Supplemental Information (SI) 4 for

Reading **your** emotions in **my** physiology? Reliable emotion interpretations in absence of a robust physiological resonance

Julia Folz[1,2], Donatella Fiacchino[1], Milica Nikolić[1,2,3], Henk van Steenbergen[1,2], Mariska E. Kret[1,2]

[1]Department of Cognitive Psychology, Institute of Psychology, Leiden University, 2333 AK Leiden, The Netherlands

[2]Leiden Institute for Brain and Cognition (LIBC), Leiden University, 2300 RC Leiden, The Netherlands

[3] Research Institute of Child Development and Education, University of Amsterdam, 1018 WS Amsterdam, The Netherlands

*Correspondence: m.e.kret@fsw.leidenuniv.nl

Analysis 3 (Classification analysis)

For creating a model to predict the emotion interpretation of perceived emotion expressions based on physiological signal patterns, we considered data from four physiological channels (two EMG channels, SC, SKT) taking into account the exclusion criteria for each channel as specified earlier (6283 trials). The pupil data was excluded from this analysis due to potential bias from stimulus characteristics. As emotion perception was observed to co-occur with physiological signals on different time-scales and in different ways in analysis 1 (i.e. variation in onsets and shapes), trial-by-trial summary measures were calculated. More specifically, since Analysis 1 showed that the earliest stable effect of emotion on EMG time courses can be observed 500ms after stimulus onset and that differences between EMG time courses can last up to 3.9s after stimulus onset, we chose this broad time window and created the means for the Corrugator signal and the Zygomaticus signal as summary measures. Further, for skin conductance and skin temperature, we created regression models with polynomials as predictors in order to get an estimate for the relative presence of each shape in the signal (again using the nlme package; Pinheiro et al., 2020). Importantly, for each measure, we chose the same time windows as well as polynomials of the same order as in Analysis 1 and we also included an autoregressive term. The resulting coefficients of the predictors as well as the intercept were taken as summary measures (3947 trials had data from all measures). Thus, we had one summary measure for each EMG channel (mean) and 4 summary measures each for skin conductance and skin temperature (intercept, linear trend, quadratic trend and cubic trend).

After creating the summary measures, we focused on trials in which prototypical facial expressions were shown since only expressions from this modality seemed to be related to meaningful (de-)activations in all signals (1515 trials). Further, we only included trials in which the participants accurately labelled the emotion displayed by the face (i.e. presented emotion = perceived emotion, 1321 trials). In the next step, the dataset was split into a training and a test dataset (similar to the EMG analysis) in order to test whether emotion labels in the test sample (N = 27, 667 trials) could be successfully predicted with the association built between physiological activity and emotion labels in the training sample (N = 26, 654 trials). To achieve this, the summary measures were used as predictors in a multinomial logistic regression, using the mlogit-package in R (Croissant, 2020), in order to train the association between the physiological signals and the five facial expressions. We then used the resulting model fit to predict the labels in the test sample. In the third step, the predicted label was compared to the pre-assigned stimulus label to evaluate model accuracy. The overall percentage of successful classification was lastly compared to a distribution of randomly generated labels (10000 samples) in order to determine whether the classifier performed significantly better than chance, i.e. whether the predicted accuracy was higher than the upper bound of the 95% confidence interval of the random-label distribution. Further, to investigate whether the performance of the classifier differed by emotion category, we compared the presence of the corresponding emotion label (0,1) for each emotion category separately in the actual presentation versus the prediction using signal detection theory. More specifically, we looked at the ratio of hits to false alarms in labelling an expression as e.g. “happy” by calculating the area under the curve (AUC) and its standard error.

