A Social Robot for Emotion Recognition and Burden Levels Assessment for Informal Caregivers

Samuel Millan Norman
School of Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
millan-normans@cardiff.ac.uk

Carolina Fuentes
School of Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
fuentestoroc@cardiff.ac.uk

***Abstract*—Emotional exhaustion and high burden levels affect the wellbeing of Informal Caregivers. Different technologies explore how to support better their carer tasks and their wellbeing. In this paper, we present the deployment of a system that will track burden levels and emotion analysis using a social robot. We describe two systems; the first system integrates multi-modal emotional analysis using visual and voice inputs. The second system is based on a multi-modal emotional analysis using visual, voice inputs and touch-screen interaction. Based on the information detected, the system analyses emotions and activates the pick-and-place movement associated with the emotions identified. Future work will include participatory design sessions and evaluations with social care workers, and informal and formal caregivers to improve the interactions with social robots in the informal care context.**

***Index Terms*—Human-Robot Interaction, Self-Report of Emotions, Emotion Detection, Informal Caregivers, Social Robots, Neural networks**

I. Introduction

The use of social robots for emotion recognition has been explored due its relevance to improve Human Robot Interaction (HRI). The Pepper robot created by SoftBank Robotics [1] is a popular social robot that aims to enable human-like robot communication whilst users can benefit from their functionalities. However, multiple HRI challenges are still unsolved to effectively integrate social robots in different industries. Some of the challenges are the integration of the systems in changing environments and the overall performance of the system in terms of processing speed and accuracy. The Healthcare sector is a particularly challenging field in which robots such as Paro [2], NAO [3], CRECA [19], Betty [21] have been explored to support stress reduction and enhance motivation.

Informal caregivers or unpaid carers are people who usually provide care to family members or friends, without receiving any type of compensation for their work. Informal caregivers commonly are affected by multiple physical, financial and emotional problems that impact their wellbeing. In the context of informal care, several studies explored how to support caregivers using technology [25], [13]. The scope varies from tangible, ambient and mobile applications to more sophisticated emerging technologies as social robots [14], [5], [12], [27]. Previous implementations of social robots explore different scenarios for elder people with depression [20]. However, in the context of informal care robots that can analyse a person emotions can support them contributing to their well-being [23].

One of the multiple challenges that informal caregivers face is the increase of burden levels which is caused by the high demanding needs that caregivers are required to undertake [18]. Due to the COVID-19 pandemic, the levels of burnout of this group increased [7] impacting burden. To identify burden levels there are several standardized tests that can be applied. In this paper, we propose a framework to be used as a mechanism for emotional recognition and self-report of burden levels using a Zarit test [30] implemented in a social robot. At this initial stage we have deployed two systems that combine data from different types of interactions that will help us to understand caregivers' well-being and could increase the test's feasibility whilst providing faster feedback to the caregiver [10], [22].

II. Related Work

The use of robots to identify human emotions has been developed and tested in laboratories [29]. Previously, the performance of NAO (a humanoid robot, developed by SoftBank Robotics) has been analysed and compared to a human examiner to produce a cognitive test [10]. This test demonstrated the improvement in the "ecological validity of the testing procedure" implying the mitigation of possible influence of emotion produced by the examiner towards the participant during the test. In addition, it has been proven that the use of Pepper robot for multimodal emotion recognition is a solution useful to classify emotions [29]. Previous analysis with Pepper robot determined that there was a 10% increase in the overall accuracy in a system using gait and thermal data in order to analyse the emotion [29]. Other applications of social robots within the healthcare industry has been designed to help elderly people to perform the tasks which due to their age they are no longer capable of doing, such as support a rehabilitation process,companionship or therapeutic support [6], [17].

