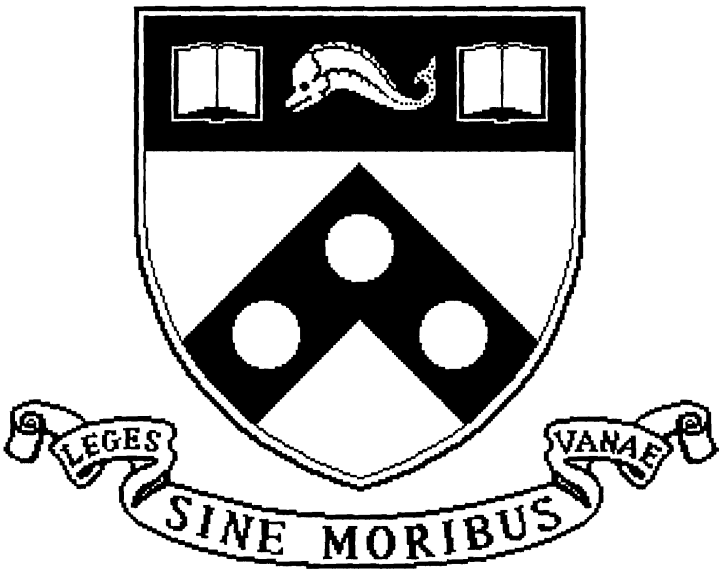
Structure-Based Animation Of The Human Face

MS-CIS-91-15
GRAPHICS LAB 38

Stephen M. Platt
Aaron T. Smith
Francisco Azuola
Norman Badler
Catherine Pelachaud



University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

1991

Structure-Based Animation Of The Human Face

MS-CIS-91-15
GRAPSHICS LAB 38

Stephen M. Platt
Aaron T. Smith
Francisco Azuola
Norman I. Badler
Catherine Pelachaud

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

February 1991

Structure-Based Animation of the Human Face

Stephen M. Platt, Aaron T. Smith,
Francisco Azuola, Norman I. Badler, Catherine Pelachaud

Stephen Platt: Storage and Retrieval, Havertown, PA 19083

Aaron Smith: Department of Computer Science, Brown University, Providence, RI,

Francisco Azuola, Norman I. Badler, Catherine Pelachaud:

Department of Computer and Information Science

University of Pennsylvania

Philadelphia, PA 19104-6389

Abstract

The face is an interesting object to animate for several reasons: it is an important channel of communication and therefore important to any human body animation, and it is a complex object in that it is composed of many nonrigid interacting nonarticulated regions. In this paper, we examine the face, and present it as as a hierarchically structured regionally defined object. Based on this regional decomposition, and a set of primitive actions, we describe an encoding of a large set of high level facial action descriptors. We also present an application which studies the interaction between intonation and facial expressions for a given emotion. It offers a higher level of representation of the action units by grouping them into specialized functions (lips shape for phonemes, eyebrow movements). An animation system linked to facial motion property is also presented.

Key words: facial structure, facial animation, speech synthesis.

1 Introduction

A majority of the efforts in animation of human figures have concentrated on the animation of the body, as it is easily representable as a rigid, jointed object. However, the most expressive communication channel of the human figure, the face, has not been analyzed to the same depth as the rest of the body.

The face exemplifies certain problems which are present, although initially ignorable, in the rest of the body. Its actions are mainly those of tissue movement – masses of flesh will move up and down, in and out, to create such motions as eyebrow raises, nostril flares, and smiles. These motions cause the flesh to bulge, bag, and pouch; in addition, the skin surface can wrinkle and furrow, and lines can appear and disappear. All of these actions may naturally occur anywhere on the body; however, on the rest of the human figure, these actions are secondary to the articulations.

The face itself functions in known and defined regions. Ekman and Friesen defined the Facial Action Coding System **FACS** [EKM78], an anatomically based notation of the expressive abilities of the face. **FACS** describes the actions of the face in terms of the visible and discernable changes to noted regions such as the eyebrows, cheeks, and lips. This is due to the inherent functional and structural nature of the face – a mass of facia, surrounded by skin, muscle, and bone, tends to move as chunks when a force is applied to it.

Our current research involves the construction of a structural model of the human face. We decompose the face into a set of regions. Each region is a section of tissue which functions as a single unit – it is

this feature which enables us to treat the region as a single functional block. The face itself is a hierarchy of these regions connected in a natural manner. A set of primitive actions have also been described; the actions described by **FACS** have been encoded on top of these primitive actions, yielding high-level facial action control. Using these high-level action units, an animation system for speech synchronization has been constructed.

Implementing the face as a hierarchy has afforded us two advantages over non-hierarchical systems:

Descriptive control : descriptive information could be placed fairly high in the hierarchy – for example, information about the furrowing properties of the Nasolabial Furrow was placed in the NASOLABIAL_FURROW region and inherited by the UPPER and LOWER Nasolabial Furrow subregions.

Expressive control : an action affecting an entire region could be applied to the entire region (and would propagate down to the constituent subregions); an action affecting only portions of a region could be applied only to those particular subregions.

Both of these approaches were used to simplify the description of the face and its actions.

In this paper, we summarize a number of techniques which have been used to model and animate the face. We introduce a structural, hierarchical model of the face which, when combined with a simple set of motion primitives, lends itself to facial action description schemes such as **FACS**. The simulation scheme used has a number of advantages over prior schemes; these are discussed in a later section. An application is finally presented.

2 Facial Models

The first major model of the human face was that of Parke [PAR74], [PAR82]. This model parametrically defined the face – a large set of parameters can be varied to redefine the shape of the face (e.g. distance between eyes) or animate it by changing certain segments (e.g. size of mouth opening). In this model, there is little differentiation between the physical description process and the animation process. Parke's model has been extended to encompass more advanced techniques of generating speech and expression.

Platt [PLA81] described an alternate approach to facial animation, one based on the underlying structure of the face. By simulating the muscles and facia of the face, this model naturally produced the bulging and bagging which occurs when actions are applied to the face. However, computational and descriptive complexity prevent this scheme from becoming more generally useful.

A muscle based scheme for animation control has also been implemented by Magnenat-Thalmann et al. [MAG87a]. Their scheme creates procedural definitions of the functions of abstract muscles on abstract faces. It is readily mappable to alternate data sets, such as those emulating the faces of Marilyn Monroe and Humphrey Bogart [MAG87b].

A third muscle-based animation scheme is described by Waters [WAT87]. Rather than defining procedures which implement muscles, he algebraically describes the function of a muscle as a regular distortion upon a

flat grid. This scheme is fast and reasonably effective, however, its generality forces it to ignore some of the peculiar nuances of the human face. Waite [WAI89] describes a scheme similar to that of Waters in that it controls a flat rubber-sheet representation of the skin surface by parametrically controlling the distortion. However, while Waters used parametric control of a simple spring-grid distortion of the skin surface, Waite controlled the distortion of the entire grid using B-spline surface. This allows more exact modification of the basic shape change at the expense of a less anatomically intuitive model.

Terzopoulos and Waters [TER90] have developed a physically based facial model. They integrate the representation of the various layers of the facial tissue with dynamic simulation of the muscles movement. The skin is constructed from a lattice whose points are connected by springs; the stiffness value of the springs on each layer depends on its biomechanical property. The model offers a point-to-point control and is able to yield more accurately subtle facial deformations such as wrinkles and furrows which are difficult to reproduce with a geometric model.

These systems have been used to generate a number of applications involving the animation of the human face. Examining these applications present a number of interesting insights into the weaknesses and strengths of particular control techniques.

Bergeron [BER85] controlled the fictional animated character Tony De Peltrie's face by mapping it point-by-point onto a test subject's face and digitizing the changes the subject's face experienced when performing certain expressions. These location changes were then applied to the De Peltrie face, resulting in a library of standard expressions. Both lip synchronization and general expression animations were created from this basic library. The animations were in essence key-frame animations based on parametric control of a fixed, caricature-style model. The tediousness of the animation specification described by Bergeron point out the weakness of directly using any low-level description of the face for final control.

Pearce et. al. [PEA86] have encoded a rule-based "phoneme compiler", which takes a sequence of phonemes and translates them into the appropriate mouth and lip actions. This scheme relies on having the animator specify the exact sequence of phonemes to be animated. Lewis and Parke [LEW87] use the alternate approach of performing limited speech analysis using linear predictive coding on spoken phrases to automatically create lip-synched animations. Hill, et. al. [HIL88] later describe a rule-based phoneme generator which also uses spoken text as input, generating phonemic output. All three of these systems produce as a final product parametric descriptions of human faces based on the work of Parke, however, Hill notes, "We should very much like to adapt our method to use a face model based on real muscle and bone structures, using the Facial Action Coding System." due to the belief that "the resulting image should be much more natural."

Pieper [PIE89] describes a physically-based model for animating the face which uses a great amount of detail of the internal structures (facia, bone, muscle) of the human face. In this case, efficiency needs to be sacrificed in favor of accuracy; since the goal of this system is simulation of reconstructive surgery, a true simulation of the face, preserving and accurately representing its actions, is necessary instead of the more common emulation, which just presents a surface image which appears to function correctly.

These applications encourage a two-layer division of the processes needed to describe the animation of

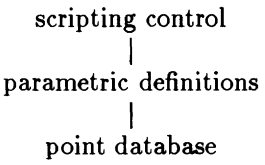


Figure 1: Layers of Parametric Control

the face. At one level, the major concern is that of the structures needed to represent the primitive actions of the face. A second level involves the techniques of describing those actions in a manner easily manipulatable, either directly or under program control.

