Prototyping a Zoomorphic Interactive Robot Companion with Emotion Recognition and Affective Voice Interaction for Elderly People

BENJAMIN SCHNITZER, Offenburg University, Germany
UMUT CAN VURAL, Offenburg University, Germany
BASTIAN SCHNITZER, KORION Interactive GmbH, Germany
MUHAMMAD USMAN SARDAR, Devomech Solutions GmbH, Germany
OREN FUERST, Interactive Companion Ltd., Israel
OLIVER KORN, Offenburg University, Germany

An aging society paired with a skilled labor shortage, particularly in European countries, requires a rethinking of deprecated structures. Intelligent assistive technologies, specifically socially assistive robots, addressing the gap between caretakers and elderly people in need of care have moved into the focus of debate due to their potentials to reduce costs, improve independence, and eventually raise quality of life. In this work, we outline the potentials of zoomorphic robot companions combining intelligent conversational abilities and emotion recognition. We then describe the prototyping of an emotion-sensing zoomorphic interactive robot companion including the development and implementation of a multimodal emotion recognition framework. This framework uses speech emotion recognition, sentiment analysis, and affective voice interaction based on a large language model. The prototyping has been accompanied by two studies on elderly peoples' design preferences regarding the proposed feature set as well as different embodiments to find the appropriate casing for the robot companion. This work provides valuable insights into the prototyping and can thus support future research endeavors in this area.

CCS Concepts: • **Computer systems organization** → *External interfaces for robotics.*

Additional Key Words and Phrases: Intelligent Assistive Technology, Socially Assistive Robots, Speech Emotion Recognition, Affective Computing, Zoomorphic Embodiment, Health, Elderly People

ACM Reference Format:
Benjamin Schnitzer, Umut Can Vural, Bastian Schnitzer, Muhammad Usman Sardar, Oren Fuerst, and Oliver Korn. 2024. Prototyping a Zoomorphic Interactive Robot Companion with Emotion Recognition and Affective Voice Interaction for Elderly People. *Proc. ACM Hum.-Comput. Interact.* 8, EICS, Article 242 (June 2024), 32 pages. https://doi.org/10.1145/3660244

1 INTRODUCTION

The global population aged 60 or above is projected to reach nearly 2.1 billion by 2050, meaning one in five people will be a senior [91]. About a quarter of these seniors already suffer from loneliness

Authors' Contact Information: Benjamin Schnitzer, benjamin.schnitzer@hs-offenburg.de, Offenburg University, Affective & Cognitive Institute, Offenburg, Germany; Umut Can Vural, Umut.vural@hs-offenburg.de, Offenburg University, Affective & Cognitive Institute, Offenburg, Germany; Bastian Schnitzer, bastian.schnitzer@korion.de, KORION Interactive GmbH, Ludwigsburg, Germany; Muhammad Usman Sardar, usman.sardar@devomech.com, Devomech Solutions GmbH, Kornwestheim, Germany; Oren Fuerst, oren@eco-fusion.com, Interactive Companion Ltd., Tel Aviv, Israel; Oliver Korn, oliver.korn@hs-offenburg.de, Offenburg University, Affective & Cognitive Institute, Offenburg, Germany.

ACM 2573-0142/2024/6-ART242
https://doi.org/10.1145/3660244

[22], defined as "the feeling of being alone, regardless of the amount of social contact" [32]. At the same time, a lack of social connections leads to social isolation that can lead to loneliness in turn. Research has shown that the mortality effect of loneliness equates to smoking 15 cigarettes a day [39]. Moreover, loneliness increases the risk of many illnesses commonly associated with aging. These risks include high blood pressure and other cardiovascular diseases, obesity and diabetes, as well as insomnia, anxiety, and depression [22]. In the mid and long term, there are also increased risks of cognitive decline and dementia [22]. While all countries face major challenges to ensure that their health systems are prepared for the imminent demographic shift [91], the COVID-19 pandemic worsened conditions for many elderlies and accelerated social isolation. Furthermore, it revealed another alarming issue particularly for European countries: a skilled labor shortage [29]. Due to these societal issues, intelligent assistive technologies (IATs) addressing the gap between caretakers and elderly people in need of care have recently become the focus of debate. This is mainly due to their potential to safe time, reduce costs, improve independence, and eventually raise quality of life. At the same time, ethical concerns are raised and the design of such IATs is questioned regarding functionality, appearance, and behavior.

In a recent advocacy brief, the World Health Organization suggests the use of digital interventions that incorporate cognitive behavioral therapy, train social skills, and befriend people to improve health and well-being amongst the elderly [97]. In general, IAT is an umbrella term often used for a broad range of digital solutions (both hardware and software) with self-contained computation capabilities that offer smart services. These services are designed to assist older adults and people with mild to moderate cognitive impairments (MCI) or dementia and their caregivers [41, 42]. Thereby, the spectrum encompasses a large variety of Internet of Things and cloud technologies including distributed systems such as ambient assisted living technologies, mobility and rehabilitation aids, handheld devices, wearables, apps, human-machine interfaces, and care robots or robot companions [42]. The latter, commonly known as socially assistive robots (SARs) are of special interest for the target group of lonely elderly people [47, 86], particularly for those with cognitive impairments or dementia, as they could both help the elderly persons to retain autonomy and serve as a companion offering distraction, entertainment, and cognitive training [34, 36, 83].

In terms of shape and appearance, SARs vary considerably, although they are frequently distinguished by humanoid, zoomorphic (also referred to as pet-like or creature-like), and object-like design metaphors [75]. For example, while SARs like Pepper [8] or Ryan [2] clearly represent anthropomorphic designs and emulate human-like behavior, e.g. through verbal communication, robots like Paro [96] or Tombot's Jennie [45] have zoomorphic embodiments and imitate animal-like behavior.

People generally seem to have positive attitudes towards SARs and are willing to interact with them [66]. However, most current devices for elderly people are too expensive for the broad adoption in healthcare centers [50], especially when more than one device is needed. Zoomorphic robot companions, which elicit lower expectations regarding human-like capabilities, may bridge a gap here. Nevertheless, the question arises whether elderly people would accept zoomorphic SARs with intelligent and affective language capabilities capable of simultaneously monitoring users' emotional well-being. Interestingly, while the acceptance of zoomorphic embodiments has been studied [9–11], to the best of our knowledge there is little research investigating the combination of design features. But considering the demographic shift and the concurrent labor-shortage, this approach appears particularly promising to assist clinicians' work.

The outlined gap and a growing demand for affordable assistive technologies encouraged us to investigate the following explorative questions: What are the potentials of a low-cost emotion-sensing zoomorphic interactive robot companion (EZIRC) for elderly people suffering from MCI or dementia? What should such a companion look like in terms of appearance and embodiment? And finally, what features should the device provide?

In this work we describe the prototyping of an EZIRC including the development and implementation of a multimodal emotion recognition framework. This framework uses speech emotion recognition (SER), sentiment analysis, and an affective voice interaction based on speech recognition and natural language processing. Furthermore, we outline the implementation of a breathing feature simulating respiration to instruct meditation and co-breathing exercises. The prototyping has been accompanied by two preliminary studies on elderly peoples' preferences regarding the proposed feature set as well as various embodiments to find the appropriate casing for the robot companion.

2 RELATED WORK

2.1 Combining Natural Language Processing and Emotion Recognition

For a long time, machine learning models for language processing were overshadowed by their visual relatives. Up to this point, the models were not semantically rich enough and the application domains often were too narrow. As a result, many users had frustrating experiences with so-called "intelligent chatbots" or virtual agents. However, natural language processing strongly improved within the last decade due to increased computation power and the latest achievements in machine learning and deep learning [49]. Nowadays, an increasing number of tools and software libraries in various programming languages are available and speech recognition, speech-to-text, and text-to-speech frameworks are becoming ubiquitous. Not least, the release of OpenAI's ChatGPT [67] in November 2022 brought large language models (LLMs) into the focus of AI debates [37]. ChatGPT is based on the GPT3.5 and most recently GPT4 model, two generative pre-trained transformer models. Evolving LLMs like GPT3.5 and GPT4 provide impressive results and the capabilities to have contextualized conversations open up new possibilities for human-robot-interaction. However, these models generally lack empathy or at least some sense of affective sensitivity which at the same time - in addition to semantic understanding - are essential factors for a "good" conversation.

The combination of emotion recognition and dialog systems allows for engaging "empathic" user interactions and thus has been a consistent area of research in the field of affective computing.

Currently three approaches dominate computational emotion recognition: (1) image-based emotion recognition (e.g., with the help of facial expressions or gesture recognition), (2) physiological signal recognition (e.g., with the help of heart-rate, electrodermal activity, or electroencephalography), and (3) audio-based emotion recognition [98]. The latter either makes use of audio features often referred to as prosodic aspects of speech [65] and asks for the question how something is said, or in the first instance transforms audio (particularly speech) into text and subsequently analyzes these data, e.g. with the help of SA [72], referring to the question what is said. Multimodal approaches combining methods are also frequently used [82, 90], mainly to increase the precision by fusing different sources of emotion recognition to one output set [5]. For instance, Bagher Zadeh et al. combine audio, video, and text-based emotion recognition into one feature set [5]. In [82], Schiffmann et al. presented an affective approach by combining emotion recognition based on audio with text-based emotion recognition.

As mentioned before, another approach to increase the robustness of emotion recognition frameworks is to include physiological signal recognition. One physiological parameter frequently used is heart rate or heart-rate-variability (HRV). Dzedzickis et al. reviewed different sensors and methods for emotion recognition and found several studies using HRV for emotion recognition [27]. Chan and Nejat [17] successfully used a heart-rate sensor in the humanoid robot Brian 2.0 to adapt to the users' emotional state during a cognitive task. In general, these approaches draw on the fact

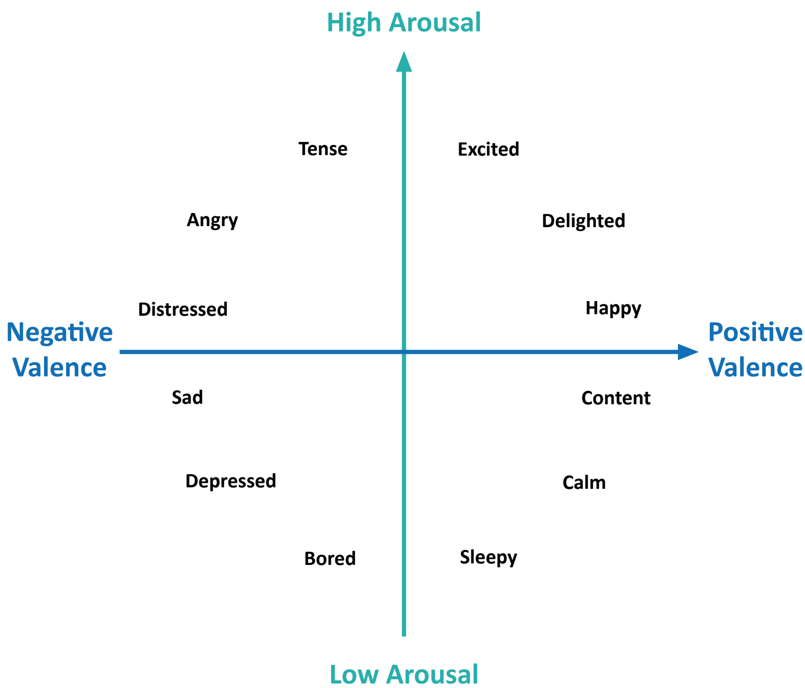

Fig. 1. Russell's circumplex model of affect. Own representation.

that different emotions are accompanied by certain physiological reactions [74]. There are a variety of models for emotion recognition in human–computer-interaction [23] but the most frequently [30] used in psychophysiology is Russell's circumplex model of affect [80] shown in Figure 1. The authors argue that affective states are based on two main neurophysiological systems: one that explains the value of emotion along a continuum of pleasant and unpleasantness and a second that refers to the corresponding physiological level of activation and arousal respectively. Thus, every emotion can be defined by a specific combination of the two dimensions varying in valence and level of activation or arousal. For instance, while sadness is an emotional state characterized by high negative valence but moderate arousal, angriness is characterized by lower negative valence but high arousal. Conversely, the emotional state allows conclusions to be drawn about the degree of arousal in such that an angry behavior normally goes along with high arousal. There are various attempts in literature to map HRV responses to the two-dimensional valence-arousal model [43, 63, 92]. An example for a state-of-the-art multi-modal architecture making use of this interdependency between emotions and physiological activation is proposed in [57]. While in Russell's Circumplex model valence and arousal are generally seen as continuous variables, it is also possible to separate the arousal-valence space into four quadrants (Figure 2): low arousal/high valence (LAHV), low arousal/low valence (LALV), high arousal/high valence (HAHV), and high arousal/low valence (HALV). This binary segmentation is commonly done for classification problems.

Generally, emotion located in HAHV and LAHV quadrants are experienced as pleasant [80]. These emotional states are therefore something that desirably last longer than negatively connoted emotions from the HALV and LALV quadrants.

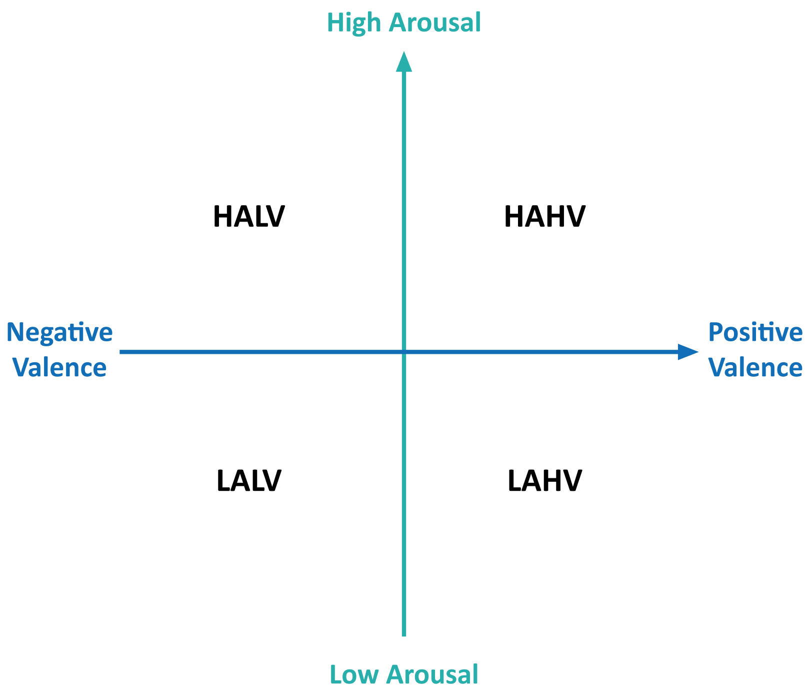

Fig. 2. Binary Segmentation of Valence-Arousal states. Own representation.

