Automated recognition of complex categorical emotions from facial expressions and head motions

Andra Adams
Computer Laboratory
University of Cambridge
Cambridge, UK
Email: Andra.Adams@cl.cam.ac.uk

Peter Robinson
Computer Laboratory
University of Cambridge
Cambridge, UK
Email: Peter.Robinson@cl.cam.ac.uk

***Abstract*—Classifying complex categorical emotions has been a relatively unexplored area of affective computing. We present a classifier trained to recognize 18 complex emotion categories. A leave-one-out training approach was used on 181 acted videos from the EU-Emotion Stimulus Set. Performance scores for the 18-choice classification problem were AROC = 0.84, 2AFC = 0.84, F1 = 0.33, Accuracy = 0.47. On a simplified 6-choice classification problem, the classifier had an accuracy of 0.64 compared with the validated human accuracy of 0.74. The classifier has been integrated into an expression training interface which gives meaningful feedback to humans on their portrayal of complex emotions through face and head movements. This work has applications as an intervention for Autism Spectrum Conditions.**

***Index Terms*—affective computing; emotion recognition**

I. Introduction

Facial expressions [1] and head motions [2] are important modalities that humans use to communicate their mental states to others. Computers that can detect and recognize emotional displays through these modalities have many interesting applications in society [3].

This paper presents a classifier which recognizes complex affective states based on facial expressions and head motions. The simplicity of the training features is then leveraged to provide feedback to humans for facial expression training.

A. Related Work

Automated emotion classification systems already exist in many forms (see [1] and [3] for surveys). However the majority of these systems focus only on the so-called basic emotions [4] (fear, disgust, surprise, joy, sadness, anger) for training [5]–[8]. Only a few papers have explored more complex taxonomies [9] [10], yet complex emotions have been found to occur more frequently in everyday human interaction than basic emotions [11].

An important consideration in automated emotion classification is the choice of features to extract. Previous research on dynamic facial expression analysis extracts features that are not inherently meaningful to a typical user (e.g. Gabor wavelets [12], dynamic textures [13], optical flow [14]). The result is that differences between dynamic facial expressions are impossible to communicate back to the user without additional processing. (To the best of our knowledge, this additional processing has also not yet been explored.)

Automated emotion classifiers must also choose between the two most commonly used models of affect: categorical and dimensional [15]. Categorical models attempt to classify each affective display into a discrete category, based on the primary-process theory that a small number of emotions are hard-wired in our brain. Dimensional models relate affective states to one another in a systematic manner, namely by plotting each affective state onto one or more chosen dimensions. The advantages and disadvantages of each approach are surveyed in [15] and [3]. Until the neurology behind emotion processing in humans is agreed upon [16], both categorical and dimensional approaches seem to have merit and should be explored equally.

B. Current Work

We present an automated classifier for complex, categorical emotions. The emotion categories considered were rated by ASC clinical experts (n=47) and parents of children with ASC (n=88) as being the most important for social interaction [17]. Furthermore, previous studies have found them to be discretely identifiable through facial expression [18] [19]. We have trained our classifier on simple extracted features (see Section II-B) which later are used to provide meaningful feedback to humans for expression training (see Section V).

II. Methodology

This section describes the three stages of our emotion classification: video analysis, feature extraction, and classification. We also describe the dataset selected for training, called the EU-Emotion Stimulus Set (EESS).

A. Video Analysis

The Cambridge face tracker [20] was used to extract head pose values (angles of rotation for yaw, pitch and roll) and facial Action Unit (AU) intensities (continuous values between 0 and 5 for each facial muscle movement) from each video in the EESS [21]. Additionally, an algorithm for finding the eye center location [22] was used to roughly estimate the horizontal eye gaze direction.

The 17 action units from the Facial Action Coding System [23] used in our classifier are:

978-1-4799-9953-8/15/$31.00 ©2015 IEEE

- AU 2 (outer brow raiser)
- AU 4 (brow lowerer)
- AU 5 (upper lid raiser)
- AU 6 (cheek raiser)
- AU 9 (nose wrinkler)
- AU 12 (lip corner puller)
- AU 17 (chin raiser)
- AU 25 (lips part)
- AU 26 (jaw drop)
- AU 51 (head turn left)
- AU 52 (head turn right)
- AU 53 (head up)
- AU 54 (head down)
- AU 55 (head tilt left)
- AU 56 (head tilt right)
- AU 61 (eyes turn left)
- AU 62 (eyes turn right)

B. Feature Extraction

Each video frame was analyzed to detemine the **intensity** and the **change in intensity** relative to the previous frame (i.e. speed of movement) of each of the 17 AUs. The range of continuous output values for AU intensity was evenly divided into 10 buckets, and the range of continuous output values for AU speed was evenly divided into 6 buckets.

