

The ENRICHME Project: Lessons Learnt from a First Interaction with the Elderly

Roxana Agrigoroaie, François Ferland, and Adriana Tapus

Robotics and Computer Vision Lab, U2IS, ENSTA-ParisTech,
828 bd des Marechaux, 91762 Palaiseau Cedex, France
`firstname.lastname@ensta-paristech.fr`

Abstract. The main purpose of the ENRICHME European project is to develop a socially assistive robot that can help the elderly, adapt to their needs, and has a natural behavior. In this paper, we present some of the lessons learnt from the first interaction between the robot and two elderly people from one partner care facility (LACE Housing Ltd, UK). The robot interacted with the two participants for almost one hour. A tremendous amount of sensory data was recorded from the multi-sensory system (i.e., audio data, RGB-D data, thermal images, and the data from the skeleton tracker) for better understanding the interaction, the needs, the reactions of the users, and the context. This data was processed offline. Before the interaction between the two elderly residents and the robot, a demo was shown to all the residents of the facility. The reactions of the residents were positive and they found the robot useful. The first lessons learnt from this interaction between Kompaï robot and the elderly are reported.

Keywords: human-robot interaction; elderly; social robots

1 Introduction

In the context of a worldwide aging population [9], the need to find solutions for helping the elderly and enhancing their quality of life arises. One of the solutions that received much attention in recent years is the use of social robots for personalized care. Social robots can be good companions for the elderly as they can provide monitoring [2], they can alert their families or caregivers in case of emergencies [8], and they can stimulate their cognitive functions [17, 19].

ENRICHME is a EU Horizon2020 project whose purpose is to develop a socially assistive robot for the elderly with mild cognitive impairment (MCI). The project aims at developing a robotic system that can help elderly to remain independent longer and enhance their everyday life. The system is designed to learn from past experiences as it contains an episodic-like memory [12]. Moreover, it should be able to adapt its behavior to the user’s profile. More specifically, it should take into consideration its personality, cognitive disability level, emotional state, and preferences.

In the past few years, many socially assistive robots have been developed [2, 18, 5, 13, 20]. However, the ENRICHME system uses both the robot and the



Fig. 1. The robot in one of the user’s home

intelligent environment equipped with sensors in order to get information on the context, the user, and the potential needs of the user. Furthermore, it is meant to be a personal robot that lives with its owner. This might be a first sustainable solution for a long-term interaction between a robot and the elderly in their own environment. During the testing phase, the robot will stay in the houses of the elderly for a period of 4 months. The total testing time for the platform is 1 year, with 9 robots in 3 testing sites across Europe (Italy, Poland, and Greece).

The current work summarizes some of the lessons learnt from a first interaction with the elderly from LACE Housing Ltd, an elderly care facility in the UK. This being the first interaction between the possible final users and the robot, only two elderly individuals were selected to interact with the robot and to test the modules that were created so far.

This paper is structured as follows. Section 2 presents a description of the robot and the interaction scenario. Section 3 describes how the data was recorded and analyzed. Section 4 gives the results from the interaction and lists lessons that were learnt, and finally Section 5 concludes the paper and offers a perspective on future work.

2 Interaction scenario

ENRICHME project uses the Kompaï robot designed by Robosoft¹ (one of the project’s partners). The robot is equipped with an ASUS XTion Pro RGB-D sensor mounted beneath its head, an Optris PI450 thermal camera mounted on the head, and different sensors for measuring environmental parameters (e.g. ambient temperature). The torso of the robot features a touchscreen to facilitate user interaction. The robot can also use speech recognition and synthesis to communicate.

For this first interaction, one of the use cases developed for the project has been implemented. In this use case, the activity sensors of the robot and the sensors located in the house detect that the person is restless, and the robot approaches the person to propose an activity for them to do. The possible activities are: playing a cognitive game, doing some physical exercises, or learning some healthy eating tips.

¹ www.robosoft.com

For this scenario, the robot uses a leg detection algorithm by using the laser range finder mounted on the robot [3] so as to track people in the room. Based on data from an online survey, where participants would define approach parameters (i.e., stopping distance, deceleration, and curvature of the path) according to the robot's personality, the robot was configured to approach the person in a submissive and friendly manner. The robot automatically approaches the closest person it detects, and keep itself oriented toward that person at a maximum distance of 1.0 m. Therefore, even if the person is moving in the room, the robot is able to continuously follow the person.

3 Interaction, data recording and analysis

A face recognition module based on [1] was trained with the images of the two residents. Therefore, the robot was able to recognize them and greet them by their name when they were in range. Next, a web-based graphical user interface (see Figure 2) was displayed on the touch screen of the robot. The font size used throughout the interface was between 4 mm (for the timer and some instructions) and 9 mm (used for titles).

