

Research Article

Multispace Behavioral Model for Face-Based Affective Social Agents

Ali Arya¹ and Steve DiPaola²

¹ Carleton School of Information Technology, Carleton University, Ottawa, ON, Canada K1S5B6

² School of Interactive Arts & Technology, Simon Fraser University, Surrey, BC, Canada V3TOA3

Received 26 April 2006; Revised 4 October 2006; Accepted 22 December 2006

Recommended by Tim Cootes

This paper describes a behavioral model for affective social agents based on three independent but interacting parameter spaces: knowledge, personality, and mood. These spaces control a lower-level geometry space that provides parameters at the facial feature level. Personality and mood use findings in behavioral psychology to relate the perception of personality types and emotional states to the facial actions and expressions through two-dimensional models for personality and emotion. Knowledge encapsulates the tasks to be performed and the decision-making process using a specially designed XML-based language. While the geometry space provides an MPEG-4 compatible set of parameters for low-level control, the behavioral extensions available through the triple spaces provide flexible means of designing complicated personality types, facial expression, and dynamic interactive scenarios.

Copyright © 2007 A. Arya and S. DiPaola. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Chuck Jones, the cocreator of such legendary animated characters as Bugs Bunny, Daffy Duck, and the Road Runner, once said [22]: “Believability. That is what we were striving for.” The history of animation, traditional or computer-generated, has shown that the most successful animated characters are not necessarily those who have been geometrically realistic, but those that are believable in behavior. As many researchers in the area of social agents have noticed [4, 5, 29], this believability of characters (i.e., acting in a realistic and “natural” way) is a key element in allowing viewers/users to relate to the agent. In our opinion, such believability depends, mainly, on proper behavioral modeling. Another aspect of behavioral modeling is the creation of non-scripted actions. A strong behavioral model allows an animated character such as a social agent to follow certain rules or high-level scripts, and define and create proper details of actions based on any dynamic situation with no need to design those details in advance.

Although many researchers have proposed behavioral models for social agents [4, 5, 17, 19, 26, 33, 36], the following essential features seem to require further improvements.

- (1) Theoretical base in behavioral psychology.
- (2) Proper parameterization to simplify the model configuration.
- (3) Scripting language specially designed for agents.
- (4) Independence of behavioral components such as tasks, personality, and mood.

In this paper, we describe the behavioral model used in our facial animation system, iFACE (interactive face animation—comprehensive environment) [2]. iFACE uses a parameterized approach where the behavior is controlled through three separate but interacting parameter spaces: knowledge, personality, and mood (see Figure 1). They are not organized as layers on top of each other; they are “parallel” which means that each one can operate (and be controlled) independently while at same time interact with the other ones. Knowledge is the primary space where all action and configuration scripts are processed. Personality and mood can be controlled by these scripts and personality itself can affect mood. A fourth parameter space, geometry, forms the visual foundation of the system with low-level parameters such as size and location of facial features. A hierarchical set of geometrical parameters provides an efficient and unified set of controls for facial actions,

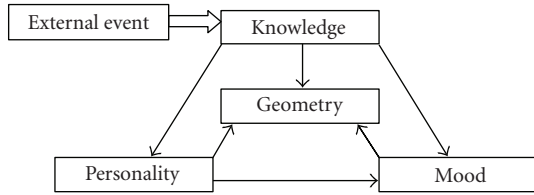


FIGURE 1: behavioral model parameter spaces.

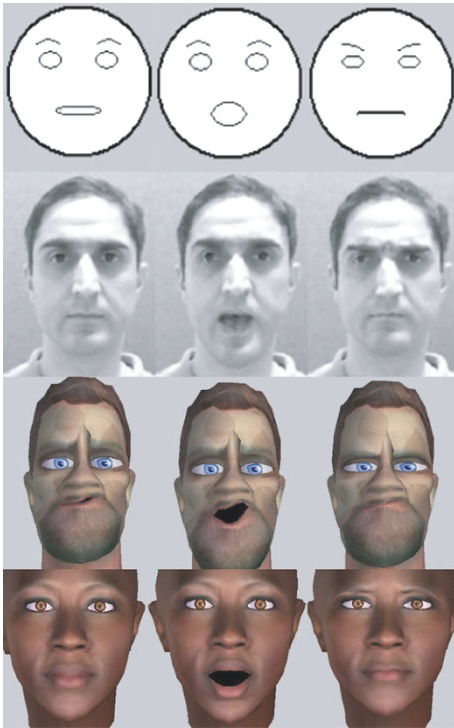


FIGURE 2: Sample animated heads from iFACE, featuring neutral, talking, and frowning states (columns 1 to 3, resp.) of 2D cartoonish, 2D photorealistic, 3D cartoonish, and 3D realistic faces (rows 1 to 4, resp.).

independent of the 2D or 3D head data type, as shown in Figure 2.

Knowledge encapsulates the tasks to be performed and general rules of behavior that are independent of the character. A specially designed XML-based language is used for knowledge space. Personality and mood are based on parameterized models in behavioral psychology and represent the characteristics and emotional state of a specific individual. Personality is related to the long-term traits such as typical head movements and mood controls short-term emotional states visualized by facial expressions.

The principal concept in our research is that parameterization allows animators and designers to create new geometries, personality types, and emotional states without being involved in technical details. For example, changing the affiliation and dominance [40] parameters can easily create new personalities, and since the parameters are associated to facial

actions, this new personality type already has proper facial actions. The existing systems either do not use well-defined and scientifically accepted parameters or have not associated the parameters to facial actions properly (e.g., random or ad-hoc selection of actions compared to our system that is based on user studies with the aid of behavioral psychologists). So our main contributions, compared to the existing research that we will review later, are the following.

- (1) The only XML-based face-specific language compatible with MPEG-4 with dynamic decision-making and temporal constructs.
- (2) Associating facial actions to the perceived personality based on user studies. Facial actions have been extensively studied with regards to emotions but not personality.
- (3) Linking facial actions to personality and emotion parameters rather than “personality types” and “emotional states” themselves. As we will see, this will cause facial actions that are more “perceptually valid” when creating new and combined types and states.
- (4) A layered geometry model that allows animation parameters and design files to be applied to a variety of data types (see Figure 2) due to abstraction and hiding details.
- (5) A unified model encapsulating all required features in one framework.

