CyberPsychology & Behavior
Volume 4, Number 4, 2001
Mary Ann Liebert, Inc.

Performance-Driven Facial Animation: Basic Research on Human Judgments of Emotional State in Facial Avatars

A.A. RIZZO, U. NEUMANN, R. ENCISO, D. FIDALEO, and J.Y. NOH

ABSTRACT

Virtual reality is rapidly evolving into a pragmatically usable technology for mental health (MH) applications. As the underlying enabling technologies continue to evolve and allow us to design more useful and usable structural virtual environments (VEs), the next important challenge will involve populating these environments with virtual representations of humans (avatars). This will be vital to create mental health VEs that leverage the use of avatars for applications that require human–human interaction and communication. As Alessi et al.[1] pointed out at the 8th Annual Medicine Meets Virtual Reality Conference (MMVR8), virtual humans have mainly appeared in MH applications to "serve the role of props, rather than humans." More believable avatars inhabiting VEs would open up possibilities for MH applications that address social interaction, communication, instruction, assessment, and rehabilitation issues. They could also serve to enhance realism that might in turn promote the experience of presence in VR. Additionally, it will soon be possible to use computer-generated avatars that serve to provide believable dynamic facial and bodily representations of individuals communicating from a distance in real time. This could support the delivery, in shared virtual environments, of more natural human interaction styles, similar to what is used in real life between people. These techniques could enhance communication and interaction by leveraging our natural sensing and perceiving capabilities and offer the potential to model human–computer–human interaction after human–human interaction. To enhance the authenticity of virtual human representations, advances in the rendering of facial and gestural behaviors that support implicit communication will be needed. In this regard, the current paper presents data from a study that compared human raters' judgments of emotional expression between actual video clips of facial expressions and identical expressions rendered on a three-dimensional avatar using a performance-driven facial animation (PDFA) system developed at the University of Southern California Integrated Media Systems Center. PDFA offers a means for creating high-fidelity visual representations of human faces and bodies. This effort explores the feasibility of sensing and reproducing a range of facial expressions with a PDFA system. In order to test concordance of human ratings of emotional expression between video and avatar facial delivery, we first had facial model subjects observe stimuli that were designed to elicit naturalistic facial expressions. The emotional stimulus induction involved presenting text-based, still image, and video clips to subjects that were previously rated to induce facial expressions for the six universals[2] of facial expression (happy, sad, fear, anger,

Integrated Media Systems Center/Andrus Gerontology Center, University of Southern California, Los Angeles, California 90089.

disgust, and surprise), in addition to attentiveness, puzzlement and frustration. Videotapes of these induced facial expressions that best represented prototypic examples of the above emotional states and three-dimensional avatar animations of the same facial expressions were randomly presented to 38 human raters. The raters used open-end, forced choice and seven-point Likert-type scales to rate expression in terms of identification. The forced choice and seven-point ratings provided the most usable data to determine video/animation concordance and these data are presented. To support a clear understanding of this data, a website has been set up that will allow readers to view the video and facial animation clips to illustrate the assets and limitations of these types of facial expression–rendering methods (www.USCAvatars.com/MMVR). This methodological first step in our research program has served to provide valuable human user–centered feedback to support the iterative design and development of facial avatar characteristics for expression of emotional communication.

INTRODUCTION

Over the last 25 years, there has been an emergence of psychological research on the human capacity to signal, recognize, and generally communicate implicit information via facial expression.[3,4] Although some controversy exists as to the limits of what information is conveyed or received facially,[4] a lively body of literature has produced findings that suggest the existence of basic universals of facial expression that are recognizable by humans regardless of sociocultural background or past exposure to visual media.[2] Some researchers take these results as indicative of underlying genetic mammalian hard-wired neural circuitry for nonverbal implicit communication,[5] as Darwin proposed over 100 years ago.[6] Others have placed less emphasis on postulating underlying mechanisms and instead have focused on the empirical analysis of the components of facial and body gestural communication.[3,4,7,8] An understanding of such issues relating to the nature of facial/gestural implicit communication is vital to support research and development efforts to create virtual human representations, or avatars, that can be integrated within virtual environments (VEs). In this regard, the creation of VEs that employ avatars in applications that require some level of simulated human–human interaction could provide new opportunities for the development of useful mental health (MH) virtual reality (VR) scenarios.

The creation of more compelling and naturalistic virtual environment applications has become possible with continuing advances in computing power, display technology, interfacing tools, graphics and image capture, immersive audio, haptics, wireless tracking, voice recognition, and VR authoring software. As these enabling technologies continue to evolve and allow for the development of more useful and usable structural VEs, the next important challenge will involve populating these environments with avatars. Indeed, Alessi et al.[1] has pointed out that until recently, virtual humans have mainly appeared in mental health scenarios to "serve the role of props, rather than humans" (p. 321). More believable virtual humans inhabiting VEs would open up possibilities for scenarios that allow for assessment and intervention strategies that leverage social interaction, naturalistic communication and more personal guidance/instruction. The existence of avatars in VEs could also serve to enhance realism that may in turn promote the experience of presence in VR.

VEs designed to target certain anxiety disorders might directly benefit from the presentation of virtual humans that are capable of some form of interaction, speech, and have the ability to recognize and emit typical nonverbal social communication via facial expressions and hand/body gesture cues. For example, early research in this area is investigating the use of video and computer graphics methods to render virtual humans for treatment of public speaking and social phobias,[9–13] as well as for a variety of social psychology applications.[14] The capacity to easily render avatars that are modeled after real persons in the users' everyday life might also create new possibilities for mental health applications that could utilize

more realistic role-playing strategies. Additionally, with advanced research developing avatars that are fueled with artificial intelligence (AI), the potential for more authentic real time interaction would further encourage exciting new MH application domains. For example, Rickel and Johnson[15] have reported success in the implementation of an AI avatar named "Steve" who serves the role as instructor for a virtual training environment targeting the operation and maintenance of equipment on a battleship. As well, similar avatar applications for testing and training tactical decision-making tasks such as crisis response in U.S. Army peacekeeping operations are under development.[16]