3 Action/Object Hierarchies

An analysis of facial animation schemes allows us to classify them into four different models of action/object hierarchy (including the model being introduced here). Each hierarchy has at its lowest level a database of points and surfaces forming the skin of the face. Likewise, each has at its highest level a set of actions performable on the face. The structures and controls defined upon the basic database determine the general flexibility of the system, and in most cases, the efficiency as well.

Note that the principle area of concern is not the higher level of control. Once a set of primitive operations has been defined (parameters, muscle actions, functions, motion primitives), any number of animation control schemes can be defined. Two schemes predominate: direct control, useful where precise control is needed such as for certain speech synthesis applications, and **FACS**, useful where simplicity of high-end control and data independence is desired. There is nothing which theoretically prevents the use of either of these specification schemes with any of the control schemes. The principle area of concern is in the structure of the underlying database, and the interaction between the primitive operation set and the database.

Parametric Control schemes such as those of Parke have the positions of the underlying point database defined in terms of controllable parameters. Control of the system is in terms of manipulated parameters; there is no interaction between low-level elements (points). Higher level animation is defined in terms of the second-level controls. Figure 1 shows the different levels of control.

Physically Based systems ([PLA81], [TER90]), partition the data set into skin (displayed) elements and muscle (nondisplayed) elements. A parametric force is applied only to the nondisplayed elements; there is significant (and costly) interaction between the entire database elements to create the desired effect. Interactions are only among the low level elements. Figure 2 shows the different levels of control.

Functional Control schemes ([WAT87], [WAI89]) define functional operations upon low-level elements. Similarly to the Partitioned Parametric Control schemes, the underlying points may interact, although in a more limited manner. However, the Functional Control schemes allow control in a manner similar

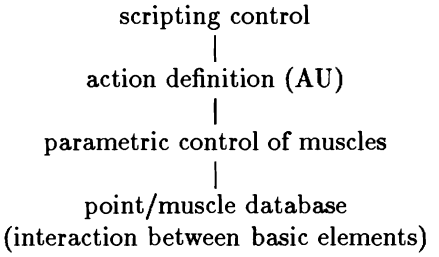


Figure 2: Layers of Physically Based System

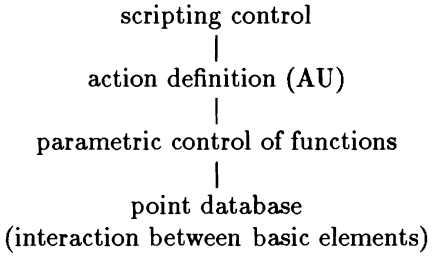


Figure 3: Layers of Functional Control

to that of basic muscle control, but with greater computational efficiency. Figure 3 describes the levels of control.

A Hierarchical Control scheme decomposes the face into a network of functional regions, allowing abstracting of the interactions and action definitions to a level above that of the raw point dataset. It is equivalent to any other abstracting scheme: for example, moving a *chair* instead of moving each point being used to define the *chair*.

In the hierarchical control scheme, points are used to define the basic shape of the skin surface. A small network of regions is defined in terms of these points. Basic actions are described as operations upon regions (instead of operations on points); higher level actions are described as collections of basic actions. The most critical difference is that object interaction occurs at the regional level, instead of the point level.

1. Properties of masses of tissue which function as a coherent whole are defined upon the region, not upon the individual points.
2. The point data set can be modified or replaced to produce new face geometries, without modifying the region database.
3. The point set can be replaced with a new data set, without modifying the region database.
4. Muscles act upon regions; a muscle action inspired system is simply defined as primitive region motions.

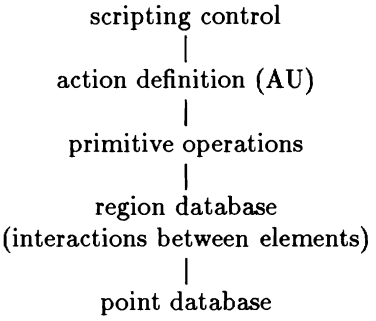


Figure 4: Layers of Hierarchical Control

Figure 4 shows the layers of control of this system.

4 Development of a Hierarchical Control Model

We have developed a model of the human face based on the hierarchy formed by its functional behavior and its physical structure. In this section, we will describe the motivation behind our functional decomposition of the face into structural regions, and outline the technique by which this was performed. In later sections, we will describe in greater depth the hierarchical structure defining the face, and the techniques used to apply actions to this structure.

FACS was developed as a descriptive system based on the basic actions (e.g. Brow Raiser) performable on the face. Each of these basic actions is the result of the contraction or relaxation of a single muscle fiber or a small group of related muscles. Although initially designed for analysis, this muscle-based approach created a simple translation scheme enabling these animation systems to specify their changes in terms of **FACS** Action Units (**AUs**).

However, **FACS** was designed as a system to be used to analyze images of human faces. It was designed for psychologists, not anatomists, and is based upon observation of the effects of muscle actions rather than the muscles themselves. The **AUs** of **FACS** are described in terms of the changes to regions of the face – for example, *AU1: Inner Brow Raiser* is described as:

1. Central and medial brow will move up;
2. Area between the brows will move up;
3. Central forehead will wrinkle in some cases;
4. Medial forehead may wrinkle or bulge;
5. Lateral brow may be pulled in extreme cases.

The nature of these descriptions indicates a number of important features about the face:

Figure 5a goes here
[(a) Major regions of the face]

Figure 5b goes here
[(b) Subregions of the face]

Figure 5: A Regional Decomposition of the Face

1. The face itself moves in predesignated chunks such as the brow, cheek, and lips.
2. These regions are well-defined on any particular human face, but may vary in exact position, size, and shape between faces.
3. The functionality of these regions and interactions between regions are reasonably consistent from face to face, regardless of the fact that the exact size and shape of the regions may differ.

This indicates that a regional representation of the face would allow both simplified **FACS** encoding as well as portability between actual facial images. To demonstrate this, the following steps were performed:

1. **FACS** was analyzed to create a functional atlas of the human face, as well as a catalog of performable actions on the regions of the face.
2. The description of **FACS** was formalized, producing an exact description of what changes each **AU** causes.

A complete atlas of the face, as well as the catalog of basic instructions and formal description of **FACS AUs**, has been described by Platt [PLA85]. During the analysis, a number of additional observations were made about the structure of the face and its functionality:

1. Most actions take place upon small regions (subregions) such as the central, medial, or lateral eyebrow.
2. For any particular **AU**, some of the actions were caused by direct muscle action, while others were caused by propagation of the initiating action to adjacent regions.
3. The secondary actions were rarely (if ever) required parts of an **AU** description – in very weak performances of the **AU**, they may not cause sufficient change to be noticed.

Figure 5a shows the general regions of the face which resulted from this analysis, and figure 5b defines in greater detail the subregions of the human face.

5 Description of Facial Regions

A region is a mass of tissue which functions as a single unit. It is more than a single point of skin – it can represent different amounts of skin surface as well as the underlying tissues.

Figure 6a goes here
[(a) Created Hierarchy of Regions]

Figure 6b goes here
[(b) Instanced Hierarchy of Regions]

Figure 6: Created and Instanced Hierarchies

Our current region map was obtained by analyzing **FACS** and cataloging all regions defined within it. This map was extended to account for nonactive skin surface patches, effectively completing a facial surface atlas. The logical hierarchy of regions is diagrammed in figure 6a; figure 6b is a more complete tree demonstrating the duplicity of instancing.

Each minimal region maintains information on the current state of that particular patch of tissue. This state consists of three types of information:

Physical information : a description of the information needed to display this region.

Functional information : the current state of the region: how far from 'home' it has been moved, etc.

Connective information : what regions it is connected to (for action propagation).

A high-level region (e.g. FOREHEAD) contains information pertinent to the entire region. Actions may be applied to an entire region or to any of its constituent subregions (down to the level of the minimal region). We initially implemented this structure in LISP, and have reimplemented them in C. From this we can notice the following:

1. In each implementation, each *type* of region (e.g. brow-lateral) had to be described only once. This description inherited properties as needed from a parent (containing) region. When instanced, the side (left or right) of the face would be assigned, determining which direction is "out".

2. In the LISP implementation, the data structure was created as part of the execution-time loading. This was done for ease of implementation.

3. In the C implementation (and unlike the LISP implementation), there was only one type defined to describe any region of the face. Since non-object-oriented C does not support object hierarchies, they were eliminated.

4. In the C implementation, the actual structural information was stored in a file and loaded at run-time, rather than in the original system source. This was done for reasons of flexibility – regionally significant paramaters could be easily modified to allow different facial behavior patterns.

In the current model, the physical information is composed of a collection of references to 3D points and surfaces. These points can be modified by the actions; it is through this modification that an action can affect the displayed image. The point structure adds the bottommost layer beneath the leaves of the trees

in the hierarchy. At present, our implementation does not take full advantage of this fact; the point is used solely to store 'current' and 'next' position information. The surfaces are not used by the animator – this information is passed unchanged to the renderer.

Each region r has a designated point, the centroid C_r. Motion of this point characterizes motions of the region, and motion of all other points can be derived from the motion of the centroid and the centroids of neighboring regions.

The functional information is used by the action descriptions to control the modification process. This information is used to quickly describe the logical state of the region and thus determine to what extent (and in what manner) the corresponding physical information should be changed.

