Report

Automatic Decoding of Facial Movements Reveals Deceptive Pain Expressions

Marian Stewart Bartlett,[1,*] Gwen C. Littlewort,[1]
Mark G. Frank,[2] and Kang Lee[3,*]
[1]Institute for Neural Computation, University of California,
9500 Gilman Drive, MC 0440, La Jolla, San Diego,
CA 92093-0440, USA
[2]Department of Communication, University at Buffalo,
State University of New York, Baldy Hall, Buffalo,
NY 14260-1020, USA
[3]Dr. Erick Jackman Institute of Child Study, University of
Toronto, 45 Walmer Road, Toronto, ON M5R 2X2, Canada

Summary

In highly social species such as humans, faces have evolved to convey rich information for social interaction, including expressions of emotions and pain [1–3]. Two motor pathways control facial movement [4–7]: a subcortical extrapyramidal motor system drives spontaneous facial expressions of felt emotions, and a cortical pyramidal motor system controls voluntary facial expressions. The pyramidal system enables humans to simulate facial expressions of emotions not actually experienced. Their simulation is so successful that they can deceive most observers [8–11]. However, machine vision may be able to distinguish deceptive facial signals from genuine facial signals by identifying the subtle differences between pyramidally and extrapyramidally driven movements. Here, we show that human observers could not discriminate real expressions of pain from faked expressions of pain better than chance, and after training human observers, we improved accuracy to a modest 55%. However, a computer vision system that automatically measures facial movements and performs pattern recognition on those movements attained 85% accuracy. The machine system's superiority is attributable to its ability to differentiate the dynamics of genuine expressions from faked expressions. Thus, by revealing the dynamics of facial action through machine vision systems, our approach has the potential to elucidate behavioral fingerprints of neural control systems involved in emotional signaling.

Results

Human Experiments

To test both human observers' and our computer vision system's abilities to discriminate real versus faked emotional expressions, we created two sets of videos. One set contained faces of individuals experiencing genuine pain as induced through a "cold pressor" method [12], whereas the other set contained faces of the same individuals pretending to be in pain. Expressions of pain were chosen because pain is a universally experienced emotive physiological state [12–15]. Additionally, both genuine and faked expressions of pain can be readily elicited using the cold pressor method, which is a

routine experimental procedure used to induce pain for research purposes [12]. Stimulus subjects either experienced genuine pain while submerging their arm in ice water (5°C) for 1 min or were instructed to fake pain while submerging their arm in warm water (20°C) for 1 min. Facial expressions in the two conditions were video recorded.

In experiment 1, we showed 170 human observers video clips of the stimulus subjects individually in a randomized order. The observers judged whether the expression shown in the video clip was real or faked. The observers distinguished genuine pain from faked pain at rates no greater than guessing (accuracy = 51.9%; SD = 14.6; chance accuracy = 50%).

Experiment 2 examined whether training could improve human observers' detection accuracy. We showed 35 new participants 24 video pairs in a training procedure to match the cross-validation training of the computer vision system described below. Observers were presented with two videos of the same person shown sequentially. In one video the individual was expressing genuine pain, and in the other the individual was expressing faked pain. Observers then judged which video of the pair showed the individual expressing genuine pain and which video showed the individual expressing faked pain. Observers received feedback about their accuracy immediately. After being trained on all 24 pairs, participants saw, in random order, 20 new videos of 20 new stimulus subjects for the test phase. Half of these new videos displayed stimulus subjects with faked pain, and the other half displayed stimulus subjects with real pain. Observers judged whether the expression shown in each of the 20 videos was real or faked, with no feedback offered. This test phase assessed whether human observers could generalize what they had learned to detect new exemplars of genuine or faked pain expressions. In the first third of the training trials (eight trials), the accuracy was 49.6% (SD = 11%). The accuracy rate for the last third of the training trials was 58.6% (SD = 8.5%), which was significantly above chance (t[34] = 2.45, p < 0.01) and showed a significant, albeit small, improvement over earlier training trial blocks (t[34] = 2.22, p < 0.02). In the test phase, detection accuracy remained just above chance level at 54.6% (SD = 15.5%) (t[34] = 1.76, p < 0.05). Thus, results from the two human experiments together suggest that human observers are generally unsuccessful at detecting differences between real and faked pain. There was a small improvement with training, but performance remained below 60%. This result is highly consistent with prior research [14].

