

Representing and Parameterizing Agent Behaviors

Norman Badler, Jan Allbeck, Liwei Zhao, Meeran Byun
Center for Human Modeling and Simulation
Computer and Information Science
University of Pennsylvania
200 S. 33rd St.
Philadelphia, PA 19104-6389
badler — allbeck — lwzhao — byun @seas.upenn.edu

Abstract

The last few years have seen great maturation in understanding how to use computer graphics technology to portray 3D embodied characters or virtual humans. Unlike the off-line, animator-intensive methods used in the special effects industry, real-time embodied agents are expected to exist and interact with us “live.” They can be represent other people or function as autonomous helpers, teammates, or tutors enabling novel interactive educational and training applications. We should be able to interact and communicate with them through modalities we already use, such as language, facial expressions, and gesture. Various aspects and issues in real-time virtual humans will be discussed, including consistent parameterizations for gesture and facial actions using movement observation principles, and the representational basis for character believability, personality, and affect. We also describe a Parameterized Action Representation (PAR) that allows an agent to act, plan, and reason about its actions or actions of others. Besides embodying the semantics of human action, the PAR is designed for building future behaviors into autonomous agents and controlling the animation parameters that portray personality, mood, and affect in an embodied agent.

1. Introduction

Given the extraordinary appearance and achievements of special effects and interactive games, the creation of effective real-time autonomous embodied agents remains one of the last research frontiers in computer animation. Whether the beings portrayed are supposed to be cartoon characters, fictional personalities, or known individuals, computer animation techniques with the greatest visual success are either crafted with relatively low level techniques such as paramet-

ric key interpolation or edited from motion captured with live performers. By necessity, such techniques are iterative, off-line, and dependent on the technical and aesthetic skills of the animator. Autonomous agents, however, are supposed to be entities that respond to human interaction in real-time and with behaviors that are perceived to be appropriate to the interaction and needs of the participants. In general, these needs are constrained by the application so that the behavioral repertoire of the agent can be created in advance and recalled and modified in real-time as needed. Thus games restrict the movements and choices of the game entities so that the user is really only exploring a finite game space.

In contrast to games, interactions between real people appear effectively infinite: at least they are real-time, not pre-determined, broad in content, highly contextual, communicative, behaviorally subtle, and even subliminal. The communications channels themselves are multi-modal and encode multiple levels of meaning. The agent may communicate with speech or other verbalizations, facial expressions, eye movements, head movements, limb gestures, body posture, and even gait. During interpersonal interactions, most people are relatively unaware of the acts their bodies are performing while they are talking, for example. Likewise, an observer picks up additional non-verbal signals from the speaker’s “body language.” Both participants are attempting to obtain or convey some sort of information via the communication, and therefore are exposing and manifesting some aspects of their internal knowledge, desires, beliefs, intentions, emotions, and feelings to the other.

For most people, the generation of all body actions associated with communication is both natural and subconscious. For actors, such actions may need to be created outside the “natural” modes of the actor and this requires skill, training, and discipline. Even for other people, however, even everyday interactions may require conscious “acting”: namely, the management of manifest behaviors to cover up

emotions, save face, dramatize a point, or simply lie. This leads to the thesis of this paper: the communicative content portrayed by an embodied agent is a function of all available body channels and consequently the observer's perception of the internal state of the agent is dependent on the relationships of those channels as well as their content. To an actor or an expert animator, this statement may be intuitively obvious. A "bad" actor may fail to control gestures or face to conform to the demeanor of his persona and his situation. He may not have internalized the goals and motivations of his character enough to use the body's own machinery to manifest these inner drives as appropriate behaviors [9]. The skilled animator will know that all aspects of her character must be consistent with its desired mental state since only voice, shape, and movement can be controlled for the final product [33]. We cannot open a dialog with the already animated character to further probe its mind or its psychological state. With a real-time embodied agent, however, we may indeed have such an opportunity [20]. When real people present multiple behavior channels we interpret them for consistency, honesty, sincerity, and for social roles, relationships, power, and intention. There is an important and subtle relationship between this collection of channel messages and believability [7, 11] a term used often and without definition in the animated agents literature. A character is believable if we can infer emotional or mental state by observing its behavior (even if is not portrayed as a human form). Since each channel contributes to the perception of internal state, they must not conflict if they are to present a consistent view. Conversely, if they do conflict the agent may be perceived as having some internal difficulty expressing itself. The agent may simply look clumsy or awkward, but it could appear insincere, confused, conflicted, "spacey" (emotionally detached), repetitious, or simply fake. Not surprisingly, these are often the criticisms directed toward animated agents. Our approach to remedying this problem is to enhance believability through coordinated and consistent expression of body movements in all possible channels.

In this discussion we limit the communication channels to arms and torso, facial expressions, and eye movements. Other work is in progress on gait [2]. Cassell is developing coordination models for gesture, speech, and body pose [12]. Perlin uses carefully tuned noise functions to move facial features [30] or body joints [31] for a look of coordinated animacy. Although these characters appear effective in this regard, without any benefit of underlying theory their mental and emotional states are ad hoc. Pelachaud has achieved nice results across a set of facial expression and speech parameters by considering performative relationships between two synthetic individuals in a conversation [32]. The application of internal agent states to speech parameters (intonation, rate, volume, clarity, etc.) is an interesting, important, and rather unexplored area.

