework, which a robust mental health management system can be based.

Indicative barrier from literature evidenced that mentally disordered candidates tend not to seek support when experiencing emotional or mental health difficulty. This barrier is decidedly prompted by stigma, negative perceptions, self-reliance, and lack of awareness as triggering factors [37]. Collectively embracing gathered facts for the development of health management systems notably englobes factors including human resources, finance, medicines, technology, service infrastructure. Also included are intangible assets of ideas and interests, relationships, policies, values and people-centered norms as identified by Glenn [38]. This encirclement can further unveil the benefits and limitations of electronic health care systems, and thus, barriers and facilitators between the users and the systems are farther acknowledged. This work combines concepts from Glenn and Mulgan's five factors of collective intelligence, to analyse mental health care services and system to derive combining factors shown in Figure 1. Furthermore, barriers and facilitators of mental health and disorder management system development are highlighted in the following section.

Figure 1: Englobed Factors of Collective Intelligence

V. BARRIERS AND FACILITATORS

Human's mental health is considered a vital aspect of general health yet, robust measures to cope with the existing burden of mental health disorders are not currently possessed by health care providers. Since mental health conditions range in a spectrum of mild to extremely severe states, different conditions require distinct management plans i.e., diagnosis and treatment plans, hence, the seamless integration of mental health services into primary care to bridge treatment gaps is necessitated [39]. For instance, the National Health Service (NHS) in the United Kingdom, UK provides mental health services (including dealing with problems such as drug, alcohol addiction and psychological therapies - IAPT) for patients, but accesses to most of these services are through local General Practitioner, GP referral [40]. Conversely, effective management of contemporary mental health burdens necessitates seamless user-treatments interface, powered by advanced technology sophistication to handle with distinct requirements robustly. While technology can facilitate seamless management of mental health conditions, some barriers are identified to hinder diagnostic and treatment procedures. Normally, these barriers and facilitators are identified using the Supporting the Use of Research Evidence (SURE) framework. The framework aims to support mental health system design through the involvement of a wide range of stakeholders [41] and was mainly developed for implementing health system changes and support policymaking in Africa [39]. Highlights in Table 1 are various barriers and facilitators for implementing mental health care systems. Among these factors are segregated care, lack of finance and resources, policymakers, privacy and acceptance among other issues are identified hindering factors (barriers). However, data access, end-user's motivation, infrastructure sophistications and effective collaboration between medical and IT professionals could facilitate a successful implementation of a mental health management system. Collectively, knowledge accrues from document analysis and the evaluation of barriers and facilitators are considered for the derivation of the components of Framework for Lifestyle Management pro Mental Health Systems (FLMMHS) as described in the result section.

Table 1: Barriers and Facilitators of Mental Health Systems

Barriers	**Evaluation from Research Work**
Segregated Healthcare	The perception that mental health is separated from mental health systems informs the lack of integration of both concepts [11], [42]
Financial Resources	The limitation in budget leads to the lack of developing integrated mental health care systems. [11]
Bureaucracy (Policy makers/planners)	High cost of medical care attributed some policies procedures and decision making makes the implementation of mental health systems harder [43]
Norms and Standards (HL7)	There are certain norms and standards that need to be followed when implementing mental health system for system interoperability which can be implemented with the right experts' skills. [39] [44] (World Health Organization., 2012)
Privacy and Security	There are concerns about the privacy and the security of the systems that deal with confidential and personal information. [46]
Technology Acceptance/ Change	There are usually mixed views about the use of new information systems i.e., organisation staff very often show unwillingness to adapt to changes or lack of time/interests.[46]
Credibility/Appro priateness	The appropriateness of technology needs to be assessed in order to solve particular problems faced by mental health professionals. There is little evidence that supports the efficiency of tools used to support patients suffering from mental health issues.[46] [47]
Facilitators	**Evaluation from Research Work**
Technological Infrastructure	Mental health systems can be integrated with other sub-systems through the exploitation of technology such as IoT, cloud and 4G. [48]
Knowledge/ Skills	The knowledge and skills of different experts such as Psychologists, Physiotherapist, GPs, Nurses, Patients and Software Companies can help to design mental health system to suit the needs of patients. (World Health Organization, 2012)
Motivation	Current healthcare professionals should be able to understand the benefits of using such a system to be motivated to use it. [49]
Data Access	Integrated mental health care systems may facilitate access of patients' data [48] at any time and from anywhere as far as there is an internet connection.
Training	Training can facilitate the knowledge exchange and helps motivate staff to use the new system. [50]
Collaboration	The collaboration of different stakeholders such as pharmacists, psychologist, nutritionists, GPs, designers, programmers, can facilitate the implementation of a mental care system. [45][50][51] [48]
Resources/ Government Strategies	With a time frame and appropriate resource allocation, strategic plan with clear objectives can facilitate the implementation of mental health system. [52]

VI. RESULTS

The outcomes of this work are perceived in two perspectives namely; FLMMHS (mental health system development framework) and its evaluation. The framework section explains the core components of FLMMHS including the Research Design Evaluation (RDE) wrapper, Guidelines and Requirement (G&R), and the Diagnosis Prevention Alleviation (DPA) layers. Subsequently, the evaluation section uncovered the rigour of framework through analysis of expert judges' submissions on the framework's suitability.

A. Framework for Lifestyle Management pro Mental Health System (FLMMHS)

An iterative document analysis of mental health literature, concepts of system development and the derivation of facilitators and barriers in conjunctions with the findings of the user survey aided the development of a Framework for Lifestyle Management pro Mental Health Systems (FLMMHS). Not only do these factors collectively derived the framework's composition, but also aided the components' classification into layers to improve the robust implementation of a mental health management system. Although FLMMHS's components are apportioned into three layers namely; RDE wrapper, Guides and Requirements (G&R), and DPA layers, the layers are flexibly integrated to encourage seamless deployment for system development. The following section describes FLMMHS's components (as depicted in Figure 2) according to their corresponding layers.

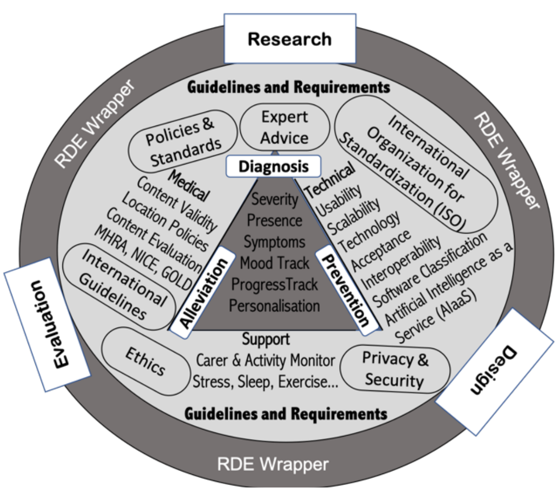


Figure 2: Framework for Lifestyle Management pro Mental Health Systems

- RDE Wrapper

The RDE wrapper is the outermost layer shielding all components of the framework. The wrapper consists of three basic features namely; Research, Design and Evaluation. Fundamentally, developers are expected to consider the wrapper layer as a development navigator of a lifestyle management system, irrespective of the type of mental health disorder or system. The **R**esearch component considers the aspect of mental health (i.e., states, disorders and types among others), users (patients, practitioners and other stakeholders) and the systems. Considering the dissimilarity in the types of mental health disorders and the heterogeneity of the stakeholders (candidates, users, medical practitioners etc.), developers are obligated to conduct intense research about specifics of mental health i.e., in terms of system requirements and the needs of the potential system users. Whereas, the **D**esign component guides the process of system development in aiming that a mental health management system is extensively supportive such that they are less demanding perhaps, moderately automated. For example, the user interface of a mental health management system should be undoubtedly simple, intuitive and usable to avoid any aggravation of the user' states. Concepts such as colour impacts, fonts, navigation, perception and overall comprehension of the system are comprehensively thought through, mostly in-line with the requirements of the intended users, i.e., significant depth of design is considered. Finally, the **E**valuation components is an appraisal mechanism for the **R**esearch and **D**esign elements, which ensures system's suitability for the intended users. The RDE wrapper is considered a kernel for developers of mental health management systems, irrespective of the disorder or intention.

- Guidelines and Requirements (G&R)

The G&R tier is an intermediate layer that binds the RDE and DPA layers of the framework. It considers technical and medical requirements of mental health management system development by employing major components namely; Policies & Standards, Experts Advice, Ethics, and Security & Privacy. The layer permits seamless integration of support, medical and technical requirements as a single module for mental health state management.

Taking that standards and medical guidelines vary by country or region [53], medical requirements of the intended location (country or region) of deployment must be appropriately implemented. For example, in the United Kingdom, the National Institute for Health Care [54] recommends that healthcare professionals provide information, advice, diagnosis and treatment for patients, while Mental Health in America (MHA) develops guidelines to identify mental health measures [55]. Also, the APA provides evidence-based recommendations regarding psychiatric disorders assessments [56] and the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) to diagnose and/or classify mental health disorders. Therefore, the G&R sector considers medical tools and contents as stringent suited for intending location.

Similarly, technical requirements of the system are considered in terms of generics and specifics. Further considered are standards in terms of software and hardware requirements, system accessibility, scalability, interoperability and technology acceptance are to be rigorously considered; and similarly, the Security & Privacy of the system users. Also, the sensitivity of mental health conditions demands robust Ethics and Security architecture for FLMMHS based management systems. Therefore, the data handling process of the system should aim user's data privacy, encryption algorithms and access level functionalities amongst other features.

Finally, developers need understand that mental health systems are not purposed to displace practitioners but to play enhanced role in handling this health situation, therefore easier communication mechanism between stakeholders, i.e., practitioner (GP, psychiatrist, career) and patients should be stringently facilitated within the system. Other support features include automatic sleep management, diet, exercise and other lifestyle factors management. The following section expands the DPA layer, its components and roles in supporting the framework particularly, in accomplishing lifestyle management for mental health conditions.

- Diagnosis Prevention Alleviation (DPA) Layer

Lastly, innermost layer of FLMMHS', DPA, consists of three foremost components namely; **D**iagnosis, **P**revention & **A**lleviation. These components consist of other sub-components including tools for determining the presence of mental health disorder, its severity, corresponding symptoms, prevention and progress managements. Also, these subcomponents embodied the traditional process of examining symptoms and severity of mental health disorders. Therefore, mental health management systems considering diagnosis should incorporate a means of deducing mental health symptoms and corresponding severity. Furthermore, the framework embraces flexible techniques to allow developers adopt a preferred standard diagnostic tool such as DSM-5 [6], ICD-10 [57], the Beck Depression Inventory, BDI [58], PHQ-9 [59] and GAD-7 [8], among other tools. To maintain the flexibleness of FLMMHS, the **D**iagnosis component connects the prevention and alleviation components. Taking that the framework focuses on lifestyle management approach of mental health disorders, emphases are laid on prevention and alleviation methods through lifestyle management, therefore factors such as nutrition, exercise and sleep rate among other factors are considered core for system based on this framework. These core factors are associated with several studies and scientific findings, hence, are deemed important constituents for lifestyle approach of management. For example, the study conducted by Jacka et al, reflected significant association between diets and mental health management [60]. Similarly, Tanaka [61] and Freeman [62] found correlations between sleep rate and mental health state. Although cues are taken from these studies, the framework flexibility allows the inclusion of further elements to the DPA component. Finally, personalisation and progress tracking are considered for all DPA components i.e., the process of diagnosis, prevention and alleviation are personified, therefore, FLMMHS based system mandates progress-track functionality to manage associated lifestyle activities, perhaps, in correlation with the user's mental health state. The following section discusses FLMMHS' evaluation to ensure its robustness for mental health system development.

B. *FLMMHS Evaluation*

To determine the framework's robustness for system development, a holistic face validity evaluation was conducted using expert judges' submissions. The evaluation adopts mixed methods (qualitative and quantitative techniques) following four key parameters namely; efficacy, effectiveness, simplicity and flexibility. Explicitly, survey and focus group instruments were independently employed, and the outcomes of both methods were triangulated to improve the confidence of the derived results. The survey procedure and results were discussed in the survey section as follows.

- Survey

Twenty-two expert-judges (software developers with varying level of expertise) participated in a validation survey comprising of a derived 22 item-question. A significant proportion of the questions are multidimensional that transcend multiple evaluating parameters, but a few were unidimensional. Table 2 illustrates the item-questions and corresponding parameters intended to evaluate. For example, the 'knowledge and expertise' dimension aims to identify the level of expertise and knowledge of expert judges through questions 1, 2, 3 and 5, while question 4 primarily focused on measuring participants' perception of the framework and its suitability. Twenty-one participants completed their evaluation by providing answers to all item-question, but one participant did not complete the item-question and thus, was excluded from the data analysis. Although some questions are multidimensional, the questions are validated to be internally consistent and reliable as Cronbach's alpha score of 0.95 was recorded. Figure 3a and 3b depict the overall perception's rating (Question 4 only) and evaluation of parameter rating respectively. The average perception rating of 0.61 was recorded to evidence judges' satisfaction. Also, correlation analysis was conducted to understand the judges' bias in terms of knowledge and understanding of mental health and software and system development.

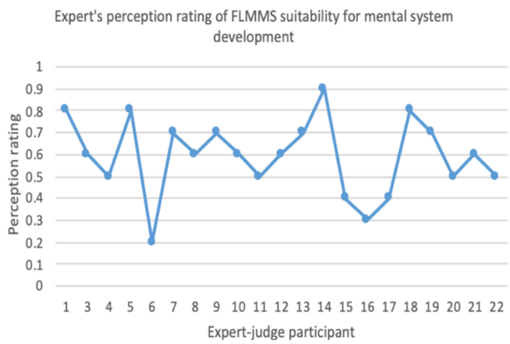


Figure 3a: Experts' Perception of FLMMHS

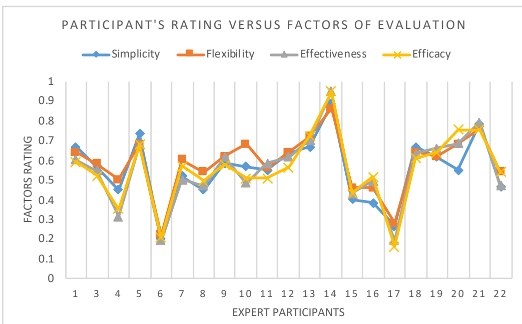


Figure 3b: Experts' Rating of Evaluation Factors

The Pearson correlation coefficient of the overall perceptions and combination of evaluating parameters (i.e., rating in flexibility, simplicity, effectiveness and efficacy) is 0.79, indicating a high positive correlation [63]. Not only does the high correlation recorded signifies robust suitability of the framework for mental health system development, but it also strengthens the success of the evaluating parameters. Figure 3c represents the correlation between the perceptions and evaluating factors, while Figure 3d represents the correlation between participant's knowledge & expertise and their perceptions.

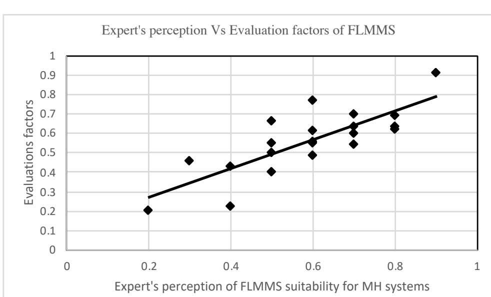


Figure 3c: Correlation Analysis of Perception

A positive correlation coefficient of 0.23 is computed for experts' knowledge/expertise versus their perceptions, indicating low or perhaps, a negligible correlation between the expert's rating of the framework's suitability for development and their knowledge/expertise. Also, this represents a positive reflection of the suitability of adopting the framework for mental health system development, irrespective of the level of expertise whether basic, intermediate or advanced.

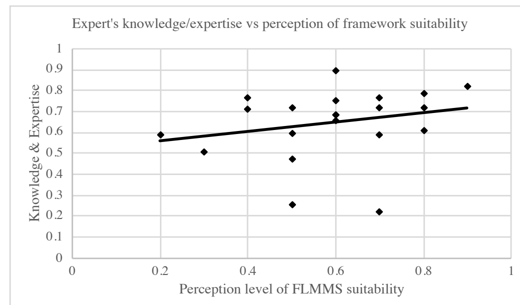


Figure 3d: Perception of FLMMHS' Suitability

Overall, the face validity of the framework is good with an acceptability index of 0.61. Besides, supplementary focus group sessions were conducted to examine the recorded results of the survey on the acceptance of the framework as discussed in the following section.

- Focus group

Following, the survey analysis, a supplementary qualitative analysis was conducted to improve evaluation confidence of FLMMHS. Three discussion-group sessions were carried out with expert-judges in groups of five, eight and seven. These sessions lasted an average of 11 minutes 46 seconds and the data was transcribed using a professional transcriber. The transcribed data was processed using NVivo software [64] for coding and thematic analysis. Firstly, the first hundred frequent words of the transcript data were deduced by taking out the sight or joining words such as *the*, *and*, *this* and *is* among others. (See Figure 4 for a cloud representation of contents) to enhance content analysis.

Figure 4: Cloud Representation of Content Analysis

Subsequently, a thematic analysis was conducted by coding the data in themes similar to the survey's key themes i.e., knowledge & expertise, flexibility, simplicity, effectiveness and efficacy. Additionally, two newly themes were derived namely; expert's perceptions and other factors, which were classified by related texts, accordingly. For example, key texts such as; think, ideal, quality and suitability relate to the perception's context thus, corresponding references were clustered appropriately for the perception of each participant group. Appendix I, Table 3 indicates the number of coding terms per participant group for each theme.

Table 2: Thematic representation of codes for each focus group

Evaluating Criteria	Group One	Group Two	Group Three
Flexibility	8	7	4
Simplicity	7	8	5
Efficacy	11	7	12
Effectiveness	19	10	21
Expert_MH_**K**nowledge	8	6	1
Expert_**E**xpertise	21	19	22
Expert_**P**erception	14	21	15
Other_**F**actors	8	8	6

Another meaningful observation is the derived correlation between the key themes. While efficacy of the framework is regarded the most important theme based on experts' submissions. Although all themes are intertwined through shared key terms, the efficacy term appeared connected with all themes as noted with tree representation of the content in Figure 5.

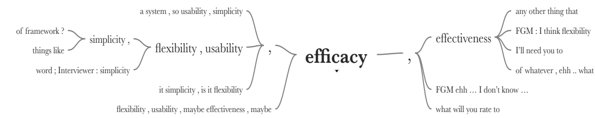


Figure 5: Tree Representation of Content Analysis

Finally, a matrix representation of the coded data shown in Appendix I, Table 4 indicates a high relationship between effectiveness and efficacy, simplicity, expertise and level of perception. There is quite low association between expertise and mental health knowledge, which further justifies easier deployment of the framework, irrespective of mental health practicing knowledge.

Overall, the results from both qualitative and quantitative evaluations indicate the suitability of the framework for mental health system deployment pro lifestyle management. The framework's robustness is further evidenced through expert judges' review on fundamental features of system development (i.e., simplicity, usability, flexibility, efficacy and effectiveness) with no preference to their level of expertise. Following is an insight on the practical deployment of FLMMHS for mental health system development.

C. *Practical deployment of FLMMHS*

Mobile systems offer a range of self-management apps, digital consultations and digital-enabled models of therapy for patients suffering from mental health conditions, but these systems often segregated from other services. FLMMHS is expected to be deployed by software engineers or Web developers working closely with healthcare or health-related professionals. The conceptual model can easily be translated to a more practical integrated solution as a software library or Web service or complete system solution. Barriers previously identified in Table 1 have thus been considered when designing the framework to include major factors such as HL7 standards for storing data. Other factors considered include privacy and security of patients' confidential information and evaluation of the system expert stakeholders to understand the deployment acceptance.

Healthcare services looking at mental health care delivery systems can adopt FLMMHS concept to implement software solutions. More so, various bodies explications can be integrated to improve the process of strategic design, delivery and development FLMMHS based systems. Systems can be reinforced with digital clinical decision-making tools, which can help healthcare professionals in early diagnosis of mental health-related issue. With an integrated care approach, improving access to psychological therapies or some other supports can be made easier. Knowledge, skills and competences can also be shared across multiple disciplines if the designed system is based on FLMMHS. A comprehensive implementation of FLMMHs based system is aimed in future work as illustrated in the following section.

VII. Conclusion and Future Work

Over the years, advancements in technology have led to its substantial deployment of clinical and health management systems. Technology adoption for managing lifestyle in line with human mental health is increasingly becoming popular in modern society. However, several existing mental health systems were developed with no known development reference. This work classifies lifestyle management as a potent approach for mental health management and presents a **F**ramework for **L**ifestyle **M**anagement pro **M**ental **H**ealth **S**ystem based on a collective intelligence approach. Indication from expert judges in terms of holistic perception, i.e., flexibilities, simplicity, efficacy and effectiveness portray a good acceptance index thus, indicating the suitability of the framework for mental health system development. Also, there is negligible correlation between judges' level of expertise and their perception of the framework's efficacy, therefore, mental health management systems' development is made easy via FLMMHS' adoption, irrespective of expert level or mental-health knowledge. Although the framework has been successfully evaluated by experts, future work aims on a comprehensive technical implementation of the framework, which can be deployable as software library, Web service or compact mobile/Web management system. Such implementation permits a further empirical validation of the effectiveness of mental health systems for managing mental health and its disorders.