Machine Learning

We then presented these videos to a computer vision system called the Computer Expression Recognition Toolbox (CERT). CERT is a fully automated system that analyzes facial muscle movements from video in real time [16]. It automatically detects frontal faces in video and codes each frame with respect to a set of continuous dimensions, including facial muscular actions from the Facial Action Coding System (FACS) [17]. FACS is a system for objectively scoring facial expressions in terms of elemental facial movements, called action units (AUs). This makes FACS fully comprehensive, given its basis in functional neuroanatomical actions. CERT

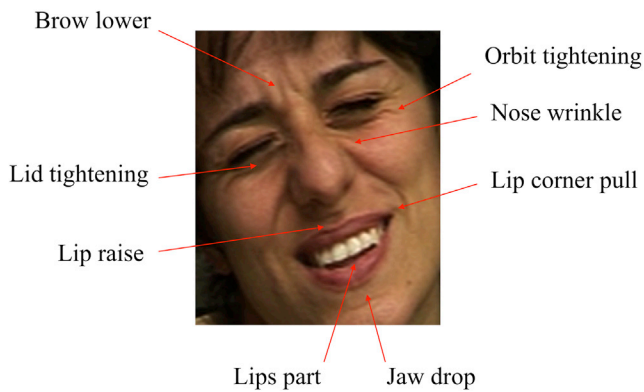*Correspondence: mbartlett@ucsd.edu (M.S.B.), kang.lee@utoronto.ca (K.L.)

Figure 1. Example of Facial Action Coding
Here, a facial expression of pain is coded in terms of eight component facial actions based on FACS.

can identify 20 out of a complete set of 46 AUs, each with its own movement and appearance characteristics (Figure 1). FACS was originally developed for manual coding by human experts. Manual coding is laborious and can take up to 3 hr to manually code 1 min of behavior, but CERT instantaneously outputs facial-movement information in real time (i.e., every 1/30th s). Furthermore, the frame-by-frame CERT output provides information on facial-expression intensity and dynamics at temporal resolutions that were previously impractical with human coding. CERT was developed at the University of California, San Diego, and is presently available at Emotient.

We used a pattern-recognition approach to assess CERT's ability to detect falsified pain expressions (Figure 2). The 60 s videos were input into the computer vision system one at a time. A set of dynamic descriptors was extracted from the output for each of the 20 AUs. The descriptors consisted of "bags of temporal features" (Figure 3). See Supplemental Experimental Procedures. Two sets of descriptors were employed: one set that describes the dynamics of facial-movement events (event descriptors) and another set that describes the intervals between events (interval descriptors). Our methods for constructing bags of temporal features represent a novel approach that can be applied generally to describe signal dynamics for other pattern-recognition problems. Bags of temporal features build upon the concept of bags of features to provide sensitivity to some aspects of the signal (such as edges or peaks at different scales) while providing invariance to aspects of the signal across which we wish to generalize, such as the specific location or time point at which the signal occurs.

Next, a classifier was trained to discriminate real pain from faked pain using these descriptors. The classifier was a support vector machine (SVM). The SVM combined information from multiple AUs. This was accomplished with a sequential feature selection procedure. In this procedure, the model began with the AU that gave the best individual classification accuracy. We then added to the model the AU that gave the best performance when combined with the previously selected AUs. This process was repeated until detection performance stopped improving. The selected AUs were the event descriptors for brow lower (AU 4), lip raise (AU 10), lip corner depress (AU 15), and lower lid tighten (AU 7) and the combined event and interval descriptors for mouth opening

(AU 26). Because of the post hoc nature of sequential feature selection, this model was tested with double cross-validation (see Supplemental Experimental Procedures). Like the human judges, the computer system was tested on each video individually, without the ability to compare real and faked data from a given stimulus subject. The system achieved detection accuracy as measured by area under the ROC curve (A') of 0.91 for detection of real versus faked pain (see Supplemental Experimental Procedures regarding A'). In terms of percent correct (PC), this corresponds to 85% correct classification at equal error rate and is significantly better than chance ($Z = 3.7$, $p < 0.0001$). Most importantly, the system outperforms human judges even after training ($Z = 3.3$, $p < 0.001$, two-tailed Z test).