Concurrent with the emergence of this line of avatar research, we have seen a revolution in our technology for telecommunication. The explosion in information technology has fundamentally impacted the way we communicate and interact with each other over distances, with no deceleration in this trend expected in the near future. With these advances in information technology forms and processes, more naturalistic human interaction within multimodal distance communication between people and in the human–computer interface via avatar integration is becoming realizable. Thus far, widespread personal telecommunication has been effectively limited to speech (telephone) and graphic text-based (telegrams, letters, e-mail) forms. Early attempts at face-to-face dialog via video-conferencing approaches have yet to be generally accepted due to limitations in bandwidth, frame rate, and poor visual quality. In this regard, people often show a preference for, and report better communication and information processing efficacy from the exclusive audio channel available with a simple phone call. However, with continued advances in computer vision, tracking and graphic rendering, the possibility is within reach to electronically deliver more comprehensive forms of communication that include human facial and gestural components via avatar representations in shared virtual spaces. Early, more basic forms of this can be seen in the use of avatar-based chat rooms (e.g., Onlive Traveler, Active Worlds, The Palace, dada worlds).[17] It will soon be possible to use computer-generated avatars that serve to provide, in real time, dynamic facial and bodily representations of individuals who are communicating with each other electronically. This could support the delivery, in shared virtual environments, of more natural human interaction styles, similar to what is used in real life between people. These techniques could enhance communication and interaction by leveraging our natural expressing, sensing, and perceiving capabilities and offer the potential to model human–computer–human interaction after human–human interaction.[18] If the dynamic characteristics of facial and gestural actions can be rendered with some degree of fidelity to prototypic expressions seen in common types of implicit signaling, then avatars could serve to enhance communication, usability, and user-acceptance. The integration of facial movements that are concordant with voice expression will also be central to producing the suspension of disbelief required to support authentic and acceptable human interaction with avatar representations. This may produce realistic options for delivering multiperson electronic communication forms that support levels of engagement or sense of presence that could be useful for a wide range of mental health purposes. For example, when low-bandwidth conditions do not allow for real-time face-to-face video representation, effective avatar representations that support the communication and detection of emotional states of people who are interacting electronically would be vital for producing future teletherapy applications that are pragmatically usable and useful. In this regard, the emergence of avatar technology, while still in its infancy, offers considerable promise for developing a more comprehensive and desirable form of human telecommunication and interaction that could have significant impact on the delivery of human and psychological services.

However, attention to basic psychological and human factors methods/principles is required to determine the relative value of these applications and for promoting system effectiveness, efficiency, enjoyment, and safety. The myriad psychological variables that influence communication and interaction need to be fully considered well in advance to prevent ineffec-

tive or undesirable system development from taking resources that might be better spent on more thoughtful human-centered application development in this area. Human research to accomplish these goals will necessarily require a multicomponent approach ranging from analysis of molecular aspects of interaction, through to naturalistic studies of user acceptance of applications in more developed forms.

The work presented in this paper examines the synchrony and functional significance of facial actions as part of a larger research program on holistic communications analysis that will incorporate verbal parameters and bodily gestures. One general aim of the overall research program is to determine how naturalistic human facial expression is integrated within total communication output and under what conditions is this source of information essential to support effective and efficient communication. While the six universals of emotion are commonly targeted in the facial domain (happy, sad, surprise, fear, anger, disgust), researchers have also examined to a lesser degree such states as pain, fatigue, alertness, boredom, interest, attention, flirtation, deception, and pre- and post–problem solving.[4] These various emotions and states are of particular relevance for humans in the course of face-to-face social communication and the capacity for people to recognize them in computer generated avatars was the specific aim of the human evaluation component of this project. The current study compared human rater judgments of emotional expression between actual video clips of facial expressions and identical expressions rendered on a three-dimensional avatar (representing the same person) using a performance-driven facial animation (PDFA) system developed at the USC Integrated Media Systems Center. PDFA offers a means for low-bandwidth communication of high-resolution and high-fidelity visual representations of human faces and bodies. The current project explored the feasibility of sensing and reproducing a wide range of facial expressions with a PDFA system and presents comparative results of human judgments of emotion expression for both video-captured and animated facial renderings.

In order to test concordance of ratings of emotional expression between video and avatar facial delivery, we first had facial model subjects observe stimuli that were designed to elicit naturalistic facial expressions. The emotional stimulus induction involved presenting to subjects text-based, still image, and video clips that were previously rated or informally conjectured to induce facial expressions for the six universals of facial expression.[2] In addition, methods were developed to elicit states of attentiveness, puzzlement, and frustration. Videotapes of these elicited facial expressions that best represented prototypic examples of the above emotional states and three-dimensional avatar animations of the same facial expressions were randomly presented to 38 human raters. The raters used open-end, forced choice, and seven-point Likert-type scales to rate the emotional expressions. The forced choice and seven-point ratings provided the most usable data to determine concordance and these results are presented. The facial video and animation sequences on which the reported ratings were based, can be viewed on our website at www.USCAvatars.com/MMVR

MATERIALS AND METHODS

Technical background, procedures, and issues

Avatar animation can be produced by many animation methods. Our animation approach is described in Fidaleo et al.[19] and designed specifically for performance driven control. It has some features in common with morphing, which is often observed to achieve the most realistic looking animations; however, morphing requires precapture of all possible facial expression states and therefore is difficult to apply directly in performance driven animation systems. Our approach requires the production of a three-dimensional head model as described in Enciso et al.[20] The three-dimensional model for each subject was obtained just before or after the experimental videos were captured. Individual model quality varied due to differing facial characteristics and the limitations of our modeling system. Most cases required about 3–6 h to produce final models. Animations were produced by selecting and tracking features in the video sequences. Correspond-

ing points were selected on rigid portions of the face in the video and the three-dimensional model to facilitate head pose estimation. Neutral and peak expressions were marked as "key frames" and the expression evolution on the Avatar we interpolated in between. In some cases additional key frames were added to better capture the temporal expression evolution. Avatar textures were fixed to the neutral expression textures captured during the head-modeling phase. Technical difficulties and the research status of the dynamic textures and classification[19] precluded their use in this study. We hope to use the collected video and three-dimensional model data in future tests that include dynamic textures and classifications. We expect that a comparison can be made at that time to demonstrate the efficacy of adding textures.

In typical facial animation, a user must tune large numbers of control parameters to achieve a desired expression. The complexity and subtlety of human expressions make it tedious to generate realistic expression sequences, and even more difficult to produce the look and dynamics of a specific individual's expressions. The goal of performance-driven animation is to automatically control facial animations from video image sequences. Applications of PDFA include character animation for entertainment and communication in shared virtual spaces.

Many of the recent entertainment industry advancements in modeling and animation transfer directly to the communications application. In the communication context, the goals of a PDFA system are (1) to sense (track) information from an image sequence (live or recorded video); (2) to faithfully reproduce the original facial expressions; (3) to compute at an interactive rate; (4) to consume low bandwidth between sensing and animation sites; (5) to easily prepare person-specific animation models; and (6) to minimize user intervention. An example produced by our sensing and animation system is shown in Figure 1. The left image shows green dots that mark the features tracked by image analysis. The right image illustrates a person-specific three-dimensional model and its animation by deformations of the eyebrow regions and wrinkle textures on the forehead.

