EMOES: Eye Motion and Ocular Expression Simulator

Nicoletta Adamo-Villani, Gerardo Beni, and Jeremy White

Abstract—We introduce, a new interactive 3D simulation system of ocular motion and expressions suitable for: (1) character animation applications to game design, film production, HCI (Human Computer Interface), conversational animated agents, and virtual reality; (2) medical applications (ophthalmic neurological and muscular pathologies: research and education); and (3) real time simulation of unconscious cognitive and emotional responses (for use, e.g., in psychological research). The system is comprised of: (1) a physiologically accurate parameterized 3D model of the eyes, eyelids, and eyebrow regions; and (2) a prototype device for real-time control of eye motions and expressions, including unconsciously produced expressions, for application as in (1), (2), and (3) above.

The 3D eye simulation system, created using state-of-the-art computer animation technology and ‘optimized’ for use with an interactive and web deliverable platform, is, to our knowledge, the most advanced/realistic available so far for applications to character animation and medical pedagogy.

Keywords—3D animation, HCI, medical simulation, ocular motion and expression.

I. INTRODUCTION

In recent years there has been considerable interest in the construction and animation of human facial models. Although several researchers have proposed different methods to represent realistic facial shape and muscle behavior with anatomical accuracy [1-5], research on faces has not specifically focused on eye movement and expression [6]. Only recently, eye movement has received more attention specifically focused on eye movement and expression [6].

In the area of character animation applied to game design and film production, 3D modelers and animators have developed realistic representations of human eyes [9]-- with no attention to medical accuracy-- but the animation of ocular motion and expression is limited to eye, eyelid, and eyebrow motions without much attention to details such as saccade magnitude, duration, blinking rate, and other parameters necessary to create higher emotional engagement between the audience and the 3D character (or avatar).

In the area of medical simulation, patient simulators have become increasingly important with the automation and telepresence of medical care [10]. An eye simulator was developed in 1997 at UC Davis Medical School [11] but only slight improvements on this model have been carried out since [12]. These simulators are designed to demonstrate the effect of disability or malfunctioning of any of the ocular muscles and/or nerves controlling them; however, the representation of the eye and ocular region is very schematic and 2 dimensional.

In the field of HCI research, virtual digital humans have been a key interest for decades. It can be anticipated that in the near future computer interaction will be characterized by natural face-to-face conversations with 3D characters that speak, emote, and gesture [13]. However, to date, the poor quality of such animated characters, also known as ‘conversational animated agents’, has been a major stumbling block to progress [14]. Researchers argue that it is premature to draw conclusions about the effectiveness of animated agents because they are still in their infancy, and nearly every major facet of their communicative abilities needs considerable research. Thus, there is a need to develop new architectures and technologies that will enable experimentation with perceptive animated agents that are more natural, believable and graceful than those available today [15]. Considering that the eyes play an essential role as a major channel of non-verbal communicative behavior, realistic representation and synchronization of eye motion and ocular expression with speech and/or signing gestures is crucial for an immersive computer generated experience.

The problem is difficult because the ocular system is anatomically very complex, and ocular expressions and motions are subtle and not univocally associated with the spoken phonemes. By comparison with the lip-sync technology, we can appreciate the complexity of what we may call ‘eye-sync’. The lip-sync process is insensitive to the semantic, i.e., the meaning of what it is spoken does not affect the lip motion (at least when no gross emotional expression is involved). Thus a one-to-one correspondence can be mechanically established between phonemes and lip motions. No such simplification is available for eye motion and ocular expression since they depend on the meaning of what is being conveyed by speech or by signing gestures.

Manuscript received August 30, 2005. This work was supported by the Purdue Research Foundation (PRF) (Summer faculty grant 2005), by the Envision Center for Data Perceptualization at Purdue, and by PHS-NIH grant 501DC005241-02 (‘Modeling the non-manuals of American Sign Language’). N. Adamo-Villani is with the Department of Computer Graphics Technology at Purdue University, West Lafayette, IN 47907, USA (phone: 765-496-1297; fax: 765-494-9267; e-mail: nadamov@purdue.edu).

