<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *10th International Conference on Social Robotics (ICSR 2018), Qingdao, China, November 28-30, 2018.*

Citation for the original published paper:

Akalin, N., Kiselev, A., Kristoffersson, A., Loutfi, A. (2018)
The Relevance of Social Cues in Assistive Training with a Social Robot
In: Ge, S.S., Cabibihan, J.-J., Salichs, M.A., Broadbent, E., He, H., Wagner, A., Castro-González, Á. (ed.), *10th International Conference on Social Robotics, ICSR 2018, Proceedings* (pp. 462-471). Springer
Lecture Notes in Computer Science
https://doi.org/10.1007/978-3-030-05204-1_45

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-70817>

The Relevance of Social Cues in Assistive Training with a Social Robot

Neziha Akalin, Andrey Kiselev, Annica Kristoffersson, and Amy Loutfi

Örebro University, SE-701 82 Örebro, Sweden,
neziha.akalin@oru.se,
<http://mpi.aass.oru.se>

Abstract This paper examines whether social cues, such as facial expressions, can be used to adapt and tailor a robot-assisted training in order to maximize performance and comfort. Specifically, this paper serves as a basis in determining whether key facial signals, including emotions and facial actions, are common among participants during a physical and cognitive training scenario. In the experiment, participants performed basic arm exercises with a Pepper robot as a guide. We extracted facial features from video recordings of participants and applied a recursive feature elimination algorithm to select a subset of discriminating facial features. These features are correlated with the performance of the user and the level of difficulty of the exercises. The long-term aim of this work, building upon the work presented here, is to develop an algorithm that can eventually be used in robot-assisted training to allow a robot to tailor a training program based on the physical capabilities as well as the social cues of the users.

Keywords: social cues, facial signals, robot-assisted training.

1 Introduction

For a social robot that shares a space with humans, being able to communicate appropriately and adapting its behavior based on the users' needs, as well as understanding social context are important skills. Social robots have the potential to provide both companionship and physical exercise therapy.

Social signals are communicative or informative signals that provide information, directly or indirectly, about social interactions [18]. One of the main advantages of social signals is that they are continuous and can be all gathered throughout the interaction. Thus, they may be used as part of an adaptation process for considering the user's point of view and preferences. Despite the fact that there are studies [12,19] using social signals as part of adaptation of robot behaviors, these studies did not consider the understanding of key social signals in a particular scenario such as robot-assisted training. In this paper, we address the understanding of key facial signals including emotions and facial actions (FA) during a physical and cognitive training with a social robot. In the

experiment, participants performed basic arm exercises with the Pepper robot. We obtained facial features by using Affdex SDK [16] and applied a recursive feature elimination (RFE) algorithm to select a subset of discriminative facial features.

The contributions of this paper include: (i) identifying the relevant facial features during a physical and cognitive training with a social robot; (ii) discovering small subset of common facial features as *valence*, *lip corner depressor*, *lip pucker*, and *eye closure*; (iii) leaving an open possibility to use the identified features in the adaptation of the robot training program. In the remainder of this paper, an overview of related studies is given in Section 2. The method, experimental design and procedure are described in Section 3. The detailed explanation of data analysis is given in Section 4. Experimental results are presented in Section 5, and the paper is concluded in Section 6.

2 Related Work

2.1 Robot-Assisted Training

Research on robot-assisted training is vast and covers applications for cognitive and physical assistance such as post-stroke rehabilitation, cognitive assistance for dementia, therapy for Autism Spectrum Disorders, tutoring systems for deaf children and for language learning.