FIG. 1. Tracked features are shown as points in video images (**left**), producing animations that include deformations and skin wrinkles (**right**) created by volume morphing and appearance classification.

Developments in many research areas relate to PDFA, including computer vision, computer graphics, and image processing. Williams[21] presents one of the earliest systems, using retroreflective markers on the performer's face to help track facial motions. Terzopoulos and Waters[22] track contour features on eyebrows and lips to animate their physically based muscle structure of a synthetic character. Model-based image coding strives for effective compression of image sequences containing human faces.[23–25] Recent systems attempt to reproduce a real person's face motion from offline optical flow methods or real-time feature tracking.

The recent success of image and geometry morphing[26] shows that both deformations and texture manipulations are important for realism. In previous PDFA approaches, animation focuses mainly on geometry deformations. Texture is either not used or static, consequently losing subtle skin details such as wrinkles and creases in expressions. These features do not track well with the tracking methods used in previous PDFA methods, precisely because of their dynamic nature. In our use of volume morphing, the three-dimensional motion field is extracted from the sensing system and directly drives the animation. These motion vectors *are* the animation parameters—we do not need to convert the motion data to physical or abstract parameters, such as muscle or rational free form deformation (RFFD) parameters.[27] Model preparation is also simplified, since we do not associate any explicit animation parameters with the model, a translation

that often leads to a need for extensive manual tuning. Classification and three-dimensional textures produce the subtle expression components arising from wrinkles or eye blinks that are hard to synthesize with only geometry deformations. Since wrinkles are hard to track, we define an appearance-sensing area on the animation model and reproject it onto the input image after pose determination. Haar Wavelet features are used to classify the appearance sample in terms of how well it matches a previously acquired library of appearance samples. The Haar features are sensitive to translations of the appearance samples. Statistical texture analysis methods (e.g., entropy, variance) may make the classification less sensitive to appearance sample variations. Other image transforms (i.e., FFT, DCT) are also possible and may offer improved stability. Analysis and testing must be done to select the optimal classification approach. Due to time constraints for completion of this project, we opted to use the deformation methods with fixed textures. We believe that the future addition of three-dimensional textures will greatly improve the expressive communications of avatars.

Evaluation procedures

The methodology for evaluating human judgments of animation (avatar) versus video presentations of facial expression required three distinct phases:

1. Development and piloting of facial emotion expression induction stimuli
2. Presentation of selected induction stimuli to human models and the video recording of the resulting facial expressions
3. Presentation of selected facial expressions via animation and video to human raters for evaluation of facial expressive communication

Phase 1: Development and piloting of facial emotion induction stimuli

This phase required the development and psychometric analysis of emotion-evoking stimuli that would be used in phase 2 to induce facial expressions for later evaluation (in both video and animation formats) by human raters in phase 3. Originally we attempted to use text-based stimuli that were to be presented to human subjects with the request that they imagine that the event they were reading about had actually happened to them. In this regard, an example of a positive emotion induction statement would be: "You have just been given the next week off with pay." A negative emotion induction statement would be: "A good friend has had a bad car accident." A fear emotion induction statement would be: "You are alone at night and you hear a loud noise in your house." And, finally, a puzzlement induction statement would be: "The weather is a book."

To develop a psychometrically based group of emotion induction statements, we brainstormed 80 statements and asked 30 research subjects to rate each statement on a one to seven scale in terms of its emotion-inducing qualities. We examined this data and selected statements that had the highest mean ratings for emotion evocation, while also selecting from among these items, the statements that produced the smallest standard deviations. This procedure supported our effort to select items that represented the most evocative stimuli with the least psychometric variability.

We next presented these statements to a small pilot group of subjects ($n = 10$) and observed their facial expressiveness. Unfortunately, while this group reported that the stimuli were indeed evocative of happy, sad, puzzled, and fearful internal states, the stimuli appeared inconsistent in their capacity to produce intense facial expressions. At this point, we decided to supplement our stimulus package by using still and video images that we qualitatively determined to have the potency to produce facial expressions of sufficient visibility. Using informal observations of pilot subjects' facial expressions when presented with a large group of still and video stimuli, we chose 16 still images and 20 short video clips to present to facial models in phase 2. Due to the time pressure to commence facial acquisition in phase 2, we did not collect ratings of the evocative strength of the stimuli and instead incorporated a stimulus rating procedure with our facial models in phase 2. This poststimulus exposure self report method served to provide a measure of the evocative nature of each of

the chosen stimuli using open-ended, forced choice, and seven-point Likert scales. This data is being used now to begin the process of honing down the number of stimuli to be used in our future research. The stimuli used and latest data on their psychometric properties for induction are available from the first author. We also decided at this time to expand our range of facial/emotion induction stimuli/methods to target puzzlement, attentiveness, and frustration in addition to stimuli chosen to elicit the six universals[2] of facial expression (happy, sad, fear, anger, surprise, and disgust).

Phase 2: presentation of selected induction stimuli to human model subjects for induction and capture of facial expressions

In this phase, we presented 50 facial stimulus trials to human model subjects while we videotaped their evoked facial expressions. Facial model subjects were instructed to view the stimuli and to react as if the event portrayed was in fact occurring in real life. The stimuli consisted of the following:

1. Thirteen original text-based stimuli
2. Sixteen still images
3. Twenty video clips

Samples of these stimuli are available on the USC IMSC facial website (www.USCAvatars.com/MMVR), and a full collection of these stimuli with rating data is available from the first author. We also included an attention/frustration induction section in an effort to elicit facial expressions in these domains. This was attempted by requesting models to focus on the display screen and press the mouse button when an "X" appeared somewhere on the screen. Three trials were used for this in order to capture attentive faces. On the fourth trial, the "X" never appeared on the screen, and the trial was concluded either 1 min later or if the model asked the test administrator if something was wrong with the system. We hypothesized that, during this waiting period, models would get frustrated while waiting for a "respond" stimulus ("X") that actually never appeared, and that in the later stages of this induction we would capture facial expressions suggestive of frustration.

In this phase, we presented the stimuli to 15 male and female facial models whose ages ranged from 19 to 58. This was done in a quiet room with a camera (Sony XC999) positioned over a computer screen on which the induction stimuli were presented. The experimenter sat behind the models and operated the stimulus delivery procedures. Models were asked to respond to the stimuli as if the statement, image, or video was really occurring. Following each stimulus presentation, the models were asked to rate the stimulus on its capacity to evoke an emotion or state. This was done to collect data on the psychometric properties of the stimuli using the rating form in Figure 2.