Multi-modal emotion recognition mechanisms have been studied from different perspectives in order to determine the accuracy of a system. The use of multi-modal mechanisms such as visual and auditory inputs for emotion recognition has been tested and identified as more accurate than mono-modal mechanisms [15]. Previous work on emotion detection

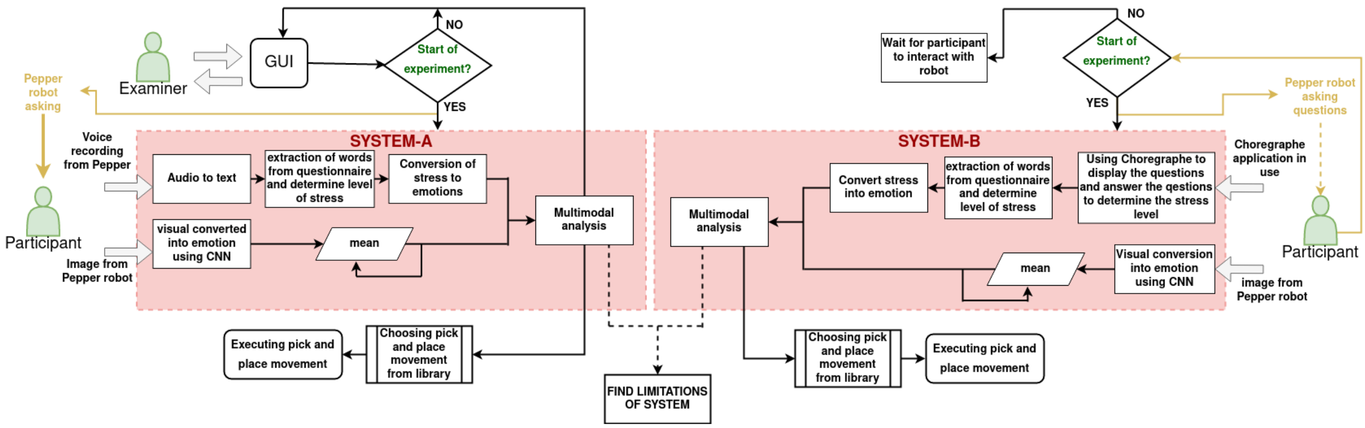

Fig. 1. The proposed framework present two systems: SYSTEM-A (shown on the left hand side) requires a examiner and a participant in order to interact with the system. It accepts both discrete (e.g. for audio analysis) and continuous (e.g. for visual analysis) types of data. On the other hand, SYSTEM-B (shown on the right hand side) only requires one user to initiate the interaction and uses other methodology to do the audio analysis.

using social robots have been attempted using a PAD model of visual and auditory sensing managed to determine the overall emotion of a group [26]. Although the robot demonstrated it is capable of analyzing and determining the overall emotion of its surrounding, still is not capable of creating actions based on the emotions detected and help to influence or represent the current emotion in the room. For the analysis of visual emotion recognition the use of Convolutional Neural Networks (CNN) [28] with Support Vector Machine classifier perform better than using a K-Nearest Neighbour classifier [16]. Implementing emotion recognition in robots that will interact with caregivers has the potential to prepare interactions that adapt to emotional caregivers' needs, serving as another element that helps caregivers to vent, reflect and calm [23]. Previous work has shown how the effect of stress can modify the emotions of a person [24]. In order to assess the burden of an informal caregiver and complement it with emotion recognition, we implement questions from the Zarit Burden Interview (ZBI), which is a self-report tool, that uses the 5-point Likert scale to measure different burden levels [9], specially designed for caregivers so they can rate each question with the pre-set option [31]. This test can be carried either with a verbal assessment or a written assessment.

In this paper we present two different systems which combine a multimodal emotion recognition. It consists of a typical method of emotion recognition using a CNN whilst the other procedure uses a new way for classifying the output of ZBI into emotions. One system gives the user the freedom to express itself and then attempt to classify the response the other method limits the user to the answers given by the questionnaire.