The connective information is essentially a list of adjacent regions. The action application process includes a step in which a modified form of the action is propagated to adjacent regions. For example, moving the Brow UP 0.1 inches will also move the Above-Brow UP 0.04 inches.

The exact parameters defining the propagation are a function of the action parameters and the functional information of the region, and were determined empirically. The parameterization of the actions was a one-time process – we studied each action unit as performed to determine to what extent it could move its initiating regions – for example, the Inner Brow Raiser (AU1, as shown in figure 10) could raise the medial brow approximately half the distance the medial brow is capable of moving; hence the motion applied to the medial brow would be half of the total magnitude of the motion. The central brow, on the other hand, recieves the full amount of the motion. Parameterization of the facial regions was also determined empirically – we examined each region of the face to determine how far it could be pushed and shoved by any action, and used this to determine the "maximal motion" amounts. Although lacking formal studies, we believe these data to be reasonably accurate as faces do share a common structure between people; it is merely the relative size of regions that changes.

Modifying the physical information will modify the shape of the face. Modifying the functional information will affect *how* the face functions when actions are applied. By changing any of these two sets, either the shape of the face (allowing the model to be "fitted" to different people) or the functionality of the face (allowing "tightening" of flesh regions, etc.) can be modified. The underlying structural definition (the region model) remains unchanged – in this manner, we have separated the method of defining the physical shape of the face from the process of creating action sequences on it. The initial LISP and C implementations used a 400-point data set to physically describe the face. We later replaced this with a 1200-point set without needing to modify any other structures. We have also added the back of the head and the inner part of the lips to this model; the eyes and eyelid are considered as separate figures since their motions involve "joint action" (rotation along an axis) and not object deformation. Images presented in this paper were generated using the later data set.

Finally, we should note that there is no practical reason to change the connective information, since the connective information is intimately tied to the regional decomposition of the face.

```
AU 1 -- Inner Brow Raiser

Primary Changes:
    PC1:  Move(dir=up): (* brow central)
    PC2:  Move(dir=up): (* brow medial)
    PC3:  Move(dir=up): (between-brow)
Secondary Changes:
    SC4:  Wrinkle(angle=0): (forehead central)
Tertiary Changes:
    TC5:  Wrinkle(angle=0): (forehead medial *)
    TC6:  Bulge: (forehead medial *)
    TC7:  Move(dir=central): (* brow lateral)
Predicates:
    P1:   TC5 -> SC4
    P2:   not (TC5 and TC6)
    P3:   TC7 -> PC2
Mandatory:
    PC1 or SC4
```

Figure 7: Action Unit 1

6 Description of Facial Actions

The **FACS AU** is the "handle" used to describe animated facial actions. A single **AU** is enacted by the contraction of one or more closely related bundles of muscle fibers. This causes an initiating motion on the skin surface which in turn may cause propagated changes. This motion is affected by the current state of the affected regions – normal flesh conditions as well as previously and concurrently applied actions may affect this condition.

One Action Unit, AU1, is presented in figure 7, and its implementation is presented in figures 8 and 9. The AU1 raises the inner portion of the brow; it can be performed both unilaterally (denoted AU1R and AU1L for right and left AU1 actions) and bilaterally (denoted AU1). A unilateral AU1, such as AU1L, would be performed by initiating the primary changes associated with the AU1 description on the appropriate side of the face (raising the (left brow central), the (left brow medial), and the (between-brow) regions). Since the (bilateral) AU1 is effectively the simultaneous performance of the actions AU1L and AU1R, we implemented the AU1 as concurrent applications of AU1L and AU1R.

An animation of the bilateral AU1 (both brows) is shown in figure 10. Figure 10a shows a neutral face (one with no expression). Figure 10b shows the performance of the AU1 – the brow has been raised, bulging the forehead and pulling up the skin below the brow. Both central brow regions have been raised, causing secondary actions in the forehead, medial and lateral brows, and the eye cover folds (above the eyes). This image clearly shows the skin areas which have been pulled to a higher-than-normal level. Likewise, figure 11a rotates the jaw to approximately half of its full extent without any other actions being performed. Figure 11b is a full "smile": the mouth is open, the lips are pulled back, and the cheek is slightly raised.

At this point we should mention a slight discrepency between the **FACS** analysis of the AU1 and the

```
(setq all-moves
  . . .
  ( AU1 . (make-action
             :propagators
               ( ( (all) (AU1R the-params)
                         (AU1L the_params) ) ) ))
  ( AU1R . (make-action
             :propagators
               ( ( (right brow central)
                   (move ( (dir up)
                         (intens (get-param intens))))
                 ( (right brow medial) . . . )
               )
           ) )
  . . .
)
```

Figure 8: Implementation of Action Unit AU1 (LISP)

from the action definition file

```
MACRO AU1R
RIGHT-BROW-CENTRAL UP 1.0
RIGHT-BROW-MEDIAL UP 0.75
BETWEEN-BROW UP 0.5
RIGHT-ABOVE-BROW-CENTRAL UP 1.0
MACRO AU1L
LEFT-BROW-CENTRAL UP 1.0
LEFT-ABOVE-BROW-CENTRAL UP 1.0
LEFT-BROW-MEDIAL UP 0.75
MACRO AU1
RIGHT-BROW-CENTRAL UP 1.0
RIGHT-BROW-MEDIAL UP 0.75
BETWEEN-BROW UP 0.5
RIGHT-ABOVE-BROW-CENTRAL UP 1.0
LEFT-BROW-CENTRAL UP 1.0
LEFT-ABOVE-BROW-CENTRAL UP 1.0
LEFT-BROW-MEDIAL UP 0.75
```

Figure 9: Implementation of AU1 (C)

Insert brow raise figures here

(a) Neutral (b) AU1

Figure 10: AU1 Performance

Insert jaw actions here

(a) Jaw Rotation (b) Smile

Figure 11: Jaw Actions

Insert lip raise, brow lower figures here

(a) AU10L (b) AU4

Figure 12: Actions Across the Face

actual changes shown in figure 10b. **FACS** was designed as a notation system, and describes the observable changes which occur when an Action Unit occurs. In the case of the AU1 (and other **AUs**, as well), additional minor (non-obvious) changes will frequently take place. In particular, there is a slight motion in the eye cover fold in both humans and the animated face when the AU1 is performed. (To sense this, we suggest placing one's finger tip lightly below the brow and above the eyelid, and raising one's brow. A small motion will be detected). We assumed this was not specified in the **FACS** description due to the practical non-observability of this change.

The face as currently implemented is capable of moving its constituent regions both individually and in concert. Figure 12 shows two more actions, the left upper lip raiser (AU10L) and the brow furrower (AU4). The asymmetric lighting allows observation of the furrowing phenomenon in two regions. When the upper lip is raised (AU10L, figure 12a), the left nasolabial furrow, above the upper lip and below the cheek, is deepened. This is evident on the right side of the figure, just above the face's left lateral upper lip. The brow furrowing (AU4, figure 12b) creates furrows between the brows (evident on the left side of the images).

6.1 Primitive Motions

The **FACS AUs** allow us to define high-level facial actions as sets of primitive changes to primitive regions. As can be seen from this example, there are a small number of primitive changes (primitive actions) which can be applied to a large number of facial regions. An **AU** is implemented as the performance of the first (primary) changes; once started, secondary actions are propagated to connected regions to create additional effects. The exact nature and extent of these secondary changes are a function of the initial action (its nature and parameters), the properties of the subject region, and the properties of the connection to the adjacent region.

FACS AUs are composed of applications of 16 primitive motions:

- MOVE
- LINE
- WRINKLE
- FURROW
- BULGE
- BAG
- POUCH
- ELONGATE

- DE-ELONGATE
- WIDEN
- PROTRUDE
- STRETCH
- NARROW
- FLATTEN
- TIGHTEN
- PUFF

as well as three independant actions:

1. Jaw motions
2. Tongue motions
3. Eye motions.

Many of these actions (BULGE, BAG, POUCH, ELONGATE, DE-ELONGATE) could be defined directly as MOVE operations and were directly implemented accordingly. An equally large set (FURROW, NARROW, WIDEN, STRETCH) are indirect consequences of the MOVE operation and also did not need distinct defining. Other actions (LINE, WRINKLE) described surface changes rather than flesh changes – these were not implemented in the current system, but could easily be added as a scalar state-variable attached to the region. Jaw motions were implemented via general ROTATE operation. The action PUFF and actions of the tongue are all effects of external events, and were not considered within the scope of this study.

A set of minimal actions, applicable to any of the face's regions, has been defined. This set allowed any basic action to be initiated; after the initiation, secondary effects were propagated to adjacent regions. Based on these minimal motions, the **FACS** Action Units have been defined as parametric sets of minimal actions.

7 Application Process

In this section, we explain the algorithm we follow to compute an animation.

7.1 Initialization

Through the concept of "macro" action, the description of any **AU** may be transcribed into a sequence of actual action applications on a set of predefined regions. The initialization process involves the setting of all internal variables done at the regional level and not at the point level. They correspond to the object name, the action name and the parameter value as defined in the definition of the **AU**. For example, in the case where AU1L is activated (see 9), the program will store the following information: the object name, *LEFT-BROW-CENTRAL*; the action name, *MOVE*; and the parameter value, *1.0*. This is done for every action in the description. The Unapplied Action List (UAL) stores these tuples; the main action resolution process will later apply the actions to the designated regions. The list of local variables which are associated with each action is then initialized. The final list of actions is obtained by evaluating the accessibility of each action, the non-accessible ones being rejected.

