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.

THE TIMING OF FACIAL MOTION IN POSED AND SPONTANEOUS SMILES

J.F. COHN* and K.L.SCHMIDT
University of Pittsburgh
Department of Psychology
4327 Sennott Square, 210 South Bouquet Street
Pittsburgh, PA 15260, USA
**jeffcohn@pitt.edu*
kschmidt@pitt.edu

Revised 19 March 2004

Almost all work in automatic facial expression analysis has focused on recognition of prototypic expressions rather than dynamic changes in appearance over time. To investigate the relative contribution of dynamic features to expression recognition, we used automatic feature tracking to measure the relation between amplitude and duration of smile onsets in spontaneous and deliberate smiles of 81 young adults of Euro- and African-American background. Spontaneous smiles were of smaller amplitude and had a larger and more consistent relation between amplitude and duration than deliberate smiles. A linear discriminant classifier using timing and amplitude measures of smile onsets achieved a 93% recognition rate. Using timing measures alone, recognition rate declined only marginally to 89%. These findings suggest that by extracting and representing dynamic as well as morphological features, automatic facial expression analysis can begin to discriminate among the message values of morphologically similar expressions.

Keywords: automatic facial expression analysis, timing, spontaneous facial behavior

AMS Subject Classification:

1. Introduction

Almost all work in automatic facial expression analysis has sought to recognize either prototypic expressions of emotion (e.g., joy or anger) or more molecular appearance prototypes such as FACS action units. This emphasis on prototypic expressions follows from the work of Darwin[10]and more recently Ekman[12] who proposed that basic emotions have corresponding prototypic expressions and described their components, such as crows-feet wrinkles lateral to the outer eye corners, in emotion-specified joy expressions. Considerable evidence suggests that six prototypic expressions (joy, surprise, anger, sadness, disgust, and fear) are universal in their performance and in their perception[12] and can communicate subjective emotion, communicative intent, and action tendencies.[18, 19, 26]

Prototypic expressions, or indeed any facial expression of interest (e.g., raised brows found in greeting displays), result from the contraction of facial muscles and their consequent effects on skin and underlying subcutaneous fascia. Work in automatic facial expression analysis has focused on the final appearance change (e.g., smile or raised brows) rather than the dynamic changes in appearance over time.[6] In the work presented here, we show that facial configurations similar in morphology but differing in psychological meaning can be discriminated based on differences in dynamic features. By extracting and representing dynamic as well as morphological features, automatic facial expression analysis can begin to discriminate among the message values of otherwise similar expressions. This is a critical step if we are to move from expression recognition to expression interpretation.

We focus on smiles because of their importance in human development and communication. Developmentally, smiles are one of the first emotion expressions to appear, they occur with relatively high frequency throughout the lifespan, and they can express a multitude of meanings, including joy, appeasement, and greetings, and they often serve to mask anger, disgust, and other negative emotions.[14, 16, 26] Ekman proposed that smiles of felt joy could be differentiated from non-felt or social smiles by whether or not smiling (contraction of *zygomaticus major*) was accompanied by contraction of the *orbicularis oculi*, which raises the cheek and causes crows-feet wrinkles to form at the eye corners) and by differences in asymmetry and timing. Lip-corner motion in felt smiles appeared smoother than that in non-felt or social smiles. This observation may be related to underlying differences in neural control to the extent that social smiles involve deliberate or volitional effort, thus differing in the relative balance of cortical and subcortical control (pyramidal and extra-pyramidal pathways[25]). In the present study, we show that differences in timing can reliably discriminate spontaneous from deliberate smiles.

Smiles consist of an initial onset phase, a peak, and an offset. We focus on the onset phase of smiling for several reasons. The onset phase provides the initial and most conspicuous change in appearance of the face as perceived by human observers.[22] Dinberg and Thunberg,[11] for instance, found that viewers activate their own *zygomaticus major* muscles within 0.40 sec after viewing an image of a smile. In spontaneous social contexts, this rapid reaction is most likely in response to the onset phase, since this time is well within the average duration of smile onsets, which is about 0.7 seconds.[4] For the purposes of producing a response in a receiver's facial muscles, therefore, the onset phase of a smile is likely to afford the initial facial signal. The onset phase also is believed to occur prior to suppressor or other movements that contribute to heterogeneity in smile configuration.[27] If the onset phase varies between types of smiles, timing must be a critical dimension.

