Research Traineeships proposal 2020-2021

1 Title of the Project

Collecting the *Haptic greetings corpus* for the study and automatic recognition of social touch.

2 Coordinators

- Dr. Merel Jung is an assistant professor in the department of Cognitive Science and Artificial Intelligence (DCA, TSHD)
- Dr. Martijn Goudbeek is an associate professor in the department of Cognition and Communication (DCC, TSHD)

3 Project summary

3.1 Project overview

Touch behavior is seen in many different forms of social interaction: a handshake as a greeting, a high-five to celebrate a joint accomplishment, or a comforting hug from a friend. In contrast to visual and vocal modalities, the haptic modality of interpersonal touch has not received much research attention [2, 6]. Consequently, the haptic modality is often overlooked in interactions with embodied or virtual agents [19]. However, as interactions with social robots will become more common in the near future, robots are expected to engage in adequate haptic interaction with humans [12, 19]. Greeting is the hallmark example of social behavior in which touch occurs frequently and serves an important function in the management of interpersonal relations by welcoming a guest or acknowledging a friend or an acquaintance [4, 8, 14]. Socially intelligent robots should therefore be able to initiate and respond to haptic greetings such as a handshake or a hug. In order to systematically study and model haptic greeting behavior the aims of this research proposal are: (I) collecting a corpus of human haptic greetings (Project 1) and (II) developing a deep learning model based on the collected corpus for the automatic recognition of haptic greeting behaviors (Project 2).

Touch is a complex modality as the behaviors involved vary in intensity, movement, velocity, abruptness, temperature, location, and duration [5]. Moreover, the meaning of touch can often not be inferred from the type of touch alone, but is also dependent on other factors such as concurrent verbal and nonverbal behavior, the type of interpersonal relationship [7, 18] and the situation [8, 11, 17]. Additionally, an observational study at an airport showed that the type of touch behavior in interpersonal greetings also depends on the gender-composition of dyads [4]. Males tended to briefly shake hands whereas female-female and cross-sex dyads displayed relatively longer physical contact involving hugging, kissing and hand-to-face/upper-body touches. Rather than using an observational study, we propose to collect a corpus of video sequences of haptic greetings for the systematic study of touch behavior in different social contexts. Moreover, deep learning models will be trained for automatic recognition based on the

video sequences of the collected haptic greeting behaviors [15]. The collection and public release of corpora of human behavior is an effective way to crowd-source the study of social signals such as facial and vocal expressions [1, 3] and has been used previously for the automatic recognition of social touch gestures [9, 10].

3.2 Project 1: Collecting the *Haptic greetings corpus*

The aim of Project 1 is to collect labeled video sequences of human greeting interactions in a variety of social scenarios. Human greeting behaviors will be systematically collected by instructing participants to perform greetings in different scenarios (e.g., greet a friend vs. someone you just met) using a within-subject design. As previous studies indicate that touch behavior in greetings is influenced by gender-composition (see, for example [4, 8, 14]), greetings will be collected from dyads with different compositions. The interaction scenarios are designed based on different roles of social robots in society, for example, greetings between friends match the peer relationship with a companion robot. Labeling will be handled by asking the participants to indicate the type of greeting (e.g., handshake, hug, high-five, or fist bump) used after each interaction.

In the field of human activity recognition, datasets containing videos of touch interactions (e.g. handshakes, high-fives, and hugs) from a third-person view have been collected to develop models for surveillance applications [20, 21]. However, since one of the future applications of the corpus will be the development of haptic greeting behaviors for social robots, a third-person view does not suffice. Social robots require touch sensing capabilities that work from a first-person point of view, since they see and feel their interaction partner from that perspective. Therefore, the greeting behaviors will be captured in a triple-view fashion: two from a first-person point of view by equipping both persons with a wide-angle camera and one with a traditional third-person view.

3.3 Project 2: Recognizing haptic greetings using deep learning

The aim of Project 2 is to train a deep learning model using the labeled triple-view video sequences from the *Haptic greetings corpus* collected in Project 1 in order to automatically recognize haptic greeting behaviors. Deep neural networks, which learn data representations from raw data (in our case from video sequences) instead of relying on hand-crafted features, have revolutionized the field of computer vision by significant improving the performance in on tasks such as object recognition and detection [15]. The models will be pre-trained on video material from related datasets (e.g. [16, 20]) before the model is fine-tuned on the *Haptic greetings corpus*. To take the spatiotemporal nature of the video sequences into account, the architecture of the deep learning model will be based on those commonly used in the field of human action recognition such as 3D convolutional neural networks or a combination of a convolutional neural network and a long short-term memory recurrent neural network [13]. The labeled *Haptic greetings corpus* will be made publicly available after the publication of our own results.

4 Project timeline

Research trainee 1 will work on Project 1 for the entire 10 month period whereas research trainee 2 will work on Project 1 for the first 6 months and will then continue to work on Project 2 for the last 4 months (see Table 1). Research trainee 1 will take the lead in preparing the instructions for the participants including the social scenarios that the participants will be acting out. Meanwhile, research trainee 2 will take the lead in preparing the set up for the video recordings in order to capture the interactions. The data collection procedure and video recording setup will be tested by running a pilot, adjustments to the procedure and technical set up will be made if necessary. During data collection research trainee 1 will take on the role of the experimenter, welcome the participants and make sure that the instructions

are followed. Research trainee 2 will be in charge of the technical setup during data collection making sure that the interactions are correctly captured. After data collection the research trainees will work together to prepare the video data for analysis (e.g. splitting the raw footage into separate interactions). Then, research trainee 1 will conduct a descriptive analysis of the video footage. Meanwhile, research trainee 2 will start working on Project 2, developing a deep learning model for the automatic recognition of the haptic greetings. In the last two months the research trainees will each write a conference paper. Research trainee 1 will write a paper on the collected corpus and the results of the descriptive analysis whereas research trainee 2 will write a paper on the classification results.

