An Expressive System for Endowing Robots or Animated Characters with Affective Facial Displays

Craig Latta, Nancy Alvarado, Sam S. Adams, Steve Burbeck

The IBM Thomas J. Watson Research Center; P.O. Box 218, Yorktown Heights, N.Y. 10598
crl@watson.ibm.com, nanalv@us.ibm.com, ssadams@us.ibm.com, sburbeck@us.ibm.com

Abstract

This paper describes an expressive system designed to link a model of emotion with socially appropriate expressive behaviors in an intelligent agent or robot. The system was developed to provide an affective interface to Joshua Blue, a computer simulation of an embodied mind that includes emotion and is designed to learn and interact effectively in social environments. The advantages of this implementation include the ability to accommodate different emotion models, to easily specify different expressive behaviors, and to directly inspect, continuously monitor, and intervene to change underlying affective states as they occur. The system currently uses a real-time 3D facial avatar with a simplified anatomical model, but may be readily modified to direct the expressive behaviors of more complex animated characters or robots.

1 Introduction

Expressive behavior in intelligent agents or robots serves two important purposes. First, it permits an entity to interact effectively with other expressive beings in social contexts. In many applications, this may be the primary function of the expressive behavior. Naturalness and realism, social effectiveness, and meaningfulness of displays to human observers are important concerns. Second, expressive behavior provides information to the designer about the internal states of the software. This latter function is essential to proving assertions that an emotion model has influenced the behavior of an agent or robot. In more complex systems with greater behavioral autonomy, the better the information conveyed by expressive behavior, the better a designer can understand the impact of affect on a machine's cognition, which is less directly observable.

Additionally, we felt that the utility of an expressive system would be greatly enhanced if it were able to: (1) support different models of emotion; (2) be configured to control a wide variety of behaviors with minimal human effort; and (3) allow the designer to easily inspect, monitor, or change the internal affective states of the underlying emotion system, through the same expressive interface. We designed this expressive system for use with Joshua Blue, a computer simulation of an embodied mind that must ultimately function in a wide variety of environments, virtual and physical (see Alvarado, Adams, Burbeck & Latta, 2001). These requirements made versatility an important design goal. However, we believe our approach may also enhance simpler, less general systems by making it easier to incorporate future improvements in animation techniques or robot technology that will make expressive systems more socially effective.

2 System Overview

The expressive system consists of the following components: (a) an emotion model; (b) an expressive behavior model; (c) a 3D facial avatar; (d) an emotional state display; (e) tools for editing emotional expressions. Each of these parts is described below, for the current implementation. The specifics of the avatar, emotion model and expressive behavior model can be changed with minimal impact on the remaining parts of the system, which make internal system values open to inspection and propagate changes between subsystems, all with fluid interactivity. The system is implemented in Squeak, a dynamic, open-source object system.

2.1 Emotion Model

The emotion model in the Joshua Blue system incorporates the two dimensions of affective experience, valence (appraisal as good or bad) and arousal (level of excitement), originally described by Osgood (1966) and later presented as an emotion model by Russell (1991; 1997). Since these dimensions are essentially abstractions derived from the dominant aspects of subjective experience, they are hypothesized to be present in any emotional phenomenon. This means that a variety of emotion models, from Ekman's (1992) basic emotions to simple pain-versus-pleasure models, can be readily mapped into the same two-dimensional space.

If reduction to two dimensions is undesirable, such as when finer distinctions are needed, multidimensional maps can be created in the same manner as was used to generate this simple two-dimensional system. While this complicates visualization, the system is not limited to using 2-D pointers but can support a 3-D controller. An alternative approach

