Spontaneous facial expression in a small group can be automatically measured: An initial demonstration

Jeffrey F. Cohn and Michael A. Sayette

Department of Psychology, University of Pittsburgh

3137 Sennott Square, 210 S. Bouquet St., Pittsburgh, PA

jeffcohn@cs.cmu.edu sayette@pitt.edu

(412-624-8825, 412-624-8818)

Spontaneous facial expression in a small group can be automatically measured:

An initial demonstration

Manual measurement of facial expression is labor intensive and difficult to standardize. Automated measurement seeks to address the need for valid, efficient, and reproducible measurement. Recent systems have shown promise in posed behavior and in structured contexts. Can automated measurement perform in more natural, less constrained settings? Previously unacquainted young adults sat around a circular table for 30 minutes of conversation. Video was selected for manual and automatic coding of Facial Action Coding System action units 6 (cheek raise) and 12 (lip corner pull), which together signal enjoyment. Moderate out-of-plane head motion and occlusion, which are challenging for automatic processing, were both common, as participants turned toward and away from each other or consumed drinks. Concurrent validity for both action units was high. This is the first study to find that automated measurement of facial action in relatively unconstrained contexts can achieve results comparable to that of manual coding.

Facial expression has been a focus of emotion research for over a hundred years (Darwin, 1872/1998). It is central to several leading theories of emotion (Ekman, 1992; Izard, 1977; Tomkins, 1962) and has been the focus of at times heated debate about issues in emotion science (Ekman, 1994; Fridlund, 1994; Russell, 1994). Facial expression figures prominently in research on almost every aspect of emotion, including psychophysiology (Levenson, Ekman, & Friesen, 1990), neural correlates (Ekman, Davidson, & Friesen, 1990), development (Malatesta, Culver, Tesman, & Shephard, 1989) perception (Ambadar, Schooler, & Cohn, 2005), addiction (Griffin & Sayette, 2008), social processes (Hatfield, Cacioppo, & Rapson, 1992), depression (Reed, Sayette, & Cohn, 2007) and other emotion disorders (Tremeau, et al., 2005), to name a few.

Because of its importance to the study of emotion, a number of observer-based systems of facial expression measurement have been developed (Cohn & Ekman, 2005). Of these various systems, the Facial Action Coding System (FACS) (Ekman, Friesen, & Hager, 2002) is the most comprehensive, psychometrically rigorous, and widely used (Ekman & Rosenberg, 2005). Using FACS and viewing video-recorded facial behavior at frame rate and slow motion, coders can manually code nearly all possible facial expressions, which are decomposed into action units (AUs). Action units, with some qualifications, are the smallest visually discriminable facial movements.

A major challenge in use of FACS and other detailed systems for annotating facial expression is the extensive time required in training and use. Training and passing the certification test for FACS can take six months, and additional training is required before coders are prepared to use FACS to annotate observational data on their own. FACS is labor intensive, requiring up to 1 hour to code a single minute of video (Cohn & Ekman, 2005). Undoubtedly, the exhaustive nature of FACS creates an obstacle to its widespread use.

Not surprisingly, there has been great interest in developing computer-based approaches to facial expression analysis that would permit FACS coding without the time-consuming aspects of doing so manually. If successful, these approaches would greatly improve the efficiency and reliability of facial expression analysis, and more importantly, make its use feasible in applied settings in addition to research. Current methods of assessing psychopathology, for instance, depend almost entirely on verbal report (clinical interviews or questionnaires) of patients, their families, or caregivers. They lack systematic and efficient ways of incorporating behavioral observations that may be strong indicators of psychological disorder. Automated FACS coding could make it possible to use this important source of information.

While the advantages of automated coding are apparent, the challenges to developing such systems are considerable. The face and facial features must be detected in video, shape or appearance information must be extracted and then normalized for variation in pose, illumination, and individual differences in face shape and texture, and then used to segment and classify facial actions. While human observers easily accommodate changes in pose, scale, illumination, occlusion, and individual differences, these and other sources of variation represent considerable challenges for a computer vision system. Then there is the machine-learning challenge of automatically detecting actions that require significant training and expertise even for human coders.

In the past decade, there has been significant effort to develop computer-vision based approaches to automatic coding of facial expression. Early work focused on posed facial expressions with frontal camera orientation, little or no head motion or occlusion, and moderate to strong expressions (Bartlett, Ekman, Hager, & Sejnowski, 1999; Cohn, Zlochower, Lien, & Kanade, 1999; Essa & Pentland, 1997; Pantic & Rothkrantz, 2000). Tian, Kanade, and Cohn (2001), for example, automatically detected 34 action units and action unit combinations in full-face frontal view images. More recently, investigators have made progress in the more demanding task of action unit detection in non-posed facial images. Valstar, Gunes, and Pantic (2007) and Cohn and Schmidt (2004) automatically discriminated posed from non-posed, naturally occurring smiles. Ambadar, Cohn, and Reed (2009) used computer-vision based measures to differentiate polite, happy, and embarrassed smiles. Metaxas and colleagues detected stress from automatic measures of facial expression (Dinges, et al., 2005). At least two groups have discriminated facial expressions or episodes of physical pain under relatively constrained conditions (Ashraf, et al., 2009; Littlewort, Bartlett, Fasel, Susskind, & Movellan, 2006). Messinger and colleagues (Messinger, Mahoor, Chow, & Cohn, 2009) demonstrated a pilot system for automatic measurement of smiles in mothers and infants during face-to-face interaction. Whitehill and colleagues (Whitehill, Littlewort, Fasel, Bartlett, & Movellan, 2009) detected smile intensity in video from five subjects while they individually watched a short video clip. At least one commercial product (Theuws, Undated) has been released that attempts to identify emotion expressions from frontal video with little or no head motion or occlusion. In each of these studies, faces were recorded from frontal or near-frontal views, and behavior samples were obtained during relatively structured tasks such as computer viewing, sitting in front of a computer display or camera, and structured face-to-face interaction.