In Section 2, we review some of the related research in the area of behavioral modeling for social agents. Sections 3 to 7 discuss our proposed behavioral model in detail. Two example applications of iFACE system and its behavioral model are the subject of Section 8, and some concluding remarks are presented in Section 9.

2. RELATED WORK

2.1. Agent and multimedia languages

The facial action coding system (FACS) was the earliest approach to systematically describe facial action in terms of small action units (AUs) such as left-eye-lid-close and jaw-open [19]. The MPEG-4 standard [6] extended this idea and introduced face definition parameters (FDPs) and face animation parameters (FAPs). FDPs define a face by giving measures for its major parts and features such as eyes, lips, and their related distances. FAPs on the other hand, encode the movements of these facial features. Together they allow a receiver system to create a face (using any graphics method) and animate that face based on low-level commands in FAPs.

Synchronized multimedia integration language (SMIL) [12] is an XML-based language designed to specify temporal relationships of components in a multimedia presentation, especially in web applications. SMIL can coexist quite suitably with MPEG-4 object-based streams. SMIL animation is a newer language (<http://www.w3.org/TR/smil-animation>) based on SMIL, which is aimed at describing animation pieces. It establishes a framework for general animation but neither of these two provides any specific means for facial

```

<vhml>
<person disposition='angry'>
<p>
First I speak with an angry voice,
<surprised intensity='50'>
then I change to look surprised.
</surprised>
</p></person>
</vhml>

```

FIGURE 3: An example of VHML script.

animation. There have also been different languages in the fields of virtual reality and computer graphics for modeling computer-generated scenes. Examples are virtual reality modeling language (VRML, <http://www.vrml.org>) and programming libraries like OpenGL (<http://www.opengl.org>).

These languages are not customized for face animation, and do not provide any explicit support for it. The absence of a dedicated language for face animation, as an abstraction on top of FACS AUs or MPEG-4 FAPs has drawn attention to the development of markup languages for virtual characters [1, 15, 30, 35]. Virtual human markup language (VHML) [30] is an XML-based language for the representation of different aspects of “virtual humans,” that is, avatars, such as speech production, facial and body animation, emotional representation, dialogue management, and hyper and multimedia information (<http://www.vhml.org>). It comprises a number of special-purpose languages for emotion and facial and body animation. In VHML, timing of animation elements in relation to each other and in relation to the realization of text is achieved via the attributes “duration” and “wait.” A simple VHML document is shown in Figure 3.

Multimodal presentation markup language (MPML) [35] is another XML-based markup language developed to enable the description of multimodal presentation in a web browser, based on animated characters (<http://www.miv.t.u-tokyo.ac.jp/MPML/en>). It offers functionalities for synchronizing media presentation (reusing parts of the synchronized multimedia integration language, SMIL) and new XML elements such as <listen> (basic interactivity), <test> (decision making), <speaking> (spoken by a TTS system), <move> (to a certain point at the screen), and <emotion> (for standard facial expressions). MPML addresses the interactivity and decision making not directly covered by VHML/FAML, but both suffer from a lack of explicit compatibility with MPEG-4 (XMT, FAPs, etc.).

2.2. Personality and perception

behavioral psychologists have studied human personality and its models and parameters for quite a while. Many personality models have been proposed, and one of the most notable examples is the big five or five-factor model [21, 39]. The big-5 model considers five major personality

dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism (OCEAN). Modeling personality as an N -dimensional space allows for navigating through the personality space by changing one parameter along each independent dimension. Although successful in many aspects, the five dimensions in the Big-5 model are (1) not independent enough and (2) hard to visualize. This results in the model being hard to use for animated characters needing user-friendly and controllable personality parameters. Wiggins et al. [40] have proposed another personality model based on two dimensions: affiliation and dominance (Figure 4). They show that different personality types can be considered points around a circular structure formed in two-dimensional space. The smaller number of dimensions allows them to be controlled more effectively and independently. Two parameters are also easier to visualize, perceive, and understand.

The perception of personality type and traits based on observation has long been a subject of research in behavioral psychology [8–10, 25]. Unfortunately, this research has not focused on facial actions, and has primarily considered the observation of full-body behaviors. Also, mainly due to logistical reasons, the observations have been mostly limited to photographs or few dynamic actions. High-quality and controllable animated characters have not been available to psychology researchers. As Borkenau et al. [9, 10] have illustrated, viewers can achieve relatively stable perceptions using short videos. Creating videos of live actors playing many different and configurable actions, however, can be expensive and difficult.

Among facial actions, the universal facial expressions of emotions (joy, sadness, anger, fear, surprise, and disgust, as described by Ekman [18]) is the only group whose effect on the perception of personality has been investigated. Knutson [25] reported on the effect of facial expression of emotions on interpersonal trait inference based on Wiggins’ model. He concludes that viewers attribute high dominance and affiliation to individuals with happy expressions, high dominance, and low affiliation to those with angry or disgusted expressions, and low dominance to those with fearful or sad expressions. Borkenau and Liebler [10] have reported one of the few studies which explicitly associated body gestures and behaviors as visual cues to the perception of personality. They have also considered audio and visual (static and dynamic) cues but facial actions were not a major focus.

2.3. Believable social agents

Badler et al. [4] proposed one of the first personality models or agents to control behavior (in their case, locomotion) based on certain individual characteristics. The proposed architecture includes a physical movement layer, a state machine for behavioral control, and an agent layer that configures the parameters of the state machine. The model is not linked to any theoretically sound personality model, and is a general architecture for configurable behavioral controllers. Other researchers (e.g., [29]) have also proposed methods for modeling agent behaviors. Among them, Rousseau and

Hayes-Roth [36] define behavior as a combination of personality, mood, and attitude. The idea of separating independent components of behavior can be very helpful in designing autonomous agents. Funge et al. [20], on the other hand, propose the idea of hierarchical modeling, which includes behavioral and cognitive modeling layers at the top.