There are studies that employ robots as exercise coaches or exercise assistants [9,10,11,13]. In [11], authors investigated the role of embodiment in exercise coaches by comparing physically and virtually embodied coaches. Their results showed that elderly participants preferred the physically embodied robot coach over the virtual coach regarding its enjoyableness, helpfulness, and social attraction. Authors also assessed the motivational aspects of the scenario [10]. The robot provided feedback based on the task success which helped to keep participants more engaged. Similarly, in [13], the Nao humanoid robot was a training coach with a positive or corrective feedback mechanism for elderly people. In [9], authors presented an engagement-based robotic coach system with one-on-one and multi-user interactions. Their experimental results showed that the robotic coach was positively accepted by older adults either with or without cognitive impairment.

As these studies emphasize, a more engaging training robot could increase the enthusiasm of elderly people to exercise more and to sustain long-term interaction. We claim that factoring in social signals in training robots is just as important as the feedback mechanism.

2.2 Facial Features

The human face affords various inferences at a glance such as gender, ethnicity, health and emotions. Moreover, it is one of the richest and most powerful tools in social communication. Facial expressions analysis has been used in a variety

of applications for understanding users and enhancing their experience. Affective facial expressions can facilitate robot learning when incorporated within a reward signal in reinforcement learning (RL) [3]. These features, which are capable of boosting the robot learning in a particular scenario, need further investigation.

The Facial Action Unit Coding System (FACS) is a taxonomy that is geared towards standardizing facial expression measurement by describing and encoding movements of the human face and FAs [7].

The human-in-the-loop approaches provide human assistance explicitly or implicitly to the learning agent. The two facial affective expressions of smile as positive and fear as negative reward were incorporated with the environment reward to enhance the agent learning in continuous grid-world mazes [3]. In [12], the authors presented a tutoring scenario. The facial affective states of the student were combined into a reward signal in order to provide personalized tutoring. Another study, [19], used social signals as a reward by considering body and head positions to obtain engagement of the user and adapt the robot’s personality accordingly to keep the user engaged in the interaction.

3 Method

Our methodology consists in a physical memory game with three difficulty levels: easy, medium and hard. By using different levels of difficulty, we expected to elicit different facial expressions in the participants. The participants’ interactions served the purpose of collecting data via video recordings and questionnaires. We recorded all of the interactions using a video camera focused at the users’ faces.

The robot used in our experiments was Pepper, a humanoid robot with 20 degrees of freedom (DOF), height of 1.2 m and a weight of 29 kg. The DOF in the arms are as follows: 2 DOF in each shoulder, 2 DOF in each elbow, and 1 DOF in each wrist [1].

3.1 The Experiment

Each session began with informing the participant about the experiment and the robot. Before starting, the participant read and filled out the consent form. Thereafter, the robot welcomed the participant and provided information about the training program. The participant started with a warm-up session to familiarize them with the experimental procedure. Participants tended to start the exercise during the demonstration and it took them some time to understand that they were supposed to wait before repeating until it was over. Therefore warm-up was helpful. Further information is given in Section 3.2.

Before conducting experiments, we had pre-experiments with two participants. The early experiments helped to adapt different experimental parameters such as the length of the warm-up session, number of exercises, the required time for each level and the questionnaires, the participants’ proximity to the robot and the positioning of the video camera. Thirteen participants (6 male, 7 female)

took part in the experiments ($\mu_{age}=32.28$, $\sigma_{age}=6.08$). An illustration of the experiment is shown in Figure 1.

Figure 1: A participant exercising with the Pepper robot.

3.2 Procedure

The following procedure was used for each participant:

1. The researcher instructed the participant about the experiment and robot.
2. The participant read and signed the consent form. The participant was also informed that there are three levels and between each level, they would fill out a questionnaire.
3. The robot introduced itself and explained the experimental procedure.
4. The participant started with a warm-up session (~ 3 mins).
5. Thereafter, the easy level including 10 exercises started, 5 of them had one (e.g. arms up, arms side, etc.) and 5 of them had two successive arm exercises (e.g. arms up and arms to the side, arms in front and arms up, etc.). At the beginning of each exercise, there was a beep sound indicating that a new exercise was about to start.
6. At the end of the easy level, the participants filled out the following Likert scale questionnaire in which each question had options ranging from ‘Strongly disagree’(1) to ‘Strongly agree’(5):
 - The exercises were easy to execute (in terms of physical effort)
 - The exercises were easy to remember
 - I would like to exercise with the robot again
7. The experiment continued with the medium and hard level consecutively. The medium level comprised three (e.g. arms up and arms in front and arms to the side, etc.) and four successive arm exercises whereas hard level comprised five and six successive arm exercises. The participant filled out the same questionnaire after each level. At the end, they also provided demographic information (age, gender).