Statistics were then aggregated on a per-video basis by counting the total number of frames which belonged to each of the buckets for each AU. This created a single 272-dimensional feature vector per video. Each feature vector was then normalized by dividing by the number of frames in the video.

An **overall average** feature vector was created by aggregating the non-normalized feature vectors for all videos in the dataset, and then normalizing by the total number of frames in the entire dataset.

We can see how a given video differs from the average by subtracting the normalized overall average feature vector from the given video's feature vector. To better visualize the information, we can ignore values below zero as the frames missing from these below-average buckets must show up as extra frames in other above-average buckets. Thus we can look only at above-average buckets without losing any information.

Two examples of feature vectors are shown in Figure 1. These feature vectors have already had the average subtracted, and only the above-average buckets are presented visually. The two example videos are in contrasting emotion categories, *Joking* and *Ashamed*. The *Joking* video tends to have higher intensities in the Lip corner puller (smile), Lips part, Cheek raiser, Jaw drop, Outer brow raiser, and Upper lid raiser (eyes widen) action units, while the *Ashamed* video has higher intensities in the Brow lowerer, Head down, and Nose wrinkler action units. The speed buckets for these two feature vectors also show contrast: the *Joking* video tends to have higher speeds in nearly all of its AUs than the *Ashamed* video.

C. Classification

An **average feature vector for each emotion category** was created by aggregating the feature vectors of the videos belonging to that emotion category, and then normalizing by the total number of frames in that category.

Classification was then determined using the Euclidean distance between a video's profile and each of the emotion category profiles, with the smallest distance indicating the emotion category to select as the classification result.

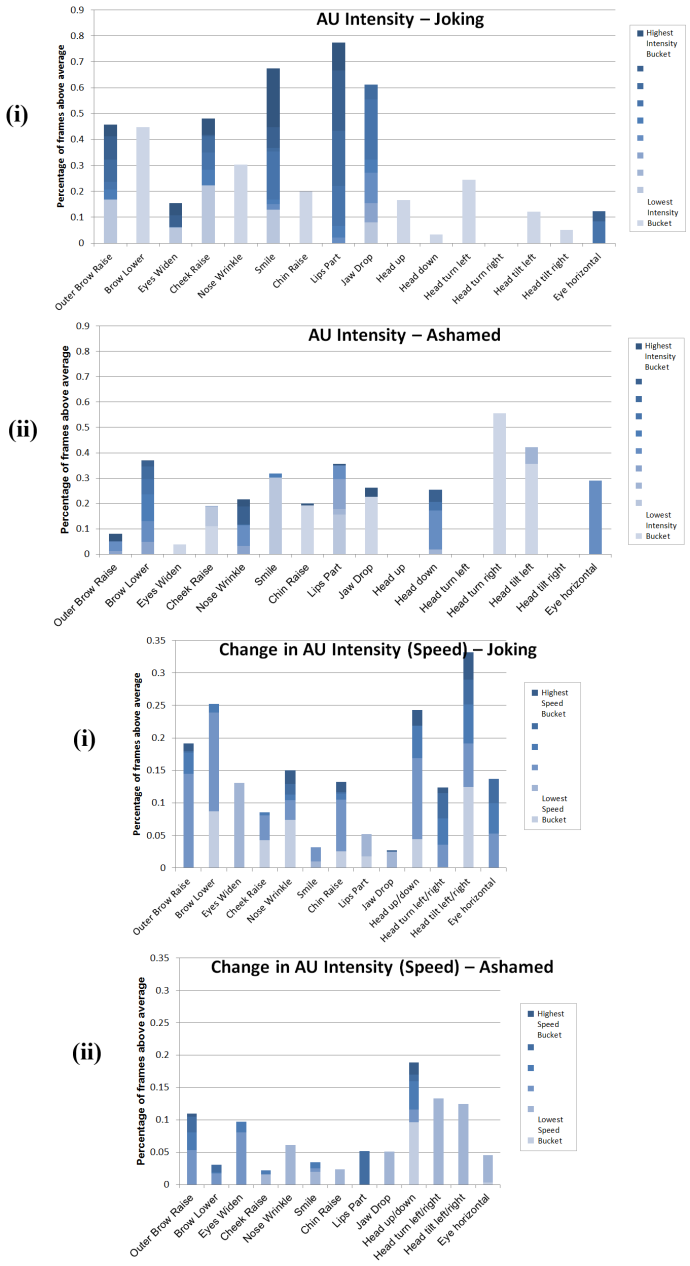


Fig. 1. Feature vectors with subtracted average for (i) a video in the emotion category *Joking*, and (ii) a video in the emotion category *Ashamed*. Only the above-average buckets are shown. The intensity and speed buckets are split into separate graphs.