We have been building a system called EMOTE to parameterize and modulate action performance [13]. It is based on a human movement observation system called Laban Movement Analysis. EMOTE is not an action selector per se; it is used to modify the execution of a given behavior and thus change its movement qualities or character. The power of EMOTE arises from the relatively small number of parameters that control or affect a much larger set, and from new extensions to the original definitions that include the non-articulated movements of the face. The same set of parameters control many aspects of manifest behavior across the agent's body and therefore permit experimentation with similar or dissimilar settings. Our working hypothesis is that behaviors manifest in separate channels with like EMOTE parameters will appear consistent to some internal state of the agent; conversely, dissimilar EMOTE parameters will convey various negative impressions of the character's internal consistency. At least in rather simple cases this latter observation is true. Arm gestures without facial expressions look odd [13]; facial expressions with neutral gestures look artificial [1]; arm gestures without torso involvement look insincere [3]; attempts at emotions in gait variations look funny without concomitant body and facial affect [2]; otherwise carefully timed gestures and speech fail to register with gesture performance and facial expressions [12]; and repetitious actions of web-based or Microsoft agents become irritating because they appear unconcerned about our changing (more negative) feelings about them.

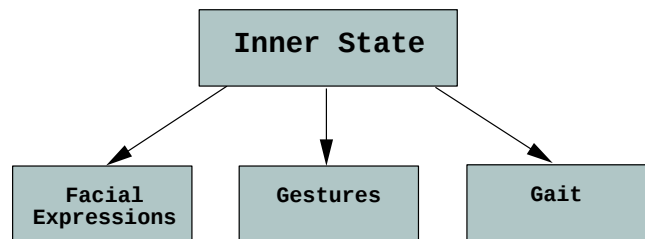


Figure 1. System Overview

As many of these criticisms are found in our own experiments as well as the current and genuinely valuable experiments of other major research groups, we were led to explore the underlying causes and remedies in order to improve visual appearance and believability in real-time autonomous animated characters. While such agents benefit from parameterizations that permit computation controls, our secondary hypothesis is that some coherent underlying models for behavior manifestation across all expressive communicative media are needed. Individual parameterizations of face (Figure 3), arms (Figure 2), gait (Figure 1), and so on may result in animated models, but these models must be related to and coordinated with some internal affect,

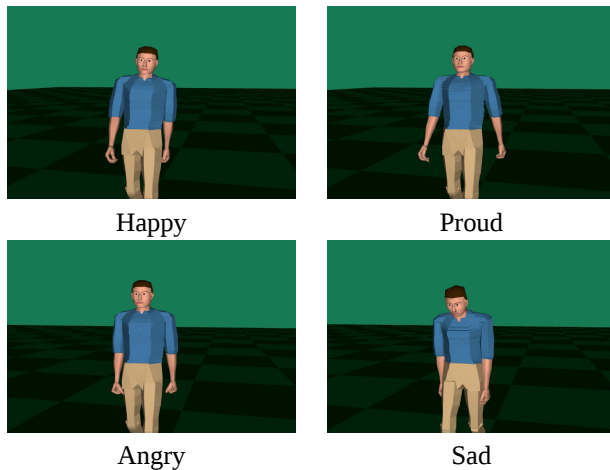


Table 1. Expressive gait without expressive face.

personality, or cultural condition of the agent. On this foundational principle we can then proceed to describe an architecture for consistent, believable, and expressive agents.

The rest of this discussion has the following structure. First we review the EMOTE parameters for gesture movements and extend them to facial expressions with FaceMOTe. Then we explore requirement for the internal state (emotion and personality) of an agent so that such states may be manifest through outward actions with behavioral consistency. Then we give an overview of parameterized actions and their relation to cognitive processes, natural language, instruction processing, planning, and action execution. We close with a discussion on building within an agent model the necessary mappings from internal state to external behaviors.

2. EMOTE Parameters

Originated by Rudolf Laban (1879-1958), Laban Movement Analysis (LMA) today is a creative method of movement study for observing, describing, notating, and interpreting human movement. LMA provides insights into one's personal movement style and increases awareness of what movement communicates and expresses. A variety of researchers have applied the LMA theories in many fields involving movement such as dance, theater, physical therapy and education, nonverbal communication and presentational skills, and management consulting [24, 16, 6, 27, 21].

LMA is composed of five major components: Body, Space, Effort, Shape, and Relationship. Together these components constitute a textual and symbolic language for describing movement. Body deals with which body parts move, where the movement initiates, and how the move-

ment spreads through the body. Space describes how large the mover's kinesphere, and what form is being revealed by the spatial pathways of the movement. Shape describes the changing forms that the body makes in space, while Effort involves the "dynamic" qualities of the movement and the inner attitude towards using energy. Relationship describes modes of interaction with oneself, others, and the environment. Each individual has a unique repertoire of and preferences for combinations of these basic elements, which can be sequenced, phrased, patterned, and orderly organized together in a particular personal, artistic, or cultural way. Our work focuses on the Effort and Shape components of LMA, because these two are the major direct specifications or indications of expressive human movements.

Effort comprises four motion factors: Space, Weight, Time, and Flow. Each motion factor is a continuum between two extremes: (1) *indulging* in the quality and (2) *fighting* against the quality. In LMA these extreme Effort Elements are seen as basic, "irreducible" qualities, meaning that they are the smallest units needed in describing an observed movement. These eight Effort Elements are: Indirect/Direct, Light/Strong, Sustained/Sudden, and Free/Bound. The eight elements can be combined and sequenced for many variations of phrasings and expressions [6].

Shape changes in movement can be described in terms of three dimensions: Horizontal, Vertical, and Sagittal. Each one of these dimensions is in fact associated with one of the three main dimensions (Length, Width, and Depth) as well as one of three main planes (Horizontal, Vertical, and Sagittal) related to the human body. Changes in Shape in the Horizontal dimension occur mainly in the side-open and side-across directions; as the movement becomes planar there would be more of a forward-backward component added to the primary side component. Changes in the Vertical dimension are manifested primarily in the upward-downward directions; the plane would add more sideward component to the up-down. Finally, changes in the Sagittal dimension are more evident in the body's depth or the forward-backward direction; planar movement would add an upward-downward component. Shape changes frequently occur in affinity with corresponding Effort Elements [24, 6, 27].

