Do You Think I Approve of That? Designing Facial Expressions for a Robot

Peter E. McKenna(✉), Mei Yii Lim, Ayan Ghosh(✉), Ruth Aylett(✉), Frank Broz, and Gnanathusharan Rajendran

Heriot-Watt University, Riccarton EH14 4AS, UK
{p.mckenna,m.lim,ayan.ghosh,r.s.aylett,f.broz,t.rajendran}@hw.ac.uk

Abstract. This paper discusses initial work with a broad age range neuro-typical sample (N = 77), towards a training system in social signal recognition for high-functioning adults with an Autism Spectrum Disorder. Outlined is a principled design of four *approval* and four *disapproval* facial expressions for the EMYS robot head and the method and conduct of a pilot study 'in the wild', testing user recognition of these expressions. Results showed that recognition varied over the eight expressions, with some expressions perceived as ambiguous and one approval expression as disapproval. Expression type, ordering effects, age, and pre-study training are identified as important issues.

Keywords: Robot expressive behaviour · Autism · Social signals

1 Introduction

In the UK, Autism Spectrum Disorder (ASD) affects 547,000 people over the age of 18 (1.3% of working age adults) according to the 2011 census. These adults encounter serious difficulties in their everyday life, particularly in securing and maintaining employment. The unemployment rate among adults with ASD is higher than 85%, nearly double the unemployment rate of 48% for the wider disabled population and compares to an overall UK unemployment rate of 5.5%.

One reason for this is that people with an ASD struggle to correctly interpret social signals, those expressive behavioural cues through which people manifest what they feel or think (facial expressions, vocalisations, gestures, posture etc.)[1]. This leads to difficulties in correctly interpreting interactions with co-workers and supervisors.

Behavioural Skills Training (BST) [2] is recognized as one of the most effective training approaches for the effects of an ASD. BST is a behaviourist training approach involving phases of instruction, modelling, rehearsal, and feedback in order to teach a new skill [3]. It has been used to teach social skills to people both with and without disabilities [4]. However, BST is too labour-intensive to be widely applied. If robots could be used to help deliver BST, this could reduce the effort required by human trainers and lower the cost of BST application.

A. Kheddar et al. (Eds.): ICSR 2017, LNAI 10652, pp. 188–197, 2017.
https://doi.org/10.1007/978-3-319-70022-9_19

In the SoCoRo project[1], work is being carried out to design such a training system. As a first step, the design of expressive facial behaviours for the EMYS robot head (Fig. 3) was investigated, a set of expressions for *approval* and *disapproval* social signals designed, and a study was carried out 'in the wild' at the Glasgow Science Centre. Thus, the present study provided empirical grounding for our proposed future BST design. Here, visitors of all ages (range = 4–77 years) individually observed the expressions to determine whether they are perceived as intended. Below we discuss the design of the expressions, the methodology, and the results derived from the analysis.

2 Designing Expressive Facial Behaviours

In this section we discuss existing work in designing facial expressions for robots and explain the approach taken for the expressions used in the study.

2.1 Relevant Work in Expressive Behaviour Design

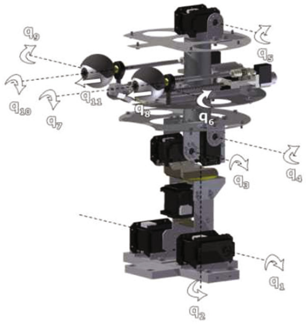
Research on facial expressions has been dominated by two judgement procedures: categorical, involving basic emotion categories [5] where universality and discreteness are central; and dimensional, involving scales or dimensions (for example: Pleasure, Arousal, Dominance) that underlie the emotion categories [6,7] where temporal dynamic is the focus. Given that a particular emotion can usually be represented by more than one facial expression with varying intensity, the discrete approach suffers from the flaw of rigidity with its one-to-one mapping. Moreover, anything that could be described as a static expression is very rarely observed in normal interaction where expressive behaviour is continuously modulated in the evolving context. The dimensional approach on the other hand is able to convey a wide range of affective messages seamlessly and supports such modulation in a natural way.

We have therefore taken this approach in the current work, focusing on two groups of expressions supporting important social signals: approval and disapproval. The aim is for EMYS to express continuous internal state so that the resulting social signals are more 'human-like', pose the advantage of increased ecological validity [8] and hence, enable transfer of learning from robot to human incrementally in line with the Reduced Generalisation Theory [9].