In-Depth Analysis of Computer Vision Accuracy: Static versus Dynamic
To test the contribution of temporal information to classification accuracy, we examined the effect of the size of the time window on classification performance. We retrained the above five-AU system on sliding time windows of varying length. Classification performance was tested for temporal integration window sizes ranging from 3.3 s (100 frames) to 60 s. Figure 4 shows that the classification accuracy increased systematically with the size of the time window.

To ascertain whether this time-related increase in accuracy was due to the nature of the temporal dynamics rather than the simple accumulation of information, we had an SVM classifier with the five selected AUs trained on the same frame-by-frame CERT outputs but with the order of the frames randomized. It did not perform at above-chance levels ($A' = 0.58$, PC = 56%, $Z = 0.6$, $p = 0.27$). A static system that did not extract temporal features but instead measured facial actions in each video frame, histogrammed them, and then passed them to an SVM was also tested. The static system was able to detect faked expressions of pain marginally above chance levels ($A' = 0.72$, PC = 66%, $Z = 1.6$, $p < 0.05$). However, this 66% accuracy rate for the static system was significantly poorer than the accuracy rate for the dynamic system reported above (85%; $Z = 2.2$, $p < 0.05$). These findings show that dynamic temporal information contributes substantially to the system's performance.

In-Depth Analysis of Computer Vision Accuracy: Most Important AUs
Next, we attempted to identify the key features for differentiating real pain from faked pain. The feature selection procedure identified which facial actions, in combination, provided the most information for detecting faked pain. We then examined which individual action units can reliably differentiate real pain from faked pain. Twenty SVM classifiers, one per action unit, were individually trained to differentiate genuine pain from faked pain using the dynamic descriptors of a single action unit. The most informative AU for differentiating real pain from faked pain was mouth opening (AU 26) using both the event and interval descriptors ($A' = 0.85$, PC = 72%, $Z = 2.3$, $p < 0.01$). Three other AUs individually differentiated genuine pain from faked pain at above-chance levels using just the event descriptors: lip raise (AU 10), lip press (AU 24), and brow lower (corrugator muscle; AU 4) ($A' = 0.75$, 0.73, and 0.73, respectively; PC = 68%, 66%, and 66%, respectively; $Z = 1.8$, 1.7, and 1.7, respectively; $p < 0.05$). All other AUs were at chance levels.
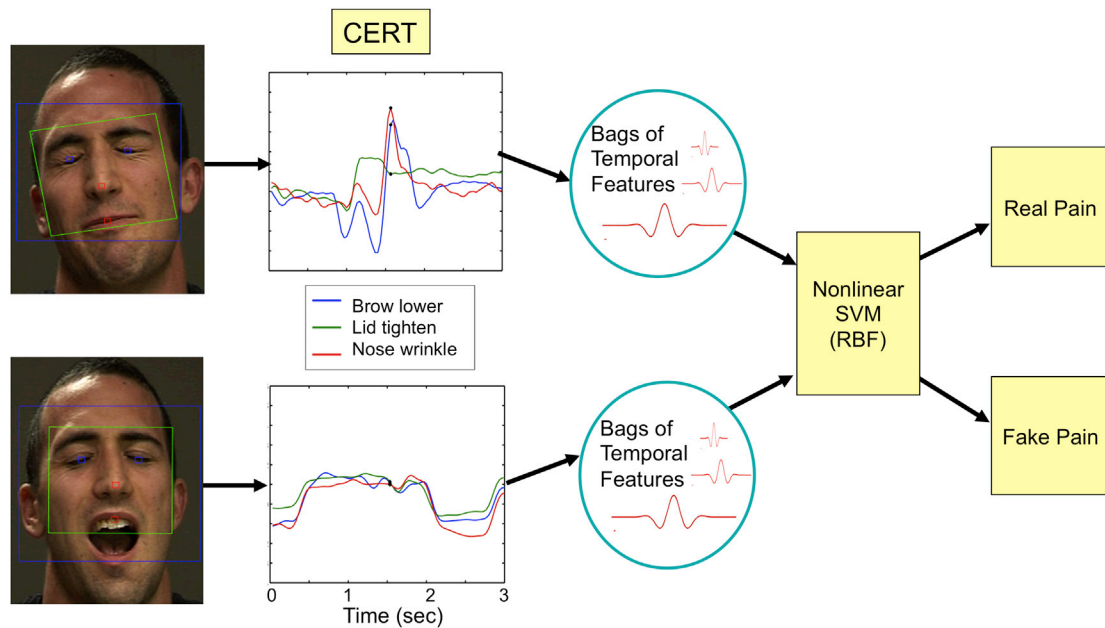
Figure 2. System Overview