Another approach in behavioral modeling for agents includes associating different facial actions with certain states and events. Cassell et al. [13] propose a method for automatically suggesting and generating facial expressions and some other gestures based on the contents of the speech. In a later work, Cassell et al. [14] propose a comprehensive toolkit with a dedicated language for generating movements based on speech, through certain configurable rules. King et al. [24] and Smid et al. [38] (among others) provide more recent examples of the automatic generation of facial actions (primarily expressions) based on speech. The main weakness of all these works is that the facial actions are (1) usually limited to the expressions, and (2) speech, and not a personality model, is the base for facial actions. A system to suggest facial actions based on personality settings has not been fully investigated.

Associating facial actions with personality requires a reasonably adequate personality model for the agent, and a thorough study of the effect of facial actions on the perception of personality. The latter, as mentioned before, has not been done properly yet, but the former has been the subject of some recent works. Kshirsagar and Magnenat-Thalmann [26] propose a multilayer personality model. It is, more precisely, a multilayer behavioral model that includes layers of personality, mood, and emotions on top of each other. Every layer controls the one below it, with the facial actions and expressions at the bottom. The model allows definition of parameters at each level to individualize the agent. At the personality level, it utilizes the Big-5 model with five parameters. The following observations can be made regarding this system.

- (i) The general issues with Big-5.
- (ii) Hierarchical dependence of emotional states to personality. The likelihood of transition between emotions can be a personality parameter, but emotional state should be also independently controllable regardless of personality.
- (iii) Lack of direct link between facial actions and personality. Speech content or a probabilistic belief networks are used to control facial actions, which may not be enough. Ideally, the facial actions (e.g., the way an agent moves his/her head or raises eye brows and how frequently he/she does it) need to be controlled by a well-defined personality type, entirely or together with speech and likelihood settings (see Section 5 for more details).
- (iv) Unnecessary separation of moods and emotions (see Section 6 for a more detailed discussion of moods and emotions).

Models proposed by Egges et al. [17] and Pelachaud and Bilvi [33] follow similar ideas. The latter uses a two-dimensional model similar to Wiggins et al. [40] for personality (called

performatives) and also separates them from emotions as two independent components activating facial actions through a belief network. The high-level personality parameters are associated to facial actions based on limited observation and arbitrary settings, rather than a well-performed user study. On the other hand, the facial actions are not limited to speech and can occur even when the agent is not talking, but they have to be set explicitly where desired, while the ideal situation is to define them as part of a personality to be activated autonomously.

2.4. Facial expression of emotions

Russell [37] has mapped emotional states onto a two-dimensional space controlled by arousal and valence. The detailed study of facial actions involved in the expression of the six universal emotions [18] has helped the computer graphics community to develop realistic facial animations. Yet the rules by which these facial expressions are combined to convey more subtle information remain less well understood by behavioral psychologists and animators. This lack of a strong theoretical basis for combining facial actions has resulted in the use of ad-hoc methods for blending facial expression in animations [27, 31, 32, 34]. These methods are usually based on a “weighted average” of facial actions caused by each expression. They are therefore computationally tractable, but the question of their “perceptual” and “psychological” validity has not yet been answered.

3. MULTISPACE BEHAVIORAL MODEL

In the previous section, we reviewed some of the related works in the area of behavioral modeling. Considering the strengths and weaknesses of these approaches, the authors have concluded that the following features are required for a comprehensive agent behavior model. It appears that none of the existing approaches provides a complete collection of them.

- (i) A behavioral model needs to be based on scientific findings and models in behavioral psychology.
- (ii) The model should have easy-to-visualize parameters for character design.
- (iii) The model should consist of separate modules for different behavioral aspects such as knowledge, personality traits, and emotions.
- (iv) These behavioral modules should be independent but able to interact with each other and with the underlying geometry.
- (v) The parameter spaces and the scripting language should be MPEG-4 compatible.
- (vi) The language has to support dynamic actions and interactive scenarios through proper decision making and event handling.

Based on these guidelines, and especially using the suggested model by Rousseau and Hayes-Roth [36], we propose a multispace behavioral model formed with four independent but interacting parameter spaces: geometry, knowledge,

personality, and mood. We replace Rousseau and Hayes-Roth’s attitude component with knowledge which includes tasks to be performed and rules of behavior and can provide a better control over agent actions. We also define these four components as parameter spaces formed with specific easily adjustable parameters. These parameter spaces are used in our comprehensive facial animation system, iFACE [2].

iFACE geometry is a hierarchical model that isolates details such as vertex/pixel information from higher-level constructs such as feature and head component, so that animation can be designed and controlled independent of the underlying geometry type. The main advantages of our knowledge space are specially designed language for facial animation, support for decision making and dynamic actions, and high-level timing control. The personality and mood spaces use current findings in behavioral psychology to relate personality traits and emotional states to facial actions, to cause the perception of intended personality type or create the perceptually valid expression. Unlike Kshirsagar and Magnenat-Thalmann’s model [26], they perform in total independence from each other (i.e., parameters set separately), but the personality parameters can also define some mood-related aspects of behavior such as the likelihood of transition between emotional states which is in fact a personality based issue (although mood settings can override personality settings temporarily). The mood space does not have any direct effect on personality settings which is again based on “real world” relationships between personality and mood. These spaces are explained in the following sections.

4. HIERARCHICAL GEOMETRY SPACE

Head/face components and regions allow grouping of head data into parts that perform specific actions together (e.g., resizing the ears or closing the eye), which results in isolating details from higher-level commands. This is a key concept in designing an efficient head model. By defining different layers of abstraction on top of actual head data (2D pixels or 3D vertices), each exposing proper interfaces for possible commands, we allow programmers/animators to access only the desired level of details, as illustrated in Figure 5. At the same time, this hierarchy allows changes in lower-level modules (e.g., the way movement of lip corner affects neighbouring points) without any change in the general behavior of higher-level parameters (e.g., an expression can still result in lip corner stretching without a need to know how that happens). Possibility of working with different types of 2D and 3D head data, using the same parameters, is another advantage of such isolation.