2.2 Design of the Expressions Used in the Study

EMYS has minimalist facial features with only 11 degrees of freedom (DOFs) as shown in Fig. 1. This contrasts with the much higher number of degrees of freedom on the human face. To generate the desired groups of expressions dynamically based on the Pleasure-Arousal-Dominance (PAD) dimensions [7], a mapping of its DOFs onto PAD dimensions was performed. We have taken a

[1] http://www.socoro.net.



Joint	Joint name
q1	Neck pitch (rotation)
q2	Head yaw (rotation)
q3	Head pitch (rotation)
q4	Lower disc (rotation)
q5	Upper disc (rotation)
q6	Left eyelid (rotation)
q7	Left eyebrow (rotation)
q8	Left eye (translation)
q9	Right eyelid (rotation)
q10	Right eyebrow (rotation)
q11	Right eye (translation)

Fig. 1. EMYS degree of freedom (source: http://doc.flashrobotics.com)

Table 1. EMYS DOFs-AUs-PAD mapping

DOF	AUs	PAD [13]	PAD [12]	PA [11]	PA [10]
q7 and q10 (eyebrows)	AU1 (inner brow raiser)	Negative pleasure, negative dominance	Low medium pleasure, low high arousal		Negative pleasure, low arousal
	AU2 (outer brow raiser)	High arousal, positive dominance	Positive pleasure		Positive pleasure, high arousal
	AU4 (brow lowerer)	Negative pleasure, positive dominance	Negative pleasure, high arousal	Low pleasure	Negative pleasure, med/high arousal
q4 (lower disc)	AU16 (lower lip depressor)	High arousal	Negative pleasure, high arousal		
	AU17 (chin raise)	Negative pleasure	Negative pleasure, high arousal	Low pleasure, low arousal	Negative pleasure, medium arousal

[a]See column heading references for more detail about PAD mappings

bottom-up approach by first describing EMYS's DOFs as comparable to single facial movements defined by the Facial Action Coding System (FACS) [5] since this approach is well-grounded in Psychology.

FACS, originally a descriptive, not a generative system, consists of measurements called Action Units (AUs) which are the contraction or relaxation of one or more muscle groups. There are around 46 AUs and combining them defines different facial expressions. Our aim however is not to focus on generating specific static emotional expressions but on dimensions of expressions. We reviewed the dimensional meaning of the relevant AUs [10–13] and found commonalities between them. An extract of our results are shown in Table 1. We then combined two AUs with similar valence value to form four approval (e.g. Head up - AU53, jaw drop - AU26) and four disapproval (e.g. Chin raise - AU17, head down - AU54) expressions. The expressions used in our study are shown in Fig. 3 below.

3 Experimental Design

The study was conducted at the Glasgow Science Centre over two days, with passing visitors invited to participate. The purpose was to have participants assess expressions designed to communicate the social signals of approval and disapproval as discussed. A basic issue with running a study in a public place like this is that it is entirely possible for participants to spectate before they take part and observe the reactions of other participants. This contagion risk was addressed by selecting a random sample of four expressions from the eight available in Fig. 3 for any given participant. The issues this raises for processing the results are discussed below.

Fig. 2. Study set-up.

Table 2. Participant data - total age range 4–77

	N	Male	Female	Mean age	Median age
Total sample	77	39	38	16.3	9
Age group					
4–8	32	20	12	5.91	6
9–13	17	8	9	10.12	10
14–24	8	3	5	19.88	20
25+	20	8	12	36.75	33

The task given to participants was to offer the EMYS head, given the gender-neutral name ‘Alyx’, items for its breakfast. These were plastic toy representations of fruit and other food items: six were laid out in front of the robot. While it is possible that food type may have influenced beliefs of the robot’s food preferences (e.g. that Alyx likes ice-cream), our focus was the users emotion recognition, not the affect of food type. After an item was offered, Alyx was made to produce one of the eight expressions as shown in Fig. 3 randomly by the

Approval		Disapproval	
	Head Up, Jaw Drop (HU-JD)		Inner Brow Raiser, Lower Lip Depressor (IBR-LLD)
	Outer Brow Raiser, Lips Part (OBR-LP)		Chin Raise, Head Down (CR-HD)
	Wink, Head Left (W-HL)		Brow Lowerer, Chin Raise (BL-CR)
	Upper Lid Raiser, Jaw Drop (ULR-JD)		Eyes Closed, Head Down (EC-HD)

Fig. 3. Approval and disapproval EMYS expressions according to PAD mapping

'Wizard' using a WoZ interface. Depending on whether they thought Alyx did or did not approve of the item, participants were instructed to place it in a box marked 'Like' or a box marked 'Dislike'. There was no relationship between the object presented and the expression generated. The experimental set-up can be seen in Fig. 2.