The cognitive games available on the robot that were developed in collaboration with the medical staff from the care facility are as follows:

- Digit cancellation: At the beginning of the cognitive game, the user is given a random digit. The purpose of the game is to find all occurrences of that digit in a random list of 40 digits in less than 30 seconds;
- Puzzle: It contains a jig-saw puzzle with three difficulty levels: easy with 12 pieces, medium with 30 pieces, and hard with 54 pieces;
- Hangman: The user has to guess a word randomly selected from a list divided into five categories: fruits, vegetables, cities, objects in the house, and animals.
- Memory game: The user is shown a series of 4 images for 10 seconds. After being shown a blank screen for 5 seconds, a new image is shown, and the user has to say if the current image is part of the previous ones or not. The game consists of 5 rounds.

Feedback was given after each game, summarizing the user's performance based on the hints used, errors made, total game time.



Fig. 2. The developed interface with the potential activities

The physical exercises activity is made of three parts, each with a duration of 1 minute. In the first part, the robot uses only speech to tell the user what exercise to do. In the second part, the user is only shown images of the exercises they have to do. In the third part, the robot uses both speech and images for the exercises. The exercises were randomly chosen from a set of 8 possible arm exercises shown in Figure 3. These exercises were selected in collaboration with the physical therapist available in the care facility. We assure ourselves that the participants were able to perform these physical exercises. After the execution of the physical exercises, the user is informed about the number of times they correctly performed each exercise.

The third proposed activity is the health information tips. The robot presented the user 8 healthy eating information such as not skipping breakfast or eating less refined sugars [6]. At the end, the amount of information retained was tested with a short quiz.

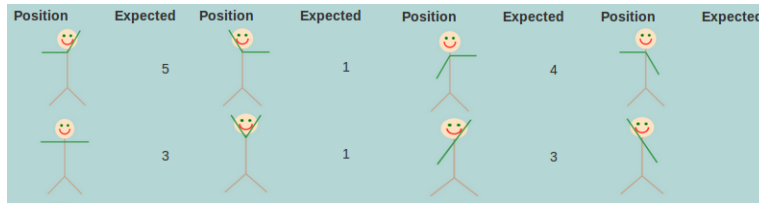


Fig. 3. The physical exercise feedback

After the interaction, each participant was given two questionnaires: a personality questionnaire, with 46 questions, based on the Big Five personality factors [7], and a custom made questionnaire, with 58 questions, developed for finding out what the participants liked or disliked about the graphical interface and the overall robot behavior. Each questionnaire took between 5 and 10 minutes to complete and consisted of yes/no questions or questions with a scale from 1 (strongly disagree) to 5 (strongly agree).

Throughout the interaction between each participant and the robot, the data from multiple sensors was recorded. For human robot interaction, the data sources used are the RGB-D data, the thermal image, the audio, and the data from the skeleton tracker. For the audio an external omni-directional microphone was used, and the skeleton tracker used the OpenNI NiTE library with the ASUS XTion Pro sensor. The data from the skeleton tracker was used for determining if the elderly executed correctly the physical exercises. Except for audio, the Robot Operating System (ROS) framework [15] was used for robot behavior coordination, data recording and analysis.

For compactness, the following notation is used: P_i for participant i (1 or 2), E_j for event j (1 to n), and P_iE_j for the j th event that was recorded for Participant i . The possible events are: performing the physical exercises, playing one of the cognitive games. Participant P1 is a 73 years old introverted male who suffers from Parkinson's disease and has cerebellar ataxia (the inability to coordinate balance, gait, extremity, and eye movements). Participant P2 is a 83 years old

extroverted female and has arthritis. Both participants started with the health tips, continued with the cognitive games, and ended with the physical exercises. The total recording times for P1 and P2 are 37 and 29 minutes, respectively. Out of the total interaction time, two events for each participant were chosen for further analysis. These events were chosen because both participants showed a strong reaction towards something they disliked or liked. Before the interaction started with the two participants, there was a short demo for all the residents of the care facility. One of the games that was tried during the demo was the puzzle game, which was found interesting, but difficult by most of the residents. The level of the game shown in the demo was the medium one, with 30 puzzle pieces. Just by seeing them they thought that they would not be able to complete the puzzle. When participant P1 saw the puzzle game among the possible cognitive games, there was a strong reaction: he found it too difficult and did not want to try it. Data during this reaction was chosen to be P1E1. P1E2 is represented by the last two minutes from the physical exercises activity. The first minute of the physical exercise was not chosen due to a lot of movement of participant P1. There was nothing shown on the screen during this time, therefore he turned towards the other people in the room.