Face video is processed by the computer vision system, CERT, to measure the magnitude of 20 facial actions over time. The CERT output at the top is a sample of real pain, whereas the sample on the bottom shows the same three actions for faked pain from the same subject. Note that these facial actions are present in both real and faked pain, but their dynamics differ. Expression dynamics were measured with a bank of eight temporal Gabor filters and expressed in terms of bags of temporal features. These measures were passed to a machine learning system (nonlinear SVM) to classify real versus faked pain. The classification parameters were learned from the 24 1 min examples of real and faked pain.

Mouth opening was the single most informative feature for discriminating genuine expressions of pain from faked expressions of pain. This feature contained dynamic information about mouth openings as well as the intervals between mouth openings. This finding led us to explore how mouth-opening dynamics differ in genuine versus faked expressions of pain using some simple statistics on the unfiltered CERT output. First, there was no difference in the overall mean CERT output for mouth opening between real versus faked expressions of pain ($t[24] = 0.006$, $p = 0.99$), implying that the crucial information was in the dynamics. A measure of the duration of mouth openings (τ), as well as an estimate of the temporal intervals between mouth openings, was then extracted (see Supplemental Experimental Procedures). There was a difference in the mean duration of mouth openings for genuine and faked expressions, with faked expressions being 5.4 frames shorter than genuine expressions on average and with the interval between mouth openings lasting 11.5 frames less for faked expressions on average ($t[24] = 2.23$ and $t[24] = 2.19$, respectively; both $p < 0.05$). The variance of τ was then computed for faked and genuine expressions. A within-subjects comparison showed that the variance of τ was 55% less for faked than for genuine expressions of pain ($t[24] = 2.7$, $p < 0.01$). Similarly, the variance of the interval length between mouth openings was 56% less for faked than for genuine expressions of pain ($t[24] = 2.11$, $p < 0.05$).

Discussion

We show for the first time that a fully automated computer vision system can be trained to detect a deceptive emotional physiological state—faked expressions of pain—from facial cues alone. The rate of accurate discrimination by the computer vision system was 85%. This is far superior to the accuracy of human observers, regardless of whether they have received training, which is consistently below 60% accuracy. This is a significant milestone for machine vision systems [18] because although computers have long outperformed humans at logic processes (such as playing chess), they have significantly underperformed compared to humans at perceptual processes, rarely reaching even the level of a human child.

Furthermore, our computer vision approach has led to the discovery of new information about facial behavior in falsified pain expressions. The single most predictive feature of falsified expressions of pain is the dynamics of the mouth opening, which alone could differentiate genuine expressions of pain from deceptive expressions of pain at a detection rate of 72%. Faked expressions were associated with a reduction in variance in terms of both the duration of mouth openings and the duration of the interval between mouth openings. In other words, the action was repeated at intervals that were too regular. The critical feature for faked pain may be this over-regularity of the dynamics of the mouth-opening action. Further investigations will explore whether over-regularity is a general feature of faked expressions.