The pending challenge is to demonstrate the ability of automatic methods to reliably detect non-posed facial actions in less constrained contexts. *We used automated facial image analysis to provide the first test of automated FACS action unit detection in a multi-person group of strangers interacting in a relatively unstructured context.* None of the participants were associated with the experiment as experimenters or confederates, and thus the facial movements were spontaneous and unscripted. Because participants were seated around a circular table, we anticipated that the video would include moderate to large head rotation as they turned toward and away from each other. We also anticipated frequent occlusion, as subjects frequently drank beverages (a glass of juice) supplied by the experimenters.

The present test focused on a critical pair of AUs (AU 6 and 12). AU 12 is caused by contraction of the *zygomatic major* muscle, which pulls the lip corners obliquely. AU 6 is caused by contraction of the *orbicularis oculi* muscle, which raises the cheeks and causes crow-feet wrinkles to form lateral to the outer eye corners. When these actions occur together they comprise what Ekman (Ekman, et al., 1990) termed the "Duchenne smile." This smile is thought to be a "true" smile and reflect happiness.

Smiles are the most frequent of all emotion expressions, occurring as often as three or more times per minute during social interaction (Schmidt & Cohn, 2001). They are one of a small number of expressions for which there is evidence of universality (Ekman, 1993). And in circumplex models of emotion (Larsen & Diener, 1992), they indicate positive valence (Cacioppo, Petty, Losch, & Kim, 1986). To demonstrate that this expression can be automatically coded in a reliable fashion during relatively unstructured, multi-person social interaction would mark an important step forward in the development of automatic approaches to coding facial expression.

Method

Digital video from three subjects who were participating in a larger study was used for the present test. These subjects were involved in a study (in progress) examining group formation processes [see (Kirchner, Sayette, Cohn, Moreland, & Levine, 2006) for details of a prior study using this design.] Although the experiment was advertised as a study examining the impact of alcohol on cognitive functioning, the group of subjects used in the present FACS analyses had been randomly assigned to a no-alcohol control condition in which they were explicitly told, and indeed provided, with cranberry juice only. They drank juice throughout the observation period. All three individuals reported that they had not consumed alcohol or other psychoactive drugs (except nicotine or caffeine) during the 24-hour period leading up the observations and (all reported a 0 on a 101 point intoxication scale during the experiment.)

Observational procedures

On arrival, participants' height and weight were recorded. They also ate a light, weight-adjusted snack (a bagel with butter) and completed a consent form describing the study. To ensure that the group was composed of three unacquainted "strangers," four people were invited to the laboratory (see Kirchner et al., 2006 for details of the overall procedures for a similar study.) Participants were told there was a slight chance that they might be asked to return on another day, in which case they would receive an extra $20. Participants were greeted separately and placed in different rooms. Then they were casually introduced to each other one at a time while being observed by two researchers for any signs of recognition. None showed any sign of recognition. Following initial greetings, they also were asked if they had ever met the others (and they reported that they had not ever met).

Setting and equipment. The three members of the group were escorted to the experimental room and seated equidistant from each other around a circular (75-cm diameter) table. They were asked to consume a control beverage consisting entirely of cranberry juice before engaging in a variety of cognitive tasks over a 36-min period. Separate wall-mounted cameras faced each person. It was explained that the cameras were focused on their drinks and would be used to monitor their consumption rate from the adjoining room. Following the drink and the cognitive tasks, subjects were debriefed, paid, and permitted to leave.

The laboratory included a custom-designed video control system that permits synchronized video output for each subject, as well as an overhead shot of the group and a quad-split image showing both the individual and group views (Figure 1). The individual view for each subject was used in this report. An example is shown below.

Insert Figure 1 about Here

Manual FACS coding

For each participant, six minutes of continuous video from the middle of the observation period was selected for analysis. Two certified FACS coders independently coded action units 6 and 12 and instances of occlusion from the digital video using Observer Video-Pro Software (Noldus, Trienes, Henriksen, Jansen, & Jansen, 2000). Occlusion was defined as any obstruction of part of the face by a hand, glass, or other object, or a portion of face moving out of the field of view. Self-occlusion, caused by non-rigid head motion, was not coded but was taken into account as described in the following paragraph. The Observer system makes it possible to manually code digital video in stop-frame and at variable speed and later synchronize codes according to digital time stamp. Inter-observer exact (30f/s) agreement was quantified using coefficient kappa, which is the proportion of agreement above what would be expected to occur by chance. Kappa coefficients were 0.68 and 0.78 for AU 6 and AU 12, respectively, and 0.94 for occlusion.

