Use of Active Appearance Models for Analysis and Synthesis of Naturally Occurring Behavior

Jeffrey F. Cohn, *Associate Member, IEEE*

***Abstract*— Significant efforts have been made in the analysis and understanding of naturally occurring behavior. Active Appearance Models (AAM) are an especially exciting approach to this task for facial behavior. They may be used both to measure naturally occurring behavior and to synthesize photo-realistic real-time avatars with which to test hypotheses made possible by those measurements. We have used both of these capabilities, analysis and synthesis, to investigate the influence of depression on face-to-face interaction. With AAMs we have investigated large datasets of clinical interviews and successfully modeled and perturbed communicative behavior in a video conference paradigm to test causal hypotheses. These advances have lead to new understanding of the social functions of depression and dampened affect in dyadic interaction. Key challenges remain. These include automated detection and synthesis of subtle facial actions; hybrid methods that optimally integrate automated and manual processing; computational modeling of subjective states from multimodal input; and dynamic models of social and affective behavior.**

***Index Terms*— AAM, facial expression, animation, depression.**

INTRODUCTION

Significant efforts have been made in the analysis and understanding of naturally occurring behavior. Active Appearance Models (AAM) [1] are an especially exciting approach to this task for facial behavior as they are robust to variation in pose and head motion that are typical of spontaneous behavior, they are able to extract shape and appearance features, critical for optimal detection of facial actions, execute rapidly enough for real-time applications, and are photorealistic [2]. They may be used both to measure naturally occurring behavior [3, 4] and to animate photo-realistic real-time avatars with which to test hypotheses suggested by those measurements or observations[5].

MEASURMENT AND SYNTHESIS

We have used both of these capabilities, analysis and synthesis, to investigate the influence of Major Depressive Disorder (MDD) [6] on face-to-face interaction. First, we used AAMs and audio signal processing to measure severity of depression in patients (n = 57) undergoing treatment for MDD. Ground truth for depression severity was measured in clinical interviews using the Hamilton Rating Scale for Depression (HRS-D) [7] at multiple assessments over the course of treatment. AAM derived features as well as acoustic parameters both demonstrated moderate concurrent validity with symptom severity as measured by the HRS-D. On basis of these findings, we anticipated that dampened facial and vocal behavior would have powerful impact on face-to-face interaction.

To pursue this question, we exploited the animation capabilities of AAM. In a two-person video-conference paradigm, we covertly perturbed head motion and facial and vocal expressiveness of one participant to simulate depression and measure its impact on each partner's behavior (n=27 dyads) [8]. Facial expression and head motion were dampened or exaggerated by AAM (See Fig. 1). For consistency with facial expression, vocalization was dampened or exaggerated by digital signal processing. Head motion and facial expression were dampened, exaggerated, or presented normally using a photorealistic, real-time avatar. Neither partner was aware that the other person's face was replaced by an avatar's face – the avatars were accepted as live video – or that the perception of their dynamics by the partner was experimentally modified.

In response to a "dampened" avatar, partners increased their own communicative head motion, as if to elicit more normal expressiveness in the avatar. This is the first time that

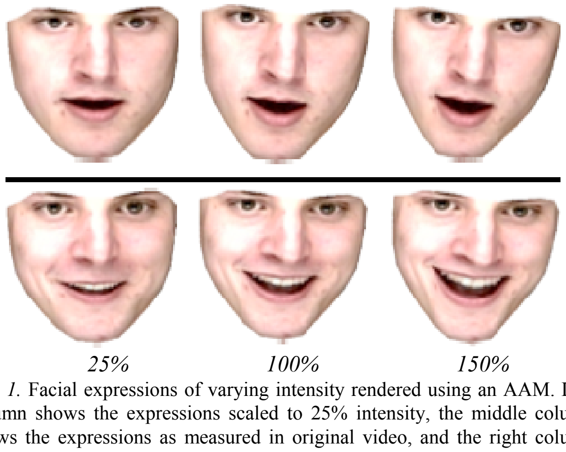

Fig. 1. Facial expressions of varying intensity rendered using an AAM. Left column shows the expressions scaled to 25% intensity, the middle column shows the expressions as measured in original video, and the right column shows expressions exaggerated to 150% of the intensity measured in the video. The effect of scaling the parameters is much more pronounced in more extreme expressions. Adapted from [5]

This work was supported in part by the U.S. National Institutes of Health grant MH R01 51435 and National Science Foundation grant BCS05 27397. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect those of the National Science Foundation or National Institutes of Health.

Jeffrey F. Cohn is Professor of Psychology at the University of Pittsburgh and Adjunct Faculty at the Robotics Institute, Carnegie Mellon University: 3137 Sennott Square, 210 S. Bouquet St., Pittsburgh, PA 15260 USA. Phone: 412-624-8825; fax: 412-624-2023; e-mail: jeffcohn@cs.cmu.edu.