With regards to the latest progress of LLMs, the question arises how a combination of these models and multimodal emotion recognition using audio-based emotion recognition and physiological emotion recognition may affect the user experience. Furthermore, to the best of our knowledge, the impact of LLMs on the reliability of audio and text-based emotion recognition, of which the essential component is a copious verbal or conversational interaction between human and machine, is still quite unexplored.

2.2 Socially Assistive Robots

A few years ago, the development of robots interacting in a social way and being empowered by machine learning opened a terminological discussion in human-robot-interaction (HRI). Feil-Seifer and Mataric defined socially assistive robots as the intersection of assistive robots and socially interactive robotics [31]. While assistive robots are defined as ones that give aid or support to a human user, socially interactive robots develop close and affective interactions with humans for the sake of interaction itself. In contrast, a SAR creates close and affective interaction with a human user for the purpose of giving assistance and simultaneously achieving measurable progress, e.g. in learning, convalescence, or rehabilitation [31].

Today's SARs are not limited to physical assistance but also provide non-tangible interaction for meditation and recreation, cognitive training, and companionship [31]. Indeed, the emergence of these "social robots" is altering the perception of robots as a whole: from tools to social entities [52, 53]. Their spectrum of services reaches from intelligent reminders [35, 64], information provision and help in carrying out everyday tasks [71], exercise advice [10, 54, 55], entertainment and cognitive stimulation, up to verbal conversation [7, 62]. Since speech is the most natural way for humans to communicate, it appears obvious that SARs' social presence and communicative competences influence technology acceptance especially in a mostly conservative and doubtfully target group [38]. Thereby, SAR embodiments vary between humanoid or anthropomorphic and animal-like or zoomorphic appearances [68, 75]. The latter are also often referred to as so-called robot companions.

Currently one of the most popular robot companions for elderly people is Paro [87]. The robot is designed by Takanori Shibata and embodied as a zoomorphic harp seal with soft fur. Paro was specifically developed for therapeutic purposes and provides haptic stimuli and pet-like companionship for elderly people, oftentimes suffering from severe dementia or Alzheimer's disease. The robot is equipped with sensors for light, tactile, posture, temperature, and audio as well as actuators enabling Paro to move its tail and flippers or to open its eyes when people pet it. Paros' functionalities basically substitute the positive effects of animal-assisted therapy. It responds to sounds and speech and shows emotions by moving accordingly or making animal-like noises as it interacts with the user [96]. Moreover, Paro is advertised to develop an individual personality over time and can learn to behave in a way that the user prefers [44]. The application of Paro and its positive impact like increasing social interaction, decreasing stress and loneliness, increasing immune system response [12, 88], promoting psychological and physiological well-being, and improving overall quality of life [85] has been shown by the literature [34, 48, 96]. Further examples of zoomorphic robot companions are AIBO by Sony [1], Joy For All Cat and Dog [11, 57] or Tombot's Jennie [45, 68, 73].

One of the most popular humanoid robots is Pepper, a SAR built to connect with people, assist them, and provide guidance. The anthropomorphic robot has a movable head with lighted eyes, two movable arms and hands with fingers, movable joints at hip level and knee height, and three rollers to navigate flat and firm undergrounds. Nevertheless, these limbs are not functional in the sense that Pepper could grab a cup of coffee. Pepper is equipped with loudspeakers and sensors for interaction, e.g. microphones, cameras (mouth and forehead), five tactile sensors (three in the head and two on the hands), as well as several sonar sensors, lasers and gyro sensors for navigation. A touch screen installed on the breast provides a graphic user interface (GUI). Moreover, the robot is controllable by voice and capable of communicating verbally through speech [24], as well as non-verbally through body language and gaze [79]. Pepper recognizes emotions and sports a multi-modal emotion module including face recognition and SER [24, 69, 82]. Pepper is frequently deployed in business contexts and not designed for a specific target group. In spite of its limited actuator capabilities, its popularity and technology readiness level has led to numerous further developments and applications in the context of care and therapy for elderly people [18, 59, 70, 81].

At the same time, any approaches (especially no low-cost solutions) exist that combine intelligent conversational features, as found in Pepper, with a zoomorphic appearance. Regarding a possible adoption in healthcare centers and considering the natural tendency of humans to relate to animals, this is surprising. Pet-like SARs such as Paro are an important best practice regarding the form factor for the embodiment of IAT for elderly people [9, 96]. For example, in a recent study Bradwell et al. revealed high acceptance of domestic cat and dog morphologies - therefore strongly supporting the use of familiar embodiments in future robot-pet designs [9–11]. This is important to avoid feelings of leeriness and possible discomfort caused by the so-called uncanny valley phenomenon [61] often experienced with humanoid robots. Although there has been evidence for a U-shaped relation between animal likeness and likeability in zoomorphic robots [60], the broad acceptance of Paro in studies underlines the potential of pet-like embodiments [36]. Apart from that, there is evidence that a "huggable" size and soft fur are preferred design features for robot companions aimed at older adults [9, 51]. Moreover, pet-like embodiments providing haptic stimuli and physical nearness seem to have positive effects particularly on patients with dementia such as increased reminiscence

Fig. 3. Beffi 1.0 sitting on an armchair.

and improved well-being featured by decreased agitation and improved communication [95] as well as positive impacts on neuropsychiatric symptoms such as apathy, depression, anxiety, and wandering [95]. Finally, there is strong evidence for the positive benefits of physical contact like stroking and petting [89].

3 PROTOTYPING AN EMOTION-SENSING ZOOMORPHIC INTERACTIVE ROBOT COMPANION

3.1 Design Objectives and Functionalities

In the previous sections we outlined the potentials of a combinatory interactive system appearing particularly promising when it comes to the target group of elderly people with cognitive impairments. We therefore reason that such a system should be represented by a socially assistive robot with a zoomorphic or animal-like embodiment, reflecting past and present experiences with familiar pets. Moreover, since verbal communication is the most natural way to communicate and simultaneously decelerates cognitive decline, such a system should combine intelligent verbal capabilities and affective features based on dialog systems using state-of-the-art LLMs, NLP, and emotion recognition. Surprisingly, hardly any approaches exist that combine the described combination of features and there are particularly no low-cost solutions available that could be adapted by research for further modification.

Therefore, in this section we describe the prototyping of an emotion-sensing zoomorphic interactive robot companion called "Beffi". First, the design decisions based on two preliminary studies are described. Afterwards, the hardware- and software architecture, that shall serve as a foundation for other researchers in the engineering community, is detailed

To begin with, we stated three design objectives that served as a guideline for the development process:

- The system should only use low-cost components and frameworks addressing the necessity of affordable SARs.
- The overall software and hardware architecture should be modular, providing options for further development or enhancements.
- The system should function as a dual IAT serving two user groups: (1) elderly people (with MCI or dementia and suffering from loneliness) by providing companionship, social presence, and haptic stimuli; (2) caregivers and therapists by sensing and monitoring patients' emotional well-being.

Based on the mentioned feature set and the stated design objectives, we aimed for four main functionalities:

(1) *Measure:* A physiological signal recognition feature that allows measuring the heart rate with the help of a photoplethysmogram (PPG) [16] integrated in the paw of the robot companion. This optical sensor detects changes in light absorption with each cardiac cycle by shining green light on extremities like a finger or an earlap and measuring the amount of reflected light with a photosensor. Changes in oxygenated hemoglobin in arterial blood result in differences in the amount of reflected light received by the photosensor producing a waveform that can be interpreted as a pulse. These data are subsequently used to calculate HRV metrics [77] which are well-known indicators for arousal and stress [46].
(2) *Cuddle:* A feature simulating respiration using an actuator placed in the belly of the robot. Since deep diaphragmatic breathing is an effective method for physiological and psychological stress reduction [6, 40], this functionality provides visual and haptic aid for guided breathing-exercises in the manner of co-sleeping known from mother-baby relations.
(3) *Listen:* A multimodal audio-based emotion recognition feature using SER and sentiment analysis. This feature allows to predict users' current emotional state based on speech. Subsequently, the predictions are used to adapt the affective voice interaction and to initiate an appropriate measure in terms of content or action. Moreover, these results can be used to create an individual wellness diary and to monitor emotional trends. The implementation of the overall multimodal emotion recognition framework (MERF) is further explained in Section 4.2.1.
(4) *Chat:* An affective voice interaction (AVI) feature providing verbal communication by combining the GPT API [61] with an empathetic dialog logic. This approach is serving mainly two purposes: As shown, verbal conversation is essential to alleviate loneliness-related cognitive decline [94]. It is therefore self-evident that language and verbal communication are important factors for the feeling of companionship and social presence apart from a physical embodiment. However, the semantic level of the used language logic and its ability to generalize and to provide a broad spectrum of topics to talk about are essential. With the help of the GPT-models research can break new ground here. At the same time, the intended multimodal emotion recognition framework and the resulting predictions heavily depend on auditive input, hence longer periods of voice interaction between the user and the technology are desirable. This led to the need for a dialog flow as natural or human as possible. The implementation of the AVI is described in Section 4.2.1 and Section 4.2.2.

Disorders grouped under the term dementia are generally caused by degenerative brain changes affecting memory, thinking, and behavior. The most common cause for dementia is Alzheimer's disease. The authors understand MCI are not a type of dementia, however, both MCI and dementia are characterized by objective evidence of cognitive decline [13]. In addition, persons with MCI are more likely to develop dementia which is typically characterized by the involvement of more than one cognitive domain. Verbal conversation decelerates loneliness-related cognitive decline and dementia and therefore is an important measure to improve life for both, elderly people suffering from MCI as well as from mild to moderate dementia [94]. From an engineering perspective, the presented functionalities are thus relevant for both groups. However, since the evaluation with people suffering from severe dementia or Alzheimer's disease is difficult, the authors of this work focused on elderly people with MCI in the following studies.

3.2 Exploring Elderly Persons' Design Preferences for the EZIRC "Beffi"

We conducted two studies (one in April 2022 and one in February 2023) to better understand the acceptability of our aspired feature set as well as preferred embodiments. In this section, we first present the study taken place in February 2023 since these results precede the study results from April 2022 as logical conclusions.

3.2.1 Study on Feature Preferences and Embodiments. In this study with elderly people (n = 28, female = 16, male = 12, age $\bar{x}$ = 83) living in a nursing home in southwest Germany, we examined the acceptability of SAR and different functionalities and compared the favorability of different zoomorphic and non-zoomorphic embodiments. The study took place in February 2023 and all seniors voluntarily participated without payment. Participants of the sample were chosen by a healthcare professional based on a previous diagnosis indicating mild to moderate cognitive impairments or dementia.

Methods. After a short introduction video explaining state-of-the-art-robots' capabilities, the intention of the study, and a short demographic assessment (age, sex, education), participants were asked to answer a questionnaire with 12 Likert-items (Table 1) ranging from 0 (fully disagree) to 4 (fully agree). Moreover, we examined embodiment preferences by showing randomly arranged pictures of ten different embodiments (Figure 5) to the participants, pretending those embodiments were actual robot companions. We also informed the seniors about the fictional verbal and affective abilities of the "robots" and told them that all of them would have the same abilities. Referring to [26], we distinguish between acceptability, one's perception of a system before use, and acceptance, one's perception of the system after use. There are several pre-validated and well-known questionnaires addressing the acceptance of robots and SAR in particular [56]. However, to the best of our knowledge there is no established questionnaire addressing the aimed feature combination as well as the target group of elderly people. Particularly elderly people with mild to moderate cognitive impairments are facing problems focusing on things for longer periods. It is therefore important not to overwhelm these people with too many questions as found in the aforementioned established questionnaires. Therefore, we consciously adapted only a small number of items from the Technology Adoption Propensity (TAP) [76], the Negative Attitudes Towards Robots Scale (NARS), and the Technology-Specific Expectation Scale (TSES) [3]. The general attitude towards a robot or robot acceptability (RA) was assessed by five items. Thereby item 1 corresponds to the TAP item 4 [76], item 3 corresponds to the NARS item 3 [93], and to provide simplicity for the target group, item 4 is a modification of NARS item 11 [93]. Regarding the aspired feature set, feature acceptability (FA) was assessed with the help of seven further items. Items 1, 2, and 4 of this construct are adapted from TSES [3] and represent TSES items 9, 8, and 6. Item 3 is an inverted item referring to NARS item 1 [93]. Assuming that different designs could influence feature acceptability, we assessed embodiment preferences after asking item 5 and before items 6-12 (Table 1).