This rating procedure (open ended, forced choice, and intensity ratings) will allow us to determine what stimuli are most useful for emotional induction purposes in our continuing work in this area. This is necessary for the

Stimulus 1
What emotion or state does this stimulus evoke in you?_____

Emotion Forced Choice (circle one):

Fear Anger Surprise Happy Sad Disgust Puzzlement Frustration Attentive

Ratings (1-7):

Fear____ Anger____ Surprise____ Happy____ Sad____ Disgust____ Puzzlement____ Frustration____ Attentive____

FIG. 2. Stimulus rating form for emotion-evoking stimuli used by facial models.

development of an emotion/face induction procedure that is more efficient via the use of a smaller number of stimuli with known targeted psychometric properties.

Following the standardized presentation of stimuli and facial expression capture, we also asked the models to feign facial expressions. This was done by saying to the models, "show me the face you make when you are happy" (as well as for all the other states that we were interested in). This was done to maximize the range of facial expressions that we had to choose from for presentation to human facial expression raters in phase 3. The same procedure was exactly followed for each of the models.

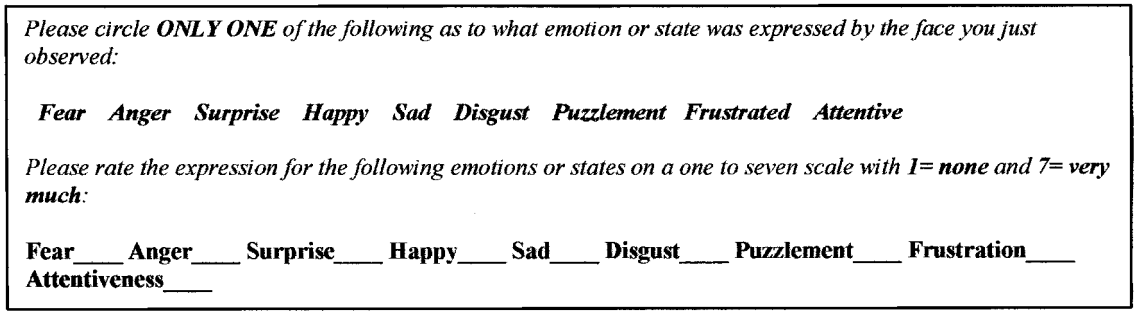
Following the acquisition of facial expressions from each model, we reviewed the video tapes and qualitatively selected facial expressions that were determined to best express the prototypic features of each of the following states: fear, anger, surprise, happy, sad, disgust, and puzzlement. Our efforts to produce attentive and frustrated faces were not successful and merely produced blank staring faces that were not incorporated into phase 3 human judgment evaluation. Two facial expression sequences (one male and one female) for each of the seven states were selected from eight of the original facial models. The other seven models produced unusable sequences due to a variety of reasons including excessive head movement during taping, lack of significant facial expressiveness, and difficulty in creating a three-dimensional model of one face due to presence of extraordinary facial hair (handlebar moustache).

Phase 3: presentation of selected facial expressions via video and animation to human raters for evaluation of facial expressive communication

This phase consisted of presenting 42 sequences, 3–5 sec in length, of facial expressions to a sample of 38 naïve human raters. The 42 sequences were comprised of three sets of the 14 facial expressions. The three sets were as follows:

1. Video clips of facial expressions as recorded in phase 2.
2. Performance-driven animations of the same facial expressions are presented on top of the video background (the AV condition). This mode of presentation preserved the background context of the model and included the some of the models' actual hair and clothing.
3. Performance-driven animations of the same facial expressions are presented alone on a gray background (the AN condition).

Thirty-eight students enrolled in an undergraduate USC computer science class served as raters of the facial sequences. Facial sequences in each of the three conditions (VID, AV, and AN) were presented in a stratified random order so that all possible equal orderings of each of the three types of stimuli were presented. This meant that, for one model's particular facial sequence, the VID condition would randomly appear within presentation of the first 14 stimuli, the AV condition would be randomly presented within the stimulus 15–28 grouping, and the AN condition would randomly appear in the 29–42 grouping. The next model's stimuli would be grouped similarly, but with the AN condition in the first 14, the


*Please circle **ONLY ONE** of the following as to what emotion or state was expressed by the face you just observed:*

Fear Anger Surprise Happy Sad Disgust Puzzlement Frustrated Attentive

*Please rate the expression for the following emotions or states on a one to seven scale with **1= none** and **7= very much**:*

Fear____ Anger____ Surprise____ Happy____ Sad____ Disgust____ Puzzlement____ Frustration____ Attentiveness____

FIG. 3. Facial expression rating form used by facial raters.

Table 1. Overall Forced Choice Ratings and Agreement for all Facial Sequences

Model	Face expression	Percentage corret			Percentage agree			First level percentage agreement	First level ratings
		VIDEO	AV	AN	VID/AV	VID/AN	AN/AV		
G.M.	Puzzled	18.2	3	9	0	0	2.8	Disgust Vid/AV = 24	Vid = 52; AV = 33
								Disgust Vid/AN = 24	Vid = 52; AN = 42
G.M.	Happy	62.9	88.6	84.4	60	50	78.1		
C.D.	Happy	100	100	100	100	100	100		
C.D.	Disgust	11.8	9.1	40	0	3.45	6.9	Puzzled Vid/AV = 25	Vid = 75; AV = 34
D.F.	Surprise	93.9	87	67.6	84	62.5	53.1		
I.Y.	Fear	18.2	12.9	3.0	3.5	3.2	0	Surprise Vid/AN = 13	Vid = 19; AN = 38; AV = 28
								Puzzled Vid/AV = 13	Vid = 28; AV = 21; AN = 23
I.Y.	Sad	29.4	55.8	53.1	18.2	22.6	41.9		
L.A.	Surprise	73.5	15.1	29.4	15.1	29.4	8.8	Vid-Sur/AV-Attn = 39.4	Vid-Sur = 73; AV-Attn = 64
								Vid-Sur/AN-Attn = 35.3	Vid-Sur = 73; AN-Attn = 56
R.E.	Angry	0	6.7	11.4	0	0	3.3	Vid-Puz/AN-Fear = 35.3	Vid-Puz = 71; AN-Fear = 47
								Vid-Puz/AV-Fear = 30	Vid-Puz = 73; AV-Fear = 40
S.R.	Angry	65.7	52.9	64.7	35.3	38.3	45.4		
S.R.	Fear	29	8.8	11.7	3.4	6.4	0	Surprise Vid/AV = 41	Vid = 45; AV = 72
								Surprise Vid/AN = 32	Vid = 48; AN = 61
S.R.	Puzzled	65.7	25	31.2	15.6	21.9	13.8	Vid-Puz/AV-Disgust = 28	Vid-Puz = 65; AV-Dis = 34
T.B.	Disgust	25	15.2	36.4	6.7	13.3	12.9	Surprise Vid/AV = 16	Vid = 40; AV = 37
								*Surprise Vid/AN = 0	Vid = 40; AN = 10
T.B.	Sad	14.7	17.1	38.2	5.9	9.1	8.2	Vid-Frus/AV-Disgust = 18	Vid-Frus = 41; AV-Dis = 32
								Vid-Frus/AN-Sad = 15	Vid-Frus = 41; AN-Sad = 36