To explore the generalizability of the classifier, we tested it on two additional data subsets: Using data from trials in which a wrong emotion label was given (194 trials), we wanted to explore the possibility that physiological signals might be more strongly related to the emotion category that a person subjectively perceives rather than the emotion category that was intended to be displayed by the stimulus/actor. Thus, we took the model fit created in the previous analysis to predict emotion labels based on the physiological patterns. The predicted emotion label was then compared to both the stimulus emotion label and the subjectively perceived emotion label and the concordance of both label pairings was again compared to a distribution of random labels (10000 samples). Lastly, even though there was no option to accurately label the neutral facial expressions with the three subtle emotional cues (blush, tears and dilated pupils), participants still indicated the emotion category which they associated the expressions with. We wanted to use this information to explore whether subtle facial cues could already elicit physiological patterns which are similar to patterns of associated full-blown facial expressions of emotion. Thus, we used the model fit again as classifier for physiological patterns during viewing of subtle emotional cues (1204 trials) and compared the resulting labels with the labels given by the participants. Once again, we compared the match between the labels to a distribution of 10000 random samples.

Reference

Croissant, Y. (2020). Estimation of Random Utility Models in R : The mlogit Package . *Journal of Statistical Software, 95*(11). https://doi.org/10.18637/jss.v095.i11

Table 1. Results of the multinomial logistic regression model describing the association between emotion category labels and summary measures of different physiological change parameters recorded during passive viewing of emotional facial expressions (neutral as reference category).

Predictors	*Odds Ratios*	CI	*t*	*p*
Intercept				
▪ angry	1.236	0.953 – 1.603	1.595	0.111
▪ happy	1.227	0.944 – 1.594	1.528	0.126
▪ sad	1.080	0.826 – 1.412	0.561	0.575
▪ fearful	1.194	0.919 – 1.550	1.329	0.184
Cor: mean activity				
▪ angry	0.371	0.196 – 0.699	-3.063	**0.002**
▪ happy	0.366	0.194 – 0.689	-3.111	**0.002**
▪ sad	0.495	0.255 – 0.964	-2.068	**0.039**
▪ fearful	0.546	0.288 – 1.035	-1.855	0.064
Zyg: mean activity				
▪ angry	1.638	0.756 – 3.549	1.250	0.211
▪ happy	1.720	0.775 – 3.820	1.332	0.183
▪ sad	1.170	0.563 – 2.431	0.420	0.675
▪ fearful	1.554	0.750 – 3.219	1.186	0.236
SCL: Intercept				
▪ angry	0.322	0.038 – 2.721	-1.041	0.298
▪ happy	0.136	0.012 – 1.523	-1.619	0.105
▪ sad	0.260	0.021 – 3.158	-1.057	0.290
▪ fearful	0.329	0.038 – 2.819	-1.014	0.311
SCL: linear slope				
▪ angry	2.634	0.267 – 25.965	0.829	0.407
▪ happy	9.832	0.791 – 122.142	1.778	0.075
▪ sad	4.414	0.317 – 61.370	1.106	0.269
▪ fearful	5.116	0.506 – 51.765	1.382	0.167
SCL: quadratic slope				
▪ angry	0.349	0.083 – 1.479	-1.428	0.153
▪ happy	0.185	0.036 – 0.953	-2.017	**0.044**
▪ sad	0.207	0.036 – 1.191	-1.764	0.078

▪ fearful	0.396	0.094 – 1.659	-1.268	0.205
SCL: cubic slope				
▪ angry	3.667	0.769 – 17.474	1.631	0.103
▪ happy	7.977	1.633 – 38.977	2.566	**0.010**
▪ sad	3.747	0.720 – 19.511	1.569	0.117
▪ fearful	3.926	0.886 – 17.391	1.801	0.072
SKT: Intercept				
▪ angry	2.382	0.155 – 36.569	0.623	0.533
▪ happy	0.101	0.007 – 1.520	-1.658	0.097
▪ sad	1.674	0.104 – 27.030	0.363	0.717
▪ fearful	0.751	0.049 – 11.559	-0.205	0.837
SKT: linear slope				
▪ angry	0.550	0.037 – 8.293	-0.432	0.666
▪ happy	13.337	0.908 – 195.868	1.890	0.059
▪ sad	0.716	0.045 – 11.328	-0.237	0.813
▪ fearful	1.617	0.107 – 24.386	0.347	0.729
SKT: quadratic slope				
▪ angry	1.871	0.438 – 7.997	0.845	0.398
▪ happy	0.361	0.086 – 1.517	-1.391	0.164
▪ sad	1.472	0.336 – 6.444	0.513	0.608
▪ fearful	0.880	0.207 – 3.743	-0.173	0.863
SKT: cubic slope				
▪ angry	0.991	0.317 – 3.102	-0.015	0.988
▪ happy	3.332	1.063 – 10.444	2.065	**0.039**
▪ sad	0.888	0.278 – 2.835	-0.200	0.841
▪ fearful	1.407	0.447 – 4.431	0.583	0.560
Observations				3270
R^2 McFadden				0.022

