SOCIALLY EXPRESSIVE COMMUNICATION AGENTS: A FACE-CENTRIC APPROACH

S. DiPaola and A. Arya
School of Interactive Arts and Technology,
Simon Fraser University
Surrey, BC, Canada
{sdipaola,aarya}@sfu.ca

Abstract – Interactive Face Animation - Comprehensive Environment (iFACE) is a general-purpose software framework that encapsulates the functionality of "face multimedia object". iFACE exposes programming interfaces and provides authoring and scripting tools to design a face object, define its behaviors, and animate it through static or interactive situations. The framework is based on four parameterized spaces of Geometry, Mood, Personality, and Knowledge that together form the appearance and behavior of the face object. iFACE capabilities are demonstrated within the context of some artistic and educational projects.

INTRODUCTION

Recent advances in computer graphics and multimedia systems have caused tremendous changes in the areas of visualization and human computer interaction. Computer-generated content is rapidly replacing "live action" in entertainment, social and business services, and many other applications. Unavailability of characters and hardship and/or expense associated with utilizing them on one hand and flexible and programmable nature of parameterized computer-generated content on the other hand, make a "virtual" agent an appealing replacement. Such an agent can communicate with viewers using a realistic (or stylistic) appearance, social interaction, and "expressiveness" (i.e. displaying familiar and expected behaviors and emotions).

Our proposed system, Interactive Face Animation - Comprehensive Environment (iFACE), is a general-purpose software framework that encapsulates the functionality of "face multimedia object" [1], freeing the application developers from details related to generating face-based multimedia presentations. iFACE exposes programming interfaces and provides authoring and scripting tools that allow users to design a face object, define its behaviors, and animate it through static or interactive situations. The framework is based on four parameterized spaces of Geometry, Mood, Personality, and Knowledge that together form the appearance and behavior of the face object.

Human narrators, guides, and instructors are major parts of many art shows, museums, and classrooms. In many of these cases, a human figure can be significant part of the presented content itself. We will demonstrate how iFACE is being used to improve the quality of presentation in two artistic and educational applications: The Evolving Faces project aims at using iFACE to make human figures that are both the content and the narrator of this evolutionary story, and Storytelling Masks allow computer-generated masks and figures of North American natives to tell their stories and sing their songs.

RELATED WORK

The parameterized models are effective ways for facial animation due to use of limited parameters, associated to main facial feature points instead of simulating physical characteristics [7]. Such parameters can be used to calibrate a standard model and animate it. Facial Action Coding System (FACS) [5] was an early and still valid study of possible facial actions related to such feature points. Although not originally a computer graphics technique, FACS has been widely used by researchers in parameterized models and others. This approach has been formalized in MPEG-4 standard by introduction of Face Definition Parameters (FAPs) and Face Animation Parameters (FAPs) [2]. The former group of parameters defines the important features and the latter actions applied to a subset of them (not every FDP can be animated).

The primary issue with such parameter space is its "flatness". All parameters are worked with in a similar way while not every application actually needs all of them. A hierarchical grouping of parameters is necessary for efficient management of parameter space. Also, the relatively huge amount of parameters (result of extensions to the original models) makes application development and authoring hard. A common solution to this issue has been defining higher-level parameters and behaviors [3]. For example, "smile" is an action that involves a few features and can be defined at a higher level of abstraction, knowing the combined effect of movements in feature points. MPEG-4 FAPs define two groups of such high level parameters for standard facial expressions and visemes (visual representation of uttered phonemes).

Although such "macro" parameters make it easier to use the underlying model, the simple two-tier model is still not very effective for managing facial activities and providing local control over level-of-details. The MPEG-4 standard allows definition of parameter groups but it is only a standard to be used by particular models, which are still mainly "flat". Some researchers [3,7] have proposed hierarchical head models that allow a more efficient parameter control through grouping and regions. Our approach, as explained later, uses this idea and extends it to multiple layers of abstraction on top of actual data points (2D pixels or 3D vertices) to ensure maximum flexibility and minimum effort when group actions are required.