Table 1: Timeline of Projects 1 and 2 specified per research trainee

Project 1: Collection of the *Haptic greetings corpus*		
Month	Research trainee 1	Research trainee 2
1-2	Preparation of participant instructions	Prepare setup video recording
3	Run pilot of data collection	
4-5	Data collection haptic greetings corpus	
6	Data cleaning and pre-processing	
7-8	Descriptive video analysis	
9-10	Writing paper on corpus	
Project 2: Recognizing haptic greetings using deep learning		
Month	Research trainee 1	Research trainee 2
7-8		Development of deep learning model
9-10		Writing paper on classification results

5 Research trainee profile

We are looking for two enthusiastic students from TSHD, preferably at the (research) master level (though excellent bachelor students are also considered). The students to be hired must have a student status throughout their employment as research trainees.

5.1 Student profile for research traineeship 1 on collecting and analyzing social interactions

The applicant should have a strong interest in studying human social behavior. Experience with conducting studies with human participants and video analysis will be helpful. The research trainee will acquire experience in setting up and running a data collection of acted social interactions, descriptive video analysis, and scientific writing.

5.2 Student Profile for research traineeship 2 on social signal processing

The applicant should have a strong interest in the automatic recognition of human social behavior (social signal processing). Moreover, the applicant should have experience in training and evaluating machine learning models using Python. Experience in video processing and deep learning will be helpful. The research trainee will acquire experience in recording human behavior, video analysis using deep learning, and scientific writing.

5.3 How to apply

Please send your application, including a motivation letter and CV, to both Merel Jung (m.m.jung@uvt.nl) and Martijn Goudbeek (m.b.goudbeek@uvt.nl).

References

[1] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon. From individual to group-level emotion recognition: Emotiw 5.0. In *Proceedings of the 19th ACM international conference on multimodal interaction*, pages 524–528. ACM, 2017.

[2] A. Gallace and C. Spence. The science of interpersonal touch: an overview. *Neuroscience & Biobehavioral Reviews*, 34(2):246–259, 2010.

[3] M. Goudbeek and M. Broersma. The demo/kemo corpus: A principled approach to the study of cross-cultural differences in the vocal expression and perception of emotion. In *7th International Conference on Language Resources and Evaluation (LREC 2010)*, pages 2211–2215. ELRA, 2010.

[4] P. E. Greenbaum and H. M. Rosenfeld. Varieties of touching in greetings: Sequential structure and sex-related differences. *Journal of Nonverbal Behavior*, 5(1):13–25, 1980.

[5] M. J. Hertenstein, R. Holmes, M. McCullough, and D. Keltner. The communication of emotion via touch. *Emotion*, 9(4):566–573, 2009.

[6] M. J. Hertenstein, J. M. Verkamp, A. M. Kerestes, and R. M. Holmes. The communicative functions of touch in humans, nonhuman primates, and rats: a review and synthesis of the empirical research. *Genetic, Social, and General Psychology Monographs*, 132(1):5–94, 2006.

[7] R. Heslin, T. D. Nguyen, and M. L. Nguyen. Meaning of touch: The case of touch from a stranger or same sex person. *Journal of Nonverbal Behavior*, 7(3):147–157, 1983.

[8] S. E. Jones and A. E. Yarbrough. A naturalistic study of the meanings of touch. *Communications Monographs*, 52(1):19–56, 1985.

[9] M. M. Jung, X. L. Cang, M. Poel, and K. E. MacLean. Touch challenge'15: Recognizing social touch gestures. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 387–390. ACM, 2015.

[10] M. M. Jung, M. Poel, R. Poppe, and D. K. J. Heylen. Automatic recognition of touch gestures in the corpus of social touch. *Journal on multimodal user interfaces*, 11(1):81–96, 2017.

[11] M. M. Jung, M. Poel, D. Reidsma, and D. K. J. Heylen. A first step towards the automatic understanding of social touch for naturalistic human-robot interaction. *Frontiers in ICT*, 4(3), 2017.

[12] M. M. Jung, L. van der Leij, and S. M. Kelders. An exploration of the benefits of an animallike robot companion with more advanced touch interaction capabilities for dementia care. *Frontiers in ICT*, 4:16, 2017.

[13] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, et al. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*, 2017.

[14] A. Kendon. *Conducting interaction: Patterns of behavior in focused encounters*, volume 7. CUP Archive, 1990.

[15] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[16] A. Patron-Perez, M. Marszalek, A. Zisserman, and I. D. Reid. High five: Recognising human interactions in tv shows. In *BMVC*, volume 1, page 33. Citeseer, 2010.

[17] D. Silvera-Tawil, D. Rye, and M. Velonaki. Artificial skin and tactile sensing for socially interactive robots: A review. *Robotics and Autonomous Systems*, 63:230–243, 2015.

[18] J. T. Suvilehto, E. Glerean, R. I. Dunbar, R. Hari, and L. Nummenmaa. Topography of social touching depends on emotional bonds between humans. *Proceedings of the National Academy of Sciences*, page 201519231, 2015.

[19] J. B. F. Van Erp and A. Toet. Social touch in human–computer interaction. *Frontiers in Digital Humanities*, 2(2):1–14, 2015.

[20] C. van Gemeren, R. Poppe, and R. C. Veltkamp. Hands-on: deformable pose and motion models for spatiotemporal localization of fine-grained dyadic interactions. *EURASIP Journal on Image and Video Processing*, 2018(1):16, 2018.

[21] K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, and D. Samaras. Two-person interaction detection using body-pose features and multiple instance learning. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 28–35. IEEE, 2012.