

When No One is Watching: Ecological Momentary Assessment to Understand Situated Social Robot Use in Healthcare

Casey C. Bennett*

Department of Intelligence
Computing, Hanyang University,
Seoul Korea, College of Computing,
DePaul University, Chicago USA
cabennet@hanyang.ac.kr

Ćedomir Stanojević

School of Public Health, Indiana
University, Bloomington USA
cestan@iu.edu

Selma Šabanović

School of Informatics, Computing,
and Engineering, Indiana University,
Bloomington USA
selmas@indiana.edu

Jennifer A. Piatt

School of Public Health, Indiana
University, Bloomington USA
piatt@indiana.edu

Seongcheol Kim*

Department of Intelligence
Computing, Hanyang University,
Seoul Korea
sckim219@hanyang.ac.kr

ABSTRACT

Socially-Assistive Robots (SARs) hold great potential to revolutionize the way we manage chronic illness outside clinical settings, but a current limitation to their broad adoption for this purpose is the lack of “ground truth” around interactions between robots and humans in in-home settings. Such ground truth is a necessity for using robotic sensor data for machine learning models of patient activity patterns or to create AI to customize robotic interactive behavior autonomously. Traditional subjective recall-based data collection methods lack the fine-grained temporal detail to support such AI development, as well as suffering from “recall bias” effects. One potential solution to this challenge is to adapt novel forms of interaction assessment, such as ecological momentary assessment (EMA), to collect patient interaction data *in real-time*. Here we describe a pilot study utilizing such an EMA system with SARs. We describe the development of the EMA framework, theoretical design issues, and lessons learned. Preliminary machine learning results indicate 75-80% accuracy for detecting specific interaction modalities. We also discuss the potential utility of EMA for exploring cross-cultural differences with *in-the-wild* robot use, and as a tool to support participatory design research on robotics in healthcare settings.

*Corresponding Author: Casey C. Bennett,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HAI '21, November 09–11, 2021, Virtual Event, Japan

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8620-3/21/11...\$15.00

<https://doi.org/10.1145/3472307.3484670>

CCS CONCEPTS

• ; • **Human-centered computing** → Human computer interaction (HCI); HCI design and evaluation methods; • **Computer systems organization** → Embedded and cyber-physical systems; Robotics; • **Applied computing** → Life and medical sciences; Health informatics;

KEYWORDS

Socially-Assistive Robots, Ecological Momentary Assessment, Healthcare, Artificial Intelligence, Robotics, Participatory Design

ACM Reference Format:

Casey C. Bennett*, Ćedomir Stanojević, Selma Šabanović, Jennifer A. Piatt, and Seongcheol Kim. 2021. When No One is Watching: Ecological Momentary Assessment to Understand Situated Social Robot Use in Healthcare. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI '21), November 09–11, 2021, Virtual Event, Japan*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3472307.3484670>

1 INTRODUCTION

1.1 Background

Socially-Assistive Robots (SARs) hold great potential to revolutionize the way we manage chronic illness outside clinical settings, both as a social support tool and as a novel form of remote monitoring via robotic sensors [1-4]. However, one of the major challenges with deploying SARs and making use of the collected sensor data is having “ground truth” labels so that machine learning algorithms can be developed to classify multi-modal sensor patterns into discernible activities. This is a similar challenge seen with mobile phone activity detection [5, 6]. With SARs this challenge is two-fold: we both want to understand the humans’ behavioral patterns, as well as use the sensor data to create more *adaptive* interactive behavior in our social robots. To do so, however, requires us to establish “ground truth” for what behaviors are occurring when robots interact with people *in-the-wild*. This is both a technological as well as design problem.

Until now, most SARs research outside lab settings has been conducted using traditional recall-based methods to collect data post-hoc (after the interaction), e.g. diaries, phone calls, and other

data-collection instruments at the end of each day/week [1, 2, 4, 7]. While these capture the general gist of interaction, they lack the fine-grained temporal detail of when specific activities occur throughout the day, and the critical *sequential* nature of human interactions [8]. Events do not occur “randomly”, but rather most events are situated within specific scenarios. As such, we need a way to capture this information, tie it to the robotic sensor data, and link that back to patient health outcomes. Here, we propose an approach to address this problem with SARs using ecological momentary assessment (EMA), which attempts to reveal how people interact with such technology in the real world in *real-time*, to gain a better understanding of the behaviors elicited in humans [9, 10].