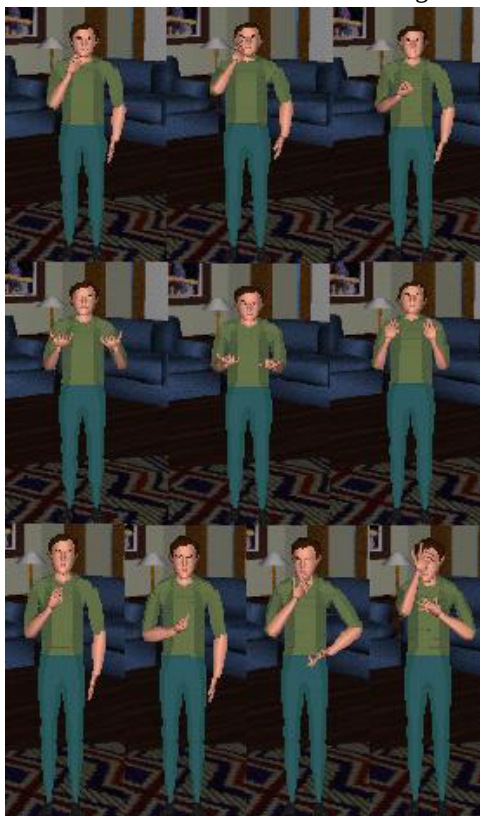
EMOTE (Expressive Motion Engine) is a computational realization of the core LMA concepts and principles. The main theme of the EMOTE system is to use high level qualitative Effort and Shape parameters for human animation control. To achieve that, the key component is to translate the qualitative Effort and Shape parameters into a set of low level quantitative parameters that are directly related to the control of the characteristics of the movement. An extensive empirical study carried out with the help of professional LMA notators developed the equations building



Car salesman



Tour guide



Sign Language with EMOTE

Tour guide

Table 2. Expressive gestures with the EMOTE system.

the connections [13]. Figure 2 shows some animations that resulted from these connections. There are four types of low level movement parameters: (1) parameters that affect the limb trajectory; (2) parameters that affect timing; (3) parameters that modify the torso shape and volume; and (4) flourishes that add to the expressiveness of the movement. The challenge lies in connecting this parameterization and action selection to the internal state of an agent.

3. FacEMOTE

Unlike skeletal gestures such as arm and hand movements, facial expressions are created mostly with soft tissue deformation. There are joints that cause head, jaw, and eyeball rotations, and these may be parameterized by the same EMOTE features. But the extension of EMOTE onto the soft tissues of the face is neither obvious nor explored in the movement observation literature. Nonetheless, it is worthwhile pursuing the possibility that facial movements may be moderated by EMOTE-like parameters.

One characteristic of soft tissue deformation is that moving a part of face has effects on other parts of the face. For example, raising corners of lips, as in smile, not only change the shape of the lips but also lift the cheeks and bunch up the lower eyelids to the effect of closing the eyes. Creating facial expressions involves moving facial muscles or regions of face so that the face expresses nuances in character's internal states of mind. The Table 3 lists some examples of facial expressions or movements showing clear manifestation of EMOTE parameters.

We add subtle changes to facial expressions by controlling these four intuitive EMOTE parameters, thus making it possible to easily obtain many different shades from a single basic expression. For example, increasing the value of Quick parameter for smile makes it fleeting, thus it may give an impression of not wanting to show pleasure. On the other hand, increasing the value of Suspended parameter (the opposite of Quick) may give an impression of putting on a polite smile at a social occasion. Increasing the value of Direct parameter may add an impression of slyness because of lowered eyebrow for focusing.

The FacEMOTE system runs on facial expressions specified as facial animation parameters (FAPs) standardized by MPEG-4. There are 66 low level parameters, each describing an atomic facial action deforming a face model from rotating head to raising a lip corner. The value of a FAP at given instance specify the amount of facial action from the model's relaxed position. For example, sequence of facial action acquired by increasing and then decreasing the value of 'close_t_l_eyelid' will make the model wink with its left eye.

FacEMOTE parameters (higher level parameter than FAPs) are applied onto the facial expressions given as FAP

Space	Indirect Direct	Scanning the party floor. Rolling eyes, trying to make a decision. Focusing on a ball player at the ball field. Squinting at the object an artist is drawing. Blowing out a candle.
Weight	Light Strong	Whispering to a child to sleep. Lightly tickled into giggling. Whining in muffled sound. Licking a spoonful of ice cream. Spelling out a word at a spelling bee. Snarling at an offender. Putting on a stern face when scolding a child.
Time	Suspended Quick	Relaxed expression while daydreaming. Taking a deep breath. Yawning. Nervous fidgeting. Coughing. Clearing the throat. Sobbing of a child after a screaming fit.
Flow	Free Bound	Crying of a baby when it is hungry. Bursting into uncontrollable laughter. Shouting in raging fury. Holding back tears. Chuckling instead of laughing loudly. Grimacing when touching a slimy object. Holding breath not to smell stench.

Table 3. Example manifestations of EMOTE parameters on the face.

streams, changing the intensity and the duration of each FAP. Increasing the value of Quick parameter value will affect most of the 66 low level parameters and will also have the following effects: The duration of facial action will be shortened for each parameter. This means that the time taken from the onset to the decay of an action is shortened, but the lengths of the FAP streams remain the same; The intensity of facial action at its local peaks will tend to be reduced. Overall, the FacEMOTE system works as a filter perturbing the value of each FAP for every time frame as it runs through the FAP streams.

4. Inner Action

The source whence perfection and final mastery of movement must flow is the understanding of that part of the inner life of man where movement and action originate. Such an understanding furthers the spontaneous flow of movement, and generates effective liveliness. [Rudolf Laban] [6].