7.2 UAL Resolution

After initialization, the UAL contains a set of action/region tuples describing the animation being performed. The resolution process applies the actions to the regions and creates propagated actions. When the UAL has been emptied, the application is complete.

The application is composed of two steps: region modification and action propagation. In the first step, region modification, the program code corresponding to the described action (MOVE or ROTATE) is called with the region and other UAL tuple contents as its parameters. This will cause a modification to the region as described by the action.

The next step at this stage is to find where an action is propagated. Each object is connected to other objects by different types of connections. Each connected object is added to the UAL and its amount of displacement is computed.

This process is repeated until the UAL is empty.

7.3 Operations Affecting Regions

We have implemented only two bottom-level operations upon the region database: MOVE and ROTATE. These have proven sufficient to produce a wide range of operations upon the face.

The MOVE operation moves a region by a specified amount within the stated limits and propagates the created motion to adjacent regions. Information contained within each region acts as a limiting constraint to the actual amount of the motion, and since the actual amount is the amount propagated, therefore indirectly affects the motion propagated to adjacent regions. The current implementation of MOVE is fairly simplistic, but effective in that it produces the desired results, both visually and functionally. It considers all actions within a set time segment to be operating in parallel – for any concurrent set of MOVE operations,

```
new_position = old_position +
                    MOVE (old_position,amt_to_move).
```

Therefore, any coincident MOVEs are summed, and as part of the process, clipped to the maximal motion ranges of the region.

A second low-level operation, ROTATE, was defined to allow axial rotation operations. This operation causes an absolute and complete rotation of a region about the mando-templar junction, without regard to region interaction. Effectively, this rotates certain key regions (the jaw regions) as they would be directly pushed by the actual jaw actions. The remaining regions (lips, cheeks, etc.) follow in sequence as the motion is propagated as a sequence of regular MOVE operations.

7.4 Post-Animation

The last phase in an animation step is the Post-Animation process. This phase involves the updating of numerous internal variables, such as the final centroid position. However, the most visually interesting part of the post-animation step is the skin stretching process, which presents the appearance of smoothly deforming regions.

The process of skin stretching is nothing more than the computation of the displacement of the rest of the face points (i.e., non-centroids). In order to accomplish this task, the movement of a given point "i" in region "k" is calculated as the weighted sum of the displacement of the centroid C_k of region "k" and the displacement of the centroid C_j, which is the closest centroid to point "i". The weights are proportional to the distance between point "i" and each of these centroids. The displacement of point "i" follows that of the centroid closer to it, but also is affected by the motion of the centroid of the region it belongs to. In this way, it is possible to have a good control on the motion of the centroid and the points of each region.

Finally, furrows are modelled as a barrier where movement is not propagated through. It is implemented as a pseudo region which deepens into the face, accordingly to the action, or set of actions, being performed. This processing is also performed in the post-animation phase.

7.5 Animation

The basic animation process allows for detailed descriptions of the **AU** sequences to be animated. However, the system does not take into consideration the nature of muscle actions upon the face, such as the contraction and release times. This phenomenon is most apparent during rapid speech, when the mouth positions created for sequences of phonemes is blurred, as the mouth cannot keep up with the speech. It is called coarticulation, and will be discussed in greater depth in a later section.

To account for the problem of blurring of actions applied rapidly to the face, a "smoothing" operation has been defined as a global post-animation step. Rather than being applied to a single parallel action application step (in the manner of the post-animation process), it is applied to the sequence of data sets generated by all of the pre-animate/resolve/post-animate **AU** applications.

The regions of the face are organized into three sets, namely, one with high movement (like the lips or the brows), one with regular movement (forehead or cheeks), and one with low movement (outer part of the face). Every point, for each keyframe, is then modified according to the associated weight of the region it belongs to. These weights are computed by looking at the centroids' position in consecutive key-frames. Weights are calculated considering, for all centroids, the rate of change for the distance traversed by a given centroid, among consecutive pairs of keyframes. Average weights are computed for each of these sets, on each of the keyframes considered. The final location of each node on the face is computed according to the node's corresponding region's (average) weights. It is important to mention that the new position for a given node lies within the convex hull of the original control polygon, which has as vertices the initial locations of that given node through time, i.e. in the different keyframes. The key-frames for the animation contain the new points, which in turn are used to perform the in-betweening. A cubic spline curve, that goes through matching points on consecutive key-frames, is used to compute the interpolated frames.[FAR90].

8 Animation Control using FACS

We can now present an application using this model: animation of the face coordinated with intonated speech. In this section we explain how we built a system of 3D animation of facial expressions based on the link between emotion and the intonation of the voice. Until now, the existing systems ([BER85], [KLE88],

Insert figure of emotions here

Figure 13: Facial Expressions of Emotion

[HIL88], [PEA86], [WAT87]) did not take into account the correlation of these two features. Some of them propose an automatic lips synchronization process but most of the other facial motions remain to be specified manually. This may be a very tedious task. While talking the face remains seldomly still; eyebrow movements accentuate a word, the head moves in harmony with the flow of speech. Our study is based on the fact that many linguists and psychologists have noted the importance of spoken intonation for conveying different emotions associated with speakers' messages. Moreover, some psychologists have found some universal facial expressions linked to emotions and attitudes [EKM84]. Our study considers these six emotions: anger, disgust, fear, happiness, sadness and surprise (see figure 13).

8.1 Facial Expressions and their Rules

In the first step, we enumerated and differentiated facial movements due to emotion from the ones due to conversation [EKM84].

emblems correspond to movements whose meaning is very well-known and culturally dependent.

conversational signals are made to punctuate a speech (occurrence on an accented item within a word), to emphasize it (signal is stretch out over a syntactic portion of the sentence).

punctuators can appear at a pause (due to hesitation) or signal punctuation marks (such as a comma or interrogation marks).

regulators are movements that help the interaction between speaker/listener.

manipulators correspond to the biological needs of the face.

affect displays are the facial expressions of emotion.

These are the different types of facial expressions we need to compute.

8.2 Steps for Computing Facial Expressions

We assume that the speech input is decomposed into a sequence of discrete units with their timing; its intonational pattern is also given [STE90]. A facial expression is expressed as a set of **AUs** and its computation is derived by a rule-based system. We introduce first the underlying property we based our expression generation system on and then how we derive the rules governing the details as a function of phoneme sequences and intonational pattern.

Insert figure "Julia perfers popcorn" here

Figure 14: "Julia prefers popcorn" – slow-speech rate

8.2.1 Synchrony

An important property linking intonation and facial expression (also extended to body movement) is the existence of synchrony between them [CON71]. Synchrony implies that changes occurring in speech and in body movements should appear at the same time. This is the basic principle which regulates the computation of the facial animation.

Therefore, the sentence is scanned at various levels. Blinks and lip shapes are computed at the phoneme level while conversational signals and punctuators are obtained at the word level, by the intonational pattern of the utterance. Synchrony describes exactly when an **AU** will be expressed.

8.2.2 Coarticulation

Lips synchronization is obtained using speech-reading technique [JEF71] which offers the possibility to define lip shapes for each cluster of phonemes. Vowels and consonant are divided into clusters corresponding to their lip shapes. Each of these groups are ranked from the highest to the lowest visible movements (for example, the phonemes 'f', 'v' are part of the top (least deformable) group, while 's', 'n' are very context dependent, and may be greatly deformed). This clustering is speech rate dependent, a person speaking fast moves much less his lips than a person talking slowly. This phonemic notation, however, does not tell us how to deal with the difficult problem of coarticulation. The problem occurs from the overlap of units during their production. The boundaries among phonemic items are blurred.

Some rules look at the context a phoneme is produced to compute the adequate lips position. Nevertheless no complete set of rules currently exists. One remedy is to consider the geometric and temporal constraints over consecutive actions. Muscles do not relax or contract instantaneously. If the time between two consecutive phonemes is smaller than the contraction time of a muscle, the previous phoneme is influenced by the contraction of the current phoneme. In a similar manner, if the time between two consecutive phonemes is smaller than the relaxation time, the current phoneme will influence the next phoneme when relaxing.

The geometric relationship between successive actions is obtained by designing a table of weights corresponding to how similar two actions are. The intensity of an **AU** is rescaled depending on its surrounding context (see figure 14).

8.2.3 Remaining movements

To find the remaining facial expressions, we scan the utterance and its given intonational pattern. The emotion gives the overall behavior of the face (global direction of the head, basic facial expression on top of which other actions are added). Pauses and pitch accents are often accompanied by facial movements such

as eyebrow movements, eyeblinks, or head motion. We compute for each of the group the corresponding actions by a set of rules [PEL91].

The occurrence of facial expressions linked to intonation are also emotion-dependent (a sad person shows few movements of low intensity, while a frighten person has very fast and abrupt facial behaviors). Nevertheless, it is still unknown when a paralanguage feature is accompanied by a facial movement. The function of these facial movements have been established (e.g. focus on one word) but the regularity of the occurence cannot be predicted. Indeed, the attitude of the speaker (what he wants to convey) and his personnality are an important factors in his facial behavior. But such points are not yet incorporated in the present study. Offering a tool to compute separately each of the above groups of facial expressions offers a better grasp and control over the final animation.

8.2.4 Summary

The computation of the facial expressions linked to one particular utterance with its intonation and emotion is done independently of the facial model. Contrary to the technique of using a stored library of expressions ([BER85], [KLE88]) which computes facial expressions for one model only, this method used the decomposition of the facial model into two levels: the physical level (described in previous sections) and the expression level. This yields a script of generated **AUs** which are then applied to a face.