Spontaneous smile onsets, which are produced by contraction of the *zygomaticus major*, subjectively appear genuine and are believed to exhibit the characteristics of automatic movement.[17] Previous work on other automatic movements has shown a consistent, deterministic relationship between maximum duration and amplitude.[1] Automatic motor routines are pre-programmed and have the characteristic of reaching their goal without interruption, leading to a consistent relationship between maximum duration and amplitude of movement.[3] Deliberate smiles, because they are mediated by greater cortical involvement,[25] are less likely to exhibit characteristics of a pre-programmed motor routine and thus would have a less predictable relation between duration and amplitude. In addition, following Fridlund[18] we predicted that smiles occurring in a social context would show greater amplitude than those occurring in a solitary context.

2. Method

2.1. *Participants*

Spontaneous smiles from two independent studies, a film-clip viewing context designed to elicit spontaneous positive emotion[24] ("emotion induction" N = 48), and a directed facial action task in which subjects were instructed to smile deliberately[20] (n = 33) were analyzed. Original videotaping and study procedures involving human participants were approved by the Institutional Review Board of the University of Pittsburgh.

Emotion induction (spontaneous smile) participants were selected from a larger group of participants recruited for a study of risk for depression conducted at the University of Pittsburgh Affect Analysis Group Laboratory. Selected participants watched a film clip of a comedy routine while seated alone in a laboratory room during the course of a study designed to measure facial activity and other psychophysiological measures of emotional response. Directed facial action task subjects (deliberate smile) were recruited for purpose of creating a large representative image database of facial expression for use in algorithm development and testing. Details may be found in Kanade, Cohn, and Tian.[20]

The gender ratios for spontaneous and deliberate smile groups did not significantly differ (% male = 47% and 33%, respectively, $X^2 = 1.51$, $df = 1$, $p > .10$) nor did minority representation (27% and 24%, respectively, $X^2 < 1$, $df = 1$, $p > .10$). Because some of the participants in the spontaneous smile condition had a history of depression, which could affect the timing of expressive behavior,[34] we assessed depressive symptoms using the Beck Depression Inventory.[2] All scores were below clinically significant cut-off values.

Videotaping procedures for these independent studies were similar with respect to pose, focal range, and lighting. In the emotion induction sample, facial electromyography using surface electrodes was conducted in the regions of the *corrugator supercilii*, *orbicularis oculi*, *levator labii superioris*, and the *zygomaticus major*. By recording facial EMG, we were able to validate the concurrent validity of automatic facial motion tracking with underlying muscle activity. Individual participant smiles were selected from the first analyzable image sequence, either following the first joke in the emotion induction context, or during the directed facial action tasks in that study.

2.2. *Materials*

2.2.1. *Image Data*

Sequences were digitized at 30 frames per second, producing a set of sequential 640x480 pixel grayscale images. The first spontaneous smile in the emotion induction image data and the first deliberate smile in the directed facial action task image data were selected for analysis. A sequence was eligible for analysis if it was free of occlusion and any out-of-plane motion was minimal. Videotape sequences were manually coded for activity of facial muscles affecting lip corner movement.[13] Agreement between two certified FACS coders on a subset of 27 participants was 0.91 overall for visible muscle activity of *orbicularis oculi* (action unit 6), *zygomaticus major* (action unit 12), d*epressor anguli oris* (action unit 15), and *mentalis* (action unit 17).[13]

2.2.2. Facial feature tracking and face analysis

The system of Tian, Kanade, and Cohn[30] was used to track facial features. These included the lip and eye corners and twelve other facial feature points (see Figure 1). Only measurements of lip and inner eye corners are included in this report. Pixel coordinates of the right lip corner in subsequent frames, relative to the initial center point of the lip corners in the initial frame, were automatically obtained using the Lucas-Kanade algorithm for feature tracking as implemented in.[23] Following Schmidt, Cohn, and Tian,[28] this initial center point was recalculated automatically by the program in each frame, relative to the stable inner eye corner feature points, allowing for accurate measure in the case of small head movements. Note that this calculated “initial center point” remains stable with respect to the image frame, not the current position of the lips, and thus is a suitable anchor point for the measurement of lip corner movement.