Our findings further support previous research on human facial expressions that has shown that the dynamics of expression are important distinguishing characteristics of emotional expressions, such as a genuine smile versus a faked smile [8]. This difference stems from the fact that expressions are mediated by two distinct neural systems, each one originating in a different area of the brain [4–7]. A genuinely felt or experienced emotion originates in the subcortical areas of the brain and is involuntarily propelled onto the face via the extrapyramidal motor system [4–7]. In contrast, posed or faked expressions originate in the cortical motor strip and are voluntarily
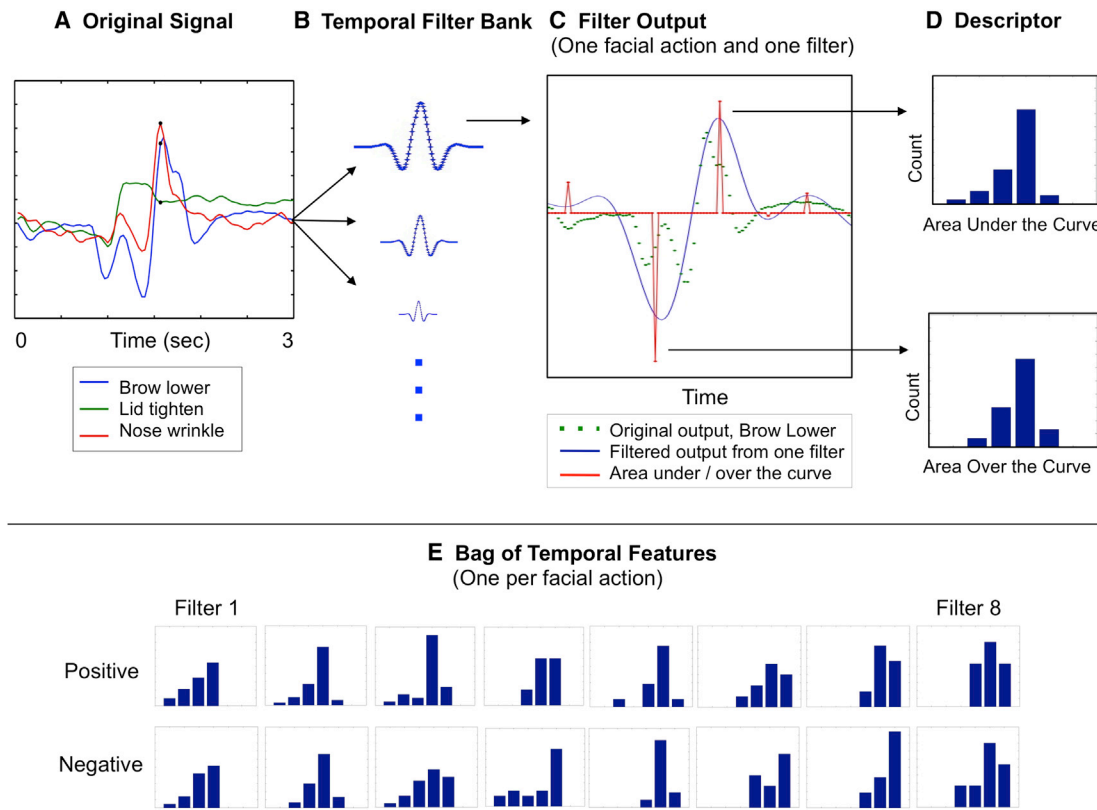
A Original Signal
B Temporal Filter Bank
C Filter Output
(One facial action and one filter)
D Descriptor

0   Time (sec)   3

- Brow lower
- Lid tighten
- Nose wrinkle

Time

- Original output, Brow Lower
- Filtered output from one filter
- Area under / over the curve

Count
Area Under the Curve

Count
Area Over the Curve

E Bag of Temporal Features
(One per facial action)

Filter 1   Filter 8

Positive

Negative

Figure 3. Bags of Temporal Features

Here, we illustrate an exemplar of one stimulus as it is processed by each step.