References

[1] S. O. Isiaq and K. Hamling, 'A Personalised Lifestyle Management Framework for Decision Support System in Mental Health', in *The Eleventh International Conference on eHealth, Telemedicine, and Social Medicine (e-TELEMED)*, 2019, pp. 96–101.

[2] National Institute of Mental Health, 'Transforming the understanding and treatment of mental illnesses', 2019. [Online]. Available: https://www.nimh.nih.gov/health/statistics/mental-illness.shtml. [Accessed: 01-Mar-2019].

[3] C. Samele, S. Frew, and N. Urquía, 'Mental health Systems in the European Union Member States, Status of Mental Health in Populations and Benefits to be Expected from Investments into Mental Health', 2013.

[4] K. Smith, 'Mental health: A world of depression', *Nature*, 2014.

[5] A. S. Zigmond and R. P. Snaith, 'The Hospital Anxiety and Depression Scale', *Acta Psychiatr. Scand.*, 1983.

[6] W. J. Earle, 'DSM-5', *Philos. Forum*, 2014.

[7] World Health Organization, 'The ICD-10 Classification of Mental and Behavioural Disorders. Clinical description and diagnostic guidelines', *Int. Classif.*, 1992.

[8] R. L. Spitzer, K. Kroenke, J. B. W. Williams, and B. Löwe, 'A brief measure for assessing generalized anxiety disorder: The GAD-7', *Arch. Intern. Med.*, 2006.

[9] World Health Organisation, 'Global Mental Health, How Are We Doing?', 2015.

[10] M. H. Trivedi *et al.*, 'Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice', *Am. J. Psychiatry*, vol. 163, pp. 28–40, 2006.

[11] D. L. Marais and I. Petersen, 'Health system governance to support integrated mental health care in South Africa: Challenges and opportunities', *Int. J. Ment. Health Syst.*, vol. 9, no. 1, 2015.

[12] G. E. A. Ndersson, P. I. M. C. Uijpers, P. E. R. C. Arlbring, H. E. R. Iper, and E. R. I. K. H. Edman, 'Guided internet-based vs. face-to-face cognitive behaviour therapy for psychiatric and somatic disorders', *J. World Psychiatr. Assoc.*, no. October, 2014.

[13] R. Alnanih, O. Ormandjieva, and T. Radhakrishnan, 'Context-based user stereotype model for mobile user interfaces in health care applications', in *Procedia Computer Science*, 2013.

[14] D. C. Mohr, S. M. Schueller, E. Montague, M. N. Burns, and P. Rashidi, 'The behavioral intervention technology model: An integrated conceptual and technological framework for ehealth and mhealth interventions', *Journal of Medical Internet Research*. 2014.

[15] Samsung, 'Samsung Galaxy Watch Active', 2019. [Online]. Available: https://www.samsung.com/global/galaxy/galaxy-watch-active/. [Accessed: 28-May-2019].

[16] D. C. Mohr, M. N. Burns, S. M. Schueller, G. Clarke, and M. Klinkman, 'Behavioral Intervention Technologies: Evidence review and recommendations for future research in mental health', *Gen. Hosp. Psychiatry*, 2013.

[17] BACtrack, 'The Leader in Breathalyzers', 2018. [Online]. Available: https://www.bactrack.com/. [Accessed: 28-May-2019].

[18] Digital Health Age, 'Wearable SmartStop to help smokers quit', 2016. [Online]. Available: http://digitalhealthage.com/smartstop-helps-smokers-quit/. [Accessed: 28-May-2019].

[19] J. Shelton, 'Top 25 Mental Health Apps for 2018: An Alternative to Therapy?', *PSYCOM*, 2018. .

[20] Google Play, 'Best Depression Apps of 2018', *Google Android Store*, 2018. [Online]. Available: https://play.google.com/store/apps/details?id=com.excelatlife.depression&hl=en_GB. [Accessed: 08-Nov-2018].

[21] Psych Central Research Team, 'Depression Test', 2019. [Online]. Available: https://psychcentral.com/quizzes/depression-quiz/. [Accessed: 28-May-2019].

[22] PsyberGuide, 'MoodTrack Diary: Mood Tracker', 2018. [Online]. Available: https://psyberguide.org/apps/mood-tracking-journal-and-diary/. [Accessed: 28-May-2019].

[23] BC Mental Health & Substance Use Servicces, 'BC Partners for Mental Health and Substance Use Information', 2018. [Online]. Available: http://www.bcmhsus.ca/health-info/mental-health-substance-use-resources/bc-partners. [Accessed: 28-Nov-2018].

[24] Pacifica Labs Inc, 'Pacifica App for Anxiety & Depression', 2018. [Online]. Available: https://www.thinkpacifica.com/. [Accessed: 22-Oct-

2018].
[25] An Roinn Slainte; Mannystrie O Poustie, ‘Service Framework for Mental Health and Wellbeing 2018-2021’, 2018.
[26] Organisation for Economic Co-operation and Development, ‘OECD Mental Health Performance Framework’, 2018.
[27] NHS, ‘The Community Mental Health Framework for Adults and Older Adults’, 2019.
[28] A. M. Kilbourne *et al.*, ‘Measuring and improving the quality of mental health care: a global perspective.’, *World Psychiatry*, vol. 17, no. 1, pp. 30–38, Feb. 2018.
[29] N. Sunderji, A. Ion, A. Ghavam-Rassoul, and A. Abate, ‘Evaluating the implementation of integrated mental health care: A systematic review to guide the development of quality measures’, *Psychiatr. Serv.*, vol. 68, no. 9, pp. 891–898, 2017.
[30] L. A. Clark, B. Cuthbert, R. Lewis-Fernández, W. E. Narrow, and G. M. Reed, ‘Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health’s Research Domain Criteria (RDoC)’, *Psychol. Sci. Public Interes.*, vol. 18, no. 2, pp. 72–145, 2017.
[31] M. Vrabel, ‘Preferred Reporting Items for Systematic Reviews and Meta-Analyses’, *Oncol. Nurs. Forum*, 2015.
[32] British Library, ‘About British Library EThOS - search and order theses online’, *EThOS e-theses online service*, 2017. .
[33] TIB, ‘Home - Technische Informationsbibliothek (TIB)’, 2019. [Online]. Available: https://www.tib.eu/en/. [Accessed: 28-May-2019].
[34] Google, ‘Google Scholar’, 2019. [Online]. Available: https://scholar.google.co.uk/. [Accessed: 28-May-2019].
[35] T. Segaran, *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. 2007.
[36] G. Mulgan, ‘Big Mind: How Collective Intelligence Can Change Our World - introduction’, in *Big Mind*, Princeton University Press, 2017.
[37] K. Salaheddin and B. Mason, ‘Identifying barriers to mental health help-seeking among young adults in the UK: a cross-sectional survey’, *Br. J. Gen. Pract.*, vol. 66, no. 651, pp. e686–e692, 2016.
[38] J. C. Glenn, ‘Collective intelligence systems and an application by The Millennium Project for the Egyptian Academy of Scientific Research and Technology’, *Technol. Forecast. Soc. Change*, vol. 97, no. February, pp. 7–14, 2015.
[39] E. K. Wakida *et al.*, ‘Barriers and facilitators to the integration of mental health services into primary health care: a systematic review protocol’, *Syst. Rev.*, vol. 6, no. 1, p. 171, Dec. 2017.
[40] NHS, ‘How to access mental health services - NHS’, 2018. [Online]. Available: https://www.nhs.uk/using-the-nhs/nhs-services/mental-health-services/how-to-access-mental-health-services/. [Accessed: 12-Oct-2018].
[41] World Health Organization, ‘Evidence-informed policy-making’, World Health Organization, 2013.
[42] Department of Health, ‘From segregation to inclusion: Commissioning guidance on day services for people with mental health problems’, 2006.
[43] R. Keynejad *et al.*, ‘Building the capacity of policy-makers and planners to strengthen mental health systems in low- and middle-income countries: a systematic review’, *BMC Health Serv. Res.*, vol. 16, no. 1, p. 601, Dec. 2016.
[44] Health Level 7 International, ‘HL7 Standards Product Brief - HL7 Cross-Paradigm Implementation Guide: Behavioral Health Data Exchange, Release 1 - US Realm’, *2017*. [Online]. Available: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=462. [Accessed: 12-Nov-2018].
[45] WHO (World Health Organization)., ‘Framework and Standards for Country Health Information Systems’, *Heal. Metrics Netw.*, 2012.
[46] D. D. Luxton, R. A. Mccann, N. E. Bush, M. C. Mishkind, and G. M. Reger, ‘mHealth for Mental Health: Integrating Smartphone Technology in Behavioral Healthcare’, *Prof. Psychol. Res. Pr.*, vol. 42, no. 6, pp. 505–512, 2011.
[47] D. E. Jake-Schoffman *et al.*, ‘Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps.’, *JMIR mHealth uHealth*, vol. 5, no. 12, p. e190, Dec. 2017.
[48] S. M. Schueller, J. J. Washburn, and M. Price, ‘Exploring mental health providers’ interest in using web and mobile-based tools in their practices’, *Internet Interv.*, vol. 4, pp. 145–151, May 2016.
[49] H. A. Pincus, C. M. Pechura, L. Elinson, and A. R. Pettit, ‘Depression in primary care: linking clinical and systems strategies’, *Gen. Hosp. Psychiatry*, vol. 23, no. 6, pp. 311–318, Nov. 2001.
[50] U. Sarkar *et al.*, ‘Usability of Commercially Available Mobile Applications for Diverse Patients’, *J Gen Intern Med*, vol. 31, no. 12, pp. 1417–1443, 2016.
[51] E. R. Surjaningrum, A. F. Jorm, H. Minas, and R. Kakuma, ‘Personal attributes and competencies required by community health workers for a role in integrated mental health care for perinatal depression: voices of primary health care stakeholders from Surabaya, Indonesia’, *Int. J. Ment. Health Syst.*, vol. 12, no. 1, p. 46, Dec. 2018.
[52] S. Siddiqi *et al.*, ‘Framework for assessing governance of the health system in developing countries: Gateway to good governance’, *Health Policy (New. York).*, vol. 90, pp. 13–25, 2009.
[53] National Collaborating Centre for Mental Health, *The NICE Guideline on the Treatment and Depression the Treatment and Management of Depression*. 2010.
[54] National Institute for Health and Care Excellence, ‘Depression in adults: recognition and management | Guidance and guidelines | NICE’, *clinical guideline*, 2018.
[55] Mental Health America, ‘Mental Health Facts, Stats, and Data’, *Mental Health America*, 2018.
[56] A. P. Association, *Diagnostic an Statistical Manual of Mental Disorders*. 2016.
[57] World Health Organisation, ‘ICD-10 International Statistical Classification of Diseases and Related Health Problems’, 2011.
[58] A. Beck, R. Steer, and G. Brown, ‘Beck Depression Inventory-II (BDI-II)’, *... Beck Depress. Invent.*, 1993.
[59] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, ‘The PHQ-9’, *J. Gen. Intern. Med.*, 2001.
[60] F. N. Jacka and M. Berk, ‘Depression, diet and exercise’,

The Medical journal of Australia. 2013.

[61] H. Tanaka *et al.*, 'An examination of sleep health, lifestyle and mental health in junor high school students', in *Psychiatry and Clinical Neurosciences*, 2002.

[62] D. Freeman *et al.*, 'Articles The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis', p. 749, 2017.

[63] M. Mukaka, 'A guide to appropriate use of correlation coefficient in medical research', *Malawi Med. J.*, pp. 69–71, 2012.

[64] A. Edwards-Jones, 'Qualitative data analysis with NVIVO', *J. Educ. Teach.*, 2014.

APPENDIX I

Table 3: A Recapitulated Survey Questions with Dimensions of Evaluation for FLMMHS

No	Expert-judge questions	Experts' knowledge & expertise	Evaluation parameters pro questions correspondence			
			Efficacy	Effectiveness	Flexibility	Simplicity
1	Expert's level of expertise of Web/Software/System development	√				
2	Expert's experience of clinical & MH system development	√				
3	Expert's knowledge of MH disorders, characteristics & symptoms	√				
4	Expert's perception of utilising the framework for developing MH management system	-	-	-	-	-
5	Expert's knowledge of MH Standards and its availability	√				
6	Technology acceptance & change within the framework		√	√		
7	Acceptability of the framework by system developers for MH system development		√	√		
8	Credibility of Framework for MH system development		√	√		
9	Framework acceptability (Question 'g' repeated)		√	√		
10	User's privacy consideration for MH system development		√			
11	Provisions of guidelines for developers of MH system		√			
12	Skill & knowledge enhancement for developers of MH system		√	√	√	√
13	Motivation and support for developer in developing MH management system			√	√	
14	Resources consideration for development of MH system		√			
15	Guide to MH Policies & Strategies for developing MH system		√	√	√	√
16	Easy of comprehending framework for MH system development					√
17	Expert's perceived usefulness of the framework for MH system development		√			
18	Effectiveness/efficacy of the framework for developing MH Diagnosis		√	√		
19	Effectiveness/efficacy of the framework for developing MH Prevention		√	√		
20	Effectiveness/efficacy of the framework for developing MH Alleviation		√	√		
21	Effectiveness of the framework for developing MH system holistically			√		
22	Expert's ideal characteristics of framework for MH system development		√	√	√	√

Table 4: A Matrix Representation of Themes Association via Coded Data

	Expertise	MH knowledge	Expert perception	Flexibility	Simplicity	Efficacy	Effectiveness	Other factors
Expertise	-	-	-	-	-	-	-	-
MH knowledge	2	-	-	-	-	-	-	-
Expert perception	16	5	-	-	-	-	-	-
Flexibility	4	0	10	-	-	-	-	-
Simplicity	4	0	11	10	-	-	-	-
Efficacy	7	3	8	5	6	-	-	-
Effectiveness	20	4	17	7	12	18	-	-
Other factors	6	0	6	5	10	8	15	-

Telehealth-based Intrapartum Monitoring: Impact of Clinical and Technical Factors on Remote Decision Making

Hiwale Sujitkumar, Krishnan Navaneetha, Addepalli Jeevan Ram, and Ulman Shrutin
Philips Research India,
Philips Innovation Campus, Manyata Tech-Park,
Bengaluru, India-560045
e-mail: {sujit.hiwale, navaneetha.krishnan, jeevan.ram.addepalli, shrutin.ulman}@philips.com

***Abstract* - Digital Labour and Delivery Solution (DLDS) is a mHealth-based solution for structured and instant communication during intrapartum care. The primary objectives of this study were – (1) to evaluate feasibility of the DLDS for information exchange among healthcare professionals for remote intrapartum monitoring and decision making; and (2) to study impact of various clinical and technical factors on decision agreements between the doctors. The inclusion criteria for the study were, a live-singleton pregnancy with cervical dilatation ≥ 4 cm but < 8 cm at the time of admission, and presenting without any complication necessitating any immediate intervention. The feasibility of the DLDS was evaluated by comparing the decisions taken by a remote doctor using the DLDS to that of decisions taken by a doctor in a labour room. Impact of clinical parameters (mother's age, parity, anemia and presence of intrapartum complications) on decision agreement between the doctors was studied by comparing agreements in different subcategories of these parameters. Similarly, the total number of observation records for a subject were also studied to find their impact on decision making. The overall agreement between the two doctors for 110 cases (220 independent decision points) was 0.764 using unweighted Cohen's kappa and 0.723 using weighted Cohen's kappa statistic. The doctors had comparable agreements in all the sub-categories of the clinical parameters, indicating minimal impact of clinical parameters on decision agreement between the doctors. A significant improvement was observed in the agreement as the total number of assessments available during the course of labour increased. The substantial agreement between the two doctors for intrapartum decision making demonstrates the feasibility of the DLDS for remote intrapartum monitoring and decision making. This also indicates that DLDS was able to convey the appropriate information to the remote doctor in the different sub-categories of the clinical parameters. The study recommends further investigation of DLDS for a general purpose remote intrapartum monitoring.**

Keywords- feasibility study; inter-observer variability; intrapartum; mHealth; obstetrics; partograph; telemedicine.

I. Introduction

In the last few decades, obstetrics care has evolved significantly from delivery at home, to delivery at a specialty center under the supervision of a trained medical or paramedical team. A short time interval between onset of complications and time to intervene makes intense monitoring and prompt decision making very important during the intrapartum phase. Structured and instant communication can play a very important role to make intrapartum care effective and safe. These are areas where Information and Communications Technology (ICT) can play a major role. In recent times, mobile devices (smartphones and tablets) have emerged as one of the most important enablers of ICT in healthcare. Considering the need gaps in intrapartum communication and the potential of mobile devices for telehealth, we have designed a Digital Labour and Delivery Solution (DLDS). DLDS is a tablet-based solution designed for systematic information gathering and sharing during intrapartum monitoring [1].

Intrapartum monitoring and decision making is a very complex procedure. Availability of structured and instant data is just one aspect of this process. Many other factors such as clinical history and complication of a patient, and technical factors such as how frequently a patient is assessed during course of labour and frequency of vital parameters' recording can also have a significant bearing on decision making. This could be a reason that obstetricians have shown to have a poor agreement during intrapartum decision making [2]. This is especially important during remote monitoring. Therefore, it is equally important to study impact of these parameters on remote intrapartum monitoring.

The primary objectives of this study were to evaluate the feasibility of the DLDS in information exchange among health care professionals for remote intrapartum monitoring and decision making, and to study impact of various clinical and technical factors on decision agreements between the doctors. The feasibility of DLDS in information exchange for remote intrapartum monitoring and decision making is covered in detail elsewhere [1]. This paper mostly focuses on study of impact of various clinical and technical factors on decision agreements between the doctors by comparing agreements in different subcategories of these parameters.

The rest of the paper is organized as follows, Section II summarizes the literature review of the communication tools used during intrapartum phase; Section III covers details of the study protocol and statistical methodology. The study results are summarized in Section IV. Section V provides commentary on overall results and their possible implications for clinical practice. Section VI concludes with the most important findings and future work directions. The acknowledgement section closes the article.

II. Review of Literature

As the complications during intrapartum phase are responsible for almost 42% of maternal mortality, 32% of still births, and 23% of neonatal deaths [3], prompt monitoring and instant communication is very important during intrapartum phase for an effective decision making. This also highlights need of an effective collaboration between doctors and midwives with well-defined roles and responsibilities. However, contrary to this requirement, it has been observed that intrapartum care suffers from a poor teamwork. A study conducted in the USA has observed that less than 50% of doctors and less than 37% of nurses in labour room rated their teamwork as adequate [4], indicating magnitude of poor teamwork in the intrapartum care.

It is well recognized that a poor teamwork or interprofessional collaboration is one of the important reasons for adverse clinical outcomes and poor delivery of health services and patient care [5][6][7]. Among various issues, which have an adverse impact on teamwork, poor communication patterns have been identified as one of the most important issues. This is evident by a fact that issues in communication have been identified as the root cause in 72% of total cases related to infant deaths and injuries during delivery [8]. Poor communication is usually a result of a poor transmission or exchange of information. Paper-based methods and telephonic communication are two conventional methods used by intrapartum monitoring team for information exchange.

Intrapartum monitoring team is usually composed of midwives/nurses, junior doctors and senior obstetricians working in a labour room. The role and responsibilities of each of them are usually defined and bounded by a hierarchical system, although overlaps do exist. Midwives are primarily assigned for intrapartum monitoring of vital parameters and for basic interventions, while doctors are responsible for clinical decision making and advanced surgical interventions. Midwives are stationed in the labour room all the time during their shifts, while doctors may or may not be constantly present in the labour room due to various other responsibilities assigned to them. In the latter case, midwives regularly update the doctors to get management guidance.

Two important conventional methods of communication during intrapartum monitoring are paper-based method and telephonic communication, a brief review of these methods is presented as follows.

A. Paper-based method

In a conventional paper-based workflow, all patient related information, history and management details are recorded on paper sheets. This time-honoured method is simple to follow, inexpensive and has wide acceptance among healthcare workers around the globe. Although, easy to use and simple, there are few significant disadvantages with this method – (1) the data recorded by this method is often unstructured and non-standardized, this makes it difficult to standardize the care practices; (2) it is not possible to share the recorded information with multiple people simultaneously; (3) this method has been shown to be prone to manual errors, which could be a cause of legal litigations [9]; and (4) over a period of time, storage and retrieval of paper records become quite challenging, which makes it very difficult to use recorded information for any analytics or predictions. The most significant disadvantage from intrapartum monitoring perspective, is the inability to remotely share the critical patient information in real time for collaborative decision making. To make the paper-based method more structured and effective various different approaches have been tried in the past; one important initiative in this regard is the use of partograph for intrapartum monitoring.