The behavioral modeling of animated characters has been studied by some researchers. Funge et al. [6], for instance, define a hierarchy of parameters. At the base of their parameter pyramid is the geometric group. On top of that come kinematic, physical, behavioral, and cognitive parameters and models. Although very important for introduction of behavioral and cognitive modeling concepts, the model may not be very suitable for face animation purposes due to the interaction of parameter groups and the need for emotional parameters as opposed to physically-based ones. Cassell et al. [4] defined behavioral rules to be used in creating character actions but do not propose a general head model integrating geometrical and behavioral aspects. This will be discussed in more details in the next section. Byun and Badler [3] propose the FacEMOTE system that allows four high-level "behavioral" parameters (Flow, Time, Weight, and Space) to control the expressiveness of an input FAP stream. Although it demonstrates how high-level behavioral parameters can control facial animation, it does not intend to be a comprehensive face object. On the other hand, three spaces of Knowledge, Mood, and Personality (each with their own parameters as explained later) can control the facial behavior in a more explicit way.

FACE MULTIMEDIA OBJECT IN iFACE SYSTEM

For a large group of applications, facial presentations can be considered a means of communication. A "communicative face" relies and focuses on those aspects of facial actions and features that help to effectively communicate a message. We believe that the communicative behavior of a face can be considered to be determined by the following groups (spaces) of parameter:

Geometry: This forms the underlying physical appearance of the face. Creating and animating different faces and face-types are done by manipulating the geometry that can be defined using 2D and/or 3D data (i.e. pixels and vertices). This geometry is based on a hierarchy of facial regions and sub-regions as illustrated in Figure 1.

Knowledge: Behavioral rules, stimulus-response association, and required actions are encapsulated into Knowledge. In the simplest case, this can be the sequence of actions that a face animation character has to follow. In more complicated cases, knowledge can be all the behavioral rules that an interactive character learns and uses. Knowledge acts through an XML-based language that defines scenarios, events, and decision-making logic.

Personality: Different characters can learn and have the same knowledge, but their actions, and the way they are performed, can still be different depending on individual interests, priorities, and characteristics. Personality encapsulates all the long-term modes of behavior and characteristics of an individual. Facial personality is parameterized based on typical head movements, blinking, raising eye-brows and similar facial actions.

Mood: Certain individual characteristics are transient results of external events and physical situation and needs. These emotions (e.g. happiness and sadness) and sensations (e.g. fatigue) may not last for a long time, but will have considerable effect on the behavior. Emotional state can be modeled as point in a 2D space where two axes correspond to energy and stress.

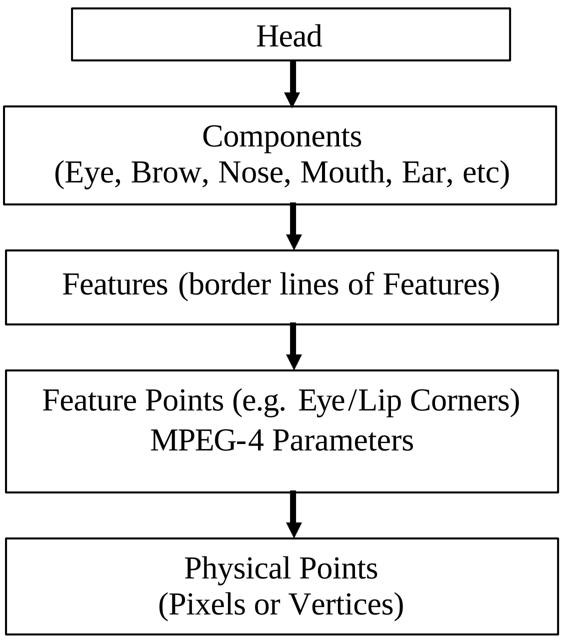

Figure1. Facial Geometry

```
<fml>
    <model>
        <event name="select" value="-1" />
    </model>
    <story>
        <action>
            <!--parallel actions-->
            <par>
                <hdmv type="yaw" value="80"
                        begin="0" end="2000" />
                <play file="Audio1.wav" />
            </par>
            <!--exclusive actions -->
            <!—-depending on the event value-->
            <!--only one of options will run-->
            <excl ev_name="select">
                <talk ev_value="0">Text One</talk>
                <talk ev_value="1">Text Two</talk>
            </excl>
        </action>
    </story>
</fml>
```