III. Framework for Emotion Detection and Burden Assessment Using a Social Robot

We propose a framework that will compare two systems to measure caregivers' well-being through emotion detection and a burden levels assessment. Both systems implemented are composed of a multi-modal input using Pepper robot's front microphone and top camera peripherals. From the data collected, the system detects the type of emotion, and the robot will produce a pick and place movement based on the type of emotion.For example, if the user is detected to be happy the robot will move hand to a certain point in space. Then it would pick an object and place it in the designated area.

We present two types of multimodal systems. As shown in Figure 1, the first system, 'SYSTEM-A', requires a guide to dictate the pace at which the test is completed and extracts keywords from the burden assessment in the phrases recorded for each question. On the other hand, the second system, 'SYSTEM-B', integrates the screen interaction to show the burden assessment but is limited to only one answer per question.

A. *Emotion recognition and Zarit Burden Interview*

To label emotions, we use Paul Ekman's categorisation of six basic emotions [11]. For obtaining the overall emotion detected visual and auditory inputs are used to correctly recognise emotions. In order for both sources of emotion recognition to have an equal weight in the final output, the result of each input is obtained independently and is then normalized and averaged. The continuous input utilises the visual component making use of a pre-existing Convolutional Neural Network (CNN) [8] for emotion detection. The discrete input utilises the auditory component for obtaining the burden level from the user based on the questions from Zarit Burden Interview (ZBI) and one last question to assess the user's self-report of his/her own emotions which then in combination with burden level is converted into an emotional representation. Hence the output of the discrete analysis would be limited to 4 different cases as the ZBI uses to represent the level of stress. To represent the feeling of stress in terms of emotions, a classification procedure is used. Since the feeling of stress combines the emotions of anger, disgusted, fearful, and sad, the overall weight of the test to obtain the final emotion is based on a conversion of the four levels of stress from 0 to 3 per group of emotions which represents stress, each emotion being assigned the same weight within the group, and vice-versa for the remaining emotions.

TABLE I
CLASSIFICATION OF EMOTIONS FROM LEVEL OF STRESS

Level of Stress	angry	disgusted	fearful	happy	sad	surprised
1	0	0	0	3	0	3
2	1	1	1	2	1	2
3	2	2	2	1	2	1
4	3	3	3	0	3	0

Although both systems utilise the same principle for extracting the emotion, there is a difference in how the ZBI answers are gathered. Below we explain each system individually.

B. SYSTEM-A

In this case the system will have an *examiner*, a person who assists the robot when doing the testing with the caregiver, in order to be used. The system uses a Graphical User Interface (GUI), manipulated by the examiner, to start the data gathering. Once the process starts, both the visual and auditory inputs are gathered and analysed. Finally, the result of the total analysis is represented in the movement of the Pepper robot. In addition, the examiner can see the statistics with the help of the GUI for each individual emotion detection procedure and the overall emotion extracted.

For this system, we used Robot Operating System (ROS) framework. Pepper robot is utilized as the social robot to interact with the participant whilst the laptop is used by the examiner in order to interact with the GUI, and to process all the information obtained from the peripherals of the robot.

1) Auditory analysis from SYSTEM-A: The way in which this data gathering is done within this system is by first collecting the answers to the questions in an audio file. First, the robot asks a question which is also indicated to the user by changing the tablet and LEDs to a blue colour, as shown in the left hand side of Figure 2. Then both, the tablet and LEDs, changes to green colour to indicate that the system is recording the audio as shown in the right hand side Figure 2. The response of the user is converted from audio to text using the speech recognition library [32] from Google Cloud Speech API. For extracting the emotion, the system uses a similar principle as explained above, although it searches for all the possible solutions which are provided by the questionnaire for the given answer and then it calculates an average value for the overall level of stress for that given question which then is translated into emotions one of the six type of emotions using the classification procedure explained before.