Orientation to the camera was quantified automatically using a non-rigid structure from motion algorithm as noted above. Head orientation is important because the face looks different from different views and parts of the face may become self-occluded. We evaluated AU detection in relation to variation in pitch (the head rotating up or down, as in head nods) and yaw (the head rotating to the left or right, as in head turns).

Automatic facial image analysis

Automatic facial image analysis included three steps. These were 1) extract the face shape and appearance using an Active Appearance Model (AAM) (Matthews & Baker, 2004); 2) normalize shape and appearance to control for variation due to rigid head motion (e.g., turning toward or away from other participants); 3) detect FACS action units.

Active Appearance Model. AAMs decouple shape and appearance of a face image. Given a pre-defined linear shape model with linear appearance variation, AAMs align the shape model to an unseen image containing the face and facial expression of interest. To train an AAM for each participant, approximately 3% of keyframes were manually labeled during a training phase. The remaining frames were automatically aligned using a gradient-descent AAM fit described in (Matthews & Baker, 2004; Xiao, Baker, Matthews, & Kanade, 2004).

Insert Figure 2 about Here

The *shape* $\mathbf{s}$ of an AAM is described by a 2D triangulated mesh. In particular, the coordinates of the mesh vertices define the shape $\mathbf{s}$ (Ashraf, et al., 2009). These vertex locations correspond to a source appearance image, from which the shape is aligned. Since AAMs allow linear shape variation, the shape $\mathbf{s}$ can be expressed as a base shape $\mathbf{s}_0$ plus a linear combination of m shape vectors $\mathbf{s}_i$:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{m} p_i \mathbf{s}_i$$

where the coefficients $\mathbf{p} = (p_1, \ldots, p_m)^{\mathrm{T}}$ are the shape parameters (See Figure 2). Additionally, a global normalizing transformation (in this case, a geometric similarity transform) is applied to $\mathbf{s}$ to remove variation due to rigid motion (i.e. translation, rotation, and scale). The parameters p_i are the residual parameters representing variations associated with the actual object shape (e.g., mouth opening and eye closing). Given a set of training shapes, Procrustes alignment is employed to normalize these shapes and estimate the base shape $\mathbf{s}_0$, and Principal Component Analysis (PCA) is then used to obtain the shape and appearance basis eigenvectors $\mathbf{s}_i$ (Matthews & Baker, 2004). A non-rigid structure from motion algorithm is used to estimate head pose parameters (e.g., pitch and yaw)(Matthews, Xiao, & Baker, 2007; Xiao, et al., 2004). Because AAMs are invertible, they can be used both for analysis, as in the current study, and for synthesizing new images (Theobald & Cohn, 2009).

AAM features. Although person-specifc AAM models were used for tracking, a global model of the shape variation across all sessions was built to obtain the shape basis vectors and corresponding similarity normalized coefficients p_i. A model common to all subjects is necessary to ensure that the meaning of each of the coefficients is comparable across sessions. 95% of the energy was retained in the PCA dimensionality reduction step, resulting in 10 principal components or shape eigenvectors.

Action unit detection

Action units were detected using support vector machine classifiers (SVM) (Hsu, Chang, & Lin, 2005). SVMs attempt to find the hyper-plane that maximizes the margin between positive and negative observations for a specified class. For AAM shape and appearance coefficients, they seek to maximize the boundary between each action unit (e.g., AU 6) and all instances of other action units including neutral faces (i.e., AU 0 in FACS).

To maximize generalizability, we trained and tested the SVMs on independent data. For training, we use used the RU-FACS (Frank, Movellan, Bartlett, & Littlewort, Undated) image database. RU-FACS consists of digitized video and manual FACS coding of 34 young adults. They were recorded during an interview of approximately 2 minutes duration in which they lied or told the truth in response to an interviewer's questions. Pose orientation was mostly frontal with small out-of-plane head motion. Image data from five subjects could not be analyzed due to image artifact. Thus, image data from 29 subjects was used for training the classifiers. Classifiers then were tested on the independent subjects from the current study.

Results

Descriptive statistics

AU 6 and AU 12 occurred 7% and 32% of the time, respectively. With the exception of only four video frames, AU 6 always occurred in the presence of AU 12. Thus, AU 6 was a reliable signal of Duchenne smiling.

Occlusion, defined as partial obstruction of the view of the face, occurred in 10.9% of video frames. Head orientation was variable. Mean orientation was 6.78 degrees from frontal view for pitch and 6.79 degrees from frontal view for yaw. (Here and following, absolute values are reported for pitch and yaw). For pitch, the 90^{th} and 95^{th} percentiles were 12.85 and 15.54 degrees from frontal, respectively. For yaw, the corresponding values were 12.53 and 15.86 degrees. Maximum pitch was 28.82 degrees; maximum yaw was 73.37 degrees.

Automatic action unit detection

We compared automatic and manual FACS coding of the three 6-min video streams of naturally occurring facial expression during the social interaction. The video included out-of-plane head motion (pitch and yaw) and partial occlusion, which are challenging for automatic coding. Figure 3 shows an example of the face tracking. (For video demo, please see supplemental materials). The face image with tracked facial features appears in the bottom panel. Across the top are the similarity transformed and piece-wise warped appearance. The former is appearance after removing rotation and translation; translation is variation due to change in horizontal and vertical motion and scale. In piece-wise normalized appearance, variation due to out-of-plane head motion has been removed and thus stabilized for all but non-rigid motion (i.e. expression). To the right of the appearance are three representations of shape. The first is 2D, the second is 3D when viewed from a ¾ of frontal pose, and the third is 3D viewed from above the face. By stabilizing the face image and estimating change in rigid head motion, potential confounds in AU detection due to rigid head motion are removed. Also, head motion may itself be an important nonverbal cue, and thus useful in its own right to measure.