Some of the embodiments shown to the participants depicted commercially available robot companions [25, 45, 58, 78, 87]. The rest are cuddle toys with characteristics resembling the commercial solutions (size, soft fur, huggable). Six pictures showed a zoomorphic embodiment (Cat, Dog, Sausage Dog – which is a popular pet in Germany - , Owl, and Panda) two a stylized zoomorphic embodiment (Teddybear and Lovot), one an anthropomorphic embodiment (Baymax from the Pixar

Table 1. Questionnaire Items 1-12

ID	Item	Reference	Construct (I1-I2 Cronbach's α=.87)
1	New technologies make my life easier.	TAP Item 4	Robot Accept. (I1-5 α = .75)
2	Intelligent social robots would make my life easier.		Robot Accept. (I1-5 α = .75)
3	I would feel relaxed talking to robots.	NARS Item 3	Robot Accept. (I1-5 α = .75)
4	I can rely on robots' help.	NARS Item 11*	Robot Accept. (I1-5 α = .75)
5	I would like to interact with a robot.		Robot Accept. (I1-5 α = .75)
6	I want robots to be able to understand me.	TSES Item 9*	Feat. Accept. (I6-12 α = .82)
7	I want robots to have a sense of humor.	TSES Item 8*	Feat. Accept. (I6-12 α = .82)
8	I want robots to show feelings.	NARS Item 1*	Feat. Accept. (I6-12 α = .82)
9	I want robots to understand my feelings.	TSES Item 6*	Feat. Accept. (I6-12 α = .82)
10	A robot is allowed to measure my pulse.		Feat. Accept. (I6-12 α = .82)
11	I would cuddle with a robot.		Feat. Accept. (I6-12 α = .82)
12	I would do breathing exercises together with a robot.		Feat. Accept. (I6-12 α = .82)

* adapted and modified for simplicity

movie "Big Hero 6"), and one an abstract or object-like embodiment (ElliQ). ElliQ was presented without its screen. Scaling was aligned and the different embodiments were presented on neutral backgrounds.

Results. The overall questionnaires' construct represents the acceptability or attitude towards a social robot companion with specific features like verbal communication, emotion recognition, physiological emotion recognition (pulse sensor), and a feature for guided breathing exercises. Reliability (Table 1) was proven by high inter-item correlation for the overall questionnaire (α = .87), for robot acceptability (α = .75), and for feature acceptability (α = .82).

Participants' overall attitude towards a social robot companion tends to be positive (overall mean attitude score $\bar{x}$ = 2.26, SD = .92, mdn = 3) including items 1-12 (Table 1) for each participant as well as an overall robot acceptability score ($\bar{x}$=2.31, SD=.96, mdn=2.75) and an overall feature acceptability score ($\bar{x}$ = 2.23, SD = .97, mdn = 2). A Friedman test revealed significant differences within subjects (Chi-squared = 70.96, df = 11, p < .001) between items 1 ("New technologies make my life easier."), item 4 ("I can rely on robots' help."), item 6 ("I want robots to be able to understand me."), item 7 ("I want robots to have a sense of humor."), item 8 ("I want robots to show feelings."), and item 11 ("I would cuddle with a robot."), and 12 ("I would do breathing exercises together with a robot."), respectively. Results of items 1-12 are shown in Figure 4. Since educational classification led to very small subsamples (n < 5 for secondary school, high school, and college/university), we aggregated the data and created two subsets: lower education or elementary school (n = 19) and higher education (n=9). Born before the year 1955, many elderly people in Germany only visited a secondary modern school, thus this approach is reasonable. We analyzed differences between gender and educational groups and the respective item with the help of Mann-Whitney-U tests and likewise investigated possible correlations regarding age with the help of Spearman's rank correlation coefficients. We did not find any significant differences between age and items 1-12. However, we found a significant difference (U = 52.5, p < .04) between females ($\bar{x}$ = 1.88, SD = 1.59, mdn = 2) and males ($\bar{x}$ = .58, SD = 1.16, mdn = 0) regarding item 11 ("I would cuddle with a robot.").

Findings revealed a significant difference (U = 135, p < .01) between lower educated ($\bar{x}$ = 1.89, SD = 1.49, mdn = 2) and higher educated participants ($\bar{x}$ = .33, SD = .79, mdn = 0) regarding item 12 ("I would do breathing exercises together with a robot.").

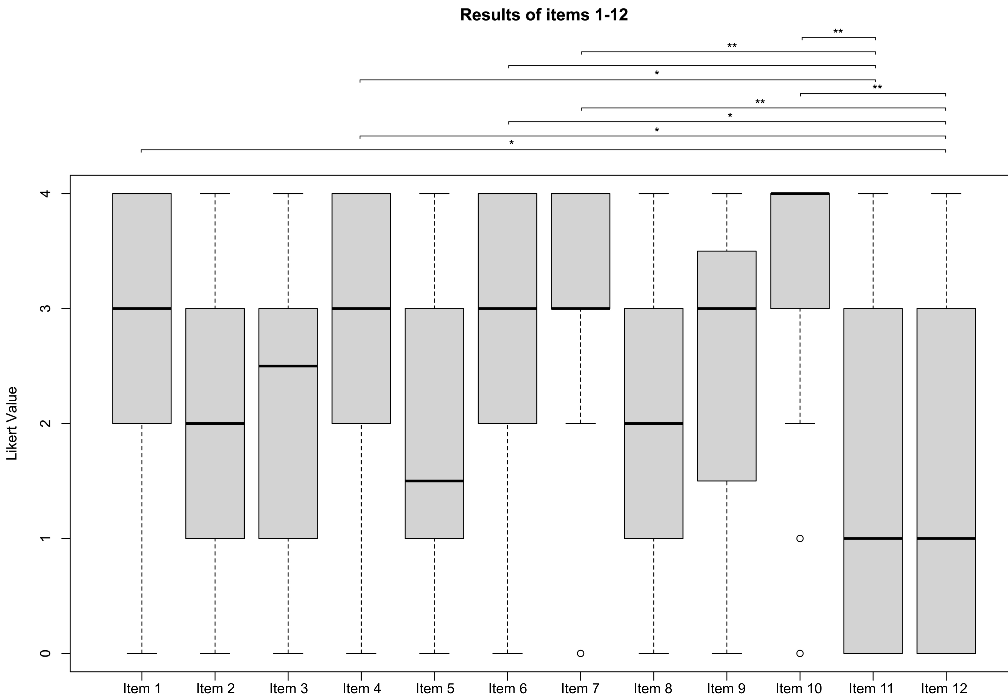

Fig. 4. Results of items 1-12.

In general, zoomorphic embodiments were rated higher than the stylized. Preferred embodiments were cat, dog, teddy bear, and panda, not differing significantly from each other, but differing significantly from the non-zoomorphic and more stylized examples (Figure 5). Moreover, we analyzed differences between females and males as well as regarding educational levels between the different embodiments. Except for the dog and the teddy bear, differences were insignificant. However, females rated the dog ($U = 55$, $p < .05$) and the teddy bear ($U = 48.5$, $p < .03$) significantly higher than males. Hence, while 100% of females rated the appearance of the dog and teddy bear rather positively (lower whisker = 2), more than 50% of males rather disliked those two embodiments (mdn < 2). Likewise, females rated the anthropomorphic embodiment significantly lower than males ($U = 132.5$, $p < .05$). Mann-Whitney-U tests regarding different embodiments and educational groups were insignificant. Finally, we found a moderate correlation (spearman $r = -0.37$, $p < .05$) between age and the ratings of Paros' appearance.

Discussion. Overall results indicate a rather positive attitude towards SAR (overall robot acceptability score $\bar{x} = 2.31$, $SD = .96$, mdn = 3). Elderly people are willing to rely on robots´ help (Figure 4, item 4, *"I can rely on robots' help."*). This is consistent with previous research [66] challenging some existing doubts about the adoption of SAR. Our findings show that the seniors appreciate robots with smart features for understanding users´ intentions (Figure 4, item 6, *"I want robots to be able to understand me."*). Although item 8 ("I want robots to show feelings.") did not show any distinct tendencies regarding robots expressing their feelings, results of item 9 ("I want robots to understand my feelings.") indicate acceptability of affective features and emotion recognition. From a design perspective, results of item 7 ("I want robots to have a sense of humor.") are interesting, showing a clear affection towards humorous robots. It is human nature to like to be amused, nevertheless,

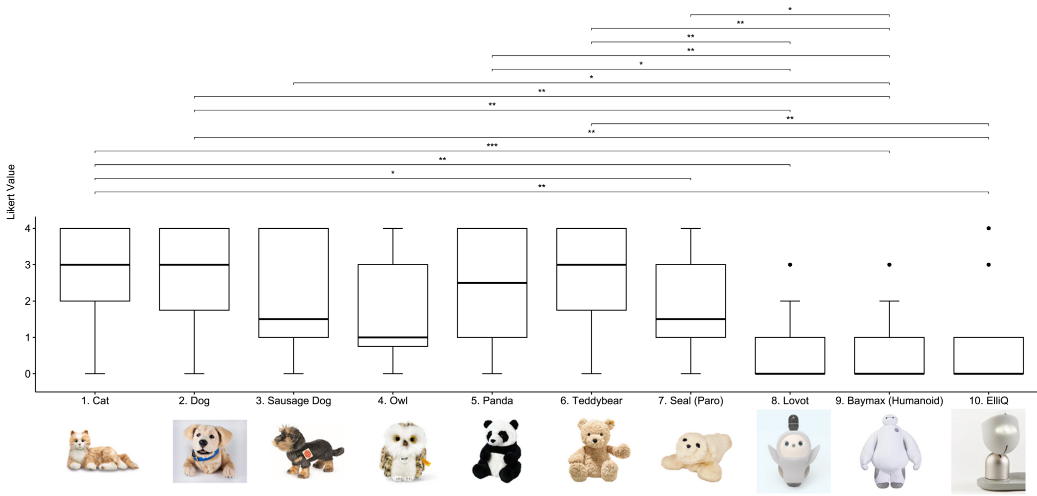

Fig. 5. Results of embodiments 1-10.

humor might be an important design factor for a user-centered experience and to engage the target group with the technology. Finally, we found strong evidence for the acceptance of robots performing routine tasks like measuring the pulse (Figure 4, item 10, *"A robot is allowed to measure my pulse."*).

At the same time, the seniors would rather not do breathing exercises with the robot. However, the experimenters realized some confusion about this question since breathing exercises and meditation weren't familiar terms to all seniors. This is also explaining the significant difference between the two educational groups. While higher educated residents clearly disagreed with item 12 ("I would do breathing exercises together with a robot"), results for lower educated elderlies are ambiguous. Likewise, the seniors would rather not cuddle with a robot. The attitude of males was clearly negative here, 50% of females (mdn = 2) at least could imagine cuddling one of the robots. Since we only assessed acceptability with hypothetical questions, results might be different for seniors who actually interact with such a robot. This assumption is confirmed by the broad acceptance of Paro shown by other studies.

Regarding embodiments, seniors preferred the zoomorphic ones. These findings underline the assumption regarding seniors' tendency to relate to familiar pets like cat and dog and a cultural habituality of the typical teddy bear. This is consistent with research by Bradwell et. al. [9–11]. However, it is questionable if interaction with a speaking zoomorphic robot companion might lead to irritations due to possible expectations regarding the behavior of such an animal. Takanori Shibata consciously chose a baby seal form for Paro, not very familiar to people but still easily accepted [87]. Although Paros' appearance was rated low in this sample, it still seems reasonable to consider not very familiar embodiments, e.g. the panda that resembles the popular teddy bear but does not break with user expectations, at least within a European target group.

3.2.2 Study on different Panda Embodiments. Apart from strong indications for the potentials of the aspired feature set described in Section 2, the previously presented results underline our decision to strive for an embodiment that on the one hand conveys familiarity and on the other hand doesn't irritate the users due to the expected behavior (e.g. irritations due to a talking cat).

Table 2. Different Panda Embodiments

P1	P2	P3	P4
			
teddy bear-like	abstracted, pillow-like	teddy bear-like	realistic, lifelike
variant 1	variant 2	variant 3	variant 4
Likert-Item (1 – 5)	Likert-Item (1 – 5)	Likert-Item (1 – 5)	Likert-Item (1 – 5)
"I like the design of the panda."	"I like the design of the panda."	"I like the design of the panda."	"I like the design of the panda."
$\bar{x}$ = 3.92	$\bar{x}$ = 2.00	$\bar{x}$ = 4.33	$\bar{x}$ = 2.83
SD = 1.51	SD = 1.35	SD = 0.89	SD = 1.59

To this end, the original formfactor was a panda bear. Although pandas do not naturally occur in western countries, they are familiar from zoos and resemble the well-established teddy bears.

Therefore, in the second study with elderly people (n = 12, female = 8, male = 4, age $\bar{x}$ = 76,7) living in a nursing home in southwest Germany, we investigated users' preferences regarding the formfactor of different panda embodiments. The study took place in April 2022 and, like the first presented study, all seniors voluntarily participated without receiving payment and after being chosen by a healthcare professional based on the diagnosis of MCI or dementia.

Methods. After a short introduction video explaining the intention of the study, we showed randomly arranged pictures of four commercially available panda cuddle toys with different design characteristics and form factors to the elderly people. Due to the hygiene and safety measures of the elderly home at that time, only pictures and not the actual cuddle toys were presented. Subsequently, the participants were asked to rate the appearance of these panda bears with the help of a single item ("I like the design of the panda.") on a Likert-scale ranging from 1 (fully disagree) to 5 (fully agree). Images of the panda bears are shown in Table 2.

Results. We performed a Friedman test (Friedman chi-squared = 14.45, df = 3, p < 0.002) and found significant differences between the panda bears' acceptance depicted by Figure 6. Subsequently, paired Wilcoxon tests proved this finding, and we identified a significant difference between P1 and P2 (p < .03), P2 and P3 (p < .005), as well as P3 and P4 (p < .02).

Discussion. As mentioned in the Methods section, a limitation of these findings is whether presenting the actual cuddle toys instead of pictures would have changed the results. Taking the results of the first study into account, the findings, nevertheless, underline our initial assumption of a cultural habituality and familiarity of the typical teddy bear and indicate that a teddy bear-like appearance of a robot companion would be preferred within the target group. Since panda P3 is a comparably big cuddle toy not capable of sitting by itself, we chose panda P1 as feasible embodiment for our robot companion. The actual appearance and the proportions of the companion robot (dimensions: 22 cm x 25 cm x 40 cm) with its soft fur are shown in Figure 3.