9 Implementation

The initial implementation of this system was in Common LISP on a VAX-785 under VAX/VMS. Output was in the form of a file of points and polygons which were rendered as a post-processing step. Several different formats were available (all polygon/point models) to meet the needs of several renderers then in use.

The complete system was then ported to Lucid Common LISP running on Apollo workstations running UNIX and AEGIS. No changes were made to the programs, and output was kept in the same format. On both of the LISP implementations, execution was rather slow. This was due in part to the nature of LISP (even in a compiled form, memory usage was both large and inefficient) as well as a number of known inefficiencies in the code. Simple animations took an average of 1 to 5 minutes of CPU time; a complex animation performed in small steps (keyframe animation) could take several hours.

During the summer of 1989, the entire system was rewritten in C on the Apollo workstations. As we now knew the structure of the program and how it used the data, a number of internal enhancements were implemented. Output formats are basically the same – points and polygons to be fed to one of several renderers.

The second implementation showed a tremendous improvement over the first. Simple animations are performed in under a second. More complex animations, ones which took roughly 30 minutes under the LISP implementation, would execute in roughly 30 seconds.

The second implementation also allowed us to experiment with different data sets. It was originally

debugged using a 400-point model of the face. After the debugging process, the 400-point model was substituted with a 1200-point model. No changes were made to the animation system, and it performed as before. Due to the hierarchical nature of the face (actions applied to regions, points moved upon completion of the action application), there was no performance loss with the higher resolution face. Regional and inter-region parameters which described the structure of the face did not need to be changed.

The C implementation has since been ported to a Silicon Graphics IRIS. Timings have not been performed, but it animates faces roughly 5 times faster than the Apollo/C version. This version has been adapted to the 'JACK' system [PHI88].

10 Conclusions

We have presented a model of the face which is hierarchical both in its structural definition and in its actions. This parallel decomposition is a result of a structural analysis of the face and its actions – as the face is broken into small functional regions, actions are defined which function at the regional level.

The regional description is independent of the particular face being animated. As the region-set is moved from one face to another, only the low-level parameters (tissue flexibility, point locations) are changed; the general structure of the face remains constant.

Likewise, the primitive actions are independent of the local variations present in different regions. An action is defined as a set of changes to parameters of the region, not as modification of any particular region. This low-level flexibility allows us to repeatedly use the same base action to describe regionally different changes in different sections of the face. High-level actions (**AUs**) complete the action hierarchy, providing a flexible and rich method of describing facial changes.

The multi-level descriptive scheme allows substitution of different point geometries (face shapes) without modification of the facial structures. Alternately, the functionality of the face can be modified on a structural basis without needing to consider the constituent points of the skin surface. This allows abstraction of facial functions and facial features – for example new face shapes (different people) can be animated without modifying the structural information of the system.

Also, We have presented an application of this model in the study of the relation between the intonation of the voice and the emotions of the face. The decomposition of the face into two independent parameters (physical and functional) and its simulation of **FACS** allows this application to be more general and work at a high-level (the level of facial expressions of emotion, conversational signals and others). We offer a tool to analyze, manipulate and integrate different channels of communication (face and voice) and to facilitate further research of human communicative faculties via animation.

11 References

References

[BER85] P. Bergeron, "Controlling Facial Experessions and Body Movements in the Computer-Generated

Animated Short 'Tony De Peltrie'". *SIGGRAPH Advanced Animation Tutorial*, 1985.

[CON71] W.S. Condon, W.D. Ogston, "Speech and body motion synchrony of the speaker-hearer", in The perception of Language, *D.H. Horton, J.J. Jenkins ed.*, 1971: 150-185.

[EKM78] P. Ekman and W. Friesen. Facial Action Coding System, it Consulting Psychologists Press, Palo Alto, CA. 1978.

[EKM84] P. Ekman, "Expression and the nature of emotion", in Approaches to emotion, *K. Scherer, P. Ekman ed.*, 1984.

[FAR90] G. Farin, Curves and Surfaces for Computed Aided Geometric Design, A Practical Guide, 2nd ed., *Academic Press*, 1990.

[HIL88] D.R. Hill, A. Pearce, and B. Wyvill, "Animating Speech: An Automated Approach using Speech Synthesis by Rules", *Visual Computer 3*:277-289. 1988.

[JEF71] J. Jeffers, M. Barley, Speechreading (lipreading), *C. C. Thomas*, 1971.

[KLE88] Kleiser-Walczak Construction Comp., "Sextone for President", *ACM SIGGRAPH'88 Film and Video Show*, issue 38/39, 1988.

[LEW87] J.P. Lewis and F.I. Parke, "Automated Lip-Synch and Speech Synthesis for Character Animation", *Computer/Human Interface + Graphics Interface '87*, 1987: 143-147.

[MAG87a] N. Magnenat-Thalmann, E. Primeau, and D. Thalmann, "Abstract Muscle Action Procedures for Human Face Animation", *Visual Computer 3*:151. 1987.

[MAG87b] N. Magnenat-Thalmann, and D. Thalmann, "The Direction of Synthetic Actors in the Film *Rendezvous a Montreal*", *IEEE Computer Graphics and Applications*, December 1987: 9-19.

[PAR74] F. Parke, "A Parametric Model of the Human Face", *PhD Thesis, University of Utah*, 1974.

[PAR82] F. Parke, "Parametrized Models for Facial Animation", *Computer Graphics and Applications*, November 1982: 61-68.

[PEA86] A. Pearce, B. Wyvill, G. Wyvill, and D. Hill, "Speech and Expression: A Computer Solution to Face Animation", *Graphics Interface '86*, 1986: 136-140.

[PEL91] C. Pelachaud, N.I. Badler, M. Steedman, "Linguitics Issues in Facial Animation", to appear in *Computer Animation'91*, 1991.

[PHI88] C.B. Phillips, N.I. Badler, "Jack: A Toolkit for Manipulating Articulated Figures", *Proceedings of ACM/SIGGRAPH Symposium on User Interface Software,* Banff, Alberta, Canada, 1988.

[PIE89] S. Pieper, "Physically-Based Animation of Facial Tissue for Surgical Simulation", *SIGGRAPH State of the Art in Facial Animation Tutorial*, 1989.

[PLA81] S. Platt and N.I. Badler, "Animating Facial Expressions", **Computer Graphics 15:3**, August 1981: 245-252.

[PLA85] S.Platt, "A Structural Model of the Human Face", *PhD Thesis Department of Computer and Information Science, University of Pennsylvania*, 1985.

[STE90] M. Steedman, "Structure and intonation", *Technical Report MS-CIS-90-45, LINC LAB 174, Computer and Information Science Department, University of Pennsylvania*, 1990. To appear in *Language*, 1991.

[TER90] D. Terzopoulos, K. Waters, "Physically-based Facial Modelling, Analysis, and Animation", **The Journal of Visualization and Computer Animation**, vol. 1, 1990.

[WAI89] C. Waite, "The Facial Action Control Editor FACE: A Parametric Facial Expression Editor for Computer Generated Animation", *M.S. Thesis, MIT*, 1989.

[WAT87] K. Waters, "A Muscle Model for Animating Three-Dimensional Facial Expression", *Computer Graphics 21:4*, July 1987: 17-24.