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

such effects have been tested experimentally in other than infants [9] and with this level of experimental control. In the study with infants, mothers simulated depression and thus were aware of the manipulation, which may have influenced their behavior. In the current study, participants were unaware of the manipulation and responded only to what they thought was the partner. The findings from this work may help explain the interpersonal rejection experienced in depression. Failing to elicit a change in the dampened individual's behavior, the partner is likely to turn away, experience negative affect, and withdrawal [10].

CURRENT CHALLENGES

In our experience to date, we have encountered five key challenges:

1) Automatic detection of subtle facial actions While high-intensity actions may be common in posed behavior[11], low intensity actions are more typical in naturally occurring behavior. Even when strong expressions occur, they may be quickly attenuated by "smile controls" or similar actions [12]. Significant gains in automatic detection of FACS [13] action units [3, 14, 15] and dynamics [16-18] notwithstanding, automatic detection of subtle facial actions remains a challenge. People are highly responsive to even extremely subtle facial actions [19]; the same acuity is needed for AAM and other automated approaches.

Rather than seek broad improvement in action unit detection, there may be value in using domain knowledge to set priorities on which ones to concentrate. As a candidate for focused efforts, consider contempt. Contempt expressions (AU 14 in FACS, resulting from contraction of the *buccinator* muscle) is a common response in depression [12], correlates with depression severity [20] and treatment outcome [21], signals violation of community standards [22], and predicts maladaptive outcomes [23]. A powerful emotion, contempt often is inhibited, occurring at low intensities that may escape automatic detection. Improved automatic detection of subtle expressions of contempt would be of considerable benefit in several areas of clinical and social psychology.

2) An equally pressing challenge is synthesis of specific facial actions. Theobald and colleagues [5, 24] successfully cloned facial expression and identity between a source person and a photorealistic avatar of another person in real time, but adding or subtracting specific facial actions in a photorealistic avatar remains a research question. While expressiveness may be dampened or exaggerated in mapping from one person's video to another's, the means to introduce or delete specific facial actions in a real-time photorealistic avatar are not yet known. As above, domain knowledge could suggest priorities for research. As a candidate action, consider cheek raising (AU 6 in FACS, resulting from contraction of the lateral portion of the *orbicularis oculi*). Cheek raising is believed to profoundly alter the meaning of a smile [17, 25], but causal evidence for this hypothesis awaits necessary progress in AAM-based synthesis. There would be broad interest in experiments of this sort.

3) How best to integrate automated and manual methods? De la Torre and Simon and colleagues [26] developed a promising hybrid approach to action unit detection. The apex or peak of FACS action units is manually annotated; an AAM then automatically identifies the onset and offset of each AU. This method has reduced human processing by 50% or more and is the first to demonstrate precision for action unit onsets and offsets. While fully automated systems are our goal, much progress may be achieved by utilizing the best aspects of both manual and automated methods.

4) How can we infer intention and emotion from facial actions? There is no 1:1 mapping between expressive or physiological signals and subjective states of intention or emotion (cf. [27]). Inferences from multiple sources and computational models for their fusion will likely prove necessary.

5) With the advent of new, more powerful tools, more informative languages of social interaction than yet exist will become possible. Unsupervised learning may suggest useful alternatives to manually based schemes such as FACS [28]. More generally, the dynamics of facial behavior and social behavior, more generally, are with few exceptions[29] unchartered. Improved understanding of dynamics would contribute to improved AU detection and synthesis and enable greater advances in applications spanning human-human, human-computer, and robot-human interaction.

SUMMARY

In summary, AAMs provide a powerful tool with which to measure facial behavior and to animate photorealistic real-time avatars in a video conference paradigm. The analysis capabilities of AAMs enrich observations and make possible new hypotheses about social behavior. The synthesis or animation capabilities of AAMs enable rigorous experimental tests of those hypotheses. Perturbations can be introduced in face-to-face interaction without either participant's becoming aware that the source behavior is manipulated. We have used these capabilities of AAM to investigate affective and interpersonal effects related to depression. In doing so, we have made discoveries and identified key challenges for current and future research.

ACKNOWLEDGMENT

The research and ideas expressed here owe deep thanks to my collaborators. These include Zara Ambadar, Steven Boker, Fernando De la Torre, Takeo Kanade, Patrick Lucey, Simon Lucey, Iain Matthews, Daniel Messinger, Ken Prkachin, and Barry John Theobald among others.