The partograph (a.k.a. partogram) is a simple, inexpensive tool, which provides health professionals a pictorial overview of the labour for early identification and diagnosis of the pathological labour. It was first proposed by Emanuel Friedman in 1954, as a cervicograph [10]. Subsequent large multi-centric prospective studies conducted by the World Health Organization (WHO) concluded that partograph was able to clearly differentiate between normal and abnormal progress in labour; the WHO recommended its universal application in all the labour rooms [11]. Following this, a number of studies were conducted across the globe to determine the effect of partograph use on perinatal morbidity and mortality; however, there is still no consensus in the literature on the effectiveness of the partograph [12].

Despite its proven effectiveness in labour monitoring, partograph has not been utilized optimally in many settings across the world [12][13]. It has been observed, that the rate of partograph utilization during intrapartum monitoring varies significantly in different setups, being as high as 97.8% in Niger study to as low as just 1.4% in a study done in Bangladesh [14]. The suboptimal utilization of partograph does not stop here, it has been observed that in most of the cases where partograph is used, the clinical parameters are either not recorded or recorded less frequently than prescribed. It is well-documented that, when information on the partograph is incomplete, misinterpretation is more likely and it may lead to delayed diagnosis, inappropriate or no action, and consequent development of serious complications [15].

Despite its effectiveness, sub-optimal utilization and poor recording of partograph parameters during labour is a matter of great concern for quality intrapartum care throughout the world. To address this issue it is important to understand the barriers for partograph use. Based on published literature, barriers to partograph use can be grouped in three broad categories – (1) implementation related barriers; (2) caregivers related barriers; and (3) clinical workflow-related barriers. Out of these barriers, the caregiver-related issues, mainly – (1) insufficient knowledge on how to use partograph [16][17]; (2) work overload due to shortage of staff [17]; and (3) time consuming nature of partograph plotting were found to be the most important barrier for effective partograph use [18].

To reduce barriers in partograph use and increase its utilization different approaches have been tried across the globe such as PartoPen, mlabour, DAKSH and E-Partograph. These digital initiatives have rejuvenated interest in partograph and have shown a good user acceptance in feasibility studies [19][20].

B. Telephonic communication

Telephonic conversations among healthcare workers are now primary mode of remote consultation and monitoring in intrapartum care. This method is very simple, universally available and offers advantages of real time communication in a cost effective manner. Availability of mobile phones has further enhanced the reach of this method. However, there are few significant disadvantages with this method – (1) effectiveness of communication is limited by education and experience of involved parties; (2) limited access to information, leading to misunderstanding or underestimation of complications [21][22][23]; (3) it is not possible to record information for analysis, or audit or for medico-legal purpose. Moreover, telephonic triaging is considered as the most complex and vulnerable part of the out-of-hospital care process and has also been shown to be associated with patient dissatisfaction [21][24].

To make telephonic conversions more structured and effective various approaches have been tried. One important initiative in this regard is use of **S**ituation–**B**ackground–**A**ssessment–**R**ecommendation (SBAR) technique. SBAR allows the medical team to communicate with each other in a standard way by using a structured method for the transfer of vital information. SBAR technique was first used in military communication, followed by aviation industry for effective communication. This has been later adopted by many health care settings as a communication tool.

SBAR has been shown to be an excellent tool for information sharing and has found application in many sub specialties of medical care. It has been observed in a number of studies that SBAR technique has not only improved communication between healthcare professionals but also has improved the overall quality of care [25][26]. However, a few studies have also reported no or little improvement in overall communication or quality of care [22].

SBAR is shown to be an effective communication tool, but it is not free of disadvantages. It has been observed that – (1) SBAR concept is difficult to learn and practice, therefore it requires extensive education and training along with frequent follow-ups for effective implementation [27]; (2) SBAR approach requires changes in nursing practice [27]; (3) it a is time consuming technique [26].

Despite its limitations, SBAR is a very effective tool for structured communication. To enhance it further there are now attempts to design it in the electronic form. The initial studies have indicated encouraging results in this regard [28].

C. mHealth applications in the intrapartum space

The use of ICT in healthcare has grown exponentially in the last two decades; however, it still lacks way behind in comparison with other sectors like finance, retail, transportation etc. This is usually because of high cost, limited evidences of benefit, non-availability or complexity associated with the new technologies, and lack of end-user centric solutions [29]. Fortunately, mHealth has overcome many of these classical barriers with their ubiquitous presence and high acceptance among end users, even in the remote parts of the world [30]. Availability of cheap and portable computing platform in the form of mobile devices has further accelerated the reach of ICT in healthcare. While mHealth solutions have overcome several barriers of conventional communication tools, other barriers like intermittent power and connectivity, low literacy levels, low levels of technical training, and maintenance and scalability costs are yet to be fully solved.

In the intrapartum space, applications like PartoPen [19], and DAKSH [20] are introduced to make communication structured and more effective. These application have mostly focused on digitization of labour records, partograph and usability aspects. However, none of them have systematically studied their applicability for remote intrapartum monitoring and decision making. Moreover, they have also not studied impact for various clinical and technical factors on decision agreement during remote intrapartum care.

D. A summary of literature review and indentification of communication needs for the intrpartum care

To improve communication during intrapartum care, many tools and techniques such as partograph, SBAR, and digital partograph have been introduced. These are shown to be effective but underutilized due to time constrains and stiff learning curves. The literature indicates that the existing methods of communication have limitations when it comes to clear and real-time information exchange during intrapartum and have been shown to be either inadequate or cumbersome for this purpose [9][21]. Moreover, none of these techniques provide an integrated solution for intrapartum monitoring and decisions, making them of limited use for a hospital setup.

An emergence of affordable smartphones and increase cellular connectivity and data transfer facility with 3G, 4G and upcoming 5G networks are likely to provide a significant boost to the use of the mobile platform for providing healthcare services. On other hand, lack of regulations or stringent regulations and data security issues are likely to be major hurdles for a wide spread use of this technology platform in healthcare. Going forward, one very important driver for this platform will be its high penetration in the developing countries, which have a high burden of diseases, limited human resources and funds to provide adequate healthcare. These countries are likely to be the primary consumers of mHealth solutions.

A digital solution, which incorporates real time data sharing capacities with standardized communication (partograph and SBAR) and decision protocols along with seamless connectivity has potential to address information sharing need gaps and is likely to be a way for future of intrapartum communication. Taking positive trends from digital health journey so far, many mHealth models for information sharing and data processing are already in various stages of development. DLDS is one such mHealth solution for intrapartum care. The DLDS is designed to serve as an integrated solution for remote intrapartum monitoring and decision making.

The primary objectives of this study were - (1) to evaluate the feasibility of the DLDS in information exchange among health care professionals for remote intrapartum monitoring and decision making; and (2) to study impact of various clinical and technical factors on decision agreements between the doctors. We had a primary hypothesis that a remote doctor can be equally adept at decision making if he/she is provided with all the necessary information. The effectiveness of the DLDS was evaluated by comparing the decisions taken by a remote doctor (outside of a labour room) using intrapartum information provided by the DLDS to that of decisions taken by a doctor in a labour room (in-charge doctor). Impact of clinical and technical factors was studied by comparing decision agreements in different subcategories of these parameters.

III. Material and Methods

This section provides details about the study protocol and statistical methodology.

A. *Study design*

This observational study was conducted in a medical college hospital in Mysuru (Mysore), India in 2016. Inclusion criteria for the study were a live-singleton pregnancy with cervical dilatation $\geq$ 4 cm but < 8 cm at the time of admission to a labour room. All the cases with planned caesarean section or cases with complication(s) or indication(s), which required immediate intervention or where a trial of labour was contraindicated were excluded. The study was conducted in accordance with local regulations after approval of an institutional review board. Subjects were enrolled only after obtaining informed consent in writing.

B. *Study protocol*

All the enrolled subjects were managed as per the established clinical workflows and protocols of the hospital. The subjects were regularly assessed by an in-charge doctor (doctor involved in an active management of a subject). After each assessment, the in-charge doctor took one management decision from four possible options - (1) “Wait and watch”, i.e., to continue the expectant management without any active intervention; (2) “Accelerate the labour”, i.e., accelerate the labour process either by means of artificial rupture of membranes or by medication; (3) go for “Assisted vaginal delivery”, i.e., use of forceps or vacuum extraction method for delivery; and (4) go for “Caesarean section”.

All the subjects and newborns were monitored up to 24 hours after delivery for any adverse outcomes. Outcomes monitored included obstructed labour, uterine rupture, post-partum haemorrhage, stillbirth, early neonatal mortality, Apgar score at five minutes, and newborn’s admission to a Neonatal Intensive Care Unit (NICU).

For each subject, complete clinical history, examination, investigation details and management decision for each assessment were entered in the DLDS. To prevent any influence of the DLDS on clinical workflow and patient management an additional nurse (not actively involved with patient management) was appointed for data entry in the DLDS. The study workflow is illustrated in Fig. 1.

C. *DLDS application*

DLDS has been developed as a monitoring and communication solution for labour, delivery and immediate post-partum care. DLDS is a tablet-based solution built on an Android platform and allows secured sharing of information over a Wi-Fi network. Its intuitive design and user interface allows systematic and easy entry of the past and present history, examination and investigation details of the patient with an option to customize entry fields. It also provides an advanced visualization for various clinical trends and partograph. For guidance (e.g., normal ranges of clinical parameters, recommended frequency of intrapartum parameters assessment etc.) and alerts (e.g., alert and action line for partograph) the well-established guidelines and protocols given by the reputed organizations such as the WHO were incorporated in the DLDS. The DLDS can be used as a stand-alone delivery solution or could be integrated with maternal telehealth platforms such as Mobile Obstetrics Monitoring [31]. For the study, two DLDS tablets were used; the one in the labour room was designed to anonymize and securely transmit information to the other tablet over a wireless network connection.

D. *Workflow of the remote doctor*

A doctor who was not involved in the management of any of the study subjects was assigned as a ‘remote doctor’. To ensure that there is no discrepancy in decision making due to skill and knowledge differences, doctors with a similar profile as the in-charge doctor was selected as a remote doctor. The remote doctor was asked to use the second DLDS application to review case records (without management decision information) and enter one of the four management decisions in the DLDS. For each subject, the remote doctor reviewed the case records at the two instances, first at the time of admission and the last record before any active intervention or delivery (refer to Fig. 1).

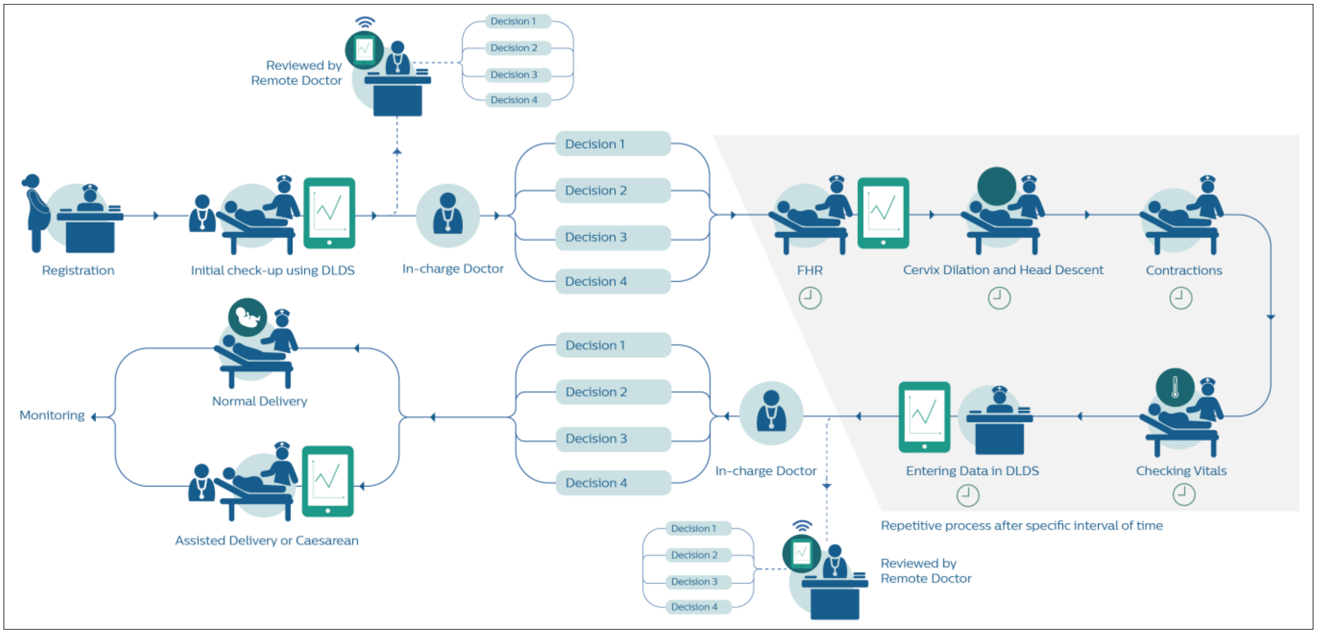


Figure 1. Workflow of the study. Decision 1 to 4 denotes one of the four possible management decisions taken by the in-charge doctor. The remote doctor reviewed the subject records using the Digital Labour and Delivery Solution (DLDS) at two instances and took one of the four possible management decisions.

E. Potenial impact of clinical and technical factors on the decision agreements between the doctors

Decision making during intrapartum phase is a complex procedure. Doctors' consider many clinical parameters during decision making such as the age of the patient, obstetric history, presence of complications and so on. To study the effectiveness of the DLDS in conveying this information, the decision agreements between the in-charge and the remote doctor in various sub-categories of clinical parameters were studied. Other than pre-mentioned adverse outcomes such as obstructed labour, uterine rupture, post-partum haemorrhage, stillbirth, early neonatal mortality, Apgar score, and newborn's admission to a NICU, other clinical parameters such as, mother's age, parity, anemia (Haemoglobin <= 11 gm/dL) and presence of intrapartum complications (leaking/bleeding per vaginam) were also studied to find out their possible impact on decision agreement between the doctors. Impact of clinical factors was studied by comparing decision agreements in different subcategories of these parameters.

Apart from the clinical parameters, technical factors such as a number of times a particular patient is assessed/observed with recording of vital parameters also has bearing on intrapartum decision making. This is especially important for remote decision making. To study the impact of a number of assessments on decision making, the cases were divided in three sub-categories based on the total number of observation records available and decision agreements between the doctors were studied in each of those sub-categories.

F. Statistical analysis methodology

The decisions taken by the doctors for each case were extracted from the two DLDS applications. The agreement between the in-charge doctor and the remote doctor on the four types of management decisions was assessed using the Cohen's kappa statistics. However, as different types of management decisions have different implications in clinical practice, it is important to study not only overall agreement between the two doctors but also an extent of disagreement for individual decisions. This is important as some decisions are closer to each other when compared to other decisions (e.g., a decision to go for "Caesarean section" is much closer to a decision to go for "Assisted vaginal delivery" in comparison to a decision of "Wait and watch").

As Kappa analysis does not account for the difference in decision types, weighted Kappa analysis was used for this purpose. The weights used to grade the differences in decisions are presented in Table I.

TABLE I. WEIGHT MATRIX FOR DECISION GRADING

Decision taken by in-charge doctor	**Decision taken by the remote doctors**			
	Wait and watch	***Accelera-te labour***	***Assisted vaginal delivery***	***Caesarea-n section***
Wait and watch	0	1	2	3
Accelerate the labour	1	0	1	2
Assisted vaginal delivery	2	1	0	1
Caesarean section	3	2	1	0

The agreement scale proposed by Landis and Koch was used to grade and compare the agreements between the doctors in various sub-categories of clinical and technical parameter [32]. All statistical analyses were done using Microsoft Office Excel-2016 and R (version 3.5.2).

IV. RESULTS

The study results are summarized in this section.

A. *Demographic characteristics of the study population*

In total, 110 subjects were enrolled for the study. The mean maternal age was 24.21 ± 2.69 year, with a mean body mass index of 24.48 ± 2.08 kg/m^2. The nulliparous women constituted 30.43% of the study population. Gestational age was in the range of 37 to 41.6 weeks (median = 39.55 weeks). The mean birth weight of the neonates was 3037.98 ± 345.25 g, with a range of 2320 g to 4040 g.

B. *Intrapartum monitoring and labour outcomes*

Throughout labour, all the subjects were monitored using the conventional workflows and protocols of the hospital. None of the cases had any significant antenatal complication. The average duration of labour was 7 hours 3 minutes (± 63 minutes). On an average, each subject was assessed 15.63 (± 0.518) times during labour, which comes out to be one assessment per 28 minutes. During each assessment, vital parameters, examination details and management decision for a subject were entered in the DLDS application.

Five cases were delivered by caesarean section. Two cases were delivered by forceps extraction method. The rest of the cases were delivered vaginally. Four cases had history of leaking/bleeding per vaginam. None of the other cases had any adverse intrapartum or immediate postpartum outcome. All the neonates had Apgar score of eight or more at five minutes and none of them required admission to a NICU.

C. *Agreement for the management decisions*

The remote doctor was asked to review 220 records (two records per case) using the DLDS. The confusion matrix of the four management decisions taken by both the doctors is summarized in Table II.

TABLE II. DECISION AGREEMENT BETWEEN THE DOCTORS

Decision taken by in-charge doctor	**Decision taken by the remote doctor (using *DLDS)***				**Total**
	Wait and watch	***Accelera-te labour***	***Assisted vaginal delivery***	***Caesarea-n section***	
Wait and watch	103	10	0	0	113
Accelerate the labour	11	88	0	1	100
Assisted vaginal delivery	1	0	1	0	2
Caesarean section	2	2	0	1	5
Total	117	100	1	2	220

It was observed that for the "Wait and watch" decision the remote doctor was in a perfect agreement with the decisions of the in-charge doctor in 91.15% of total records; for "Accelerate the labour" this agreement was 88%. Agreements for "Assisted vaginal delivery" and "Caesarean section" were 50% and 20%, respectively. Nevertheless, as these two categorizes had very few samples (not even 30 samples, a general requirement for statistical analysis) it is difficult to comment on their statistical significance.

The overall agreement between the two doctors for all the decisions combined was 0.764 using unweighted Cohen's kappa statistics. The weighted Cohen's kappa between the two doctors was 0.723.

D. *Impact of clincial parameters on the agreement between the doctors*

In this study following clinical parameters, mother's age parity, anemia (Haemoglobin <= 11 gm/dL) and presence of intrapartum complications (leaking/bleeding per vaginam) were studied to find out their possible impact on decision agreement between the doctors. All of these parameters were divided in two sub-categories to compare decision agreements. In study population, none of the subject was over 35 years of age and only a few were above 30 years; therefore, 25 years was used as a threshold to divide subjects in two sub-categories.

The unweighted Cohen's kappa statistics between the in-charge and remote doctor for various sub-categories of clinical parameters are summarized in Table III. It was observed that the in-charge and remote doctor had comparable agreements for all the studied clinical parameters with substantial agreements in all the sub-categories.

TABLE III. IMPACT OF CLINCIAL PARAMETERS ON AGGREMENT BETWEEN THE DOCTORS

Clinical parameters	**Sub category**	**No. of decision points**	**Cohen's kappa**
Age (year)	Age <= 25	132	0.775
	Age > 25	88	0.745
Parity	Nulliparous	80	0.747
	Multiparous	140	0.774
Anaemia (Hb <= 11 gm/dL)	Yes	22	0.748
	No	198	0.759
Complications (leaking/bleeding PV)	Yes	8	0.800
	No	212	0.755

Hb = Haemoglobin, PV = Per vaginam.

As none of the cases had obstructed labour, uterine rupture, post-partum haemorrhage, stillbirth, early neonatal mortality, low Apgar score, and newborn's admission to a NICU we could not study the possible impact of these complications on decision agreement between the doctors.

E. Impact of total number of assessments records on the agreements between the doctors

For remote monitoring it is important that a doctor has frequent data from the labour room. To study impact of a number of assessment records on the decision agreements, the cases were divided in three subgroups based on the total number of assessments for each case. The unweighted Cohen's kappa statistics between two doctors for the three sub-categories are summarized in Table IV.

TABLE IV. IMPACT OF TOTAL ASSESSMENTS RECORDS ON THE AGREEMENT BETWEEN THE DOCTORS

Parameter	**Sub category**	**No. of decision points**	**Cohen's kappa**
Number of assessment (records)	<= 10 assessments	24	0.576
	11 - 20 assessments	160	0.779
	>= 21 assessments	36	0.837

As per the Landis and Koch scale, a significant improvement was observed in the agreement between the in-charge and remote doctor as the total number of available assessments records increased. For the group with less than or equal to 10 assessments the doctors had just a moderate agreement among themselves, whereas for the group with more than 20 assessments the agreement was almost perfect.

