Moveable Facial Features in a Social Mediator

Muhammad Sikandar Lal Khan[1], Shafiq ur Réhman*✉[1,2], Yongcui Mi[1], Usman Naeem[2], Jonas Beskow[3], and Haibo Li[3]

[1] Umeå University, SE-90187 Umeå, Sweden.
[2] University of East London, E16 2RD, United Kingdom.
[3] The Royal Institute of Technology (KTH), SE- 114 28 Stockholm, Sweden.

Abstract. Human face and facial features based behavior has a major impact in human-human communications. Creating face based personality traits and its representations in a social robot is a challenging task. In this paper, we propose an approach for a robotic face presentation based on moveable 2D facial features and present a comparative study when a synthesized face is projected using three setups; 1) 3D mask, 2) 2D screen, and 3) our 2D moveable facial feature based visualization. We found that robot's personality and character is highly influenced by the projected face quality as well as the motion of facial features.

1 Introduction

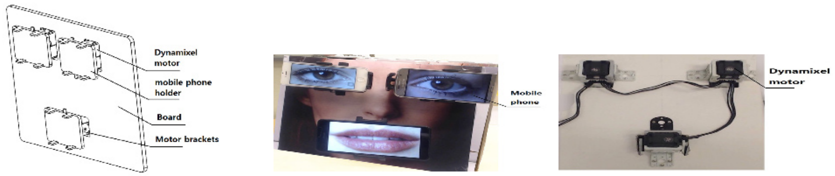
When it comes to judging the human personality traits, the certain face regions and their motions may contain more information than others; for instance the eyes and the lips regions [5, 4]. Therefore, human face with moveable facial features modeling as an interface has always been a major interest for human-robot interactions. Social information (e.g., personality traits) can be accurately perceived from dynamic body motion [6], however it is *unclear* how the motion of facial features are related to the perceived personality and/or projected face. In this work, we consider *how well can a human observer recognize a presented face*? To investigate this, we have prototyped a moveable facial feature robotic (MFFR) system and performed a comparative study of face (personality) recognition considering various faces (and moveable facial features) projection methods; 1) 3D facial mask based Furhat [1], 2) 2D screen, and 3) our proposed *moveable facial feature robotic (MFFR) system.*

2 Moveable Facial Feature Robotic (MFFR) System

We have designed a moveable facial feature robotic (MFFR) system which represents a brief moveable display of non-verbal facial feature behavior. It has three components; i) a mechanical platform, ii) facial features capturing module and iii) a control module. The MFFR system is designed to represent a pilot users facial features in different social settings. The developed electromechanical setup mimics the eyes and lips motion of a pilot user as shown in Fig. 1. The

* Corresponding authors: shafiq.ureheman@umu.se

J. Beskow et al. (Eds.): IVA 2017, LNAI 10498, pp. 205-208, 2017.
DOI 10.1007/978-3-319-67401-8_23



(a) A CAD model. (b) A View of Electromechanical platform. (c) Back View of our prototype.

Fig. 1: The Moveable Facial Feature Robotic (MRRF) system which provides an active 'visualization' - a synchronous and simultaneous display of eyes and lips regions to represent a human face.

(a) Furhat [1] - a 3D projection. (b) 2D screen - a flat projection. (c) MFFR - a moveable feature projection.

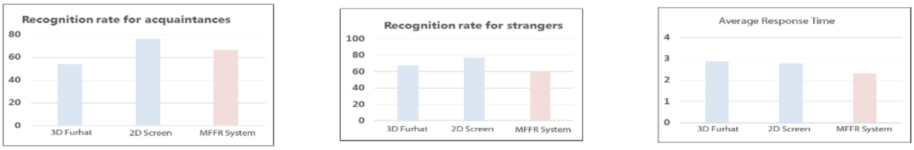
Fig. 2: An Experimental Setup for three projection methods.

smart phones 2D screens are used to display vital parts of a human face. These 2D screens are attached to actuators for mimicking the facial expressions. To project the facial expressions of a person on our MFFR system, the head pose estimation algorithm [3] and facial features (such as eyes, nose, lips, eye-brows, and face boundary)segmentation using Haar-feature based cascade classifiers for human face detection [7] and w Constrained Local Model (CLM) approach [2]. We then present the most important face regions, more specifically eyes and lips regions on to the MFFR system.

3 Experimental Study

In this work, we consider a comparative experiment among three systems, i) 3D Furhat, ii) 2D screen and iii) MFFR system. The experiment consisted of three sessions in which participants were asked to recognize the projected face representations. The session 1 and 2 were held at the Royal Institute of Technology (KTH), Sweden, and the 3rd session was conducted at Umeå University, Sweden. The three setups for this experiment are shown in Fig. 2. We had n=20 participants (ranging in age from 18 - 40) for each session of an experiment. The participants were briefed about the experimental setup and aim of the study. The participants were also explained how the face projection/visualizations would be displayed. Each session of our experiment is divided into two phases, i) training phase and ii) testing phase.