Insert Figure 3 about Here

As noted above, classifiers were trained on video from an independent database (RU-FACS) and tested on the subjects from the experiment in progress. Automatically coded AU 6 and AU 12 for the current video then were compared on a frame-by-frame basis with manual FACS coding. Following previous literature (Ashraf, et al., 2009; Pantic & Bartlett, 2007), we quantified accuracy using receiver operator characteristics curves (ROC). ROC curves illustrate the relation between true and false positive rates of classifiers as the decision threshold varies. Area under the curve (*A'*) can vary from 0 to 1.00, with 0.50 representing the expected value of random guessing.

ROC curves are widely used in signal detection, analysis of diagnostic systems, and machine learning (Fawcett, 2005). They are especially useful for skewed distributions, such as those for action units, and unequal classification costs.

True positive rate (*TPR*), also known as “sensitivity” or “recall,” is defined as:

$$TPR = TP / (TP + FN)$$

where *TP* is “true positive” and *FN* is “false negative.” False positive rate (*FPR*), also known as “false alarm rate” or “1 – Sensitivity” is defined as:

$$FPR = FP / (FP + TN)$$

where *FP* is “false positive” and *TN* is “true negative.”

We first report results for the entire video. We then report results in relation to occlusion and non-rigid head-motion.

Insert Figures 4

Concurrent validity for the entire video. Concurrent validity was high for both AU 6 and AU 12 (Figure 4). For AU 6, *A'* was .96 (standard error = .002, p < .0001); for AU 12, *A'* was .88 (standard error = .004, p < ,0001). For comparison with inter-observer agreement, *A'* to measure false positive rates corresponding to 90% and 80% true positive rates. A hit rate of 70% is required to pass the FACS certification test; hit rates of 70% or above are common in research that uses FACS (Ekman & Rosenberg, 2005).

Insert Table 1 about Here

For AU 6, false positive rates were well within acceptable limits even when true positive rate was set to 90% (Table 1). For AU 12, an 80% true positive rate yielded an acceptable false positive rate.

Concurrent validity as a function of occlusion and non-rigid head motion. To assess the influences of occlusion on AU detection, we computed ROC curves separately for video with and without occlusion. For AU 6, occlusion reduced *A'* from .97 to .91. For AU 12, occlusion reduced *A'* from .90 to .75.

To assess robustness to pitch and yaw, we computed ROC curves separately for every 5 degrees of pitch and yaw variation. AU 6 occurred through 20 degrees pitch and all intervals of yaw (Table 2). AU 12 occurred across the full range for both pitch and yaw. Most variation in both pitch and yaw variation was within intervals between ± 0 to 20 degrees; results for intervals outside of this range should be interpreted with caution.

With respect to pitch, *A'* for AU 6 was stable (mean = .96) through ±15 degrees pitch and decreased to .89 at ±15 to 20 degrees pitch (Table 2). For AU 12, *A'* was stable through ±20 degrees pitch (mean = .88) and then decreased to .70 over the interval between ±20 to 25 degrees pitch. For pitch variation greater than ±25 degrees, there was an uptick for AU 12.

With respect to yaw, AU 6 was stable (mean = .97) through ±20 degrees and then decreased to .89 and .83 at ±20-to-25 and ±25 degrees or greater. For AU 12, *A'* was stable (mean = .90) through ±15 degrees yaw and then decreased gradually to .78 at yaw greater than ±25 degrees. Overall, the general pattern for both AU 6 and AU 12 for pitch and yaw was for *A'* to remain high for intervals through ±15 or 20 degrees and then to decrease. For no intervals did *A'* go below .70, and in all but one case was .78 or higher.

Insert Table 2 about Here

Discussion

This study provides initial support for the use of automatic facial image analysis for the detection of emotion expression to detect spontaneous facial expressions arising during unstructured social interaction. Specifically, using automated facial image analysis, two action units critical to positive emotion, AUs 6 and 12, were automatically detected with high reliability as compared with two independent certified FACS coders. For AU 6, setting true positive rate as high as 90% resulted in only a small error rate of

6%. For AU 12, setting a true positive rate of 80% resulted in an acceptable error rate of 15%. These values may actually underestimate the effective reliability of automatic coding. That is because they are based on exact agreement. Most users of FACS, and many similar coding systems, estimate agreement within a precision window of plus/minus 0.5 to 1 second (Sayette, Cohn, Wertz, Perrott, & Parrott, 2001), which would effectively increase intersystem agreement.

When hits and misses were reviewed in the video, it appeared that many false positives occurred with talking or were confusions with talking. Talking has always been a challenge to manual FACS coding. Guidelines for manual FACS coding were initially to code “Talk” (AU 50) in place of other action units in the mouth region when present. While this instruction often proves too costly in lost information, criteria for coding action units in the presence of talking remain lacking. For an automatic system, it may be helpful to code talking on a continuous time base and to include that in training AU 12 detectors.