D. Dataset

The dataset of videos used to train the classifier is a subset of the EU-Emotion Stimulus Set (EESS), recently collected and validated as part of the European ASC-Inclusion project [17]. Summary details of the EESS are found in Table I.

The EESS contains 247 video clips of acted non-verbal facial expressions and head motions. Each video is 2–14 seconds in length and is labelled with one of 20 emotion categories or *Neutral*. Crowd-sourced labels were gathered by O'Reilly et al. [17] using a six-option forced-choice format, and each video

TABLE I
EMOTION CATEGORIES IN THE EU-EMOTION STIMULUS SET

Emotion category	Raters	Total Videos	Accepted Videos
Afraid	2,113	17	17
Angry	1,997	17	10
Ashamed	829	8	5
Bored	895	8	7
Disappointed	1,260	10	6
Disgusted	2,025	18	14
Excited	985	9	9
Frustrated	2,017	12	11
Happy	1,498	14	11
Hurt	1,106	10	8
Interested	1,360	11	8
Jealous	774	7	0
Joking	1,083	9	9
Kind	969	9	0
Neutral	1,927	17	17
Proud	1,348	11	7
Sad	1,506	14	13
Sneaky	1,221	11	8
Surprised	2,249	18	16
Unfriendly	1,156	9	0
Worried	759	8	5
Total	29,077	247	181

TABLE II
FOILS PER EMOTION

Target	Foil #1	Foil #2	Foil #3	Foil #4
Afraid	Ashamed	Unfriendly	Disappointed	Kind
Angry	Jealous	Disgusted	Surprised	Happy
Ashamed	Disappointed	Worried	Unfriendly	Proud
Bored	Frustrated	Sad	Hurt	Excited
Disappointed	Worried	Bored	Afraid	Joking
Disgusted	Afraid	Frustrated	Sad	Interested
Excited	Interested	Joking	Hurt	Bored
Frustrated	Sad	Jealous	Sneaky	Kind
Happy	Interested	Surprised	Bored	Angry
Hurt	Worried	Unfriendly	Surprised	Happy
Interested	Excited	Proud	Joking	Disappointed
Jealous	Disappointed	Disgusted	Interested	Kind
Joking	Kind	Interested	Proud	Angry
Kind	Interested	Proud	Excitement	Frustrated
Neutral	Bored	Kind	Surprised	Frustrated
Proud	Excited	Interested	Kind	Afraid
Sad	Afraid	Jealous	Disgusted	Proud
Sneaky	Angry	Disappointed	Ashamed	Kind
Surprised	Happy	Joking	Worried	Bored
Unfriendly	Frustrated	Hurt	Bored	Surprised
Worried	Angry	Disappointed	Disgusted	Happy

was labelled by 57–580 people. The six choices given for each video were: one target emotion that the actor was intending to express, four foils (control emotions selected per-emotion based on similarity scores between emotion categories [17]), and "None of the above" to prevent artifactual agreement. Table II shows the specific foils used for each target emotion category.

Based on the crowd-sourced labelling results, only 181 of the 247 videos met the validation requirement and were accepted as reasonable portrayals of their labelled emotions by O'Reilly et al. This excluded all videos from the emotion categories *Jealous*, *Kind*, and *Unfriendly*.

Table III, modified from [5], gives a comparison of the EESS against several other databases that have been used for various emotion recognition challenges.

III. RESULTS

The classifier was tested using a leave-one-out approach on the 181 accepted videos from the EESS.

There are a number of ways to report the performance of a classifier (see [1] for an overview). In this study, we have considered AROC (Area under the Receiver Operating Characteristics curve), 2AFC (Two-Alternative Forced Choice), F1-score, and overall accuracy. Results are reported in Table IV.

AROC is calculated by computing the True Positive Rate (TPR) and False Positive Rate (FPR) for the system at a range of thresholds. The points are then interpolated into a curve and the area beneath is calculated through integration. When TPR=FPR (i.e. there is an equal chance of accepting a negative sample or a positive sample, essentially chance), the AROC is 0.5. A perfect classifier has TPR=1 and FPR=0 at all thresholds, giving an AROC of 1. **The AROC for our classifier was 0.84**. The ROC curve is shown in Figure 2.