G. Beni, is with the Department of Electrical Engineering at University of California Riverside, CA 92507, USA (e-mail: beni@ee.ucr.edu).

J. White is with the Department of Computer Graphics Technology at Purdue University, West Lafayette, IN 47907 (e-mail: Jeremy@purdue.edu).
Thus, much research and development is needed for both the problem of the anatomically accurate representation of ocular dynamics for physiological studies (medical applications) and the problem of the realistic and semantically accurate representation of ocular expressions and motions for character animation and psychological/cognitive studies. A first step toward the solution of both problems is the development of an anatomically realistic 3D simulator of eye motion and ocular expression, which is the object of this research.

The overview of the paper is as follows: in section 2 we describe the 3D eye simulator, including the eye model and its parameters; in section 3 we discuss the prototype controller device; conclusive remarks and examples of future applications are presented in section 4.

II. THE EYE SIMULATOR

We have developed a physiologically accurate 3D model of the eye, eyelids and eyebrow region using state-of-the-art 3D modeling/texturing tools (Alias Maya 6.0 software) and techniques. Fig. 1 shows the model of the eyes including the six eye muscles. Fig. 2 shows the eyes with eyelids and eyebrow areas.

Based on [16] we considered 4 variables of the ocular system: pupil, lens, eye movements (saccadic, pursuit, vergence) and eyelid movements (startle or reflex blinks, voluntary blinks, periodic blinks). For each oculomotor variable we considered the following parameters (in italics):

Pupil: Size (diameter)
Lens: Curvature and thickness (back radius (r), surface power (P), thickness (d) from [17])
Eye movements, 3 modes: saccades, pursuit, vergence (a) Saccades:
a.1 Saccade latency: time interval between presenting a visual (or other) stimulus and the initiation of a saccade to foveate the stimulus
a.2 Saccade amplitude: distance traveled by the eye between saccade initiation and termination
a.3 Direction of movement: (polar angle)
a.4 Velocity: average speed of the eye to saccade a specified distance
(b) Pursuit movements:
b.1 Direction of movement: (polar angle)
b.2 Velocity: average speed while tracking a slowly moving object
(c) Vergence (disconjugate rotations) movements:
c.1 Direction of movement: (angle of convergence)
c.2 Velocity: average speed while foveating an object

Eyelids motion:
1. Elevation motion: (angle of elevation, positive value)
2. Closure motion: (angle of elevation, negative value)
3. Duration: (speed)
4. Blink Latency: time interval between presenting a visual (or other) stimulus and the initiation of a blink

For the eyebrow region, the Facial Action Coding System (FACS) [18] provides a complete list of possible facial deformations relative to the eyebrow and eyelid areas. In particular, we have focused on AUs (Action Unit) 1 (Inner Brow Raiser), 2 (Outer Brow Raiser), 4 (Brow Lowerer), 7 (Lid Tightener), and 44 (Squint). AUs 1, 2 and 7 are shown in Fig. 2.

For the eyelid region, Fig. 2 shows the eyes with eyelids and eyebrow areas. In particular, we have focused on AUs (Action Unit) 1 (Inner Brow Raiser), 2 (Outer Brow Raiser), 4 (Brow Lowerer), 7 (Lid Tightener), and 44 (Squint). AUs 1, 2 and 7 are shown in Fig. 2.

A. Eye Movements

The six extrinsic eye muscles, whose function is to move the eyeball, have been modeled as extruded surfaces and their relaxation and contraction are simulated using soft modification deformers. Such deformers allow the 3D muscles to deform while remaining attached to the outer tunic of the eye, and thus to rotate the eyeball (see Fig. 3).