Partial occlusion was relatively common, occurring in about 11% of video frames. Face touching, holding a drink in front of or touching the face, and the face moving out of view were frequent causes. Had drinking not been part of the experimental protocol, partial occlusion may have occurred less frequently. Nevertheless, partial occlusion had minimal effect on accuracy for AU 6, perhaps because face touches and drinking were more likely to occlude only the lower face. For AU 12, area under the ROC curve decreased by about 13% when occlusion occurred. The effects of occlusion were not uniform, but varied with AU.

Accuracy of action detection was stable within a range of about ±15 to 20 degrees. Beyond that, accuracy decreased, although still remaining within acceptable limits. The findings with respect to larger values of pitch and yaw must be considered with caution, in that head orientation at higher ranges was relatively uncommon. With that caveat, the findings suggest that automatic facial image analysis is capable of performing well within a much larger range of pitch and yaw than has been demonstrated previously. Indeed, this is the first study to report AU detection results in relation to parametric variation in pitch and yaw. The findings provide initial evidence that automatic facial image analysis can perform well within the range of head orientation that is likely to occur in unscripted, spontaneous facial behavior in a social setting.

Future research with larger data sets is needed to replicate and extend these preliminary findings. Many factors may potentially influence system accuracy. The influence of skin color, glasses, facial jewelry, and lighting will require careful evaluation. The face models appear robust to differences in ethnic or racial background but dark skin may require attention to illumination. With respect to glasses, in our experience glasses only interfere with tracking when the lenses are highly reflective. When that occurs, eye closure may be difficult to detect. In the current study, all participants were men; it is possible that results could have differed for women, although that seems unlikely. While men have more facial texture in the lower face, which could contribute to extraction of appearance features, sex differences have not emerged in previous research using posed facial behavior. Also, in so far as women are more expressive or smile more than men (e.g.,LaFrance, Hecht, & Paluck, 2003), their facial expression in naturally occurring behavior should be more easily detected. With respect to pose and occlusion, we found that AU detection was relatively robust. Future work will want to examine these issues and include a larger number of action units and emotions.

Despite the limitations of the current study, this initial demonstration of the efficacy of automated facial image analysis suggests that deployment of automated facial image analysis in behavioral research may be close at hand. Initial efforts have used automated facial image analysis to study pain (Ashraf, et al., 2009; Cohn, Lucey, et al., 2009; Littlewort, Bartlett, & Lee, in press, 2009), smiling (Ambadar, et al., 2009; Cohn & Schmidt, 2004; Schmidt, Lui, & Cohn, 2006), and measurement of depression severity (Cohn, Simon Kreuz, et al., 2009; Wang, et al., 2008) over relatively brief periods under more controlled conditions. The current paper is the first to use automated facial image analysis in relatively unconstrained small-group interactions over relatively long spans of many minutes. In this expanded use, we found that automated facial image analysis had high concurrent validity with manual FACS coding. Automated facial image analysis appears on the verge of impacting a wide range of clinical and research applications. For the first time, precise and valid measurement will be possible without reliance on laborious training and coding. Efficiencies of scale, including real-time applications (Ryan, Cohn, & Hamerski, 2009), are about to significantly boost research productivity and open new areas of investigation.

Acknowledgements

This research was supported in part by NIH grants MH 051435 and AA015773. Tomas Kruez Simon, Kasey Griffin, Iain Matthews, Nicole Ridgeway, and John Dimoff provided technical assistance.

References

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior, 33*, 17-34.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions. *Psychological Science, 16*, 403-410.

Ashraf, A. B., Lucey, S., Cohn, J. F., Chen, T., Prkachin, K. M., & Solomon, P. (2009). The painful face: Pain expression recognition using active appearance models. *Image and Vision Computing, 27*, 1788–1796.

Bartlett, M. S., Ekman, P., Hager, J. C., & Sejnowski, T. J. (1999). Measuring Facial Expressions by Computer Image Analysis. *Psychophysiology, 36*, 253-263.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H.-S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50*(2), 260-268.

Cohn, J. F., & Ekman, P. (2005). Measuring facial action by manual coding, facial EMG, and automatic facial image analysis. In J. A. Harrigan, R. Rosenthal & K. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences* (pp. 9-64). New York: Oxford.

Cohn, J. F., Lucey, P., Ambadar, Z., Lucey, S., Solomon, P., & Prkachin, K. (2009). *The face of pain: Automatic recognition and analysis using computer vision and machine learning*. Paper presented at the American Psychological Society.

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing, 2*, 1-12.

Cohn, J. F., Simon Kreuz, T., Matthews, I., Yang, Y., Nguyen, M. H., Tejera Padilla, M., et al. (2009). *Detecting depression from facial actions and vocal prosody. Affective Computing and Intelligent Interaction (ACII 2009),* 1-7.

Cohn, J. F., Zlochower, A. J., Lien, J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology, 36*, 35-43.

Darwin, C. (1872/1998). *The expression of the emotions in man and animals (3rd edition)*. New York: Oxford University.

Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Ziga Cizman, et al. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, Space, and Environmental Medicine, 76*(6), 172-182.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3/4), 169-200.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384-392.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Review, 115*(2), 268-287.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain Psychology II. *Journal of Personality and Social Psychology, 58*(2), 342-353.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system*: Research Nexus, Network Research Information, Salt Lake City, UT.