Features are special lines/areas that lead facial actions, and feature points (corresponding to MPEG-4 parameters) are control points located on features. Only the lowest level (physical point) depends on the actual (2D or 3D) data. iFACE geometry object model corresponds to this hierarchy and exposes proper interfaces and parameters for client programs to access only the required details for each action. iFACE authoring tool (iFaceStudio) allows users to select feature points and regions-of-influence for them. Each level of

geometry accesses the lower levels internally, hiding the details from users and programmers. Eventually, all the facial actions are performed by applying MPEG-4 FAPs to the face.

5. PARAMETERIZED PERSONALITY SPACE

The primary objective of personality modeling is to make it possible for the agent to perform facial actions that cause the viewer to perceive certain personality types, as intended by the character designer. As discussed in Section 2, Wiggins’ circumplex model provides an effective parameterized framework for modeling and defining personality types. On the other hand, the effect of dynamic facial actions on personality perception has not been studied properly, partly due to difficulty of hiring actors to record variety of head and face movements [3, 8–10]. Using a realistic facial animation system can help researchers to perform a wider range of experiments.

In order to design a perceptually valid personality model (i.e., one that initiates actions that most likely cause the intended personality perception in viewers), we performed a four-step process.

- (1) Define sets of facial actions and expressions that may affect personality perception (visual cues).
- (2) Run experiments with a large enough user base to study the effect of these visual cues on personality perception.
- (3) Associate visual cues to personality parameters, affiliation and dominance.
- (4) Create a model that defines parameterized personality profiles and initiates proper facial actions based on that.

Table 1 shows the visual cues selected at step 1 and the results of our experiments with 31 undergraduate students at the Department of Psychology, University of British Columbia, Vancouver, Canada. Details of experiments have been published in an earlier paper [3].

The personality model controls strength and the timing of initiating facial actions based on personality settings. We give each personality parameter three linguistic values: low, medium, and high. For example, for parameter dominance these correspond to dominant, neutral, and submissive, as shown in Figure 4. After performing the experiments, visual cues are associated with each one of these parameter values, to form sets like the following: $C_{i,j} = \{c_{i,j,n}\}$, where $c_{i,j,n}$ is the n th visual cue associated with the j th value of i th parameter.

Each visual cue is defined as an individual MPEG-4 FAP [6] or a combination of them. If p_i is the value of i th personality parameter ($i = 0$ or 1), $v_{i,j}$ (the strength of the j th linguistic values of that parameter) will be calculated using a fuzzy membership function based on p_i . These strengths are then used to activate the visual cues to certain levels:

$$a_{i,j,n} = v_{i,j} \times m_{i,j,n} \quad (1)$$

$a_{i,j,n}$ and $m_{i,j,n}$ are activation level of the visual cue (or the related FAP) and its maximum value, respectively.

TABLE 1: Affiliation and dominance scores for facial actions (min = -5, max = 5).

Facial action	Affiliation	Dominance
Joy	4.7	2
Sadness	0.2	-0.2
Anger	-2.6	0.6
Fear	2	-0.8
Disgust	0.9	1
Surprise	2.9	-1
Contempt	-5.7	1.4
Neutral	0.8	-0.8
Slow turn	1.7	1.2
Slow tilt	0.9	0.2
Slow nod	-0.5	-3.1
Slow blink	2.5	-0.7
Slow avert	0.1	-0.7
Slow one brow	-0.1	0
Slow two brows	4.2	-0.9
Fast turn	2.5	1.7
Fast tilt	2.1	1.9
Fast nod	-0.6	-2.8
Fast blink	2.7	-0.8
Fast avert	-0.2	-2.9
Fast one brow	-1.6	3.3
Fast two brows	3.8	0.9
Head rest down	-0.1	-1.4
Head rest side	0.4	-3.4

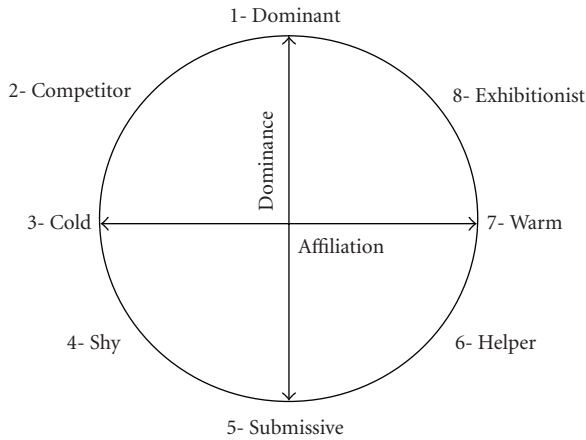


FIGURE 4: Wiggins' personality circumplex.

The timing for activating visual cues is also set in the personality profile. It can be random, periodic, or based on speech energy level. The content of the speech can also be used as suggested by other researchers [38]. Some measures of speech energy can be calculated by analyzing the speech signal. Two strength thresholds of impulse and emphasis can be defined for this energy. Different visual cues (or different versions of them with varied maximum values) can be associated with these thresholds. Once a threshold is reached,

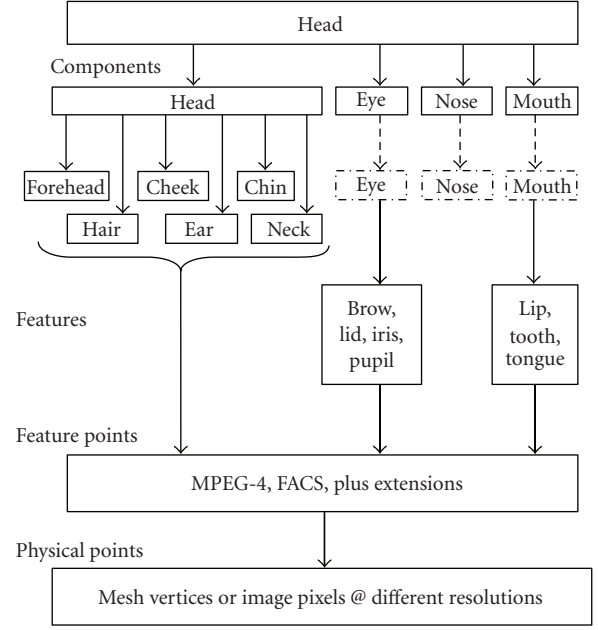


FIGURE 5: Hierarchical facial geometry.

one of the associated cues that matches the agent personality is randomly selected and activated based on the value of $a_{i,j,n}$.