The paper is laid out as follows. In Section 1, we provide a general argument for why EMA should be used with SARs. In Section 2, we describe our current ongoing research and technical framework design for how EMA can be used with SARs. In Section 3, we present some of our results so far, including preliminary machine learning analyses and lessons learned during deployment. Finally in Sections 4 and 5, we discuss our conclusions, as well as the broader cross-cultural issues that may impact the use of EMA to design interactive behavior for social robots.

1.2 EMA and Social Robot Design

Previous participatory design studies have suggested that robots as artificial agents need to be designed not only from a physical standpoint, but also as artifacts geared towards specific situated contexts and uses, in order to elicit a more social stance amongst the users [2, 11]. Interestingly, recent research on fMRI neural activity during social interaction between human-human versus human-agent (conversational digital avatar) found significant differences in several parts of the brain, including areas that are hypothesized to associate with the difference between a more social *intentional* stance versus a more mechanistic *design* stance [12]. However, it is an open question as to how exactly a robot’s behavior should change autonomously in order to produce this effect in humans during social interaction.

The use of EMA as a novel form of psychological assessment during *real-time* social interaction opens up intriguing possibilities for attempting to uncover how agent behavior influences social cognition during human-agent interaction in more naturalistic settings [13]. This is particularly true if the EMA is combined with other devices, such as a wearables and smartphones [14]. Furthermore, EMA helps to characterize variations in behaviors *within* a participant over time [15]. From a design standpoint, successful human-robot interaction (HRI) in the wild requires a *social fluidity* that can adapt to both intra- and inter- person variation, and the harmonization of behaviors between humans and evocative objects in their environment [16, 17]. EMA can help us unravel this complex set of dynamics to create better interaction models for SARs in various healthcare settings, both clinical and community-based.

1.3 Use of SARs for Healthcare

Within the context of healthcare, EMA has been shown to be a powerful tool for monitoring everyday patient behaviors [9, 10]. One challenge with many artificial intelligence approaches in healthcare has been their reliance on clinical data, which is limited to times

when the patient visits a clinic or doctor’s office [18, 19]. SARs, on the other hand, are often used outside those specific care settings, such as in patients’ homes where they can collect data on everyday patient routines and health status. This has wide potential applicability to chronic illnesses, aging-related issues, and other pressing healthcare issues. For instance, the growing older adult population along with longer life expectancies in many countries presents a challenge as how to best take care of older adults in their own homes and communities, without resorting to costly institutionalized care (e.g. assisted living, nursing homes) [20]. Moreover, new health care issues are emerging out of the COVID-19 shelter-in-place situation with the older adult population. Regardless if the older adult was living independently at home or a retirement community, recent research has shown this population is experiencing lower levels of quality of life encompassing a decreased social life, less in-person social interactions, lower levels of physical activity, and an increased risk of depression post-pandemic [21]. More broadly, prior research has demonstrated the impact SARs can have on addressing clinical outcomes in a wide array of populations, such as managing symptoms associated with mental illness and promoting an overall higher quality of life [4, 22, 23]. As health care has shifted from in-person to telehealth platforms, creative interventions that can be adapted to the home environment have become more relevant and critical in managing health [24]. Moving forward, the challenge lies in how to better extend these findings beyond research settings.

2 TECHNICAL FRAMEWORK FOR USE OF EMA WITH SOCIAL ROBOTS

2.1 Theoretical Design Considerations

As mentioned in Section 1, the biggest advantage of the EMA approach is to reduce the recall bias effect in understanding real-world interactions with healthcare AI technology. However, the challenge is designing an EMA framework that collects *relevant* information at *appropriate* times. From a theoretical standpoint, this requires careful consideration of both sampling strategies and stimulus design.