4 SYSTEM ARCHITECTURE

Figure 7 shows the overall system architecture consisting of three main parts: The actual robot companion, embodied as the zoomorphic cuddle toy, a mobile app serving as a GUI to manually

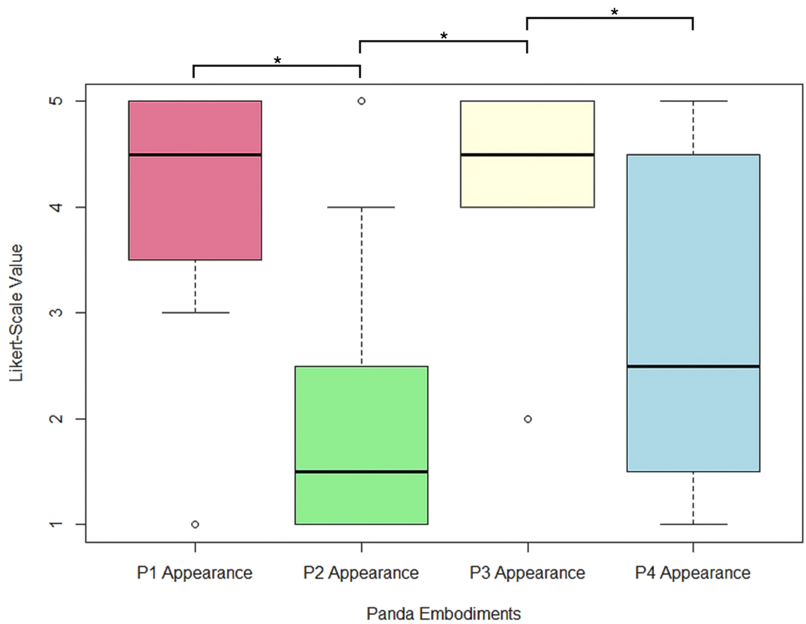

Fig. 6. Boxplot of study results "Appearance of different Embodiments".

start functionalities or change settings, and the cloud services where emotional predictions and HRV metrics are processed and an individual wellness diary is stored.

4.1 Hardware

The architecture core is a Raspberry Pi 4 Model B (Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz; 4GB LPDDR4-3200 SDRAM) running the Python main program and handling the I/O management with the peripheral devices like the PPG sensor and the actuator.

A simple microphone and a speaker required for the Listen and Chat feature are connected to the Raspberry Pi via USB. The PPG sensor is connected to a MCP3008 A/D converter communicating with the GPIO board of the Raspberry Pi. We used a pulse sensor from World Famous Electronics llc. combining an APDS-9008 light photo sensor with a reverse mounted LED emitting light with a wavelength of approximately 565 nm.

This low-cost optical heart-rate sensor is easy to use and well-documented. We designed and 3D-printed a simple casing enclosing the sensor and serving as a mechanical interlocking to keep the finger in place (Figure 8) similar to a pulse oximeter. Specifications of the pulse sensor are shown in Table 3 in the appendix (A.1). To simulate respiration, we used a simple DC motor flange-mounted to a con-rod-like mechanism converting the rotational motor movement into a linear movement of a plastic disc sewed to the inside of the belly fur. An L298N driver is connected upstream of the motor. The overall circuit diagram is shown in Figure 9.

All electrical components except the microphone, the speaker, and the sensor are placed inside a box on which the motor casing with the con-rod-like mechanism is mounted. The box and the motor are surrounded by a roundish polystyrene casing which in turn is surrounded by soft filling and thus only noticeable when the robot is hugged strongly. Figure 10 shows the placement of the hardware components.

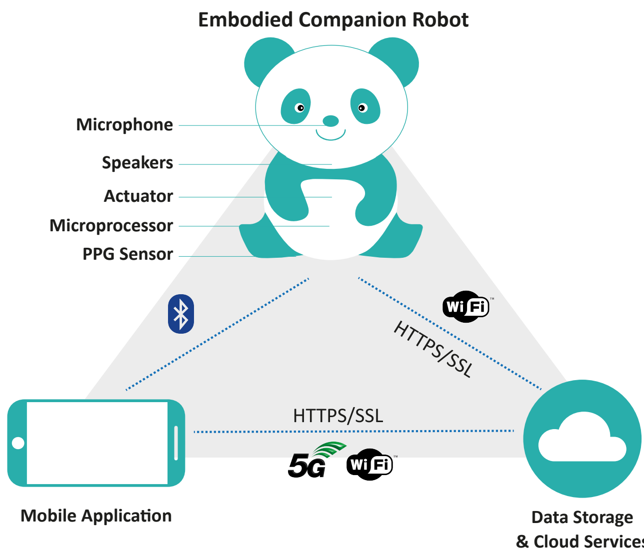

Fig. 7. EZIRC Architecture.

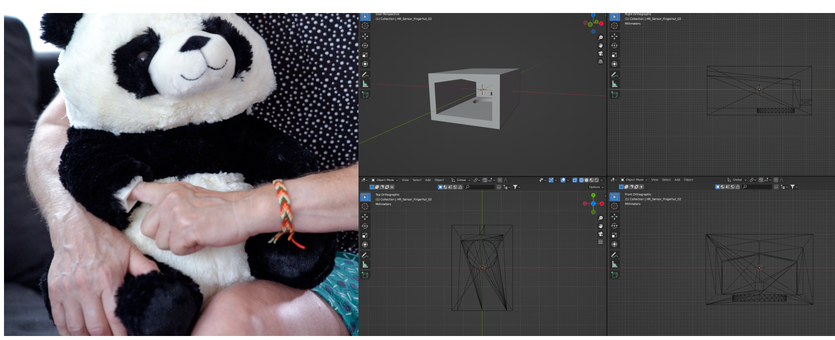

Fig. 8. Sensor casing and mechanical interlocking placed in the paw.

4.2 Software

The Raspberry Pi is running a Python (version 3.10.4) main program controlling all the functionalities described in Section 3.1. A list of the external Python libraries used for this prototype can be found in the appendix A.2. To access the used web APIs, the Raspberry Pi needs a Wi-Fi connection. A "Panda" class is initializing the main program loop including recording the users' speech, sensing the pulse, classifying and storing emotions, and activating the actuator. This main class references five essential classes in turn:

- **EmotionRecognition** class classifying emotions based on audio from speech and text. The MERF including the training process is detailed in the next section (Section 4.2.1).
- **Conversation** class including the dialog flow by using *Google SR*, *ChatGPT*, *Google TTS* and a customized keyword detection. For *ChatGPT* the *GPT3.5* model is used by the writing date of this work.
- **IODevicesControl** class handling the GPIO management to start and stop the actuator.

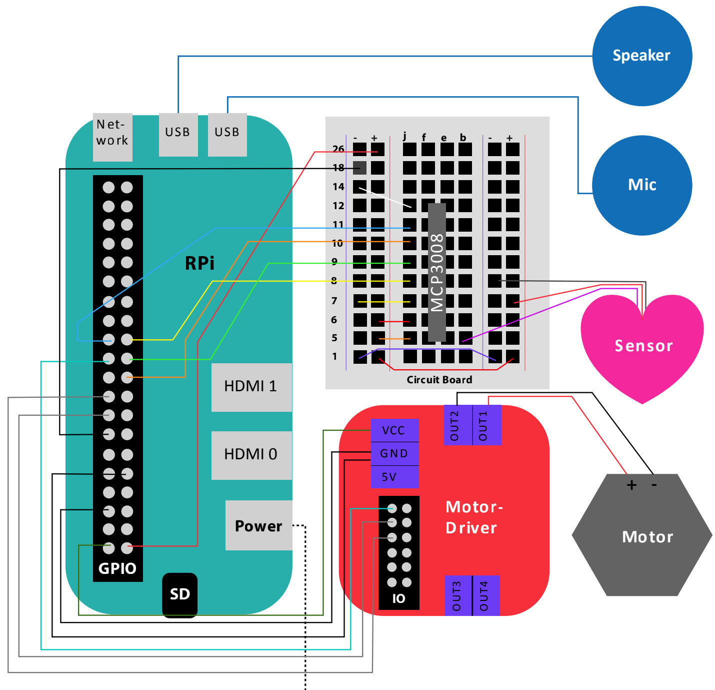

Fig. 9. EZIRC Circuit Diagram.

- **Pulsesensor** class controlling pulse sensor measurement and computing HRV metrics.
- **BLESetup** class managing the Bluetooth Low Energy (BLE) connection with the mobile application.

4.2.1 Multimodal Emotion Recognition Framework. Generally, the idea behind the multimodal emotion recognition framework is combining emotion recognition based on speech and auditory input with text-based sentiment analysis and a reliable and well-known physiological indicator for arousal. To be able to derive suitable system actions based on simple cases (`if emotion e is dominant → appropriate action`), we aimed for six values: four categorical emotion values, each of which can be assigned to a quadrant of Russell's Circumplex model. As well as further two for valence (as an indicator of pleasure representing the x-axis of the Circumplex model) and the degree of arousal (as an indicator of activation and stress representing the y-axis of the Circumplex model). Summing up, the MERF consists of three parts: (1) the categorical emotion detection using SER, (2) the valence detection using sentiment, and (3) an arousal detection using HRV metrics.

Categorical Emotion Detection. The SER makes use of a recurrent neural network (RNN) implemented with *TensorFlow* and is based on prosodic features of audio files found in the frequency spectrum of speech. Specifications of the RNN are as follows:

- batch size: 64
- epochs: 500
- loss: categorical_crossentropy
- optimizer: adam
- dropout: 0.3

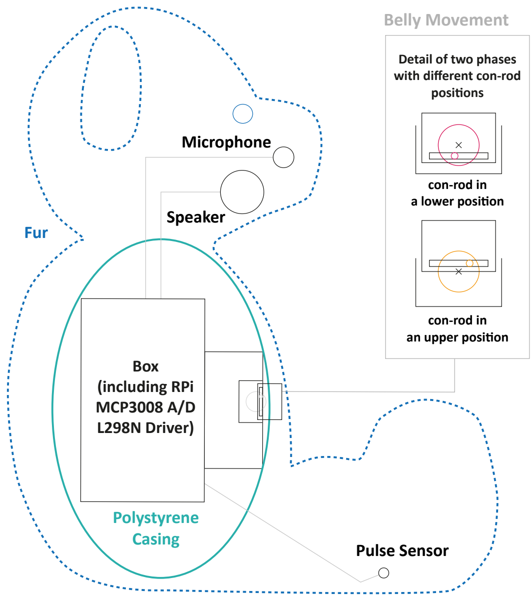

Fig. 10. Placement of Hardware Components.

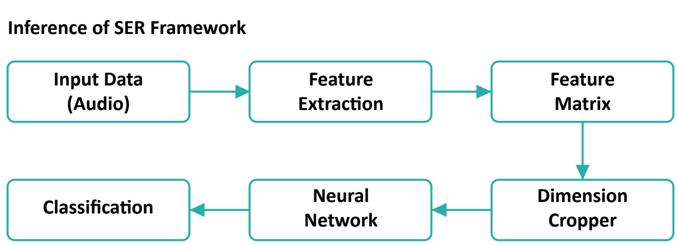

Fig. 11. Inference of the SER.

As a starting point, we used the emodb dataset [14] to train the RNN (80% training data, 20% test data). In the first instance, Mel Frequency Cepstral Coefficients (MFCC) are extracted with the *Python Speech Features* library and assigned to a matrix. Thereby, the dimensionality of the matrix is varying based on the corresponding features (in this case the MFCC) and the length of the recorded audio file. However, this variety is not ideal for the RNN, hence the matrix is cropped in case it's becoming too long. For a matrix that is too short for the dimension cropper, zero padding is used to complete it. Figure 11 visualizes the inference of a single prediction. A single prediction results in values between 0 and 1 for four emotions: happy, sad, angry, and neutral. Thereby, the neutral emotional condition is located in the LAHV quadrant (Section 2.1) in Russell's model and is equivalent to feeling content or calm.

Valence Detection. Google SR is used to transform speech to text and subsequently analyze it with the help of the SentimentIntensityAnalyzer() function of the python library *NLTK(VADER)*.

Subsequently, the compound value of the polarity scores ranging from -1 to 1 is used as an overall valence score of a particular sentiment.

Arousal Detection. Generally, HRV metrics relate to time intervals between consecutive heartbeats or interbeat intervals (IBI). Thereby, time-domain HRV intervals could range from ultra-short-term (seconds) to long-term (hours) and the root mean square of successive differences (RMSSD) is a standard statistical measure for HRV [84]. The RMSSD is obtained by calculating each successive time difference between heartbeats in milliseconds. Subsequently, each of the values is squared and the result is averaged before obtaining the square root of the total [84]. We assume that average interaction time with the robot companion will approximately range from three to fifteen minutes, so we focus on calculating IBI and RMSSD values for rather short intervals of 30 [4] and 60 [28] seconds, respectively. Geovanini et al. conducted a study with 543 healthy participants and found that the average RMSSD value for people aged 60 or above is 40ms [33]. Since RMSSD values and stress are negatively correlated, we reason that RMSSD values lower than 40ms indicate a higher level of arousal and therefore align with an emotional state in the HALV and HAHV quadrants. Likewise, RMSSD higher than 40ms indicate a lower level of arousal aligning with an emotional state in the LALV and LAHV quadrants of the Arousal-Valence model (Section 2.1). The overall arousal value of the MERF equals the users' current RMSSD value.

Deduction of the Current Emotion. Assuming that there is only one dominant emotion at a time, the current emotion derives from the following four cases. In the presented pseudo code $E_{CATEGORICAL}$ equals the predominant emotion of the SER (neutral, happy, angry, sad) and $E_{CURRENT}$ equals the predominant current emotion located in one of the four quadrants of the valence-arousal model (HALV, LALV, HAHV, LAHV).

...

if ($E_{\text{Categorical}}$ = neutral) AND ($valence > 0$) OR ($arousal >$ 40ms)

E_{CURRENT} = LAHV

else if ($E_{\text{Categorical}}$ = happy) AND ($valence > 0$) OR ($arousal <$ 40ms)

E_{CURRENT} = HAHV

else if ($E_{\text{Categorical}}$ = sad) AND ($valence < 0$) OR ($arousal >$ 40ms)

E_{CURRENT} = LALV

else if ($E_{\text{Categorical}}$ = angry) AND ($valence < 0$) OR ($arousal <$ 40ms)

E_{CURRENT} = HALV

...