6. PARAMETERIZED MOOD SPACE

The distinction between moods and emotions has been discussed by many researchers. The major differences seem to be duration and cause, and the emotions are believed to be more external and visible [7]. Due to complicated relation between moods and emotions, and between moods and visual appearance, it is hard to create mood parameters (independent of emotions) that can effectively and clearly control the facial actions. Some researchers [26] have tried to define such parameters for an agent's mood in which the result is simply three types of moods (bad, normal, and good) which only change the likelihood of transition between emotions and have no extra functionality (e.g., direct effect on facial actions).

In our model, we consider emotions and mood part of one parameter space called mood. This space controls the emotional state through two parameters (see Figure 6), and also includes probability settings for random or event-based transition between emotions. With better understanding of how moods affect emotions and other visual aspects, we hope to separate moods and emotions into two parameter spaces, but at this time a simple "likelihood setting" does not seem enough for such separation.

The emotional state of the agent can be set in three different ways.

- (1) Explicitly in the course of an action (see FML scripts).
- (2) Randomly/periodically as configured in the personality profile.

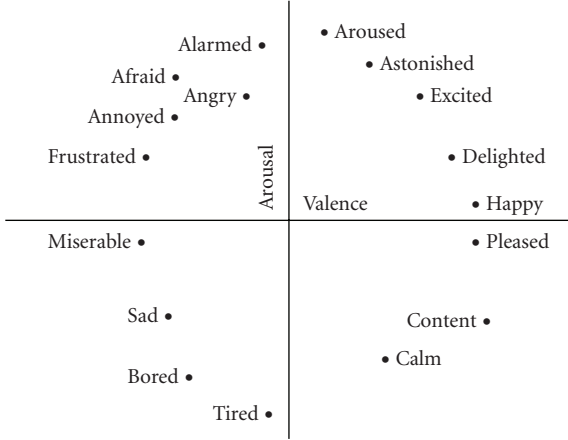


FIGURE 6: Parameterized mood space [37].

- (3) Randomly/periodically as configured in the mood space which overrides personality setting.

In either case, the mood (or emotional state) is set by specifying a universal emotion and its level of activation, or by setting the values of two mood parameters: valence and arousal (see Figure 6). Ekman has described the facial actions associated with the expression of universal emotions in detail [18]. For example, the expression of joy involves tightening of eyelids, raising cheeks, lowering eyebrows slightly, and wrinkles around the eyes especially the corners. For single universal emotions, we activate the associated actions based on the level of emotional state. For blending two expressions, we differentiate between two cases: transition from one expression to another, and activation of two expressions at the same time, that is, a combined expression.

The facial actions for transitions are simply the weighted average of the source and destination expressions:

$$a_i = k \times a_{i,s} + (1 - k) \times a_{i,d},$$

$$k = \frac{N - f}{N}. \quad (2)$$

N is the number of frames to create for the transition, f is the current frame, a_i is the activation of i th action at frame f , and $a_{i,s}$ and $a_{i,d}$ are the activation of that action in source and destination expressions.

The combined expressions are created by either selecting two universal expressions, or by setting arousal and valence parameters. In the first case, the activation levels of two expressions are first mapped into a pair of arousal-valence parameters. The resulting values of arousal and valence are then used to activate facial actions associated with each parameter as shown in Table 2. These facial actions are selected by analyzing the Ekman’s description of universal expressions and their facial actions, and by clustering similar actions based on arousal and valence parameters.

These two cases are illustrated in Figure 7. In this figure, (a) and (b) show surprise and anger expressions. The middle frame for transition (c) is between (a) and (b). We see that

due to the raised mid lower lip in anger (target of linear interpolation), the middle of the mouth closes while the sides are not closed yet. This may be acceptable for a transition but in case of a combined expression like aroused, it is better to locate the source and target on arousal-valence map, and then find the proper (perceptually valid) facial actions for a point between them. This is shown in (d) where the jaw is slightly dropped, upper and lower eyelids are raised a little, and brows are slightly lowered and drawn together. The effectiveness of the parameter-based expression blending compared to the simple weighted average method is the subject of an extensive user study in the University of British Columbia. The details of this study will be presented in a separate paper.

7. FACE MODELING LANGUAGE

7.1. Design ideas

To describe the tasks to be performed, the timing, and event handling mechanism, a special-purpose language for facial animation has been designed for iFACE that performs proper configuration and controls the main sequence of actions. The need for such a high-level language, as opposed to low-level parameters such as those in MPEG-4, can be shown using an example. Figure 8 illustrates a series of facial actions. A “wink” (closing eye lid and lowering eyebrow), a “head rotation,” and a “smile” (only stretching lip corners, for simplicity). These actions can be described by the following MPEG-4 FAPs.

Wink: FAP-31 (raise-l-i-eyebrow),
FAP-33 (raise-l-m-eyebrow),
FAP-35 (raise-l-o-eyebrow),
FAP-19 (close-t-l-eyelid).

Head rotation: FAP-49 (head rotation -yaw).

Smile: FAP-6 (stretch-l-lipcorner),
FAP-6 (stretch-r-lipcorner).

Although simple and powerful, the use of MPEG-4 FAPs for behavioral description lacks the following features.

- (1) Parameters at facial component level (e.g., one eye wink instead of four FAPs).
- (2) Proper timing mechanism (e.g., duration and dependencies).
- (3) Event handling and decision-making.

Face modeling language (FML) [2] is an XML-based language designed for facial animation. It combines MPEG-4 compatibility with higher-level features such as those mentioned above. Also, FML is independent of the underlying animation system. The actions of Figure 8 can be done by an FML script such as lines shown in Figure 9 (elements are discussed later).

FML defines a timeline of events (Figure 10) including head movements, speech, and facial expressions, and their combinations. Temporal combination of facial actions is

TABLE 2: Sample facial actions and the expressions that include them.