Saccadic, pursuit and convergence motions are also controlled by three directional (aim) constraints represented in Fig. 3 by the two green circles and the four-arrow symbol (which we named Focus Ctrl). The green circles control the rotation of each eye independently, and the four-arrow symbol controls the rotation of both eyes simultaneously. Saccade amplitude and direction, pursuit direction, and vergence direction can be manipulated by changing the 3D location and/or weight of the constraints; saccade, pursuit, and vergence velocities are controlled by the velocity settings programmed in the prototype controller device (described in section III).
B. Lens Thickness and Curvature

The lens of the human eyes can accommodate to light waves from distant or close objects. Light waves from distant objects (farther than 6 meters from the eye) travel in nearly parallel paths and require very little bending by the lens, after they have passed through the cornea, to be focused on the retina. Light waves from near objects (nearer than 6 meters from the eye) are more divergent and require greater bending by the lens to be brought into focus on the retina. When the eyes are focused on a near object, the ciliary body muscles contract and allow the lens to become thicker. When the eyes are focused on a distant object, the ciliary body muscles relax and allow the lens to be thinner.

To simulate the process of lens accommodation, we have applied a cluster deformer to the surface of the ciliary muscle and a blendshape deformer to the surface of the lens. By altering the value of the ciliary muscle cluster scale parameter, we produce its relaxation (positive values) or contraction (negative values). The value of the cluster scale parameter drives the value of the envelope of the blendshape deformer applied to the lens. Positive values of the cluster scale parameter output a blendshape envelope value between 0 and 0.5 (thin lens); negative values of the cluster scale parameter output a blendshape envelope value between 0.6 and 1 (thick lens) (see Fig. 3, right frame). The relation between thickness and curvature of the lens and object distance is documented in several publications. We based our model on [19] and [20].

altering the value of the cluster scale parameter. In addition, we have programmed a ‘light intensity’ parameter whose value controls the amount of light in the scene. The value of the light intensity parameter drives the size of the pupil (see Fig. 4). The mathematical relation between light intensity and pupil size is derived from the servo-analytic study of pupil reflex to light by Stark and Sherman [21].

D. Eyelid Motion

The eyelid motion of each eye is produced by rotating a weighted joint deformer whose pivot point is placed in the center of the eyeball. The elevation motion and the closure motion are determined respectively by positive and negative values of the joint rx parameter. The duration (speed) is controlled by the velocity settings programmed in the prototype controller device (described in section III). Examples of eyelids motion are shown in frames g,h, and i of Fig. 2.

E. Accommodation Parameters

As mentioned previously, the lens changes shape in order to accommodate to light waves from distant and close objects. During this process of accommodation two other events occur that help to create a sharp image of a near object: the pupils constrict and the eyes converge, as both medial rectus muscles contract slightly to direct both eyes toward the near object.

We have programmed three accommodation attributes (boolean) (“AutoConvergence”, “AutoPupillaryConstriction”, and “AutoLensAccommodation”, see Fig. 3 on the right). When the attributes have a value equal to 1 (on), the eyes accommodate automatically as “Focus Control” is moved closer or farther from the eyes; when the parameters are off, the eyes do not accommodate and convergence, pupil size, and lens thickness can be altered manually (See attributes in the left frame of Fig. 3).

F. Dynamic Cross-Section

To reveal the internal structure of the eyes we alter the position of the near clipping plane of the virtual camera. The location of the near clipping plane is controlled by two custom parameters: “Cross-section toggle” and “Cross-section Move”. When the value of “Cross-section toggle” is equal to 1, the virtual camera near clipping plane is placed in the center of the eye and “Cross-section Move” is enabled. By changing the value of “Cross-section Move” the user alters the location...
of the near clipping plane and thus the position of the cross section (see Fig. 5).

G. Eyebrow Region Deformations

All deformations relative to the eyebrow region (including AU 1, 2, 4, 44, and 7) were produced using weighted joint deformers. The values of the joints’ rx parameters control the eyebrow motions and relative deformations of the skin.