Accordingly, the current emotion is defined if the emotion prediction with the highest percentage value ($0 \sim 1$) coming from the speech emotion recognition aligns with the corresponding valence or arousal value. As discussed earlier, the goal of the MERF is to increase the accuracy of emotion recognition. Therefore, the derivation shown was chosen because it can be assumed that two otherwise independent recognition systems that provide the same tendency with respect to the valence-arousal quadrant are sufficient indication for the presence of a certain stress level and well-being, respectively.

Affective Measures and Actions. While interacting with the companion robot the user is assigned to one of the four emotions (neutral, happy, angry, sad) during runtime. Each emotion corresponds to one of the four quadrants in the segmented valence-arousal model: neutral equals LAHV, happy equals HAHV, angry equals HALV, and sad equals LALV. As described in Section 2.1, the user's

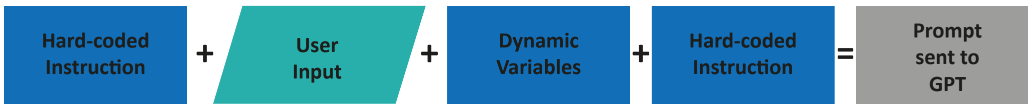

Fig. 12. Prompt sent to the GPT API consisting of different static and dynamic strings.

emotional state should be predominantly located in the two quadrants (HAHV and LAHV) to the right of the y-axis of the valence-arousal model. Emotional states located in the two quadrants (HALV and LALV) to the left of the y-axis should be addressed by a specific measure, e.g. a conversation about the reasons of this emotional state or an appropriate action like a meditation or breathing-exercise bringing the user's emotional state back more towards the two right quadrants.

Based on the current emotional state, four corresponding measures are therefore initiated by the software:

(1) Longer periods of $E_{CURRENT}$ = LAHV initialize a task that cognitively stimulates the user. In our case it's a quiz game.
(2) Longer periods of $E_{CURRENT}$ = HAHV initialize a mindful conversation about the reasons for the current emotional state. This is done to create awareness for the "good" things.
(3) Longer periods of $E_{CURRENT}$ = HALV initialize the Cuddle feature including a guided breathing-exercise to calm down the user.
(4) Longer periods of $E_{CURRENT}$ = LALV initialize conversation asking about the reasons for the current emotional state. This is done to create awareness for the things that might depress the user.

On a continuum of well-being, the different quadrants are either considered more unpleasant (LALV, HALV) or more pleasant (LAHV, HAHV). Although melancholy (associated with the LALV quadrant) can be experienced as a "beautiful" emotional state by some people, from a general perspective, longer periods of an emotional state in the LALV quadrant are not desirable. Therefore, the measures and robot actions have been consciously chosen to either improve the current emotional state of the user by trying to shift $E_{CURRENT}$ from the two left quadrants in one of the two right ones (measures 3 and 4) or to maintain the more desirable emotional states in the right quadrants (measures 1 and 2). The latter shall be achieved by either cognitively stimulating the user (measure 1) and shift $E_{CURRENT}$ towards the HAHV quadrant or by reflecting the current emotional state in terms of mindfulness (measure 2) and to maintain $E_{CURRENT}$ in the HAHV quadrant.

On the Use of GPT. A characteristic of ChatGPT is the fact that output results strongly depend on the input prompt. Apart from general API parameters that can be adjusted to create certain behaviors of the GPT model, the "system" can be directed to react in a specific way or with specific "personality" with the help of a prompt at the beginning of a single chat (API request). Such an initial prompt is also referred to as a system prompt. As an example, an initial prompt could be:

"Behave as "Beffi", the friendly companion robot for elderly people with MCI or dementia, embodied as a cute panda cuddle toy."

In addition, a single prompt sent to the API may consist of several different pieces of text that are first joined together before being sent to GPT. This allows for prompts consisting of the actual user input as well as further hard-coded instructions or instructions based on variables coming from dynamic code parts. Figure 12 depicts such an engineered prompt.

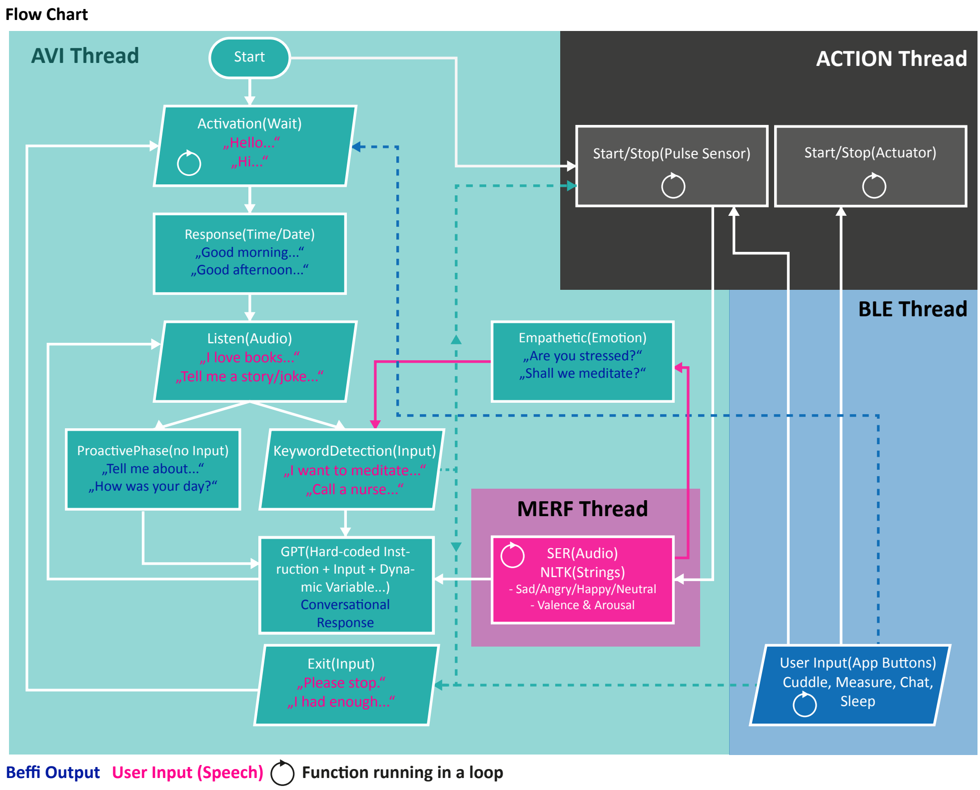

Fig. 13. Flow Chart of the main program.

We use this method to enrich the prompt with the predicted emotional state of the user and to give GPT a sense of empathy. In the following section, we detail how the current emotional state coming from the MERF is used to directly affect the conversation flow.

4.2.2 Software Flow and Threading. Since different functions are used in a parallel manner and diverse user input is checked through independent loops, we used multithreading to achieve concurrency and to avoid blocking code by longer processes. Figure 13 depicts the overall flow chart and the four different threads. All threads are initialized as soon as the *Panda* class started the main program loop.

The BLE-thread establishes BLE connection between the Raspberry PI and the smartphone running the mobile app and continuously checks for user input through the app (e.g. the user presses the button "Cuddle" to manually start the actuator). As mentioned before, the functionalities *Cuddle*, *Measure*, and *Chat* can be manually started independent from the dialog flow. Moreover, the BLE-thread updates the start and stop values in the Panda class for *Chat*, *Measure*, *Cuddle*, and *Sleep*.

The MERF-thread is running a `SER()` function that constantly uses new audio files created from voice input for emotion prediction as well as the **SentimentIntensityAnalyzer()** function from the *NLTK(VADER)* library as described in the previous section (Section 4.2.1). Subsequently, a string variable of the conversation class holding the predominant emotional state based on the derivation described in Section 4.2.1 is updated.

The AVI-thread contains seven linearly arranged functions and one additional function working in parallel. `Activation()` is a function (loop) constantly waiting for a keyword like “hello”, “hi”, or “good morning”. If triggered `Activation()` leads to the next step in the flow. Additionally, `Activation()` can be triggered via the BLE thread with the help of the Chat button within the mobile app. This loop is reentered after a conversation with Beffi is finished. `Response()` is a function triggering the first answer in a conversation with Beffi depending on date and time. This hard-coded response is always starting the conversation (“Good Morning, good to see you. Let’s talk a little bit.”).

Due to the consideration of date and time, the `Response()` function provides a sense of individualization, topicality, and awareness of the current surrounding. After the first response `Listen()`, a function recording audio and creating strings using Google SR, is activated. If no voice input by the user is detected for a certain time, `Listen()` returns “None” and `ProactivePhase()` function is triggered, sending a hard-coded prompt like “Ask me to tell you something about myself” or “Ask me, how my day has been so far”.

If voice input is detected, the resulting string is analyzed by a `KeywordDetection()` function on trigger words like “yes”, “no”, “stop”, “help” etc. If triggered, this function skips the forwarding of the string to GPT and allows for custom functionalities or actions. For instance, a user could immediately activate the breathing exercise by saying “I want to meditate”. If `KeywordDetection()` is not triggered, the `GPT()` function constructs a prompt as described in Section 4.2.1, sends the input string to the OpenAI web API (model GPT3.5), and receives a response. Based on the prompt engineering approach described in the previous section, GPT receives an initial hard-coded system instruction in the first iteration of the dialog flow:

Behave as "Beffi", the friendly companion robot for elderly people with MCI or dementia, embodied as a cute panda cuddle toy. Never say anything bad or depressing. Always try to cheer me up. Start to chat with me by simply responding "How do you feel?"

In the second iteration, GPT receives a prompt constructed of three parts:

(1) A hard-coded instruction (*“In the following, you will always get my emotional state as a single word first. React accordingly in a sensitive way, but don’t name the emotion stated. Ask why I’m feeling this way.”*)
(2) A string variable holding the recent emotional state coming from the MERF thread, e.g. “angry” or $E_{CURRENT}$ = HALV
(3) The actual user input coming from the Conversation class using Google SR and Google TTS.

From the third iteration with GPT, the prompt consists only of the current emotional state and the actual user input. The number of iterations with GPT is influenced by `Empathetic()`, a function that is constantly fed with emotional values by the MERF thread. The function replaces the conversational responses of GPT after a predetermined number of rounds. This process reflects the current emotional state of the user and allows for an appropriate measure as described in the previous section. Hence, if the predominant emotional result is $E_{CURRENT}$ = HALV , the suitable measure, in this case the breathing exercise, is triggered by the Action-thread.

In this case, the companion robot asks to start the breathing exercise and the corresponding `Cuddle()` function in the Action-thread after the user confirmed the current prediction result (Beffi: “It seems that you’re stressed.” → User: “Yes, indeed.” → Beffi: “Shall we do a breathing exercise together?”). If the user does not confirm the result, e.g. by saying “no”, the `ProactivePhase()` function is triggered to enter another conversational round and keep the dialog ongoing. In addition, a maximum number limiting the iterations with GPT is set by a variable in case the predominant

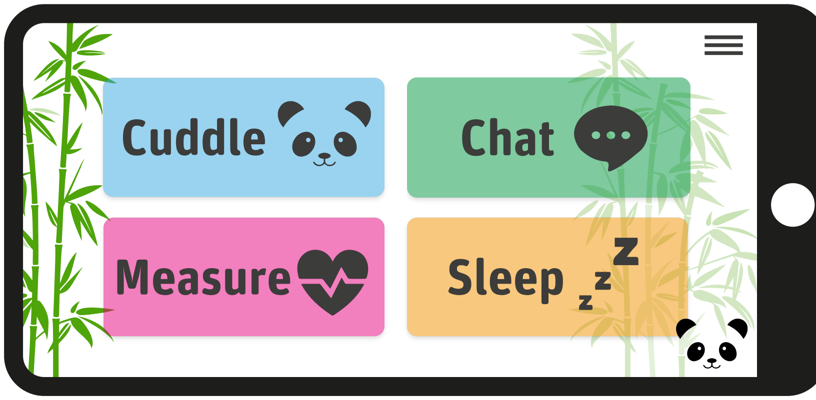

Fig. 14. Home Screen of the Mobile App.

prediction result is "neutral" or $E_{CURRENT}$ = LAHV over a longer period. Eventually `Exit()` is a function either triggered by a keyword or user input within the mobile app (pressing the Chat button again) in order to quit the dialog flow and stop at a predetermined endpoint.

The Action-thread handles the measurement and cuddle (co-breathing exercise or meditation) actions and includes a `Start/Stop()` function, once activated, running in a loop. To be able to constantly get the users' arousal values, the `Measure()` function is running in a loop as soon as the Action-thread is started. Moreover, this function is either stopped or started again when the `Measure()` function is initialized by the AVI thread or by pressing the corresponding button Measure within the mobile app. The cuddle function is not a loop but controlled by the driver status. The driver is simply activated by closing the corresponding circuit (GPIO Pinout == HIGH).

4.2.3 Mobile App. The mobile app is designed to work with a Bluetooth Low Energy (BLE) device. When a user connects to BLE, the app detects all services the two entities can communicate with. Once the device is connected, the user is shown a sleep mode screen. At this point the individual device ID is sent to the Raspberry Pi. On a home screen (Figure 14), the main functionalities *Chat*, *Measure*, and *Cuddle* can be activated and deactivated.

When the user presses a button, a corresponding signal is sent to the connected Raspberry Pi via BLE. The app also supports two-way communication between the mobile and the BLE device. This functionality allows the app to receive information from the Raspberry Pi as well as send information to the Raspberry Pi to allow the main functions to be triggered either by voice or the buttons on the mobile app. Furthermore, when *Measure* is started, the real time value of the heart-rate sensor is read and displayed in a dialog box on the app.

Once the dialog flow of *Chat* is stopped, all values are uploaded as a CSV file to the cloud storage for further processing. Vice versa, the mobile app displays aggregated SER and HRV metrics by fetching the data from the cloud storage. Moreover, a `Writewifi()` function is used to allow users to switch the wireless network of the Raspberry Pi with the help of the mobile app. Since the Raspberry Pi is used headless, the mobile app is the only GUI for both user groups: elderly people and their caregivers. The overall flowchart of the mobile app is shown in Figure 15.