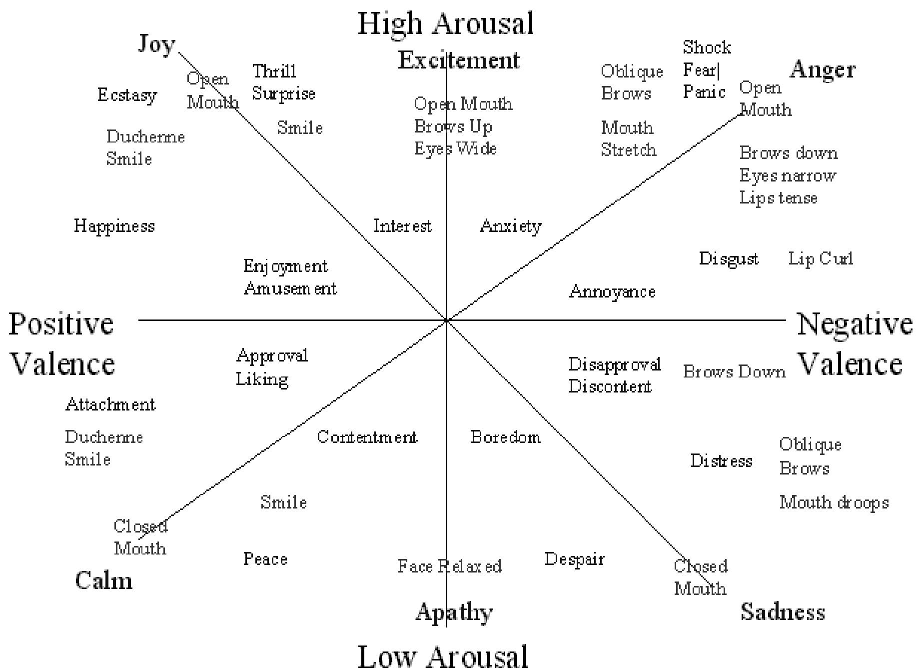

Figure 1. A Mapping of Discrete Emotions and Facial Expressions onto the Two Dimensions of Affective Experience

includes specifying higher dimensionality in a series of 2-D maps, each including different specified pairs of dimensions.

In this implementation, regions of the affective space are partitioned by ten equal increments along each axis. Discrete emotions are mapped to the space, as shown in Figure 1. The terms shown in Figure 1 are descriptive and suggest what occurs facially in each region. The actual specification of facial movement is defined by the expressive behavior model, described below.

2.2 Expressive Behavior Model

Facial expressions for animating the avatar were specified using the definitions of action units corresponding to basic emotional states (Ekman and Friesen, 1978). Action units for additional emotional states were identified based on empirical studies by Snodgrass (1992) and Alvarado (1996). We determined contraction values for muscle groups in each expression using action units defined by the Facial Action Coding System (Ekman & Friesen, 1978). Intensities of muscle movement (expressed as slider position) were specified for each step in the two-dimensional affective space of the emotion model. A separate map was defined for each muscle group. This specified a facial expression for each state possible in the Joshua Blue's emotion model. Basic expressions concurrent with extreme emotional arousal have been tested for reliable interpretation by Ekman and Friesen. To the extent that our versions conform to the specified basic expressions, we expect to obtain similar results in judgment studies. The less intense expressions and those expressions corresponding to states not defined as basic emotions are being tested for interpretability by human observers. Following that validation, all expressions will be tested for naturalness of transition and believability.

2.3 Facial Avatar

The current expressive display employs an animated facial avatar, rendered from a three-dimensional mesh of external facial skin points deformed under the influence of underlying musculature, as described by Waters (1987) (see Figure 2). We adapted a Smalltalk implementation of Waters' avatar (Notarfrancesco, 2001) for use as an expressive display driven by the affective states of the Joshua Blue system. A user may manually modify the tensions of 12 major facial muscle groups, by opening and manipulating sliders. While not anatomically precise, this simplified muscle model permits simulation of reasonably believable affective facial expressions. It is the first of three main elements in the expressive system's user interface, the other two being the emotional state display and modification tools.

2.4 Emotional State Display

The user may view and manipulate the procession of Joshua Blue's emotional state through the affect space over time with a specialized monitor. This monitor samples Joshua Blue's current emotional state at a user definable rate, and displays it as a point in the space. Previously displayed samples persist on the display, but change color