We believe that modeling and simulating the “inner life” will provide us with parameterization to create consistent movement. There seem to be no absolutes when dealing with psychological aspects of humans and their manifestations in behavior. At best we can try to model trends or tendencies. Another difficulty arises in the number of variables or influences on human behavior. We can say that a person who is happy tends to smile and have light movements, but this is probably not the tendency if the person is extremely tired or around others who are sad. Although there are many properties that can be associated with individuality, including gender, age, culture, experiences, status, and role, much research has been done on personality and emotions and their application to virtual humans.

The manifestation of emotions, particularly in facial expressions, are quite recognizable [17]. The presence of emotional expression in virtual humans has long been shown to increase the believability or interest in virtual beings [7]. The presence of emotional expression, however, does not create the appearance of individuality. In the same situation, different people may feel different emotions, and even the same emotions may have different intensities and be displayed differently. Although personality is not the only determiner of emotions, their intensity, and their manifestation, we will use it as the first step in the creation of consistent individuality.

4.1. Modeling Emotion

Several different emotion models have been applied to virtual beings [15, 19, 25]. The most popular model is the OCC model, named after the authors [28]. In this model, emotions are generated through the agent’s construal of and reaction to the consequence of events, actions of agents, and aspects of objects. Although many researchers have based their work on this model [18, 7, 20], none have systems which display all 22 emotional states represented in the model in one communication channel, yet alone more than one.

4.2. Modeling Personality

Personality is a pattern of behavioral, temperamental, emotional, and mental traits that distinguish people from one another. Traits are basic tendencies that remain stable across the life span, but characteristic behavior can change through adaptive processes. The ways in which a person perceives, acts, and reacts is influenced by his or her personality. There has been an increasing research interest in personalities for virtual humans [34].

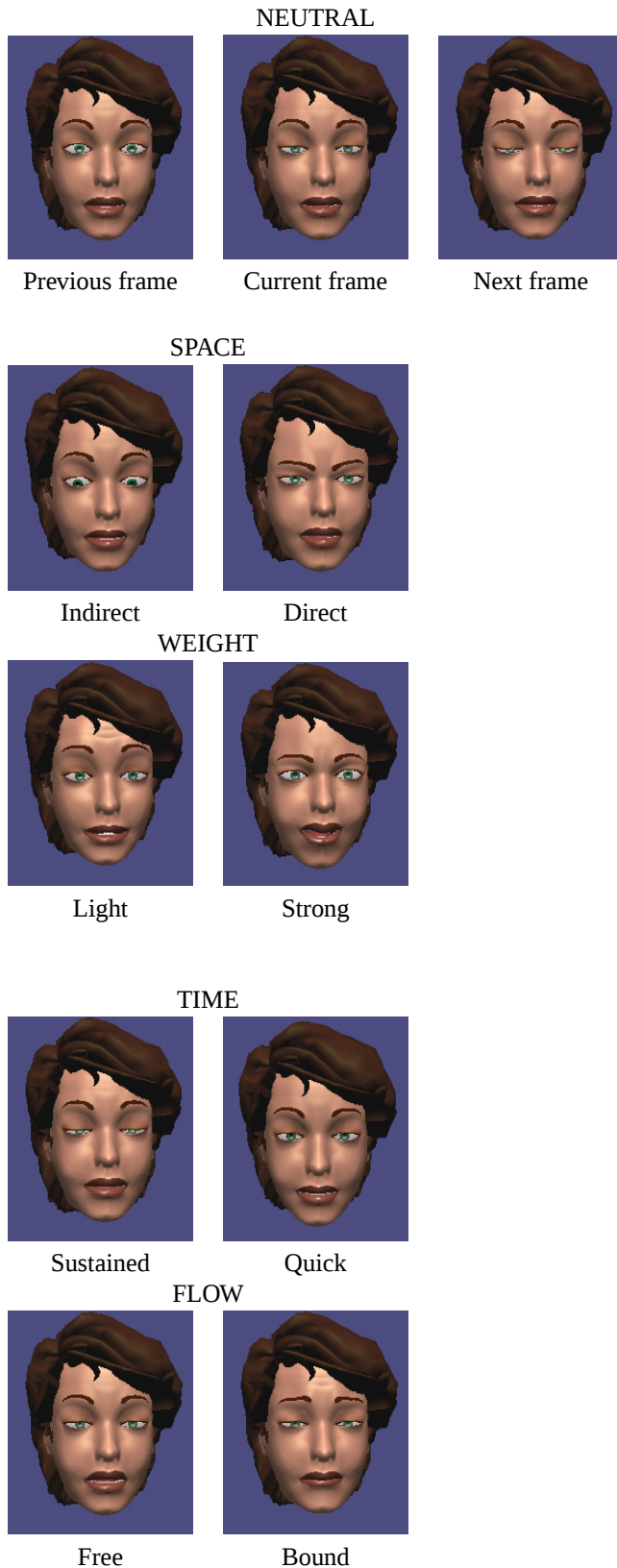


Table 4. FacEMOTE applies EMOTE parameters to facial expressions.

There are many psychological theories of personality. While there is no universally accepted theory, the Big Five or OCEAN model has gained some acceptance [35]. The Big Five represent a taxonomy of traits that some personality psychologists suggest capture the essence of individual differences in personality. The traits of the Big Five model are shown in Table 5.

Openness means a person is imaginative, independent minded and has divergent thinking. Openness to experience describes the breadth, depth, originality, and complexity of an individual's mental and experiential life. Conscientiousness means a person is responsible, orderly, and dependable. Conscientiousness describes socially prescribed impulse control that facilitates task and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks. Extraversion means that a person is talkative, social, and assertive. It implies an energetic approach to the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality. Agreeableness means a person is good natured, co-operative, and trusting. Agreeableness contrasts a prosocial and communal orientation toward others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty. Neuroticism means a person is anxious, prone to depression, and worries a lot. It contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad, and tense.

