Interactive Multimodal Social Robot for Improving Quality of Care of Elderly in Australian Nursing Homes

Rajiv Khosla
Research Centre for Computers, Communication and Social Innovation (RECCSI)
La Trobe University, Vic. 3086, Australia
+61 3 94793064
r.khosla@latrobe.edu.au

Mei-Tai Chu
Research Centre for Computers, Communication and Social Innovation (RECCSI)
La Trobe University, Vic. 3086, Australia
+61 3 9479 1259
m.chu@latrobe.edu.au

Reza Kachouie
Research Centre for Computers, Communication and Social Innovation (RECCSI)
La Trobe University, Vic. 3086, Australia
+61 3 9479 5310
rkachouie@students.latrobe.edu.au

Keiji Yamada
C&C Innovation Initiative
NEC Corporation
8916-47 Takayama-cho, Ikoma-Shi, Nara, 630-0101, Japan
kg-yamada@cp.jp.nec.com

Fujita Yoshihiro
C&C Innovation Initiative
NEC Corporation
8916-47 Takayama-cho, Ikoma-Shi, Nara, 630-0101, Japan

Tomoharu Yamaguchi
C&C Innovation Initiative
NEC Corporation
8916-47 Takayama-cho, Ikoma-Shi, Nara, 630-0101, Japan

ABSTRACT
This paper describes the design of multimodal robotic system, embodiment of multimodal interaction (voice, gestures, emotion, touch panel and dance) in assistive social robot (Matilda) for modeling group based and one-to-one interactions with technology adverse elderly in nursing homes. It describes the human-centered evaluation of Matilda based on quality, naturalness, user satisfaction and predictive accuracy based on first ever field trials in Australia. The multimodal social robots have facilitated breaking the technology barriers with the elderly leading to several aged care facilities and community centers showing interest in future trials.

Categories and Subject Descriptors
I.2.9 [**Artificial Intelligence**]: Robotics – *Commercial robots and applications*

General Terms
Human Factors, Design, Experimentation

Keywords
Assistive social robot, multimodal interaction, dementia, nursing homes

1. INTRODUCTION
Like most of the developed countries, Australia's population is ageing. According to Australia Bureau of statistics, in 1970 there were nearly 1.1 million people (8.3 percent) in Australia over 65. It increased to 3.01 million people (13.5 percent) in 2010. It is estimated that by a steady increase in 2050 around 8.1 million (22.7 percent) of Australian population to be 65 and over [1]. Over the next several decades, the need for aged care services is expected to grow at the rate of 68 percent and supply of health care workers is expected to grow at the rate of 14.8 percent, housing and demand for skilled labor [2]. On the other hand, a number of researches point out that in general elderly have difficulty in learning to use and operate new technologies. Thus, an important task in the field of gerontology is to develop tools that can be used easily by elderly. One solution is using multimedia and multimodal systems, which can be perceived as social entities.

In this paper the authors report on embodiment of multimodal attributes in an assistive social robot, named Matilda to support elderly in Australian nursing homes. Our contributions include a) use of human-like robot based multimodal interaction to break technology barriers between elderly and Matilda; b) improving efficiency and effectiveness of social activity (diversion therapy) programs in nursing homes; c) improving personalization of care in nursing homes.

2. MULTI-MODAL ROBOT SYSTEM ARCHITECTURE FOR AGED CARE
Multimodal Human Computer Interaction (MMHCI) lies at the crossroads of several research areas including computer vision, psychology, artificial intelligence, and many others [3]. On the other hand, telehealth and e-health research [4-7] besides collecting and integrating data has focused on a person-centered approach by using nurse avatars for improving the socio-emotional communication channel between the care receiver and care giver. One of the limitations of this approach has been that the present generation of elderly in nursing homes is not tuned into the concept of avatars and technology in general. In designing assistive technologies for elderly, particularly mentally or physically impaired, since they have limited expert knowledge of the technology, it should be able to interact in a natural manner. Because of that, they still prefer embodiment of care in physical form. This falls under the emerging area of healthcare robotics.

MM'12, October 29–November 2, 2012, Nara, Japan.