V. DISCUSSION

One very crucial part of communication is transmission or exchange of information in a structured way for effective decision making. Unfortunately, the existing modes of intrapartum communications are shown to be insufficient for this purpose [21][22] making them a less reliable medium for information exchange [9][23]. The main objective of this study was to evaluate feasibility of the DLDS application for remote intrapartum monitoring and decision making. This was done by comparing decisions taken by a remote doctor using the DLDS to that of the in-charge doctor. In this regard, a substantial agreement was observed between the two doctors for intrapartum decision making. This demonstrates the feasibility of the DLDS for remote intrapartum monitoring and decision making.

It was observed that the agreement between the doctors for non-operative mode of deliveries was significantly higher than for operative deliveries. This finding is in line with the published literature, where complete agreement for caesarean section decision has been observed to be about 65% [2]. Nevertheless, the lower agreement for operative deliveries (in particular more decisions of "Assisted vaginal deliveries" and "Caesarean section" by the in-charge doctor) needs further investigation. This could be due to the remote doctor missing some crucial information or the doctor in-charge getting negatively influenced by real-life factors such as stress of other emergencies to attend, lack of sleep, or pressure from the healthcare workers or patients. However, as only seven cases were delivered by a non-vaginal route, it is difficult to generalize findings of this study to mode of deliveries other than vaginal.

A number of clinical and technical factors can have a major impact of intrapartum decision making. To study effectiveness of the DLDS in conveying this information, the decision agreement between the doctors in the various sub-categories of the clinical parameters were assessed. It was observed that the in-charge and remote doctor had comparable agreements in all the sub-categories of the clinical parameters, indicating minimal impact of clinical parameters on decision agreement between the doctors. This also indicates that DLDS was able to convey the appropriate information to the remote doctor in all the sub-categories of the clinical parameters. For total number of assessment, a significant improvement was observed in the agreement between the in-charge and remote doctor as the total number of assessments increased. This clearly indicates advantage of having more data points in decision making and also makes a strong case for having frequent and automated monitoring of vital parameters during labour.

Small sample size from a single center and recruitment of just one doctor in the labour room and one for remote assessment are two important limitations of our study. However, as this was a feasibility study we first wanted to test and verify our concept before conducting a large study with multiple doctors. Despite having a small sample size, we compared 220 independent decisions points between the two doctors. Furthermore, as none of the cases in our study had any adverse outcome, it was not possible to assess adequacy and quality of information provided by the DLDS to the remote doctor in such situations. Nevertheless, it was observed that the remote doctor could use the DLDS application for decision making for all the sub-categories of the clinical parameters.

On the study design, the use of an additional nurse for data entry is likely to have contributed to better and more comprehensive data gathering, which may not have been possible in conventional workflows. However, having a complete and accurate data entry is prerequisite for any digital solution and it is bound to have some change in the existing workflow. It also brings the advantage of enhanced patient safety by improving the communication, comprehensiveness, and organization of patient notes [33]. Moreover, it has been also indicated that introduction of digital records are likely to reduce risk and liability for obstetric providers, especially in the intrapartum care [34]. This study has also demonstrated that more data led to a better decision agreement, which further supports advantage of complete and accurate data entry in decision making.

VI. CONCLUSION

The strength of this study lies in being one of the first studies where the feasibility of a telehealth solution for remote intrapartum monitoring and decision making has been studied systematically. The finding of this study could serve as an important input for further research in this area. In the future, we would like to extend this work on a larger sample size with recruitment of more remote and in-charge doctors. Moreover, we would also like to conduct a dedicated usability study to understand and improve user interactions with the DLDS.

To conclude, our study has demonstrated a substantial agreement in the intrapartum decisions taken by a remote doctor using the DLDS and decisions taken by a doctor in a labour room. The study has also demonstrated that the in-charge and remote doctor had comparable agreements in all the sub-categories of the clinical parameters. This indicates that DLDS was able to covey the appropriate information to the remote doctor in the different sub-categories of the clinical parameters. The study has also clearly indicated advantage of more data points in decision making. This supports the hypothesis that it is possible to remotely monitor intrapartum labour progress and take appropriate decisions if a remote doctor is provided with all necessary information. It further supports use of telehealth solutions such as DLDS for remote intrapartum monitoring. Considering limited resources and shortage of trained healthcare workers in the developing countries, we believe that there is a huge need for intrapartum telehealth solutions in such countries. The study recommends further investigation of DLDS for a general purpose remote intrapartum monitoring.

ACKNOWLEDGMENT

We duly acknowledge help of the staff of JSS, Medical college, Mysuru, India for this study.

REFERENCES

[1] S. Hiwale, N. Krishnan, J. R. Addepalli, and S. Ulman, “mHealth Solution for Remote Intrapartum Monitoring: A Feasibility Study,” in *eTELEMED 2019 : The Eleventh International Conference on eHealth, Telemedicine, and Social Medicine*, Athens, Greece, 2019, pp. 118–122.

[2] S. A. Pillai *et al.*, “Decisions to Perform Emergency Caesarean Sections at a University Hospital: Do obstetricians agree?,” *Sultan Qaboos Univ. Med. J.*, vol. 16, no. 1, pp. e42-46, Feb. 2016.

[3] J. E. Lawn *et al.*, “Two million intrapartum-related stillbirths and neonatal deaths: where, why, and what can be done?,” *Int. J. Gynaecol. Obstet. Off. Organ Int. Fed. Gynaecol. Obstet.*, vol. 107 Suppl 1, pp. S5-18, S19, Oct. 2009.

[4] J.-M. Guise and S. Segel, “Teamwork in obstetric critical care,” *Best Pract. Res. Clin. Obstet. Gynaecol.*, vol. 22, no. 5, pp. 937–951, 2008.

[5] L. Lingard *et al.*, “Communication failures in the operating room: an observational classification of recurrent types and effects,” *Qual. Saf. Health Care*, vol. 13, no. 5, pp. 330–334, 2004.

[6] S. Kvarnström, “Difficulties in collaboration: A critical incident study of interprofessional healthcare teamwork,” *J. Interprof. Care*, vol. 22, no. 2, pp. 191–203, 2008.

[7] M. Zwarenstein, J. Goldman, and S. Reeves, “Interprofessional collaboration: effects of practice-based interventions on professional practice and healthcare outcomes,” *Cochrane Database Syst. Rev.*, no. 3, p. CD000072, 2009.

[8] J. Commission, “Sentinel Event Alert, Issue# 30, Preventing Infant Death and Injury During Delivery,” *Wash. DC Jt. Comm.*, 2004.

[9] T. Schabetsberger *et al.*, “From a paper-based transmission of discharge summaries to electronic communication in health care regions,” *Int. J. Med. Inf.*, vol. 75, no. 3–4, pp. 209–215, Mar. 2006.

[10] E. Friedman, “The graphic analysis of labor,” *Am. J. Obstet. Gynecol.*, vol. 68, no. 6, pp. 1568–1575, Dec. 1954.

[11] “World Health Organization partograph in management of labour. World Health Organization Maternal Health and Safe Motherhood Programme,” *Lancet*, vol. 343, no. 8910, pp. 1399–1404, Jun. 1994.

[12] T. Lavender, A. Hart, and R. M. D. Smyth, “Effect of partogram use on outcomes for women in spontaneous labour at term,” *Cochrane Database Syst. Rev.*, vol. 7, p. CD005461, 2013.

[13] E. Ollerhead and D. Osrin, “Barriers to and incentives for achieving partograph use in obstetric practice in low- and middle-income countries: a systematic review,” *BMC Pregnancy Childbirth*, vol. 14, no. 1, p. 281, Aug. 2014.

[14] E. Landry, C. Pett, R. Fiorentino, J. Ruminjo, and C. Mattison, “Assessing the quality of record keeping for cesarean deliveries: results from a multicenter retrospective record review in five low-income countries,” *BMC Pregnancy Childbirth*, vol. 14, no. 1, p. 139, Apr. 2014.

[15] F. Ngabo *et al.*, “Quality of Care for Prevention and Management of Common Maternal and Newborn Complications,” USAID, Mar. 2012.

[16] A. C. Umezulike, H. E. Onah, and J. M. Okaro, “Use of the partograph among medical personnel in Enugu, Nigeria,” *Int. J. Gynaecol. Obstet. Off. Organ Int. Fed. Gynaecol. Obstet.*, vol. 65, no. 2, pp. 203–205, May 1999.

[17] B. Kushwah, A. P. Singh, S. Singh, B. Kushwah, S. Campus, and R. Huzur, “The partograph: an essential yet underutilized tool,” *J Evol Med Dent Sci*, vol. 2, no. 24, pp. 4373–4379, 2013.

[18] M. M. Opiah, A. B. Ofi, E. J. Essien, and E. Monjok, “Knowledge and utilization of the partograph among midwives in the Niger Delta Region of Nigeria,” *Afr. J. Reprod. Health*, vol. 16, no. 1, pp. 125–132, Mar. 2012.

[19] H. Underwood, J. Ong’ech, G. Omoni, S. Wakasiaka, S. R. Sterling, and J. K. Bennett, “Improving partograph training and use in Kenya using the partopen digital pen system,” in *Biomedical Engineering Systems and Technologies*, Springer, 2014, pp. 407–422.

[20] M. Singh, “DAKSH: Digital Partograph and Intrapartum Monitoring Mobile Application,” p. 7, 2019.

[21] P. Giesen *et al.*, “Safety of telephone triage in general practitioner cooperatives: do triage nurses correctly estimate urgency?,” *Qual. Saf. Health Care*, vol. 16, no. 3, pp. 181–184, 2007.

[22] E. Joffe, J. P. Turley, K. O. Hwang, T. R. Johnson, C. W. Johnson, and E. V. Bernstam, “Evaluation of a problem-specific SBAR tool to improve after-hours nurse-physician phone communication: a randomized trial,” *Jt. Comm. J. Qual. Patient Saf.*, vol. 39, no. 11, pp. 495–501, 2013.

[23] K. Nagpal *et al.*, “Information transfer and communication in surgery: a systematic review,” *Ann. Surg.*, vol. 252, no. 2, pp. 225–239, 2010.

[24] R. Leibowitz, S. Day, and D. Dunt, “A systematic review of the effect of different models of after-hours primary

medical care services on clinical outcome, medical workload, and patient and GP satisfaction," *Fam. Pract.*, vol. 20, no. 3, pp. 311–317, 2003.

[25] K. De Meester, M. Verspuy, K. G. Monsieurs, and P. Van Bogaert, "SBAR improves nurse–physician communication and reduces unexpected death: A pre and post intervention study," *Resuscitation*, vol. 84, no. 9, pp. 1192–1196, 2013.

[26] S. M. Renz, M. P. Boltz, L. M. Wagner, E. A. Capezuti, and T. E. Lawrence, "Examining the feasibility and utility of an SBAR protocol in long-term care," *Geriatr. Nur. (Lond.)*, vol. 34, no. 4, pp. 295–301, 2013.

[27] B. B. Pope, L. Rodzen, and G. Spross, "Raising the SBAR: how better communication improves patient outcomes," *Nursing2014*, vol. 38, no. 3, pp. 41–43, 2008.

[28] R. S. Panesar, B. Albert, C. Messina, and M. Parker, "The Effect of an Electronic SBAR Communication Tool on Documentation of Acute Events in the Pediatric Intensive Care Unit," *Am. J. Med. Qual.*, p. 1062860614553263, 2014.

[29] S. Hiwale, S. Ulman, and K. Subbaraman, "Barriers to Adoptions of IoT-Based Solutions for Disease Screening," in *Pre-Screening Systems for Early Disease Prediction, Detection, and Prevention*, Hershey, PA, USA: IGI Global, 2019, pp. 50–68.

[30] S Vishwanath, K Vaidya, R Nawal, A Kumar, "Touching lives through mobile health-Assessment of the global market opportunity." Pricewaterhouse Coopers, 2012.

[31] "Philips - Mobile Obstetrics Monitoring Maternal telehealth software," *Philips*. [Online]. Available: https://www.philips.co.id/healthcare/product/HC867055/mobile-obstetrics-monitoring-maternal-telehealth-software. [Accessed: 23-Jul-2019].

[32] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, Mar. 1977.

[33] J. Cowan, "Clinical governance and clinical documentation: still a long way to go?," *Br. J. Clin. Gov.*, vol. 5, no. 3, pp. 179–182, 2000.

[34] J. Stausberg, D. Koch, J. Ingenerf, and M. Betzler, "Comparing paper-based with electronic patient records: lessons learned during a study on diagnosis and procedure codes," *J. Am. Med. Inform. Assoc.*, vol. 10, no. 5, pp. 470–477, 2003.

A System for Collecting Motion Data for Use in Quantitatively Evaluating Activities of Daily Living

Yoshitoshi Murata, Oky Dicky Ardiansyah Prima, Rintaro Takahashi
Graduate School of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
e-mail: {y-murata, prima}@iwate-pu.ac.jp, g231r019@s.iwate-pu.ac.jp

Yukihide Nishimura, Hiroyuki Tsuboi,
Department of Rehabilitation Medicine
Iwate Medical University
Morioka, Japan
e-mail: ynishi@iwate-med.ac.jp, hiroyuki.tsuboi@j.iwate-med.ac.jp

***Abstract*— A system is needed for quantitatively evaluating the activity recovery level of functional disable people. Although functional recovery is administered to hemiplegic patients during rehabilitation, some patients who have recovered function in a rehabilitation facility are still unable to perform daily activities at home. Therefore, recovering activities of daily living (ADL) has become more important than functional recovery. Since existing ADL recovery level indices are based on responses to questionnaires, judgment of recovery level is easily affected by an evaluator's subject. We have developed a system for collecting and storing motion data on daily life activities for use in quantitatively evaluating ADL recovery levels. Evaluation of the system using data measured for a healthy participant with restricted movement and two actual hemiplegic patients demonstrated that slight differences in disability levels can be detected. This system is thus well suited for quantitative ADL assessment for patients with a disability.**

Keywords-rehabilitation; functionary recovery; activities in daily living; ADL; BLE beacon; Google Firebase.

I. INTRODUCTION

Most patients suffering from cerebrovascular disease have paralysis on one side of the body, and their bodies lean and twist to the paralyzed side. Also, because of unusual muscle strain, their hands and feet become stiff. In some cases, muscles of the upper body go into convulsions. Functionary recovery is administered to hemiplegic patients as rehabilitation. However, some patients are not always to live less inconveniently in their home. Some patients who recover hand and arm functionality better in a rehabilitation facility cannot eat meals better in their home. Therefore, recovering Activities of Daily Living (ADL) has recently become more significant than recovering functionaries. We proposed a system to collect motion data on patient's ADL in IARIA eTELEMED 2019 [1].

The Barthel Index, which is based on questionnaires, is popularly used to quantitatively evaluate ADL recovery levels [2][3]. With questionnaires, however, recovery level judgments easily change in accordance with the evaluator's subject. Each recovery level is digitized to a few levels. For example, answers for feeding include "unable," "needs help cutting, spreading butter, etc., or requires modified diet" and "independent." Each answer is scored 0, 5, or 10. However, the recovery level for feeding with help ranges from "a patient eating food directly from dishes without using a spoon or fork" to "a patient eating a meal with a knife and fork in almost the same way as a healthy person." Also, it takes too much time to ask and observe whether a patient can do an activity independently without needing help.

Functional Independence Measure (FIM) [4] and Katz Index [5]-[7] scores are also used to evaluate ADL. FIM scores cover not only functional disease but also mental disease. Scores are broken down into seven levels for each activity, including feeding. Katz Index scores are usually applied to cure elder patients or those suffering from chronic disease.

The question formats for these evaluation methods are basically the same, and an evaluator needs much time to ask questions and observe a patient. We think that a quantitative evaluation system with a computer is needed to evaluate patients objectively without needing to ask them any questions and/or observe them.

Judging ADL recovery levels is based on whether patients can do tasks, such as eating, getting dressed, bathing, washing, and discharging bodily waste by themselves. Therefore, a system that collects motion data of patients in daily life needs to not only measure and collect the motions of body parts but also detect which activities are performed. However, it is very difficult to estimate these merely from changes in acceleration and/or gyro sensor data obtained from devices attached to body parts. Therefore, we estimate activities by using information about places, such as a dining table, bathroom, dressing room, or bedroom. We used the BLE beacon [8] to detect places in this system.

Most surgeons also think that postoperative patient functions assessed by ADL and quality of life have become especially important ways to measure surgical treatment outcomes for the elderly [9].

In this study, we developed a system to collect and store patients' motion data to quantitatively judge **the recovery level of activities in daily living**. The system can use up to seven sensors for simultaneously measuring the motions of seven body parts. A patient's name, measured location, sensor-attached body parts and time-stamps are described as the file names for each measured data file in this system.

The system only requires that recognized medical doctors or physiotherapists can access measured data to maintain security. To ensure this, we developed a data collecting system based on Google Firebase [10]. Since the Firebase application can be independently implemented for any organization, high level security can be maintained.

To confirm whether this cloud-based system can distinguish between normal and restricted movements, we first had a healthy participant rotate both lower arms and then extend them forward to ensure that the system measured these movements correctly. We then restricted movement of the person's elbows and collected movement data during teeth brushing, face washing, and eating.

Next, we applied this system to tow actual hemiplegic patients. Drinking and walking motions were measured.

The measurement data differed slightly depending on the level of movement restriction or disability; this system is thus suitable for quantitatively assessing the ADL level of patients with a disability.

After introducing related work in Section II, we describe the system's design concept and its implementation in Sections III and IV. Confirmation of its performance for a healthy participant with restricted movement is presented in Section V, and that for actual hemiplegic patients is presented in Section VI. Section VII concludes with a summary of the key points and a mention of future work.

II. Related work

To develop a quantitative evaluation system for the recovery level of activities in daily living of hemiplegic patients, we have to know how to evaluate ADL quantitatively, existing life log systems and healthcare information cloud service.

A. *Evaluation index for function level*

Three indexes to evaluate function level in daily living are widely used: the Barthel Index, the FIM and the Katz Index. They are basically questionnaires for daily life activities, such as feeding. The Barthel Index and FIM are popularly applied to evaluate function levels for rehabilitation patients, such as those afflicted with cerebrovascular disease. There are ten question items in the Barthel Index: Feeding, Moving from wheelchair to bed and return, Personal toilet (washing face, combing hair, shaving, cleaning teeth), Getting on and off the toilet (handling clothes, wiping, flushing), Bathing self, Walking on level surfaces, Ascending and descending stairs, Dressing (includes tying shoes, fastening fasteners), Controlling bowels and Controlling bladder [2] [3]. A score of independently doing an activity is usually 10 points, doing it with help is usually 5 points, and not doing it is 0 points.

FIM evaluates not only physical functions but also social abilities, such as communication or social recognition [3]. The number of questions covers 18 issues; 13 for physical functions and five for social abilities. Questions about physical functions are more segmentized. For example, the dressing function is divided into dressing the upper body and the lower body, moving activities are divided into the moving between a wheelchair and a bed/chair, and sitting on a toilet seat and moving to a bathtub. Scores are given on a seven-point system. Independently doing an activity gets seven points, doing it with full help gets one point, and doing it with partial help gets scores ranging from two to six points.

The Katz Index is usually applied to elder patients or those suffering from chronic disease in a variety of care settings [4 - 6]. The index ranks adequacy of performance in six activities: bathing, dressing, toileting, transferring, continence, and feeding. Clients are scored yes/no for independence in each of the six functions.

Every three indexes evaluate whether a patient can do activities in daily living. Therefore, our proposed system must know what kinds of activities a patient tries to do.

In addition, one of the most widely recognized and clinically relevant measures of body function impairment after stroke is the Fugl-Meyer (FM) assessment. Of its 5 domains (motor, sensory, balance, range of motion, joint pain), the motor domain, which includes an assessment of the upper extremity (UE) and lower extremity (LE), has well-established reliability and validity as an indicator of motor impairment severity across different stroke recovery time points. Consistently, greater motor severity as indicated by lower UE and LE FM motor scores is correlated with lower functional ability, such as spontaneous arm use for feeding, dressing and grooming, or walking at functional gait speeds. [11].

B. *Life log system*

Over the years, many researchers have tried to estimate daily life human activities, such as walking and sitting up and down from acceleration and/or gyro sensor data obtained from wearable devices and/or smartphones. In this paper, we refer to the research done respectively by Zhan et al. and Wang et al. [12] [13]. Only a few motions were given in this research; distinctions among activities were not recognized. In contrast, Debraj et al. tried to recognize 19 daily living activities [14]. They collected environment information, such as that for temperature and location in addition to activity information. They used GPS and BLE beacons to identify places. However, they did not consider the Barthel Index or other indices and consequently their target activities did not correspond to activities in the index of function recovery levels.