Alyx was operated in Wizard-of-Oz (WoZ) fashion, the wizard sitting close to the experiment and behind a screen visible in Fig. 2 top-right. The experiment was introduced by Alyx itself, using a female Scottish-accented unit-selection voice, with a pre-scripted statement. After each of the first three items Alyx would state that it was still hungry and would like another item; after the fourth item Alyx made a positive statement about having had an enjoyable breakfast.

Other than the approval/disapproval expressions, the head did not move; in particular it did not visually track the presented object. Generally, participants found the interaction easy. Table 2 gives their age and gender distribution. Gender was split almost evenly; child participants were the largest group followed by the 25+ group: their were few in the age range 14–24. This is not a surprise for a weekend in a Science Centre.

After interacting, each participant was asked to fill in a short questionnaire. The chosen objects and their order were logged manually with the expressions, contents of the boxes were logged using digital photos and the expressions and their order were also logged by the WoZ software.

4 Results

4.1 Participant Evaluations

As the expressions were generated in the WoZ style, the distribution of approval and disapproval types was uneven. Additionally, attempts to avoid social

Table 3. Facial expression trial and response summary

	n trials	% trials	Like	Dislike
Approval				
Head up, jaw drop	51	16.35	45	6
Outer brow raiser, lips part	35	11.22	16	19
Wink, head left	51	16.35	8	43
Upper lid raiser, jaw drop	47	15.06	46	1
Total	184	58.97	115	69
Disapproval				
Inner brow raiser, lower lip depressor	32	10.26	16	16
Chin raise, head down	32	10.26	5	27
Brow lowers, chin raise	32	10.26	9	23
Eyes closed, head down	32	10.26	6	26
Total	128	41.03	36	92
Overall total	312	–	151	161

contagion - as described above - meant participants experienced a varying number of approval and disapproval expressions. Table 3 demonstrates this bias, showing that 58.97% of trials contained approval expressions. Moreover, the expressions HU-JD (16.35%), W-HL (16.35%) and ULR-JD (15.06%) were the most frequently generated expressions. The implications of this imbalance are discussed later.

[2]To give a clearer representation of performance trends we investigated the relative difference between the number of Like and Dislike responses for each expression (see Fig. 4). Two expressions were clearly interpreted as approval: HU-JD and ULR-JD. Expressions W-HL, CR-HD, BL-CR and EC-HD were interpreted as disapproval. Notably, W-HL was consistently interpreted as disapproval if shown on the first trial. The lack of variation for OBR-LP and IBR-LLD suggests these two expressions were the most ambiguous.

All data analyses were completed using R Statistics. Firstly, stepwise regression modelling was used to identify the significant contributors to the model variance. As the dependent variable of interest was binary (Like, Dislike), binary logistic regressions were performed. Preliminary analyses comparing models using analysis of variance (ANOVA) yielded main effects of facial expression and order of presentation. As such, the final model contained these factors.

This final model found significant effects of the expressions HU-JD, $z = 5.10$, $p < 0.001$ and ULR-JD, $z = 4.34$, $p < 0.001$. There was near significant contribution of presentation order ($p = 0.06$). Conversion of the log-odds to estimated probabilities showed that HU-JD would be interpreted as approval 81% of the time and ULR-JD 96% of the time.

[2] One participant left the session without providing demographic information; Table 3 above includes a sample N = 78.

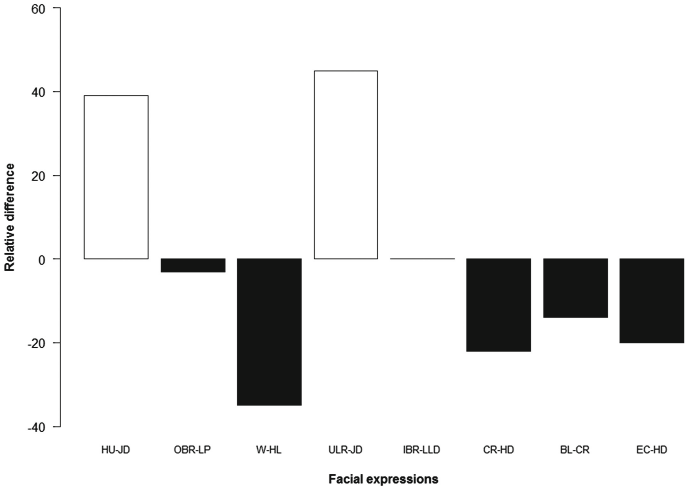


Fig. 4. Relative difference between approval and disapproval evaluations. White bars with positive values represent expressions interpreted as approval. The black bars with negative values represent disapproval interpretations.