2AFC is calculated by asking the classifier to make a forced-choice between each possible pair of samples where one of the samples is positive and the other is negative. For each pair, a score of 1 is given if the classifier correctly selects the positive sample, a score of 0.5 is given if the two samples were equally likely, and a score of 0 is given if the negative sample is chosen. A 2AFC score of 0.5 means that the classifier is performing at chance. **The 2AFC score for our classifier was 0.84**. The confusion matrix for all 2AFC pairs is in Figure 3.

F1-score is well-known for measuring classification performance and is the harmonic mean of *precision* and *recall*. **The F1-score for our classifier was 0.33**. The F1-score for a classifier that randomly selects emotion labels according to the proportion of videos per emotion category would be 0.063.

Overall accuracy was 0.47 on the 18-choice classification problem, and 0.64 on the 6-choice classification problem.

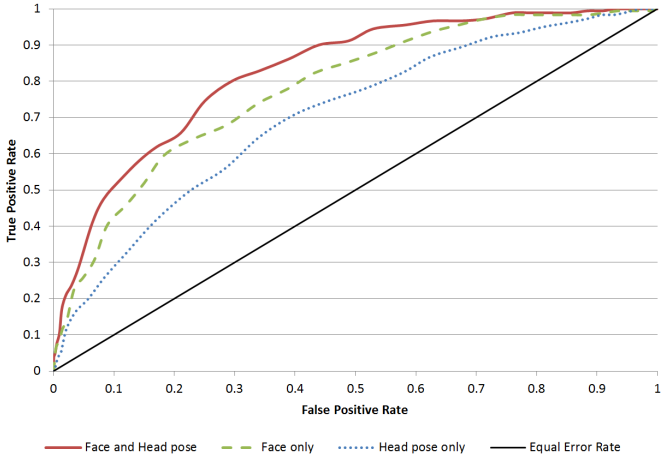


Fig. 2. Receiver Operating Characteristics (ROC) curve for the complex emotion classifier. The Area under the ROC curve (AROC) is 0.84.

TABLE III
COMPARISON OF EESS WITH DATABASES USED IN PREVIOUS EMOTION RECOGNITION CHALLENGES

Database	Challenge	Natural?	Labels	Environment	Subjects Per Sample	Construction Process
EESS [17]	–	Posed	Discrete – Afraid, Angry, Ashamed, Bored, Disappointed, Disgusted, Excited, Frustrated, Happy, Hurt, Interested, Jealous, Joking, Kind, Neutral, Proud, Sad, Sneaky, Surprised, Unfriendly, Worried	Lab	Single	Manual
AFEW [5]	EmotiW	Spontaneous (Partial)	Discrete – Anger, Disgust, Fear, Happiness, Neutral, Sadness, Surprise	Wild	Single and Multiple	Semi-Automatic
Cohn-Kanade+ [7]	–	Posed	Discrete – Angry, Disgust, Fear, Happy, Sadness, Surprise, Contempt	Lab	Single	Manual
GEMEP-FERA [6]	FERA	Spontaneous	Discrete – Anger, Fear, Joy, Relief, Sadness	Lab	Single	Manual
MMI [8]	–	Posed	Discrete – Angry, Disgust, Fear, Happy, Sadness, Surprise	Lab	Single	Manual
SEMAINE [10]	AVEC	Spontaneous	Continuous – Valence, Arousal, Power, Expectation, Intensity, 6 basic emotions, and others	Lab	Single	Manual

More details on the 6-choice classification problem can be found in the next section (Section III-A).

A. Comparison to Human Classification

The crowd-sourced labelling data from the EESS allows us to compare our classifier's accuracy with the accuracy of human raters. In the original study, human raters were given six options for each forced choice response: the target emotion, four foils, and "None of the above". Details of the crowd-sourcing have been given in Section II-D.

We recreated the six-option forced choice experiment for the machine classifier using the leave-one-out approach with each of the 181 videos. Results are presented per emotion category in Table IV.

Calculating the classification distance to the target emotion and to each of the 4 foil emotions is self-explanatory. The "None of the above" option is a special case. We represented "None of the above" by a new category which was the mean feature vector of all of the videos that did not belong to the target emotion nor to any of the four foil emotions. For example, if the video belongs to *Bored*, the foils are *Ashamed*, *Sad*, *Hurt* and *Excited*, and thus the new category is the mean of the other 13 emotion categories (*Afraid*, *Angry*, *Disappointed*, *Disgusted*, *Frustrated*, *Happy*, *Interested*, *Joking*, *Neutral*, *Proud*, *Sneaky*, *Surprised* and *Worried*). For a given video, if the distance to the new category is less than the distance to the target emotion or to any of the foils, then "None of the above" is selected by the classifier.