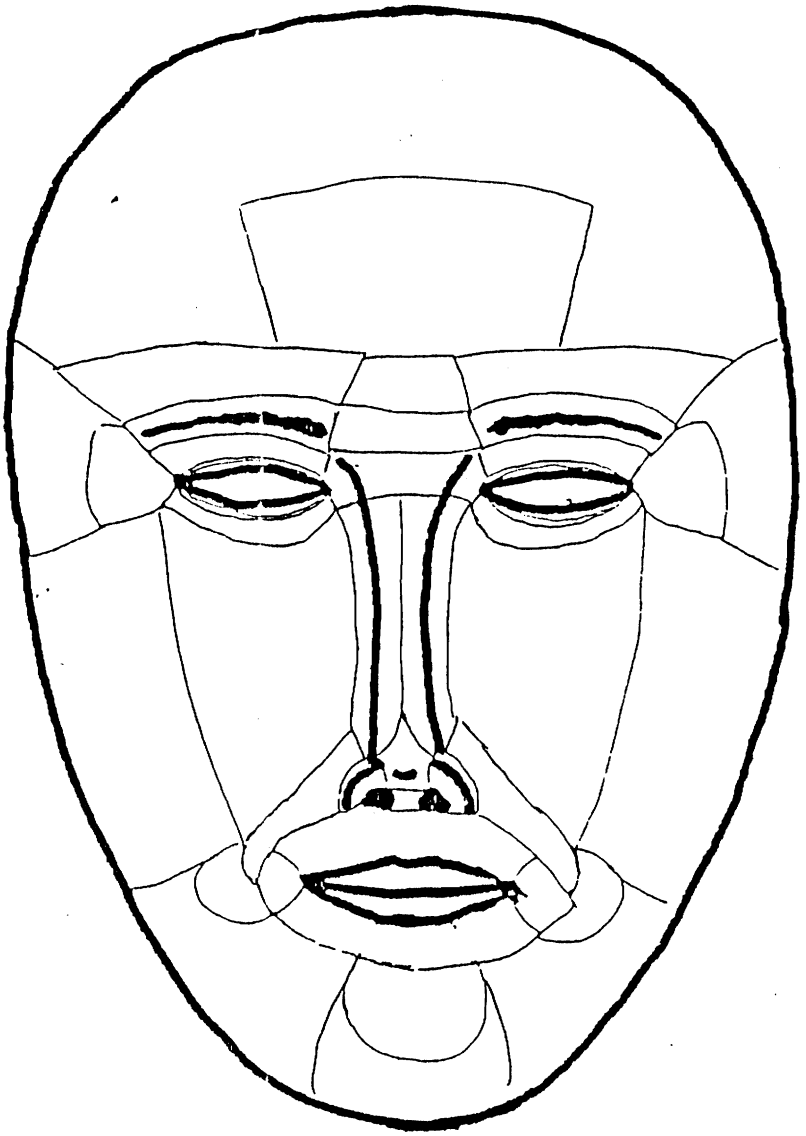


Figure 5a: Major Regions of the Face

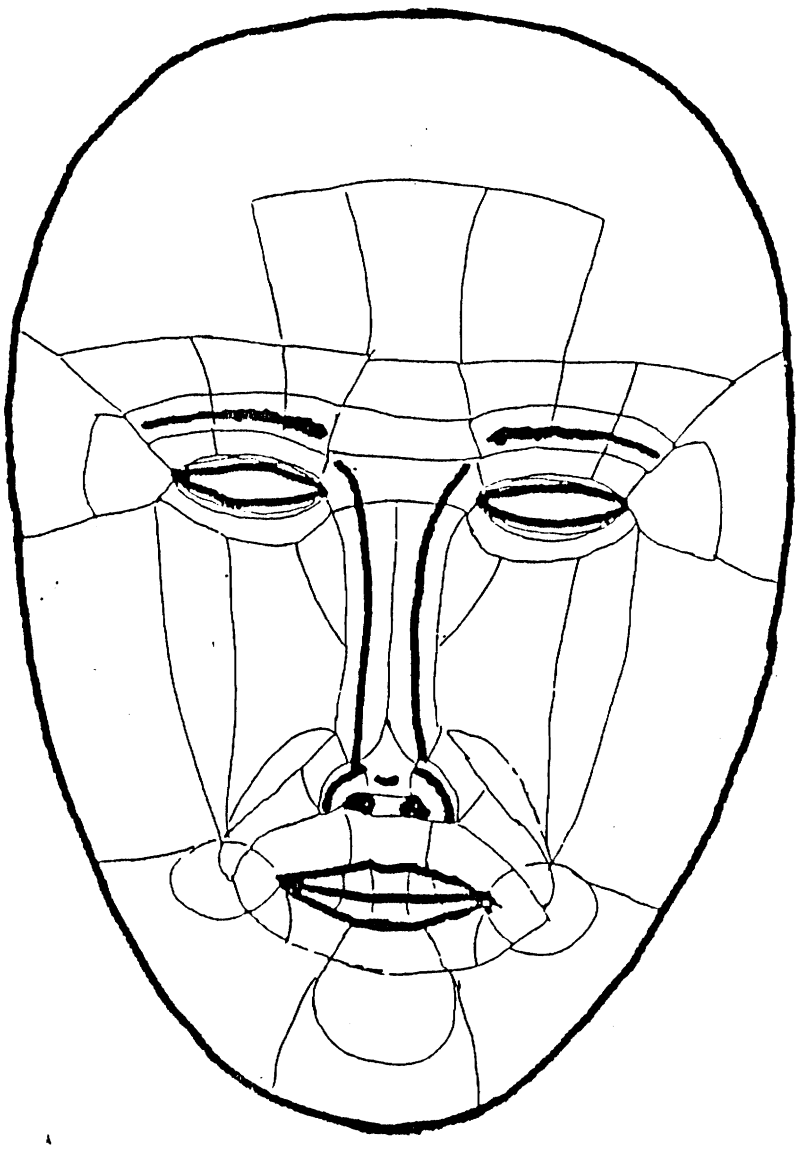
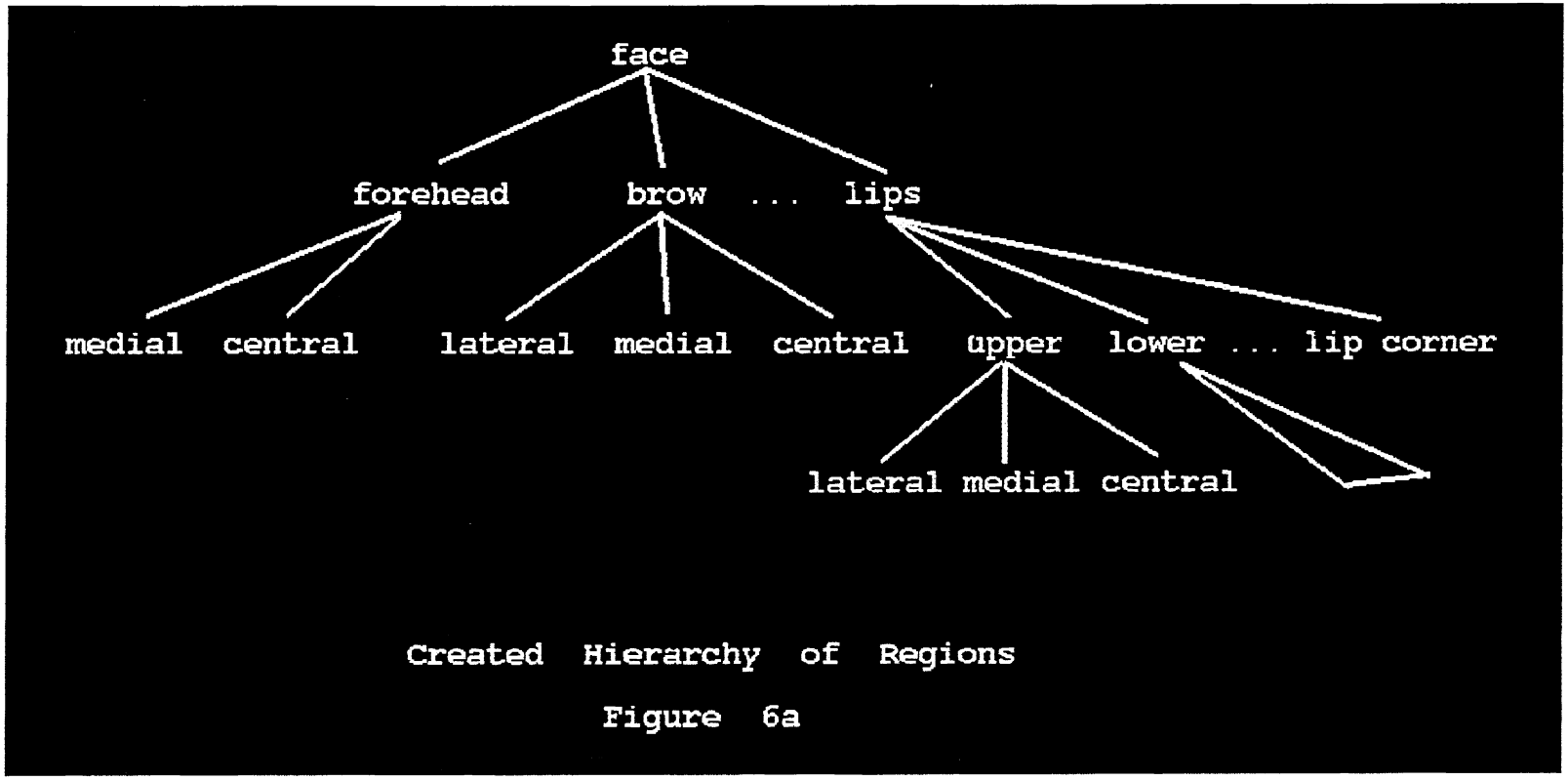
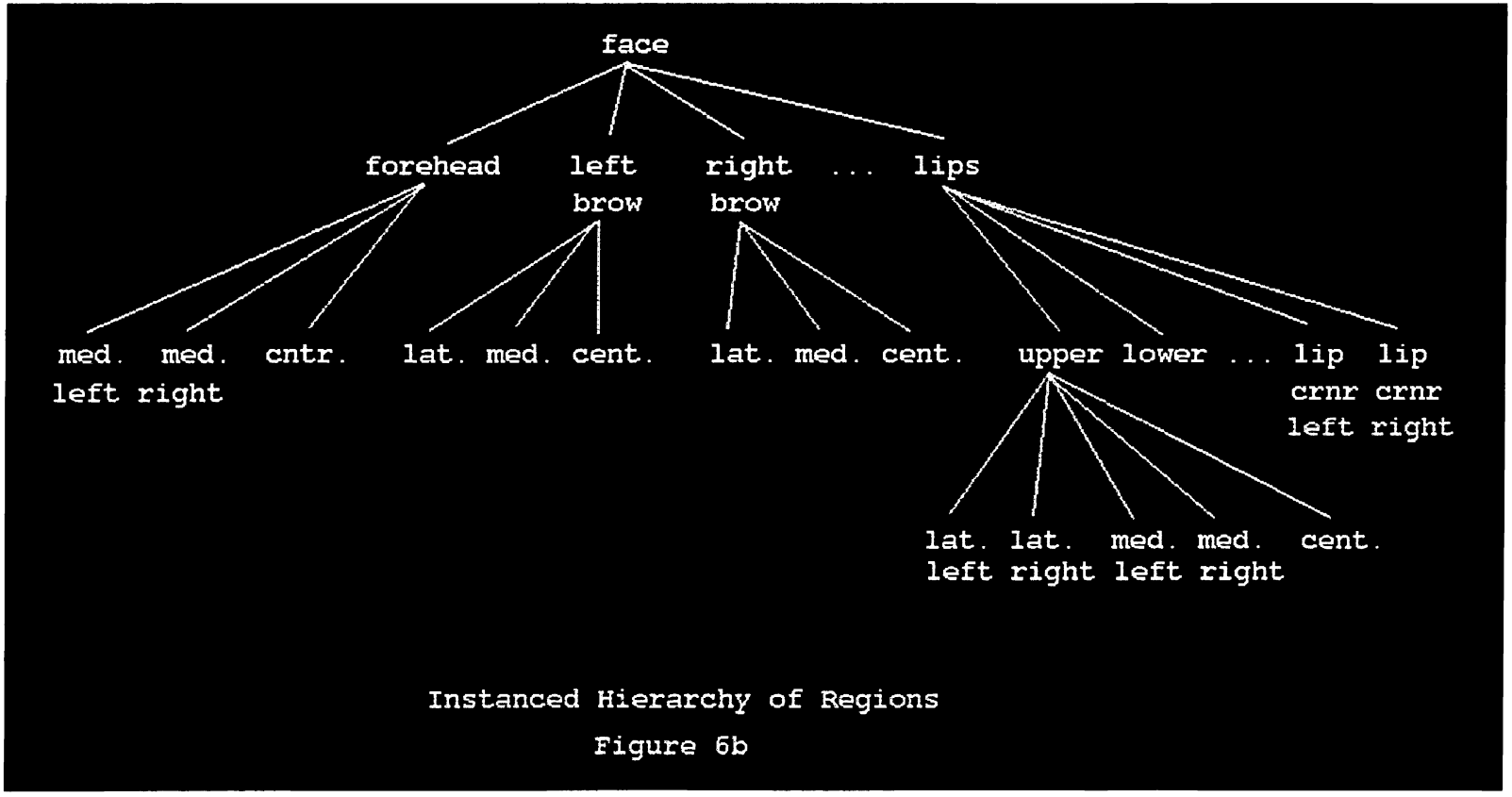