Ekman, P., & Rosenberg, E. (2005). *What the face reveals* (2nd ed.). New York: Oxford.

Essa, I., & Pentland, A. (1997). Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 7*, 757-763.

Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861-874.

Frank, M., Movellan, J., Bartlett, M. S., & Littlewort, G. (Undated). RU-FACS-1 Database, *Machine Perception Laboratory, U.C. San Diego*.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. New York: Academic Press.

Griffin, K. M., & Sayette, M. A. (2008). Facial reactions to smoking cues relate to ambivalence about smoking. *Psychology of Addictive Behaviors*, *22*(4), 551-556.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1992). Primitive emotional contagion. In M. S. Clark (Ed.), *Emotion and Social Behavior* (Vol. 14, pp. 151-177). Newbury Park, CA: Sage Publications.

Hsu, C. W., Chang, C. C., & Lin, C. J. (2005). *A practical guide to support vector classification*: Department of Computer Science, National Taiwan University.

Izard, C. E. (1977). *Human emotions*. New York: Plenum.

Kirchner, T. R., Sayette, M. A., Cohn, J. F., Moreland, R. L., & Levine, J. M. (2006). Effects of alcohol on group formation among male social drinkers. *Journal of Studies on Alcohol, 67*, 785-793.

LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: A meta-analysis of sex differences in smiling. *Psychological Bulletin, 129*(2), 305-334.

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Emotion* (Vol. 13, pp. 25-59). Newbury Park, CA: Sage Publications.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*(4), 363-384.

Littlewort, G. C., Bartlett, M. S., Fasel, I. R., Susskind, J., & Movellan, J. (2006). Dynamics of facial expression extracted automatically from video. *Journal of Image & Vision Computing, 24*(6), 615-625.

Littlewort, G. C., Bartlett, M. S., & Lee, K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing, 27*(12), 1797-1803.

Malatesta, C. Z., Culver, C., Tesman, J. R., & Shephard, B. (1989). The development of emotion expression during the first two years of life. *Monographs of the Society for Research in Child Development, 54*(219).

Matthews, I., & Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision, 60*(2), 135-164.

Matthews, I., Xiao, J., & Baker, S. (2007). 2D vs. 3D deformable face models: Representational power, construction, and real-time fitting. *International Journal of Computer Vision, 75*(1), 93–113.

Messinger, D. S., Mahoor, M. H., Chow, S. M., & Cohn, J. F. (2009). Automated measurement of facial expression in infant-mother interaction: A pilot study. *Infancy, 14*(3), 285-305.

Noldus, L. P. J. J., Trienes, R. J. H., Henriksen, A. H. M., Jansen, H., & Jansen, R. G. (2000). The Observer Video-Pro : New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments, and Computers, 32,* 197-206.

Pantic, M., & Bartlett, M. S. (2007). Machine analysis of facial expressions. In K. Delac & M. Grgic (Eds.), *Face recognition* (pp. 377-416). Vienna: I-Tech Education and Publishing.

Pantic, M., & Rothkrantz, M. (2000). Expert system for automatic analysis of facial expression. *Image and Vision Computing, 18,* 881-905.

Reed, L. I., Sayette, M. A., & Cohn, J. F. (2007). Impact of depression on response to comedy: A dynamic facial coding analysis. *Abnormal Psychology, 116*(4), 804-809.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115*(1), 102-141.

Ryan, A. H., Cohn, J. F., & Hamerski, R. (2009). *Automated facial expression recognition system. Proceedings of the IEEE International Carnahan Conference on Security Technology,* 172-177.

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Parrott, D. J. (2001). A psychometric evaluation of the Facial Action Coding System for assessing spontaneous expression. *Journal of Nonverbal Behavior, 25,* 167-186.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary perspectives in facial expression research. *Yearbook of Physical Anthropology, 116*, 8-24.

Schmidt, K. L., Lui, Y., & Cohn, J. F. (2006). The role of structural facial asymmetry in asymmetry of peak facial expressions. *Laterality, 11*(6), 540-561.

Theobald, B. J., & Cohn, J. F. (2009). Facial image synthesis. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (pp. 176-179). NY: Oxford University Press.

Theuws, H. (Undated). Face Reader: Noldus Corporation.

Tian, Y., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(2), 97-115.

Tomkins, S. S. (1962). *Affect, imagery, consciousness*. New York: Springer.

Tremeau, F., Malaspina, D., Duval, F., Correa, H., Hager-Budny, M., Coin-Bariou, L., et al. (2005). Facial expressiveness in patients with schizophrenia compared to depressed patients and nonpatient comparison subjects. *American Journal of Psychiatry, 162*(1), 92-101. .

Valstar, M. F., Gunes, H., & Pantic, M. (2007). How to distinguish posed from spontaneous smiles using geometric features. *Proceedings of the ACM International Conference on Multimodal Interfaces*, 38-45.

Wang, P., Barrett, F., Martin, E., Milonova, M., Gurd, R. E., Gur, R. C., et al. (2008). Automated video-based facial expression analysis of neuropsychiatric disorders. *Journal of Neuroscience Methods, 168*, 224–238.

Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M. S., & Movellan, J. (2009). Towards practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(*11), 2106 - 2111.

Xiao, J., Baker, S., Matthews, I., & Kanade, T. (2004). Real-time combined 2D+3D active appearance models. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2*, 535-542.

Table 1

Selected corresponding true- and false positive rates for AU 6 and AU 12.

Action Unit	True Positive Rate	False Positive Rate
AU 6	90%	6%
	80%	3%
AU 12	90%	37%
	80%	15%

Table 2

Area under the ROC curve in relation to partial occlusion, pitch, and yaw.

	Percentage of time	Area under the ROC	
		AU 6	AU 12
Occlusion			
Absent	89.22	.97 (.002)	.90 (.002)
Present	10.78	.91 (.012)	.75 (.012)
Pitch (absolute value)			
0 to 5 degrees	41.08	.96 (.004)	.86 (.004)
5 to 10 degrees	36.58	.97 (.004)	.90 (.003)
10 to 15 degrees	16.81	.96 (.006)	.90 (.005)
15 to 20 degrees	4.50	.89 (.061)	.87 (.020)
20 to 25 degrees	.71	NA	.70 (.037)
> 25 degrees	.32	NA	.94 (.027)
Yaw (absolute value)			
0 to 5 degrees	42.85	.96 (.004)	.87 (.004)
5 to 10 degrees	34.75	.97 (.004)	.91 (.004)

10 to 15 degrees	16.69	.97 (.004)	.91 (.005)
15 to 20 degrees	3.37	.98 (.004)	.84 (.013)
20 to 25 degrees	1.10	.89 (.036)	. 80 (.038)
> 25 degrees	1.23	.83 (.064)	.78 (.030)

Note. All $p < .0001$.

NA = No occurrences of the AU for this range.

Figure legends

Figure 1. An example of the image capture. Separate wall-mounted cameras were directed at each participant. A ceiling-mounted camera recorded an overview. Face meshes from the AAM are superimposed on the source video.

Figure 2. An example of the computation of AAM shape and appearance. The figure shows the mean and first two modes of variation of 2D AAM shape (a–c) and appearance (d–f) variation and the mean and first two modes of 3D AAM shape. From Xiao, Baker, Matthews, & Kanade (2004). © IEEE.

Figure 3. Screen shot of automated face tracking and AU detection. Row 2 shows a subject's tracked face and frame-by-frame detection results for AU 6 and AU 12. Row 1 column A shows corresponding 2D face appearance normalized for head translation, scale, and rotation. Row 1 column B shows the appearance after normalizing for pitch (e.g., head nodding) and yaw (e.g., head turning). Row 1 columns C through E show 2D and 3D representations of the corresponding face shapes. Row 1 column F shows the estimated 3D parameters (pitch, roll, and yaw). Please see Supplementary Materials for video examples (The video examples also are available at http://www.pitt.edu/~jeffcohn/D102_G006A1_trim.mov and http://www.pitt.edu/~jeffcohn/D102_G014A1_trim.mov).

Figure 4. The receiver operating characteristics (ROC) curves for AU 6 and AU 12 in comparison with random guessing (depicted by the diagonal lines). The corresponding areas under the ROC were 0.94 (standard error = .002, $p < .0001$) and 0.85 (standard error = .004, $p < .0001$).

Figure 1

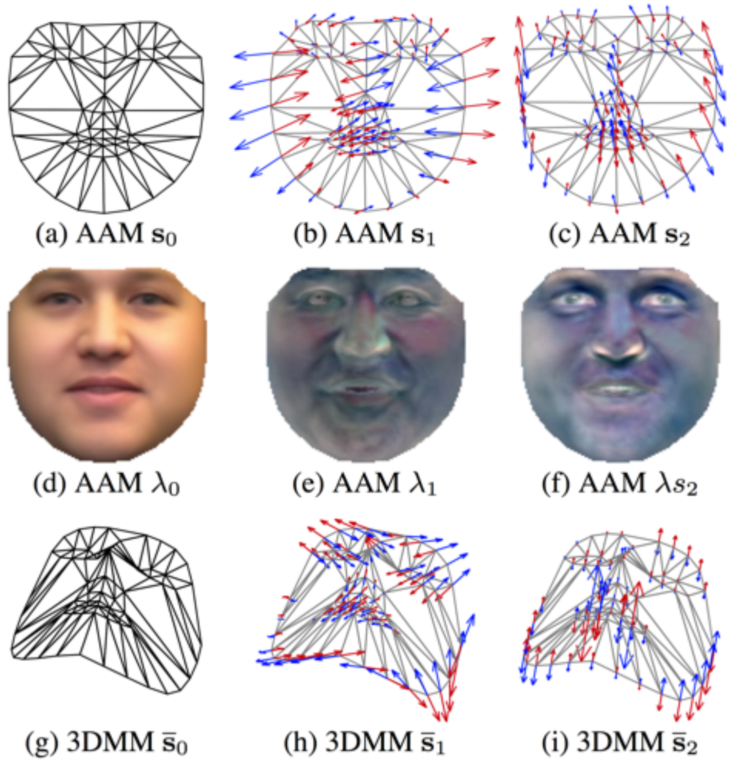

(a) AAM $\mathbf{s}_0$
(b) AAM $\mathbf{s}_1$
(c) AAM $\mathbf{s}_2$
(d) AAM λ_0
(e) AAM λ_1
(f) AAM λs_2
(g) 3DMM $\bar{\mathbf{s}}_0$
(h) 3DMM $\bar{\mathbf{s}}_1$
(i) 3DMM $\bar{\mathbf{s}}_2$