2) Graphical User Interface (GUI): The GUI provides the examiner a certain level of control of the starting procedure of the data collection and to show graphically the data collected. The way it is designed is to have half of the screen which displays the image imported from the robot with the emotion detected. The other side of the screen allows users to interact where the user is intended to go through three different windows. The first window, provides a button for starting

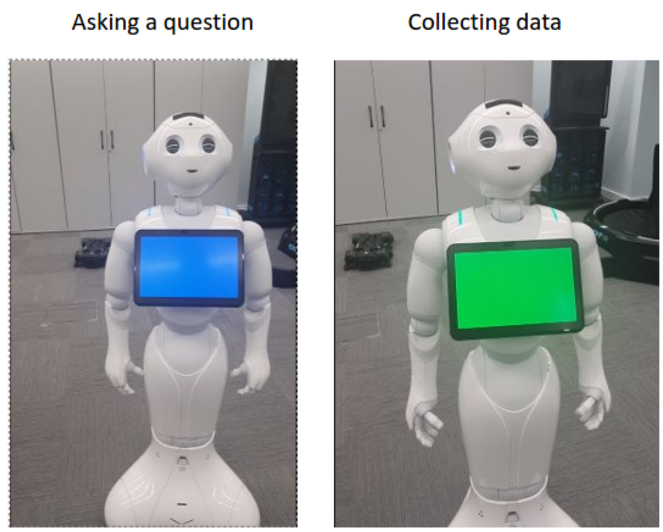

Fig. 2. Pepper robot responds to different auditory inputs, indicating to the user when the robot is asking a question and when is ready to collect the data in order to obtain a record of the response to the question asked.

the test. The second screen is a window which shows the timing left for the test to gather all the necessary data for analysing the person's emotion. And the final emotion is the graphical representation of the data collected and a button is the one which allows the process to start again. Four types of information are represented in four different bar charts which are accessed by clicking a button. The first data type is the instantaneous facial recognition emotion. This data represents the current emotional output of the face detection mechanism. The second data type is the overall facial recognition, Which is the averaged facial emotion detected. The third data type is the data extracted from the voice recognition mechanism. Finally, the last data type is the overall emotion recognition which represents the combination of the facial and voice recognition in one output.

C. SYSTEM-B

This system requires the participant to initiate the test, and once the test is started, the system activates data gathering. In order to start the data gathering, the user interacts with the tablet which Pepper robot has incorporated. At the end of the analysis, the user will see displayed in the tablet the answer obtained from the discrete input analysis and in addition the movement selected from the pick-and-place library representing the final emotion of the multi-modal system.

The system utilises a ROS framework in conjunction with Choregraphe [4]. Similar to the SYSTEM-A, this system also uses Pepper robot as the social robot to interact with the participant, however, the discrete input analysis is incorporated into Choregraphe whilst the continuous input analysis, multimodal analysis, and movement of the robot is incorporated into ROS.

1) Auditory analysis from SYSTEM-B: The emotion is extracted by utilising the tablet screen as a means of input in conjunction with the auditory system. Therefore the robot

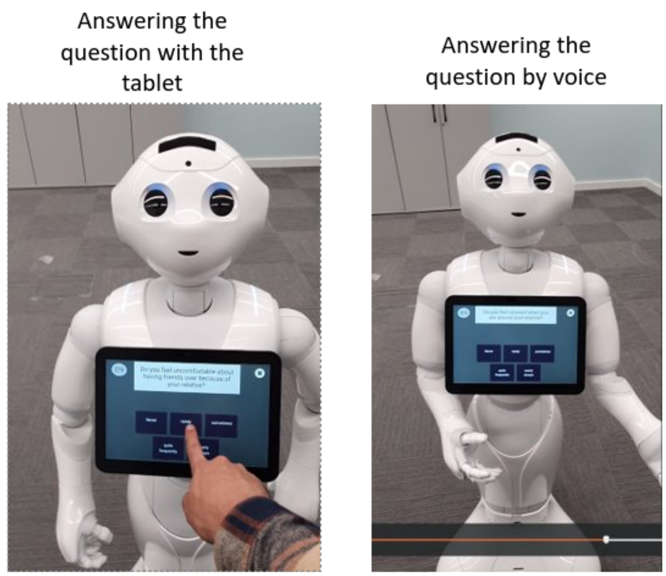

Fig. 3. This shows an image of system which utilises the ZBI can be interacted using the voice or the tablet in the screen