In order to use the parameterized model in an interactive and web deliverable platform, the model was imported in Macromedia Director MX via the Maya 3D Shockwave Exporter. To achieve portability to Director and high speed of response in a web deliverable environment, the poly count of the 3D model was reduced and certain deformers were modified because not supported by the platform (i.e., clusters, blendshapes, and soft modification deformers were replaced by joint deformers). In addition, various parameters (i.e., “accommodation” and “dynamic cross-section”), originally programmed in MEL (Maya Embedded Language), were reprogrammed in Lingo.

Demos of the simulator are available at http://www.tech.purdue.edu/cgt/13/eye_sim.html

The 3D eye simulator has been evaluated for anatomical accuracy throughout its development by Purdue and Butler University faculty and graduate students in biology who have provided continuous feedback on accuracy and readability of the animated model. In addition, they have contributed to the design of the interface in order to use the eye simulator as an interactive visualization tool for medical education.

Although designed primarily for WASD (WebSphere Studio Application Developer) we have found that the n52 can be adapted for the eye-movement simulation and can interact with both Maya and Director; adaptation required software development for both programs.

The Nostromo key layout and different modes were customized by the use of a profile which allowed for remapping single-key assignments, creating macros, setting mouse control functions, programming control (direct input) functions, and setting shift states. All the oculomotor and expression parameters with corresponding controller types are listed in Table I. The key layouts relative to each mode of operation (shift state) are represented in Fig. 6. Mode 4 is used to control the velocity of eye, eyelid, eyebrow, and camera motion. When in mode 4, a “Speed Window” displaying two text boxes appears within the 3D simulator window. The user can use the Nostromo controller as a numeric keypad to enter different speed values for eye/eyelid/eyebrow and camera movement. Entering different speed values corresponds to setting the controllers with different discrete steps of user-defined resolution.

III. THE PROTOTYPE CONTROLLER DEVICE

The degrees of freedom (parameters) to control the simulator are 30 including 4 degrees of freedom for the virtual camera (rx, ry, tz and d --near clipping plane distance--) and 1 degree of freedom for the light (intensity). This is of the same order of the number of parameters controlled in state-of-art videogames. Taking advantage of the technology developed for game controllers, we adapted a commercial versatile game controller (Belkin Nostromo Speedpad n52) [22], and we reconfigured it for eye motion/expression simulation. Using different modes, it was possible to achieve 104 functions.

![Fig. 5 Four different cross-sections produced by changing the value of the “Cross-section_Move” parameter; “Cross-section_Control” menu on the right](image-url)

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>OCULOMOTOR PARAMETERS WITH CORRESPONDING CONTROLLER TYPES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Parameter</td>
</tr>
<tr>
<td>Mode 1</td>
<td>1.1 Pupil Size*</td>
</tr>
<tr>
<td>Mode 2</td>
<td>2.1 Saccade latency</td>
</tr>
<tr>
<td>Mode 3</td>
<td>3.1 AU 1 Direction of movement</td>
</tr>
<tr>
<td>Mode 4</td>
<td>4.1 Saccade velocity</td>
</tr>
</tbody>
</table>
4.5 AU 1 velocity
4.6 AU 2 velocity
4.7 AU 4 velocity
4.8 AU 7 velocity
4.9 AU 44 velocity

Fig. 6 Belkin Nostromo Speedpad n52 key layout for each mode of operation; screenshot of “Speed Window”, bottom right

IV. CURRENT APPLICATIONS

EMOES is currently being used in ‘Mathsigner 2.0™’, an interactive computer animation program for learning K-6 arithmetic skills, aimed at deaf children [23-25]. The program includes 3D animated signers who teach mathematical concepts and terminology in ASL (American Sign Language) (see fig. 7). The 3D eye simulation system is used to generate, in real-time, realistic animation of the avatars’ eyes and eyebrows region in sync with the signing motion.