(A) Sample CERT signals from one subject (black circles indicate the time point of the face image shown in Figure 2). Three seconds of data are illustrated, but processing is performed on the full 60 s of video.

(B) The CERT signals were filtered by temporal Gabor filters at eight frequency bands.

(C) Filter outputs for one facial action (brow lower) and one temporal frequency band (the highest frequency).

(D) Zero crossings are detected, and area under the curve and area over the curve are calculated. The descriptor consists of histograms of area under the curve for positive regions and separate histograms for area over the curve for negative regions. (Negative output is where evidence indicates absence of the facial action.)

(E) Full bag of temporal features for one action (brow lower). Consists of eight pairs of histograms, one per filter.

expressed in the face via the pyramidal motor system. Research documenting these differences was sufficiently reliable to become the primary diagnostic criteria for certain brain lesions prior to modern imaging methods (e.g., [4, 6, 7]). These two systems may correspond to the distinction between biologically driven versus socially learned facial behaviors [8]. The facial expressions mediated by these two pathways have been shown to differ in some dynamic properties. Extrapyramidal motor system-based expressions have been associated with synchronized, smooth, symmetrical, and ballistic-like facial muscle movements, whereas pyramidal motor system-based expressions are subject to volitional real-time control and tend to be less smooth, less synchronized, and less symmetric [3]. Accordingly, smiles that were spontaneously generated have been shown to have smoother dynamics, as well as faster onset and offset velocity, than smiles that are posed or faked [3, 8]. Here, we show a new difference in variance between the two systems. Pyramidally driven expression of falsified pain showed a reduced variance in the timing of mouth openings relative to the spontaneous expressions of pain driven by the extrapyramidal system. A survey study revealed that humans have an intuitive knowledge of differences between controlled and automatic responses to pain [15].

However, our findings show that despite this understanding, people could not detect differences between controlled and automatic facial responses when presented with them visually.

In highly social species such as humans and other primates, the face has evolved to convey a rich array of information for social interaction. Although facial expressions are mainly evolved as cooperative social signals to communicate one's genuinely felt emotions to others, and hence behavioral intentions [1], sometimes individuals may wish to control their expressions to mislead others. Indeed, deceptions are a part of everyday life [8, 9, 19], and there are considerable adaptive advantages, including social acceptance, to deliberately manipulating, suppressing, and dissembling emotional expressions [9]. Such voluntary facial control may have been refined (for adaptive purposes or for the purposes of being polite and facilitating interaction) so much that it made it very difficult for observers to discern honest signals from controlled or falsified ones. In studies of deception, untrained human judges are typically only accurate at or near chance levels when detecting deceptive facial behaviors [11]. This inaccuracy persists despite the fact that (albeit imperfect) diagnostic signals exist [12]. In some domains, genuine and faked expressions of emotion have shown not only morphological differences but
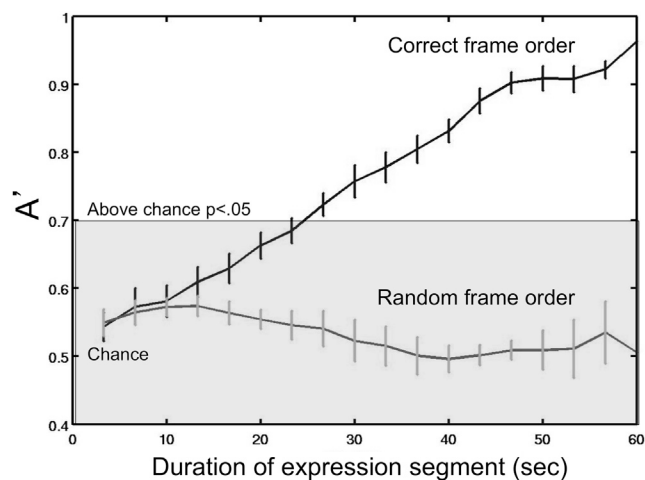
Figure 4. Contribution of Temporal Information

Classification performance (A′) is shown for temporal integration window sizes ranging from 10 s to 60 s. Samples were pooled across temporal position for training an SVM. The region above the shaded region is statistically significant at the p < 0.05 level. Error bars show 1 SE of the mean.

also dynamic differences. Although human judges were better than chance at detecting these morphological markers, they were unable to detect spatiotemporal dynamic markers [8].