4.2.4 Structure of Cloud Storage and Database. Currently, ***Google Firebase*** is used as a cloud service and the storage is arranged according to the device IDs. For each device ID, there are three

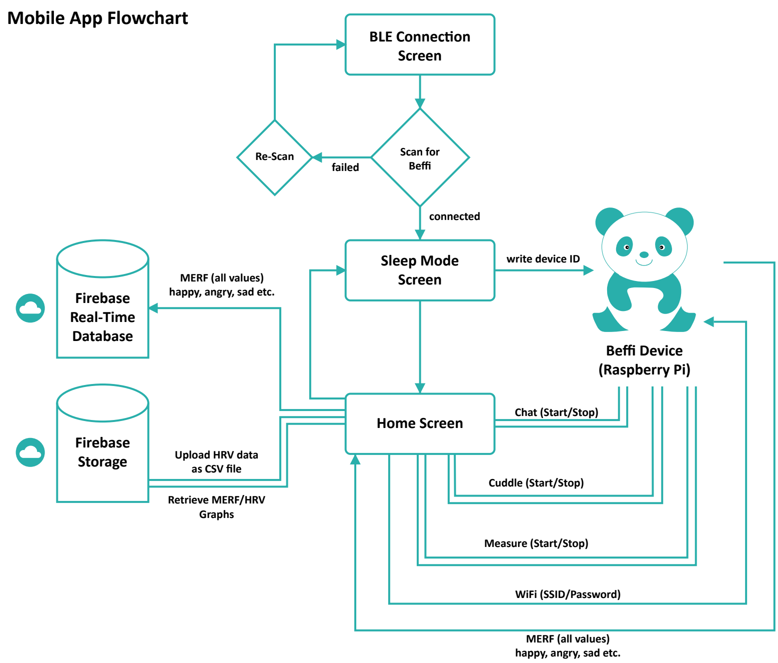

Fig. 15. Flowchart of the Mobile App.

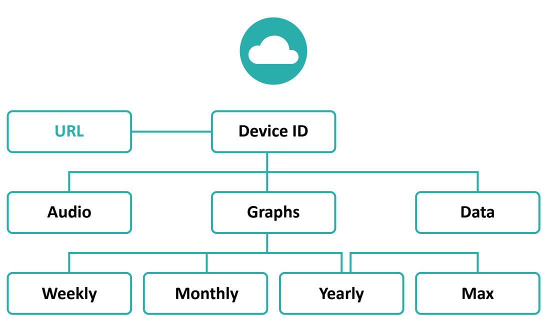

Fig. 16. Structure of the Cloud Storage.

main directories for audio files, graphs, and emotion and HRV data files. The graph files are further arranged according to week, month and year (Figure 16).

The real time database also holds the results of the MERF (Figure 17). Similarly, this database is structured based on the device ID. For every device ID, the predicted values of emotions “happy”, “neutral”, “sad”, and “angry” as well as values for arousal and valence are stored. These data is subsequently used to generate aggregated graphs monitoring emotional trends over time.
Authentication is required before any data is communicated back and forth between the cloud storage and the Raspberry Pi. Therefore, the device ID used for the data storage is based on the actual mobile phone the app is running on. The device ID generated by the mobile phone is subsequently

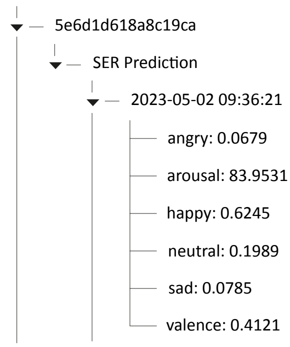

Fig. 17. Entry of MERF data in the real-time data base for a particular ID.

sent back to the Raspberry Pi and stored in the configuration file for use in case the mobile phone is out of range. Firebase also provides offline capabilities to cache the data locally and upload it later when network connectivity is re-established.

4.2.5 Data Visualization. In general, the data visualization consists of two parts: SER and HRV data. After fetching the data from the Google Firebase server, a Python script plots the graphs. For this process, the NumPy and Pandas libraries are used for data analysis, and the Matplotlib library is used for plotting.

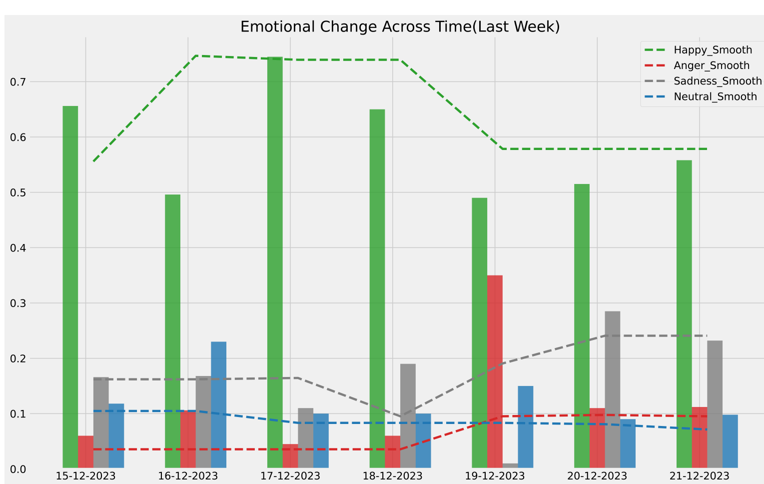

Fig. 18. Visualization of fictional weekly SER data.

The SER visualizations are supposed to show emotional variation as a function of time within weekly, monthly, and yearly intervals. Hence, in a first step, the fetched data is aggregated based on the corresponding time interval. For instance, daily aggregation is used for a weekly interval showing the last seven days. The same logic applies to the other time intervals. Each emotion is represented with a color code. Simple plots and histograms are used to visualize changing emotional patterns (Figure 18). These visualizations of the data are only intended for therapists and clinicians.

5 DISCUSSION

In the previous sections we described the development of our EZIRC prototype named "Beffi". Our goal is to improve and fine-tune this prototype in further steps. Regardless of these developments, we want to emphasize some limitations.

The studies with elderly people (Section 3.2) indicated a high acceptance of a teddy bear-like robot companion embodiment and the aspired feature set. However, since there was no interaction with the actual robot companion yet, it is unclear if elderly people appreciated the talking panda. Moreover, our recent prototype neither provides facial expression or lip sync, nor eye movement. This might lead to leeriness and an uncanny experience. Apart from the standardized Likert-items, we did not collect any qualitative data which could have enriched the value of the studies.

Regarding the usage of *ChatGPT*, possible ethical concerns [15, 99] must be considered. Simply using the Open AI API is limited to the settings and parameters provided by the API itself. To address this "black-box" phenomenon, the keyword detection and the `Empathetic()` function leading to predefined answers and actions is logically upstream GPT.

In addition, the usage of several web APIs sometimes leads to non-manipulable delays that can disrupt the dialog flow. However, actual interaction between the companion robot and elderly people may unveil delays as neglectable, especially for elderly people who generally speak slowly and appreciate small pauses in dialogues.

We used the EMO-DB dataset to train the RNN which leads to several limitations. First, possible differences in the elderly's speech patterns such as a different accent, slower speech, or a different vocabulary could affect the accuracy of the predictions. These speech patterns might differ significantly from the patterns of the actors in the EMO-DB dataset.

Generally, a user test with older adults to assess the accuracy of emotional detection with this user typology should be conducted. A possible solution to avoid overfitting and to make the model more robust for outliers is to include more diverse data, such as the Mozilla Common Voice database.

Regarding HRV metrics of the user, we set a threshold of 40ms (RMSSD above 40ms indicate a lower level of arousal, lower than 40ms indicate a higher level of arousal) as a starting point to map the metric onto the Arousal-Valence model (Section 2.1). This threshold is based on an average RMSSD values for healthy people aged 60 or above [33]. Since HRV is highly dependent from individual factors and pre-existing conditions, this hypothetical assumption should be further evaluated and verified.

Assuming that there is only one predominant emotion at a time, arousal values may be computed by distinguishing four cases and based on the logic assumption that speech identified as angry or sad corresponds with negative valence scores of the SA and speech identified as neutral or happy corresponds with positive valence scores:

- $|\text{angry}| > \text{happy, sad, neutral} \longrightarrow |\text{valence} - \text{angry}| = \text{arousal}$
- $|\text{sad}| > \text{happy, angry, neutral} \longrightarrow |\text{valence} - \text{sad}| = \text{arousal}$
- $|\text{happy}| > \text{angry, sad, neutral} \longrightarrow |\text{valence} + \text{happy}| = \text{arousal}$
- $|\text{neutral}| > \text{happy, angry, sad} \longrightarrow |\text{valence} + \text{neutral}| = \text{arousal}$

This leads to *arousal* values ranging from -2 to 2. Doubtless, the calculation of *arousal* is a hypothetical approach and not impeccable. However, since heart rate is measured as an additional physiological indicator and subsequently used to compute HRV metrics, *arousal* can be utilized as a quality measure of the MERF at a later point. Hence, possible correlations between *arousal* and RMSSD values conversely indicate that emotional predictions of the SER and SA might be acceptable. These correlations allow for conclusions to be drawn about the overall robustness of this multimodal approach without including laborious user validation of emotions. Since it is commonly known that people experience difficulties when evaluating their own emotions, such an approach yields great potential to create reliable and robust emotion recognition frameworks.

The data currently collected by the prototype is stored anonymously and does not allow analyzing individual persons. However, collecting sensitive data, data privacy standards complying with the European strategy for data [20], the European Data Governance Act [19], and the European Health Data Space [21] should be considered for a deployment in real-life scenarios. This includes the utilization of data encryption while using BLE. Generally, these concerns may also apply to different web APIs such as *Google SR*.

Finally, we focused on using low-cost hardware components to meet the need for affordable IAT. However, reliability of this hardware (such as the pulse sensor) must be evaluated and compared with professional solutions before they can be applied in real-life scenarios.

6 CONCLUSION

In this work we outlined the potentials of zoomorphic robot companions combining intelligent conversational abilities and computational emotion recognition. Moreover, we described in detail the design and development of such a low-cost prototype robot and the implementation of a multimodal emotion recognition framework using speech emotion recognition and sentiment analysis as well as an affective voice interaction feature. Especially, we described the system architecture and the way spoken language can be processed both semantically and on an affective level. Thereby, we enriched an NLP architecture using *Google SR* and *GPT* with a SER framework, sentiment analysis, and physiological emotion recognition. This approach shows how *Open AIs' GPT*, one of the most advanced language models, can be included to provide semantically unrestricted conversations but limit the overall conversational flow with an upstream keyword detection and SER to prevent the conversation going into the wrong direction.

Likewise, we illustrated how physiological activation can be computed based on dimensional and categorical affect analysis and subsequently verified with the help of HRV metrics. The prototyping has been accompanied by two studies with elderly people on various embodiments to find the appropriate casing for the robot companion. The results indicate a high acceptance of teddy bear-like embodiments and appreciated features like intelligent verbal communication and emotion recognition. Moreover, two cases of a user evaluation showed that the companion robot was able to give space to feelings such as joy or deep emotions that are otherwise only shared with very familiar people or professional therapists.

In future work, this prototype will be used in user studies to gather more practical experiences in the field. Ideally, the newly gained knowledge will then lead to further enhancements of the prototype. Likewise, we want to encourage future research endeavors to include the presented approaches and architectures.

REFERENCES

[1] Jordan Abdi, Ahmed Al-Hindawi, Tiffany Ng, and Marcela P Vizcaychipi. 2018. Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open* 8, 2 (Feb. 2018), e018815. https://doi.org/10.1136/bmjopen-2017-018815

[2] Hojjat Abdollahi, Ali Mollahosseini, Josh T. Lane, and Mohammad H. Mahoor. 2017. A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. IEEE, Birmingham, 541–546. https://doi.org/10.1109/HUMANOIDS.2017.8246925

[3] Patrícia Alves-Oliveira, Tiago Ribeiro, Sofia Petisca, Eugenio Di Tullio, Francisco S. Melo, and Ana Paiva. 2015. An Empathic Robotic Tutor for School Classrooms: Considering Expectation and Satisfaction of Children as End-Users. In *Social Robotics*, Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi (Eds.). Vol. 9388. Springer International Publishing, Cham, 21–30. https://doi.org/10.1007/978-3-319-25554-5_3

[4] Hyun Jae Baek, Chul-Ho Cho, Jaegeol Cho, and Jong-Min Woo. 2015. Reliability of Ultra-Short-Term Analysis as a Surrogate of Standard 5-Min Analysis of Heart Rate Variability. *Telemedicine and e-Health* 21, 5 (May 2015), 404–414. https://doi.org/10.1089/tmj.2014.0104

[5] AmirAli Bagher Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2018. Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 2236–2246. https://doi.org/10.18653/v1/P18-1208

[6] Melis Yilmaz Balban, Eric Neri, Manuela M. Kogon, Lara Weed, Bita Nouriani, Booil Jo, Gary Holl, Jamie M. Zeitzer, David Spiegel, and Andrew D. Huberman. 2023. Brief structured respiration practices enhance mood and reduce physiological arousal. *Cell Reports Medicine* 4, 1 (Jan. 2023), 100895. https://doi.org/10.1016/j.xcrm.2022.100895

[7] Timothy W. Bickmore, Lisa Caruso, Kerri Clough-Gorr, and Tim Heeren. 2005. 'It's just like you talk to a friend' relational agents for older adults. *Interacting with Computers* 17, 6 (Dec. 2005), 711–735. https://doi.org/10.1016/j.intcom.2005.09.002

[8] Kari Blindheim, Mads Solberg, Ibrahim A. Hameed, and Rigmor Einang Alnes. 2023. Promoting activity in long-term care facilities with the social robot Pepper: a pilot study. *Informatics for Health and Social Care* 48, 2 (April 2023), 181–195. https://doi.org/10.1080/17538157.2022.2086465