Our research that falls in the area of assistive robotics has attempted to marry person-centered approach and the perception of physical embodiment of care with artificial intelligence, soft computing and computer vision techniques in design of proactive emotionally intelligent systems for improving mental and physical well being of the elderly. Matilda, the multi-modal socially innovative robot, is illustrated in Figure 1. As shown it is ~39 cm tall and weighs 6 kg. It also comes with an intelligent docking station for automatic parking and recharge of its battery. Matilda is embodied with a range of multimodal attributes, e.g., voice vocalization and recognition, gestures, music, dance, emotions, motion and movement, face tracking and registration, touch sensors. Applications of Matilda in aged care nursing homes involve a combination of modalities like voice, gestures, music, dance, emotions, motion and movement. Its wireless network connectivity allows it to be integrated with other devices and cloud computing infrastructure thus seamlessly becoming a part of a unified communication network.

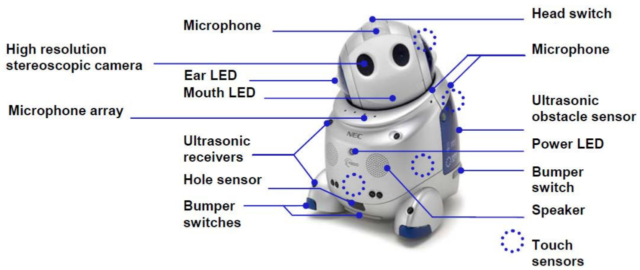

Figure 1- Matilda and its specifications

The connectivity and unified communication between a range of Information and communication technologies (e.g., Internet, mobile phones, PCs, robots) facilitates its scalability and ability to harness other technological resources of an organization, families and networks. The social robot and its various components are shown in Figure 2.

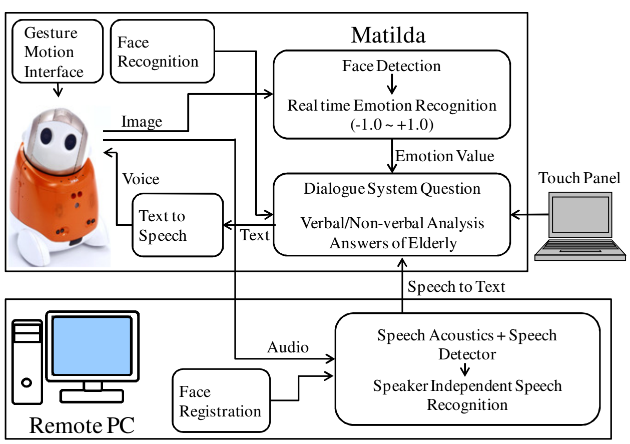

Figure 2- Assistive Social Robot and Components

It shows a shared intelligence architecture where speech recognition, face registration and touch panel are remotely connected to Matilda through its wireless network.

3. Application in Nursing Homes

Two applications related to social engagement and healthy living are described. They represent group based and one-to-one interaction with Matilda.

3.1 Hoy game

In our research we studied the requirements of elderly from this perspective to facilitate acceptability of Matilda as part of their environment in nursing homes. A large majority of elderly wanted Matilda to participate in group activities and play games like Hoy (see Figure 3). Matilda's ability to mix calling of numbers with gestures (nod), expressions, music and dance brought in more variety, enthusiasm and wider range of social interaction among the elderly. This was measured through video recording of body posture, visual and voice interaction, facial expressions and independent use of touch panel communication modality with the robot by the participants to control the pace at which the cards are called and displayed by the robot.

Matilda's ability to simultaneously call and visually display the called cards freed up the caregiver to personalize care (given that participating residents suffer from range of disabilities including dementia, depression, deafness), attend to each one of them, and help them win the game.