Specifically, with regard to pain, lay adults and even experienced physicians cannot reliably differentiate real expressions of pain from faked expressions of pain [13, 14, 20, 21] (D.M. Siaw-Asamoah, 2011, North American Primary Care Physicians Research Group, conference). As shown in experiment 2 and by others [13], immediate feedback might enable perceptual learning and improve detection accuracy to above-chance levels. However, accuracy remains modest. Previous research using a laborious manual coding method has shown that there is no telltale facial action that can indicate faked pain by its presence or absence because real and faked expressions of pain include the same set of facial actions. However, these earlier studies hinted at differences in the dynamics of facial expression [14]. In the current study, the computer vision system was able to analyze facial-expression dynamics at a much higher temporal resolution and with richer description than was feasible with manual coding methods. Thus, it revealed aspects of a pain expression that have been previously unavailable to observers.

Taken together, our findings suggest that in spite of the pyramidal motor system's sophisticated voluntary control over facial expressions, its control is imperfect; the system cannot fully replicate the genuine expressions of pain driven by the extrapyramidal motor system, particularly in their dynamics. Thus, our findings support the hypothesis that information that can differentiate experienced spontaneous expressions of emotion driven by the extrapyramidal motor system from posed or falsified expressions controlled by the pyramidal motor system exists, particularly in facial dynamics [4]. Although the present study addressed one psychophysiological state—pain—, the approach presented here may be generalizable to the comparison of other genuine and faked emotional states, which may differentially activate the cortical and subcortical facial motor pathways. Thus, our automated facial-movement coding system provides a new paradigm for the study of facial dynamics and has the potential to

elucidate behavioral fingerprints of neural control systems involved in emotional signaling.

There are some practical implications of the present findings. Falsified pain can be a lie told by patients to their physicians so that they can commit insurance fraud or receive unneeded prescription narcotics. Some healthcare professionals perceive such lies to be common (D.M. Siaw-Asamoah, 2011, North American Primary Care Physicians Research Group, conference), whereas others perceive them to be relatively rare. Our findings suggest that it might be possible to train physicians to specifically attend to mouth-opening dynamics to improve physicians' ability to differentiate real pain from faked pain. In addition to detecting pain malingering, our computer vision approach may be used to detect other real-world deceptive actions in the realm of homeland security, psychopathology, job screening, medicine, and law. Like pain, these scenarios also generate strong emotions, along with attempts to minimize, mask, and fake such emotions [19], which may involve dual control of the face. In addition, our computer vision system can be applied to detect states in which the human face may provide important clues about health, physiology, emotion, or thought, such as drivers' expressions of sleepiness [16, 18] and students' expressions of attention and comprehension of lectures [18], or to track response to treatment of affective disorders [22].

Several limitations to our findings should be noted. First, the pain manipulation task (cold pressor) is a good, but not perfect, analog to all varieties of clinical pain. Thus, future research will be needed in order to explore expression of other types of pain (e.g., back pain) collected in clinical settings. Second, pain is a complicated concept involving attitudes, movements elsewhere in the body, and so forth [23]. This paper addresses just one element (facial expression) and shows proof of principle. Future studies will need to address the other elements of the pain phenomenon.

Main Findings and Implications
In summary, the present study demonstrated the effectiveness of a computer vision and pattern-recognition system for detecting faked pain from genuine expressions. The computer system outperformed human observers, achieving significantly better accuracy. Moreover, the automated system revealed new information about facial dynamics that differentiate real expressions of pain from faked expressions of pain. Our findings demonstrate the ability of the computer system to extract information from spatiotemporal facial-expression signals that humans either cannot or do not extract. Automated systems such as CERT may bring about a paradigm shift by analyzing facial behavior at temporal resolutions previously not feasible with manual coding methods. This novel approach has succeeded in illuminating basic questions pertaining to the many social situations in which the behavioral fingerprint of neural control systems may be relevant.