Small differences in the percent values seen in different accuracy ratings for the same emotion reflect slight variations in sample sizes involved in ratings of each combination of variables and do not represent significant interpretable differences in overall ratings.

Model is the person appearing in the video. Face expression is the emotion that was expressed in the facial sequence based on phase 2 emotional stimuli/facial acquisition trials. Percentage correct VID is the percentage of raters who selected the emotion that was designated in the face expression column out of the forced choice alternatives following observation of the video presentation of the facial sequence. Percentage correct AV is the percentage of raters who selected the emotion that was designated in the face expression column out of the forced choice alternatives following observation of the animation impressed over the video background presentation of the facial sequence. Percentage correct AN is the percentage of raters who selected the emotion that was designated in the face expression column out of the forced choice alternatives following observation of the animation alone (gray background) presentation of the facial sequence. Percentage agree VID/AV is the percentage of raters that selected the emotion that was designated in the face expression column out of the forced choice alternatives for both the video and the animation impressed over the video background presentation of the facial sequence. Percentage agree VID/AN is the percentage of raters that selected the emotion that was designated in the face expression column out of the forced choice alternatives for both the video and the animation alone (gray background) presentation of the facial sequence. Percentage agree AN/AV is the percentage of raters that selected the emotion that was designated in the face expression column out of the forced choice alternatives for both the Animation impressed over the video and the animation alone (gray background) presentation of the facial sequence. First level % agreement is presented for sequences only when the percentage of rater agreement for the emotion that was designated in the face expression column was particularly low and exploration of the data found higher levels of agreement for other emotions/modes of presentation. For example, in the first entry (G.M.—puzzled), the percentage agree for both VID/AV and VID/AN were zero. However, 24% of raters agreed on both of these sequence pairings that the face presented a "disgusted" expression. In four cases—L.A./surprise; R.E./anger; S.R./puzzled; T.B./sad—data in this column is presented when subjects had noticeably higher agreement ratings across both emotion and mode of presentation (for example, Vid—Surprise/AV—Attention = 39.4). First level ratings present the percentage of raters that selected the emotion that was designated in the first level percentage agreement column for the specified emotion and mode of presentation.

Dis, disgust; Frus, frustration; Puz, puzzled; Attn, attentive.

Table 2. Overall Mean Ratings[a] for Targeted Emotion for Each Form of Facial Sequence

Model	*Face expression*	*VIDEO rating*	*AV rating*	*AN rating*
G.M.	Puzzled	2.32	2.11	1.96
G.M.	Happy	4.59	5.17	4.29
C.D.	Happy	6.4	5.2	6
C.D.	Disgust	3.14	2.43	3.13
D.F.	Surprise	6.66	5.28	4.5
I.Y.	Fear	3.03	2.04	1.78
I.Y.	Sad	3.9	3.75	3.06
L.A.	Surprise	5.77	2.70	3.16
R.E.	Angry	1.70	1.92	2
S.R.	Angry	5.31	3.53	4.24
S.R.	Fear	3	1.72	1.72
S.R.	Puzzled	5.15	3.07	3.61
T.B.	Disgust	2.90	2.07	2.83
T.B.	Sad	2.83	2.65	3.22

[a]Ratings: 1 = none; 7 = very much.

VID condition in the 15–28 grouping, and the AV condition would appear in the last grouping (stimuli 29–42). This manner of presenting the facial sequences was fully counterbalanced across models to control for possible order of presentation effects.

The raters were shown each sequence three times in succession. After the third presentation, raters were asked to judge the faces' expressiveness using a forced choice and rating format as presented in Figure 3.

While we ran the risk of lowering our accuracy and agreement ratings by having a large number of rating options (nine) and thereby spreading the range of potential variability, we opted for the inclusion of frustration and attentiveness so as to limit the possibility that we were corralling the ratings with a restricted range of response choices. This more conservative approach, while lowering the probability of higher agreement accuracy and concordance, offers better general coverage of response choices that is needed at this early stage of research to more rationally guide future efforts.

The 36 human raters had the following demographic characteristics: 29 males, seven females; average age of 23.8 (standard deviation of 4.4); average years' education of 16 (standard deviation of 2.1).

RESULTS

The data were analyzed by the following:

1. Comparing the accuracy of forced choice facial ratings with the predetermined targeted

Table 3. Significant Correlations for Ratings[a] of Targeted Facial Emotions between VIDEO and AV/AN

VIDEO	*Correlated with*	*r*	*p*	*n*
C.D.—happy (S33)	AN (S31)	0.39	0.02	35
G.M.—puzzled (S39)	AV (S11)	0.50	0.01	24
I.Y.—sad (S9)	AN (S41)	0.46	0.01	29
L.A.—surprise (S35)	AN (S28)	0.45	0.01	31
R.E.—anger (S20)	AV (S40)	0.76	0.001	24
S.R.—fear (S37)	AV (S10)	0.43	0.04	24
S.R.—fear (S37)	AN (S25)	0.53	0.005	25
T.B.—disgust (S13)	AN (S29)	0.42	0.04	25
T.B.—sad (S30)	AV (S6)	0.41	0.04	26
T.B.—sad (S18)	AN (S18)	0.41	0.04	27

[a]Ratings: 1 = none; 7 = very much.

emotion type for all three modes of presentation
2. Comparing agreement between the video (VD) presentation forced choice ratings and animated avatars pasted over the video clips (AV) containing background context
3. Comparing agreement between the video (VD) presentation forced choice ratings and animated (AN) avatars alone presented on a gray background
4. Comparing agreement on forced choice ratings between the animated avatars pasted over the video clips (AV) containing background context and animated (AN) avatars alone presented on a gray background
5. Determining the average numerical rating (1–7) of the targeted emotion for each model across subjects regardless of the emotion that they endorsed in the forced-choice rating condition. This gives an emotion attribute measure that provides separate information about perception of emotional expression in the facial sequences that is not picked up using the forced choice "all or none" categorical selection
6. Correlation of the numerical ratings (1–7) of the targeted emotion for each model. This resulted in 28 pairs of correlation's, 14 each for Video/AV and Video/AN, and this statistic produced a measure of rating concordance between conditions on the raters perception of the amount of the targeted emotion seen in the facial sequence, regardless of the emotion they selected in the forced-choice condition. This is essentially a relatedness analysis between the ratings that are presented in average form in number 5 above, and is more a measure of unity of perception regardless of strength of emotion.

