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

JEFFREY F. COHN AND MICHAEL A. SAYETTE University of Pittsburgh, Pittsburgh, Pennsylvania

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? In the present study, previously unacquainted young adults sat around a circular table for 30 min of conversation. Video was selected for manual and automatic coding of Facial Action Coding System action units (AUs), examining, in particular, AU 6 (cheek raise) and AU 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 AUs was high. This is the first study to find that automated measurement of facial action in relatively unconstrained contexts can achieve results comparable with those of manual coding. Video demos of our software may be downloaded from http://brm.psychonomic-journals.org/content/supplemental.

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, & Shepard, 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 the importance of facial expression to the study of emotion, a variety of observer-based systems of facial-expression measurement have been developed (Cohn & Ekman, 2005). Of these 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), which, with some qualifications, are the smallest visually discriminable facial movements.

A major challenge in the 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 6 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 h 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 timeconsuming 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.

Although 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, then normalized for variation in pose, in illumination, and in individual differences in face shape and texture, and



then used to segment and classify facial actions. Although 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, Hager, Ekman, & Sejnowski, 1999; Cohn, Zlochower, Lien, & Kanade, 1999; Essa & Pentland, 1997; Pantic & Rothkrantz, 2000). Tian, Kanade, and Cohn (2001), for example, automatically detected 34 AUs and AU combinations in full-face frontal view images. More recently, investigators have made progress in the more demanding task of AU detection in nonposed facial images. Valstar, Gunes, and Pantic (2007) and Cohn and Schmidt (2004) automatically discriminated posed from nonposed, naturally occurring smiles. Ambadar, Cohn, and Reed (2009) used computer-vision-based measures to differentiate polite, happy, and embarrassed smiles. Dinges et al. (2005) detected stress from automatic measures of facial expression. 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, Mahoor, Chow, and Cohn (2009) demonstrated a pilot system for automatic measurement of smiles in mothers and infants during face-toface interaction. Whitehill, Littlewort, Fasel, Bartlett, and Movellan (2009) detected smile intensity in video from 5 participants while they individually watched a short video clip. At least one commercial product (FaceReader; Theuws, 2007) 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 nonposed facial actions in less constrained contexts. We used automated facial image analysis to provide the first test of automated FACS AU detection in a multiperson 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, because participants frequently drank beverages (a glass of juice) supplied by the experimenters. The present test focused on a critical pair of AUs: 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 et al. (1990) termed the *Duchenne smile*, which is thought to be a "true smile" and to 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, multiperson social interaction would mark an important step forward in the development of automatic approaches to coding facial expression.

METHOD

Digital video from 3 participants, who were participating in a larger study, was used for the present test. These participants were involved in a study (in progress) examining group formation processes (see Kirchner, Sayette, Cohn, Moreland, & Levine, 2006, for details of a study using this design). Although the experiment was advertised as a study examining the impact of alcohol on cognitive functioning, the group of participants used in the present FACS analyses had been assigned randomly to a no-alcohol control condition, in which they were told explicitly that they would receive—and they were indeed served—cranberry juice only. They drank juice throughout the observation period. All 3 individuals reported that they had not consumed alcohol or psychoactive drugs (except nicotine or caffeine) during the 24-h period leading up to 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 3 unacquainted "strangers," 4 people were invited to the laboratory (see Kirchner et al., 2006, for details of the overall procedures for a similar study). Participants were told that 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 were placed in different rooms. Then, they were casually introduced to each other one at a time while two researchers observed them for any signs of recognition. None showed any sign of recognition. Following initial greetings, they also were asked whether they had ever met the others (they reported that they had not).

Setting and Equipment. The 3 members of the group were escorted to the experimental room and were 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, participants were debriefed, paid, and permitted to leave.

The laboratory included a custom-designed video control system that permits synchronized video output for each participant, as well



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 active appearance model are superimposed on the source video.

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 participant was used in this report. An example is shown below.