Table IV presents both the human accuracy and machine accuracy for each emotion category for the 6-option forced-choice responses. Figure 4 depicts the confusion matrices for the 6-option forced choice task for humans and for the machine classifier, respectively.

TABLE IV
CLASSIFICATION RESULTS PER EMOTION CATEGORY

Emotion category	18-choice Machine 2AFC	18-choice Machine Accuracy	6-choice Machine Accuracy	6-choice Human Accuracy
Afraid	0.77	0.24	0.71	0.74
Angry	0.73	0.4	0.5	0.72
Ashamed	0.88	0	0.4	0.75
Bored	0.79	0.43	0.43	0.72
Disappointed	0.98	0.67	0.67	0.64
Disgusted	0.79	0.36	0.57	0.81
Excited	0.81	0.44	0.67	0.79
Frustrated	0.84	0.3	0.7	0.78
Happy	0.87	0.5	0.83	0.8
Hurt	0.67	0.13	0.5	0.66
Interested	0.91	0.5	0.75	0.73
Joking	0.87	0.44	0.78	0.9
Neutral	0.85	0.94	0.94	0.81
Proud	0.94	0.71	0.86	0.76
Sad	0.93	0.62	0.85	0.79
Sneaky	0.80	0.5	0.5	0.68
Surprised	0.84	0.56	0.75	0.86
Worried	0.76	0.2	0.2	0.73
Overall	**0.83**	**0.47**	**0.64**	**0.76**

B. Facial Expressions vs. Head Motions

There has been recent interest in the influence of head motions in conveying affective states [2] [24]. To distinguish the influence of facial expressions and head motions from each other, we have run our classifier in three separate ways: (1)

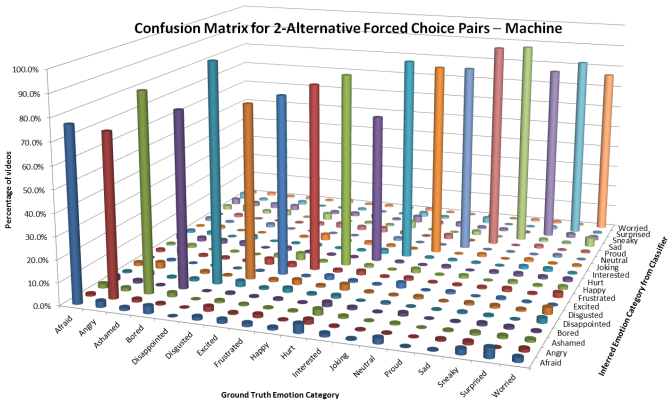
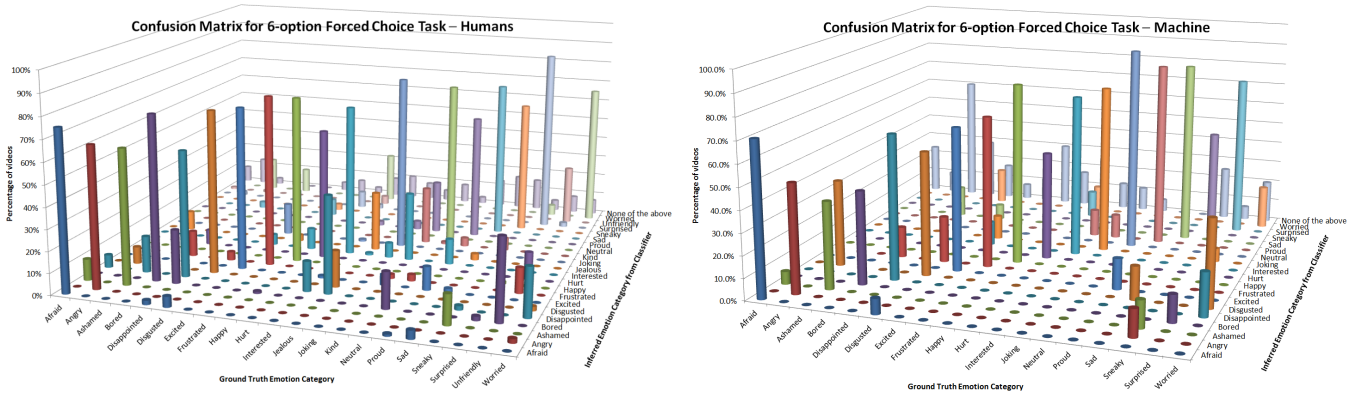


Fig. 3. Confusion matrix for automated emotion classification on 2-Alternative Forced Choice (2AFC) pairs for the 18-choice problem.