Figure 2. Sample FML Script

As shown in Figure 1, face geometry is a hierarchy of object that allow high-level or low level access to facial elements. Components are high-level functionally related regions of the face. Facial actions managed at Component level allow applying more complicated constraints such as effect of feature movements on each other. They also hide details of controlling movement from users. Features and Feature Points on the other hand make finer control of movement when necessary.

The behavior of an iFACE character is determined primarily by Knowledge. It provides the scenario that the character has go through as an XML-based script. iFACE uses Face Modeling Language (FML) [1] that is specifically designed for face animation. FML document can be a simple set of sequential actions such as speaking and moving the head, or a complicated scenario involving parallel actions and event-based decision-making as illustrated in Figure 2. Although scripts can select a new personality or modify the mood, but Knowledge space is generally independent of the "character". Mood and Personality spaces deal with character-dependent parameters. Mood controls short-term emotional state that can affect the way a certain action is animated. For instance actions in a part of script can be performed in a "happy" mood and in another part in a "sad" one and be visually different. In general, facial actions are modulated with the current mood of the character.

iFACE APPLICATIONS

Once a believable, controllable and communicative face environment is available to application developers, we believe a new range of social-based applications are possible. Most computer based communication systems such as internet-based websites (using HTML, PHP, ASP), information kiosks or ebooks (using PDF, Director, Flash) for example are informational in nature, not socially-based. People however (like teachers, aquarium or science experts, or museum guides) use a more social-based techniques to convey their message – they use their passion for the subject, narrative techniques, lesson plans, flexible content depending on audience or audience feedback, eye contact, humour, voice modulation. We believe these socially-based techniques using a communicative face system can open up more human centric applications in many areas, such as:

- *Video Games* that can convey the subtle dramatic nuances more common to cinema thereby extending games to a wider audience and into the educational and adult realms,
- *Chat Systems* that use voice and facial expression for better and deeper communication,
- *Education Systems* that bring the passion of a teacher into distance education.

For this paper however we will concentrate on applications that support more engaging art and science educational systems and we will discuss two ongoing application prototypes where it is hoped that expressive agents can engage the viewer to the deeper or complicated back-story of an artefact or science concept.

Figure 3. Frames from "storytelling mask" interactive, where a real artist's voice, passion, stories and expression first introduces himself and his work (A), begins to transform into his artwork (B), has his work tells it's back story with full voice and expression (C,D) and can return to his persona to interactively answer questions or give other educational content (A).

Most art or science museums (including zoos and aquariums) often still use static displays of text and graphics to explain the deeper historical or scientific concepts about the nearby artefact (i.e. a portrait, a model of a planet) and often the display is not read. The situation is very different when a human guide gives a presentation about that same artefact, engrossing the viewers in that subject as they use narrative, real-time and socially based deliveries. Can this experience be mimicked with interactive systems allowing students, who do not have geographically or financial access to a science facility a similar level of engagement and educational experience. Can a facility create a better level of engagement when a human guide is not available? We will describe two active prototypes:

Storytelling Masks

Museums of anthropology, especially in North America, display a variety of artefacts from "first nations" (Native Americans). Among the most attractive of these artefacts are masks and head figures presented on objects such as totem poles. Songs, myths, and stories relate these figures to the history of the people who made them. Computer-generated characters with those figure-forms who also tell their stories and sing their songs are appealing and informative for the viewers and also provide a new means of creativity and expression to the native artists. Combination of iFACE design and scripting tools provide such a creative environment. In this specific case we have begun working with the Parks Department of British Columbia, Canada and the Native community to create a museum display where a virtual version of an artist appears and tells the story of his work, can virtually turn into the artwork – a native mask – and have a virtual version of the art tell it's story (see Figure 3). Because all of this is under computer control it is possible to create many of the perceptual and educational techniques that a live human guide/artist could achieve including:

- *Introduction:* the ability to announce and bring the audience to the work,
- *Narrative Style:* conveying the back-story, passion, timing and expressiveness,
- *Multiple Contexts:* via interactive control, the material can be tailored to different age levels, different perspectives and focus areas, including updating the material,
- *Presentation*: the exhibit can feel more like a live presentation: for instance the interplay between artist and artefact – i.e. the mask is not displayed until the artist gives sufficient context; afterward the mask returns back to the artist/guide for additional commentary
- *Q&A:* at session end, the viewers can pick questions for more tailored commentary

Figure 4. Screenshot of 'Evolving Faces' where emotive talking faces describe the DNA science of human migration out of Africa, actively "morphing" facial types accordingly.

Evolving Faces

With similar goals and techniques of the Storytelling Mask project, *Evolving Faces* attempts to use facial agents to better engage viewers into the content, but in this case the agents are used to describe complicated scientific details, as well as create agents that are also an integral part of the content, evolving their appearance to tell the story of man's migration out of Africa based on new DNA techniques. iFACE allows a designer to create head models that correspond to various stages of human evolution, and assign different types of behaviour (e.g. coarse or fine) to them to be expressed during talking or interaction. Such characters are ideal for science booths or online learning. Adding simple or complicated artificial intelligence can improve the behavioural capability of the characters for real-time interaction. The display uses voices, change and expressive faces and maps rather than charts and text.

Screen shot from the "Human Migration" interactive is shown in Figure 4. It shows how complicated subject matter, how we migrated from Africa some 50-100,000 years ago with its DNA marker and facial type migration evidence; can be engagingly put forth. Viewers can click on a specific face/area and have it tell the story of that DNA marker or click on a migratory path, having an evolving face explain the science facts of the journey of man.

CONCLUSIONS

In this paper, we introduce the iFACE as a framework for face multimedia object. iFACE encapsulates all the functionality required for face animation into a single object with proper application programming interface, scripting language, and authoring tools. iFACE use a hierarchical head model that hides the modeling details and allows group functions to be performed more efficiently. Multiple layers of abstraction on top of actual head data make the client objects and users independent of data type (3D or 2D) and provide the similar behaviour regardless of that type. Application of iFACE in two art and science educational presentations has been illustrated. Future research directions for iFACE include a better parameterization of personality space (especially for nonverbal cases), adding more realistic features to head model without loss of real-time performance, and a more realistic combination of emotional mood and normal face actions [*].

References

[1] A. Arya and S. DiPaola, "Face As A Multimedia Object," *5th Intl Workshop on Image Analysis for Multimedia Interactive Services*, Lisbon, Portugal, April 21-23, 2004.

[2] S. Battista, et al., "MPEG-4: A Multimedia Standard for the Third Millennium," *IEEE Multimedia*, vol. 6, no. 4, IEEE Press. 1999.

[3] M. Byun and N.I. Badler, "FacEMOTE: Qualitative Parametric Modifiers for Facial Animations," *Proceedings of ACM SIGGRAPH/ Eurographics symposium on Computer Animation,* ACM Press. 2002.

[4] J. Cassell, et al., "BEAT: The Behavior Expression Animation Toolkit," *Proceedings of ACM SIGGRAPH*, ACM Press, 2001.

[5] P. Ekman and W.V. Friesen, *Facial Action Coding System,* Consulting Psychologists Press Inc, 1978.

[6] J. Funge, et al., "Cognitive Modeling," *Proceedings of ACM SIGGRAPH*, ACM Press, 1999.

[7] F. I. Parke and K. Waters, *Computer Facial Animation*. A. K. Peters, 2000.

[*] Simon Fraser University, Interactive Visualization Lab: http://ivizlab.sfu.ca/research