Note. Cor = Corrugator supercilii, Zyg = Zygomaticus major, SC = Skin conductance level, SKT = skin temperature. Bold font highlights p-values below the significance level of 0.05.

Results. To build the emotion label classifier, we calculated a multinomial logistic regression on the emotion labels using physiological summary measures of accurately recognized trials in the training sample. The resulting model fit suggested that the contribution of different summary measures varied widely (see Table 1). However, we were not mainly interested in which summary measure was most predictive in the data set at hand, but rather, whether the integrated pattern could be used to make a prediction for new observations. Thus, we did not further evaluate the model fit but directly moved on to applying the fit as a classifier to a new independent test dataset and to datasets of the same subjects but with their inaccurate trials and subtle cues trials, respectively. The comparison between predicted and observed emotion labels in the test sample revealed that, overall, the accuracy of the classifier was not higher than the accuracy in the 95th percentile of random labels distribution ($Acc_{\text{Classifier}} = 0.222$ versus $Acc_{\text{Random 95\%}} = 0.231$, see Table 2). Being in the 92nd percentile of the random distribution, the classifier performance was not significantly better than chance. When looking at accuracy in the receiver operating characteristic analysis (ROC) contrasting the occurrence of predicted and observed labels per emotion, this overall finding was confirmed. Even though there were differences in the Area Under the Curve (AUC) between emotions, the confidence interval consistently included chance level (see Fig. 1). The classifier performance for the subtle emotional cue set and the inaccurate trials, with regard to both stimulus and perceived labels, was even worse (see table 2). Given that in all three data sets more happy and angry expressions were predicted than presented/observed, this classifier bias might have been specifically detrimental for sets with little cases of these emotions due to their easy recognition or strong intensity.