The sequence of exercises is a random combination of basic arm exercises given in Figure 2. Our aim is both physical and cognitive training. Since participants need to remember the sequence of exercises during the scenario, the level of difficulty is based on the number of exercises. It starts with one exercise and increases up to six exercises based on the average human memory span, 7 ± 2 [17].

Figure 2: Pepper robot: performing basic arm exercises.

4 Data Analysis

The facial features analysis was carried out over datasets created by participants exercising with the Pepper robot. The facial features of the participants were obtained from their video recordings by using Affdex SDK [16], and manually annotated based on different criteria such as:

- determination (first exercise of the sequence: the participants is challenged or the last exercise of the sequence: the participant is unchallenged);
- performance of the participant (the participant performed the exercise correctly or not);
- the difficulty level of the exercise (easy, medium and hard).

Determination Dataset: Motivation is an important factor for sustained exercising. The motivation of participants was not the same in the first and the last exercises of the sequence. Self-determination theory (SDT) examines the effects of different types of motivation that is a basis of behavior [6]. Optimally challenging activities are intrinsically motivating [5]. We also observed that the motivation of participants decreased towards the last sequence of the exercise. We isolated the first (challenged) and the last (unchallenged) sequence of exercises to constitute the determination dataset.

Performance Dataset: To constitute the performance dataset, an annotator watched and annotated the video recordings based on the sequence of exercise correctness of the participants.

Difficulty Dataset: The difficulty dataset includes the beginning and ending of each level (e.g. beginning time of easy level and ending time of easy level, etc.)

We segmented each participant’s data by correlating the time stamps in the features file and the annotation file. We applied moving average (window size 10) and random sampling. All participants’ data were combined to form the resulting dataset. Based on the different criteria mentioned above, the datasets described in Table 1 were created. An example line from the raw data is given in Figure 3.

The set of features in the datasets include:

- emotions: anger, contempt, disgust, fear, joy, sadness, and surprise

- engagement: facial expressiveness of the participant
- valence: the pleasantness of the participant
- FAs: brow furrow, brow raise, cheek raise, chin raise, dimpler, eye closure, eye widen, inner brow raise, jaw drop, lid tighten, lip corner depressor, lip press, lip pucker, lip stretch, lip suck, mouth open, nose wrinkle, smile, smirk, and upper lip raise
- head orientation: pitch, yaw, and roll
- label: the corresponding criteria

Table 1: Summary of datasets.

Dataset	Features	Examples	Classes
Determination	33	18382	challenged, unchallenged
Performance	33	39984	correct, wrong
Difficulty	33	198420	easy, medium, hard

TimeStamp	pitch	yaw	roll	joy	fear	disgust	sadness	anger	surprise	contempt	valence
13.4	-20.4097	-4.9017	1.67	0.002	0.0018	0.3955	0.0829	0.002	0.1952	0.1881	-27.423
	engagement	smile	innerBrowRaise	browRaise	browFurrow	noseWrinkle	upperLipRaise	lipCornerDepressor	chinRaise	lipPucker	lipPress
	3.3273	0.6481	0.0002	0.0002	0.0904	0.0006	0	52.741	0.264	0.0019	0.1669
	lipSuck	mouthOpen	smirk	eyeClosure	attention	eyeWiden	cheekRaise	lidTighten	dimpler	lipStretch	jawDrop
	0.9608	0.7347	0.2151	0.0004	96.8313	0	0.0069	0.5775	0.5248	0.0085	0.0116

Figure 3: An example line from raw data.