The two most prominent methods for EMA sampling are “event-based” and “time-based” [15]. Event-based focuses on specific moments in a participant’s day without capturing the entire experience during the data collection period. It can be either reactive (based on a sensed event in the environment) or trigger-based (prompted purposely by the researchers). For example, behaviors can be recorded by triggering participants to interact with SAR at random intervals and reminding them to complete an event-based survey at the end of the interaction. Conversely, the time-based sampling approach is dependent on random surveys during set time windows (e.g. morning, afternoon), without interaction prompts [25]. The inquiries about the performed behavior, as well as the frequency of interactions, provide data about ongoing, organic SAR interactions. There are trade-offs with any of these sampling strategies: event-based provides richer interaction data via prompts but misses out on smaller, less-defined interactions, whereas time-based provides more organic interaction data but can often result in more “sparse” datasets because the surveys arrive at times when no interaction is

occurring. There is also debate around the appropriate frequency of EMA prompts, from just a few times a day to over 20 times [15].

Stimulus design (i.e. prompt content) in terms of SARs must be rooted in two concerns: 1) the interactions we expect to occur, and 2) the sensor-based activities capable of being detected [26, 27]. Poorly designed stimuli can omit critical features of the interaction, and/or potentially bias the results towards researchers' *a priori* preconceptions. To avoid these issues requires a grounded approach to stimulus design, typically rooted in prior research to establish a data-driven analytical framework [28]. In our case, we divided our EMA stimulus for SARs into two primary components: modality (the type of behavior) and proximity (whether the interaction occurred near/far to the robot), based on our prior research using SARS with elderly patients with Major Depressive disorder [1-3, 29, 30]. The modalities included both active interactions (e.g. petting, talking, playing) and passive interactions (e.g. watching television, eating together with the robot). Furthermore, beyond the interaction-focused stimulus questions above, we incorporated additional psychological assessment questions to gauge user perception and emotional response post-interaction. The stimuli were further divided into baseline, intervention, post-intervention phases (see Section 2.2).

2.2 EMA Protocol Design

Designing the EMA protocol was influenced by the aims of the study, which are as follows:

- To use EMA to distinguish specific kinds of human-robot interaction, rather than just general “interaction”
- To evaluate the effectiveness of an EMA data gathering procedure in establishing health-related daily activity routines, as well as emotional and behavioral changes within participants using SAR
- To establish which EMA sampling strategy (e.g. time-based, event-based) provides the most accurate data for modeling robotic sensor data against
- To evaluate how EMA sampling strategy and stimulus design influence participants' data collection compliance (response rate) during health-focused HRI interventions

To achieve the aforementioned aims we tested various mobile software and text-messaging systems that offer EMA services (full list omitted here for brevity). After that evaluation, the research team decided to utilize the EMA PilR platform (<https://pilrhealth.com/>), which provides a customizable mobile app available for both Android and iOS. The app can be used to send out a signal “ping” to the participants to collect baseline information, complete scheduled assessments, and prompt interaction triggers. The platform was configured to test our hypotheses around stimulus design and sampling strategies, as mentioned in the aims above. Before deploying SARs (in this case the Hasbro Joy-For-All robotic therapy pet, <https://joyforall.com/>) (Figure 1), detailed instructions were sent to the participants regarding the use of the PilR EMA app and the nature of the study, as well as an 18-question baseline survey which established current daily habits and behavioral routines of the participant. A similar post-intervention survey was administered after the main intervention phase, to gauge behavioral

changes and user perceptions of the SAR+EMA intervention. During the intervention phase, survey questions consisted of inquiries that assessed interaction status between the participant and the SAR (modality, proximity), as described below. Full questionnaires for all phases, referred to as the Social Robot EMA (SoREMA) survey, are available online (http://www.caseybennett.com/uploads/SoREMA_Survey_Questionnaire.docx).

In terms of intervention study design, both time-based and event-based sampling strategies were scheduled at different periods of the intervention for each participant (within-subjects design), incorporating both random sampling multiple times per day within specific time interval windows (morning, afternoon, night) for the former as well as event triggers for the latter. The participants received time-based survey prompts (7-question) throughout the course of one week to assess their interactions with the SAR during the prior 15 minutes. After this, the same routine would be repeated with event-based (trigger) sampling strategy for the second week. During these two weeks of gathering interaction data via the PilR EMA app, a robotic sensor collar was simultaneously in-use onboard the SAR (Figure 1). The collar, which was developed at Indiana University's R-house robotics lab, was equipped with sensors to collect information about sound, ambient light, and motion of the robot (via a 3D accelerometer). This information can be used to detect interaction patterns between the robot and subject [1-2]. The data was collected continuously, roughly once per second. The eventual aim is to match the sensor data to the EMA data, to see if interaction patterns align between the two.