REFERENCES

[1] Cootes, T.F., Edwards, G., and Taylor, C.J.: 'Active Appearance Models', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23, (6), pp. 681-685

[2] Matthews, I., and Baker, S.: 'Active appearance models revisited', International Journal of Computer Vision, 2004, 60, (2), pp. 135-164

[3] Lucey, P., Cohn, J.F., Lucey, S., Sridharan, S., and Prkachin, K.: 'Automatically detecting action units from faces of pain: Comparing shape and appearance features', 2nd IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB), 2009

[4] Messinger, D.S., Cassel, T.D., Acosta, S.I., Ambadar, Z., and Cohn, J.F.: 'Infant smiling dynamics and perceived positive emotion', Journal of Nonverbal Behavior, 2008, 32, (3), pp. 133-155

[5] Theobald, B.J., Matthews, I., Mangini, M., Spies, J.R., Brick, T., Cohn, J.F., and Boker, S.M.: 'Mapping and manipulating facial expression', Language and Speech, 2009, 52, (2 & 3), pp. xxx-xxx

[6] American Psychiatric Association: 'Diagnostic and statistical manual of mental disorders' American Psychiatric Association, Fourth, 1994.

[7] Hamilton, M.: 'A rating scale for depression', Journal of Neurology and Neurosurgery, 1960, 23, pp. 56-61

[8] Boker, S.M., Cohn, J.F., Theobald, B.-J., Matthews, I., Brick, T., and Spies, J.R.: 'Effects of damping facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars', Royal Society, Computation of Emotions in Man and Machine, 2009

[9] Cohn, J.F., and Tronick, E.Z.: 'Three month old infants' reaction to simulated maternal depression', Child Development, 1983, 54, pp. 185-193

[10] Cohn, J.F., and Tronick, E.Z.: 'Specificity of infants' response to mothers' affective behavior', Journal of the American Academy of Child and Adolescent Psychiatry, 1989, 28, pp. 242-248

[11] Kanade, T., Cohn, J.F., and Tian, Y.: 'Comprehensive database for facial expression analysis', Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, FG '00, pp. 46-53

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

[13] Ekman, P., Friesen, W.V., and Hager, J.C.: 'Facial action coding system' Research Nexus, Network Research Information, Salt Lake City, UT, 2002.

[14] Tong, Y., Liao, W., and Ji, Q.: 'Facial action unit recognition by exploiting their dynamic and semantic relationships', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29, (10), pp. 1683 - 1699

[15] Pantic, M., and Bartlett, M.S.: 'Machine analysis of facial expressions', in Delac, K., and Grgic, M. (Eds.): 'Face recognition' (I-Tech Education and Publishing, 2007), pp. 377-416

[16] Valstar, M.F., Gunes, H., and Pantic, M.: 'How to distinguish posed from spontaneous smiles using geometric features', ACM International Conference on Multimodal Interfaces, 2007, pp. xxx-xxx

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

[18] Cohn, J.F., and Schmidt, K.L.: 'The timing of facial motion in posed and spontaneous smiles', International Journal of Wavelets, Multiresolution and Information Processing, 2004, 2, pp. 1-12

[19] Ambadar, Z., Schooler, J.W., and Cohn, J.F.: 'Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions', Psychological Science, 2005, 16, pp. 403-410

[20] Cohn, J.F.: 'Automated facial image analysis for affective and clinical science', NIMH Mini-Symposium on Computational Analysis of Behavior, 2009

[21] Ekman, P., Matsumoto, D., and Friesen, W.V.: 'Facial expression in affective disorders', in Ekman, P., and Rosenberg, E. (Eds.): 'What the face reveals' (Oxford, 2005, 2nd edn.), pp. 331-341

[22] Rozin, P., Lowery, L., Imada, S., and Haidt, J.: 'The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity)', Journal of Personality & Social Psychology, 1999, 76, (4), pp. 574-586

[23] Gottman, J., Levenson, R., and Woodin, E.: 'Facial expressions during marital conflict', Journal of Family Communication, 2001, 1, (1), pp. 37-57

[24] Theobald, B.J., Bangham, J.A., Matthews, I., and Cawley, G.C.: 'Near-videorealistic synthetic talking faces: Implementation and evaluation', Speech Communication, 2004, 44, pp. 127-140

[25] Frank, M.G., Ekman, P., and Friesen, W.V.: 'Behavioral markers and recognizability of the smile of enjoyment', Journal of Personality and Social Psychology, 1993, 64, (1), pp. 83-93

[26] De la Torre, F., Simon, T., Ambadar, Z., and Cohn, J.F.: 'Fast FACS: A computer vision assisted system to increase speed and reliability of manual FACS coding', Unpublished

[27] Cacioppo, J.T., and Tassinary, L.G.: 'Inferring psychological significance from physiological signals', American Psychologist, 1990, 45, (1), pp. 16-28

[28] De la Torre, F., and al., e.: 'Unsupervised discovery of facial events: Learning a dynamic vocabulary for facial analysis', Submitted, 2009

[29] Ashenfelter, K.T., Boker, S.M., Waddell, J.R., Vitanov, N., and Abadjieva, E.: 'Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation', Journal of Experimental Psychology: Human Perception and Performance, 2008