These data comparisons for number 1–4 above will be presented in global form in Table 1 below. This will be followed by Tables 2 and 3, which contain data on the ratings discussed in numbers 5 and 6 above. The complexity of the data set as well as the variability across models, emotions and sequences, make attempting to produce one summary statistic across all ratings less than meaningful. Discussion of assets and limitations for each model's facial sequences will appear in separate breakouts of data for each model. It is this data that is being used to support our efforts to iteratively evolve our avatar development for our next series of studies in this area.

DISCUSSION

Due to the variability of the findings across sequences, we now present the results from each facial clip and provide specific commentary as to what was observed, as contrasted to the good overview of general findings in Tables 1–3.

Model	*Face expression*	*Percentage correct VIDEO*	*Percentage correct AV*	*Percentage correct AN*	*Percentage agree VID/AV*	*Percentage agree VID/AN*	*Percentage agree AN/AV*	*First level percentage agreement*	*First level ratings*
G.M.	Puzzled	18.2	3	9	0	0	2.8	Disgust Vid/ AV = 24 Disgust Vid/ AN = 24	Vid = 52; AV = 33 Vid = 52; AN = 42

The results from this set of stimuli indicate that raters had difficulty correctly detecting puzzlement in the model's face. This was seen particularly in view of the low 18% success rate on the video stimuli. Consequently, very low rating success or agreement was seen for puzzlement in all conditions for this sequence. Twenty-four percent of subjects perceived all three of the stimulus sequences to suggest disgust as seen in the first level agreement results. In these cases,

the Video produced higher ratings than both AV and AN conditions, although the AN condition was only 10% less accurate.

Model	Face expression	Percentage correct VIDEO	Percentage correct AV	Percentage correct AN	Percentage agree VID/AV	Percentage agree VID/AN	Percentage agree AN/AV	First level percentage agreement	First level ratings
G.M.	Happy	69.9	88.6	84.4	60	50	78.1		

In this condition, it appears that the animation may have actually enhanced perception of the features of the face that conveyed happiness in this model. The AV and AN conditions were significantly better with 26% and 22% points higher detection, respectively. Agreement ratings between Video and the AV and AN conditions were pretty good, but were limited by the lowered accurate perception within the video condition. It should be noted that the agreement rating could never be higher than the lowest percent correct value for any of the matched components, 62.9% in this case. In view of that, the agreement ratings appear noteworthy and are also supported by the similarity of the mean numerical ratings for this sequence across conditions as seen in Table 2.

Model	Face expression	Percentage correct VIDEO	Percentage correct AV	Percentage correct AN	Percentage agree VID/AV	Percentage agree VID/AN	Percentage agree AN/AV	First level percentage agreement	First level ratings
C.D.	Disgust	11.8	9.1	40	0	3.45	6.9	Puzzled Vid/ AV = 25	Vid = 75; AV = 34

Again, we see limited recognition accuracy between the targeted expression and forced choice emotion selections. This led to very low agreement ratings. Although similarity of the overall mean ratings for these sequences can be seen in Table 2, they indicate a lack of perceived intensity for the disgust aspect. Many raters saw these sequences as better representing a puzzled expression with a high level of endorsement on the video condition and moderate agreement observed with the AV condition.

Model	Face expression	Percentage correct VIDEO	Percentage correct AV	Percentage correct AN	Percentage agree VID/AV	Percentage agree VID/AN	Percentage agree AN/AV	First level percentage agreement	First level ratings
C.D.	Happy	100	100	100	100	100	100		

This represents the ideal match up whereby 100% concordance between all conditions was found. To be fair, this particular model's happy face was quite obvious and this judgment was rather easy. It does indicate that at the current level of the technology, it is possible to have success with the more obvious expressions and that the distracting characteristics of the avatars did not impair judgments.

Model	Face expression	Percentage correct VIDEO	Percentage correct AV	Percentage correct AN	Percentage agree VID/AV	Percentage agree VID/AN	Percentage agree AN/AV	First level percentage agreement	First level ratings
D.F.	Surprise	93.9	87	67.6	84	62.5	53.1		

This model displayed a good rendition of the prototypic surprise face (raised eyebrows, eyes open wide, wrinkled forehead, mouth open). The video conveyed this quite well (although lacking in the forehead wrinkles) and the AV condition, although producing a lower percentage, the difference between these groups was not statistically significant. The AN condition had a lower hit rate, but this may have been in part due to its position in the order of stimulus presentation—it was the first animation presented to the naïve raters and subjects may have been distracted from the rating task due to the perceived novelty of the animation sequence.

Model	*Face expression*	*Percentage correct VIDEO*	*Percentage correct AV*	*Percentage correct AN*	*Percentage agree VID/AV*	*Percentage agree VID/AN*	*Percentage agree AN/AV*	*First level percentage agreement*	*First level ratings*
I.Y.	Fear	18.2	12.9	3	3.5	3.2	0	Surprise Vid/ AN = 13 Puzzled Vid/ AV = 13	Vid = 19; AN = 38; AV = 28; Vid = 28; AV = 21 AN = 23

The results from this set of stimuli indicate that raters were unable to correctly detect fear in the model's face particularly in view of the 18% success rate on the video stimuli. This was roughly equivalent to the AV condition (the difference between the two was not significant). Relatively poorer ratings were seen in the avatar alone condition (AN). As well, the nature of the model's facial expression may have produced considerable variability as evidenced by the low agreement ratings among conditions and by the higher recognition and agreement ratings that were seen for ratings of surprise and puzzled. The relatively equal recognition ratings (for surprise and puzzled) coupled with minimal agreement, suggest that this model's expression was too vague to be recognizable on a consistent basis.