2.3 Data Analysis & Modeling

An initial evaluation of this system described in Section 2.2 was conducted in Fall/Winter 2020 with 4 participants from the US Midwest. The participants were drawn from the general population and living alone. Data collected was parsed into an analysis dataset to test the feasibility of machine learning (ML) using EMA data. In this case, the EMA data became the “targets” while the sensor data became the “features” for ML models. These EMA targets were the modalities described in Section 2.1. For simplicity we collapsed the dataset into a series of binary classification predictions (e.g. petting vs. not petting) rather than attempt the more difficult multi-class classification problem. Due to target class imbalance, the data was re-balanced using SMOTE [31]. Additionally, some modalities were rarely performed and thus were excluded from further analysis, which left us with a sample of 80 interactions across five modalities: petting, playing, moving the SAR (from one location to another), talking to the SAR, watching TV/radio. Each “interaction” represented a 15-minute time period (based on the EMA protocol described in Section 2.2), **so the 80 interactions constituted 20 hours of total interaction data**. The modalities were not mutually exclusive, so for instance a participant could be petting the SAR while talking to it. Indeed, participants reported slightly over 2 modality types per interaction. We expected that the pattern *signatures* in the data for both modalities would be present in that case, and with appropriate modeling techniques be detectable.

Modeling was performed using the python package Keras (<https://keras.io/>), which is a deep learning library based on TensorFlow.



Figure 1: (a) Joy For All robot wearing a sensor collar; (b) view of sensors inside the collar

For predicting the EMA target, the feature data for that 15-minute time period was sliced into 15-second-long overlapping *windows*, with 50% overlap (similar to [5]). The resulting data was then fed into a deep learning model consisting of a single 2D convolutional neural network layer (CNN) with kernel size set to 1 and using a ReLU activation function, followed by a single recurrent layer (LSTM) with 50 units [32]. The idea was that the CNN could parse out "invariant representations" of pattern signatures occurring anywhere in the interaction, followed by the LSTM detecting critical "sequences" of those patterns over time. A final fully connected "Dense" layer using a sigmoidal activation function was used to make the final binary classification predictions. To evaluate performance, 20% of the data was held out as a "test set" for each classification run. Feature selection to identify was performed to identify which features were important for predicting specific modalities, using wrapper-based methods in Scikit-learn [33]. Those features included motion in the x/y/z directions, rotational motion (arc), light/sound values, orientation, and transitions between sound/orientation categories (e.g. loud to quiet, portrait to landscape).

3 RESULTS

3.1 Design Lessons Learned

The study described here is still ongoing, but there have been some important early lessons learned, which we believe may be useful for the human-computer interaction (HCI) and HRI communities in attempting to use EMA to assess social robot deployments in real-world settings. These fall into two broad categories: EMA design recommendations and platform hardware recommendations.

In terms of EMA design, feedback from participants revealed several things. First, the baseline survey was critical for establishing target time intervals for intervention pings, which were both appropriate for the participant to receive notifications (pinging them when sleeping is not useful, for instance) and most probable time periods when participants were willing to engage the SAR (different individuals have different schedules and "unstructured leisure times"). Without setting such parameters, EMA data will

likely be much too sparse to allow for robust modeling or analysis. Secondly, ping frequency needs to achieve a delicate balance between capturing relevant interactions and being "annoying" to the user. It appears better to allow a missed interaction to occur, than including "reminder prompts", particularly when already pinging participants 5+ times per day. All participants except one reported a sense of *alert fatigue* with frequent reminder prompts, similar to that observed in clinicians and patients in other settings [34, 35]. All participants also expressed a desire for incentives, such as badges for completion or leaderboards to compare their own interaction patterns to others' robot use.

On the hardware front, we encountered challenges with time synchronization between the various devices, which is of course a known issue in the sensor fusion and internet-of-things communities [36, 37]. We found it useful to run separate internal clocks on each device, keeping logs of the exact activation time, then cross-walking between unix time-stamps. In situations where internet/Wi-Fi fails for instance, this can serve as a useful fall-back for automated time synchronization. In real-world healthcare applications, especially in populations such as rural older adults where internet connectivity is limited, this is an important system design consideration.