Manual FACS Coding

For each participant, 6 min of continuous video from the middle of the observation period was selected for analysis. Two certified FACS coders independently coded AUs 6 and 12 and instances of occlusion from the digital video using Observer Video-Pro Software (Noldus, Trienes, Hendriksen, Jansen, & Jansen, 2000). Occlusion was defined as any obstruction of part of the face by a hand, glass, or other object or as a portion of face moving out of the field of view. Self-occlusion, caused by nonrigid 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 to later synchronize codes according to digital time stamp. Interobserver exact (30 f/sec) agreement was quantified using coefficient kappa, which is the proportion of agreement above what would be expected to occur by chance. Kappa coefficients were .68 and .78 for AU 6 and AU 12, respectively; it was .94 for occlusion.

Orientation to the camera was quantified automatically by using a nonrigid structure from motion algorithm as noted above. Head orientation is important, because the face looks different from different views and because 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 comprised three steps: (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), and (3) detect FACS AUs.

AAM. AAMs decouple shape and appearance of a face image. Given a predefined 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 key frames were manually labeled during a training phase. The remaining frames were automatically aligned using a gradient-descent AAM fit described in Matthews and Baker, 2004, and in Xiao, Baker, Matthews, and Kanade, 2004.

The *shape* **s** of an AAM is described by a 2-D triangulated mesh. In particular, the coordinates of the mesh vertices define the shape **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 **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 $\mathbf{p} = (p_1, \ldots, p_m)^T$ are the shape parameters (see Figure 2). Additionally, a global normalizing transformation (in this case, a geometric similarity transform) is applied to **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 to 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 nonrigid 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 present study, and for synthesizing new images (Theobald & Cohn, 2009).



Figure 2. An example of the computation of active appearance model (AAM) shape and appearance. The figure shows the mean and first two modes of variation of 2-D AAM shape (A–C) and appearance (D–F) variation, and the mean and first two modes of 3-D AAM shape. From "Real-Time Combined 2D + 3D Active Appearance Models," by J. Xiao, S. Baker, I. Matthews, and T. Kanade, 2004, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 535–542), Piscataway, NJ: IEEE Press. Copyright 2004 by IEEE. Reprinted with permission.

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 participants is necessary to ensure that the meaning of each of the coefficients is comparable across sessions. In the PCA dimensionality reduction step, 95% of the energy was retained, resulting in 10 principal components or shape eigenvectors.

AU Detection

AUs were detected using support vector machine classifiers (SVMs; Hsu, Chang, & Lin, 2005). SVMs attempt to find the hyperplane 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 AU (e.g., AU 6) and all instances of other AUs including neutral faces (i.e., AU 0 in FACS).

To maximize generalizability, we trained and tested the SVMs on independent data. For training, we used the RU–FACS (Frank, Movellan, Bartlett, & Littlewort, n.d.) 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-min 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 5 participants could not be analyzed due to image artifact. Thus, image data from 29 participants were used for training the classifiers. Classifiers then were tested on the independent participants from the present 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° from frontal view for pitch and 6.79° from frontal view for yaw. (Here and following, absolute values are reported for pitch and yaw.) For pitch, the 90th and 95th percentiles were 12.85° and 15.54° from frontal, respectively. For yaw, the corresponding values were 12.53° and 15.86°. Maximum pitch was 28.82°; maximum yaw was 73.37°.

Automatic AU 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 face tracking. (For video demos, see the supplemental materials.) The face image with tracked facial features appears in the bottom panel. Across the top are the similarity-transformed and piecewise-warped appearance. The former is the appearance after removal of rotation and translation; translation is variation due to change in horizontal and vertical motion and scale. In piecewise normalized appearance, variation due to out-of-plane head motion has been removed and, thus, stabilized for all but nonrigid motion (i.e., expression). To the right of the appearance are three representations of shape. The first is 2-D, the second is 3-D (a threequarter view of frontal pose), and the third is 3-D 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 itself may be an important nonverbal cue and, thus, useful in its own right to measure.