(a) Humans

(b) Machine

Fig. 4. Confusion matrices for the 6-option forced choice surveys.

Facial expressions only, (2) Head motions only, and (3) Both facial expressions and head motions.

Figure 2 shows the original classification results as well as the results for **Facial Expressions Only (AROC=0.79, 2AFC=0.79, F1=0.28, Accuracy=0.41)**, and for **Head Pose Only (AROC=0.72, 2AFC=0.70, F1=0.19, Accuracy=0.22)**.

IV. Discussion

It is important to note when considering the relatively low accuracy of the full 18-choice classifier that similar emotion categories were often confused for each other. For example, *Ashamed* and *Hurt* had low classification accuracy rates (0 and 0.13 respectively) and were often confused for each other in the 2AFC pairs. Indeed, for the original validation surveys of the EESS videos, humans were deliberately not given the control response which was most similar to the target emotion (e.g. the target emotion *Disappointed* did not have *Sad* as a control response) as it was believed that "without sufficient context it would be difficult to distinguish these two emotions from one another" [17].

Comparing the human results to the machine results on the 6-option forced choice task (rather than the full 18-choice task) is therefore a more fair comparison. We see that the machine classification is more comparable in this category: 0.76 human accuracy vs. 0.64 machine accuracy.

Facial expressions seemed to have a stronger discriminating effect on classification than head pose information, though the combining both modalities outperformed the face-only classification, suggesting that not all of the emotional information present in the head motions is redundant.

TABLE V
Overall Classification Performance

Classification problem	AROC	2AFC	F1	Accuracy
18-choice, Face and Head	0.84	0.84	0.33	0.47
18-choice, Face only	0.79	0.79	0.28	0.41
18-choice, Head only	0.72	0.70	0.19	0.22
412-choice, Face and Head	0.71	0.69	0.02	0.02

A. Expanding to Larger Numbers of Complex Emotions

One commonly-raised complaint about categorical emotions is the inability of a single label to convey all of the complex information contained in an emotional expression [25]. One approach is therefore to expand the lexicon of emotion labels to capture the subtle differences between affective states.

The Cambridge mindreading (CAM) face-voice battery [18] is another categorically-labelled dataset with facial expression videos. CAM has 412 distinct emotion labels with 6 videos per emotion label, totalling 2,472 videos.

We used a leave-one-out approach to train the classifier on the **CAM dataset (AROC=0.71, 2AFC=0.69, F1=0.018, Accuracy=0.02)**. Figure 6 shows the ROC curves for the CAM classifier based on (1) Facial expressions only, (2) Head motions only, and (3) Both face and head.

Even among these 412 emotion categories, our classifier far outperforms a naive classifier that selects labels proportionally (F1=0.000012, Accuracy = 0.002). However the low accuracy in this scenario may limit its practical use.

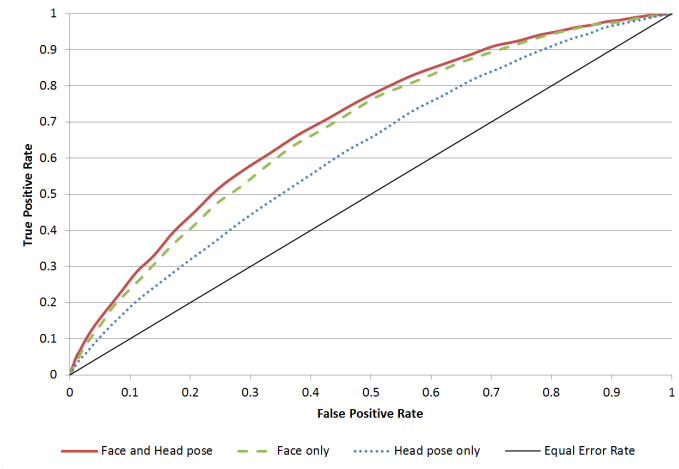


Fig. 6. The Receiver Operating Characteristics (ROC) curve for classification on the 412-emotion category CAM dataset. The Area under the ROC curve (AROC) is 0.71.