Another noteworthy phenomenon reported by participants during the research was the possible emergence of Hawthorne effect [38]. During interviews conducted in the post-intervention phase, several participants (3 out of 4) reported that they felt like they were starting to specifically perform behaviors with the SAR corresponding to the modalities being presented as a multiple-choice list during EMA prompts, even though they were given the option to add behaviors that were not in the list. As such, we analyzed the data looking for this, but only found possible evidence for it in one participant (i.e. 25% of the time). Nevertheless, it is interesting that the Hawthorne Effect is being reported by some participants, and it remains something researchers in this domain should be aware of in the future. The Hawthorne effect has been well-researched and may affect research participation through participants' increased

Table 1: Machine Learning Classification Results

Modality	Accuracy	AUC
Playing	86.2	0.9285
Talking	68.4	0.7638
Petting	75.6	0.7234
Listening TV/Radio	74.8	0.7781
Moving It	66.0	0.6938
Average	75.3	0.7752

awareness of being studied. In addition, feelings of wanting to conform to expectations and social desirability could also induce a behavior change within the participant [39]. According to McCambridge et al. (2014), the Hawthorne effect can affect participants' behavior but the mechanisms of emergence of this phenomenon and its magnitude are still unclear [38].

3.2 Data Analysis Results

The results of machine learning analysis described in Section 2.3 are shown in Table 1, based on performance on the test set. The main aim was to test the feasibility of this EMA approach, in particular the ability to use robotic sensor data for activity recognition of human-robot interactions based on EMA in real-world settings.

As can be seen in the table, the overall average performance across all modalities is approximately 75% accuracy, with an AUC (area-under-curve) of .7705. The performance is variable, however, with the active interaction modalities (Playing/Petting) appearing to have better performance than the other modalities. Moreover, a more detailed feature selection analysis indicated that specific sensors may be related to the pattern *signatures* of specific modalities, enabling the predictability. For instance, petting the robot seemed more dependent on orientation and motion features, while playing depended more on sound and motion features, TV/radio depended on orientation and sound, and talking depended on sound and motion. Rotational motion (arc) was also present for playing and talking, but not the other modalities. In other words, light/sound sensor data may be more important for detecting some modalities, while motion data and orientation may be more important for others.

Obviously, given the limited sample size in the pilot study (only 4 participants), these results should be taken with caution. They are suggestive of potential, but not sufficient to draw any firm conclusions yet. The high degree of variability across modalities, as well as discrepancies between accuracy and AUC values, indicates additional data collection is needed, particularly in cases like this where a large amount of class imbalance is to be expected, which would allow for a broader array of re-balancing techniques to be explored (e.g. under-sampling, hybrid methods). Larger sample sizes may also enable exploration of simultaneous multi-class classification, averse to the simpler binary classification here (see Section 2.3). In short, the results here point to a number of potential future research avenues related to modeling of EMA/sensor data with SARs.

4 DISCUSSION

4.1 Implications

Our preliminary results in Section 3 suggest the potential for EMA in social robotics and understanding real-world human-agent interactions, as well as a number of implications for the future related to current HRI challenges and health care practices. For instance, one of the challenges in social robot design is that robot use by humans is grounded in the situational context, which is heavily influenced by cultural factors [40]. Culture shapes not only the *practice* of technology, but also human cognition during interaction. The interplay between the two is not static, but emergent. As such, robotic technologies should not only identify and mimic cultural forms, but also be culturally robust and adaptive to the dynamic, situated performance of human culture [41].

Broader HCI research has identified differences in many forms of technology use, including robots, but also pointed out that such distinctions tend to suffer from over-simplification [42, 43]. SARs, in that regard, can be conceptualized as “companion objects” or “evocative objects” that alter our perceptions of self, our environment, and the socio-technical systems we inhabit [44, 45]. These effects operate on multiple levels and dimensions. Examples include high-context vs. low-context cultures, communal vs. individualistic cultures, Confucian power hierarchies vs. Western power hierarchies, and so forth . . . all of which impact societal structure down to individual behavior [40, 46]. Beyond simply categorizing these differences, the challenge is how we can capture the impact of these factors on *situated robot use*. For instance, prior research suggests participants in South Korea envision robots as companions for the family, while US participants see home robots as individual assistants and modern appliances [40]. Cultural differences in participants perceptions of robots may lead to differential therapeutic effects, particularly for those conditions that impact quality of life and addressed through social interaction, such as depression, loneliness, and dementia.

