Human perception baseline for anticipating haptic greetings in human-robot interaction

Merel M. Jung[1] and Martijn Goudbeek[2]

I. Introduction

Greeting is the hallmark example of social behavior in which touch occurs frequently and serves an important function in managing interpersonal relations by welcoming a guest or acknowledging a friend [1]. Socially intelligent robots should therefore be able to initiate and respond to haptic greetings such as a handshake or a hug. While anticipating touch interactions from an exocentric perspective [2] is useful for some applications, social robots would require anticipation of human actions from an egocentric perspective. To establish a baseline, a perception study was conducted to compare how accurately humans can anticipate different types of haptic greetings from an egocentric and exocentric view.

II. Methods

A corpus of video recordings of dyads performing haptic greetings was collected from an egocentric perspective (head-mounted camera) and exocentric perspective (see Fig. 1) which served as stimuli for the perception study to establish the human baseline. Participants ($n = 49$) were presented with video fragments from the viewpoint of the responder, who was unaware of the greeting that would be initiated, starting from the approach by the initiator of the instructed greeting till the first moment of physical contact ($\approx$ 1-2 seconds). Participants watched 54 interactions, 27 from each perspective. After each video, participants were asked to predict the action based on nine multiple-choice options: high five, fist bump, hug, arm touch, handshake, elbow bump, hold hands, shoulder tap, just passing (no physical interaction).

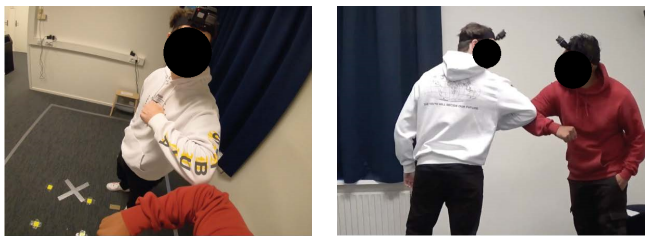

Fig. 1. Screenshots of an elbow bump greeting (initiator in white sweater, responder in red sweater): left - egocentric view of the responder; right - greeting from the exocentric perspective.

[1]Merel M. Jung is with the department of cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, The Netherlands. e-mail: m.m.jung@tilburguniversity.edu.

[2]Martijn Goudbeek, is with the department of Communication and Cognition, Tilburg University, Tilburg, The Netherlands. e-mail: m.b.goudbeek@tilburguniversity.edu.

III. Results

Average accuracies for the egocentric perspective were found to be significantly higher than for the exocentric perspective ($t(48) = 7.42, p < .001$), with a large-sized effect ($d = 1.06$). Table I shows the accuracies per greeting for the two perspectives (guessing baseline is $1/9 \approx .11$).

TABLE I

Accuracy scores and differences between greeting predictions from an egocentric and exocentric perspective

	Mean accuracy (SD)	
Greeting	Egocentric	Exocentric
Fist Bump*	.52 (.30)	.17 (.25)
Arm Touch*	.28 (.24)	.12 (.18)
Handshake*	.72 (.32)	.51 (.27)
Elbow Bump*	.25 (.22)	.05 (.12)
Just passing*	.55 (.41)	.26 (.35)
High Five	.19 (.22)	.14 (.17)
Hug	.54 (.36)	.59 (.28)
Hold Hands	.31 (.31)	.39 (.28)
Shoulder Tap	.21 (.20)	.25 (.28)
Overall*	.40 (.49)	.28 (.45)

*indicates significant differences.

IV. Discussion & conclusion

Predictions of greetings from the egocentric perspective were found to be more accurate even though the exocentric perspective provides more information regarding the postures and movements of both persons involved. Our findings might be explained by the closeness of the egocentric videos to how humans naturally perceive their environment. Generally, low accuracies seem to have been the result of confusions between similar greetings (e.g., fist bump and elbow bump). In future work we will compare the human baseline results to the performance of machine learning models trained on our haptic greeting dataset. Automatic prediction of greetings will enable robots to anticipate greetings leading to more natural human-robot interaction.

Acknowledgment

The authors would like to thank Debbie Schoone, Sam Aras, Nataliia Vdovychenko, and Frank Arts for their help with data collection.

References

[1] S. E. Jones and A. E. Yarbrough, "A naturalistic study of the meanings of touch," *Communications Monographs*, vol. 52, no. 1, pp. 19–56, 1985.

[2] H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, "Predicting the future: A jointly learnt model for action anticipation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5562–5571.