The right lip corner position, measured as r :

$$r = \sqrt{x^2+y^2} \tag{2.1}$$

was recorded for each frame of the sequence (see Figure 1.). A similar approach to

Figure 1. Measurement of Right Lip Corner Movement (r). a: Initial center point, located midway between lip corners in the initial frame, is used to anchor the radius (r). This initial center point is recalculated automatically in each frame, relative to the stable inner eye corner feature points, allowing for accurate measure in the case of small head movements. b-c: r is the distance (in pixels) between the initial center point and the right lip corner, in each frame. Reproduced from Schmidt, Cohn, & Tian.[28]

facial feature measurement, using manually obtained feature point position values, was followed by Van Swearingen and colleagues.[31] Reliability for automatic feature tracking was assessed using the values of r (i.e., right lip corner movement) obtained from two independent observations, based on the initial placement of feature tracking points by two different researchers. Measurements of r produced from these independently tracked sequences were highly correlated over series ranging from 200-300 frames in each of 5 image sequences, indicating that feature tracking recorded similar movement patterns, even when initial feature point placement varied. Correlation coefficients ranged from 0.94 to 1.00, comparable to tracking reliability values obtained in other studies using the same system.[9, 32]

For the sake of comparison among participants, individual participant's r values were standardized on the initial width of each participant's mouth in the initial image. This standardized value is hereafter referred to as "radius". Standardization resulted in a starting frame radius value of 0.5 (half of the distance from center of mouth to right lip corner) for each participant. Radius values were collected for each frame in the video sequence, forming a time series of lip corner position. Results for time series data were smoothed, using the T4253H algorithm[29] (see Figure 2 for an example of an automatically tracked smile and smoothed radius values).

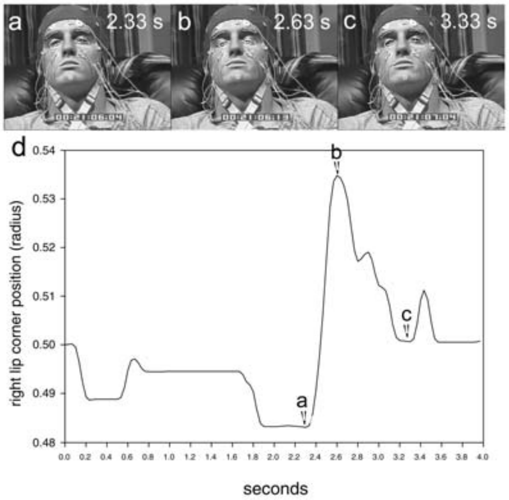

Figure 2. Facial feature tracking of lip corner movement in the spontaneous smile condition and change in radius over time. a) start of smile onset; b) end of smile onset; c) end of smile offset; d) selected frames from tracked image sequence. Adapted from Schmidt, Cohn, & Tian.[28]

The longest continuous increase in lip corner position (radius value) was obtained for each participant. This period of lip corner movement was defined as the smile onset (see Figures. 2 for examples of beginning and ending frames of smile onsets and ending frame of smile offsets). Duration and amplitude of lip corner movement during this onset period were calculated, based on the smoothed data. The duration of lip corner movement was measured as the difference in time (s) between the last and first frames of the onset, and the amplitude of lip corner movement was measured as difference in values of lip corner movement between beginning and end of longest continuous increase.

The longest continuous decrease in lip corner position (radius value) was also obtained for each participant. This period of lip corner change was defined as the smile offset, and the peak duration of the smile was defined as the distance between ending onset frame and beginning offset frame.

3. Results

3.1. Concurrent validity of radius of lip-corner with activation of the zygomaticus major.

Zygomaticus major EMG and lip corner motion were in agreement for AU 12 in 72% of cases with distinct EMG onset. In smiles in which both EMG and visually tracked onset were detectable, onsets were highly correlated (*r* = 0.95, *p*<.01), with visible onset an average of .23 seconds after the EMG onset (See Figure 3 for an example).