Action	Expressions	Valence	Arousal
Brows drawn together	Fear, anger	Low	High
Brows lowered	Joy, anger, disgust	—	High
Brows raised	Fear, surprise	—	High
Brows-inner raised	Sadness	Low	Medium
Eye-corner wrinkled	Joy	High	—
Eye-lid-lower raised	Joy, sadness	—	Medium/low
Eye-lid-lower tensed and raised	Fear, anger	Low	High
Eye-lid-upper lowered	Joy, sadness	—	Medium/low
Eye-lid-upper raised	Fear, anger, surprise	—	High
Jaw dropped	Surprise	Medium	High
Jaw thrust forward	Anger	Low	High
Lip-corners lowered	Sadness	Low	Low
Lip-corners raised	Joy	High	Medium
Lip-lower raised	Sadness	Low	Low
Lips pressed and narrowed	Anger	Low	High
Lips stretched	Joy, fear, anger	—	Medium

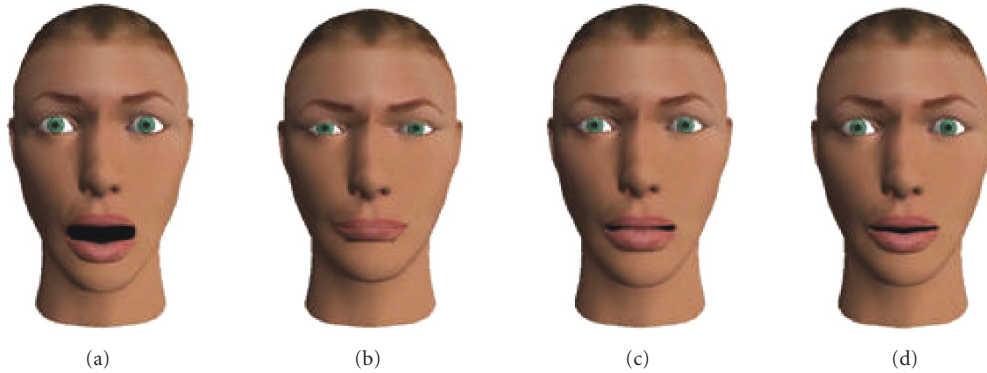


FIGURE 7: Samples of expression blending: (a) surprise, (b) anger, (c) transition between surprise and anger, and (d) blending based on valence and arousal and their associated facial actions.

done through time containers which are XML tags borrowed from SMIL (other language elements are FML specific). Since a face animation might be used in an interactive environment, such a timeline may be altered/determined by a user. So another functionality of FML is to allow user interaction and in general event handling (decision making based on external events and dynamic generation of scenarios).

7.2. FML document structure

An FML document consists, at the higher level, of two types of elements: *model* and *story*. A *model* element is used for defining face capabilities, parameters, and initial configuration. This element groups other FML elements (model items) such as configuration data and predefined actions. A *story* element, on the other hand, represents the timeline of events in face animation in terms of individual actions (FML action elements). The face animation timeline consists of

facial activities and their temporal relations. These activities are themselves sets of simple “moves.” Sets of these moves are grouped together within “time containers,” that is, special XML tags that define the temporal relationships of the elements inside them. FML includes three SMIL time containers: *exc1*, *seq*, and *par* representing exclusive, sequential and parallel move sets. Other XML tags are specifically designed for FML.

FML supports three basic face moves: talking, expressions, and 3D head movements. Combined through time containers, they form an FML action which is a logically related set of activities. Details of these moves and other FML elements and constructs will be discussed in the next subsections. The special *fap* and *param* elements are also included for MPEG-4 FAPs and other system-dependent parameters. Time containers are FML elements that represent the temporal relation between moves. The basic time containers are *seq* and *par* corresponding to sequential and



FIGURE 8: Series of facial actions: (a) start, (b) wink, (c) head rotation, (d) smile.

```

<seq>
  <param type='comp' name='eye-wink'
    duration='1s' />
  <hdmv type='yaw' value='20'
    duration='1s' />
  <expr type='smile' value='50'
    duration='1s' />
</seq>

```

FIGURE 9: FML script for actions in Figure 8.

```

<action>
  <seq begin='0'>
    <talk>Hello</talk>
    <hdmv end='5s' type='0'
      val='30' />
  </seq>
  <par begin='0'>
    <talk>Hello</talk>
    <expr end='3s' type='3'
      val='50' />
  </par>
</action>

```

FIGURE 11: Time containers and basic moves.

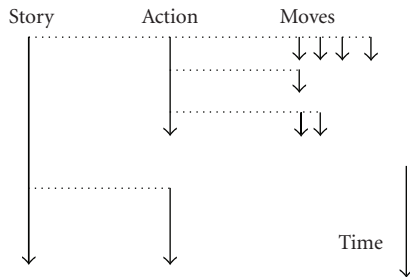


FIGURE 10: FML timeline and temporal relation of face activities.

```

<!-- in model part -->
<event name='user' val='-1' />

<!-- in story part -->
<excl ev_name='user'>
  <talk ev_val='0'>Hello</talk>
  <talk ev_val='1'>Bye</talk>
</excl>

```

FIGURE 12: Decision making and event handling.

parallel activities. The former contains moves that start one after another, and the latter contains moves that begin at the same time. Time containers include primitive moves and also other time containers in a nested way. The repeat attribute of the time container elements allows iteration in FML documents as illustrated later in sample applications.

Similar to SMIL, FML also has a third type of time containers, `excl`, used for implementing exclusive activities and decision making as discussed later. All `story` elements have four timing attributes: `repeat`, `begin`, `duration`, and `end`. In a sequential time container, `begin` is relative to start time of the previous move, and in a parallel container it is relative to the start time of the container. In case of a conflict, duration of moves is set according to their own settings rather than the container. The `repeat` attribute is considered for defining definite (when having an explicit value) or indefinite loops (associated with events). FML time containers and basic moves are illustrated in Figure 11.