4.3. Behavioral Consistency

We propose a system for the determination and display of emotions with consistency over the channels of communication, influence of personality for added individuality, and expandability to other psychological traits and complexity. The idea is to have the agent's inner state displayed on multiple channels using the same parameterizations (Figure 1).

Figure 2 shows a more detailed view of our proposed system. We start with distributions of EMOTE parameters that correspond to personality types. In order to make movement qualities consistent over different channels such as facial expressions and gestures, they should have a compatible parameterization: that is, they should have some relationship to one another otherwise one would have to derive yet another level of transformation from inner state to each external manifestation of behavior or affect. Basing this parameterization in LMA means that our parameterization is grounded in movement observation science and has been interpreted as a computational model for animation.

Another choice would be to use a parameterization based on psychological properties. This might work if we only

	High Score Traits	Low Score Traits
Openness	Creative, Curious, Complex	Conventional, Narrow interests, Uncreative
Conscientiousness	Reliable, Well-organized, Self-disciplined, Careful	Disorganized, Undependable, Negligent
Extraversion	Sociable, Friendly, Fun-loving, Talkative	Introverted, Reserved, Inhibited, Quiet
Agreeableness	Good natured, Sympathetic, Forgiving, Courteous	Critical, Rude, Harsh, Callous
Neuroticism	Nervous, High-strung, Insecure, Worrying	Calm, Relaxed, Secure, Hardy

Table 5. OCEAN Model of Personality

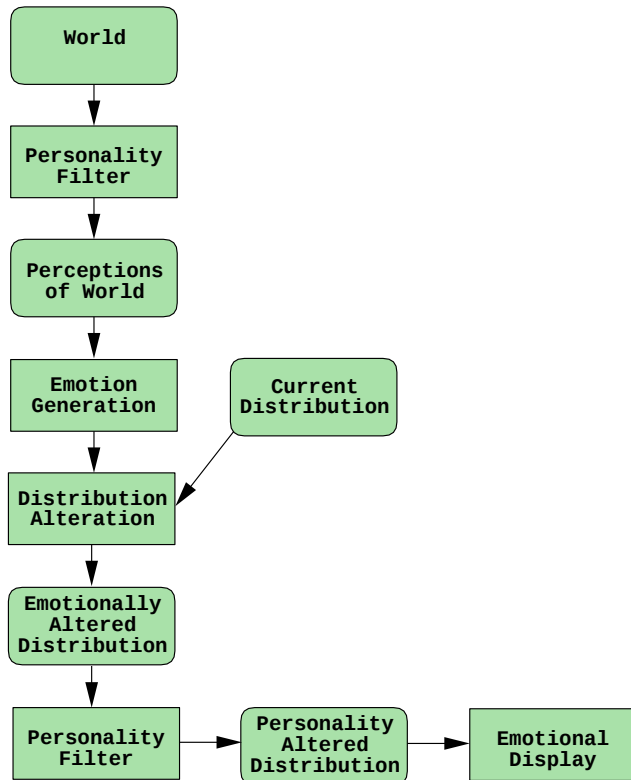


Figure 2. System Architecture

needed to model one psychological property such as emotion. We would, however, like to simulate other mental states such as anxiety and confusion. We would also like to predicate movements on culture and other social attributes. Creating a parameterization based on one psychological property would be too limiting and basing a parameterization on all properties too complex. Additionally, the parameterization needs to be compatible with animation parameters such as position, orientation, velocity, and acceleration. The EMOTE parameters fulfill this need.

Natural language is another alternative parameterization. While natural language has a lot of expressivity, it does not contain the necessary granularity in the concise form

needed for this application. In our work, we use natural language as a means for expressing action selection: adverbs and manner modify actions to yield expressive qualities [37].

The representation of the parameters as statistical distributions provides two advantages. The first advantage is the ability to shift and scale the distributions while avoiding discontinuities in the joint movements of the agents. The second advantage is the ability to make the agent more complex by adding other components. Here we describe a model which has personality and emotion, but when a culture component is developed its effects on movement qualities will be determined through manipulation of the EMOTE parameter distributions. A similar computational model has been used by Ball and Breese to model user mood based on user interface behaviors [5].

We wish to model personality in order to create characters that are distinguishable from one another in actions as well as in appearance. Bartenieff noted that:

Most people have predilections for particular Effort elements. ... The reappearance of the same Effort components finally characterize the activity for each person, and may also throughout different kinds of activities, eventually characterize the person. The “preferred” qualities (Effort choices) of individuals become aspects of their individuality, character attributes that are recognizable and remembered. [6]

Research in psychology and non-verbal communication provides us with some generalities about movements and personality types. For example, people who are more neurotic and introverted have more restrained and rigid behavior, and display more uncoordinated, random movements [10].

Once the distributions of EMOTE parameters for different personality traits have been created, the simulation begins with the agent perceiving its world. Every individual sees the world differently. There are many reasons for this, including physical location, personality, emotional state, culture, goals, and motivations. In our architecture we

currently represent this difference in world view by a personality filter. We realize, however, that this is an extreme over-simplification. Once the agent's view or beliefs about the world have been established, the construals of the OCC model can be used to obtain an emotional state.

The emotional state can then be used to alter the current distribution of EMOTE parameters to reflect the character's emotional state. This emotionally altered distribution is passed through a personality filter before EMOTE settings are chosen and displayed in the various channels of communication (Figure 1). Note that the previous personality filter was used to personalize the agent's view of the world. This personality filter is used to personalize the agent's movements. The original distributions were altered by personality traits at the beginning of the simulation, but the display of emotion also needs to be conditioned by personality.

5. Action Selection and Parameterization

Recently we have been designing and elaborating a Parameterized Action Representation or PAR [4, 8]. It is expanding into a useful ontology for actions, and links natural language with animation. An action representation is important for an agent as it holds the semantics of actions to be performed as well as a database of actions to be recognized. It is also conceivable that an action representation can function in an agent as a cognitive model for thought processes such as planning and reasoning [14]. Agents need to "do things" and these actions will be selected by needs and desires, while their execution is moderated by culture, personality, mood, affect, and skill. This section discusses the representation of actions for an agent.