Participant P2 firmly expressed a strong dislike towards quizzes, therefore at the announcement of having to complete a quiz after the healthy eating tips a strong reaction was observed. Therefore, this was considered as P2E1. P2E2 consisted of the participant playing the puzzle game. As P1E1 consisted of a strong reaction on a game, for P2 was extracted an event related to a game in order to see if the temperature variation is similar for both participants P1 and P2, respectively.

We wanted to investigate if facial temperature variations, facial expressions, and speech variations could be observed while the participants reacted to these events, and thus could be recognized by the robot in the future.

Participant P1 wore glasses throughout the interaction, during the memory game he changed his original glasses with another pair to better see the text on the screen. Participant P2 started the interaction without glasses, but during the puzzle game glasses were put on. During both interactions with the robot, other people were present in the room, and sometimes interacted with the participants. This resulted in situations such as a loss of facial temperature data because the participant turned his/her head, or voices other than the participants' being recorded by the external microphone. These events were not discouraged by the experimenters as it was believed that they would provide realistic information on situations that were certainly going to re-occur in future tests.

The following subsections describe the data that was recorded by each sensor.

3.1 Action Units

RGB-D data was used both for offline monitoring of the experiments and facial expression analysis. From the four events, facial action units (AU) [4] were extracted. The extracted AUs are: AU1 (inner brow raiser), AU2 (outer brow raiser), AU4 (brow lowerer), AU5 (upper lid raiser), AU6 (cheek raiser), AU7

(lid tightener), AU12 (lip corner puller), AU15 (lip corner depressor), and AU25 (lips part). Using these AUs, several emotions can be detected [14]: happiness (AU6 + AU12), sadness (AU1 + AU15, optionally AU4 and AU17), surprise (AU1 + AU2 + AU5 + AU25, or AU26).

3.2 Thermal data

Temperature was extracted from 6 regions of interest (see Figure 4): the forehead, the region around the eyes, the nose, the tip of the nose, the perinasal region, and the mouth. In post-analysis, the whole face was selected manually at the beginning of each event, and a correlation tracker based on facial landmarks implemented with dlib [11] was used to track the regions of interest.

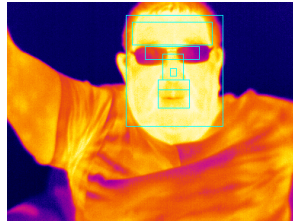


Fig. 4. The regions of interest shown for the first participant. Temperatures range from 20°C in dark purple to 40°C in light yellow.

When participants turned their heads to speak with other people present in the room, sudden drops in facial temperature would appear. These were caused by background pixels entering briefly the tracked regions of interest. Furthermore, as it can be seen from Figure 4 in the eyes region when the participant is wearing glasses, the temperature is very low. Therefore, readings associated with the glasses or the background have to be ignored. To do this, a histogram of the temperatures was plotted. It was observed that the histogram was trimodal, with a minimum between the last two modals at 33°C . The first modal represents the background temperature (with a mean of 26°C), while the second one represents the glasses (with a mean of 30°C). Therefore all temperatures lower than 33°C were discarded. The same consideration was made for the nose region, as part of that region also contains the glasses. For the other regions the same processing was applied. The histogram was bimodal with the minimum between the two modal at around 30°C . The first modal (with a mean of 27°C) represented the background, therefore all values below 30°C were discarded. The average temperature of each region for each frame was computed and then filtered using a low pass Butterworth filter. The sampling frequency for the data is 9Hz, and a cutoff frequency of 0.64 Hz was chosen ([21], equation(9)) so as to filter out the variations in the temperature that are due to the small movements of the head or of the camera. For a better understanding of the general trend of the temperature over time a linear least-squares regression was applied (e.g., the eyes region for P1: slope of 0.00034, regression intercept of 35.27 and correlation coefficient of 0.15).

4 Results and comments

4.1 Action Units

First, regarding action units, the most frequent AU present during all four events is AU4 (brow lowerer). This can be explained by the fact that the two participants complained that the font size was a little too small and they had to concentrate on reading what was written on the screen. The most frequent combination of two AUs is represented by AU4+AU15 for participant P1 and AU2+AU5 for participant P2. Happiness (AU6 + AU12)[14] was not detected at all, sadness (AU1 + AU15 + AU4) was detected for both participants (for participant P1 for both events, for participant P2 just during the game), and surprise (AU1 + AU2 + AU5 + AU25) was detected for participant P2 during the quiz.