Figure 5b: Subregions of the Face



Created Hierarchy of Regions

Figure 6a



Instanced Hierarchy of Regions

Figure 6b

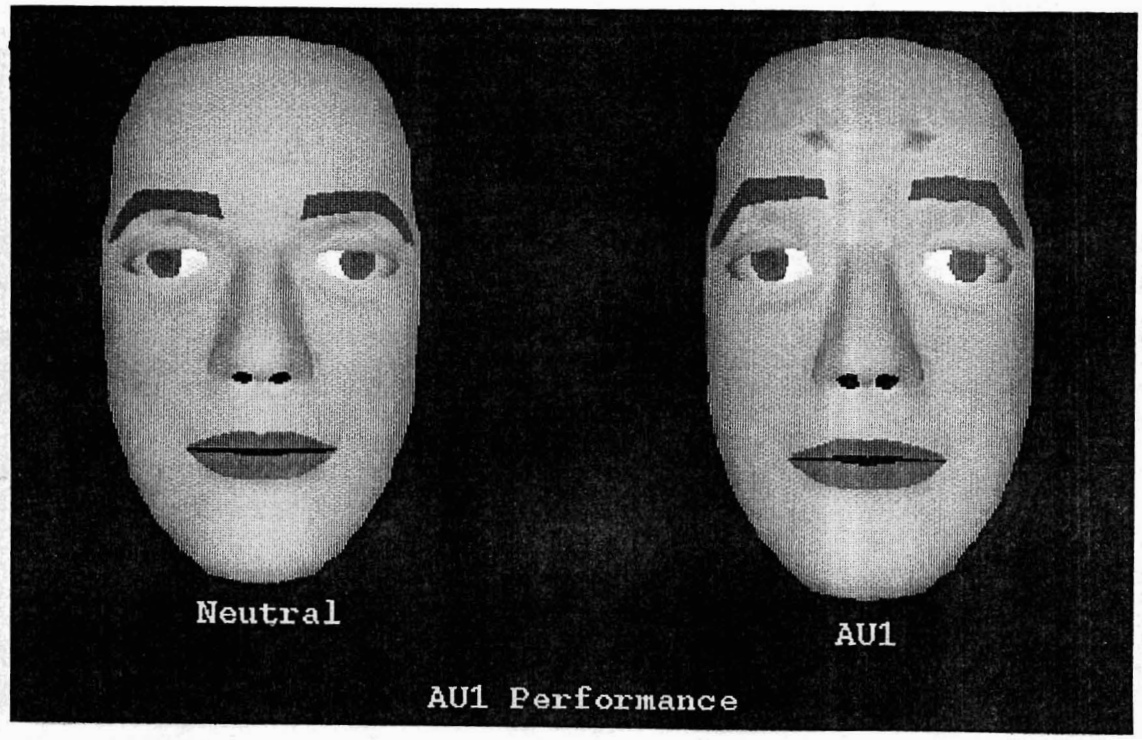


Figure 10: AU1 Performance

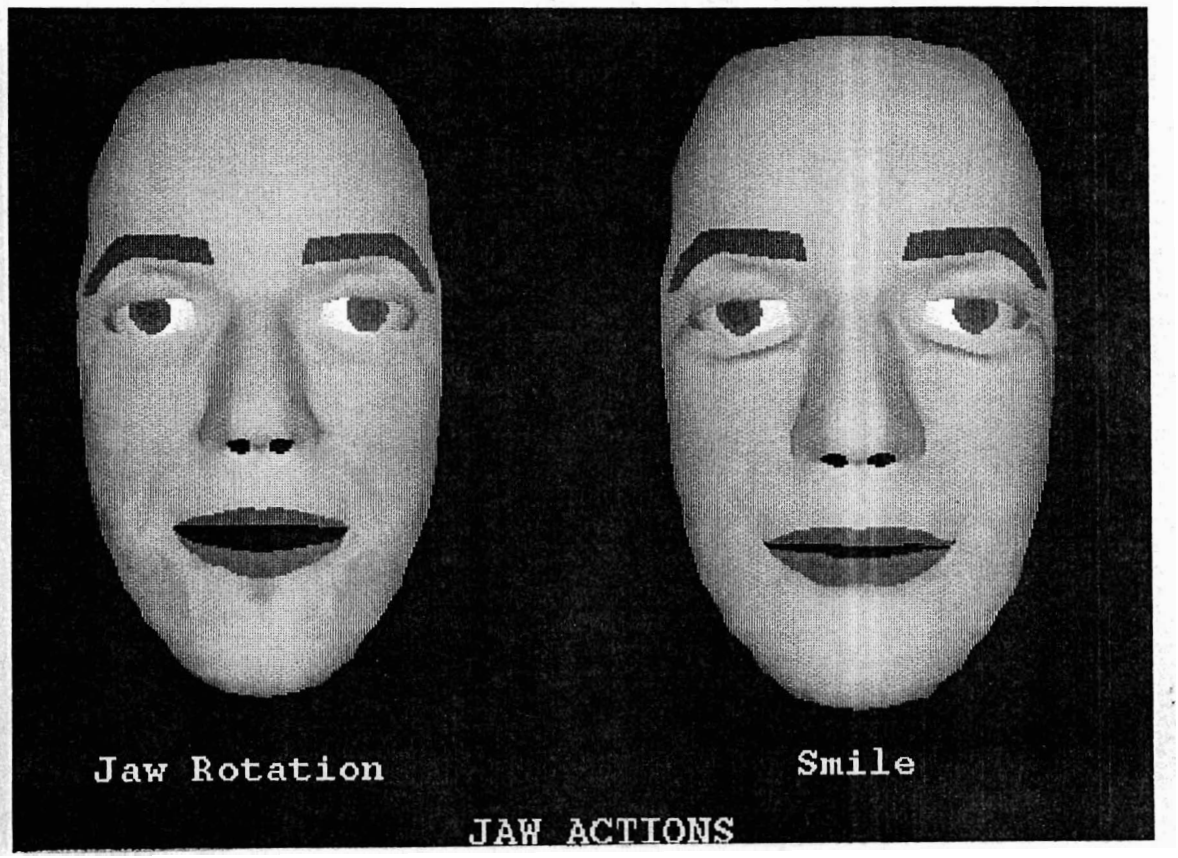


Figure 11: Jaw Actions