C. *Healthcare cloud service*

Zhang et al. developed a cyber-physical system for patient-centric healthcare applications and services [15]. They called it Health-CPS. It was built on cloud and big data analytics technologies. It consisted of a data collection layer, a data management layer and an application service layer to collect and follow up on many kinds of big data. It used a security tag to maintain security.

Doukas et al. proposed a mobile system that enables electronic healthcare data storage, update and retrieval using cloud computing [16]. A mobile application was developed using Google's Android OS and Amazon's S3 to provide management of patient health records and medical images.

We developed a cloud service whose collecting function for medical data is basically the same as that for the above systems. However, our system is specialized so that it can collect activity and place information to functionally evaluate recovery levels that correspond to existing evaluation methods, such as the Barthel Index. In this paper, we show how we implemented the system with Eri BLM620 [17] as the sensor node, as well as Android smartphone, BLE beacon, and Google Firebase.

III. SYSTEM DESIGN CONCEPT

We designed the proposed system so that it could not only evaluate ADL for a patient, but also develop algorithms for detecting whether a patient can do a designated activity. The system collects and stores sensor data and video data synchronously and allows appropriate persons to access stored data. We designed the system while taking the following issues into consideration:

1) Suppressing battery consumption for wearable sensor devices,
2) Suppressing recorded data and collecting necessary data,
3) Maintaining security.

Google Firebase service provides many functions, including authentication and real-time database functions, to enable systems to be managed effectively, such as through the means of allowing access to authorized persons. Since any organization can independently implement Firebase applications, it becomes possible to maintain high level security. This is why we implemented our data collecting system on Google Firebase.

The image of a data collecting system that collects data about the motions that a patient performs daily is shown in Figure 1. The system we propose consists of sensor devices, a sensor relay unit (smartphone), BLE beacons, and Google Firebase. A smartphone is used as the sensor relay unit that controls sensor devices and temporarily stores and forwards measured data to the Firebase.

BLE beacons are placed in various locations: under a dining table, on top of a toilet, in a bathroom, in a bedroom, in a closet. When the smartphone receives a BLE beacon signal level that exceeds the threshold level, it sends a message to sensor devices to start measuring data. And when the smartphone receives a receiving signal level lower than the threshold level, it sends a message to sensor devices telling them to stop measuring data. Sensor devices and smartphones are managed by the Realtime Database on Google Firebase. Security is maintained by enabling only authorized persons using the system, including patient, readers, such as medical doctor and installation personnel, such as nurse are also managed by the Realtime Database is used to maintain security. In this system, measured data are downloaded for pre-registered persons from the web server.

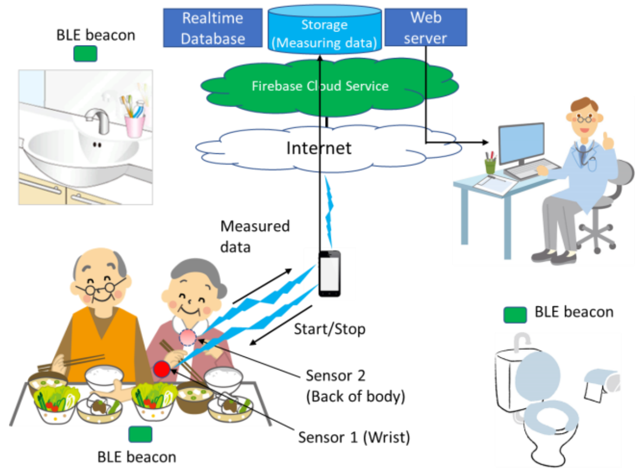


Figure 1. Image of the data collecting system for patient's daily life motions.

IV. SYSTEM IMPLEMENTATION

We developed a PatientApp program that works on the sensor relay unit and a DataCollectionServer program that works on Firebase. The PatientApp manages sensor devices, gets measured data from sensor devices and uploads the data file to the DataCollectionServer.

A. PatientApp

This time, we developed a PatientApp program based on the Android Framework. With this program, a developer must first access the Firebase and download a configuration file. An Android application package file (Apk File) is then made as a building application and is connected to Firebase. This makes it possible to securely download the Apk File for each organization.

Before starting to measure sensor data and/or video data, it is necessary to enter a patient's name, bind a sensor with a body part, bind a BLE beacon with a place of activity and select a video recording on/off function. Therefore, we designed a transition diagram of UI pages as shown in Figure 2. There were three alternatives for a user name at the login; the patient's name, the medical worker's name with measuring devices set up, and the medical professional's name with measured data analyzed. For the latter two cases, a patient's name must be entered after the login. Therefore, we decided on the first one, login with a patient's name.

After login, a "List of setting up" page is presented. An example of this page is shown in Figure 3. With it, a user can confirm a state of setting. When the "Change" button is clicked, the page will change to the "Sensor" page to bind a sensor with a body part. When the "Next" button is clicked, the page will change to the "Beacon" page to bind a BLE beacon with a place in activity. When the "Next" button is clicked, the page will change to the "Video" page to select video ON/OFF. When the "Next" button is clicked, the page will change to the "List of setting up" page. When the "Next" button is clicked in the "List of setting up" page, the page will change to the "Measuring" page. When the "Start" button on this page is clicked, the PatientApp sends a message to the sensor messages to start measuring, and the "Start" button changes to the "Stop" button. When the "Stop" button is clicked, the PatientApp sends a message to the sensor messages to stop measuring, and the "Stop" button changes to the "Start" button. When the "End" button is clicked, the PatientApp finishes.

When a sensor receives a BLE beacon signal, it starts measuring, and, when a sensor loses a BLE beacon signal, it stops measuring. After clicking the "Stop" button, measured data are changed to a measured data file. Its file name is "Patient name_place_body part_timestamp" to recognize its properties. The file is uploaded to the storage in Firebase.

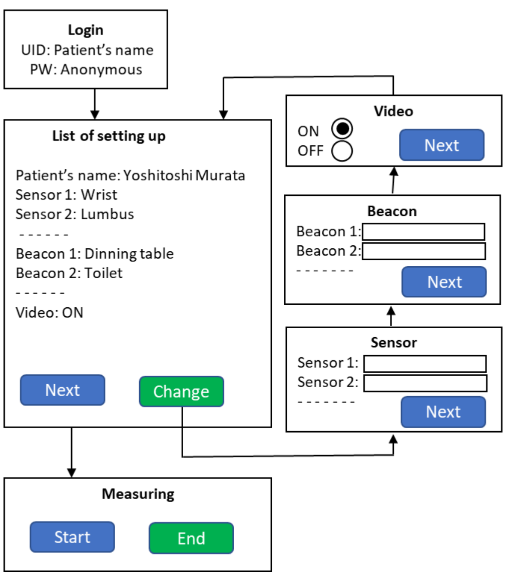


Figure 2. Transitions of UI pages in the PatientApp.

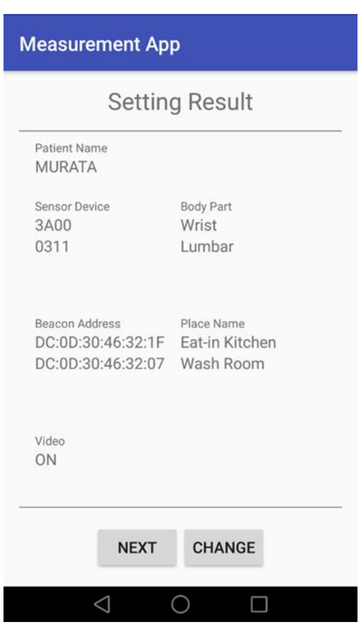


Figure 3. Example of setting up page list.

We developed the following six packages of classes to achieve the above proceedings:

- Beacon: receiving beacon signals and handing their information to other classes.
- Mobile2wear: controlling a sensor device and receiving measured data.
- Camera: managing a video camera.
- Firebase: converting measured data and transferring the data to the Firebase storage.
- View: managing transition of pages
- Viewmodel: listening events on buttons or input boxes and handing, such information to other classes.

B. DataCollectionServer

The DataCollectionServer has the following functions;

- Data upload function: The sensor relay unit temporarily stores measured data and forwards them to the server.
- Data download function: Authorized persons, such as medical doctors can access the DataCollectionServer and download measured data files securely.

It consists of the Storage and WebSite. The WebSite collaborates with the Storage and provides a file download function to a medical professional through the Web browser.

In this subsection, we mainly introduce how to upload and download measured data file.

1) Data upload function (Figure 4)

After a measured file has been made, the PatientApp uploads the file to the storage server in Firebase as shown in Figure 4. The storage server generates the file download URL, which is managed in the Realtime Database.

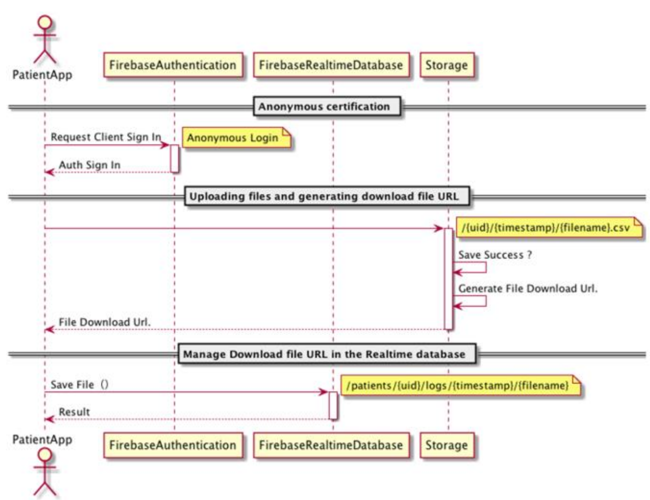


Figure 4. Sequence flow to upload measured files.

2) Measured data download function (Figure 5)

Supervisors input the access account of medical professionals from the management page in Firebase. The sequence flow with which medical professionals download their patients' files is shown in Figure 5. When medical professionals access the Website, they log in with their assigned ID and password on the page of Figure 6 (a).

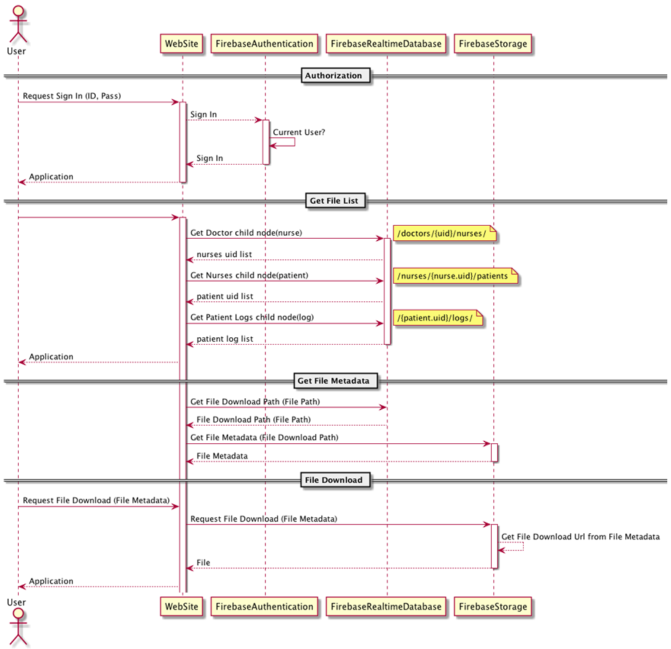
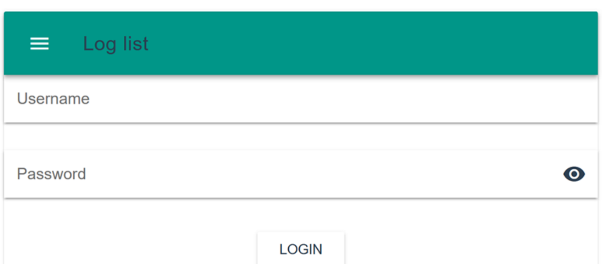


Figure 5. Sequence flow to download measured files.



(a) Login page

(2) File list page

Figure 6. WebSite user interface.

After login, the Website application accesses the Realtime Database to get information related to nurses and patients. The Website application also gets meta-data such as an access path to a stored file. When a medical professional clicks a file on the page of Figure 6 (b), the Website application accesses the indicated file on the Storage through the access path. Finally, the indicated file is downloaded.

C. *Wearable device*

We initially used a SONY Smart Watch III [18] as the wearable device, as reported at eTELEMED 2019. However, SONY no longer produces this device, so we worked with Eri, Inc. [17] to develop the BLM620 wearable sensor to ensure a stable supply.

Its appearance and configuration are shown in Figures 7 and 8. It contains a 3D digital accelerometer and a 3D digital gyroscope packaged together (LSM6DSL, STMicroelectronics [19]) and a Bluetooth and CPU module packaged together (HRM1062, Hosiden [20]). Its acceleration and gyro axes are shown in Figure 9. The maximum number of simultaneous connections is seven.

Its connection performance was measured using an Android terminal wirelessly connected to seven wearable devices, as shown in Figure 10, for two types of connection: multi-thread and sequential. The flow for each type is shown in Figure 11. The data were measured in lower and higher radio interference environments. The interference in the latter one was generated by a nearby 2.8 GHz WiFi access point. There was no such interference source in the former one. The terminal connected to each device 30 times. The connection error rate (CER) is shown in Figure 12. While there were differences in the CER between devices, the CER was generally higher in the higher radio interference environment than in the lower one. It was also higher for the multi-thread connections than for the sequential connections. Therefore, we used sequential connections in this application program.

Figure 7. Appearance of the developed wearable device

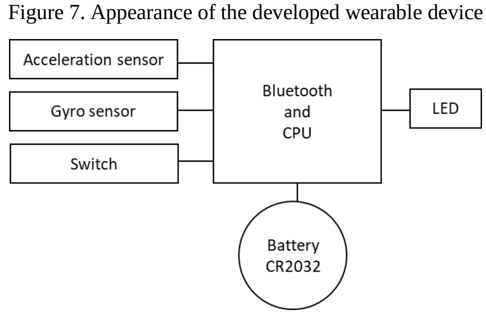


Figure 8. Configuration of the developed wearable device

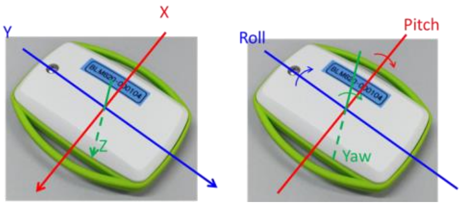
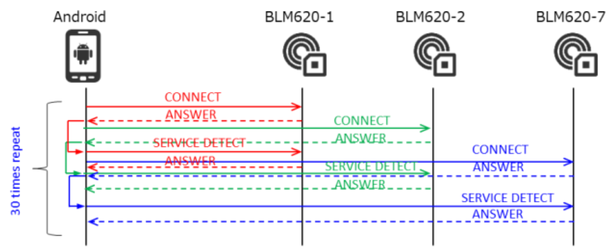
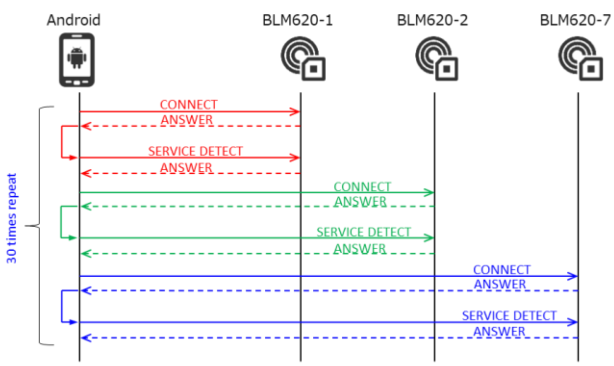


Figure 9. Acceleration and gyro axis

Figure 10. A scene of experiment



(a) Multi-thread connection



(b) Sequential connection

Figure 11. Connection flow

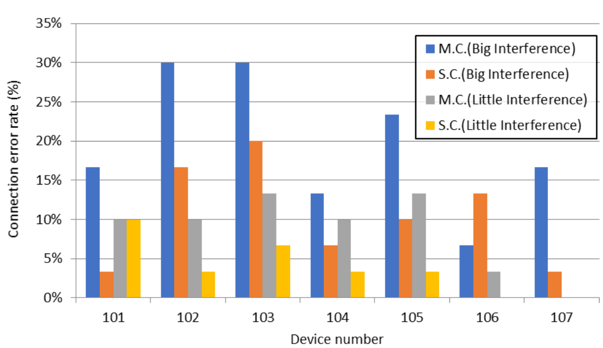


Figure 12. Connection error rate

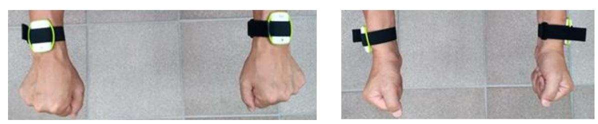
V. COMFIRMATION OF SYSTEM PERFORMANCE

Prior to collecting motion data for actual hemiplegic patients, we collected motion data for a healthy participant with and without elbow restrictions to confirm that the proposed system can detect differences between normal and restricted joint movement.

A. *Simple motions*

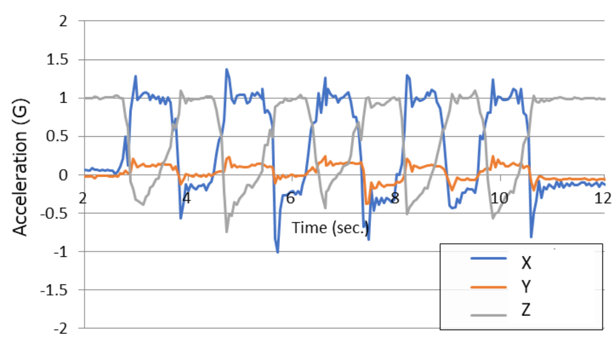
We started with simple motions for which it is easy to confirm the accuracy of measured data. We first had the participant rotate both lower arms 90°, as shown in Figure 13, five times. The measured acceleration and yaw/roll/pitch angle data are shown in Figure 14. The acceleration along the X and Z axis basically changed from 0 to 1 G alternately (Figure 14 (a)) five times. The acceleration along the X axis changed from 0 to −1 G while that along the Z axis changed from 0 to 1 G alternately (Figure 14 (c)) five times.

The roll angles for the two arms (Figures 14 (b) and (d)) were symmetrically opposite due to their symmetrical motions. The acceleration along the Z axis when both thumbs were in the "right up" position (Figures 14 (a) and (c)) was not zero. This reason is caused by over actions of the arms. The yaw/roll/pitch angles in Figures 14 (b) and (d) include the drift error.

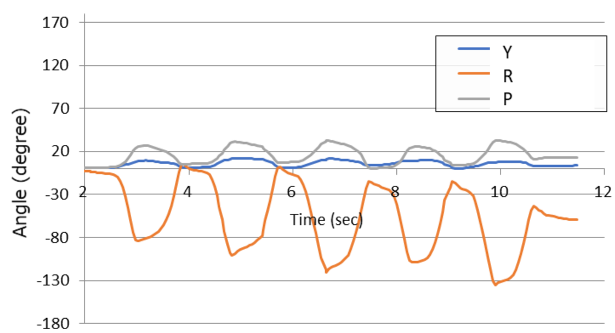


(a) Start (b) End

Figure 13. Rotating the lower arm



(a) Acceleration (left arm)



(b) Yaw/Roll/Pitch (left arm)

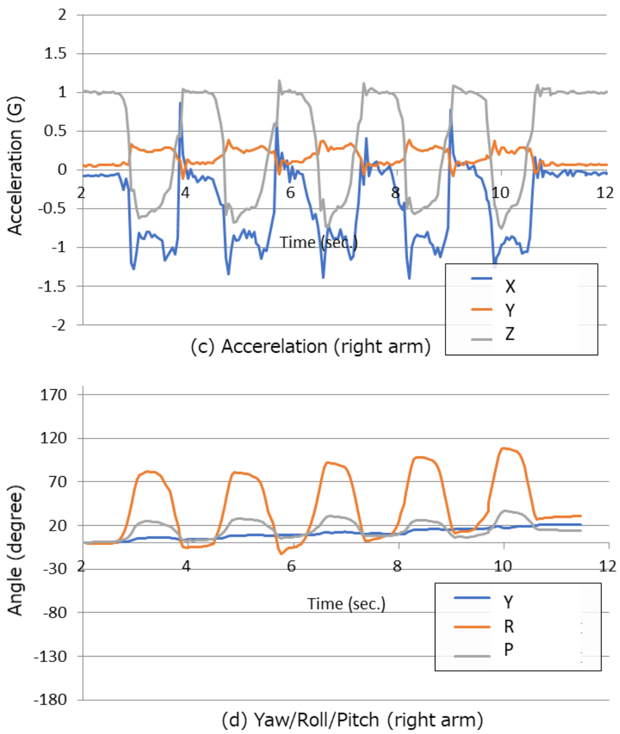
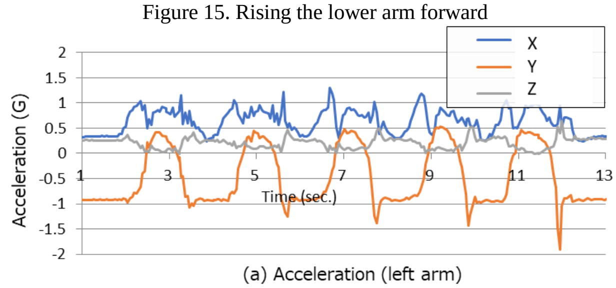


Figure 14. Measured data of rotating the lower arm

The participant raised both lower arms from the straight down position to the forward position, as shown in Figure 15, five times. The measured acceleration and yaw/roll/pitch angle data are shown in Figure 16. The acceleration along the Y axis for the straight down position was roughly −1 G (Figures 16 (a) and (c)) since the sensors simply measured gravity. The acceleration along the X axis corresponded to the centrifugal force. The symmetric differences in yaw angle between Figures 16 (b) and (d) were due to the symmetrical motion. The yaw/roll/pitch angles in Figures 16 (b) and (d) include the drift error, the same as in Figures 14 (b) and (d). The measured data in Figures 14 and 16 basically represent the changes in motion accurately.