There is a need for careful consideration when extracting AUs from elderly people mostly because of wrinkles that appear with age. Therefore, for some users even if AU15 is present it does not necessarily mean that the user has lowered the corners of the lips, it can be the neutral state for that person. The same consideration has to be applied when extracting emotions from the facial expressions. The data that was gathered during this experiment permitted the extraction of important information that can be used by the episodic memory module and to better understand the needs and the reactions of the users.

4.2 Thermal data

When looking at the raw thermal data there are some things that are clear from the very beginning. First of all, it can be easily observed that the presence of glasses produces a drop of temperature in the orbital region. However, this can be used to detect glasses and adapt the behavior of the robot accordingly. For example, if the user wants to play a cognitive game and the robot detects that the user has no glasses it can suggest wearing them based on previous information that the user sees and performs better when wearing them.

When having a closer look at the mouth temperature variation in correlation with the audio, and the RGB-D data, it can be detected when a user has spoken or kept their mouth open. For example, in Figure 5 all the variations are due to the participant speaking. Even if all temperatures that correspond to the background have been eliminated, at timestamp 00:28 there is still a small drop due to the turning of the person. This can appear due to measurement errors as the tracker does not follow properly the facial features. The region of interest can include other parts of the face (e.g. for the mouth it can also include parts of the cheeks or the chin). The thermal data can be used individually, but in correlation with the RGB-D and audio data it can provide more robust information to the robot and it can enable it to better change its behavior.

This information could help the robot better adapt its behavior in different situations (e.g., to not interrupt if the owner is speaking).

When comparing the two events in which both participants, P1 and P2, had a visible reaction to an unwanted event the following were observed. The duration

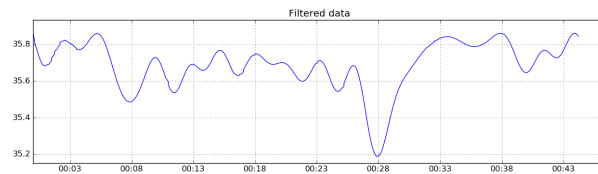


Fig. 5. Mouth temperature variation

of the event for participant P1 was of 25 seconds, while for participant P2 it was of 39 seconds. For participant P1, the temperature in all 6 regions increased between 0.05°C for the eyes region and up to 0.6°C for the tip of the nose. For participant P2, the temperature for the nose and the tip of the nose remained constant over time and it decreased for the other regions; between 0.01°C for the eyes and up to 0.5°C for the mouth region. According to [10] the temperature variation for participant P2 are indicators of mild stress, while for participant P1 the increase of temperature in the forehead region is an indicator of stress. These are some preliminary results. More tests will take place in the near future with more participants and for longer periods of time.

4.3 First lessons learnt and discussion

Some of the general lessons that were learnt from the interaction can be summarized as follows:

Technical considerations

- The position and the orientation of both RGB-D and thermal camera is very important as in some situations it can miss important information. The cameras should be able to cover the face of the person regardless if it is standing or sitting.
- There will be situations in which the user will not face the camera directly. In those situations the data should be discarded.
- Users might have glasses; the approach for extracting the periorbital temperature should take this into consideration as this region provides paramount information about the user.
- For a better detection of the facial landmarks, the RGB image should be used. With proper inter-camera calibration, coordinates of the facial features can be transposed from the RGB-D camera to the thermal camera. Thus, a smaller region of interest can be defined so other parts of the face are not included in case of sudden movements.

Discussion

Based on the questionnaire regarding their likes and dislikes, the two participants enjoyed playing the games (approval of 4.2/5) except the puzzle game

(approval of 1/5). They liked the layout, and the size of the buttons was large enough for them. One complaint was that the font size of 4 mm used for some instructions was too small. Another complaint was regarding the puzzle game. The participants and other residents found that 30 pieces of puzzle for medium level were too many. Therefore the number of pieces needs to be adjusted, 30 pieces of puzzle will be used for the hard level, as suggested by the participants. Based on the questionnaire regarding their thoughts on the overall behavior of the robot, both participants found the robot's approach behavior natural (approval of 5/5), in its approach it was neither dominant nor submissive (approval of 3/5), and appeared to be friendly (approval of 4.5/5). Participant P1 found the robot a little extroverted, while participant P2 found it neutral.

5 Conclusion

This paper presented some first lessons learnt from a first interaction between the ENRICHME robot and two residents in a care facility (Lace Housing Ltd in the United Kingdom). The robot interacted with the participants in their own homes. Events like turning their heads and talking to other people in the room were not discouraged by the experimenters. These events provide useful information on what kind of situations might occur in the long-term testing of the system. From the results of this work, we can posit that the facial temperature variations and facial expressions variations can be observed. The robot could adapt its behavior based on these variations to provide a more natural interaction with the elderly. For example if during a game the robot detects a high stress level [16] it can adjust the difficulty level of the game.