4.1 Feature Selection Method

Feature selection is widely used to select a subset of the original inputs by eliminating irrelevant inputs. RFE is an iterative procedure based on the idea of training a classifier, computing the ranking criterion, and removing the feature that has the smallest ranking criterion [14]. The most common version of RFE uses a linear Support Vector Machine (SVM-RFE) as classifier [14]. An alternative method is using Random Forest (RF) [2] instead of SVM as a classifier in the RFE. RF [2] is an effective prediction tool used in classification and regression. It consists of a combination of decision tree predictors such that each tree depends on the randomly sampled examples.

The RFE with RF was applied to all three datasets. The top five features were selected based on the feature importance. The selected features in each dataset are given in Table 2.

One of the variable importance measures in RF is mean decrease accuracy (MDA). The index MDA is used for quantifying the variable importance that is measured by change in prediction accuracy in exclusion of a single variable [4]. The variables that generate a large mean decrease in accuracy are more important for classification. MDA for each dataset is given in Figure 4.

Table 2: Selected features.

Dataset	Determination	Difficulty	Performance
Features	Valence	Lip corner depressor	Valence
	Cheek raise	Lip pucker	Lip pucker
	Lip corner depressor	Inner brow raise	Lip corner depressor
	Eye closure	Lid tighten	Cheek raise
	Lip pucker	Cheek raise	Eye closure

Figure 4: Mean decrease accuracy of features in each dataset.

The four selected features of *valence*, *lip corner depressor*, *lip pucker* and *eye closure* are common in each dataset. A figure depicting the selected features on a time scale is given in Figure 5. RF-RFE was experimentally found to be the best option for our approach. Due to the size of the difficulty dataset, applying another feature selection algorithm such as SVM-RFE and genetic algorithm was challenging.

5 Results and Discussion

This study sets out to understand the facial signals during a physical and cognitive training scenario. It is important to understand the facial signals that may help in the adaptation process of the robot behaviors. Using social signals in the adaptation process is beneficial as discussed in Section 2.2. The discriminative features were found to be *valence*, *lip corner depressor*, *lip pucker* and *eye closure*. Four selected features out of five are common in datasets. *Lip corner depressor* has been shown to be related to lack of focus [20] and *eye closure* is related to boredom [15]. *Lip pucker* is associated to thinking [8]. It means that

Figure 5: Selected features on a time scale (on the side a frame from the video at the given moment).

the participants lost their attention and got bored throughout the interaction. The robot may catch these signals to re-engage the participant to the interaction. This is the first step towards the adaptation process of robot behaviors. In [12], authors used facial valence and engagement as a part of the reward signal, they found that engagement did not change in response to the robot’s behaviors. We believe that investigating the social signals in a particular scenario is important for understanding the informative features that can improve the adaptation process.

The purpose of employing the questionnaire was to validate that the level of difficulty was increased from easy level to hard level. The results have statistically significance in the question “the exercises were easy to remember”. A one-way between subjects analysis of variance (ANOVA) was conducted to compare the effect of number of exercises on perceived difficulty; easy, medium and hard. There was a significant effect of number of exercises on perceived difficulty for the three conditions [$F(2, 30) = 14.06, p < .05$]. A post hoc Tukey HSD test showed that the perceived difficulty in easy level and hard level; and medium level and hard level differed significantly at $p < .05$; the perceived difficulty in easy level and medium level was not significantly different. Taken together, these results suggest that increasing number of exercises really do have an effect on perceived difficulty. Even though the mean is decreasing through hard level, there is no statistically significant difference in the questions “the exercises were easy to execute” and “I would like to exercise with the robot again”. A Pearson analysis between the performance of participants and the perceived difficulty

(“the exercises were easy to remember”) indicated a strong relationship between the variables. The performance decreases with perceived difficulty ($r = .98$, $p < .05$).