(a) Start (b)End

Figure 15. Rising the lower arm forward



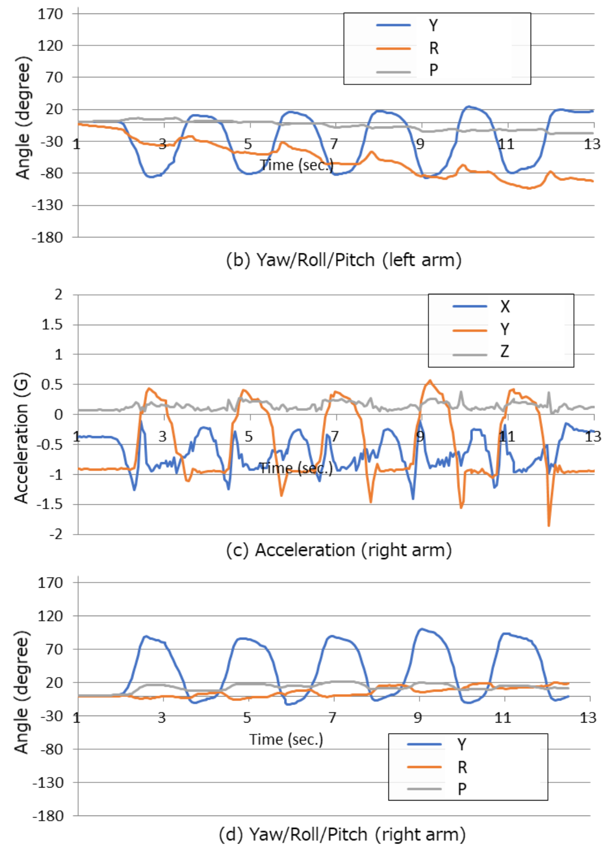


Figure 16. Motion data of rising the lower arm forward

B. *Restricted motions*

To confirm whether the proposed system can detect differences between different motion restrictions, we measured the participant's motions during eating, face washing, and teeth brushing under three conditions;

a. Bending of the right elbow was restricted by placing it in a plaster cast (Figure 17), wrapping it in a bandage, and fixing it to his upper body with a bandage,
b. Bending of the right elbow was restricted by placing it in a plaster cast and wrapping it in a bandage, without fixing it to anything.
c. No restriction.

Wearable devices were attached to the head, the mid-lumbar region, both lower arms, and both upper arms, as shown in Figure 18. Under conditions a and b, the participant brushed his teeth with his right hand, as shown in Figure 19 (a). Under condition c, he brushed his teeth with his left hand since he usually brushes with his left hand. The acceleration data for the lower arms and head are presented in Figure 20. The other data are not presented as they did not have any particular features of interest. Since there was a lot of right and left or up and down motion and little rolling motion in brushing teeth, the angle data for the tooth-brushing arm had little variation. There were big differences in the data between conditions a and b, as shown in Figures 20 (a-2) and (b-2) while there was little difference between conditions b and c, as shown in Figures 20 (b-2) and (c-1). There was little head movement under any condition, as shown in Figures 20 (a-3), (b-3), and (c-3).

Figure 17. Plaster cast

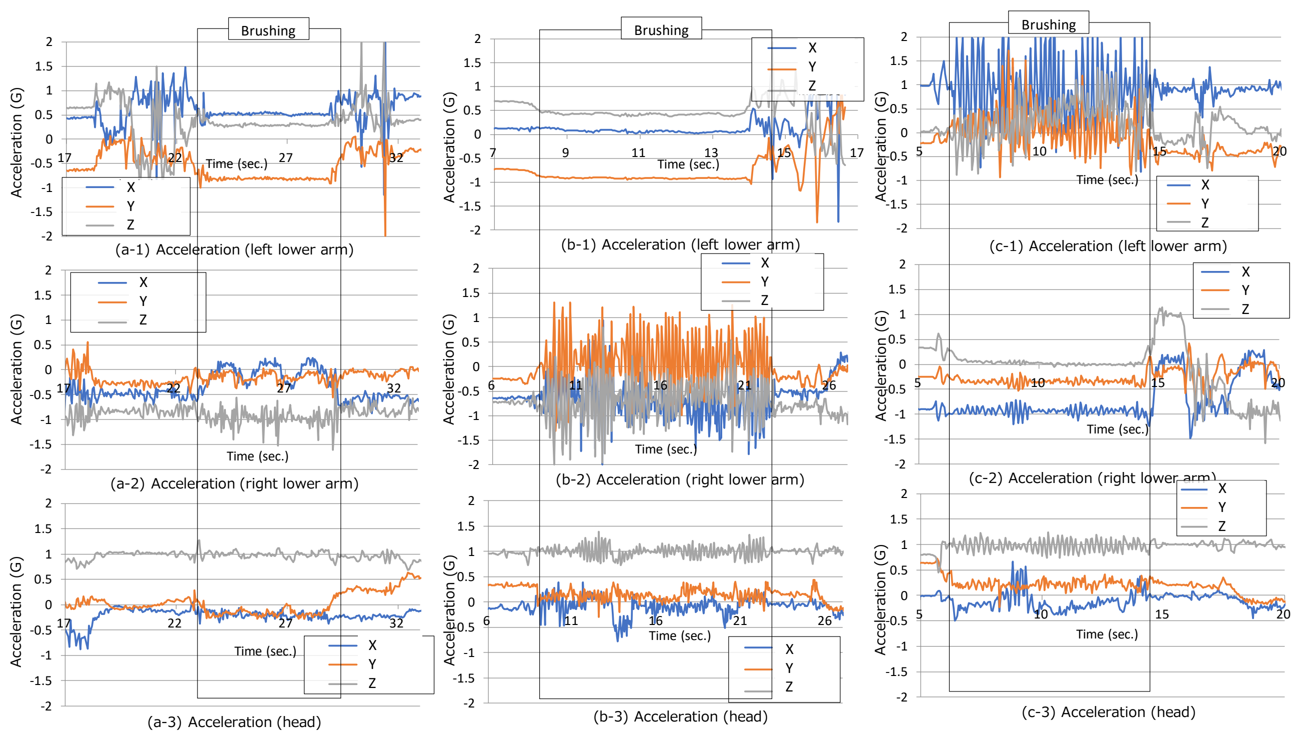
Figure 18. Participant with sensors and restrictions

(a)Brushing teeth (b)Washing face (c)Eating food

Figure 19. Restricted motions

The participant washed his face with his right hand under conditions a and b, as shown in Figure 19 (b). He washed his face with both hands under condition c. The acceleration data for the lower arms and head are presented in Figure 21. The other data are not presented as they did not have any particular features of interest. Since there was a lot of up and down motion and little rolling motion in washing face, the angle data for the face washing arm had little variation. There were not any big differences in the data between conditions a and b, as shown in Figures 21 (a-2) and (b-2). There was a big difference in the data between condition c and the other two conditions: the acceleration data for the lower arms varied widely, as shown in Figures 21 (c-1) and (c-2) due to using both hands. There was little head movement under any condition, as shown in Figures 21 (a-3), (b-3), and (c-3).

The participant ate curry rice with his right hand under all three conditions, as shown in Figure 19 (c). The acceleration and angle data for the right lower arm are presented (Figure 22), since eating food with a spoon involves much rolling motion. The angle of head for the direction of gravity is also presented (Figure 22 (a-3), (b-3), and (c-3)). The range of change in the acceleration Y, Z, and pitching of his right hand are bigger, as there was less motion restriction. The angle of head for the direction of gravity during eating is bigger, as there was less motion restriction. The participant had to close his face to curry and was hard to roll his hand during eating in condition a.



(a) Condition a (b) Condition b (c) Condition c

Figure 20. Data collected during teeth brushing.

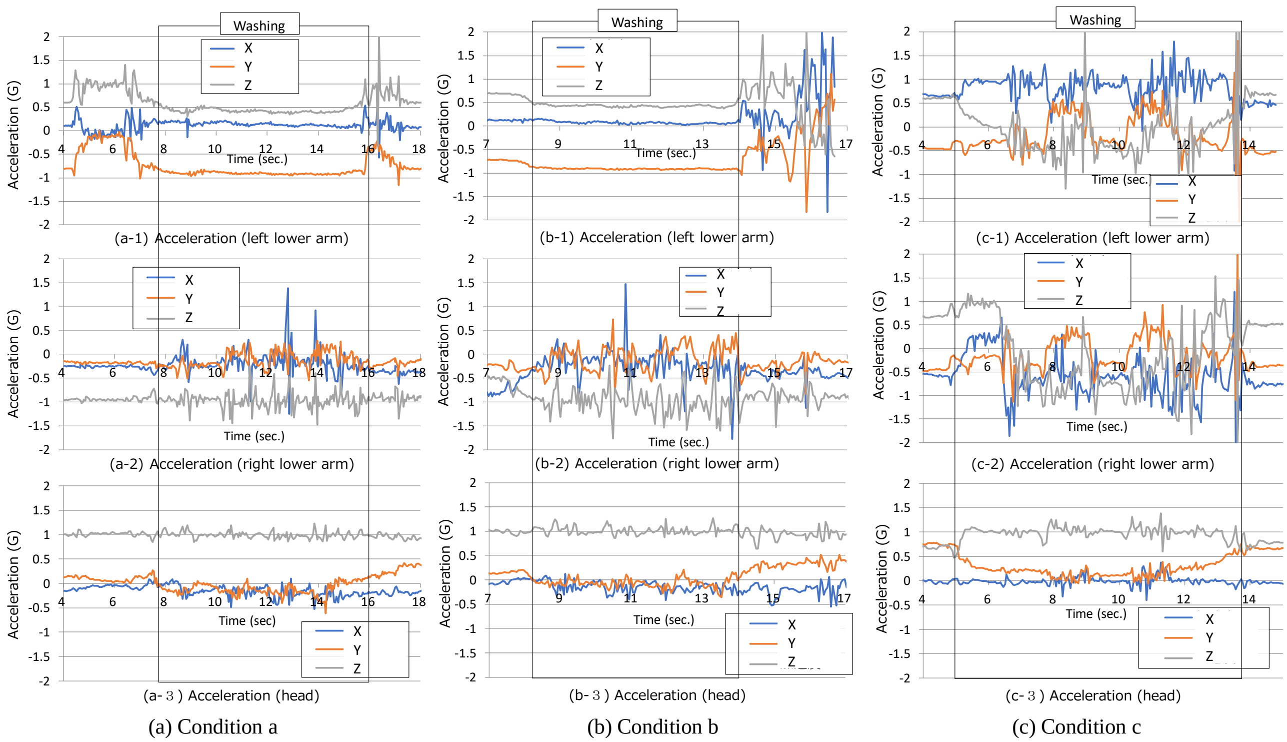


Figure 21. Data collected during face washing.

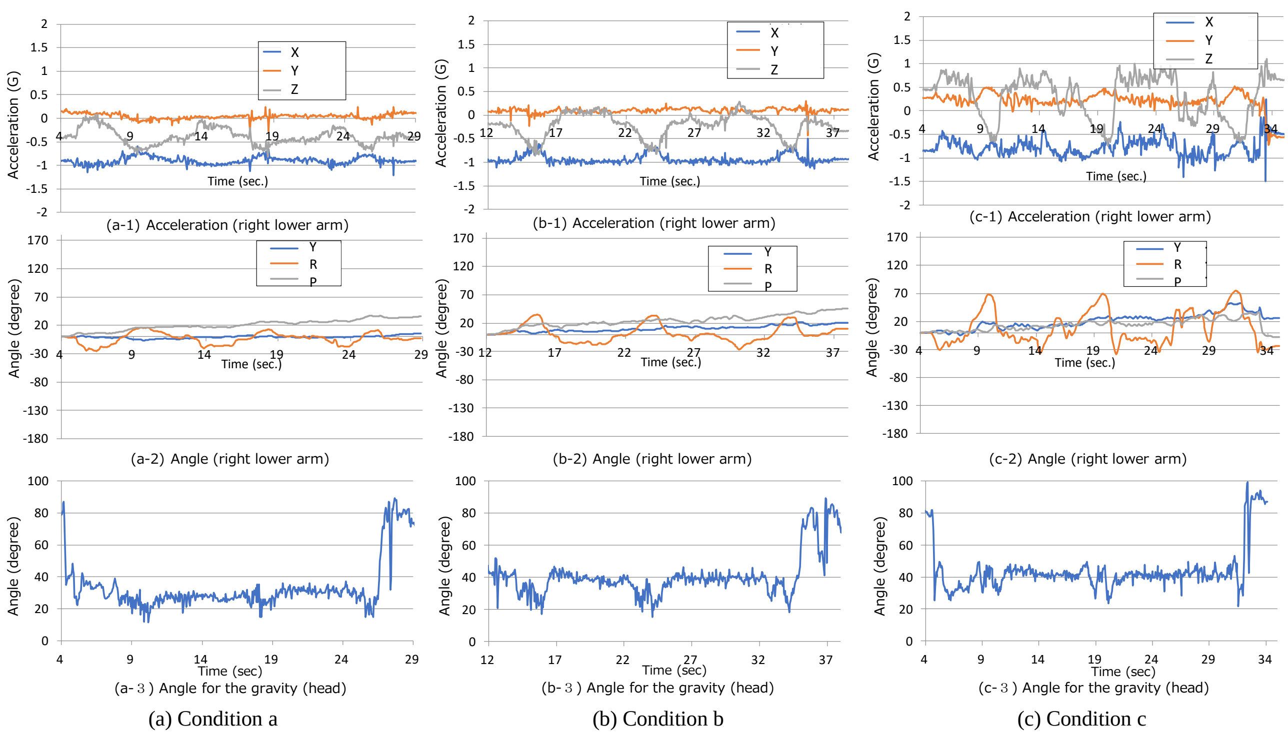


Figure 22. Data collected during eating.

These results demonstrate that measured data for lower arm motions are effective for detecting differences in motion restriction levels. Although it impossible to detect a difference using data for a single activity, such as teeth brushing or face washing, it is possible to detect one using data for a combination of activities, such as teeth brushing, face washing, and/or eating.

VI. MEASUREMENT FOR HEMIPLEGIC PATIENTS

We collected and analyzed data for the walking and drinking motions of two hemiplegic patients and three healthy participants who wore seven wearable devices for collecting data. Their placements for each motion are shown in Figure 23. Data were collected for three stable walking cycles and for one drinking motion on the paretic side. The UE and LE functionalities of the two hemiplegic patients were assessed on the basis of FMA by physical and occupational therapists. Hemiplegic patient A had severe impairment on the paretic side (FMA UE score: 25; LE score: 14) while hemiplegic patient B had mild impairment on the paretic side (FMA UE score: 58; LE score: 26).

The data were collected safely and smoothly for both the hemiplegic patients and healthy participants. The walking and drinking motions during collection were the same as their usual motions. The periods were longer for the patients due to their severe impairment. Since the period of time during walking and drinking and the acceleration and angle data for every healthy participant were similar, data for a typical healthy participant were presented in this paper.

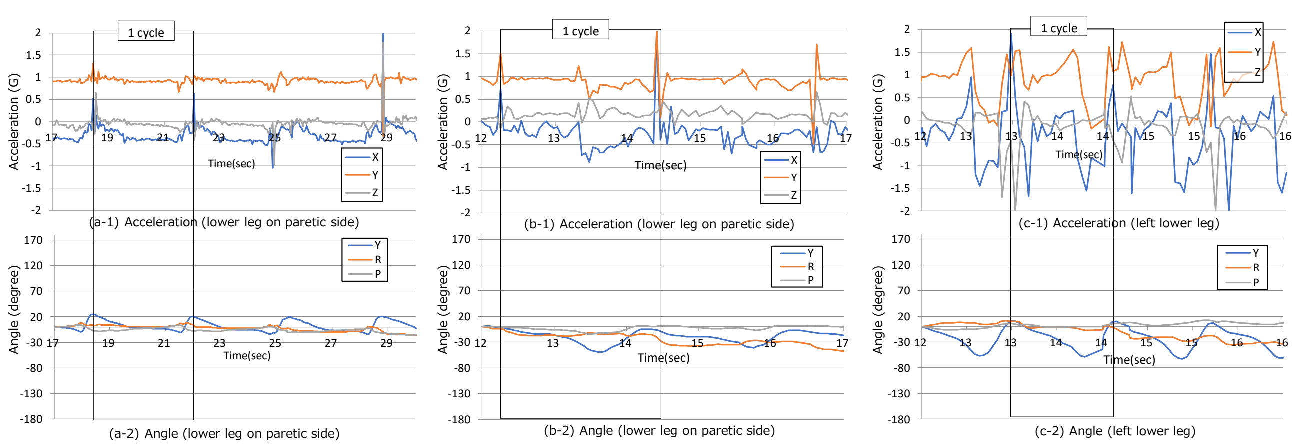
Figure 24 shows the raw acceleration and angle data for the paretic-side lower leg for hemiplegic patients A (a-1 and 2) B (b-1 and 2), and for the left lower leg for the healthy participant (c-1 and 2) for the walking motion. While it is difficult to recognize walking gait cycles from the acceleration data for hemiplegic patients B and the health participant, the walking gait cycles are clearly recognized in the yaw angle data for all participants. The yaw angle indicates forward movement in the lower leg. The range for the healthy participant is biggest in three participants. The range is smaller for the hemiplegic patients due to their severe impairment.

Figure 25 shows the raw acceleration and angle data for the paretic-side lower arm and for the head for hemiplegic patients A (a-1, 2, and 3) and B (b-1, 2, and 3), and for the left lower arm and head for the healthy participant (c-1, 2, and 3) for the drinking motion. The data indicate a larger forward movement of the head for the hemiplegic patients than for the healthy participant. And, the range of yaw angle of the lower arm of patient A is smaller than that of patient B and the healthy participant. This difference is attributed to the severe impairment and compensatory movements of the patients.

This experiment demonstrated that this device and system can safely and smoothly collect motion data for hemiplegic patients as well as healthy individuals. They are thus suitable for quantitative assessment of ADL for hemiplegic patients.