References

1. Darwin, C. (1872). The Expression of the Emotions in Man and Animals (London: Murray).
2. Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In Psychological Methods in Criminal Investigation and Evidence, D.C. Raskin, ed. (New York: Springer Publishing Company), pp. 297–332.
3. Frank, M.G., Ekman, P., and Friesen, W.V. (1993). Behavioral markers and recognizability of the smile of enjoyment. J. Pers. Soc. Psychol. *64*, 83–93.
4. Rinn, W.E. (1984). The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. Psychol. Bull. *95*, 52–77.
5. Kunz, M., Chen, J.I., Lautenbacher, S., Vachon-Presseau, E., and Rainville, P. (2011). Cerebral regulation of facial expressions of pain. J. Neurosci. *31*, 8730–8738.
6. Brodal, A. (1981). Neurological Anatomy: In Relation to Clinical Medicine (New York: Oxford University Press).
7. Tschiassny, K. (1953). Eight syndromes of facial paralysis and their significance in locating the lesion. Ann. Otol. Rhinol. Laryngol. *62*, 677–691.
8. Ekman, P., and Friesen, W. (1982). Felt, false, and miserable smiles. J. Nonverbal Behav. *6*, 238–252.
9. DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., and Epstein, J.A. (1996). Lying in everyday life. J. Pers. Soc. Psychol. *70*, 979–995.
10. Bond, C.F., Jr., and DePaulo, B.M. (2006). Accuracy of deception judgments. Pers. Soc. Psychol. Rev. *10*, 214–234.
11. Frank, M.G., and Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. J. Pers. Soc. Psychol. *72*, 1429–1439.
12. Hadjistavropoulos, H.D., Craig, K.D., Hadjistavropoulos, T., and Poole, G.D. (1996). Subjective judgments of deception in pain expression: accuracy and errors. Pain *65*, 251–258.
13. Hill, M.L., and Craig, K.D. (2004). Detecting deception in facial expressions of pain: accuracy and training. Clin. J. Pain *20*, 415–422.
14. Hill, M.L., and Craig, K.D. (2002). Detecting deception in pain expressions: the structure of genuine and deceptive facial displays. Pain *98*, 135–144.
15. McCrystal, K.N., Craig, K.D., Versloot, J., Fashler, S.R., and Jones, D.N. (2011). Perceiving pain in others: validation of a dual processing model. Pain *152*, 1083–1089.
16. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J. (2005). Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior. IEEE International Conference on Computer Vision and Pattern Recognition, 568–573.
17. Ekman, P., and Friesen, W. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement (Palo Alto, CA: Consulting Psychologists Press).
18. Bartlett, M., and Whitehill, J. (2010). Automated facial expression measurement: recent applications to basic research in human behavior, learning, and education. In Handbook of Face Perception, A.J. Calder, G. Rhodes, J.V. Haxby, and M.H. Johnson, eds. (New York: Oxford University Press).
19. Ekman, P. (2001). Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (New York: W.W. Norton).
20. Jung, B., and Reidenberg, M.M. (2007). Physicians being deceived. Pain Med. *8*, 433–437.
21. Poole, G.D., and Craig, K.D. (1992). Judgments of genuine, suppressed, and faked facial expressions of pain. J. Pers. Soc. Psychol. *63*, 797–805.
22. Ekman, P., Matsumoto, D., and Friesen, W. (1997). Facial expression in affective disorders. In What the Face Reveals, P. Ekman and E.L. Rosenberg, eds. (New York: Oxford University Press), pp. 331–341.
23. Turk, D.C., and Okifuji, A. (1999). Assessment of patients' reporting of pain: an integrated perspective. Lancet *353*, 1784–1788.