4.2 Participatory Design Approaches

The challenge mentioned above suggests that there may be a role for EMA to play in the cross-cultural study and design of robotics, as a method for understanding nuances of how robots are adapted to culturally-specific uses in *real time*, outside the purview of static post-hoc questionnaires or other forms of analysis. Combined with participatory-design (PD) approaches to social robotics [11, 47], this may enable us to create more adaptive forms of behavior tuned to latent cultural attributes that are apparent in subtle differences in the temporal dynamics of robot use by engaging users in EMA-based *data-driven* discussions during the PD process.

EMA could also serve as a form of feedback to help users understand their own robot-use patterns outside of PD workshops in order to stimulate ideas during later workshops. Interestingly, all of the participants in our pilot study expressed a desire during post-intervention interviews for that sort of feedback, as well as possible comparisons to other users (e.g. “leaderboard” or “social feed” that reports who else in their social network is currently interacting with their SAR and how), in order to *contextualize* their own robot use. Such an approach may help address some of the known challenges with PD research, such as maintaining focus during

the design process and envisioning abstract forms of interaction beyond the participant's immediate personal experience [48]. This may be particularly true during longer-term PD research studies in the field.

5 CONCLUSION

Understanding what people “really do” with robots is a key first step in designing better robot behavior. This paper describes the use of EMA as a novel form of real-time interaction assessment with SARs *in-the-wild* for healthcare purposes. Such real-time assessment is critical for enabling better interactive behavior in social robots and other forms of AI, as well as serving as the basis for machine learning models using robotic sensor data to track in-home patient status. Here, we detail a pilot study utilizing such an EMA system with SARs, including the development of the EMA framework and theoretical design issues. There were a number of lessons learned which may be valuable for the HRI/HCI communities, as well as preliminary results showing 75–80% accuracy for detecting specific interaction modalities. More broadly, we contend that EMA holds potential utility for exploring cross-cultural differences with *in-the-wild* robot use, and as a tool to support participatory design research on robotics in healthcare settings.

ACKNOWLEDGMENTS

This work was supported by the research fund of Hanyang University (HY-2020) in Korea, as well as the National Science Foundation in the United States (Grant# IIS-1900683). We would also like to thank our various collaborators over the years who have contributed to different aspects of this work.

