Ehrlich, S.M., Schiano, D.J., & Sheridan, K. (2000). Communicating facial affect: It's not the realism, it's the motion. *Proceedings of ACM CHI 2000 Conference on Human Factors in Computing Systems* (pp. 252-253). NY: ACM.

Communicating Facial Affect: It's Not the Realism, It's the Motion

Sheryl M. Ehrlich, Diane J. Schiano, & Kyle Sheridan
Interval Research Corporation
1801 Page Mill Road, Palo Alto, CA 94304
+ 1 650 842 6099
{ehrlich, schiano, sheridan}@interval.com

ABSTRACT

Designers of video-mediated communication and affective computing applications must make tradeoffs to deal with limited bandwidth. Typically spatial resolution and color are preserved at the expense of temporal resolution and accuracy. Our data suggest that this may not be the appropriate tradeoff for communicating facial affect; preserving motion is critical and may even compensate for major losses in image realism.

Keywords

Facial affect, facial expression of emotion, face, nonverbal communication, image degradation, video conferencing.

INTRODUCTION

The face is the primary carrier of emotion, and is thus a major contributor to both effective interpersonal communication and affective computing [7]. As bandwidth for such applications increases, adding a video channel becomes possible--but only under constrained (or compressed) conditions. Typically, this limited video bandwidth has been used to show highly realistic facial images (i.e., high spatial resolution, full color) at the expense of motion (e.g., low temporal resolution, lags). Yet it is becoming increasingly clear that this may not be the appropriate tradeoff [3].

Facial recognition is robust across severe image degradations [1]. However, the perception of emotional content relies heavily on motion information [3]. Emotions can be identified in displays consisting solely of fields of moving dots (i.e., no facial features whatsoever) [2]. The timing and trajectories of facial expressions are highly precise (some are in the millisecond range), and people show great sensitivity to temporal parameters in trying to determine felt emotion, its intensity and sincerity [4]. This suggests that if a bandwidth tradeoff is required, one should consider preserving high-fidelity motion information at the expense of image realism, not the other way around.

This paper presents an initial study exploring this issue, performed in the context of a larger research program on facial affect [8]. Static and moving images depicting emotional expressions were presented to participants for varying lengths of time. The moving images showed an actor transitioning from a "neutral" expression toward a very intense depiction of a specific emotion (e.g., "happy"). The static images were the endpoints from these transitions. The images were either highly realistic (high resolution color) or degraded (filtered to resemble sketches). The primary question of this research concerns the role of motion. Which will be perceived as showing higher emotionality: A static image portraying a given emotion at highest intensity for a fixed duration, or a video transitioning between neutral (i.e., no emotion) and increasingly intense emotion for the same duration? If image information content is the primary basis for perceiving emotion, the static image should seem more expressive. But if motion is the more critical factor, the moving images should appear more expressive, despite the fact that on average lower-intensity expressions are seen in the moving images condition. We predict that motion is indeed the more important factor. Moreover, we predict that motion cues may well be effective enough to compensate for substantial image degradation.

METHOD

Participants

Eighteen Stanford University students, aged 18 to 35.

Stimuli

A trained actor transitioned between neutral and the each of Ekman's 6 "basic" emotions (anger, disgust, fear, happiness, sadness, surprise) [5]. Eight-second video clips were digitized at full resolution, 30 fps, onto a Macintosh PowerPC using a Targa DTX digital video card. Four sets of stimuli were created. High-quality "realistic" images included 1) realistic still images: still frames taken at 2-, 5-, and 8-seconds into the clip, and presented as still images for 2, 5, and 8 seconds respectively; and 2) realistic moving images: the first 2, 5, and 8 seconds of each clip presented as a movie. The third and fourth sets of stimuli were "degraded" versions of those described above.

Degradation was produced by using a sequence of Adobe Premiere filters (Black & White, Find Edges, Invert, Blur, Brightness (112) and Contrast (101)) intended to transform the photograph-like realistic images into something closer to a sketch. Most contours were preserved, as when using an edge-detection filter (see Figure 1).

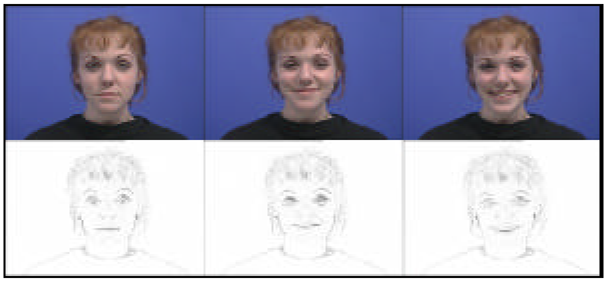

Figure 1: Realistic & degraded images taken from still frames 2, 5, and 8 seconds into the "neutral to happy" transition.