As noted above, classifiers were trained on video from an independent database (RU–FACS) and were tested on the participants from the experiment in progress. Automatically coded AU 6 and AU 12 for the present video were then compared frame-by-frame with manual FACS coding. Following previous literature (Ashraf et al., 2009; Pantic & Bartlett, 2007), we quantified accuracy using receiver–operating characteristic (ROC) curves. 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 .50 representing the expected value of random guessing.

ROC curves are used widely in signal detection, analysis of diagnostic systems, and machine learning (Fawcett, 2006). They are especially useful for skewed distributions, such as those for AUs and for unequal classification costs.



Figure 3. Screenshot of automated face tracking and action unit (AU) detection. Row 2 shows a participant's tracked face and frame-by-frame detection results for AU 6 and AU 12. Row 1 column A shows corresponding 2-D 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–E show 2-D and 3-D representations of the corresponding face shapes. Row 1 column F shows the estimated 3-D parameters (pitch, roll, and yaw). Please see the supplementary materials for video examples (the video examples are also available at www.pitt.edu/~jeffcohn/D102_G006A1_trim.mov and www.pitt.edu/~jeffcohn/D102_G014A1_trim.mov).

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 and then report results in relation to occlusion and nonrigid head motion.

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 (SE = .002, p < .0001); for AU 12, A' was .88 (SE = .004, p < .0001). For comparison with interobserver agreement, we calculated 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).

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 nonrigid 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° variation of pitch and yaw. AU 6 occurred through 20° pitch and through 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 $\pm 0^{\circ}-20^{\circ}$; results for intervals outside of this range should be interpreted with caution.

With respect to pitch, A' for AU 6 was stable (M = .96) through $\pm 15^{\circ}$ pitch and decreased to .89 at $\pm 15^{\circ}-20^{\circ}$ pitch (Table 2). For AU 12, A' was stable through $\pm 20^{\circ}$



Figure 4. The receiver-operating characteristic (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 .94 (SE = .002, p < .0001) and .85 (SE = .004, p < .0001).

pitch (M = .88) and then decreased to .70 over the interval between $\pm 20^{\circ}$ -25° pitch. For pitch variation greater than $\pm 25^{\circ}$, there was a small increase for AU 12.

With respect to yaw, AU 6 was stable (M = .97) through $\pm 20^{\circ}$ and then decreased to .89 and .83 at $\pm 20^{\circ}-25^{\circ}$ and $\pm 25^{\circ}$ or greater. For AU 12, A' was stable (M = .90) through $\pm 15^{\circ}$ yaw and then decreased gradually to .78 at yaw greater than $\pm 25^{\circ}$. Overall, the general pattern for both AU 6 and AU 12 for pitch and yaw was for A' to remain high for intervals through $\pm 15^{\circ}$ or 20° and then to decrease. For no intervals did A' go below .70, and, in all but one case, it was at least .78.

DISCUSSION

The present study provides initial support for the use of automatic facial image analysis for the detection of spontaneous facial expressions arising during unstructured social interaction. Specifically, when automated facial image analysis was used, two AUs critical to positive emotion, AUs 6 and 12, were automatically detected with high reliability as compared with that for 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

Table 1 Selected Corresponding True and False Positive Rates for AU 6 and AU 12

Faise Fositive Rates for the 6 and the 12						
AU	True Positive Rate	False Positive Rate				
AU 6	90%	6%				
	80%	3%				
AU 12	90%	37%				
	80%	15%				

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 ± 0.5 –1.0 sec (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 that there 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 AUs in the mouth region when it was present. Although this instruction often proves too costly in lost information, criteria for coding AUs 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 often. 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 15% 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 $\pm 15^{\circ}$ -20°. Beyond that, accuracy decreased, although it still remained 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