REFERENCES

- [1] Casey C. Bennett, Selma Šabanović, Jennifer A. Piatt, Shinichi Nagata, Lori Eldridge, and Natasha Randall. 2017. A robot a day keeps the blues away. *IEEE International Conference on Healthcare Informatics (ICHI)*, 536–540.
- [2] Natasha Randall, Casey C. Bennett, Selma Šabanović, Shinichi Nagata, Lori Eldridge, Sawyer Collins, and Jennifer A. Piatt. 2019. More than just friends: in-home use and design recommendations for sensing socially assistive robots (SARs) by older adults with depression. *Paladyn, Journal of Behavioral Robotics*, 10, 1, 237–255.
- [3] Jennifer Piatt, Shinichi Nagata, Selma Šabanović, Wan-Ling Cheng, Casey C. Bennett, Hee Rin Lee, and David Hakken. 2016. Companionship with a robot? Therapists' perspectives on socially assistive robots as therapeutic interventions in community mental health for older adults. *American Journal of Recreation Therapy*, 15, 4, 29–39.
- [4] Lihui Pu, Wendy Moyle, Cindy Jones, and Michael Todorovic. 2019. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. *The Gerontologist*, 59, 1, e37–e51.
- [5] Yufei Chen and Chao Shen. 2017. Performance analysis of smartphone-sensor behavior for human activity recognition. *IEEE Access*, 5, 3095–3110.
- [6] Godwin Ogbuabor, and Robert La. 2018. Human activity recognition for healthcare using smartphones. *International Conference on Machine Learning and Computing*, 41–46.
- [7] Hayley Robinson, Bruce MacDonald, and Elizabeth Broadbent. 2014. The role of healthcare robots for older people at home: A review. *International Journal of Social Robotics*, 6, 4, 575–591.
- [8] Casey C. Bennett and Kris Hauser. 2013. Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine*, 57, 1, 9–19.
- [9] Claudia Vesel, Homa Rashidisabot, John Zulueta, Jonathan P. Stange, Jennifer Duffecy, Faraz Hussain, Andrea Piscitello *et al.* 2020. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: A BiAffect iOS study. *Journal of the American Medical Informatics Association*, 27, 7, 1007–1018.
- [10] John Zulueta, Andrea Piscitello, Mladen Rasic, Rebecca Easter, Pallavi Babu, Scott A. Langenecker, Melvin McInnis *et al.* 2018. Predicting mood disturbance severity with mobile phone keystroke metadata: A BiAffect digital phenotyping study. *Journal of Medical Internet Research* 20, 7, e241.
- [11] Selma Šabanović, Wan-Ling Chang, Casey C. Bennett, Jennifer A. Piatt, and David Hakken. 2015. A robot of my own: participatory design of socially assistive robots for independently living older adults diagnosed with depression. *International Conference on Human Aspects of IT for the Aged Population*, 104–114. Springer, Cham.
- [12] Birgit Rauchbauer, Bruno Nazarian, Morgane Bourhis, Magalie Ochs, Laurent Prévot, and Thierry Chaminade. 2019. Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society B*, 374, 1771. 20180033.
- [13] Stephen Intille, Caitlin Haynes, Dharam Maniar, Aditya Ponnada, and Justin Mankjourides. 2016. μEMA: Microinteraction-based ecological momentary assessment (EMA) using a smartwatch. *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1124–1128.
- [14] Ingar Brinck, and Christian Balkenius. 2020. Mutual recognition in human-robot interaction: A deflationary account. *Philosophy & Technology* 33, 1, 53–70.
- [15] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. 2008. Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32.
- [16] Malte Jung, and Pamela Hinds. 2018. Robots in the wild: A time for more robust theories of human-robot interaction. *ACM Transactions on Human-Robot Interaction*, 7, 1, 1–5.
- [17] Selma Šabanović, Marek P. Michalowski, and Reid Simmons. 2006. Robots in the wild: Observing human-robot social interaction outside the lab. *IEEE International Workshop on Advanced Motion Control*, 596–601.
- [18] Jamie Bennett, Osvaldas Rokas, and Liming Chen. 2017. Healthcare in the smart home: A study of past, present and future. *Sustainability*, 9, 5, 840.
- [19] Kun-Hsing Yu, Andrew L. Beam, and Isaac S. Kohane. 2018. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2, 10, 719–731.
- [20] Charles F. Reynolds III, Pim Cuijpers, Vikram Patel, Alex Cohen, Amit Dias, Neerja Chowdhary, Olivia I. Okereke *et al.* 2012. Early intervention to reduce the global health and economic burden of major depression in older adults. *Annual Review of Public Health*, 33, 123–135.
- [21] Audrey Lebrasseur, Noémie Fortin-Bédard, Josiane Lettre, Emilie Raymond, Eve-Line Bussières, Nolwenn Lapierre, Julie Faita *et al.* 2021. Impact of the COVID-19 pandemic on older adults: rapid review. *JMIR Aging*, 4, 2, e26474.
- [22] Hannah Louise Bradwell, Katie Jane Edwards, Rhona Winnington, Serge Thill, and Ray B. Jones. 2019. Companion robots for older people: importance of user-centred design demonstrated through observations and focus groups comparing preferences of older people and roboticists in South West England. *BMJ Open*, 9, 9, e032468.
- [23] Arielle A.J. 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, e13322.
- [24] Lori Eldridge, Shinichi Nagata, Jennifer Piatt, Cedomir Stanojevic, Selma Šabanović, Casey Bennett, and Natasha Randall. 2020. Utilization of socially assistive robots in recreational therapy. *American Journal of Recreation Therapy*, 19, 2, 35–45.
- [25] Philippe AEG Delespaul. 1995. *Assessing schizophrenia in daily life: The experience sampling method*. Universitaire Pers Maastricht, doctoral thesis,
- [26] Liming Chen, Jesse Hoey, Chris D. Nugent, Diane J. Cook, and Zhiwen Yu. 2012. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42, 6, 790–808.
- [27] Ye Liu, Liqiang Nie, Li Liu, and David S. Rosenblum. 2016. From action to activity: sensor-based activity recognition. *Neurocomputing*, 181, 108–115.
- [28] Marika B. Solhan, Timothy J. Trull, Seungmin Jahng, and Phillip K. Wood. 2009. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. *Psychological Assessment*, 21, 3, 425.
- [29] Selma Šabanović, Casey C. Bennett, Wan-Ling Chang, and Lesa Huber. 2013. PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. *IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, 1–6.
- [30] Sawyer Collins, Selma Šabanović, Marlena Fraune, Natasha Randall, Lori Eldridge, Jennifer A. Piatt, Casey C. Bennett, and Shinichi Nagata. 2018. Sensing Companions: Potential Clinical Uses of Robot Sensor Data for Home Care of Older Adults with Depression. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 89–90.
- [31] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- [32] Kun Xia, Jianguang Huang, and Hanyu Wang. 2020. LSTM-CNN architecture for human activity recognition. *IEEE Access*, 8, 56855–56866.
- [33] Verónica Bolón-Canedo, Noelia Sánchez-Marono, Amparo Alonso-Betanzos, José Manuel Benítez, and Francisco Herrera. 2014. A review of microarray datasets and applied feature selection methods. *Information Sciences*, 282, 111–135.
- [34] Jessica S. Ancker, Alison Edwards, Sarah Nosal, Diane Hauser, Elizabeth Mauer, and Rainu Kaushal. 2017. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Medical Informatics and Decision Making*, 17, 1, 1–9.