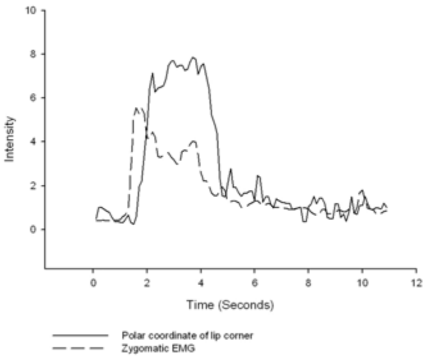

Figure 3. Relation between z*ygomaticus major* EMG and radius of lip-corner.

3.2. Differences in timing and amplitude between spontaneous and deliberate smiles.

Spontaneous and deliberate smiles differed on two of three measures evaluated. Duration of the onset phase was comparable. Amplitude of the onset phase and the ratio of duration to amplitude significantly differed. Amplitude was smaller in the spontaneous condition and the ratio of duration to amplitude substantially larger (see Table 1).

Consistent with the hypothesis that spontaneous but not deliberate smiles represent automatic pre-programmed motor routines, the relation between amplitude and duration was highly consistent in spontaneous smiles and essentially random for deliberate smiles. Separate linear regressions of amplitude on duration accounted for 69% of the variance in spontaneous smiles ($p < .001$) but only 9% of the variance in deliberate smiles (p = N.S.) (see Figure 4). While the slope for spontaneous smiles was highly significant (F =

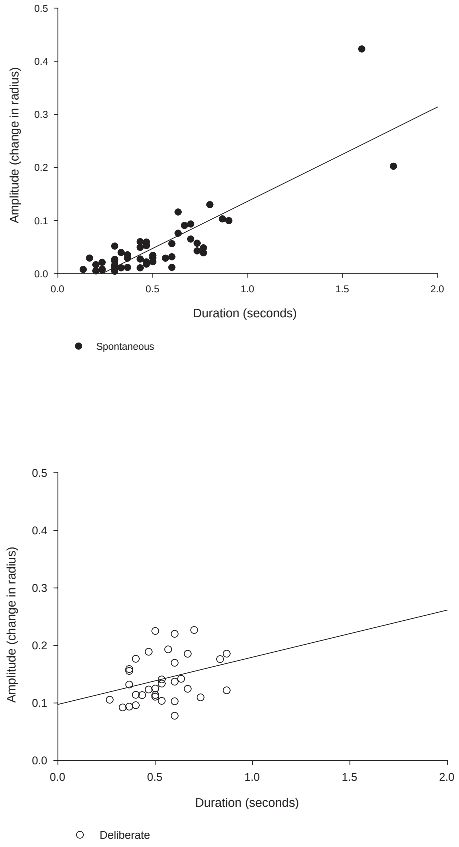

Figure 4. Relation between duration and amplitude of spontaneous and deliberate smile onsets. $R^2 = .69$ for spontaneous smiles and .09 for deliberate smiles.

103.68, *df* = 1,46, p <.001), for deliberate smiles it did not significantly differ from zero (*F* = 3.186, *df* = 1, 31, N.S.). The difference in slope was significant (*t* = 2.00, *df* = 77, *p* < .05).

Table 1. Descriptive statistics.

	Spontaneous		Deliberate		
	Mean	(SD)	Mean	(SD)	*t (df* = 79*)*
Duration (seconds)	.521	(.315)	.538	(.153)	*p* = N.S.
Amplitude (change in radius)	.051	(.067)	.142	(.041)	*p* < .001
Ratio of duration to amplitude	17.963	(13.49)	4.017	(1.39)	*p* < .001

To evaluate the utility of using the timing of smile onsets in distinguishing between spontaneous and deliberate smiles, we entered amplitude, duration, and ratio of amplitude to duration into a linear discriminant analysis with cross-validation by leave-one-out procedure. The combination of onset duration, amplitude, and ratio of duration-to-amplitude resulted in correct classification rate of 93% (Table 2). Using only duration-to-amplitude ratio, the rate decreased only marginally, to 89% (Table 3).

Table 2. Discrimination between spontaneous and deliberate smiles using linear discriminant classifier with three predictors, duration, amplitude, and ratio of duration to amplitude.

Predicted	True	
	Spontaneous	Deliberate
Spontaneous	46	2
Deliberate	4	29

Note. Wilks' $\lambda = 0.425$, $X^2 = 66.35$, $p < .001$. Classification rate = 93%

Table 3. Discrimination between spontaneous and deliberate smiles using linear discriminant classifier with single predictor, the ratio of duration to amplitude.