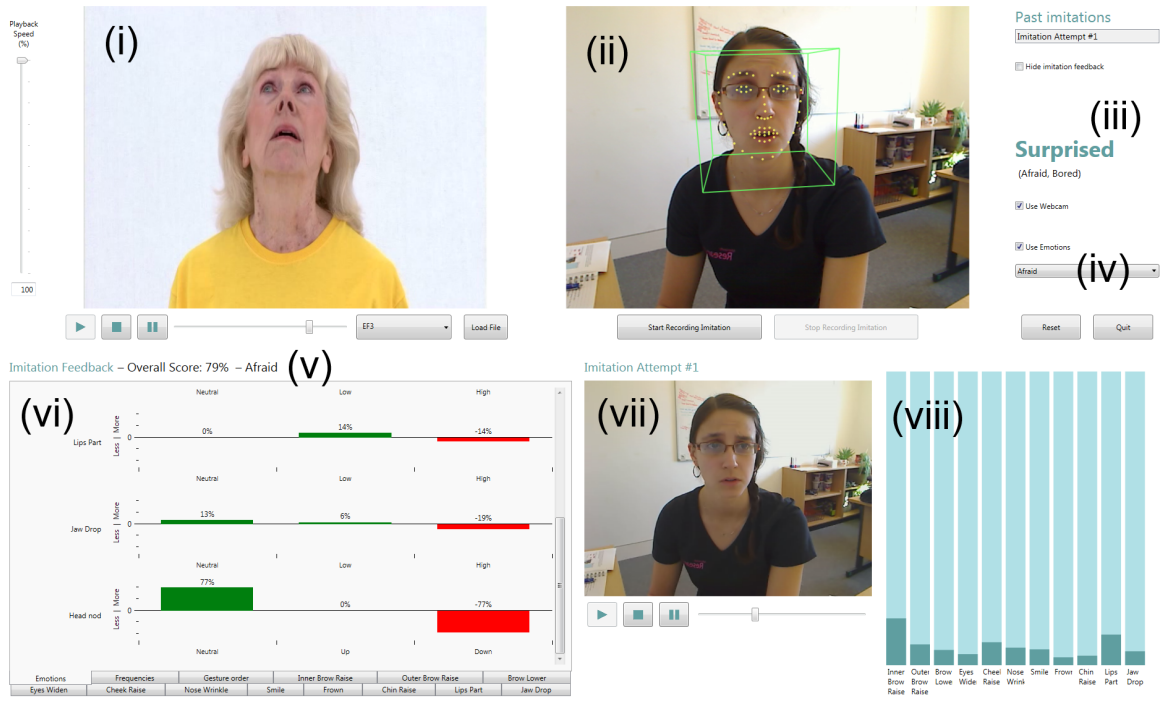


Fig. 5. The emotion classifier was integrated into an expression training interface. (i) A collection of videos belonging to the selected emotion category can be perused. (ii) The webcam video feed is analyzed in real-time using a sliding window to collect the intensities and speed of face and head movements. When the subject is ready, she can record her expression attempt. (iii) The emotion classification for the current real-time sliding window is displayed. (iv) The subject can select a target emotion from 18 emotion categories. (v) The overall emotion classification for the most recent recorded attempt is displayed. (vi) Large differences in the intensities of particular action units for the most recent recorded attempt are displayed to the subject. Green represents "Do more", red represents "Do less". (vii) The most recent recorded attempt can be re-watched. (viii) The action unit intensities for each frame of the recorded attempt from vii are displayed in a moving bar graph as the video plays.

It is unclear whether humans are able to correctly identify videos from the 412 category labels without access to further social context. The validation approach taken in the original CAM study is more similar to a thresholding approach than a 412-choice classification: if 8 out of 10 judges agreed that the label was appropriate for the video, it was accepted [26].

V. Application: Expression Training

The emotion classifier described above has been integrated into an expression training interface that provides humans with feedback on the complex emotions they are portraying through their face and head movements. A screenshot of the system is shown in Figure 5.

The classifier runs in real-time on a webcam video feed. A sliding window of previous frames is used to collect the intensities and speeds of face and head movements from which the emotion classification for the current window is calculated.

The interface provides a database of sample videos for each of the 18 complex emotion categories (the 181 validated videos from the EESS). The subject selects a complex emotion category she wishes to portray and has the option to peruse the sample videos in that category.

When the subject is ready, she presses "Record" and then uses her face and head to portray the selected emotion. When she presses "Stop", any large differences in particular action units between her portrayal and the target emotion are explained to her using simple graphs.

Expression training can be useful as an intervention for individuals with Autism Spectrum Conditions [27], or as a feedback tool for neurotypical individuals wishing to hone their emotion synthesis capabilities (such as actors, customer service representatives, etc).