[9] Hannah Bradwell, Rhona Winnington, Serge Thill, and Ray B. Jones. 2021. Prioritising Design Features for Companion Robots Aimed at Older Adults: Stakeholder Survey Ranking Results. In *Social Robotics*, Haizhou Li, Shuzhi Sam Ge, Yan Wu, Agnieszka Wykowska, Hongsheng He, Xiaorui Liu, Dongyu Li, and Jairo Perez-Osorio (Eds.). Vol. 13086. Springer International Publishing, Cham, 774–779. https://doi.org/10.1007/978-3-030-90525-5_70

[10] Hannah Louise Bradwell, Katie Edwards, Deborah Shenton, Rhona Winnington, Serge Thill, and Ray B Jones. 2021. User-Centered Design of Companion Robot Pets Involving Care Home Resident-Robot Interactions and Focus Groups With Residents, Staff, and Family: Qualitative Study. *JMIR Rehabilitation and Assistive Technologies* 8, 4 (Nov. 2021), e30337. https://doi.org/10.2196/30337

[11] Hannah L. Bradwell, Rhona Winnington, Serge Thill, and Ray B. Jones. 2020. Longitudinal Diary Data: Six Months Real-world Implementation of Affordable Companion Robots for Older People in Supported Living. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Cambridge United Kingdom, 148–150. https://doi.org/10.1145/3371382.3378256

[12] J. Broekens, M. Heerink, and H. Rosendal. 2009. Assistive social robots in elderly care: a review. *Gerontechnology* 8, 2 (April 2009), 94–103. https://doi.org/10.4017/gt.2009.08.02.002.00

[13] Maddalena Bruscoli and Simon Lovestone. 2004. Is MCI really just early dementia? A systematic review of conversion studies. *International Psychogeriatrics* 16, 2 (June 2004), 129–140. https://doi.org/10.1017/S1041610204000092

[14] Felix Burkhardt, A. Paeschke, M. Rolfes, Walter F. Sendlmeier, and Benjamin Weiss. 2005. A database of German emotional speech. In *Interspeech 2005*. ISCA, 1517–1520. https://doi.org/10.21437/Interspeech.2005-446

[15] Anastasia Chan. 2023. GPT-3 and InstructGPT: technological dystopianism, utopianism, and "Contextual" perspectives in AI ethics and industry. *AI and Ethics* 3, 1 (Feb. 2023), 53–64. https://doi.org/10.1007/s43681-022-00148-6

[16] Edward D. Chan, Michael M. Chan, and Mallory M. Chan. 2013. Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations. *Respiratory Medicine* 107, 6 (June 2013), 789–799. https://doi.org/10.1016/j.rmed.2013.02.004

[17] Jeanie Chan and Goldie Nejat. 2012. Social Intelligence for a Robot Engaging People in Cognitive Training Activities. *International Journal of Advanced Robotic Systems* 9, 4 (Oct. 2012), 113. https://doi.org/10.5772/51171

[18] Carlos A. Cifuentes, Maria J. Pinto, Nathalia Céspedes, and Marcela Múnera. 2020. Social Robots in Therapy and Care. *Current Robotics Reports* 1, 3 (Sept. 2020), 59–74. https://doi.org/10.1007/s43154-020-00009-2

[19] European Commision. 2022. European Data Governance. https://digital-strategy.ec.europa.eu/en/policies/data-governance-act publisher: European Commision.

[20] European Commision. 2022. A European Strategy for data. https://digital-strategy.ec.europa.eu/en/policies/strategy-data publisher: European Commision.

[21] European Commision. 2023. European Health Data Space. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en publisher: European Commision.

[22] Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults, Board on Health Sciences Policy, Board on Behavioral, Cognitive, and Sensory Sciences, Health and Medicine Division, Division of Behavioral and Social Sciences and Education, and National Academies of Sciences, Engineering, and Medicine. 2020. *Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System*. National Academies Press, Washington, D.C. https://doi.org/10.17226/25663

[23] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine* 18, 1 (Jan. 2001), 32–80. https://doi.org/10.1109/79.911197

[24] Enrique De La Cal, Javier Sedano, Alberto Gallucci, and Paloma Valderde. 2023. A Comparison of Two Speech Emotion Recognition Algorithms: Pepper Humanoid Versus Bag of Models. In *17th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2022)*, Pablo García Bringas, Hilde Pérez García, Francisco Javier Martinez-de Pison, José Ramón Villar Flecha, Alicia Troncoso Lora, Enrique A. De La Cal, Álvaro Herrero, Francisco Martínez Álvarez, Giuseppe Psaila, Héctor Quintián, and Emilio S. Corchado Rodriguez (Eds.). Vol. 531. Springer Nature Switzerland, Cham, 635–644. https://doi.org/10.1007/978-3-031-18050-7_62

[25] Birthe Dinesen, Helle Kidde Hansen, Gry Bruun Grønborg, Anne-Kirstine Dyrvig, Sofie Dalskov Leisted, Henrik Stenstrup, Cathrine Skov Schacksen, and Claus Oestergaard. 2022. Use of a Social Robot (LOVOT) for Persons With Dementia: Exploratory Study. *JMIR Rehabilitation and Assistive Technologies* 9, 3 (Aug. 2022), e36505. https://doi.org/10.2196/36505

[26] Verena Distler, Carine Lallemand, and Thierry Bellet. 2018. Acceptability and Acceptance of Autonomous Mobility on Demand: The Impact of an Immersive Experience. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–10. https://doi.org/10.1145/3173574.3174186

[27] Andrius Dzedzickis, Artūras Kaklauskas, and Vytautas Bucinskas. 2020. Human Emotion Recognition: Review of Sensors and Methods. *Sensors* 20, 3 (Jan. 2020), 592. https://doi.org/10.3390/s20030592

[28] Michael R. Esco and Andrew A. Flatt. 2014. Ultra-Short-Term Heart Rate Variability Indexes at Rest and Post-Exercise in Athletes: Evaluating the Agreement with Accepted Recommendations. *Journal of Sports Science & Medicine* 13, 3 (Sept. 2014), 535–541. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4126289/

[29] European Commission. Directorate General for Employment, Social Affairs and Inclusion. 2021. *Long-term care report: trends, challenges and opportunities in an ageing society. Volume I.* Publications Office, LU. https://data.europa.eu/doi/10.2767/677726

[30] Stephen Fairclough and Kiel Gilleade. 2012. Construction of the biocybernetic loop: a case study. In *Proceedings of the 14th ACM international conference on Multimodal interaction*. ACM, Santa Monica California USA, 571–578. https://doi.org/10.1145/2388676.2388797

[31] D. Feil-Seifer and M.J. Mataric. 2005. Socially Assistive Robotics. In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.* IEEE, Chicago, IL, USA, 465–468. https://doi.org/10.1109/ICORR.2005.1501143

[32] Centers for Disease Control and Prevention. 2021. Loneliness and Social Isolation Linked to Serious Health Conditions. https://www.cdc.gov/aging/publications/features/lonely-older-adults.html

[33] Glaucylara Reis Geovanini, Enio Rodrigues Vasques, Rafael De Oliveira Alvim, José Geraldo Mill, Rodrigo Varejão Andreão, Bruna Kim Vasques, Alexandre Costa Pereira, and Jose Eduardo Krieger. 2020. Age and Sex Differences in Heart Rate Variability and Vagal Specific Patterns – Baependi Heart Study. *Global Heart* 15, 1 (Oct. 2020), 71. https://doi.org/10.5334/gh.873

[34] Moojan Ghafurian, Jesse Hoey, and Kerstin Dautenhahn. 2021. Social Robots for the Care of Persons with Dementia: A Systematic Review. *ACM Transactions on Human-Robot Interaction* 10, 4 (Dec. 2021), 1–31. https://doi.org/10.1145/3469653

[35] H.-M. Gross, Ch. Schroeter, S. Mueller, M. Volkhardt, E. Einhorn, A. Bley, Ch. Martin, T. Langner, and M. Merten. 2011. Progress in developing a socially assistive mobile home robot companion for the elderly with mild cognitive impairment. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2430–2437. https://doi.org/10.1109/IROS.2011.6094770 ISSN: 2153-0866.

[36] Susel Góngora Alonso, Sofiane Hamrioui, Isabel De La Torre Díez, Eduardo Motta Cruz, Miguel López-Coronado, and Manuel Franco. 2019. Social Robots for People with Aging and Dementia: A Systematic Review of Literature. *Telemedicine and e-Health* 25, 7 (July 2019), 533–540. https://doi.org/10.1089/tmj.2018.0051

[37] Mubin Ul Haque, Isuru Dharmadasa, Zarrin Tasnim Sworna, Roshan Namal Rajapakse, and Hussain Ahmad. 2022. "I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data. (2022). https://doi.org/10.48550/ARXIV.2212.05856

[38] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2008. The influence of social presence on acceptance of a companion robot by older people. *Journal of Physical Agents (JoPha)* 2, 2 (2008), 33–40. https://doi.org/10.14198/JoPha.2008.2.2.05

[39] Julianne Holt-Lunstad, Timothy B. Smith, Mark Baker, Tyler Harris, and David Stephenson. 2015. Loneliness and Social Isolation as Risk Factors for Mortality: A Meta-Analytic Review. *Perspectives on Psychological Science* 10, 2 (March 2015), 227–237. https://doi.org/10.1177/1745691614568352

[40] Susan I. Hopper, Sherrie L. Murray, Lucille R. Ferrara, and Joanne K. Singleton. 2019. Effectiveness of diaphragmatic breathing for reducing physiological and psychological stress in adults: a quantitative systematic review. *JBI Database of Systematic Reviews and Implementation Reports* 17, 9 (Sept. 2019), 1855–1876. https://doi.org/10.11124/JBISRIR-2017-003848

[41] Marcello Ienca, Jotterand Fabrice, Bernice Elger, Maurizio Caon, Alessandro Scoccia Pappagallo, Reto W. Kressig, and Tenzin Wangmo. 2017. Intelligent Assistive Technology for Alzheimer's Disease and Other Dementias: A Systematic

Review. *Journal of Alzheimer's Disease* 56, 4 (Feb. 2017), 1301–1340. https://doi.org/10.3233/JAD-161037
[42] M. Ienca, M. Lipps, T. Wangmo, F. Jotterand, B. Elger, and R.W. Kressig. 2018. Health professionals' and researchers' views on Intelligent Assistive Technology for psychogeriatric care. *Gerontechnology* 17, 3 (Oct. 2018), 139–150. https://doi.org/10.4017/gt.2018.17.3.002.00
[43] Yuhei Ikeda, Ryota Horie, and Midori Sugaya. 2017. Estimating Emotion with Biological Information for Robot Interaction. *Procedia Computer Science* 112 (2017), 1589–1600. https://doi.org/10.1016/j.procs.2017.08.198
[44] PARO Robots Inc. 2003. PARO Therapeutic Robot. http://parorobots.com/?utm_source=robots.ieee.org publisher: PARO Robots Inc..
[45] Tombot Inc. [n. d.]. Meet Jennie - Tombot Puppies - Robotic Emotional Support Animals. https://tombot.com/pages/meet-our-puppies publisher: Tombot Inc..
[46] Syem Ishaque, Naimul Khan, and Sri Krishnan. 2021. Trends in Heart-Rate Variability Signal Analysis. *Frontiers in Digital Health* 3 (Feb. 2021), 639444. https://doi.org/10.3389/fdgth.2021.639444
[47] Reza Kachouie, Sima Sedighadeli, Rajiv Khosla, and Mei-Tai Chu. 2014. Socially Assistive Robots in Elderly Care: A Mixed-Method Systematic Literature Review. *International Journal of Human-Computer Interaction* 30, 5 (May 2014), 369–393. https://doi.org/10.1080/10447318.2013.873278
[48] Hee Sun Kang, Kiyoko Makimoto, Rie Konno, and In Soon Koh. 2020. Review of outcome measures in PARO robot intervention studies for dementia care. *Geriatric Nursing* 41, 3 (May 2020), 207–214. https://doi.org/10.1016/j.gerinurse.2019.09.003
[49] Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. 2023. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications* 82, 3 (Jan. 2023), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4
[50] Wei Qi Koh, Faith Xin Hui Ang, and Dympna Casey. 2021. Impacts of Low-cost Robotic Pets for Older Adults and People With Dementia: Scoping Review. *JMIR Rehabilitation and Assistive Technologies* 8, 1 (Feb. 2021), e25340. https://doi.org/10.2196/25340
[51] Wei Qi Koh, Elaine Toomey, Aisling Flynn, and Dympna Casey. 2022. Determinants of implementing pet robots in nursing homes for dementia care. *BMC Geriatrics* 22, 1 (Dec. 2022), 457. https://doi.org/10.1186/s12877-022-03150-z
[52] Oliver Korn (Ed.). 2019. *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction.* Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-17107-0
[53] Oliver Korn, Neziha Akalin, and Ruben Gouveia. 2021. Understanding Cultural Preferences for Social Robots: A Study in German and Arab Communities. *ACM Transactions on Human-Robot Interaction* 10, 2 (June 2021), 1–19. https://doi.org/10.1145/3439717
[54] Maya Krakovski, Naama Aharony, and Yael Edan. 2022. Robotic Exercise Trainer: How Failures and T-HRI Levels Affect User Acceptance and Trust. (2022). https://doi.org/10.48550/ARXIV.2209.01622
[55] Maya Krakovski, Shikhar Kumar, Shai Givati, Moshe Bardea, Oded Zafrani, Galit Nimrod, Simona Bar-Haim, and Yael Edan. 2021. "Gymmy": Designing and Testing a Robot for Physical and Cognitive Training of Older Adults. *Applied Sciences* 11, 14 (July 2021), 6431. https://doi.org/10.3390/app11146431
[56] Christian U. Krägeloh, Jaishankar Bharatharaj, Senthil Kumar Sasthan Kutty, Praveen Regunathan Nirmala, and Loulin Huang. 2019. Questionnaires to Measure Acceptability of Social Robots: A Critical Review. *Robotics* 8, 4 (Oct. 2019), 88. https://doi.org/10.3390/robotics8040088
[57] Chia-Yu Li, Daniel Ortega, Dirk Väth, Florian Lux, Lindsey Vanderlyn, Maximilian Schmidt, Michael Neumann, Moritz Völkel, Pavel Denisov, Sabrina Jenne, Zorica Kacarevic, and Ngoc Thang Vu. 2020. ADVISER: A Toolkit for Developing Multi-modal, Multi-domain and Socially-engaged Conversational Agents. (2020). https://doi.org/10.48550/ARXIV.2005.01777
[58] Ageless Innovation LLC. [n. d.]. Joy For All - Companion Pet Cat. https://joyforall.com/products/companion-cats publisher: Ageless Innovation LLC.
[59] Xiaofei Lu and Ying Xu. 2020. Simple Analysis of the Modelling and Function Design of the Partner Robots for the Elderly. In *2020 International Conference on Intelligent Design (ICID).* IEEE, Xi'an, China, 248–252. https://doi.org/10.1109/ICID52250.2020.00059
[60] Diana Löffler, Judith Dörrenbächer, and Marc Hassenzahl. 2020. The Uncanny Valley Effect in Zoomorphic Robots: The U-Shaped Relation Between Animal Likeness and Likeability. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction.* ACM, Cambridge United Kingdom, 261–270. https://doi.org/10.1145/3319502.3374788
[61] Melinda A. Mende, Martin H. Fischer, and Katharina Kühne. 2019. The Use of Social Robots and the Uncanny Valley Phenomenon. In *AI Love You*, Yuefang Zhou and Martin H. Fischer (Eds.). Springer International Publishing, Cham, 41–73. https://doi.org/10.1007/978-3-030-19734-6_3
[62] Juliana Miehle, Ilker Bagci, Wolfgang Minker, and Stefan Ultes. 2019. A Social Companion and Conversational Partner for the Elderly. In *Advanced Social Interaction with Agents*, Maxine Eskenazi, Laurence Devillers, and Joseph Mariani (Eds.). Vol. 510. Springer International Publishing, Cham, 103–109. https://doi.org/10.1007/978-3-319-92108-2_12