Table 2									
Area Under the Receiver-Operating Characteristic (ROC) Curve in Relation to Partial Occlusion, Pitch, and Yaw									
		Area Under the ROC							
	Percentage of Time	AU 6		AU 12					
Condition		M	SE	M	SE				
Occlusion									
Absent	89.22	.97	.002	.90	.002				
Present	10.78	.91	.012	.75	.012				
Pitch (absolute value)									
0°-5°	41.08	.96	.004	.86	.004				
5°-10°	36.58	.97	.004	.90	.003				
10°-15°	16.81	.96	.006	.90	.005				
15°-20°	4.50	.89	.061	.87	.020				
20°-25°	0.71	NA		.70	.037				
>25°	0.32	NA		.94	.027				
Yaw (absolute value)									
0°-5°	42.85	.96	.004	.87	.004				
5°-10°	34.75	.97	.004	.91	.004				
10°-15°	16.69	.97	.004	.91	.005				
15°-20°	3.37	.98	.004	.84	.013				
20°-25°	1.10	.89	.036	.80	.038				
>25°	1.23	.83	.064	.78	.030				

Note—All ps < .0001. NA, no occurrences of the AU for this range.

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. In our experience, glasses interfere with tracking only if the lenses are highly reflective. When that occurs, eye closure may be difficult to detect. In the present study, all participants were men; results might have differed for women, although that seems unlikely. Although 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, insofar 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 detected more easily. With respect to pose and occlusion, we found that AU detection was relatively robust. Future work should examine these issues and include a larger number of AUs and emotions.

Despite the limitations of the present 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, 2009), smiling (Ambadar et al., 2009; Cohn & Schmidt, 2004; Schmidt, Liu, & 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 present article is the first to use automated facial image analysis in relatively unconstrained smallgroup 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 soon to significantly boost research productivity and open new areas of investigation.

AUTHOR NOTE

This research was supported in part by NIH Grants MH 051435 and AA015773. Tomas Simon Kreuz, Kasey Griffin, Iain Matthews, Nicole Ridgeway, and John Dimoff provided technical assistance. Address correspondence to J. F. Cohn, Department of Psychology, University of Pittsburgh, 3137 Sennott Square, 210 S. Bouquet St., Pittsburgh, PA 15260 (e-mail: jeffcohn@cs.cmu.edu).

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. doi:10.1007/s10919-008-0059-5
- 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 & Vision Computing*, 27, 1788-1796.
- BARTLETT, M. S., HAGER, J. C., EKMAN, P., & 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 & Social Psychology*, **50**, 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 University Press.
- COHN, J. F., LUCEY, P., AMBADAR, Z., LUCEY, S., SOLOMON, P., & PRKACHIN, K. (2009, May). *The face of pain: Automatic recognition and analysis using computer vision and machine learning*. Paper presented at the American Psychological Society Convention, San Francisco.
- COHN, J. F., & SCHMIDT, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution, & 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 & 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. (1998). The expression of the emotions in man and animals

(3rd ed.). New York: Oxford University Press. (Original edition published 1872)

- DINGES, D. F., RIDER, R. L., DORRIAN, J., MCGLINCHEY, E. L., ROG-ERS, N. L., CIZMAN, Z., ET AL. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. Aviation, Space, & Environmental Medicine, 76, 172-182.
- EKMAN, P. (1992). An argument for basic emotions. *Cognition & Emotion*, **6**, 169-200.
- EKMAN, P. (1993). Facial expression and emotion. American Psychologist, 48, 384-392.
- EKMAN, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Review*, **115**, 268-287. doi:10.1037/0033-2909.115.2.268
- EKMAN, P., DAVIDSON, R. J., & FRIESEN, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Per*sonality & Social Psychology, 58, 342-353.
- EKMAN, P., FRIESEN, W. V., & HAGER, J. C. (EDS.) (2002). Facial action coding system. Salt Lake City, UT: Research Nexus, Network Research Information.
- EKMAN, P., & ROSENBERG, E. (EDS.) (2005). What the face reveals (2nd ed.). New York: Oxford University Press.
- ESSA, I., & PENTLAND, A. (1997). Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 7, 757-763.
- FAWCETT, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27, 861-874.
- FRANK, M., MOVELLAN, J., BARTLETT, M. S., & LITTLEWORT, G. (n.d.). RU–FACS-1 database. San Diego: U.C. San Diego, Machine Perception Laboratory.
- FRIDLUND, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego: Academic Press.
- GRIFFIN, K. M., & SAYETTE, M. A. (2008). Facial reactions to smoking cues relate to ambivalence about smoking. *Psychology of Addictive Behaviors*, 22, 551-556.
- HATFIELD, E., CACIOPPO, J. T., & RAPSON, R. L. (1992). Primitive emotional contagion. In M. E. Clark (Ed.), *Emotion and social behavior* (Vol. 14, pp. 151-177). Thousand Oaks, CA: Sage.
- 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, 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). Thousand Oaks, CA: Sage.
- LEVENSON, R. W., EKMAN, P., & FRIESEN, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, **27**, 363-384.
- LITTLEWORT, G. C., BARTLETT, M. S., FASEL, I. R., SUSSKIND, J., & MOVELLAN, J. (2006). Dynamics of facial expression extracted automatically from video. *Image & Vision Computing*, 24, 615-625.
- LITTLEWORT, G. C., BARTLETT, M. S., & LEE, K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. *Image & Vision Computing*, 27, 1797-1803.
- MALATESTA, C. Z., CULVER, C., TESMAN, J. R., & SHEPARD, B. (1989). The development of emotion expression during the first two years of life. *Monographs of the Society for Research in Child Development*, 54, 1-104.
- MATTHEWS, I., & BAKER, S. (2004). Active appearance models revisited. International Journal of Computer Vision, 60, 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**, 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, 285-305.