4.2 Accuracy of Interpretation

The initial analysis of participant evaluations then informed a further analysis examining accuracy, based our EMYS mappings (see Fig. 3). The expression W-HL was predominantly interpreted as disapproval, contradicting our intended response pattern. So, W-HL data were omitted from this analysis. Imposing the PAD mapping on the evaluation data showed different levels of accuracy between the different age groups. The young group (4–8 years) provided the least correct interpretations (67.89%), followed by the Adult (25+) group (78.13%). The Middle (9–13 years) and Adoles/Adult (14–24 years) group performed similarly, 84.13% and 88% correct respectively.

The same analysis procedure was followed instead using the binary dependent variable of accuracy (Correct, Incorrect). Stepwise assessment of the fixed effects contribution to the model variance using ANOVA yielded two main effects: facial expression and age group. Table 4 shows the binary logistic regression results of this model of accuracy.

The binary logistic regression of participant accuracy indicated a similar trend to participant evaluations, with significant effects of the expressions HU-JD, $z = 2.22$, $p < 0.05$ and ULR-JD, $z = 2.83$, $p < 0.01$. Also, data from the young age group (ages 4–8 years) significantly contributed to the variance in the model, $z = -2.31$, $p < 0.05$. Conversion of the log-odds to estimated probabilities again showed a high success rate for HU-JD, interpreted correctly 96% of the time and ULR-JD 99% of the time. The young group were predicted to be correct 57% of the time.

Table 4. Modelling the predictors of accuracy

	β	SE β	z-value
(Intercept)	1.91	0.74	2.58**
CR-HD	0.93	0.64	1.45
EC-HD	0.78	0.63	1.25
HU-JD	1.34	0.60	2.22*
IBB-LLD	−0.90	0.55	−1.63
OBR-LP	−0.96	0.54	−1.77
ULR-JD	3.09	1.09	2.83**
Adult	−1.09	0.74	−1.48
Middle	−0.48	0.75	−0.64
Young	−1.62	0.70	−2.31*

Signif. codes: $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

5 Discussion

The objective of this pilot experiment was to investigate user recognition of eight (four approval and four disapproval) EMYS head facial expressions, for future work training employment and social skills to adults with an ASD. The data analysis demonstrated the utility of the EMYS head for this purpose and considerations for research in this area.

Most notably, users consistently perceived the expressions HU-JD and ULR-JD as approval. This suggests that expressions using a combination of head or upper disc and jaw AUs are processed favourably by users. A possible explanation lies in the overtness of these expressive compositions, making their social signals easier to process. In future, we plan to use these expressions to model positive workplace interactions for social skills training.

Some expressions however were viewed differently to our proposed mappings. For example, participants frequently judged W-HL as an expression of disapproval. Further investigation of this trend revealed an order effect, whereby the expression was consistently interpreted this way if shown on the first trial; explaining why the effect of presentation order almost reached significance. We therefore recommend that similar work include a familiarity period prior to experimentation. This could include non-experimental expressions that engage each of the AUs of interest to ensure participants are suitably primed for testing.

Another interesting feature of W-HL was it was temporally quicker (2 s) to all other expressions (set to 3 s). Anecdotal feedback from participants and experimenters identified issues with this duration: an *approval* W-HL would need to be significantly faster. Duration of expression is therefore important for future studies of this nature. It is also possible that the *head-left* movement was interpreted as a *head-shake* or no rather than a head-tilt. Thus, the amplitude as well as the speed of a movement should be examined.

Furthermore, it appears that expressions engaging the brow units (OBR-LP and IBR-LLD) produced a similar number of both approval and disapproval evaluations. This indicates that users found subtler AUs of the EMYS particularly difficult to disambiguate. An incremental examination of each AUs intensity with respect to user judgement would provide a useful index of EMYS expression interpretation.