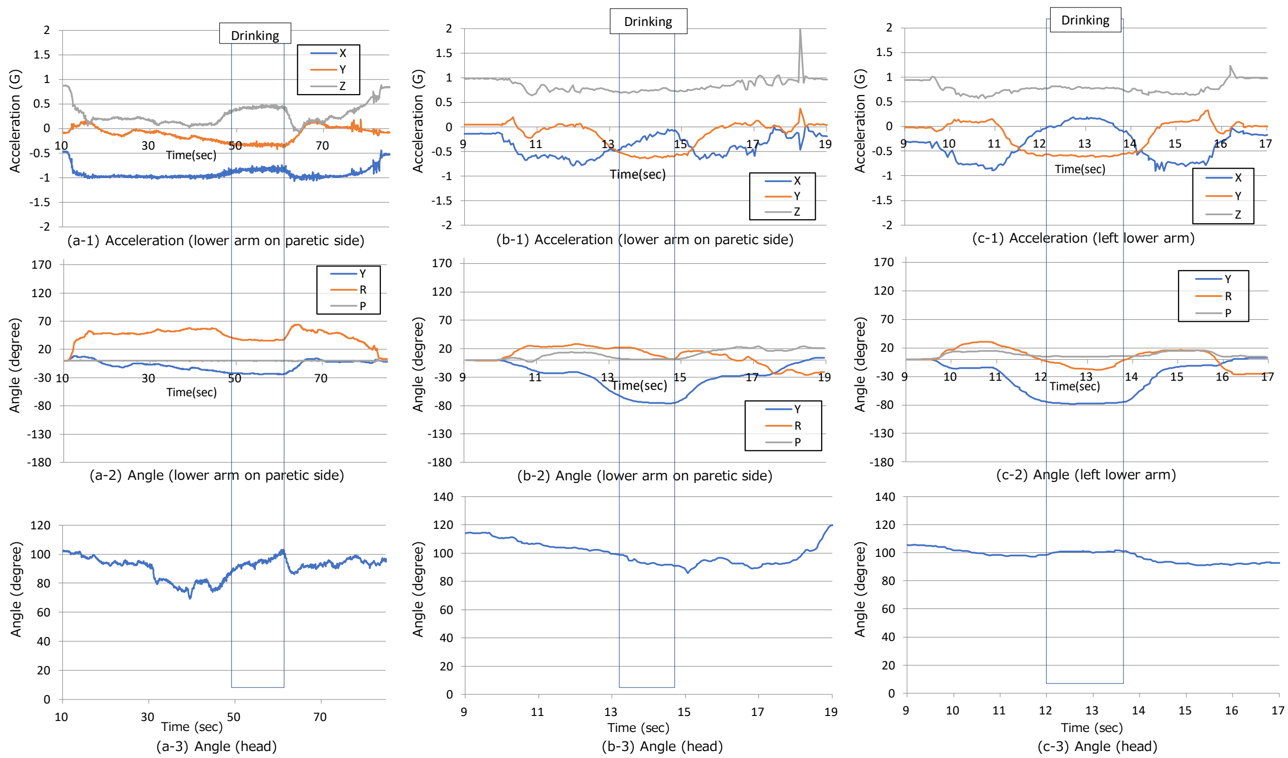
(a) Drinking a cup of water (b) Walking

Figure 23. Experimental scenes with a hemiplegic patient



(a) Hemiplegic patient A (b) Hemiplegic patient B (c) Healthy participant

Figure 24. Data collected walking



(a) Hemiplegic patient A (b) Hemiplegic patient B (c) Healthy participant

Figure 25. Data collected drinking

VII. CONCLUSION AND FUTURE WORK

Existing evaluation indexes for activities in daily living (ADL) recovery levels such as the Barthel Index are based on responses to questionnaires. Therefore, the judging of recovery levels can be easily affected by an evaluator's subject. We have presented a system for collecting and storing motion data about daily life activities for use in quantitatively evaluating ADL recovery levels. The system was developed on the basis of Google Firebase. We used information about places such as a dining room and a bathroom to estimate the type of activity. The places are detected using Bluetooth beacons.

Measurement results obtained for a healthy volunteer with restricted movement demonstrated that it is possible to detect slight differences in the restriction level. However, it is difficult to estimate whether the motions can be performed without help.

Through the experiment measuring for hemiplegic patients, the proposed system can collect motion data safely and smoothly. Measurement results obtained from two hemiplegic patients whose severity of impairment were different shows that it is possible to detect slight differences in the severity.

Planned improvements to the proposed system include uploading video and GPS data to a cloud server. GPS data will enable measurement of motion during walking or running outdoor.

Our goal is to develop a new index for evaluating ADL recovery levels on the basis of big motion data measured for people performing various activities.

ACKNOWLEDGEMENT

The authors extend thanks to Mr. Kazuhiro Yoshida, Mr. Fumiaki Yamaguchi, and Mr. Shunpei Okimura for the help he provided in performing this research. The research and development was supported by the MIC/SCOPE #181602007.

REFERENCES

[1] R. Takahashi, Y. Murata, and O. D. A. Prima, "A System for Collecting Motion Data on Patients' Activities of Daily Living," IARIA, Proceedings of eTELEMED 2019, pp. 7-12, 2019.

[2] The Barthel Index; http://www.strokecenter.org/wp-content/uploads/2011/08/barthel.pdf [retrieved: November, 2019].

[3] F. I. Mahoney and D. Barthel, "Functional evaluation: The Barthel Index.", Maryland State Medical Journal 1965, vol.14, pp. 56-61, 1965.

[4] D. Chumney et al., "Ability of Functional Independence Measure to accurately predict functional outcome of stroke-specific population: Systematic review," Departmentof Veterans Affairs, Journal of Rehabilitation Research & Development, Volume 47, Number 1, pp. 17-29, 2010.

[5] M. Shelkey and M. Wallace, "Katz index of independence in activity of daily living (ADL)," https://www.semanticscholar.org/paper/Katz-Index-of-Independence-in-Activities-of-Daily-Shelkey-Wallace/fb433f328b79f56d82a52551960c93da74469baa

[retrieved: November, 2019].

[6] S. Katz et al., "Progress in development of the index of ADL," Gerontologist 1970, vol. 10, pp. 20 –30, 1970.

[7] M. Shelkey, and M. Wallace, "Katz Index of Independence in Activities of Daily Living (ADL)," The Hartford Institute for Geriatric Nursing, New York University, College of Nursing, Issue Number 2, 2012.

[8] Bluetooth Technology; https://www.bluetooth.com/bluetooth-technology/radio-versions, [retrieved: November, 2019].

[9] T. Amemiya et al., "Activities of Daily Living and Quality of Life of Elderly Patients After Elective Surgery for Gastric and Colorectal Cancers," Lippincott Williams & Wilkins, Annals of Surgery, Volume 246, Number 2, pp. 222-228, 2007.

[10] Google Firebase Cloud Messaging, https://firebase.google.com/docs/cloud-messaging/?hl=en, [retrieved: November, 2019].

[11] K. J. Sullivan et al., "Fugl-Meyer assessment of sensorimotor function after stroke: standardized training procedure for clinical practice and clinical trials." Stroke, vol. 42, pp. 427-432, 2011

[12] K. Zhan, S. Faux, and F. Ramos, "Multi-scale Conditional Random Fields for First-Person Activity Recognition on Elders and Disabled patients," ELSEVIER, Pervasive and Mobile Computing, Volume 16, Part B, pp. 251-267, 2015..

[13] L. Wang, T. Gu, X. Tao, H. Chen, and J. Lu, "Recognizing Multi-User Activities Using Wearable Sensors in a Smart Home," ELSEVIER, Pervasive and Mobile Computing, Volume 7, Number 3, pp. 287-298, 2011.

[14] D. Debraj, B. Pratool, K. D. Sajal, and C. Sriram, "Multimodal Wearable Sensing for Fine-Grained Activity Recognition in Healthcare," IEEE, Internet Computing, pp. 26-35, 2015.

[15] Y. Zhang et al., "Health-CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data," IEEE, Systems Journal, Volume 11, Number 1, pp. 88-95, 2017.

[16] C. Doukas, T. Pliakas, and I. Maglogiannis, "Mobile Healthcare Information Management utilizing Cloud Computing and Android OS," IEEE, 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 1037-1040, 2010.

[17] Eri, Inc., https://www.erii.co.jp/en/, [retrieved: November, 2019].

[18] SONY Smart Watch 3, https://www.sonymobile.com/global-en/products/smart-products/smartwatch-3-swr50/#gref, [retrieved: November, 2019].

[19] LSM6DSL, STmicroelectronics, https://www.st.com/en/mems-and-sensors/lsm6dsl.html, [retrieved: November, 2019].

[20] HRM1062, Hosiden, https://www.hosiden.co.jp/news/product/hrm1062.html [in Japanese, retrieved: November, 2019].

Evaluation of The KINECT-Based Auscultation Practice System
―Turning simulated patients into real patients―

Natsuko Miura, Yoshihito Endo
Faculty of Nursing
Iwate Prefectural University
Takizawa, Japan
e-mail: natsuko@iwate-pu.ac.jp
e-mail: y-endo@iwate-pu.ac.jp

Yoshitosi Murata, Kazuhiro Yoshida
Faculty of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
e-mail: y-murata@iwate-pu.ac.jp,
e-mail: kyoshida@ipu- office.iwate-pu.ac.jp

Abstract **- Our group developed a system for practicing breath sounds auscultation using KINECT. This system solves several problems associated with past simulation education models. The system is inexpensive, and simply operated. We evaluated the process of learning breath sounds auscultation with a nurse and a nursing student. In this paper, we introduce practical exercises, using the KINECT system. We also report the student's evaluation of the system. We additionally completed an auscultation test using an existing breathing sound file and breathing sound data obtained using the KINECT system, and compared the results. We obtained responses from 78 students. All students replied that they were interested in the KINECT system, and 83.3% of students were able to distinguish an accessory murmur from normal breath sounds. In all, 97.4% of students reported feeling motivated to learn using the KINECT system. The result of the examination to identify a kind of the respiratory sound using the apparatus which we developed did not have a result and the change using existing sound data. The KINECT system was useful for learning breath sounds. Use of the system was interesting to the students because of the simulated patient interaction, similar to those encountered in "real life" clinical settings. Using the realistic KINETIC system, the nursing students were able to develop the skills necessary to distinguish breath sounds. Additionally, the system was motivating for nursing students.**

Keywords-simulation; auscultation; nursing; physical assessment.

I. INTRODUCTION

We developed an auscultation practice system using KINECT to solve various problems related to simulation-based education [1] [2]. KINECT is a relatively inexpensive, easy-to-operate system that can produce respiratory sounds (e.g., wheezes) in synchronization with the respiration of a simulated patient.

In Japan, slowing birth rates and an aging population, as well as advanced medical treatments limit the scope and opportunities available for nursing students to practice nursing techniques within internship settings at hospitals; additionally, enhanced consciousness of medical care safety can limit opportunities to practice [3]. For example, most inpatients in Japan are elderly. Therefore, in clinical practice it is difficult to be in charge of adolescents and children. In addition, nursing students cannot perform invasive procedures, such as blood sampling or intravenous IV injection.

Amid this climate, simulation-based nursing education is poised for wide dissemination, allowing students to repeatedly experience realistic practical settings without risking patients' safety. Simulations represent learner-oriented education, and the equipment has been introduced to our university [4].

There are three types of trainings conducted in simulation education, i.e., task training for acquiring techniques, algorithm training like learning Basic Life Support, and situation-based training in which various clinical situations are reproduced [5].

Education with a highly functional simulator capable of computerized control of vital signs, breath sounds, and heart sounds has been practiced in areas such as intensive care [5] and in operating rooms [6]. Simulation training is necessary for learning team-based cooperative skills involving nurses and physicians [7].

This highly functional simulator is unable to simulate conversation; however, it has a limited capacity to elicit simulated communication via an integrated microphone (operated remotely) in cases where conversation is warranted [8]. Communication is an important skill for nurses, and communicative competence is very important to patients' care [9]. Therefore, simulated patients designed to replicate patient-specific sentiments and personalities, not only in terms of clinical history and physical findings, are widely used for medical staff and allied health students [5] [9].

However, we cannot reproduce abnormal breath sounds because a simulated patient is a healthy person. In exercises for learning diseases such as pneumonia, a hybrid simulation is conducted where conversation is made with a simulated patient and respiratory sounds are auscultated by a simulator placed nearby, but this is also unnatural. These problems interfere with the natural flow of clinical settings.

Furthermore, high-performance simulators and existing auscultation training equipment are very expensive. At our university, more than 90 nursing students are registered in one class. The number of expensive simulators necessary to provide training and practice opportunities to this many students is cost-prohibitive.

We therefore developed the auscultation practice system using KINECT to address these issues. In our system, a simulated patient play the role of a patient, instead of a humanoid, and stethoscope locations on the body are measured with KINECT. Movements of the upper body from breathing can also be detected by KINECT. Also, appropriate disease sounds including normal ones can be assigned at four points on the upper body. Practicing students hear such disease sounds, synchronized with the movement of breathing, through earphones when a stethoscope is placed on the assigned area [1] [2].

This paper describes our early experiences using the KINECT system. Additionally, we report student assessments of the system. In Section Ⅱ, we describe related simulators and their problems. In Section Ⅲ, the method employed in the study is explained. In Section Ⅳ, the results of student assessments of the system, are outlined. A brief discussion is offered in Section Ⅴ. In Section Ⅵ, the method of simulation training using this system is described, conclusions and recommendations for future research are presented in Section Ⅶ.

II. Related Work

In this section, we describe related simulators and their problems.

There are several varieties of equipment currently used for learning auscultation of respiratory sounds. Kyoto Science's breathing sound auscultation simulator "Lang" [10] is an upper body instrument (Fig. 1). Users can auscultate from the anterior chest and back. Heart sounds are also audible. Also, the pedestal illuminates in accordance with inspiration and expiration.

The Sakamoto model "Choushin kun" is a similar upper body simulator (Fig. 2) with seven built-in speakers [11]. For each speaker, breathing sounds can be selected and the volume adjusted. Like Kyoto Science's "Lang," heart sounds are also audible.

The person-like simulator used in this research was the ALS simulator, made by Laerdal Medical [12] (Fig. 3). Speakers are built into both sides of the chest. The thorax can be moved up and down by injection of air from the outside. Heart sounds are also audible. The ALS is a high-performance simulator capable of displaying values such as electrocardiogram and pulse oximetry on the monitor.

In the model released from Cardionics, a seal is placed on the chest of a simulated patient [13] (Fig. 4). When a stethoscope is placed on the seal, respiratory sounds can be heard.

Cardionics's simulator evolves into a suit with built-in seal. A simulated patient wears this suit and exercises [14] (Fig. 5).

However, existing simulators are extremely expensive, ranging from one million yen to several million yen. Also, the cost is high because the seal is disposable. It is necessary to prepare a suit for each simulated patient, and extra expenses such as washing are required.

The system we developed is designed only with KINECT and PC controlling it. Therefore, simulated patients can sit on a chair and express abnormal respiratory sounds, just like real patients, by breathing normally. Moreover, the system is relatively inexpensive [1] [2] (Figs. 6 and 7).

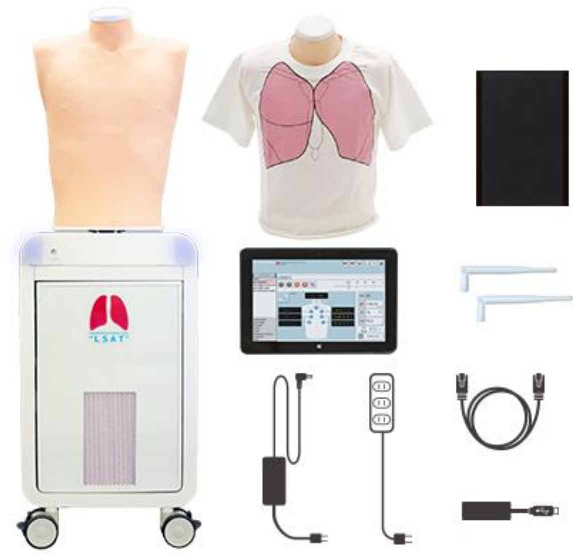


Figure 1. Breathing sound auscultation simulator "Lang." Kyoto kagaku.

Figure 2. “Chosin Kun,” Sakamoto Model Co., Ltd.,

Figure 3. ALS simulator, Laerdal Medical

Figure 4. Simscope-Wi-Fi-the-hybrid-simulator, Cardionics

Figure 5. SimShirt System, Cardionics

Figure 6. The KINECT-Based Auscultation Practice System (Face-to-face arrangement)

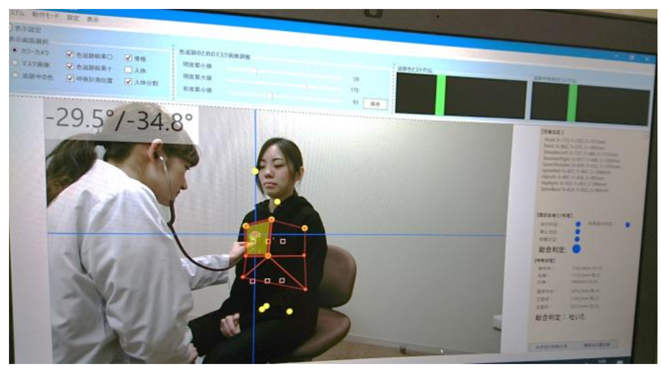


Figure 7. The KINECT-Based Auscultation Practice System (Screen of control computer)

III. METHOD

In this section, we discussed the auscultatory exercise practice method and research method.

A. Participants and data collection

The subjects were 91 first graders at the School of Nursing. The first training was held in July 2017. The students were previously instructed on respiratory anatomy and physiology; however, they had not yet practiced conducting physical assessments. The next training was held in December 2017 after they learned the physical assessment.

B. Practice method and evaluation

a) First training

We practiced distinguishing an accessory murmur from normal breath sounds using a system which we developed and a realistic, person-like simulator (Figs. 8 and 9).

The students broke up into groups of 15 to practice. At first, an upper-class student explained how to use the system and the person-like simulator, then explained the breath sounds auscultation method.

After practice, we questioned the students using a Likert-like scale. The questions included:

- Did you develop the ability to distinguish breath sounds?
- Were you interested in a system and person-like simulator?
- Did this exercise motivate you to learn nursing?

Each student had an additional free response option for recording his or her impressions.

b) The second training

The second training focused on distinguishing accessory murmurs using the new system and existing breathing sound data. After listening to the breath sounds, the nursing students labeled the type of breath sounds using clickers.

Figure 8. Practice with the newly-developed system

Figure 9. Practice using the person-like simulator

C. Statistical analysis

We compared the results following use of the person-like simulator with the subsequently-developed system using the Mann-Whitney U test. In the test to distinguish the type of breath sounds, we compared the first and second training results using the McManey test. Using the chi-square test, we compared the existing data with the test of the breathing sound of the newly-developed system. We used the SPSS Ver.22 statistical software program for all analyses. The level of significance was set at 5%.

For the free description responses, we organized the various responses around several response categories.

IV. Result

In this section, the effect of simulation training and the result of distinguishing breathing sounds were mentioned.

A. Participant and Questionnaire Responses

Response questionnaires were obtained from 84 (92.3%) students. Ultimately, 78 (85.7%) of those agreed to participate in the study.

All students expressed interest ("interested" – 74.4%; "moderately interested" – 25.6%) in the newly-developed system (Table 1). The mean ± standard deviation was 3.74 ± 0.44. Additionally, all students expressed interest ("interested" - 85.9%; "moderately interested" - 14.1%) in the person-like simulator (Table Ⅰ). The mean ± standard deviation was 3.86 ± 0.35. There was no significant difference in the level of interest expressed, with regard to the developed system and the humanoid simulator.

When asked if the newly-developed system motivated the students to learn nursing, 71.8% responded "agree." Additional responses included "moderately agree" (25.6%), "moderately disagree" (1.3 %), and "disagree" (1.3%; Table Ⅱ). The mean ± standard deviation was 3.68 ± 0.57. When asked if the person-like simulator motivated them to learn nursing, responses included "agree" (87.2%), "moderately agree" (11.5%), and "moderately disagree" (1.3%; Table Ⅱ). The mean ± standard deviation was 3.86 ± 0.39.

TABLE I. STUDENT'S INTEREST IN THE DEVELOPED SYSTEM AND HUMANOID SIMULATORABLE (%)

Scale	*Newly-developed system*	*Person-like simulator*
Yes :4	74.4	85.9
3	25.6	14.1
2	0	0
No: 1	0	0

TABLE II. STUDENT MOTIVATION ATTRIBUTABLE THE NEWLY-DEVELOPED SYSTEM AND HUMANOID SIMULATOR (%)

Scale	*Newly-developed system*	*Person-like simulator*
Agree :4	71.8	87.2
3	25.6	11.5
2	1.3	1.3
Disagree :1	1.3	0

TABLE III. COMPARISON OF THE NEWLY-DEVELOPED SYSTEM AND THE PERSON-LIKE SIMULATOR

	Interest	*Motivation*	*Distinction of the respiratory sound*
Newly-developed system	3.74	3.68*	3.05
Person-like simulator	3.86	3.86*	2.95
p value	0.072	0.018	0.421

TABLE IV. DISTINGUISHING BETWEEN NORMAL BREATH SOUNDS AND ACCESSORY MURMURS

Scale	*Newly-developed System*	*Person-like simulator*
Distinguishable :4	23.1	19.2
3	60.3	60.3
2	15.4	16.7
Indistinguishable :1	1.3	3.8

With regard to the motivation to learn nursing, there was a significant difference between the newly-developed system and the person-like simulator (Table Ⅲ).

In the exercise using the developed system, we asked if the students could distinguish between normal breath sounds and accessory murmurs. The students responded that the differences were "distinguishable" (23.1%), "moderately distinguishable" (60.3%), "moderately indistinguishable" (15.4%), and "indistinguishable" (1.3%; Table Ⅳ). The mean ± standard deviation was 3.05% ± 0.66.

In the exercise using the person-like simulator, we asked if they could distinguish between normal breath sounds and accessory murmurs. The students responded that the differences were "distinguishable" (19.2%), "moderately distinguishable" (60.3%), "moderately indistinguishable" (16.7%), and "indistinguishable" (3.8%; Table Ⅳ). The mean ± standard deviation was 2.95±0.72. There was no significant difference in the reported ability to distinguish between respiratory sounds when comparing the newly-developed system and the person-like simulator.