- NOLDUS, L. P.J. J., TRIENES, R. J. H., HENDRIKSEN, 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, & 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.
- PANTIC, M., & ROTHKRANTZ, L. J. M. (2000). Expert system for automatic analysis of facial expression. *Image & 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**, 804-809.
- RUSSELL, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of cross-cultural studies. *Psychological Bulletin*, **115**, 102-141.
- RYAN, A. H., COHN, J. F., & HAMERSKI, R. (2009). Automated facial expression recognition system. In *Proceedings of the IEEE International Carnahan Conference on Security Technology* (pp. 172-177). Piscataway, NJ: IEEE Press.
- SAYETTE, M. A., COHN, J. F., WERTZ, J. M., PERROTT, M. A., & PAR-ROTT, D. J. (2001). A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior*, 25, 167-185.
- SCHMIDT, K. L., & COHN, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. *Yearbook of Physical Anthropology*, **116**, 3-24.
- SCHMIDT, K. L., LIU, Y., & COHN, J. F. (2006). The role of structural facial asymmetry in asymmetry of peak facial expressions. *Laterality*, **11**, 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). New York: Oxford University Press.
- THEUWS, H. (2007). FaceReader [Computer software]. Wageningen, The Netherlands: Noldus.
- TIAN, Y. L., KANADE, T., & COHN, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis* & Machine Intelligence, 23, 97-116.
- TOMKINS, S. S. (1962). *Affect, imagery, consciousness*. New York: Springer.
- TREMEAU, F., MALASPINA, D., DUVAL, F., CORRÊA, 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, 92-101. doi:10.1176/appi.ajp.162.1.92
- VALSTAR, M. F., GUNES, H., & PANTIC, M. (2007). How to distinguish posed from spontaneous smiles using geometric features. *Proceedings* of the Ninth International ACM Conference on Multimodal Interfaces (pp. 38-45). Nagoya, Japan.
- WANG, P., BARRETT, F., MARTIN, E., MILONOVA, M., GUR, 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., & MOVEL-LAN, J. (2009). Toward practical smile detection. *IEEE Transactions* on Pattern Analysis & Machine Intelligence, **31**, 2106-2111.
- XIAO, J., BAKER, S., MATTHEWS, I., & KANADE, T. (2004). Real-time combined 2D+3D active appearance models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition* (Vol. 2, pp. 535-542). Piscataway, NJ: IEEE Press.

SUPPLEMENTAL MATERIALS

Video demos of automated face tracking and action unit detection with our software may be downloaded from http://brm.psychonomic -journals.org/content/supplemental.

(Manuscript received November 25, 2009; revision accepted for publication June 2, 2010.)