As a representation for actions as instructions for an agent, the PAR has to specify (parameterize) the agent, any relevant objects, and information about paths, locations, manners, and purposes. Natural language often describes actions at a high level, leaving out many of the details that need to be specified for animation. There are linguistic constraints on how this information can be conveyed by the language – agents and objects tend to be verb arguments, paths are often prepositional phrases, and manners and purposes might be in additional clauses [29, 26]. For instance, the instruction *walk to the door and turn the handle slowly*, lacks any explicit information about grasping the handle, or which direction it will need to be turned, yet these are necessary to the action execution. The example does include a movement quality term ("slowly") and such a modifier must be applied to the generic "turn" instruction. The PAR has to include information about *applicability*, *preparatory*, and *terminating* conditions in order to fill in these gaps. It also has to be *parameterized* because other details of the action depend on its participants and method of performance

– agents, objects, and other attributes such as direction and manner. The handle object "knows" what actions it can perform and what state changes they cause (it is a "smart" object" [22]: instantiating the "turn" PAR on the handle yields normative values against which "slowly" can be evaluated and then simulated. The agent receiving the walk PAR decides on its direction, path, and gait depending on the environment and her needs; e.g., she may have to turn around first or navigate past furniture and may approach the door cautiously or aggressively.

In the remainder of this section, we will describe the components of the PAR system and how it processes and represents instructions for animations. Fig. 3 shows the architecture of the PAR system.

5.1. Natural Language Processing

The user inputs natural language instructions for a specific agent through a graphical user interface in the *Execution Engine*. The *NL transducer* parses the instructions, translates them into situation calculus expressions encapsulating references to PAR schemas, and sends them to the *Agent Process*. PAR schemas describe actions in terms of a conjunction of semantic predicates and its arguments. A verb like *hit* can be specified by predicates which establish that there is an agent and an object involved, that the agent causes the event, and that there is contact with force at the end of the event.

The PAR Schema hierarchy of actions exploit the idea that verbs can be represented in a lattice that allows semantically similar verbs, such as motion verbs or verbs of contact, to be closely associated with each other under a common parent that captures the properties these verbs all share. The highest nodes in the hierarchy are occupied by generalized PAR schemas which represent the basic predicate-argument structure for entire groups of subordinate actions. The lower nodes are occupied by progressively more specific schemas that inherit information from the generalized PARs, and can be instantiated with arguments from natural language to represent a specific action.

PAR schemas have also been shown [23] to be provide a kind of interlingual representation for translation of actions between verb-frame and satellite-frame languages.

5.2. Conditional Instructions and Goal Processing

Not every instruction a user may want to give should take place immediately. *Conditional instructions*, such as "When you enter the room, turn the light on," happen only when the condition is met. It is also possible that they should happen every time the condition is met, not just the first time. We call these type of instructions *standing orders*.

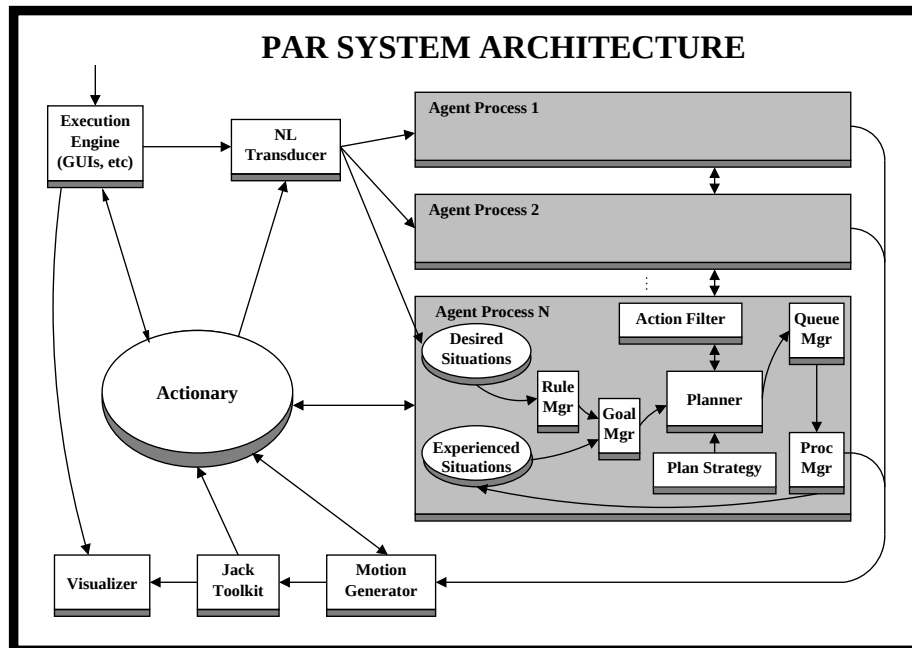


Figure 3. PAR Architecture.

In order to process these types of instructions, we need to maintain a reservoir of the instructions in their logical form. Essentially, this reservoir is a collection of situations the user wishes to see in the virtual environment both in the present and in the future. Hence, we call this collection of goal states, the *Desired Situations*. This collection is periodically processed by the *Rule Manager* in order to determine the current set of goal states (immediate instructions and applicable conditional instructions).

It is also necessary to ensure that goal states that have already been processed are not processed again. Take for example, "Check every room." Essentially, this requires that the agent performing the action remembers the rooms it has checked. This information is stored in the *Experienced Situations*. Before an agent starts planning, its *Goal Manager* checks its current goal states against those in the *Experienced Situations* and uses the results for planning.