7.3. Event handling and decision making

In dynamic and interactive applications, the FML document needs to make decisions, that is, to follow different paths based on certain events. To accomplish this, `excl` time container and event element are added. An event represents any external data, for example, the value of a user selection. The new time container associates with an event and allows waiting until the event has one of the given values, then it continues with exclusive execution of the action corresponding to that value, as illustrated in Figure 12. The system component processing FML scripts exposes proper interface function to allow event values to be set in run time. `event` is the FML counterpart of familiar if-else constructs in normal programming languages.

```

<!-- in model part -->
<event name='userChoice' val='-1' />
<param name='dominance' val='60' />
<param name='affiliation' val='90' />
<data name='reply-1' val='Hello' />
<data name='reply-2' val='Fine' />

<!-- in story part -->
<excl ev_name='userChoice'
  repeat='userChoice;4'>
  <talk ev_val='1' name='reply-1'>
    </talk>
  <talk ev_val='2' name='reply-2'>
    </talk>
  <talk ev_val='3' name='reply-3'>
    </talk>
</excl>

```

FIGURE 13: FML script for interactive agent.

8. SAMPLE APPLICATIONS

In this section, we review sample application using iFACE system and our proposed behavioral model. For more information, sample applications, and videos please see our research web site <http://ivizlab.sfu.ca/research>.

8.1. Interactive agent

Typical examples of an interactive agent are game characters and online customer service representatives. In such cases, the agent needs to follow a main scenario, allow non-linear sequences of events (e.g., making a decision based on a user input and going through different paths as the result), show emotions, and have a certain personality. Figure 13 demonstrates a sample FML script for such an agent.

This script creates a character that waits for user questions and replies to them. The user interface is controlled by the GUI application. It provides four options: “Hello,” “How are you?” a user-typed question, and “Bye.” The reply to options 1 and 2 are hard coded in the script (data elements in model). The reply to the third (user-typed) question will be provided by the background application (i.e., the intelligence behind the script). The fourth user option ends the script.

In the model part of the script, the personality parameters are set, a user event has been declared and set to -1 (default value, meaning not defined), and finally two data items have been set for user options 1 and 2. The main actions are controlled in the `excl` element. The `repeat` attribute defines the ending condition. The `excl` options look for the appropriate reply, either in the script or from the background application (through the iFACE API not shown here).

TABLE 3: Example relations between music features and emotions [11, 23, 28].

Emotion	Feature	Value
Fear	Tempo	Irregular
	Sound level	Low
	Articulation	Mostly nonlegato
Anger	Tempo	Very rapid
	Sound level	Load
	Articulation	Mostly nonlegato
Happiness	Tempo	Fast
	Sound level	Moderate or load
	Articulation	Airy



FIGURE 14: Sample animated heads from MusicFace.

8.2. MusicFace

Music-driven emotionally expressive face (MusicFace) [16] is a multimedia application based on iFACE to demonstrate the concept of affective communication remapping, that is, transforming affective information from one communication medium to another. Affective information is extracted from a piece of music by analyzing musical features such as rhythm, energy, timbre, articulation, and melody (see Table 3).

After setting general personality type and parameters based on the music, the emotional state is determined and updated continuously using the following algorithm (sample animation frames in Figure 14).

- (1) Select high or low arousal emotions based on music power level.
- (2) Select positive or negative valence emotions based on timbre and rhythm.
- (3) Fine tune emotional state based on other musical features.

9. CONCLUSION

We have described a behavioral model for social agents that consists of four independent but interacting parameter spaces: geometry, knowledge, personality, and mood. Personality and mood are modeled based on current findings in behavioral psychology, relating the perception of personality and the emotional states to facial actions and expressions. The character knowledge and tasks to be performed, in addition to the rules of behavior and decision making, are encapsulated in a specially designed language that is also

compatible with the MPEG-4 standard. Associating facial actions to parameters (affective or personality dimensions) rather than “basic emotions” or “personality types” allows a designer to easily change the parameters and create new personality types and combined expressions that are perceptually valid. Further research is needed to study the effect of cultural background on such perception.

