Supplementary Material: Exploring Anticipatory Human Reactions for Outcome Prediction in HRI

MARIA TERESA PARREIRA, SUKRUTH GOWDRU LINGARAJU, ADOLFO RAMIREZ-ARISTIZABAL, ALEXANDRA BREMERS, MANASWI SAHA, MICHAEL KUNIAVSKY, and WENDY JU

1 FEATURES USED FOR TRAINING

Below, we list the features included in the dataset. The facial activation features were extracted using OpenFace [1]. While this tool provides a very large set of features, we reduced the feature space by only including the most relevant markers. We obtained a total of 35 per participant facial activation features:

```
vis_features = [
'AU01_r', 'AU02_r', 'AU04_r', 'AU05_r', 'AU06_r', 'AU07_r', 'AU09_r', 'AU10_r', 'AU12_r',
'AU14_r', 'AU15_r', 'AU17_r', 'AU20_r', 'AU23_r', 'AU25_r', 'AU26_r',  'AU45_r', 'AU01_c',
'AU02_c', 'AU04_c', 'AU05_c', 'AU06_c', 'AU07_c', 'AU09_c', 'AU10_c', 'AU12_c', 'AU14_c',
'AU15_c', 'AU17_c', 'AU20_c', 'AU23_c', 'AU25_c', 'AU26_c', 'AU28_c', 'AU45_c'
]
```

2 SHORT-LISTED DATASET DEMOGRAPHICS

Our short-listed dataset is comprised of data from 14 participants. Ages range from 20-37 (27.85 ± 5.01). 5 participants identified as female, 8 as male and 1 who self-identified. Racial/ethnical distribution includes 8 Caucasian/White or Asian/White, 4 African/African American/Black, 1 Hispanic/Latino and 1 Asian/Asian American participants. Participants took an average of $30m42s \pm 6m30s$ to complete the full data collection process.

3 MODEL TRAINING - HYPERPARAMETER TUNING

On Table 1, we leave details of the hyperparameter tested when implementing our models. The final set of hyperparameters was selected based on the best-performing model on the test set, according to the accuracy.

4 STATISTICAL ANALYSIS

In Table 2 we show the p-values for all the AUs tested.

REFERENCES

[1] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. 59–66. https://doi.org/10.1109/FG.2018.00019

Table 1. Hyperparameters that were used for hyperparameter tuning. More details on the main manuscript.

Model	HyperParamaters tested
LSTM	lookback = [5, 10, 15] units = [32, 64, 128] dropouts = [0.0, 0.2, 0.4, 0.6] activations = [‘sigmoid’, ‘relu’, ‘tanh’, ‘softmax’] optimizers = [‘SGD’, ‘Adam’, ’Adadelta’] batch_sizes = [512, 1024]
GRU	lookback = [5, 10, 15] units = [32, 64, 128] dropouts = [0.0, 0.2, 0.4, 0.6] activations = [‘sigmoid’, ‘relu’, ‘tanh’, ‘softmax’] optimizers = [‘SGD’, ‘Adam’, ’Adadelta’] batch_sizes = [512, 1024]
BiLSTM	lookback = [5, 10, 15] units = [32, 64, 128] dropouts = [0.0, 0.2, 0.4, 0.6] activations = [‘sigmoid’, ‘relu’, ‘tanh’, ‘softmax’] optimizers = [‘SGD’, ‘Adam’, ’Adadelta’] batch_sizes = [512, 1024]

Table 2. Facial Unit Activation features tested. AU – action unit. Grey rows indicate non-significant (p >0.05) features.

Feature	AU	p-value
AU01_r	Inner Brow Raiser	$7.34e-05$
AU02_r	Outer Brow Raiser	7.82
AU04_r	Brow Lowerer	$4.87e-26$
AU05_r	Upper Lid Raiser	16.07
AU06_r	Cheek Raiser	$1.04e-82$
AU07_r	Lid Tightener	3.37
AU09_r	Nose Wrinkler	0.0002
AU10_r	Upper Lip Raiser	$5.82e-46$
AU12_r	Lip Corner Puller	$1.22e-213$
AU14_r	Dimpler	$7.63e-80$
AU15_r	Lip Corner Depressor	14.24
AU17_r	Chin Raiser	$2.82e-38$
AU20_r	Lip Stretcher	3.20
AU23_r	Lip Tightener	$3.64e-05$
AU25_r	Lips Part	$8.05e-27$
AU26_r	Jaw Drop	$8.17e-27$
AU45_r	Blink	0.011