Model	*Face expression*	*Percentage correct VIDEO*	*Percentage correct AV*	*Percentage correct AN*	*Percentage agree VID/AV*	*Percentage agree VID/AN*	*Percentage agree AN/AV*	*First level percentage agreement*	*First level ratings*
I.Y.	Sad	29.4	55.8	53.1	18.2	22.6	41.9		

The results for this sequence suggest that the animated conditions produced better recognition of the sad components. However, this could be an artifact of the slight exaggeration in the rendering of the mouth in the downturned position in the animated expressions.

Model	*Face expression*	*Percentage correct VIDEO*	*Percentage correct AV*	*Percentage correct AN*	*Percentage agree VID/AV*	*Percentage agree VID/AN*	*Percentage agree AN/AV*	*First level percentage agreement*	*First level ratings*
L.A.	Surprise	73.5	15.1	29.4	15.1	28.4	8.8	Vid-Sur/ AV-Attn = 39.4 Vid-Sur/ AN-Attn = 34.3	Vid-Sur = 73; AV-Attn = 64 Vid-Sur= 73; AN-Attn = 56

This sequence produced significantly better recognition in the video condition (73.5%) and less correct recognition and limited agreement in the animation conditions. This may have to do

with limitations in the animation eye representation. On the video, the classic eyes wide open look of surprise is evident with a large amount of white appearing between the pupil and the bottom of the upper eyelid. However, this component as well as the slight opening of the mouth, was not effectively rendered on the avatar. Also, a selective emphasis of the eyebrow's upward movement in the animation conditions was not seen in the video may have affected the ratings on this expression. Instead, raters saw the animations as more representing attentiveness than surprise as is evidenced in the first order agreement findings linking the surprise video to the attentive animations by nearly 40%. This was also supported in Table 2, where 1–7 intensity ratings for the video were nearly twice the value of the AV/AN conditions. The animations were unable to represent the intense dynamic features of surprise sufficiently and this may have caused raters to perceive the expression to represent the more mild arousal state of attentiveness.

Model	*Face expression*	*Percentage correct VIDEO*	*Percentage correct AV*	*Percentage correct AN*	*Percentage agree VID/AV*	*Percentage agree VID/AN*	*Percentage agree AN/AV*	*First level percentage agreement*	*First level ratings*
R.E.	Angry	0	6.7	11.4	0	0	3.3	Vid-Puz/ AN-Fear = 35.3 Vid-Sur/ AV-Fear = 30	Vid-Puz = 71; AN-Fear = 47 Vid-Sur = 73; AV-Fear = 40

This sequence conveyed none of the anger content in the video condition and very little with the animation conditions. Hence, the agreement between video and animation was zero. When presented with this set of facial sequences, raters had higher best guess agreement and ratings for puzzled and fear. Perhaps the absence of signaling features around the mouth seen in prototypic displays of anger affected these ratings. Similar weak ratings in the 1–7 format for anger is evident in Table 2, while a high significant correlation for Vid/AV (0.76) in Table 3 merely suggests a lot of unity by the raters on their agreement that the face did not present anger components in an effective manner.

Model	*Face expression*	*Percentage correct VIDEO*	*Percentage correct AV*	*Percentage correct AN*	*Percentage agree VID/AV*	*Percentage agree VID/AN*	*Percentage agree AN/AV*	*First level percentage agreement*	*First level ratings*
S.R.	Angry	65.7	52.7	64.7	35.3	45.4	3.3		

Raters correctly determined anger moderately the same across sequences. Fairly equal agreement ratings were found as well with no significant first order agreements on any of the other rating choices. Intensity ratings for anger (Table 2) were similar but with a significant edge seen with the video sequence (over one point). This video advantage on 1–7 intensity ratings in the presence of equal forced choice ratings and agreement may indicate that the subtlety of anger facial components that include forehead wrinkles added to video ratings. These wrinkles were absent in the texture mapping of the animations, but will be possible to render in future iterations of these animation creation procedures.

Model	Face expression	Percentage correct VIDEO	Percentage correct AV	Percentage correct AN	Percentage agree VID/AV	Percentage agree VID/AN	Percentage agree AN/AV	First level percentage agreement	First level ratings
S.R.	Fear	29	8.8	11.7	3.4	6.4	0	SurpriseVid/ AV = 41 SurpriseVid/ AN = 32	Vid = 45; AV = 72 Vid = 48; AN = 61

On this set of sequences, somewhat better accuracy in rating for fear was seen on the video condition. Very little agreement between the video and animation conditions was seen due to the limited accuracy seen in animation conditions. This is also seen in Table 2 with relatively low intensity ratings for fear. Raters significantly endorsed these facial stimuli to represent surprise rather than fear with reasonable agreement ratings evidenced in the first level percentage agreement column. This was seen notably in the AV condition with 72% of raters selecting this emotion choice. This can perhaps be explained in that prototypic fear faces contain many of the components included in surprise. To go beyond facially expressed surprise to expressed fear, additional eye components are needed that were not rendered in the animations. Thus, subjects tended to rate the surprise components better in the animation sequences due the absence of full-blown rendering of the eye area movements that might better signal anger.

Model	Face expression	Percentage correct VIDEO	Percentage correct AV	Percentage correct AN	Percentage agree VID/AV	Percentage agree VID/AN	Percentage agree AN/AV	First level percentage agreement	First level ratings
S.R.	Puzzled	65.7	25	31.2	15.6	21.9	13.8	Vid/Puz/ AV-Disgust = 28	Vid = 65; AV-Dis = 34

Fairly high accuracy ratings were found in the Video condition with relatively fair to moderate recognition seen in the animation conditions. This was also seen in the Table 2 intensity ratings. Puzzlement is not viewed as a universal of facial expression, and we included it in this project in an effort to go beyond what had been done previously with this highly relevant for communication state. As with the other puzzled condition (G.M.), raters may have had difficulties detecting this state since a puzzled face may contain many idiosyncratic differences between expressers, and may be better noticed when someone is actually familiar with the expresser's face. The subtlety of this facial expression will require better eye and wrinkle rendering before it may be reliably recognized in animation-based communication.

Model	Face Expression	Percentage correct VIDEO	Percentage correct AV	Percentage correct AN	Percentage agree VID/AV	Percentage agree VID/AN	Percentage agree An/AV	First level percentage agreement	First level ratings
T.B.	Disgust	25	15.2	36.4	6.7	13.3	12.9	SurpriseVid/ AV = 16 SurpriseVid/ AN = 0	Vid = 40; AV = 37 Vid = 40; AN = 10

Video and animation accuracy was similar, but low overall with weak agreement ratings and low intensity scores (Table 2). There was a lot of variability across all raters for all the choices as seen in the first level percentage agreement for surprise also producing low values, although surprise had the highest forced ratings for video and AV conditions. Indeed, the rating accuracy/agreement scores were low and equally diluted across the various nine choice alternatives. This may be due again to limited rendering of wrinkle textures in the nose and forehead areas that help to signal disgust.