Research on signed communication shows that hand shapes and motions are not the only carriers of linguistic information in signed discourse. Some lexical items are formed by facial movements (not hand movements) and specific motions of the eyes and the facial musculature are important in marking negatives, interrogatives, and relative clauses. For example, a signer’s gaze in the direction of a spatial referent can function as a pronominal reference, and a tight closure of the eyes during a sign can serve to make the sign more emphatic. Lowered eyebrows can indicate negation, while raised eyebrows signal a question [26, 27].

In the Mathsigner application, the virtual signers’ hand/arm configurations and motions are generated from a library of signing clips recorded directly from a signer wearing a 19-markers motion capture optical suit and a pair of 18-sensors cybergloves. Smoothness of motion (in real time) and interactivity are achieved via programmable blending of the motion captured animation segments.

Eye motions and expressions are instead produced and recorded using EMOES, and subsequently stored in an animation database. Real-time playback of the animation clips, in response and in sync with the concept being signed, is programmed using calls to the animation database.

Although with EMOES it is possible to represent anatomically accurate eye motions and expressions, still it is not easy to automatically time the animations in response to the meaning of what is being signed. Presently, we are following research results by Baker and Padden [26], and by Wilbur [28] to formulate an appropriate linguistic generalization for the occurrence of (i) eye gaze behavior (i.e., saccadic and pursuit movements), (ii) eyeblinks, and (iii) eyebrow motions during signed communication of mathematical concepts.

In addition, the 3D eye simulation system is being adapted for display and interaction in a virtual immersive environment (i.e., the Fakespace FLEX virtual theater [29]) for use in science education. To display the eye model, with realistic deformations of eyelids and eyebrow region, in the FLEX, we are using the Cal3D character animation library [30] along with OsgCal. OsgCal is an adaptor for using Cal3D within OpenSceneGraph [31] the format necessary for use with VRJuggler [32]. A pair of Fakespace Labs [29] pinch gloves, along with a 6 degrees-of-freedom tracker by Intersense [33], allow the user to interact directly with the 3D eye simulator by detecting when the user makes gestures that can affect the virtual eyes, such as moving an object closer or farther away from the eyes, thus causing ‘accommodation’ events to occur.
V. CONCLUSION AND FUTURE WORK

The 3D eye simulator presented above is still to be considered a prototype since many of its features are only at a first stage of development. But in spite of its limitations, EMOES is, to our knowledge, the most advanced and physiologically accurate eye simulator available for applications to medical education and HCI. The 3D model is characterized by 30 parameters (Table I). The 30 parameters can be controlled manually with: initial settings, continuous controller types (cartesian, XY, or polar), and discrete controller types. In simulation, the controllers can be set using discrete steps of user-defined resolution. This method of controlling the 30 parameter model can be implemented practically by using various input devices. As an example we have adapted a multi-use game controller [22] which has a sufficient number of functions achievable with 4 different shift states.

Many applications to medical simulations are conceivable using this simulator even at this stage of development. Pedagogical applications are immediately possible. Applications to game design are also easy to envision. For example, realistic visual tracking of various targets (e.g. projectiles, sports items, etc.) can be incorporated in many situations to increase the immersive appeal of the game.

On the other hand, much remains to be done before eye and ocular expression simulators may become commonplace in HCI and face to face computer human interaction. The main challenge, in our opinion, remains what we have called the ‘eye-sync’ problem. Although, with the current simulator, it is possible to represent accurately significant emotional ocular expressions (including pupil dilation and eyebrow movements) still it is not easy to see how such expressions can be automatically generated in response to verbal and or signed communication. As discussed in the introduction, the main reason is the difficulty to extract from the verbal message its emotional and semantic value, and to quantify it. The solution to this problem lies primarily in applied cognitive science. In fact we can see that the coupling between ocular simulation and cognitive science is two-fold: from one side ocular simulation is a tool for cognitive science as it can be used to study ocular response to emotions and meaning; from the other side, cognitive science should provide the tools for translating emotional and semantic content into quantifiable variable to operate the ocular simulator. Thus, the future of face-to-face human computer interaction hinges critically on this hybrid area of future research.

REFERENCES