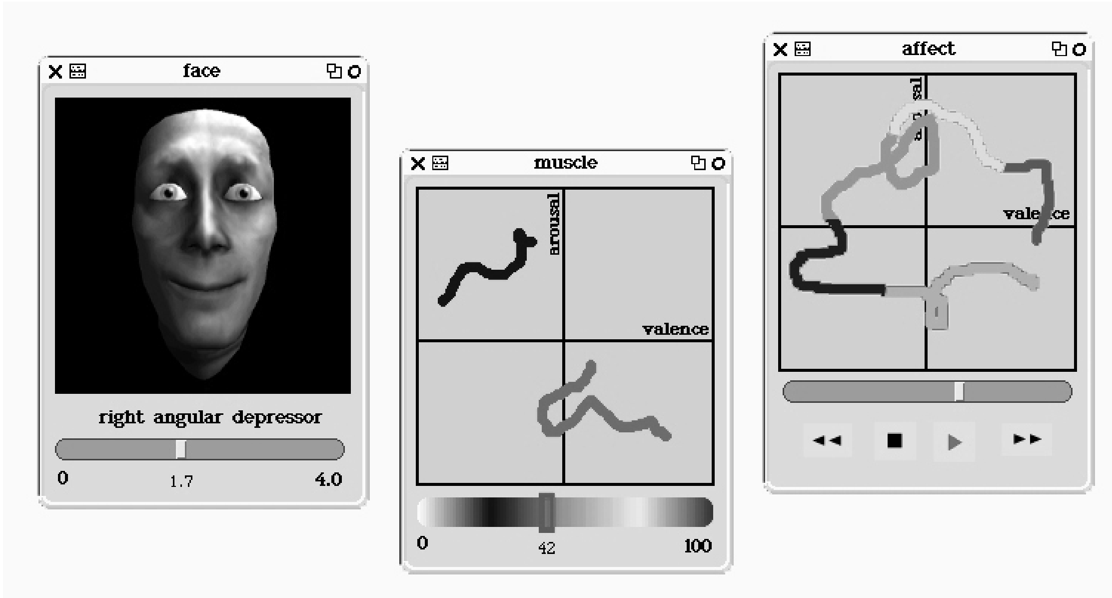

Figure 2. Facial Avatar, Affective Muscle Response Editor, and Emotional State Display

over time as they age. Over time, the monitor accumulates a "trace" of the changes in Joshua Blue's emotional state.

The monitor provides transport controls, so that the user may see the expression corresponding to a previous emotional state, and control the sampling of subsequent states. When the user clicks on a point in the space, the monitor highlights that point in outline, and the avatar takes on the expression corresponding to the associated emotional state. A slider below the affect space display provides random access to particular points in the trace's history. Finally, buttons at the bottom of the monitor provide traditional playback control, toggle the recording of new samples, and enable storage and loading of traces.

2.5 Editing Tools

In addition to rendering Joshua Blue's current emotional state as a facial expression, the avatar provides access to tools for editing muscular responses. As the user moves the pointer over the avatar, the system indicates the muscle groups muscles in the vicinity, and displays their names. The user may then select a particular muscle group, which displays its current contraction value along with a slider showing the range of possible contraction values.

Having selected a muscle group, the user may also open an editor on the mapping between affect and muscle contraction. This editor shows a particular muscle group's contraction response for each point in the affect space, indicated by color. The user may rapidly specify the response for many points by selecting a contraction value from a palette and painting its associated color in the affect space. Direct manipulation of the muscle in the avatar also selects contraction values in the muscle editor; the user is free to interpret the palette selection as a number or as a visible expressive cue. The user may open any number of muscle editors concurrently, supporting quick composition of complex facial expressions over the affect space.

3 Squeak Implementation Benefits

The Squeak Smalltalk system (see Ingalls, Kaehler, Maloney, Wallace & Kay, 1997) affords us a great deal of flexibility on several fronts. Squeak's message-oriented organization allows us to establish a strong separation of concerns, such that interdependencies between major functional parts are few. As a result, it is straightforward for us to change the emotion model, expressive behavior, and human interface subsystems independently. Since Squeak is a "late-binding" system, in which objects' message responses are determined at runtime, we may make such changes during operation. This allows us to incorporate new insights while continuing to monitor expressions and record emotional states.

The Squeak architecture is completely open. This gives us the ability to make fundamental operational changes, such as might be required when profiling performance. It also permits complete understanding of the system, both by us and by other researchers, and promotes a broad consistency throughout the system. Largely due to its openness, Squeak has an active worldwide development community. Several contributions from this community have saved us a great deal of development time, and we have been able to contribute significant works in return.

We have noticed most of these features in other systems, but only in Squeak have we found them in the same system.

4 Problems Resolved

Our attempts to produce believable facial expressions using this method led to two concerns. First, how is a transition from one affective state to another accomplished in a natural-looking manner? Second, how does the emotion model's intensity relate to the intensity of muscle movements in the facial avatar? Answers to these questions do not exist in the facial expression literature and thus must be addressed empirically. This model provides a means for doing so. An advantage of using an animated facial avatar to address such questions, rather than photographed human facial activity, is that transitions between expressions and levels of intensity can more readily be manipulated in stimuli presented to observers for evaluation. The perceived naturalness of manipulated expressions can be used to predict what is likely to be typical in human facial behavior, as well as to dictate choices about how to implement natural-looking expression in artificial faces.