Predicted	True	
	Spontaneous	Deliberate
Spontaneous	39	9
Deliberate	0	33

Note. Wilks' $\lambda = 0.694$, $X^2 = 28.72$, $p < .001$. Classification rate = 89%

4. Discussion

We found strong evidence that spontaneous and deliberate smiles could be distinguished from differences in the timing of smile onsets. In spontaneous but not deliberate smiles, amplitude of lip corner motion was a strong linear function of duration. Tight coupling of duration and amplitude has been described previously for saccadic eye movements,[1] which are pre-programmed motor routines. These findings suggest that the onset phase of spontaneous smiles is an automatic "unbidden" action, as initially proposed by Ekman.[12] In contrast, there was no consistent relation between duration and amplitude in deliberate smiles, which is consistent with pyramidal motor control. A linear discriminant function using only ratio of duration to amplitude discriminated 89% of spontaneous and deliberate smiles. When morphological features were added, discrimination increased marginally, to 93%. This is the first study to use automatic facial expression analysis to discriminate differences of psychological significance between facial expressions similar in appearance.

In the smiles we analyzed, head motion was limited primarily to the image plane. The use of a high-backed chair in the spontaneous smile condition may have limited smiles accompanied by out-of-plane motion, and the few that occurred were omitted from analysis. In future work, it will be important to consider simultaneously the dynamics of rigid (head) and non-rigid (face) motion as well as direction of visual regard. Covariation across modalities may inform interpretation of facial expression. "Surprise" expressions when they occur with upward pitch of the head and eyes has been found to indicates visual tracking rather than the emotion of surprise.[5] Quantitative measurement of the timing of head and facial motion may help disambiguate the meaning of morphologically similar facial actions (e.g., surprise or smile). Indeed, for some emotions, such as embarrassment, head motion is a key component of the display.[21] Observers report that smiles of embarrassment occur as the head pitches down.

Xiao, Kanade, and Cohn[33] recently developed a cylindrical head model that recovers full 6 *df* of head movement as well as permitting stabilization of the face image for detailed analysis of facial motion, independent of head motion.[8] The approach is sufficiently robust to quantify differences between the downward and upward phases of blinks occurring during spontaneous facial behavior with moderate out-of-plane head motion. Applying this system to smiles thought to indicate embarrassment, we found the negative correlation between lip-corner displacement and head motion[7] as predicted from human-observer based descriptions by Keltner.

In summary, we used automatic feature tracking to measure the relation between amplitude and duration of smile onsets in spontaneous and deliberate smiles of 81 young adults of Euro- and African-American background. Spontaneous smiles were of smaller amplitude and had a larger and more consistent relation between amplitude and duration than deliberate smiles. A linear discriminant classifier using timing and amplitude measures of smile onsets achieved a 93% recognition rate. Using timing measures alone, recognition rate declined only marginally to 89%. These findings suggest that by extracting and representing dynamic as well as morphological features, automatic facial expression analysis can discriminate among the message values of morphologically similar expressions.

Acknowledgements

This work was supported by grants R01 MH51435, P01 MH56193, and F32 MH 012579 to the University of Pittsburgh from the U.S. National Institute of Mental Health.

References

1. Baloh, R. W., Sills, A. W., Kumley, W. E., & Honrubia, V. (1975). Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology, 25,* 1065-1070.
2. Beck, A., Steer, R., & Garbin, M. (1988). Psychometric properties of the Beck depression inventory: *Clinical Psychology Review, 8,* 77-100.
3. Brooks, V. B. (1986). *The Neural Basis of Motor control*. New York: Oxford University Press.
4. Bugental, D. B. (1986). Unmasking the "Polite Smile": Situational and Personal Determinants of Managed Affect in Adult-Child Interaction. *Personality and Social Psychology Bulletin, 12,* 7-16.
5. Camras, L.A., Lambrecht, L., & Michel, G.F. (1996). Infant "surprise" expressions as coordinative motor structures. *Journal of Nonverbal Behavior, 20,* 183-195.
6. Cohn, J. F. (in press). Automated analysis of the configuration and timing of facial expression. In P. Ekman & E. Rosenberg, *What the face reveals* (2nd edition): *Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press Series in Affective Science.* New York: Oxford.
7. Cohn, J.F., Reed, L., Moriyama, T., Xiao, J., Schmidt, K., & Ambadar, Z. (In press). Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FG'04)*, Seoul, Korea, xxx-xxx.
8. Cohn, J.F., Xiao, J., Moriyama, T., Ambadar, Z., & Kanade, T. (2003). Automatic recognition of eye blinking in spontaneously occurring behavior. *Behavior Research Methods, Instruments, and Computers, 35,* 420-428.
9. Cohn, J. F., Zlochower, A., Lien, J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology, 36*, 35-43.
10. Darwin, C.R. (1872/1965). *The expression of emotions in man and animals*. London: John Murray (Chicago, Illinois: University of Chicago).
11. Dimberg, U., & Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology, 39,* 39-45.

12. Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48,* 384-392.
13. Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System*. Palo Alto: Consulting Psychologists Press.
14. Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable Smiles. *Journal of Nonverbal Behavior, 6*, 238-252.
15. Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion: In U. Segerstrale & P. Molnar (Eds.), *Nonverbal Communication* (pp. 27-46). Mahwah, NJ: LEA.
16. Ekman, P. & Rosenberg, E. (Eds.) (in press). What the face reveals. 2nd edition. NY, NY: Oxford.
17. Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology, 64,* 83-93.
18. Fridlund, A. (1994). *Human Facial Expression: An evolutionary view*. New York: Academic Press.
19. Frijda, N.H. & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In Russell & & Fernandez-Dols, *The psychology of facial expression,* pp. 78-102.
20. Kanade, T., Cohn, J.F., & Tian, Y. (March 2000). Comprehensive database for facial expression analysis. *Fourth IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 46-53. Grenoble, France.
21. Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68,* 441-454.
22. Leonard, C. M., Voeller, K. K. S., & Kuldau, J. M. (1991). When's a smile a smile? or how to detect a message by digitizing the signal. *Psychological Science, 2*, 166-172.
23. Lien, J. J.-J., Kanade, T., Cohn, J. F., & Li, C.-C. (2000). Detection, tracking, and classification of subtle changes in facial expression. *Journal of Robotics and Autonomous Systems, 31,* 131-146.
24. Miller, A. Fox, N.A., Cohn, J.F., Forbes, E.E., Sherrill, J.T., & Kovacs, M. (2002). Regional brain activity in adults with a history of childhood-onset depression. *American Journal of Psychiatry, 159*, 934-940.
25. Rinn, W.E. (1984). The neuropsychology of facial expression. *Psychological Bulletin, 95*, 52-77.
26. Schmidt, K., & Cohn, J.F. (2001). Human facial expressions as adaptations: Evolutionary perspectives in facial expression research. *Yearbook of Physical Anthropology, 116*, 8-24.
27. Schmidt, K. & Cohn, J.F. (2001). Dynamics of facial expression: Normative characteristics and individual differences. *IEEE International Conference on Multimedia and Expo (ICME 2001)*, Tokyo, Japan, 728-731.

28. Schmidt, K., Cohn, J. F., & Tian, Y. L. (2003). Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles. *Biological Psychology, 65, 49-66*
29. SPSS Inc. (1999). SPSS Base 10.0 for Windows User's Guide. SPSS Inc., Chicago IL.
30. 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.
31. Van Swearingen, J. M., Cohn, J. F., & Bajaj-Luthra, A. (1999). Specific impairment of smiling increases the severity of depressive symptoms in patients with facial neuromuscular disorders. *Aesthetic and Plastic Surgery, 23,* 416-423.
32. Wachtman, G., Cohn, J. F., Van Swearingen, J. M., & Manders, E. K. (2001). Automated tracking of facial features in patients with facial neuromuscular dysfunction. *Plastic and Reconstructive Surgery, 107,* 1124-1133.
33. Xiao, J., Kanade, T., & Cohn, J.F. (May 2002). Robust Full Motion Recovery of Head by Dynamic Templates and Re-registration Techniques. *Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FG'02),*. 163-169. Washington, DC.
34. Zlochower, A. & Cohn, J.F. (1996). Vocal timing in face-to-face interactions of clinically depressed and nondepressed mothers and their 4-month-old infants. *Infant Behavior and Development, 19,* 373-376.