[63] Sadaf Moharreri, Nader Jafarnia Dabanloo, and Keivan Maghooli. 2018. Modeling the 2D space of emotions based on the poincare plot of heart rate variability signal. *Biocybernetics and Biomedical Engineering* 38, 4 (2018), 794–809. https://doi.org/10.1016/j.bbe.2018.07.001

[64] Michael Montemerlo, Joelle Pineau, Nicholas Roy, Sebastian Thrun, and Vandi Verma. 2002. Experiences with a Mobile Robotic Guide for the Elderly. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence, July 28 - August 1, 2002, Edmonton, Alberta, Canada*, Rina Dechter, Michael J. Kearns, and Richard S. Sutton (Eds.). AAAI Press / The MIT Press, 587–592. http://www.aaai.org/Library/AAAI/2002/aaai02-088.php

[65] Chuchi S. Montenegro and Elmer A. Maravillas. 2015. Acoustic-prosodic recognition of emotion in speech. In *2015 International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*. IEEE, Cebu City, Philippines, 1–5. https://doi.org/10.1109/HNICEM.2015.7393229

[66] Stanislava Naneva, Marina Sarda Gou, Thomas L. Webb, and Tony J. Prescott. 2020. A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust Towards Social Robots. *International Journal of Social Robotics* 12, 6 (Dec. 2020), 1179–1201. https://doi.org/10.1007/s12369-020-00659-4

[67] OpenAI. 2022. Introducing ChatGPT. https://openai.com/blog/chatgpt/ publisher: OpenAI.

[68] Deniz Ozdemir, Jaroslav Cibulka, Olga Stepankova, and Iva Holmerova. 2021. Design and implementation framework of social assistive robotics for people with dementia - a scoping review. *Health and Technology* 11, 2 (March 2021), 367–378. https://doi.org/10.1007/s12553-021-00522-0

[69] Amit Kumar Pandey and Rodolphe Gelin. 2018. A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind. *IEEE Robotics & Automation Magazine* 25, 3 (Sept. 2018), 40–48. https://doi.org/10.1109/MRA.2018.2833157

[70] Chris Papadopoulos, Tetiana Hill, Linda Battistuzzi, Nina Castro, Abiha Nigath, Gurch Randhawa, Len Merton, Sanjeev Kanoria, Hiroko Kamide, Nak-Young Chong, David Hewson, Rosemary Davidson, and Antonio Sgorbissa. 2020. The CARESSES study protocol: testing and evaluating culturally competent socially assistive robots among older adults residing in long term care homes through a controlled experimental trial. *Archives of Public Health* 78, 1 (Dec. 2020), 26. https://doi.org/10.1186/s13690-020-00409-y

[71] Hugo Pinto, Roberta Catizone, Yorick Wilks, and Alexiei Dingli. 2008. *The senior companion multiagent dialogue system.* The ACM Digital Library. https://www.um.edu.mt/library/oar/handle/123456789/17226

[72] Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. 2019. Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances. *IEEE Access* 7 (2019), 100943–100953. https://doi.org/10.1109/ACCESS.2019.2929050

[73] Elena Portacolone, Jodi Halpern, Jay Luxenberg, Krista L. Harrison, and Kenneth E. Covinsky. 2020. Ethical Issues Raised by the Introduction of Artificial Companions to Older Adults with Cognitive Impairment: A Call for Interdisciplinary Collaborations. *Journal of Alzheimer's Disease* 76, 2 (July 2020), 445–455. https://doi.org/10.3233/JAD-190952

[74] Jonathan Posner, James A. Russell, and Bradley S. Peterson. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology* 17, 03 (Sept. 2005). https://doi.org/10.1017/S0954579405050340

[75] Valentina Ramírez, Dominique Deuff, Xela Indurkhya, and Gentiane Venture. 2022. Design Space Survey on Social Robotics in the Market. *Journal of Intelligent & Robotic Systems* 105, 2 (June 2022), 25. https://doi.org/10.1007/s10846-022-01624-1

[76] Mark Ratchford and Michelle Barnhart. 2012. Development and validation of the technology adoption propensity (TAP) index. *Journal of Business Research* 65, 8 (Aug. 2012), 1209–1215. https://doi.org/10.1016/j.jbusres.2011.07.001

[77] Sergio Rinella, Simona Massimino, Piero Giorgio Fallica, Alberto Giacobbe, Nicola Donato, Marinella Coco, Giovanni Neri, Rosalba Parenti, Vincenzo Perciavalle, and Sabrina Conoci. 2022. Emotion Recognition: Photoplethysmography and Electrocardiography in Comparison. *Biosensors* 12, 10 (Sept. 2022), 811. https://doi.org/10.3390/bios12100811

[78] Intuition Robotics. [n. d.]. Meet ElliQ: ElliQ. https://elliq.com/pages/features publisher: Intuition Robotics.

[79] SoftBank Robotics. 2023. Meet Pepper - The Robot Built For People. http://us.softbankrobotics.com/pepper publisher: SoftBank Robotics.

[80] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (Dec. 1980), 1161–1178. https://doi.org/10.1037/h0077714

[81] Miki Sato, Yuko Yasuhara, Kyoko Osaka, Hirokazu Ito, Michael Joseph S. Dino, Irvin L. Ong, Yueren Zhao, and Tetsuya Tanioka. 2020. Rehabilitation care with Pepper humanoid robot: A qualitative case study of older patients with schizophrenia and/or dementia in Japan. *Enfermería Clínica* 30 (Feb. 2020), 32–36. https://doi.org/10.1016/j.enfcli.2019.09.021

[82] Michael Schiffmann, Aniella Thoma, and Anja Richert. 2021. Multi-modal Emotion Recognition for User Adaptation in Social Robots. In *Advances in Human Factors in Robots, Unmanned Systems and Cybersecurity*, Matteo Zallio, Carlos Raymundo Ibañez, and Jesus Hechavarria Hernandez (Eds.). Vol. 268. Springer International Publishing, Cham, 128–134.

https://doi.org/10.1007/978-3-030-79997-7_16

[83] Arielle Aj Scoglio, Erin D Reilly, Jay A Gorman, and Charles E Drebing. 2019. Use of Social Robots in Mental Health and Well-Being Research: Systematic Review. *Journal of Medical Internet Research* 21, 7 (July 2019), e13322. https://doi.org/10.2196/13322

[84] Fred Shaffer and J. P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 5 (Sept. 2017), 258. https://doi.org/10.3389/fpubh.2017.00258

[85] Amanda J. C. Sharkey and Natalie T. Wood. 2014. The Paro seal robot: demeaning or enabling? https://api.semanticscholar.org/CorpusID:34801556

[86] Aparna Sharma, Yash Rathi, Vibhor Patni, and Deepak Kumar Sinha. 2021. A systematic review of assistance robots for elderly care. In *2021 International Conference on Communication information and Computing Technology (ICCICT)*. IEEE, Mumbai, India, 1–6. https://doi.org/10.1109/ICCICT50803.2021.9510142

[87] Takanori Shibata, Joseph F. Coughlin, Human Technology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan, The AgeLab., Massachusetts Institute of Technology, 77 Massachusetts Ave, E40-279, Cambridge, MA 02139, USA, and Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Japan. 2014. Trends of Robot Therapy with Neurological Therapeutic Seal Robot, PARO. *Journal of Robotics and Mechatronics* 26, 4 (Aug. 2014), 418–425. https://doi.org/10.20965/jrm.2014.p0418

[88] Takanori Shibata and Kazuyoshi Wada. 2011. Robot Therapy: A New Approach for Mental Healthcare of the Elderly – A Mini-Review. *Gerontology* 57, 4 (2011), 378–386. https://doi.org/10.1159/000319015

[89] Shoshana Shiloh, Gal Sorek†, and Joseph Terkel. 2003. REDUCTION OF STATE-ANXIETY BY PETTING ANIMALS IN A CONTROLLED LABORATORY EXPERIMENT. *Anxiety, Stress & Coping* 16, 4 (Dec. 2003), 387–395. https://doi.org/10.1080/1061580031000091582

[90] Shamane Siriwardhana, Tharindu Kaluarachchi, Mark Billinghurst, and Suranga Nanayakkara. 2020. Multimodal Emotion Recognition With Transformer-Based Self Supervised Feature Fusion. *IEEE Access* 8 (2020), 176274–176285. https://doi.org/10.1109/ACCESS.2020.3026823

[91] M. Steverson. 2022. Ageing and health. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health publisher: World Health Organization.

[92] Kei Suzuki, Tipporn Laohakangvalvit, Ryota Matsubara, and Midori Sugaya. 2021. Constructing an Emotion Estimation Model Based on EEG/HRV Indexes Using Feature Extraction and Feature Selection Algorithms. *Sensors* 21, 9 (April 2021), 2910. https://doi.org/10.3390/s21092910

[93] Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2009. The Negative Attitudes Towards Robots Scale and reactions to robot behaviour in a live Human-Robot Interaction study. https://api.semanticscholar.org/CorpusID:15335957

[94] Masaaki Tanaka, Akira Ishii, Emi Yamano, Hiroki Ogikubo, Masatsugu Okazaki, Kazuro Kamimura, Yasuharu Konishi, Shigeru Emoto, and Yasuyoshi Watanabe. 2012. Effect of a human-type communication robot on cognitive function in elderly women living alone. *Medical Science Monitor* 18, 9 (2012), CR550–CR557. https://doi.org/10.12659/MSM.883350

[95] Sofia Thunberg, Lisa Rönnqvist, and Tom Ziemke. 2020. Do Robot Pets Decrease Agitation in Dementia Patients?: An Ethnographic Approach. In *Social Robotics*, Alan R. Wagner, David Feil-Seifer, Kerstin S. Haring, Silvia Rossi, Thomas Williams, Hongsheng He, and Shuzhi Sam Ge (Eds.). Vol. 12483. Springer International Publishing, Cham, 616–627. https://doi.org/10.1007/978-3-030-62056-1_51

[96] Xinxia Wang, Jun Shen, and Qiu Chen. 2022. How PARO can help older people in elderly care facilities: A systematic review of RCT. *International Journal of Nursing Knowledge* 33, 1 (Jan. 2022), 29–39. https://doi.org/10.1111/2047-3095.12327

[97] WHO. 2021. Social isolation and loneliness among older people: advocacy brief. https://iris.who.int/bitstream/handle/10665/343206/9789240030749-eng.pdf?sequence=1

[98] Jianhua Zhang, Zhong Yin, Peng Chen, and Stefano Nichele. 2020. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion* 59 (July 2020), 103–126. https://doi.org/10.1016/j.inffus.2020.01.011

[99] Terry Yue Zhuo, Yujin Huang, Chunyang Chen, and Zhenchang Xing. 2023. Red teaming ChatGPT via Jailbreaking: Bias, Robustness, Reliability and Toxicity. (2023). https://doi.org/10.48550/ARXIV.2301.12867

A APPENDICES

A.1 Pulse Sensor Specifications

Table 3. Pulse Sensor Specifications

Description	Specification
Maximum Ratings	*VCC 3.0 – 5.5V*
	Imax (Maximum Current Draw) <4mA
	VOut (Output Voltage Range) 0.3 to VCC
Wavelength	LED Output 565nm
	Sensor Input 525nm
Dimensions & Temperature	L x W (PCB) 15.8mm (0.625”)
	Lead Length 20cm (7.8”)
	Operating Temperature Range -40 up to 85°C

A.2 Python Libraries Used

- dbus_python==1.2.16
- firebase_admin==6.0.1
- gobject==0.1.0
- gTTS==2.3.0
- numpy==1.19.5
- openai==0.26.0
- PyAudio==0.2.13
- pygame==1.9.6
- PyGObject==3.38.0
- python_speech_features==0.6
- pyttsx3==2.90
- sensai==0.1.9
- soundfile==0.11.0
- SpeechRecognition==3.9.0
- spidev==3.5
- nltk==3.8.1
- pandas==1.4.2
- matplotlib==3.5.1
- tensorflow==2.11.0

Received February 2023; revised July 2023; accepted October 2023