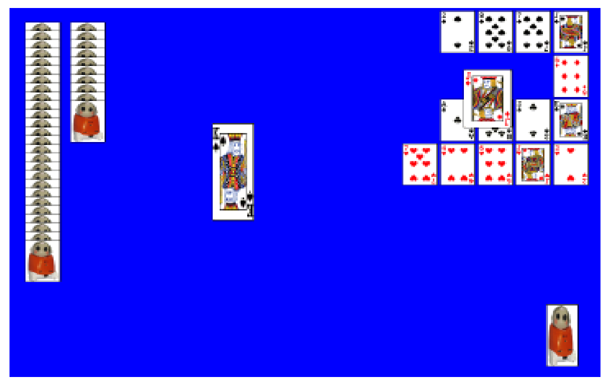

Figure 3- Hoy Game Displaying the Cards Called by Matilda

The recorded videos clearly demonstrate that residents use different modes of communication at different times when they play the game. In the recorded video residents with dementia use the visual display as they cannot only see the number, which has been called, but also all the previous numbers which have been called. The visual display confirms/reminds them of the spoken information and enables them to interact in a more relaxed manner.

The different modes of communication also demonstrate that Matilda can engage residents with different medical conditions (e.g., deafness, dementia, short-term memory loss) more effectively than a health care worker can.

3.2 Persuasive Diet Improvement system

Existing systems for improving dietary habits use arguments to inform the users on possible consequences of their behavior, or argue against their arguments, and in this way try to persuade the user of the 'correct' behavior. These systems resist on considering sentiments and emotions expressed in communication and faces as in the multimodal social robot system described in this paper.

The diet improvement dialog system is based on trans-theoretical model of behavior change suggested by [8] and adopted by health practitioners. The dialogues are designed around five mental states of behavior change. These are Pre-contemplation, Contemplation, Preparation, Action and Maintenance. In addition to the mental state estimation, the dialog system integrates human emotion and speech with Matilda's facial expression, gestures, and voice vocalization for dialog adaptation. At the time of writing this paper, speech acoustics of the elderly is also being integrated for dialog adaptation.

A sample interaction is provided below.
Waiting for a remote connection...
Accepted connection from 127.0.0.1:2539
Matilda (M) - *Welcome to the Diet improvement dialog system*
M - How are you?
Elderly (E) - *I'm good*
Non-verbal Response (NVR) - emotion value: -0.999489
You looked negative
(This is an example of how the emotion detected from the elderly has affected the dialog)
M - *It's good to hear that you are well. It would be even better to see you smiling as well! Let's talk about healthy diet.*
M- *How many meals do you eat on an average day?*
E - About three
NVR- Got emotion value: 0.271914
M - *Do you like eating French fries and other fried foods?*
E - Yes
NVR- emotion value: -0.999999
M - *Do you eat many sweets?*
E - Sometimes
NVR - emotion value: 0 …..
M - *You indicated before that you sometimes like to eat sweets. It is good to limit the amount of sweets we eat.*
M- Now that you know some of the problems that may arise, would you consider reducing the amount of sweets you eat?
E- *Yes I would*
NVR- emotion value: -0.999609
The robot records emotional state changes (instead of emotional states) of the elderly while they are responding to the questions asked by the robot (Figure 4). In initial part of the interaction, the emotional state changes of the elderly are frequently negative (i.e., spikes below the line). Later on the emotions become gradually positive (i.e., spikes above the line).

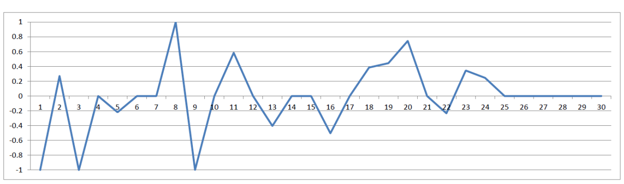

Figure 4- Emotional State Change of Elderly

4. HUMAN-CENTERED EVALUATION

The human-centered evaluation is based on constructs defined by Jaimes and Sebe [3].

4.1 Efficiency, Quality and Naturalness

Efficiency and Quality as measured by us through video recordings is described in this section.

Efficiency: The elderly (71-98 years old) involved in playing the Hoy game suffer from various disabilities. So efficiency is determined in terms of range of people with varying disabilities participating in the game and successfully identifying the called card in their Hoy (card) sheet without help from the care giver. In our video recordings only two people in a group of 15 elderly with disabilities needed some support from the care giver for coordinating the identification of visually displayed card on their card sheet.