[35] Mustafa I. Hussain, Tera L. Reynolds, and Kai Zheng. 2019. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review. *Journal of the American Medical Informatics Association*. 26, 10, 1141-1149.

[36] Raffaele Gravina, Parastoo Alinia, Hassan Ghasemzadeh, and Giancarlo Fortino. 2017. Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*. 35, 68–80.

[37] Enea Cippitelli, Samuele Gasparrini, Ennio Gambi, Susanna Spinsante, Jonas Wåhslén, Ibrahim Orhany, and Thomas Lindh. 2015. Time synchronization and data fusion for RGB-depth cameras and inertial sensors in AAL applications. *IEEE international conference on communication workshop (ICCW)*. 265-270.

[38] Jim McCambridge, John Witton, and Diana R. Elbourne. 2014. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*. 67, 3, 267-277.

[39] Gustav Wickström, and Tom Bärdix. 2000. The Hawthorne effect - what did the original Hawthorne studies actually show? *Scandinavian Journal of Work, Environment & Health*. 363-367.

[40] Hee Rin Lee, and Selma Šabanović. 2014. Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 17-24.

[41] Selma Šabanović, Casey C. Bennett, and Hee Rin Lee. 2014. Towards culturally robust robots: A critical social perspective on robotics and culture. In *Proc. HRI Workshop on Culture-Aware Robotics*, vol. 2014.

[42] Christina Bröhl, Jochen Nelles, Christopher Brandl, Alexander Mertens, and Verena Nitsch. 2019. Human–robot collaboration acceptance model: development and comparison for Germany, Japan, China and the USA. *International Journal of Social Robotics*. 11, 5, 709-726.

[43] Kerstin Sophie Haring, et al. 2014. Perception of an android robot in Japan and Australia: A cross-cultural comparison. *International Conference on Social Robotics (ICSR)*. Springer, Cham, 2014.

[44] Rosalind W. Picard. 2000. *Affective Computing*. MIT press.

[45] Selma Šabanović. 2021. Designing Companion Artifacts: The Relational Construction of Culture and Technology in Social Robotics. *Culturally Sustainable Social Robotics: Proceedings of Robophilosophy*. 335, 3.

[46] Tatsuya T. Nomura, Dag Sverre Syrdal, and Kerstin Dautenhahn. 2015. Differences on social acceptance of humanoid robots between Japan and the UK. In *Procs 4th int Symposium on New Frontiers in Human-Robot Interaction*. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB).

[47] Emma J. Rose, and Elin A. Björling. Designing for engagement: using participatory design to develop a social robot to measure teen stress. 2017. *Proceedings of the 35th ACM International Conference on the Design of Communication*. 1-10.

[48] Stephen Lindsay, Daniel Jackson, Guy Schofield, and Patrick Olivier. 2012. Engaging older people using participatory design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. 1199-1208.