VI. Conclusion

We have demonstrated that an automated classifier based on simple, human-understandable features can successfully recognize 18 different emotion categories in acted data.

A. Future Work

Classification accuracy might be improved by expanding the set of human-understandable features in the feature vectors used for training. Possible extensions include detecting a greater number of AUs (e.g. lip biting, cheek blowing, tongue show, jaw clencher), and detecting temporal sequences of AUs (e.g. head nodding, head shaking, blinking, eyes darting).

Acknowledgement

This work is generously supported by the Gates Cambridge Trust. The authors would like to thank Erroll Wood for his eye gaze code, and David Dobias, Tadas Baltrušaitis and Marwa Mahmoud for many helpful discussions on this research.

REFERENCES

[1] J. Whitehill, M. S. Bartlett, and J. R. Movellan, "Automatic facial expression recognition," *Social Emotions in Nature and Artifact*, vol. 88, 2013.

[2] Z. Hammal and J. F. Cohn, "Intra-and interpersonal functions of head motion in emotion communication," in *Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges*. ACM, 2014, pp. 19–22.

[3] J. F. Cohn and F. De la Torre, "Automated face analysis for affective computing," *The Oxford Handbook of Affective Computing*, p. 131, 2014.

[4] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[5] A. Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon, "Emotion recognition in the wild challenge 2013," in *Proceedings of the 15th ACM International Conference on Multimodal Interaction*. ACM, 2013, pp. 509–516.

[6] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first Facial Expression Recognition and Analysis challenge," in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*. IEEE, 2011, pp. 921–926.

[7] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2010, pp. 94–101.

[8] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2005, pp. 5–pp.

[9] R. El Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," in *Real-time vision for human-computer interaction*. Springer, 2005, pp. 181–200.

[10] G. McKeown, M. F. Valstar, R. Cowie, and M. Pantic, "The SEMAINE corpus of emotionally coloured character interactions," in *IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2010, pp. 1079–1084.

[11] P. Rozin and A. B. Cohen, "High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans," *Emotion*, vol. 3, no. 1, p. 68, 2003.

[12] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *Journal of neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.

[13] D. Chetverikov and R. Péteri, "A brief survey of dynamic texture description and recognition," in *Computer Recognition Systems*. Springer, 2005, pp. 17–26.

[14] K. Mase, "Recognition of facial expression from optical flow," *IEICE Transactions on Information and Systems*, vol. 74, no. 10, pp. 3474–3483, 1991.

[15] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," *Image and Vision Computing*, vol. 31, no. 2, pp. 120–136, 2013.

[16] P. Zachar and R. D. Ellis, *Categorical versus dimensional models of affect: a seminar on the theories of Panksepp and Russell*. John Benjamins Publishing, 2012, vol. 7.

[17] H. O'Reilly, D. Pigat, S. Fridenson, S. Berggren, S. Tal, O. Golan, S. Blte, S. Baron-Cohen, and D. Lundqvist, "The EU-Emotion Stimulus Set: A validation study," *In press*, 2015.

[18] O. Golan, S. Baron-Cohen, and J. Hill, "The Cambridge mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger syndrome," *Journal of autism and developmental disorders*, vol. 36, no. 2, pp. 169–183, 2006.

[19] T. Bänziger, M. Mortillaro, and K. R. Scherer, "Introducing the geneva multimodal expression corpus for experimental research on emotion perception." *Emotion*, vol. 12, no. 5, p. 1161, 2012.

[20] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Continuous conditional neural fields for structured regression," in *European Conference on Computer Vision*. Springer, 2014, pp. 593–608.

[21] T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic Action Unit detection," *Facial Expression Recognition and Analysis Challenge, Ljubljana, Slovenia*, 2015.

[22] F. Timm and E. Barth, "Accurate eye centre localisation by means of gradients," *International Conference on Computer Vision Theory and Applications*, vol. 1, pp. 125–130, 2011.

[23] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 1997.

[24] H. Gunes and M. Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *Intelligent virtual agents*. Springer, 2010, pp. 371–377.

[25] H. R. Markus and S. Kitayama, "Culture and the self: Implications for cognition, emotion, and motivation," *Psychological review*, vol. 98, no. 2, p. 224, 1991.

[26] O. Golan and S. Baron-Cohen, "Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia," *Development and psychopathology*, vol. 18, no. 02, pp. 591–617, 2006.

[27] J. A. DeQuinzio, D. B. Townsend, P. Sturmey, and C. L. Poulson, "Generalized imitation of facial models by children with autism," *Journal of applied behavior analysis*, vol. 40, no. 4, pp. 755–759, 2007.