5.3. The Actionary™

The *Actionary* (Figure 4) is the core component of our system. It contains persistent, hierarchical databases of agents, objects, and actions. The agents are treated as special objects and stored within the same hierarchical structure as the objects. Actions are represented as PARs (Parameterized Action Representation). Each PAR can either be uninstantiated (UPAR), contain only default properties for the action or be instantiated (IPAR), containing specific information about the agent, objects, and other prop-

erties. All the UPARs are stored hierarchically within the *Actionary*. PAR schemas have their own hierarchically organized tree in the *Actionary* which is derived from natural language semantics. PAR schemas may map to one or more UPARs. For example, a PAR schema representing *enter* would correspond to many UPARs, including walk, run, skip, swim, and crawl depending on the adjunctions and the context of the animation.

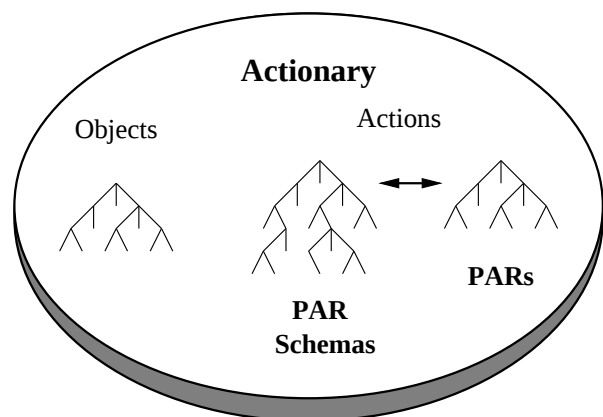


Figure 4. Actionary™.

The *Actionary* is a persistent potentially large database. In order to increase the speed of searches and traversals of both the action and object hierarchies, a *Database Manager* loads only the portions of the databases necessary for the

current environment. The *World Model* represents the current state of objects in the environment and actions which have been loaded both use in the environment.

5.4. Planning

The UPARs in the action hierarchy are used as building blocks for plans. The *Planner* evaluates the situation calculus expressions and retrieves the PAR schemas from them. For each PAR schema, the *Planner* needs to retrieve the set of all relevant UPARs from the *World Model*. The *Action Filter* may first remove or sort some of actions based on the characteristic of the agent before returning the list of UPARs to the *Planner*. For example, the *Action Filter* might prioritize *walking* over *skipping* as a translatory action for a business woman.

The *Planner* solves an abstract planning problem, where the initial state comes from the *World Model*, the goal state from the *Desired Situations*, a preference order of available actions from the *Action Filter*, parameters to select plan structure from *Plan Strategy*, and constraints, also from *Desired Situations*, are used to eliminate possible plans from consideration. The result of the planning process is a plan constructed of IPARs.

5.5. Action Execution

Once the plan of IPARs is determined and action parameterization such as the addition of manner specification completed, the IPARs are placed on the *Agent Process's* action queue. The *Queue Manager* and *Process Manager* monitor the execution of actions. These managers ensure that conditions necessary for performing actions are met. For example, the *applicability conditions* of each PAR generally check certain properties of the objects, the abilities of the agent, and other unchangeable or uncontrollable aspects of the environment.

Preparatory specifications contain conditions which must be true in order for the PAR to be performed and actions that may be performed in order to enable the current action to proceed. In general, preparatory specifications may involve the full power of motion planning to determine for instance, that a handle must be grasped before it can be turned.

The *Queue Manager* and *Process Manager* also call the *Motion Generators* associated with the action to be performed. Any manner parameters are interpreted as motion qualities in the EMOTE or FacEMOTE subsystems. These managers also monitor the performance of the action and check for termination conditions and failure states. If a failure occurs, the action can be aborted, sent to a specialized error handler for recovery, replanning can occur, or the entire plan can be aborted. The *Motion Generators* currently

work on top of JackTM.

6. Discussion and Future Directions

The agent architecture we have described still requires that inner states be mapped onto selections of EMOTE parameters from distributions relative to personality, culture, mood, and emotion. We do not believe such a mapping is obvious, though it might be elucidated in either a constructive or deconstructive fashion. In the constructive case, we would have an agent “learn” the mapping from inner agent states to face and body actions by observation. This is at least partly the way humans learn to “read” each others’ affect and correlate behaviors with expressed desires and intentions. This takes time and large experimental samples. In the deconstructive case, the mapping must be laboriously developed from specific experiments to tease apart the relationship of each movement variable from the surrounding interaction context. Our experimental tools seem ill-prepared to tackle the high dimensionality and context-dependency of this approach. We are presently more sanguine about the constructive approach. In recent work, Zhao [36] has shown that the appearance of significant EMOTE parameters may be observed in human movements captured with electromagnetic or video sensors. By observing enough human behaviors and having them correlated with some observational ground truth, one may begin to build a mapping from inner state to outward manifestation. In fact, such experiments will be needed to validate our claim that consistency across body communication channels is a prerequisite for believable behaviors.

Beyond facial expressions and arm gestures lie relatively unexplored areas of parameterizations that manifest affect and personality on eye movement, head movement, gait, and even communal (group) behaviors. We are proceeding with some models based on EMOTE but specialized onto these components. We hope to create a fully integrated agent model in the near future so that we can better test the hypotheses we propose here.

As we were told so many times before we learned: It is the change of shape that shows the character is thinking. It is the thinking that gives the illusion of life. It is the life that gives meaning to the expression. [33]

7. Acknowledgments

This research is partially supported by U.S. Air Force through Delivery Order #8 on F33615-99-D-6001, Office of Naval Research K-5-55043/3916-1552793, NSF IIS99-00297, and NASA NRA NAG 9-1279. Any opinions, findings, and conclusions or recommendations expressed in this

material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor any other sponsoring organization.