Based on the information gathered in this first interaction, changes will be applied to the system before starting the next phase of the project, a preliminary test with 30 participants in two Ambient Assisted Living (AAL) laboratories in the Netherlands (Stichting Smart Homes) and Italy (Fondazione Don Carlo Gnocchi).

Acknowledgement

This work was funded and done in the context of the EU Horizon2020 ENRICHME project, Grant Agreement No: 643691

References

1. Amos, B., Ludwiczuk, B., Harkes, J., Pillai, P., Elgazzar, K., Satyanarayanan, M.: Openface: Face recognition with deep neural networks. GITHUB (2016), <http://github.com/cmusatyalab/openface>
2. Coradeschi, S., et al: Giraffplus: A system for monitoring activities and physiological parameters and promoting social interaction for elderly. *Human-Computer Systems Interaction: Background and Applications 3 Advances in Intelligent Systems and Computing* 300, 261–271 (2014)

3. Dondrup, C., Bellotto, N., Jovan, F., Hanheide, M.: Real-time multisensor people tracking for human-robot spatial interaction. In: Workshop on Machine Learning for Social Robotics at International Conference on Robotics and Automation - ICRA (2015)
4. Ekman, P., Friesen, W.: Facial action coding system. In: Palo Alto, CA: Consulting Psychologists Press (1978)
5. Fischinger, D., Einramhof, P., et al: Hobbit - the mutual care robot. In: ASROB-2013 in conjunction with IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Japan (2013)
6. Foundation, B.N.: 8 healthy eating tips. British Nutrition Foundation (2014), <https://www.nutrition.org.uk/healthyliving/healthyeating/8tips>
7. Goldberg, L.: An alternative 'description of personality': the big-five factor structure. In: Journal of Personality and Social Psychology. vol. 59 (1990)
8. Graf, B., Reiser, U., Haegele, M., Mauz, K., Klein, P.: Robotic home assistant care-o-bot 3-product vision and innovation platform. IEEE Workshop on Advanced Robotics and Its Social Impacts (2009)
9. He, W., Goodkind, D., Kowal, P.: An aging world: 2015. International Population Reports (2016)
10. Ioannou, S., Gallese, V., Merla, A.: Thermal infrared imaging in psychophysiology: Potentialities and limits. In: Psychophysiology, Wiley Periodicals (2014)
11. King, D.: Dlib-ml: A machine learning toolkit. In: Journal of Machine Learning Research. vol. 10, pp. 1755–1758 (2009)
12. Leconte, F., Ferland, F., Michaud, F.: Design and integration of a spatio-temporal memory with emotional influences to categorize and recall the experiences of an autonomous mobile robot. Autonomous Robots (2015)
13. Lorenz, T., Weiss, A., Hirche, S.: Synchrony and reciprocity: Key mechanisms for social companion robots in therapy and care. In: International Journal of Social Robotics (2015)
14. Matsumoto, D., Ekman, P.: Facial expression analysis. In: Scholarpedia (2008), http://www.scholarpedia.org/article/Facial_expression_analysis
15. Quigley, M., Conley, K., Gerkey, B.P., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A.Y.: Ros: an open-source robot operating system. In: ICRA Workshop on Open Source Software (2009)
16. Sorostinean, M., Ferland, F., Tapus, A.: Reliable stress measurement using face temperature variation with a thermal camera in human-robot interaction. In: 15th IEEE-RAS International Conference on Humanoid Robots, Humanoids (2015)
17. Tanaka, M., Ishii, A., et al: Effect of a human-type communication robot on cognitive function in elderly women living alone. Medical Science Monitor 18 (2012)
18. Tapus, A., Tapus, C., Mataric, M.J.: Music therapist robot for individuals with cognitive impairments. In: Proceedings of the ACM/IEEE Human-Robot Interaction Conference (HRI) (2009)
19. Tapus, A., Tapus, C., Mataric, M.J.: The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In: Proceedings of the International Conference on Rehabilitation Robotics (ICORR) (2009)
20. Wada, K., Shibata, T., Saito, T., Tanie, K.: Psychological, physiological and social effects to elderly people by robot assisted activity at a health service facility for the aged. In: Proceedings of IEEE/ASME International Conference on Advanced Intelligent Mechatronics (2003)
21. Yu, B., Gabriel, D., Noble, L., An, K.N.: Estimate of the optimum cutoff frequency for the butterworth low-pass digital filter. In: Journal of Applied Biomechanics. vol. 15 (1999)