TABLE V. FREE RESPONSE CATEGORIES

Category title	*Newly-developed system*	*Person-like simulator*
Communication	17	7
Learning abnormalities	17	8
Interested in equipment	9	2
Reproduction of the clinical site	14	11
Motivation to learn	15	39
Think as human	0	6
Total	72	73

B. Free response field

We compared the thematic categories extracted from the free response field for both the newly-developed system and the person-like simulator (Table Ⅴ).

【Can learn communication】 included the following items: “I can learn how to respond to patients”; “I can learn while taking communication with before and after auscultation”; “I knew how to talk to patients”; and “I was able to learn how to attend to breath sounds and conversation.”

There were 17 students who responded that “I can learn communication” using the newly-developed system. There were 7 students who expressed the same believe for the person-like simulator.

【Ability to learn abnormal breath sounds】 included the following items: “Learn the difference between murmur and normal breath sounds”; “Can learn care for abnormality.” There were 17 students who responded that “I can learn abnormal breath sounds” using the newly-developed system and 8 students with the same response to the person-like simulator.

【Reproduction of the clinical situation】 includes the following items. "I can learn skills that resemble practical nursing"; "I can learn how to avoid causing the patient discomfort" and; "The respiratory sounds I heard were realistic." Fourteen students felt the newly-developed system accurately reproduced.

There were 14 students who responded that the newly-developed system accurately reproduced clinical situations and 11 students who had the same impression of the person-like simulator.

【Motivation to learn】 included the following items. "I learned about advanced nursing care"; "I want to learn more about nursing"; "I felt motivated"; and "I am interested in nursing.”

Fifteen students felt that the newly-developed system motivated them to learn. There were 39 students who responded that the person-like simulator motivated them to learn.

Free responses for the person-like simulator included, “I was able to practice thinking as an actual human” In addition, "There is a need to think humanoid simulator as human"; "There are similarities between people, but they cannot actually speak, there is no real person's weight."

Improvements on the developed system were proposed by six students. Critiques included: "The five ranges varied depending on the posture of the patient"; "It was difficult to react"; "The breath sounds were more realistic if you hear them through a stethoscope"; "The sound felt small"; "The sound was difficult to hear "; and "I thought that it would sound smooth if breath sensing improved. "

C. *Determining the type of respiratory sound*

Nursing students received an examination using existing data and the newly-developed system to examine their abilities to identify respiratory sound types. Following the first and second trainings, we compared the existing breath sound data with that obtained using the newly-developed system.

In testing the existing sound files, the students were able to accurately distinguish among the various kinds of respiratory sounds. Fourth breathing sounds (normal breath sounds, wheezes, coarse crackles, rhonchi) were the most frequent answers in the second test (Table Ⅵ).

In testing the newly-developed system, the students were able to accurately distinguish among the various breath sounds, with the exception of wheezes. Following the second training, the ability to distinguish between coarse crackles and normal respiratory sounds had improved.

However, the ability to distinguish between fine crackles and rhonchi was somewhat diminished following the second training (Table Ⅶ).

Following the first training, the number of correct answers for coarse crackles and normal respiratory sounds was significantly better when existing breath sound files were used. Using the newly-developed system, wheezes and rhonchi had significantly more correct answers (Table Ⅷ).

Following the second training, correct responses to three kinds of respiratory sounds improved significantly more with use of the existing breath sound files, compared to the newly-developed system (Table Ⅸ).

TABLE VI. CROSSTABULATION OF THE NUMBER OF TESTS AND THE NUMBER OF CORRECT ANSWERS (EXISTING SOUND FILES)

Coarse crackles	After second training			
After initial training	*FALSE*	*TRUE*	*p*	
FALSE	3	24	<.001	**
TRUE	2	61		

Fine crackles	After second training			
After initial training	*FALSE*	*TRUE*	*p*	
FALSE	4	3	<.001	**
TRUE	27	56		

Wheezes	After second training			
After initial training	*FALSE*	*TRUE*	*p*	
FALSE	4	75	<.001	**
TRUE	0	11		

Rhonchi	After second training			
After initial training	*FALSE*	*TRUE*	*p*	
FALSE	1	12	0.003	**
TRUE	1	76		

Normal respiratory sounds	After second training			
After initial training	*FALSE*	*TRUE*	*p*	
FALSE	3	35	<.001	**
TRUE	2	50		

**= p < .01.

TABLE VII. CROSSTABULATION OF THE NUMBER OF TESTS AND THE NUMBER OF CORRECT ANSWERS (NEWLY-DEVELOPED SYSTEM)

Coarse crackles	After second training			
After initial training	*FALSE*	*TRUE*	*p*	
FALSE	7	33	<.001	**
TRUE	1	49		

Fine crackles	After second training			
After initial training	*FALSE*	*TRUE*	*p*	
FALSE	9	4	<.001	**
TRUE	44	18		

Wheezes / After initial training	After second training *FALSE*	*TRUE*	*p*	
FALSE	0	0	.063	
TRUE	5	85		

Rhonchi / After initial training	After second training *FALSE*	*TRUE*	*p*	
FALSE	1	4	.012	*
TRUE	16	69		

Normal respiratory sounds / After initial training	After second training *FALSE*	*TRUE*	*p*	
FALSE	12	13	.002	**
TRUE	1	6		

**= p < .01. *= p < .05

TABLE VIII. COMPARISON OF EXISTING BREATH SOUND FILE DATA AND THE NEWLY-DEVELOPED SYSTEM (FOLLOWING INITIAL TRAINING)

Coarse crackles / The number of tests	The true-false test FALSE	TRUE	χ^2	
Existing breath sound files	27	63	4.02	*
	(-6.5)	(6.5)		
Newly-developed system	40	50		
	(6.5)	(-6.5)		

Wheezes / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2	
Existing breath sound files	79	11	140.79	**
	(39.5)	(-39.5)		
Newly-developed system	0	90		
	(-39.5)	(39.5)		

Normal / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2	
Existing breath sound files	38	52	11.87	**
	(-11.5)	(11.5)		
Newly-developed system	61	29		
	(11.5)	(-11.5)		

Rhonchi / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2	
Existing breath sound files	13	77	3.95	*
	(4.0)	(-4.0)		
Newly-developed system	5	85		
	(-4.0)	(4.0)		

Fine crackles / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2
The existing data	7	83	3.15
	(-3.7)	(3.7)	
the development system	13	65	
	(3.7)	(-3.7)	

**= p < .01.
Adjusted standardized residuals appear in parentheses below group frequencies.

TABLE IX. COMPARISON OF EXISTING BREATH SOUND FILE DATA AND THE NEWLY-DEVELOPED SYSTEM (FOLLOWING THE SECOND TRAINING)

Coarse crackles / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2	
Existing breath sound files	5	85	0.75	
	(-1.5)	(1.5)		
Newly-developed system	8	82		
	(1.5)	(-1.5)		

Wheezes / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2	
Existing breath sound files	4	86	0.12	
	(-0.5)	(0.5)		
Newly-developed system	5	85		
	(0.5)	(-0.5)		

Normal / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2	
Existing breath sound files	5	85	23.08	**
	(-8.3)	(8.3)		
Newly-developed system	13	19		
	(8.3)	(-8.3)		

Rhonchi / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2	
Existing breath sound files	2	88	13.24	**
	(-7.5)	(7.5)		
Newly-developed system	17	73		
	(7.5)	(-7.5)		

Fine crackles / The number of tests	The true-false test *FALSE*	*TRUE*	χ^2	
Existing breath sound files	31	59	25.61	**
	(-16.8)	(16.8)		
Newly-developed system	63	24		
	(16.8)	(-16.8)		

**= p < .01.
Adjusted standardized residuals appear in parentheses below group frequencies.

V. DISCUSSION

In our previous research, 50 clinical nurses felt that the newly-developed system was useful for learning to distinguish respiratory sounds [2].

In other words, this system is effective for 95.8% of nurses to learn by students, 87.8% is effective for nurses to learn. In this research, we compared the newly-developed system with a person-like simulator, which has been used conventionally.

Comparing the two systems, we found a significant difference in students' motivation to learn nursing. They appeared to be more motivating because the heart sounds and pulse were measurable, and the output of the electrocardiogram could be confirmed via monitor. Future development of the system should focus on enabling both respiratory and heart sounds. In addition, the result of the developed system is considered to be caused by students acting as simulated patients. The response of students and educators to high fidelity patient simulation has been extremely positive [15]. In the future, we plan to conduct training using simulated patients, rather than students.

There was also no significant difference between the newly-developed system and the person-like simulator for facilitating the ability to distinguish between normal respiratory sounds and accessory murmurs. An auscultatory learning equivalent to the person-like simulator is possible using the newly-developed system.

However, students made many mistakes when attempting to distinguish among detailed respiratory sounds. In the second auscultatory test of fine crackles, many students answered incorrectly, both when using existing sound data files and the newly-developed system. These mistakes appear to be attributable to hearing fine crackles only during inhalation and mistaking them as coarse crackles. We also did not observe improved accuracy at detecting rhonchi using the newly-developed system. The previous study suggested that highly contextualized learning environments may not be uniformly advantageous for instruction and may lead to ineffective learning by increasing extraneous cognitive loading in novice learners [16]. In addition, respiratory sounds were perceived as "difficult to hear." Other students reported, "We thought that it would sound smooth if breath sensing improved." We think that it is caused by the delay incurred when switching between inspiration and expiration. According to our study, the detection delay of respiration in KINECT v2 is 1.47-0.17 seconds for inspiration and 1.21-0.37 seconds for expiration [1] [2]. Under present circumstances, it is difficult to desire further detection capabilities. Therefore, respiratory sounds were not synchronized with breathing. After detecting the inspiration and reproducing the sound of inspiration, exhalation should be reproduced continuously.

Next, we discuss the issues clarified from student free comments.

–*"The five ranges varied depending on the posture of the patient" and "It was difficult to react."*

Although the installation angle and the height of KINECT are constant, the position of the chair shifts when students practice one after another. Measures that can mark the position of the simulated patient and the nurse's chair are needed.

– *"It sounds more realistic if you hear it through a stethoscope" and "The sound felt small";*

This was likely caused by using a speaker connected by Bluetooth so that respiratory sounds could be heard by simulated patients and students other than those playing the role of nurse. The stethoscope auditory resolution can likely be improved by using a wireless earphone.

Also, the newly-developed system reproduces respiratory sounds synchronized with respiration of a simulated patient. Therefore, it would be unsuitable for use in scenarios such as cardiopulmonary resuscitation of patients with no response. However, the ability to communicate is an important ability for medical personnel such as nurses and doctors and requires entraining early on in clinical education.

From the free comments of this survey, it was revealed that the developed system is useful 【for learning abnormalities】 while taking 【communication】 that 【reproduction of a clinical situation】.

For example, in a scenario where communication between medical personnel and patients is essential, we believe that the effectiveness of this system will be evident. Some examples may include:

- ✧ Physical assessment of patients with convalescent pneumonia
- ✧ Guidance for discharge from the hospital
- ✧ Cases requiring physical assessment and medication guidance for elderly patients
- ✧ A case of a febrile home care patient

VI. EXAMPLE OF A SIMULATION EXERCISE USING THE NEWLY-DEVELOPED SYSTEM

We introduced a simulation-based on a scenario of a patient with pneumonia before clinical practice.

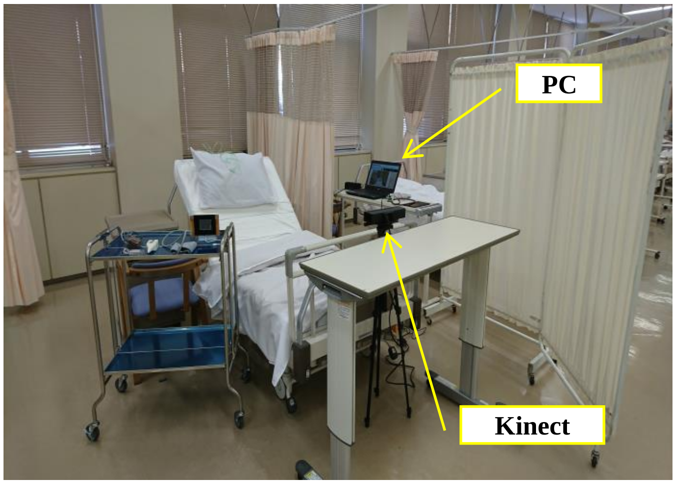


Figure 9. Simulation equipment arrangement of pneumonia patients

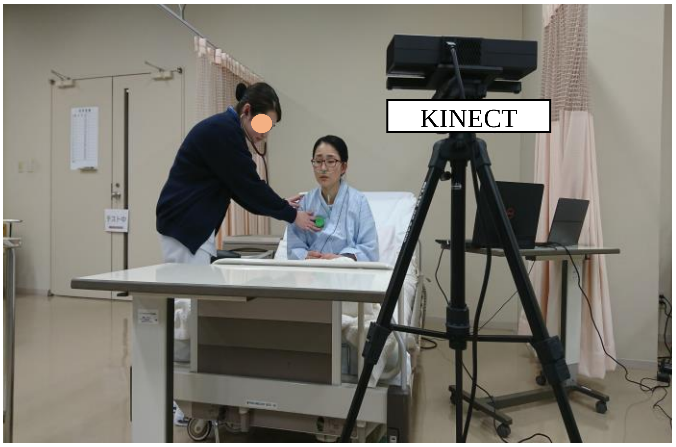


Figure 10. Simulation of respiratory sound auscultation

A KINECT was placed on the foot side of the bed with the head lifted (Fig. 9). This KINECT placement is in front of the patient, so the thorax can be sensed without distortion. Therefore, when the stethoscope contacts the thorax, it can detect respiratory sounds in the correct position. Also, since students approach from the side of the bed, they will not block the KINECT's detection field, which is a good arrangement. To assist students with concentrating on the simulation, a personal computer was placed behind the curtain. Nasal cannulas for oxygen therapy and devices for vital signs measurement were also available.

Patients with pneumonia generally receive oxygen therapy using a nasal cannula (Fig. 10). As advance preparation, the patient's breath sound sets featured accessory murmurs (coarse crackles or wheezes, rhonchi) in one or both lungs. This sound reproduces respiration that is affected by sputum over-production. The simulated patient wears typical clothing and sits on the bed.

The simulation training time was 10 minutes. First year nursing students entered the hospital room and greeted the patient. After that, they checked the nasal cannula, performed auscultation of respiratory sounds, and measured vital signs. The students asked the patient about signs and symptoms of dyspnea, the presence of sputum, and perceived pain. In addition, the students communicated with the patient to get a comprehensive sense of the patient's complaints. After the simulation, the students reviewed the session with the group members using the checklist. Debriefing time was carried out in 15 minutes using a 3-Phase Conversation Structures like the GAS method [17]. During debriefing, the students discussed what they were doing as nurses and how to improve, as a group. The simulation and debriefing were repeated three times. In these scenarios, the patient's respiratory state gradually worsened little by little. The patient also developed a medical device-related pressure ulcer behind the ear. Medical device-related pressure ulcers of the ear, results from contact with oxygen tubing, is included in the scenario for additional training [18].

TABLE X. CONTENTS OF SIMULATION TRAINING

Time allocation	Details of contents	
15 minutes	Advance preparation	・Kinect, PC, nasal cannula, devices for vital sign measurement, extension cords, clothing of patient
10 minutes	Briefing	・Sharing study objectives ・Guidance on using the equipment
10 minutes	Training 1	・The patient does not complain of breathing difficulties although some sputum is evident. ・One student plays the role of a nurse. ・The other students observe based on the checklist.
15 minutes	Debriefing 1	・Using GAS method
10 minutes	Training 2	・The patient has breathing difficulty because of cough and phlegm.
15 minutes	Debriefing 2	・Using GAS method
10 minutes	Training 3	・The patient has breathing difficulty because of cough and phlegm. ・The patient develops a medical device-related pressure ulcer behind the ear.
15 minutes	Debriefing 3	・Using the GAS method
10 minutes	Self-evaluation Summary	

Nursing students were thereby trained to examine worsening respiratory conditions and determine necessary care.

VI. CONCLUSION

Our group developed a system for practicing breath sounds auscultation using KINECT. This system solves several problems associated with past simulation education models. The system is inexpensive, and simply operated. We evaluated the process of learning breath sounds auscultation with a nurse and a nursing student. In this paper, we introduced practical exercises, using the KINECT system. The newly-developed system was equally useful as a person-like simulator for assisting students with developing their ability to distinguish normal breathing sounds and accessory murmurs.

In addition, the system proved useful for learning anomalies while communicating in an environment that accurately reproduced a clinical setting. Additionally, the system was motivating for nursing students.

Future efforts should address environmental settings, including the system, improving the stethoscope, and the timing of the respiratory sound reproduction in order to enhance training reproducibility within the simulated clinical setting.

Acknowledgment

This research was conducted using research expenses of Iwate Prefectural University (Iwate Monodzukuri · Software Integration Technology Center).

References

[1] Y. Murata, K. Yoshida, and N. Miura, "Demonstration of The KINECT-Based Auscultation Practice System," in *eTELEMED 2017 : The Ninth International Conference on eHealth, Telemedicine, and Social Medicine*, 2017, pp. 99-100.

[2] Y. Murata, K. Yoshida, N. Miura, and Y. Endo, "KINECT-Based Auscultation Practice System," Journal on Advances in Life Sciences, vol. 8, no. 3 &4 pp. 277-288, 2016.

[3] Ministry of Health Labor and Welfare. (2011). *Investigative Commission report regarding enhancement of nursing basic education*. Available: http://www.mhlw.go.jp/shingi/2007/04/dl/s0420-13.pdf 2018.05.26

[4] Y. Endo, N. Miura, H. Chieda, and T. Takeda, "A trial of using simulation in undergraduate nursing education," Journal of the Faculty of Nursing, Iwate Prefectural University, vol. 16, pp. 43-46, 2014.

[5] Y. Abe, "Simulation education in medical care," Journal of the Japanese Society of Intensive Care Medicine, vol. 23, pp. 13-20, 2016.

[6] R. Takashiki, K. Abe, C. Saito, and T. Horiguchi, "Effect of simulation education for operation department nurses to cope with cases of intubation difficulty," Journal of Japanese Association for Operating Room Technology, vol. 37, no. 2, pp. 39-41, 2016.

[7] M. Flentje, T. Müßel, B. Henzel, and J.-P. Jantzen, "Simulating a patient's fall as a means to improve routine communication: Joint training for nursing and fifth-year medical students," GMS Journal for Medical Education, vol. 33, no. 2, 2016.

[8] Ministry of Health Labor and Welfare. (2011, 2017.07.27). *Investigative Commission report regarding contents and methods of nursing education*. Available: www.mhlw.go.jp/stf/houdou/2r98520000013l0q-att/2r98520000013l4m.pdf 2018.05.26

[9] N. Coyle *et al.*, "Discussing Death, Dying, and End-of-Life Goals of Care: A Communication Skills Training Module for Oncology Nurses," Clinical journal of oncology nursing, vol. 19, no. 6, pp. 697-702, 2015.

[10] Kyoto Kagaku Co. Ltd. *Breath sound auscultation simulator*. Available: http://www.kyotokagaku.com/jp/educational/products/detail01/mw28.html 2018.05.26

[11] Sakamoto Model Co. Ltd. *Chosin Kun*. Available: http://www.sakamoto-model.co.jp/product/physical/m164/, 2018.05.26

[12] Laerdal Co. Ltd. *ALS Simulator* Available: http://www.laerdal.com/jp/ALS 2018.05.26

[13] Cardionic Co. Ltd. (2017.08.02). *Simscope-wifi-hybrid-simulator*. Available: https://www.cardionics.com/product/learning-systems/simscope-wifi-the-hybrid-simulator 2018.05.26

[14] Cardionic Co. Ltd. *SimShirt System*. Available: https://www.cardionics.com/product/learning-systems/simshirt-system 2018.05.26

[15] J. A. Gordon, W. M. Wilkerson, D. W. Shaffer, and E. G. Armstrong, "“Practicing” Medicine without Risk: Students' and Educators' Responses to High-fidelity Patient Simulation," Academic Medicine, vol. 76, no. 5, pp. 469-472, 2001.

[16] R. Chen, L. E. Grierson, and G. R. Norman, "Evaluating the impact of high- and low-fidelity instruction in the development of auscultation skills," Medical education, vol. 49, no. 3, pp. 276-85, Mar 2015.

[17] T. Sawyer, W. Eppich, M. Brett-Fleegler, V. Grant, and A. Cheng, "More Than One Way to Debrief: A Critical Review of Healthcare Simulation Debriefing Methods," Simulation in Healthcare, vol. 11, no. 3, pp. 209-217, 2016.

[18] M. Barakat-Johnson, C. Barnett, T. Wand, and K. White, "Medical device-related pressure injuries: An exploratory descriptive study in an acute tertiary hospital in Australia," (in eng), Journal of tissue viability, vol. 26, no. 4, pp. 246-253, Nov 2017.