asks the question to the user and at the same time it shows the possible answers as if they were doing the questionnaire. In comparison to SYSTEM-A, this system also allows the user to input the answer using the tablet screen but limiting the user to only one possible solution per question (See Figure 3). Then the level of stress is converted to emotion using the classification procedure explained before.

D. Pick and place library

The pick and place library used contains seven different predefined and hard-coded movements which prove the robot with the commands to pick an object and place it in another place. The movements are intended to use are to pick an object whose colour represents one of the possible emotions the multi-modal analysis can detect. Then, placing the object into a base so that the user can have a clear visual output of the emotion detected.

‘

IV. Discussion

Both System-A and System-B aim to combine an emotional assessment and burden levels from a user based in their responses, moderated by means of interactions with a social robot.

This work is an initial stage to exploring ZBI and emotion recognition integration mediated by a social robot. We have deployed a system to analyse emotions and produce an output successfully, however there are aspects of this project to be assessed and refined based on expert feedback and participatory design sessions that will help to determine how and when this solution would be suitable to integrate within the informal care context.

SYSTEM-A is purely based on voice and visual detection, and the SYSTEM-B uses quiz-like answers which collect information from the ZBI using visual and voice interaction combined with the screen as the output of the visual representation. The SYSTEM-B also provides the user with the option of pressing the multiple choice answer or to input it by voice. All of this whilst doing a visual analysis of the face for emotion detection. Hence the difference between both systems could impact the result of the experiment since the data gathered is different. Whilst SYSTEM-A does not suggest any type of answer and searches for keywords in sentences from each question, SYSTEM-B limits the user to the given words in the screen and only allows to input one of the word as an answer per question. Therefore it reduces the possibilities of self-reporting emotions allowing to express how the informal caregiver feels.

A. Testing of both systems in simulation

A simulation of the complete system has been tested in laboratory environment using the developers as the participants. The simulation environment used was Gazebo which provides an easy interaction with the ROS system. To gather the data outputted by the multimodal system the Pepper robot was utilized. Therefore the user would interact with Pepper robot for gathering all the information and the result of the experiment would be shown in the simulation environment.

B. Implementation of both systems in real environment

Next steps of this works aim to conduct participatory design sessions, involving experts as social care workers and formal and informal caregivers to refine the interactions with the robot, identifying appropriate mechanisms on when to trigger Zarit burden questions, and analyse which scenarios would be the most appropriate to deploy this solution.

For the current systems, the experiment has to be undertaken in a known environment where all the objects around it must be placed in specific coordinates in order for the pick-and-place function to work effectively. Enabling a high level control of the robot and incorporating visual and depth sensors for determining object positions could enable the mechanism to pick and place an object without having a fixed position. This will eliminate the need for using the pick-and-place library and will enable the robot to be deployed in a wider range of environments.

In addition, for SYSTEM-A, the GUI is currently manipulated by an external examiner to dictate the pace of the experiment. Although a better implementation of this system would be to convert the GUI to a web page allowing the GUI to be implemented into Pepper robot whilst having an external database. Allowing for further analysis of each experiment undertaken while the robot is still deployed. For SYSTEM-B, the screen can be used to show the overall emotion detected by the multi-modal analysis instead of only showing the output of one detection mechanism.

C. Further Work

Future work will include participatory design sessions to improve and refine interactions, in addition, we will automatise and incorporate the GUI into the robot for self-deployment in SYSTEM-A, followed by data analysis of the data gathered. A possible solution could be to develop an algorithm to analyse the data throughout different days, months or years for assessing and tracking the participant level of burden follow-through time. The output interfaces will require improving the visualisations to facilitate understanding. Furthermore, both systems can be used to determine and analyse the support provided in order to overcome burnout. Finally, is required a methodology in order to analyse each system.