During the training phase, the participants were presented with 8 *known* projected faces with 3 orientations - Left (L), Center (C), Right (R); i.e.,*24 known*



(a) Acquaintance faces recognition

(b) Stranger (non-familiar) faces recognition

(c) The Average time for one face.

Fig. 3: Face recognition Accuracy and Response time of three face presentations.

stimuli. For memory building, these projected faces were repeated 3 times. In total we had 8 x 3 orientations x 3 repetitions = 72 projected face presentations. Each face presentation is about 3 seconds. The time of training phase was 72 x 3 seconds = 216 sec for each participant. During testing phase, 24 known presentations (from training phase 8C, 8L, 8R) were mixed with 24 unknown presentations (8C, 8L, 8R)) and were presented randomly to the participants and asked to press *'Y'* for "Yes"- for already seen face, otherwise *'N'* for "No"- not seen.

4 Results and Concluding Remarks

The participants completed the experimental study in a reasonable time duration and response rate was 100%. Our parameter of interest are *the accuracy of face-recognition* and *the response time of each participant*. The accuracy measure indicates how accurately a participant can recognize a projected face (Tab. 1). Response time indicates how long time a participant takes in making his/her decisions (Tab. 2). It turns out that the recognition rate of the MRRF is between 3D Furhat and 2D screen when it comes to acquaintances. The recognition rate of MRRF is lower than both 3D Furhat and 2D screen for strangers (to successfully say that they have not seen this presentation during training phase). It is very interesting that the moveable features help the users to make the decision faster as the response time drops as seen in Fig. 3c. From the Fig. 3a, it is clear that 2D screen performs better in face recognition compare to 3D face mask and MRRF system. Despite 2D screen performs better in face recognition, it is abortive in presenting different social cues (such as head gesture, gaze direction, etc.) which are important for different applications such as, video teleconferencing. The 3D face mask presents better social cues compare to 2D screen but it lacks the property of face recognition which is an important feature of social settings. However, the introduction of our MRRF system contains the properties of both systems. However, when it comes to the strangers (unknown presentations), the face recognition rate drops as compared to acquaintances as shown in Fig. 3b. In future studies, we will address this issue. It is very interesting that the moveable features help the users to make the decision faster as there is a decrease in response time (Fig. 3c).

The motivation of this work was to examine how well can a human observer recognize a presented face and what is an impact of moveable facial features in

Visualization Method	Faces	Test		Rate (%)
3D Furhat	ACQ (525)	YES	283	54
		NO	242	46
	STR (435)	YES	141	32.4
		NO	294	67.9
2D-Screen	ACQ (519)	YES	394	76
		NO	125	24
	STR (441)	YES	102	23.1
		NO	339	76.9
MFFR System.	ACQ (480)	YES	319	66.5
		NO	161	33.4
	STR (480)	YES	188	39,2
		NO	292	60.8

Table 1: Projected face recognition accuracy for three Systems. ACQ= Acquaintances, STR = Strangers

Time (s)	Face mask	2D 2D	MFFR System
Total	2755	2666	2225
Average	2.87	2.78	2.31

Table 2: All participants' Response time for all three systems.

social settings. This work presented a design of novel MFFR system. MFFR not only presents the important social gestures and cues (which 2D screen is *not* capable of) but also helps in face recognition (which 3D mask is *not* capable of). In our future work, we will optimize the program to reduce the response time of a system and employ an eye-tracker sensor to improve the accuracy of the eyes movements. Furthermore, we are also interested in identifying the key components for humanizing (human-like interface) the social robot when it comes to face and facial expressions.

References

1. Al Moubayed, S., Beskow, J., Skantze, G., Granström, B.: Furhat: a back-projected human-like robot head for multiparty human-machine interaction. Cognitive behavioural systems pp. 114–130 (2012)
2. Cristinacce, D., Cootes, T.F.: Feature detection and tracking with constrained local models. In: British Machine Vision Conference (BMVC). p. 3. No. 2 (2006)
3. Khan, M.S.L., ur Réhman, S., Lu, Z., Li, H.: Head orientation modeling: Geometric head pose estimation using monocular camera. In: IEEE/IIAE Int'l Conf. Intelligent Sys. and Image Proc. pp. 149–153 (2013)
4. ur Réhman, S., Liu, L.: ifeeling: Vibrotactile rendering of human emotions on mobile phones. In: Mobile multimedia processing, pp. 1–20. Springer (2010)
5. Schurgin, M., Nelson, J., Iida, S., Ohira, H., Chiao, J., Franconeri, S.: Eye movements during emotion recognition in faces. Journal of vision 14(13), 14–14 (2014)
6. Thoresen, J.C., Vuong, Q.C., Atkinson, A.P.: First impressions: Gait cues drive reliable trait judgements. Cognition 124(3), 261–271 (2012)
7. Viola, P., Jones, M.J.: Robust real-time face detection. International journal of computer vision 57(2), 137–154 (2004)