Quality has been defined by nursing home management in terms how Matilda proves to be effective in context of stated and perceived expectations of the elderly. These expectations are described next.

Control: Elderly can play at their own pace using the touch panel to instruct Matilda when to call the next card. Prior to introduction of Matilda a caregiver controlled calling the cards.

Accounting for various disabilities: Elderly suffer from various disabilities including hearing and visual impairment, dementia (cognitive impairment), depression, etc. Matilda through voicing and visually displaying the cards as well as relinquishing control to the elderly (in terms of the pace at which the cards are called by allowing elderly to using to use the touch panel for sending the control commands) helps to improve effectiveness of elderly with various disabilities to play the game successfully. Prior to introduction of Matilda, only one care giver was employed for calling the cards thus limiting the effectiveness in terms of elderly with visual and cognitive impairments.

Personalization of Care: The video recording of elderly playing Hoy game with Matilda showed that the care givers were freed from the task of call the cards and were successfully able to help residents on one to one basis thus improving the quality of care for the nursing home.

Social Interaction: The use of Matilda's gestures (nod) and music accompanied by dance movements (after some one wins the game) was welcomed by the elderly and improved its acceptability and social interaction.

Naturalness: The physical embodiment of baby like face, human-like multimodal attributes facilitates natural interaction with the elderly who seem to forget they are actually interacting with technological device.

4.2 User Satisfaction Survey

Thirty-four residents from three nursing homes in Australian states of Queensland and Victoria participated in the field study and trial respectively. Low care and high care residents were involved in the trial. The residents involved also suffered from various medical conditions including depression, dementia, Parkinson disease, deafness short-term memory loss etc.

A field survey consisting of 19 questions was designed and administered to the elderly to elicit their feedbacks (degree of concurrence to each question) of 3 field trials of Matilda. Questionnaire is designed with the help of Director of Nursing and Nursing Home Managers who have more than 25+ years of experience in the aged care industry. Keeping the frail health of elderly (most of whom are 71-98 years old); the questions designed are short, and simple. Three questions were open ended, 4 questions about personal demographical state e.g. gender and age and for the remainder twelve questions respondents invited to indicate their feedback in each criteria based on a five-point Likert rating scale (Strongly disagree=1, disagree=2, Not sure=3, Agree=4, Strongly agree=5). SPSS software was used for analyzing the data along this research. Out of total 34 respondents, 28 (85 %) are female and only 6 (15 percent) are male. They had different medical conditions, some had dementia, and some suffered from depression. We conducted reliability analyses to check up the internal consistency of the instrument. Of twelve adapted items of 34 questionnaires, reliability analysis resulted to a Cronbach's alpha of 0.874. Descriptive statistics is shown in Table 1.

Table 1- Descriptive statistics of questions

Question	Mean	Std. deviation	Std. error mean
1. Do you enjoy the contact with Matilda?	4.26	.963	.165
2. Do you like to participate in group activity (e.g. bingo, hoy) with Matilda?	3.85	1.077	.185
3. Do you enjoy one to one activity (e.g., calendar, diet improvement) with Matilda?	3.55	1.227	.214
4. Did Matilda make you feel better (e.g. make you smile)?	4.12	.946	.162
5. Do you think Matilda can help you to make more friends?	3.21	1.122	.192
6. Do you feel concerned in the presence of Matilda?	1.94	1.434	.246
7. Do you like Matilda to be your friend?	3.85	1.034	.180
8. Do you like the way Matilda respond to you?	3.79	1.053	.183
9. Do you think Matilda can improve your daily life?	3.38	1.231	.211
10. Do you like to touch Matilda?	3.26	.931	.160
11. Are you comfortable with Matilda (e.g. dancing, blushing)?	4.26	.963	.165
12. Do you feel relaxed talking to Matilda?	4.03	1.262	.220

The standard error of the mean is the standard deviation of the sample mean and is an indication of how well the mean of the target population could be estimated by the mean of a sample. The meaning of a small standard error portrays that the majority of sample means are similar to the mean of population, thus it is expected that the population could be accurately described by the sample [9]. The standard error of all the means is close to zero signifying that the selected sample describes the population accurately. Moreover, the fact that the standard deviations are virtually 1.0 or less indicates that there is little variability in the gathered data among the respondents is observable [9].