V. Conclusion

This paper presents a preliminary deployment stage to prepare a social robot to interact with and support informal caregivers. We deploy a system that will track burden levels and emotion analysis to later evaluate and integrate it as a well-being support tool in the context of informal care. We describe two systems; the first system is based on a multi-modal emotional analysis using visual and auditory inputs. The second system is based on a multi-modal emotional analysis that uses visual, voice inputs and touch. Based on the information detected, the system analyses emotions. The output uses a library to activate pick-and-place movements as a result of the emotions detected by the multi-modal system. Hence, the framework proposed aims to find the possible limitations when users interact with the robot for emotion detection. Future work includes conducting participatory design sessions with informal and formal caregivers and social care workers to improve the interactions with social robots in the informal care context.

References

[1] Accessed: 15-07-2022. [Online]. Available: https://www.softbankrobotics.com/emea/en/pepper

[2] Accessed: 15-07-2022. [Online]. Available: http://www.parorobots.com/

[3] Accessed: 15-07-2022. [Online]. Available: https://www.softbankrobotics.com/emea/en/nao

[4] Accessed: 15-07-2022. [Online]. Available: https://developer.softbankrobotics.com/pepper-naoqi-25/naoqi-developer-guide/choregraphe-suite

[5] I. Astudillo, C. Fuentes, and V. Herskovic, "Analízate: towards a platform to analyze activities and emotional states of informal caregivers," *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 2, no. 19, p. 1205, 2018.

[6] S. Bedaf, H. Draper, G.-J. Gelderblom, T. Sorell, and L. de Witte, "Can a service robot which supports independent living of older people disobey a command? the views of older people, informal carers and professional caregivers on the acceptability of robots," *International Journal of Social Robotics*, vol. 8, no. 3, pp. 409–420, 2016.

[7] D. Burrows, J. Lyttleton-Smith, L. Sheehan, and S. Jones, *Voices of carers during the COVID-19 pandemic: messages for the future of unpaid caring in Wales*, Cardiff University, June 2021.

[8] T. Chusri, "Facial-expression-recognition-1," *GitHub repository*, 2019. [Online]. Available: https://github.com/chusri/Facial-Expression-Recognition-1

[9] A. DeCastellarnau, "A classification of response scale characteristics that affect data quality: a literature review," *Quality & quantity*, vol. 52, no. 4, pp. 1523–1559, 2018.

[10] L. Desideri, C. Ottaviani, M. Malavasi, R. di Marzio, and P. Bonifacci, "Emotional processes in human-robot interaction during brief cognitive testing," *Computers in Human Behavior*, vol. 90, pp. 331–342, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563218303881

[11] P. Ekman, "Basic emotions," in *Handbook of Cognition and Emotion*, T. Dalgleish and M. J. Powers, Eds. Wiley, 1999, pp. 4–5.

[12] C. Fuentes, C. Hernandez, L. Escobedo, V. Herskovic, and M. Tentori, "Promoting self-reflection of social isolation through persuasive mobile technologies: the case of mother caregivers of children with cancer," *International Journal of Human-Computer Interaction*, vol. 30, no. 10, pp. 802–814, 2014.

[13] C. Fuentes, V. Herskovic, I. Rodríguez, C. Gerea, M. Marques, and P. O. Rossel, "A systematic literature review about technologies for self-reporting emotional information," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 4, pp. 593–606, 2017.

[14] C. Fuentes, I. Rodríguez, and V. Herskovic, "Making communication frequency tangible: How green is my tree?" in *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, 2016, pp. 434–440.