REFERENCES

- [1] Y. Arafa, K. Kamyab, E. Mamdani, et al., “Two approaches to scripting character animation,” in *Proceedings of the 1st International Conference on Autonomous Agents & Multi-Agent Systems, Workshop on Embodied Conversational Agents*, Bologna, Italy, July 2002.
- [2] A. Arya, S. DiPaola, L. Jefferies, and J. T. Enns, “Socially communicative characters for interactive applications,” in *Proceedings of the 14th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG '06)*, Plzen-Bory, Czech Republic, January-February 2006.
- [3] A. Arya, L. N. Jefferies, J. T. Enns, and S. DiPaola, “Facial actions as visual cues for personality,” *Computer Animation and Virtual Worlds*, vol. 17, no. 3-4, pp. 371–382, 2006.
- [4] N. Badler, B. D. Reich, and B. L. Webber, “Towards personalities for animated agents with reactive and planning behaviors,” in *Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents*, R. Trapp and P. Petta, Eds., pp. 43–57, Springer, New York, NY, USA, 1997.
- [5] J. Bates, “The role of emotion in believable agents,” *Communications of the ACM*, vol. 37, no. 7, pp. 122–125, 1994.
- [6] S. Battista, F. Casalino, and C. Lande, “MPEG-4: a multimedia standard for the third millennium—part 1,” *IEEE Multimedia*, vol. 6, no. 4, pp. 74–83, 1999.
- [7] C. J. Beedie, P. C. Terry, and A. M. Lane, “Distinctions between emotion and mood,” *Cognition and Emotion*, vol. 19, no. 6, pp. 847–878, 2005.
- [8] D. S. Berry, “Accuracy in social perception: contributions of facial and vocal information,” *Journal of Personality and Social Psychology*, vol. 61, no. 2, pp. 298–307, 1991.
- [9] P. Borkenau, N. Mauer, R. Riemann, F. M. Spinath, and A. Angleitner, “Thin slices of behavior as cues of personality and intelligence,” *Journal of Personality and Social Psychology*, vol. 86, no. 4, pp. 599–614, 2004.
- [10] P. Borkenau and A. Liebler, “Trait inferences: sources of validity at zero acquaintance,” *Journal of Personality and Social Psychology*, vol. 62, no. 4, pp. 645–657, 1992.
- [11] R. Bresin and A. Friberg, “Synthesis and decoding of emotionally expressive music performance,” in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, vol. 4, pp. 317–322, Tokyo, Japan, October 1999.
- [12] D. C. A. Bulterman, “SMIL 2.0—part 1: overview, concepts, and structure,” *IEEE Multimedia*, vol. 8, no. 4, pp. 82–88, 2001.
- [13] J. Cassell, C. Pelachaud, N. Badler, et al., “Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents,” in *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '94)*, pp. 413–420, New York, NY, USA, July 1994.
- [14] J. Cassell, H. H. Vilhjálmsson, and T. Bickmore, “BEAT: the behaviour expression animation toolkit,” in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*, pp. 477–486, Los Angeles, Calif, USA, August 2001.
- [15] B. De Carolis, C. Pelachaud, I. Poggi, and M. Steedman, “APML, a markup language for believable behaviour generation,” in *Proceedings of the 1st International Conference on Autonomous Agents & Multi-Agent Systems, Workshop on Embodied Conversational Agents*, Bologna, Italy, July 2002.
- [16] S. DiPaola and A. Arya, “Affective communication remapping in musicface system,” in *Proceedings of the 10th European Conference on Electronic Imaging and the Visual Arts (EVA '04)*, London, UK, July 2004.
- [17] A. Egges, S. Kshirsagar, and N. Magnenat-Thalmann, “A model for personality and emotion simulation,” in *Proceedings of the 7th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES '03)*, pp. 453–461, Oxford, UK, September 2003.
- [18] P. Ekman, *Emotions Revealed*, Consulting Psychologists Press, San Francisco, Calif, USA, 1978.
- [19] P. Ekman and W. V. Friesen, *Facial Action Coding System*, Consulting Psychologists Press, San Francisco, Calif, USA, 1978.
- [20] J. Funge, X. Tu, and D. Terzopoulos, “Cognitive modeling: knowledge, reasoning and planning for intelligent characters,” in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)*, pp. 29–38, Los Angeles, Calif, USA, August 1999.
- [21] L. R. Goldberg, “An alternative “description of personality”: the big-five factor structure,” *Journal of Personality and Social Psychology*, vol. 59, no. 6, pp. 1216–1229, 1990.
- [22] C. Jones, *Chuck Amuck: The Life and Times of Animated Cartoonist*, Farrar, Straus, and Giroux, New York, NY, USA, 1989.
- [23] P. N. Juslin, “Cue utilization in communication of emotion in music performance: relating performance to perception,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 6, pp. 1797–1813, 2000.
- [24] S. A. King, A. Knott, and B. McCane, “Language-driven non-verbal communication in a bilingual conversational agent,” in *Proceedings of the 16th International Conference on Computer Animation and Social Agents (CASA '03)*, pp. 17–22, New Brunswick, NJ, USA, May 2003.
- [25] B. Knutson, “Facial expressions of emotion influence interpersonal trait inferences,” *Journal of Nonverbal Behavior*, vol. 20, no. 3, pp. 165–181, 1996.
- [26] S. Kshirsagar and N. Magnenat-Thalmann, “A multilayer personality model,” in *Proceedings of the 2nd International Symposium on Smart Graphics*, pp. 107–115, Hawthorne, NY, USA, June 2002.
- [27] W.-S. Lee, M. Escher, G. Sannier, and N. Magnenat-Thalmann, “MPEG-4 compatible faces from orthogonal photos,” in *Proceedings of Computer Animation (CA '99)*, pp. 186–194, Geneva, Switzerland, May 1999.
- [28] D. Liu, L. Lu, and H.-J. Zhang, “Automatic mood detection from acoustic music data,” in *Proceedings of the 4th International Symposium on Music Information Retrieval (ISMIR '03)*, Baltimore, Md, USA, October 2003.
- [29] A. B. Loyall and J. B. Bates, “Personality-rich believable agents that use language,” in *Proceedings of the 1st International Conference on Autonomous Agents*, pp. 106–113, Marina del Rey, Calif, USA, February 1997.
- [30] A. Marriott and J. Stallo, “VHML: uncertainties and problems. A discussion,” in *Proceedings of the 1st International Conference*

- on Autonomous Agents & Multi-Agent Systems, Workshop on Embodied Conversational Agents*, Bologna, Italy, July 2002.
- [31] J.-Y. Noh and U. Neumann, "Expression cloning," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*, pp. 277–288, Los Angeles, Calif, USA, August 2001.
- [32] A. Paradiso, "An algebra of facial expressions," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*, New Orleans, La, USA, July 2000.
- [33] C. Pelachaud and M. Bilvi, "Computational model of believable conversational agents," in *Communication in Multiagent Systems: Background, Current Trends and Future*, M.-P. Huget, Ed., pp. 300–317, Springer, New York, NY, USA, 2003.
- [34] K. Perlin, "Layered compositing of facial expression," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*, Los Angeles, Calif, USA, August 1997.
- [35] H. Prendinger, S. Descamps, and M. Ishizuka, "Scripting affective communication with life-like characters in web-based interaction systems," *Applied Artificial Intelligence*, vol. 16, no. 7-8, pp. 519–553, 2002.
- [36] D. Rousseau and B. Hayes-Roth, "Interacting with personality-rich characters," Report KSL 97-06, Knowledge Systems Laboratory, Stanford University, Stanford, Calif, USA, 1997.
- [37] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [38] K. Smid, I. Pandzic, and V. Radman, "Autonomous speaker agent," in *Proceedings of Computer Animation and Social Agents Conference (CASA '04)*, Geneva, Switzerland, July 2004.
- [39] D. Watson, "Strangers' ratings of the five robust personality factors: evidence of a surprising convergence with self-report," *Journal of Personality and Social Psychology*, vol. 57, no. 1, pp. 120–128, 1989.
- [40] J. S. Wiggins, P. Trapnell, and N. Phillips, "Psychometric and geometric characteristics of the revised interpersonal adjective scales (IAS-R)," *Multivariate Behavioral Research*, vol. 23, no. 3, pp. 517–530, 1988.