Based upon previous research (reviewed by Alvarado & Jameson, 2001), it seems likely that the implementation of facial activity will depend upon whether the goal is to simulate naturalistic human expression or to produce meaningful displays for human observers. It is clear to us from many studies (c.f., Ekman & Rosenberg, 1997) that there is far from a straightforward connection between what is produced on the face in various contexts and the meaning that is interpreted by observers from such facial displays. To date, only Ekman and Friesen's basic expressions have produced consensual interpretations among human observers, while the array of behavior produced on the face is much wider than that subset of basic expressions. Wierzbicka (1999) and others have called for a component-based semantics of the face, but as yet there is no empirical support for component-based meaning or any context-independent meaning beyond Ekman and Friesen's basic facial expressions. Further, whatever might be expressed by humans in naturalistic settings is mediated by a variety of factors not typically included in expressive models, including display rules, instrumental intentions, personality, and coping style (Alvarado & Harris, 2001).

Because there is no straightforward relationship between human internal states, expressive behavior and communicative intentions, optimizing expressive behavior to attain naturalism does not necessarily maximize communication effectiveness (see Russell & Fernandez-Dols, 1997; Alvarado & Harris, 2001). One solution to this problem is to include a model of social intention that incorporates expressive display rules and instrumental goals. Otherwise, designers may be required to select one or the other goal, depending on the intended use of their agent or robot.

5 Test Results

We plan to test whether the system functions as intended, and whether the avatar displays recognizable versions of Ekman and Friesen's basic expressions. We will then use the system to explore the issues raised concerning, transitions, tradeoffs between naturalism and interpretability, and the relationship between intensity of movement and intensity of affect, as perceived by observers. This will likely constitute an ongoing research program, made possible by the flexibility of this expressive system.

6 Related Work

Several others have made implementations of expressive behavior, directed by various models of emotion. Further, the use of topographic mapping to link an emotion or motivational space with specified behaviors is also not new. For example, Breazeal's (2001) robot Kismet is directed by a topographic emotion space based upon Ekman's basic emotions, with defined behaviors corresponding to each region of a motivational space. A thorough review of the many projects with similarity to this one is beyond the scope of this paper. The main contribution of the current work is the flexible association of these components into a system where internal states can be inspected and compared with facial activity and where emotion models and the expressive output of an avatar or robot can be readily changed without redesign or extensive programming.

7 Conclusion

While the relationship between cognition and expression is not yet well understood, we believe that many useful elements for its pursuit are in hand. We have created an environment for fluid, interactive evaluation of these elements, which we hope will yield new insights into their composition and development.

References

Alvarado, N. (1996). Congruence of Meaning Between Facial Expressions of Emotion and Selected Emotion Terms. *Motivation and Emotion, 20*, 33-61.

Alvarado, N., Adams, S. S., Burbeck, S. & Latta, C. (2001). Integrating Emotion and Motivation into Intelligent Systems. Manuscript in submission.

Alvarado, N. & Harris, C. (2001). Facial Expression and Coping Style. Manuscript in submission.

Alvarado, N. & Jameson, K. (2001). Varieties of Anger: The Relation Between Emotion Terms and Components of Anger Expressions. Manuscript in submission.

Breazeal, C. 2001. *Designing Sociable Machines*. The MIT Press.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition & Emotion*, 6, 169-200.

Ekman, P. & Friesen, W. (1978). *Facial Action Coding System* (FACS). Consulting Psychologists Press.

Ingalls, D., Kaehler, T. Maloney, J., Wallace, S., Kay, A. (1997). Back to the Future: The Story of Squeak, a Practical Smalltalk Written in Itself. *Proceedings of the 1997 ACM Conference on Object-Oriented Programming Systems, Languages, and Applications*, 318-326.

Notarfrancesco, L. (2001). An adaptation of the Waters facial animation system. Unpublished source code.

Osgood, C.E. (1966). Dimensionality of the Semantic Space for Communication Via Facial Expression. *Scandinavian Journal of Psychology, 7*, 1-30.

Russell, J. (1991). The contempt expression and the relativity thesis. *Motivation and Emotion, 15*, 149-168.

Russell, J. (1997). Reading Emotions From and Into Faces: Resurrecting a Dimensional-Contextual Perspective. In J. Russell & J.M. Fernandez-Dols (Eds.), *The Psychology of Facial Expression* (pgs 295-320). Cambridge University Press.

Snodgrass, J. (1992). Judgment of Feeling States From Facial Behavior. Reported by Russell (1997). Unpublished doctoral dissertation, University of British Columbia.

Waters, K. (1987). A muscle model for animating three-dimensional facial expressions. *IEEE Computer Graphics (SIGGRAPH'87), 21(4).*

Wierzbicka, A. (1999). *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge University Press.