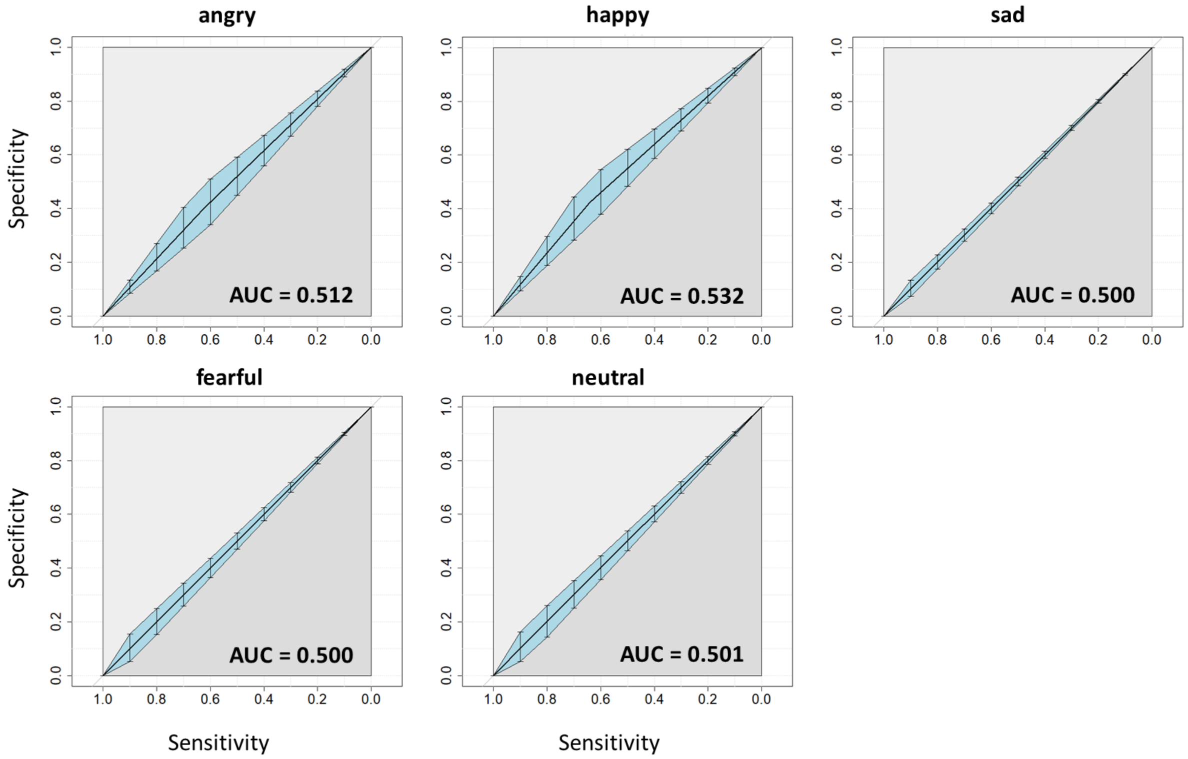

Fig. 1 Receiver operating characteristic curves for the classification of the 5 different expression categories. The blue shaded areas and black bars indicate the 95% confidence interval of sensitivities at different specificities. The dark grey area represents the area under the curve, with the exact value in the right button corner of each plot

Table 2. Results of the classifier performance evaluation in the three new datasets (test set, inaccurate trials set and subtle cues set) as defined by comparison with a random sample distribution (overall) and the Area Under the Curve (AUC; per expression category).

	Observations	Predictions	% Overlap	Percentile (rand. distr.)	95% CI of random distribution	AUC	95% CI of AUC
Test sample							
Overall	667	667	0.222	0.922	0.169 –0.231		
Angry	140	254	0.4			0.512	0.466 – 0.558
Happy	144	249	0.424			0.532	0.487 – 0.577
Sad	131	20	0.031			0.500	0.484 – 0.517
Fearful	138	63	0.094			0.500	0.472 – 0.527
Neutral	114	81	0.123			0.501	0.468 – 0.534
Inaccurate trials							
Stimulus label (overall)	194	194	0.155	0.065	0.144 –0.258		
Angry	21	74	0.381			0.5149	0.400 –0.629
Happy	8	55	0.25			0.5914	0.403 –0.780
Sad	56	11	0.054			0.4819	0.466 –0.498
Fearful	32	12	0.063			0.5318	0.464 –0.600
Neutral	77	42	0.195			0.5203	0.468 –0.572
Perceived label (overall)	194	194	0.165	0.129	0.144 – 0.258		
Angry	41	74	0.463			0.508	0.422 –0.594
Happy	10	55	0.2			0.3288	0.294 –0.363
Sad	57	11	0.035			0.4942	0.472 –0.516
Fearful	48	12	0.063			0.4869	0.440 –0.534
Neutral	38	42	0.158			0.4921	0.431 –0.553
Subtle cues							
Perceived label (overall)	1204	1204	0.159	0	0.178 – 0.223		
Angry	63	454	0.476			0.5523	0.487 –0.616
Happy	12	368	0.333			0.514	0.374 –0.654
Sad	339	57	0.071			0.5163	0.501 –0.531
Fearful	104	111	0.135			0.5232	0.489 –0.557
Neutral	686	214	0.173			0.495	0.473 –0.517