[15] D. Gharavian, M. Bejani, and M. Sheikhan, "Audio-visual emotion recognition using fcbf feature selection method and particle swarm optimization for fuzzy artmap neural networks," *Multimedia Tools and Applications*, vol. 76, pp. 2331–2352, 2015.

[16] T. Iqbal and S. T. Javed, "A new perspective towards analysis of human facial expression using supervised classification algorithms," in *2019 8th International Conference on Information and Communication Technologies (ICICT)*, 2019, pp. 94–99.

[17] S. Jenkins and H. Draper, "Care, monitoring, and companionship: Views on care robots from older people and their carers," *International Journal of Social Robotics*, vol. 7, no. 5, pp. 673–683, 2015.

[18] D. G. Kinman, "Managing stress, burnout and fatigue in health and social care," in *Supporting occpuation health and wellbeing professionals*. Produced in partnership with Public Health England.

[19] K. Kurashige, S. Tsuruta, E. Sakurai, Y. Sakurai, R. Knauf, and E. Damiani, "Design of counseling robot for production by 3d printer," in *2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, Dec 2017, pp. 56–62.

[20] H. R. Lee, S. Šabanović, W.-L. Chang, D. Hakken, S. Nagata, J. Piatt, and C. Bennett, "Steps toward participatory design of social robots: Mutual learning with older adults with depression," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 2017, pp. 244–253.

[21] S. M. Loi, A. Bennett, M. Pearce, K. Nguyen, N. T. Lautenschlager, R. Khosla, and D. Velakoulis, "A pilot study exploring staff acceptability of a socially assistive robot in a residential care facility that accommodates people under 65 years old," *International Psychogeriatrics*, vol. 30, no. 7, p. 1075–1080, 2018.

[22] S. N. Mohammed and A. K. A. Hassan, "A survey on emotion recognition for human robot interaction," *J. Comput. Inf. Technol.*, vol. 28, no. 2, pp. 125–146, 2020. [Online]. Available: http://cit.fer.hr/index.php/CIT/article/view/4841

[23] S. Moharana, A. E. Panduro, H. R. Lee, and L. D. Riek, "Robots for joy, robots for sorrow: community based robot design for dementia caregivers," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 458–467.

[24] S. Paulmann, D. Furnes, A. M. Bøkenes, and P. J. Cozzolino, "How psychological stress affects emotional prosody," *PLOS ONE*, vol. 11, no. 11, pp. 1–21, 11 2016. [Online]. Available: https://doi.org/10.1371/journal.pone.0165022

[25] M. Petrovic and A. Gaggioli, "Digital mental health tools for caregivers of older adults—a scoping review," *Frontiers in Public Health*, vol. 8, p. 128, 2020.

[26] J. Rincon, A. Costa, P. Novais, V. Julian, and C. Carrascosa, "Detecting social emotions with a nao robot," in *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer, 2016, pp. 286–289.

[27] M. Spezialetti, G. Placidi, and S. Rossi, "Emotion recognition for human-robot interaction: Recent advances and future perspectives," *Frontiers in Robotics and AI*, vol. 7, p. 532279, 2020.

[28] P. Tzirakis, G. Trigeorgis, M. Nicolaou, B. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, 04 2017.

[29] C. Yu and A. Tapus, "Interactive robot learning for multimodal motion recognition," in *Social Robotics: 11th International Conference, ICSR 2019, Madrid, Spain, November 26–29, 2019, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2019, p. 633–642.

[30] S. H. Zarit, K. E. Reever, and J. Bach-Peterson, “Relatives of the impaired elderly: correlates of feelings of burden,” *The gerontologist*, vol. 20, no. 6, pp. 649–655, 1980.
[31] ——, “Relatives of the Impaired Elderly: Correlates of Feelings of Burden1,” *The Gerontologist*, vol. 20, no. 6, pp. 649–655, 1980.
[32] A. Zhang, “Speech recognition,” *GitHub repository*, 2017. [Online]. Available: https://github.com/Uberi/speech$_recognition$