Figure 2

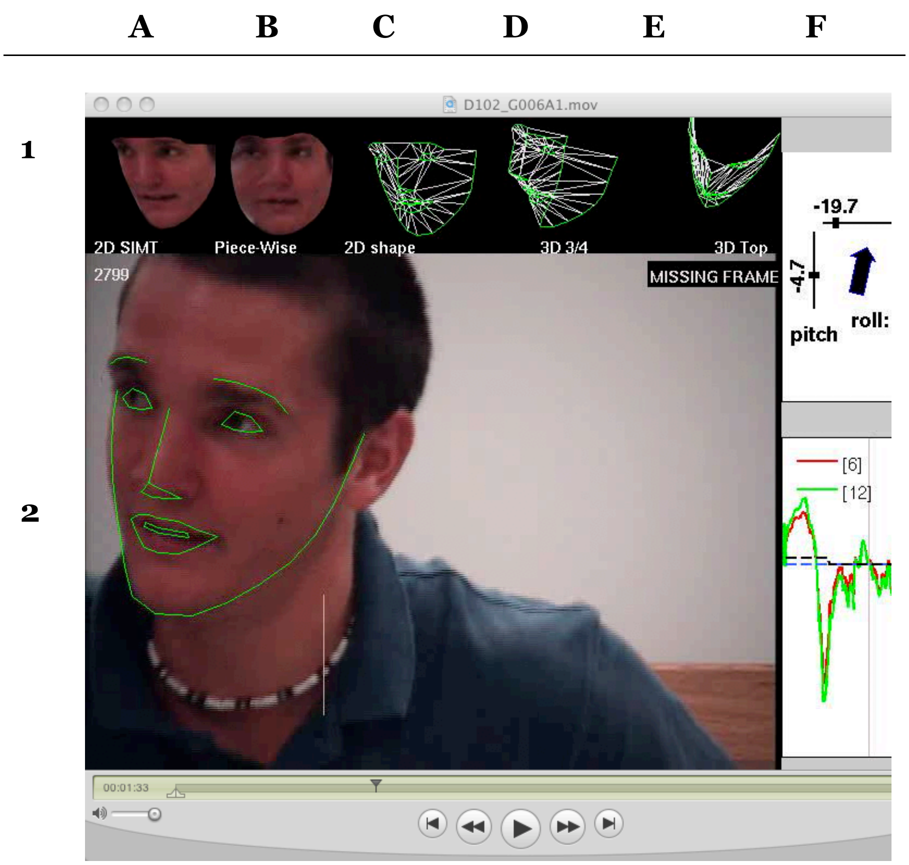
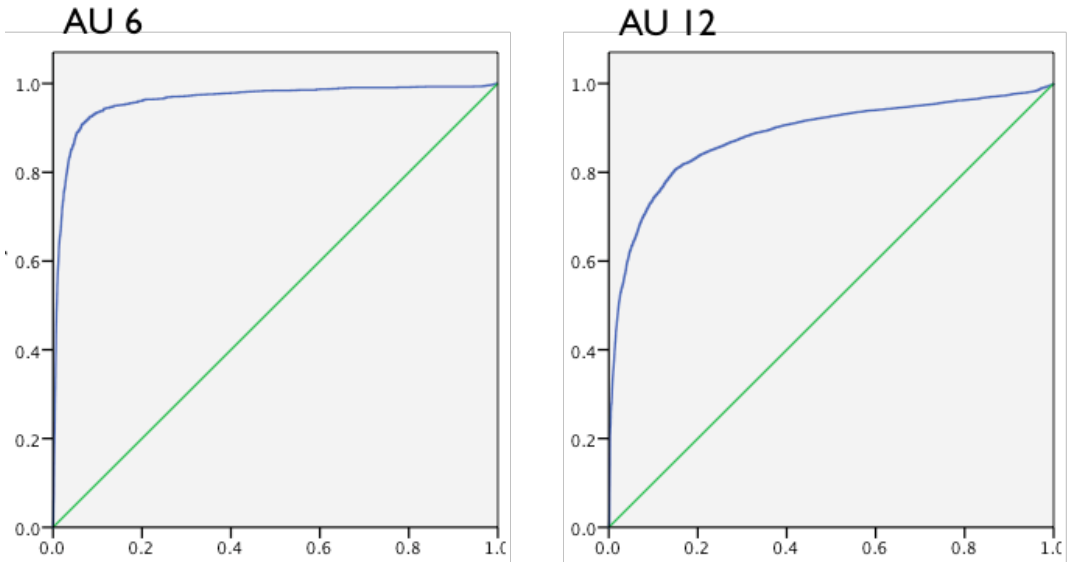

A
B
C
D
E
F
1
2
D102_G006A1.mov
2D SIMT
Piece-Wise
2D shape
3D 3/4
3D Top
2799
MISSING FRAME
-19.7
-4.7
pitch
roll:
[6]
[12]
00:01:33

Figure 3

True
Positive
Rate



False Positive Rate

Figure 4