6 Conclusion and Future Work

In this work, we focused on the analysis of user facial expressions during a physical and cognitive training scenario. We applied the RF-RFE feature selection method to obtain informative features in the given scenario. The interaction videos of the participants were annotated and created three different datasets to study facial expressions under different conditions. This study is a first step towards understanding the social signals that will enable us to incorporate appropriate facial signals in the adaptation process and formulating the problem as reinforcement learning. The most common discriminative features are: *valence*, *lip corner depressor*, *lip pucker* and *eye closure*. The features, other than *valence*, are related to negative feelings so they can be used as a punishment in the adaptation of the robot behavior. The contribution of the present work lies in providing a basis for the choice of different facial features in different conditions during a training scenario. In our future work, we will perform closed-loop experiments involving implicit human-robot interaction based on multi-modal social cues including the facial features obtained in this study. We will formulate the problem as reinforcement learning and alter the robot’s behavior in order to address human needs.

7 Acknowledgement

This work has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 721619 for the SOCRATES project.

References

1. Pepper humanoid robot. <https://www.softbankrobotics.com/emea/en/robots/pepper/find-out-more-about-pepper>. Accessed: 2018-06-30.
2. Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.
3. Joost Broekens. Emotion and reinforcement: affective facial expressions facilitate robot learning. In *Artificial intelligence for human computing*, pages 113–132. Springer, 2007.
4. M Luz Calle and Víctor Urrea. Letter to the editor: stability of random forest importance measures. *Briefings in bioinformatics*, 12(1):86–89, 2010.
5. Fred W Danner and Edward Lonky. A cognitive-developmental approach to the effects of rewards on intrinsic motivation. *Child Development*, pages 1043–1052, 1981.
6. Edward L Deci and Richard M Ryan. The” what” and” why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4):227–268, 2000.

7. Paul Ekman and Erika L Rosenberg. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.
8. Rana El Kaliouby and Peter Robinson. Real-time inference of complex mental states from facial expressions and head gestures. In *Real-time vision for human-computer interaction*, pages 181–200. Springer, 2005.
9. Jing Fan, Dayi Bian, Zhi Zheng, Linda Beuscher, Paul A Newhouse, Lorraine C Mion, and Nilanjan Sarkar. A robotic coach architecture for elder care (rocare) based on multi-user engagement models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(8):1153–1163, 2017.
10. Juan Fasola and Maja J Mataric. Using socially assistive human-robot interaction to motivate physical exercise for older adults. *Proceedings of the IEEE*, 100(8):2512–2526, 2012.
11. Juan Fasola and Maja J Matarić. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2(2):3–32, 2013.
12. Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. Affective personalization of a social robot tutor for children’s second language skills. In *AAAI*, pages 3951–3957, 2016.
13. Binnur Görer, Albert Ali Salah, and H Levent Akin. An autonomous robotic exercise tutor for elderly people. *Autonomous Robots*, 41(3):657–678, 2017.
14. Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.
15. Bethany McDaniel, Sidney D’Mello, Brandon King, Patrick Chipman, Kristy Tapp, and Art Graesser. Facial features for affective state detection in learning environments. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 29, 2007.
16. Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 3723–3726. ACM, 2016.
17. George A Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.
18. Isabella Poggi and Francesca D’Errico. Social signals: A psychological perspective. In *Computer analysis of human behavior*, pages 185–225. Springer, 2011.
19. Hannes Ritschel, Tobias Baur, and Elisabeth André. Adapting a robots linguistic style based on socially-aware reinforcement learning. In *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*, pages 378–384. IEEE, 2017.
20. Alexandria K Vail, Joseph F Grafsgaard, Kristy Elizabeth Boyer, Eric N Wiebe, and James C Lester. Predicting learning from student affective response to tutor questions. In *International Conference on Intelligent Tutoring Systems*, pages 154–164. Springer, 2016.