4.3 Predictive accuracy

Matilda monitors the emotional state changes over time in order to determine the social attitude of elderly towards Matilda (during interaction) as well as to give it the ability to adapt its persuasive dialogs based on emotional state change feedback. The changes in facial features are tracked/recognized and mapped to positive and negative or neutral emotional states [10, 11] as shown in figure 4 (simplified diagram) thus overriding the inaccuracies associated with stereotypes states. The process of classifying human face expressions as positive, negative and neutral emotional state is not based on any absolute benchmark of happy, angry or sad stereotypical expressions. We only measure the relative change in the direction and magnitude of facial expression in order to determine the change and intensity of change of positive, negative or neural (no change) emotional state. A simplified diagram of mapping the facial action units to neutral, +ive and –ive emotional states is shown in Figure 5. It also captures the intensity of emotional states (high, medium and low) based on the magnitude of movement of facial features.

5. Conclusion

This paper reports on first-ever field trials of multimodal assistive social robot in 3 Australian nursing homes. It illustrates the success of the field trials through multimodal robotic system design, and effectiveness of multimodal interaction through group based and one-to-one interaction of elderly with the social robot. It evaluates the multimodal system through measures like quality, naturalness and user satisfaction. The social robots have proved as very effective device for improving the quality of care of elderly, breaking down the technology barriers, and bringing novelty in the social activities.

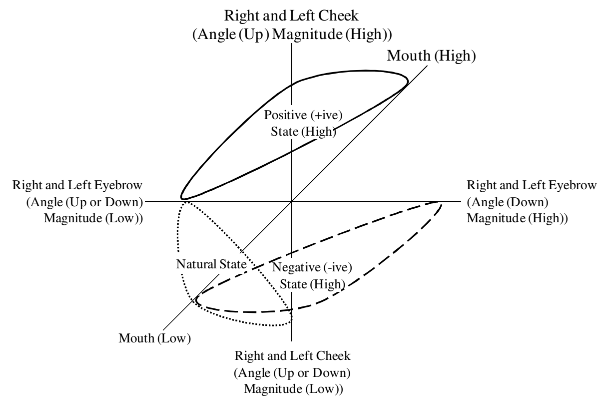

Figure 5- Visual Representation of Emotional State Change

REFERENCES

[1] Australian-Bureau-of-Statistics. "Population by Age and Sex, Regions of Australia (cat no. 3235.0)," .

[2] Australian-Bureau-of-Statistics. "Feature Article 1: Population by Age and Sex, Australian States and Territories ".

[3] A. Jaimes, and N. Sebe, "Multimodal human–computer interaction: A survey," *Computer Vision and Image Understanding,* vol. 108, no. 1–2, pp. 116-134, 2007.

[4] K. M. Buckley, B. Q. Tran, and C. M. Prandoni, "Receptiveness, use and acceptance of telehealth by caregivers of stroke patients in the home," *Online J Issues Nurs,* vol. 9, no. 3, pp. 9, 2004.

[5] C. Lisetti, and C. LeRouge, "Affective computing in tele-home health," 2004.

[6] C. Lisetti, F. Nasoz, C. LeRouge *et al.*, "Developing multimodal intelligent affective interfaces for tele-home health care," *International Journal of Human-Computer Studies,* vol. 59, no. 1-2, pp. 245-255, 2003.

[7] J. Naish, "21st century nurse," *Nursing standard (Royal College of Nursing (Great Britain): 1987),* vol. 24, no. 8, pp. 22, 2009.

[8] R. Cooper, S. Larsson, C. Matheson *et al.*, "Coding instructional dialogue for information states," *Deliverable D,* vol. 1, 1999.

[9] A. P. Field, *Discovering statistics using SPSS*: SAGE publications Ltd, 2009.

[10] P. Ekamn, and W. Friesen, "Facial action coding system (FACS): manual," Consulting Psychologists Press, 1978.

[11] R. W. Picard, *Affective computing*: The MIT press, 2000.