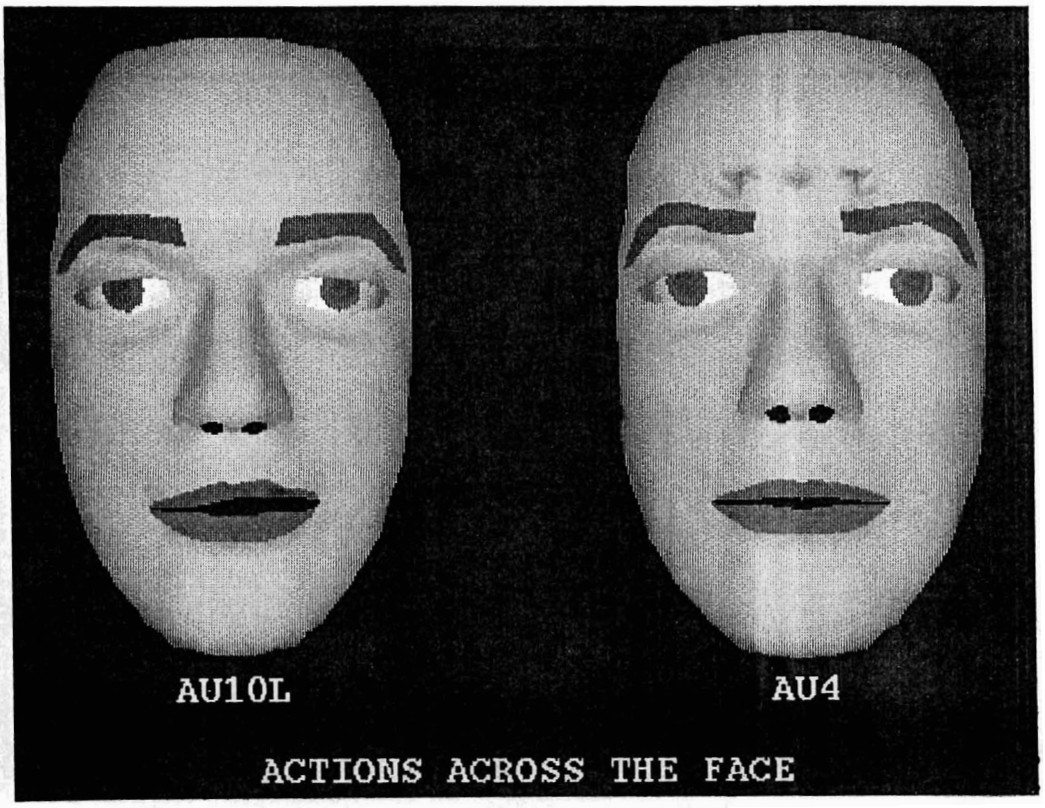


Figure 12: Actions Across the Face.

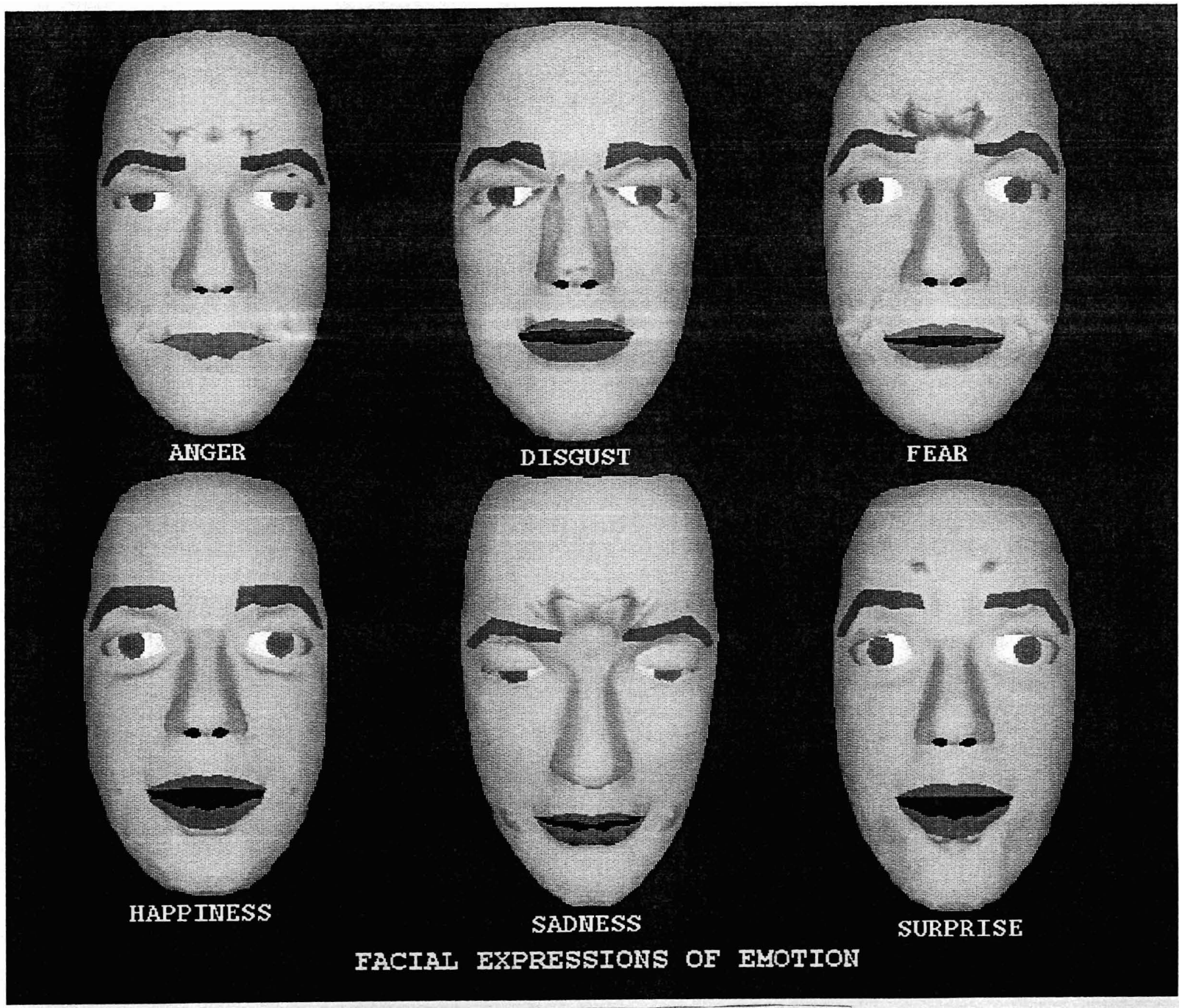


Figure 13: Facial Expressions of Emotion

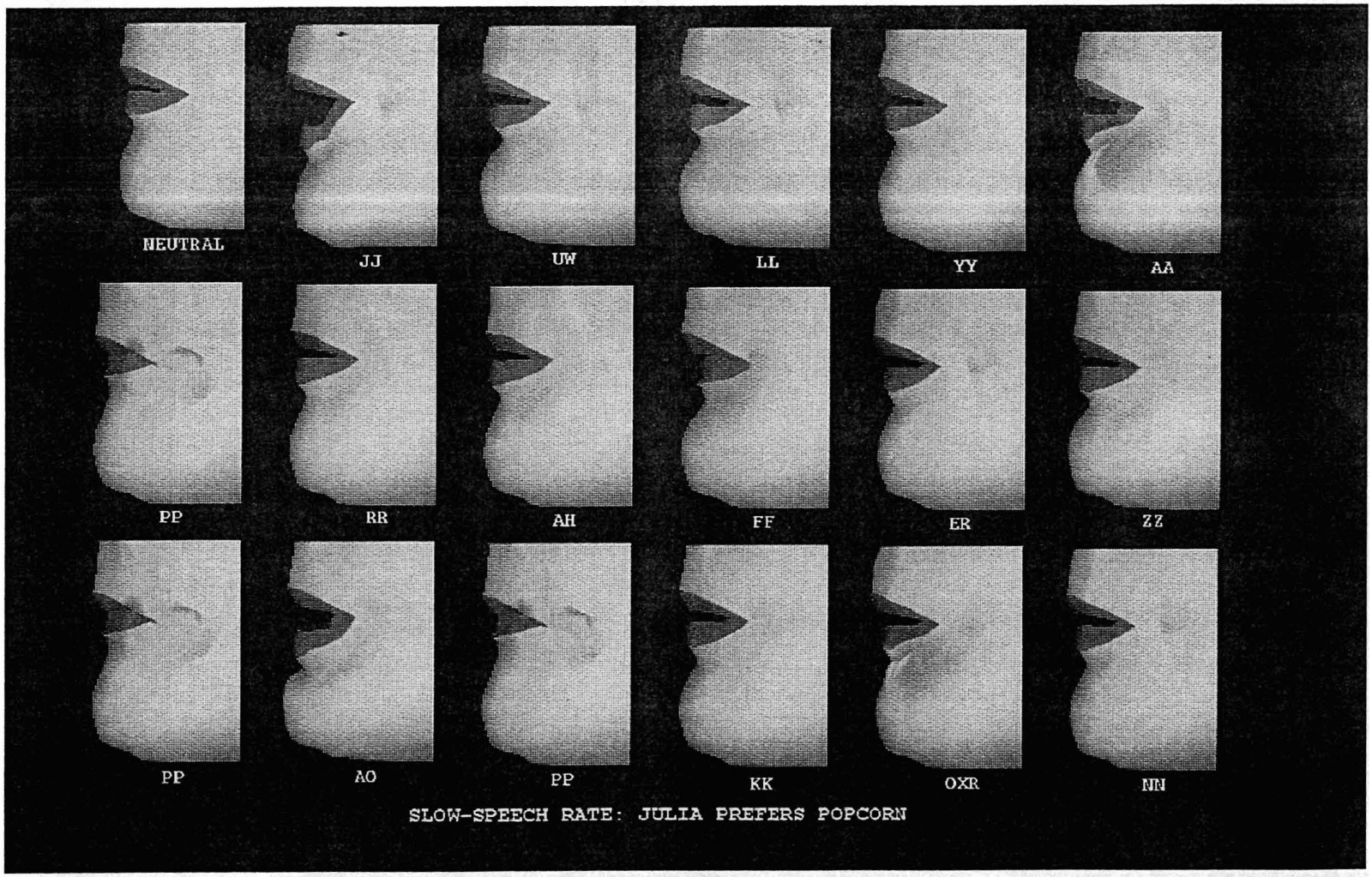


Figure 14: Slow-Speech Rate: "Julia prefers popcorn"