References

- [1] Ananova. <http://www.ananova.com/>.
- [2] K. Ashida, S.-J. Lee, J. Allbeck, H. Sun, N. Badler, and D. Metaxas. Pedestrians: Creating agent behaviors through statistical analysis of observation data. In *Proc. Computer Animation*, pages 84–92, Seoul, Korea, 2001. IEEE Computer Society.
- [3] N. Badler, M. Costa, L. Zhao, and D. Chi. To gesture or not to gesture: What is the question? In *Proc. Computer Graphics International*, pages 3–9, Geneva, Switzerland, June 2000. IEEE Computer Society.
- [4] N. Badler, M. Palmer, and R. Bindiganavale. Animation control for real-time virtual humans. *Comm. of the ACM*, 42(8):64–73, August 1999.
- [5] G. Ball and J. Breese. Emotion and personality in a conversational agent. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 189–219. MIT Press, Cambridge MA, 2000.
- [6] I. Bartenieff and D. Lewis. *Body Movement: Coping with the environment*. Gordon and Breach, New York, 1980.
- [7] J. Bates. The role of emotion in believable agents. *Comm. of the ACM*, 7(37):122–125, 1994.
- [8] R. Bindiganavale, W. Schuler, J. Allbeck, N. Badler, A. Joshi, and M. Palmer. Dynamically altering agent behaviors using natural language instructions. In *Autonomous Agents*, pages 293–300, New York, June 2000. ACM Press.
- [9] B. Bruehl. *The Technique of Inner Action: The Soul of a Performance Work*. Heinemann, Portsmouth, NH, 1996.
- [10] J. Burgoon, D. Buller, and W. Woodall. *Nonverbal Communication, the unspoken dialogue*. Harper and Row, New York, 1989.
- [11] J. Cassell. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 1–27. MIT Press, Cambridge MA, 2000.
- [12] J. Cassell, H. Vilhjalmsson, and T. Bickmore. BEAT: The Behavior Expression Animation Toolkit. In *Proc. ACM SIGGRAPH*, pages 477–486, Los Angeles, CA, 2001. ACM Press.
- [13] D. Chi, M. Costa, L. Zhao, and N. Badler. The emote model for effort and shape. In *Proc. ACM SIGGRAPH*, pages 173–182, New Orleans, LA, 2000.
- [14] W. Clancey. *Conceptual Coordination*. Lawrence Erlbaum Associates, Mahwah, NJ, 1999.
- [15] A. Damasio. *Descartes' error: Emotion, reason, and the human brain*. G.P. Putnam, 1994.
- [16] M. Davis. Effort-Shape analysis: Evaluation of its logic and consistency and its systematic use in research. In I. Bartenieff, M. Davis, and F. Paula, editors, *Four Adaptations of Effort Theory in Research and Teaching*, New York, 1970. Dance Notation Bureau, Inc.
- [17] P. Ekman and W. Friesen. Constraints across cultures in the face and emotion. *J. of Personality and Social Psychology*, 17:124–129, 1971.
- [18] M. El-Nasr, J. Yen, and T. Ioerger. FLAME – Fuzzy Logic Adaptive Model of Emotions. *Autonomous Agents and Multi-Agent Systems*, 3:219–257, 2000.
- [19] D. Goleman. *Emotional Intelligence*. Bantam Books, New York, 1995.
- [20] J. Gratch and S. Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proc. Autonomous Agents*, pages 278–285, Montreal, Quebec, 2001. ACM Press.
- [21] J. Hodgson. *Mastering Movement: The Life and Work of Rudolf Laban*. Routledge, New York, 2001.
- [22] M. Kallmann and D. Thalmann. Smart objects. In *Proc. Computer Animation Conf.*, pages 138–146, Geneva, Switzerland, 1999. IEEE Computer Society.
- [23] K. Kipper and M. Palmer. Representation of actions as an interlingua. In *Proceedings of the Third Workshop on Applied Interlinguas, held in conjunction with ANLP-NAACL 2000*, Seattle, WA, April 2000.
- [24] W. Lamb and D. Turner. *Management Behavior*. International Universities Press, Inc., New York, 1969.
- [25] J. LeDoux. *The Emotional Brain*. Simon and Schuster, New York, 1996.
- [26] S. Narayanan. Talking the talk is like walking the walk. In *Proceedings of the 19th Meeting of the Cognitive Science Society*, 1997.
- [27] J. Newlove. *Laban for Actors and Dancers: Putting Laban's Movement Theory into Practice*. Theatre Arts Books, New York, 1993.
- [28] A. Ortony, G. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.
- [29] M. Palmer, J. Rosenzweig, and W. Schuler. Capturing motion verb generalizations with synchronous tag. In P. S. Dizier, editor, *Predicative Forms in NLP: Text, Speech and Language Technology Series*. Kluwer Press, Dec. 1998.
- [30] K. Perlin. Responsive face. <http://www.mrl.nyu.edu/~perlin/facedemo/>.
- [31] K. Perlin. Real time responsive animation with personality. *IEEE Trans. on Visualization and Computer Graphics*, 1(1):5–15, 1995.
- [32] I. Poggi and C. Pelachaud. Performative facial expressions in animated faces. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 155–189. The MIT Press, Cambridge, MA, 2000.
- [33] F. Thomas and O. Johnson. *The Illusion of Life*. Abbeville Press, New York, 1981.
- [34] R. Trappl and P. Petta, editors. *Creating personalities for synthetic actors: Towards autonomous personality agents*. Springer Verlag, Berlin, Germany, 1997.
- [35] J. Wiggins. *The Five-Factor Model of Personality: Theoretical Perspectives*. The Guilford Press, New York, 1996.
- [36] L. Zhao. *Synthesis and acquisition of Laban Movement Analysis qualitative parameters for communicative gestures*. PhD thesis, Computer and Information Science, Univ. of Pennsylvania, Philadelphia, PA, 2001.
- [37] L. Zhao, M. Costa, and N. Badler. Interpreting movement manner. In *Proc. Computer Animation Conf.*, pages 112–120, Philadelphia, PA, May 2000. IEEE Computer Society.