Procedure

Stimuli were presented on a 20" video monitor connected to a Macintosh PowerPC. The task was to select the emotional expression portrayed at the end of each sequence from an alphabetized list (which included all 6 emotions plus neutral) and rate its perceived intensity on a 0-6 (none-highest) scale. Stimuli were randomized, counterbalanced, and presented in HyperCard. Participants proceeded at their own pace; the session lasted under 1 hour.

RESULTS

Moving images were indeed rated higher than static images ($F(1, 17)=5.112$, $p=.0372$) in this study, despite the fact that on average lower-intensity expressions were seen in the moving images condition. Perceived expressiveness increased with clip duration (means = 1.09, 3.83 and 4.31 for 2-, 5- and 8-second clips, respectively; $F(2, 34)=574.3$, $p<.0001$), as expected.

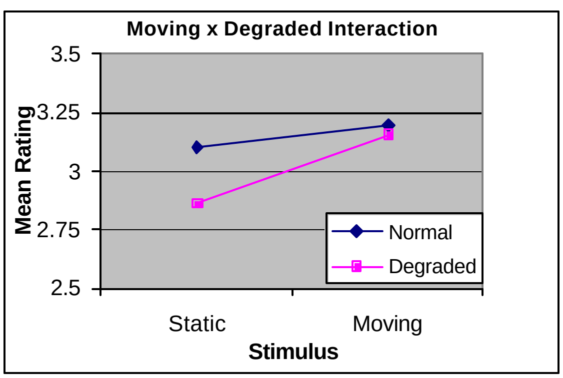

Figure 2: Moving x degradation interaction

Image degradation decreased perceived intensity for static but not moving images (see Figure 2), also as predicted ($F(1, 17)=8.943$, $p=.0082$). That is, motion cues effectively compensated for image degradation. While the size of this overall effect is not large, it is highly meaningful given that major facial features remained visible in the degraded images, and despite the high variability of emotion intensity ratings [8]. The effect is greater at longer motion durations. Finally, further analyses examined the magnitude of these effects for specific emotions. Most notably, in previous work with still images, the "fear" expression was found to be fairly ambiguous, and was consistently rated lowest [8]. This pattern was replicated in the static condition of this study (mean = 1.86), but adding motion raised the fear ratings to the level of the other expressions (mean = 3.12).

DISCUSSION

The results of this initial study underscore the importance of motion for facial affect, even under conditions strongly favoring the static images. The motion cues in this study were derived from a transition between neutral and a specific emotion, thus on average the motion clips showed fairly low-intensity stimuli (in terms of degree of emotion). The static images showed the most intense expressions at all times. Nonetheless, emotional intensity ratings were higher for moving images. This effect may be related to "representational momentum", in which the further trajectory of a moving stimulus is perceived from only partial viewing [6]. Further research is clearly needed. The present data further show that motion cues can compensate for image degradation. We find this intriguing, with potentially important design implications. A further study to replicate this finding and address the type of motion required for the effect, by comparing biologically plausible and implausible (morphed) motions, is in progress.

Designers of video-mediated communication and affective computing applications must make tradeoffs to deal with limited bandwidth. Our data suggest that preserving motion is critical for communicating facial affect and may even compensate for major losses in spatial resolution and color. Thus, decisions affecting visual dynamics (video framerate, lag, compression algorithm) must be approached with caution to avoid timing biases which can dramatically affect communication [3, 4]. Parameterization of the spatial-temporal tradeoffs for facial affect was not attempted in this exploratory study, but such work is planned as part of our current project. Bandwidth demand has always exceeded supply and we expect it to continue to do so, even in the broadband future. The critical issue is how to use the bandwidth that is available most effectively.

REFERENCES

1. Bachmann, T. Identification of spatially quantised tachistoscope images of faces. *European Journal of Cognitive Psychology, 3,* (1991), 87-103.
2. Bassili, J.N. Emotion Recognition. *Journal of Personality and Social Psychology, 37,* (1979), 2049-2058.

3. Bruce, V. The role of the face in communication. *Interacting with Computers, 8,* (1996), 166-176.
4. Colston, H.L. & Schiano, D.J. Looking and Lingering as Conversational Cues in Video-Mediated Communication. Proceedings of CHI 1995 (Denver, CO, May 1995)
5. Ekman, P. & Friesen, W.V. *The Facial Action Coding System.* Consulting Psychologists Press: Palo Alto, CA, 1978.
6. Kelly, M.H. & Freyd, J.J. Explorations of representational momentum. *Cognitive Psychology,* 19 (3), (1987), 369-401.
7. Lisetti, C. & Schiano, D.J. Automatic facial expression interpretation. *Pragmatics,* in press.
8. Schiano, D.J., Ehrlich, S.M., Rahardja, K. & Sheridan, K. Face to Interface: Facial Affect in (Hu)Man and Machine. *Proceedings of CHI2000* (The Hague, Netherlands, April 2000), in press.