Model	*Face expression*	*Percentage correct VIDEO*	*Percentage correct AV*	*Percentage correct AN*	*Percentage agree VID/AV*	*Percentage agree VID/AN*	*Percentage agree AN/AV*	*First level percentage agreement*	*First level ratings*
T.B.	Sad	14.7	17.1	38.2	5.9	9.1	8.2	Vid-Frus/AV-Disgust = 18 Vid-Frus/AN-Sad = 15	Vid-Frus = 41; AV-Dis = 32 Vid-Frus = 41; AN-Sad = 36

Interestingly, the AN condition produced significantly higher ratings for sadness compared to video and AV conditions. This sequence accentuated the downturned mouth similar to what was seen in the other sadness animation condition (I.Y.) and this may have served to increase rating accuracy. However, low intensity ratings for sadness in Table 2 and the perception of frustration and disgust by a large number of raters, suggest a lack of specificity in the signaling features of these animations. Again, better textures around the eyes and forehead might produce more specific animation ratings.

In conclusion, we have used these results to provide information as to what is needed to advance our research program in the development of performance driven facial animation methods that can produce more useful avatar representations of human facial expression and communication. It is hoped that the details of our methodology for initially addressing these issues along with the results and multimedia examples on the accompanying website (www.USCAvatars.com/MMVR) will be of value to others who are interested in this area. As we advance mental health applications and standard communication forms into the 21st century, the design and development of avatars that can represent key human characteristics will continue to be a vital and important research endeavor!

ACKNOWLEDGMENTS

This research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152 and by the U.S. Defense Advanced Research Projects Agency (DARPA).

REFERENCES

1. Alessi, N.E., & Huang, M.P. (2000). Evolution of the virtual human: from term to potential application in psychiatry. *CyberPsychology and Behavior*, 3:321–326.
2. Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In: Wagner, H., & Manstead, A. (Eds.), *Handbook of social psychophysiology.* Chichester, U.K.: Wiley, pp. 143–164.
3. Ekman, P., & Rosenberg, E.L. (1997). *What the face reveals*. New York: Oxford University Press.
4. Russell, J.A., & Fernandez-Dols, J.M. (1997). *The psychology of facial expression.* Cambridge, U.K.: Cambridge University Press.

5. Iverson, J.M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, 396:238.
6. Darwin, C. (1872). *The expression of the emotions in man and animals*. Reprinted 1998. New York: Oxford University Press.
7. Goldin-Meadow, S., Alibali, M.W., & Church, R.B. (1993). Transitions in concept acquisition: using the hand to read the mind. *Psychological Review*, 100:279–297.
8. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J.G. (2001). Emotion recognition in human–computer interaction. *IEEE Signal Processing Magazine*, 18:32–80.
9. Anderson, P., Rothbaum, B.O., & Hodges, L.F. (2000). Social phobia: virtual reality exposure therapy for fear of public speaking. Presented at the Annual Meeting of the American Psychological Association, Washington, DC.
10. Pertaub, D.P., & Slater, M. (2001). An experiment on fear of public speaking in virtual reality. Presented at the 9th Annual Medicine Meets Virtual Reality Conference, Newport Beach, CA.
11. Rizzo, A.A., Neumann, U., Enciso, R., Fidaleo, D., & Noh, J.Y. (2001). Judgment of emotional expression in 3D facial avatars created using performance driven animation: towards populating mental health VR applications with believable virtual human representations. Presented at the 9th Annual Medicine Meets Virtual Reality Conference, Newport Beach, CA.
12. Riva, G. (2001). The VEPSY Project: virtual reality in clinical psychology. Presented at the 9th Annual Medicine Meets Virtual Reality Conference, Newport Beach, CA.
13. Wiederhold, B., Riva, G., Choi, Y.H., & Wiederhold, M. (2000) Virtual reality exposure therapy in the treatment of panic disorder with agoraphobia. Presented at the 34th Annual Convention of the Association for the Advancement of Behavior Therapy, New Orleans, LA.
14. Blascovich, J., Loomis, J., Beall, A.C., Swinth, K.R., Hoyt, C.L., & Bailenson, J.N. (2001). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry* (in press).
15. Rickel, J., and Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: perception, cognition, and motor control. *Applied Artificial Intelligence*, 13:343–382.
16. Swartout, W., Hill, R., Gratch, J., Johnson, W.L., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J., Rickel, J., Thiébaux, M., Tuch, L., and Whitney R. (2001). Toward the Holodeck: integrating graphics, sound, character and story. In: *Proceedings of the 5th International Conference on Autonomous Agents* (in press).
17. Damer, B. (1998). *Avatars!* Berkeley, CA: Peachpit Press.
18. Turk, M., & Robertson, G. (2000). Perceptual user interfaces. *Communications of the ACM*, 43:33–34.
19. Fidaleo, D., Noh, J., Kim, T., Enciso, R. & Neumann U. (1999). Classification and volume morphing for performance-driven facial animation. Presented at the International Conference on Digital and Computational Video.
20. Enciso, R., Li, J., Fidaleo, D., Kim, T., Noh, J., & Neumann, U. (1999). Synthesis of 3D faces. Presented at the International Conference on Digital and Computational Video.
21. Terzopoulos, D., & Waters K. (1993). Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE PAMI*, 15:569–579.
22. Williams, L. (1990). Performance-driven facial animation. *Proceedings of the Siggraph*, 235–242.
23. Li, H., Roivainen, P., & Forchheimer, R. (1998). 3-D motion estimation in model-based facial image coding. *IEEE PAMI*, 15:545–555.
24. Eisert, P., & Girod, B. Analyzing facial expressions for virtual conferencing. *IEEE CG & A*, 18:70–78.
25. Zhong, J. (1998). Flexible face animation using MPEG4/SNHC parameter streams. *Proceedings of the IEEE ICIP*,
26. Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., & Salesin, D.H. (1998). Synthesizing realistic facial expressions from photographs. *Siggraph Proceedings*, 75–84.
27. Esher, M., & Thalmann, N. M. (1997). Automatic 3D cloning and real-time animation of a human face. *Proceedings of IEEE Computer Animation.*

Address reprint requests to:
Integrated Media Systems Center/Andrus Gerontology Center
University of Southern California
3715 McClintock Ave., MC-0191
Los Angeles, CA 90089

E-mail: arizzo@mizar.usc.edu