We found age related effects on accuracy: children aged 4–8 years provided fewer correct expression judgements relative to the other age groups. This finding supports previous work with the EMYS head indicating that older participants are superior at decoding EMYS facial expressions [14]. As we aim to develop a system for adults this finding was not of primary interest, but provides impetus for those studying developmental mediators of emotion recognition.

Study limitations were born out of the setting. One major issue was the noise level, a feature that is difficult to control outside of the laboratory. Concerns over social contagion led to an imbalance in expression presentation. Logistic regression was chosen reduce some of this bias. A future iteration of the experiment will include a counterbalanced, equal number of expressions per participants. Synchronisation issues were also evident as participants attempted to capture the gaze of the robot with the food items. As such, motion tracking hardware and an RFID reader will be augmented with our FLASH platform create a more autonomous system.

6 Conclusion and Future Work

This work was intended to inform future robot design for work with adults with an ASD, as part of the SoCoRo project. The analysis show that two of the eight expressions tested were judged unambiguously by participants. These expressions will therefore be used in our future work, examining the effect of expressive behaviour on the users social signal processing. For example, we intend to create a set of simulated office scenarios where a robot 'boss' interacts with adults with ASD and provides task feedback.

Acknowledgements. This work was supported by the Research Councils UK, project EP/N034546/1. We thank the staff of the Glasgow Science Centre who provided invaluable assistance and materials to ensure successful data collection over two days.

References

1. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: survey of an emerging domain. Image Vis. Comput. **27**(12), 1743–1759 (2009)
2. Hillier, A., Campbell, H., Mastriani, K., Izzo, M.V., Kool-Tucker, A.K., Cherry, L., Beversdorf, D.Q.: Two-year evaluation of a vocational support program for adults on the autism spectrum. Career Dev. Transit. Except. Individ. **30**(1), 35–47 (2007)
3. Dib, N., Sturmey, P.: Behavioral skills training and skill learning. In: Seel, N.M. (ed.) Encyclopedia of the Sciences of Learning, pp. 437–438. Springer, Boston (2012). doi:10.1007/978-1-4419-1428-6_644

4. Stewart, K.K., Carr, J.E., LeBlanc, L.A.: Evaluation of family-implemented behavioral skills training for teaching social skills to a child with Asperger's disorder. Clin. Case Stud. **6**(3), 252–262 (2007)
5. Ekman, P., Friesen, W., Ellsworth, P.: What emotion categories or dimensions can observers judge from facial behavior? In: Ekman, P. (ed.) Emotion in the Human Face, pp. 39–55. Cambridge University Press, New York (1982)
6. Jack, R., Garrod, O., Schyns, P.: Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. Curr. Biol. **24**(2), 187–192 (2014)
7. Mehrabian, A.: Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. Curr. Psychol. **14**(4), 261–292 (1996)
8. Wingenbach, T.S.H., Ashwin, C., Brosnan, M.: Validation of the Amsterdam dynamic facial expression set - bath intensity variations (ADFES-BIV): a set of videos expressing low, intermediate, and high intensity emotions. PLoS ONE **11**(1), e0147112 (2016). doi:10.1371/journal.pone.0147112
9. Plaisted, K.C.: Reduced generalization in autism: an alternative to weak central coherence. In: Burack, J.A., Charman, T., Yirmiya, N., Zelazo, P.R. (eds.) The Development of Autism: Perspectives from Theory and Research, pp. 149–172. LEA (2001)
10. Snodgrass, J.: Judgment of feeling states from facial behavior: a bottom-up approach. Unpublished doctoral dissertation, University of British Columbia (1992)
11. Hadar, D.: An interrelating model between action units and emotional descriptives: pleasure, arousal and dominance. A lab research project, Hebrew University, Jerusalem (2015)
12. Grammer, K., Oberzaucher, E.: The reconstruction of facial expressions in embodied systems. ZiF: mitteilungen **2**, 14–31 (2006)
13. Boukricha, H., Wachsmuth, I., Hofstätter, A., Grammer, K.: Pleasure-arousal-dominance driven facial expression simulation. In: Mühl, C., Heylen, D., Nijholt, A. (eds.) ACII 2009, pp. 119–125. IEEE (2009)
14. Kędzierski, J., Muszyński, R., Zoll, C., Oleksy, A., Frontkiewicz, M.: EMYS–emotive head of a social robot. Int. J. Soc. Robot. **5**(2), 237–249 (2013)