November 1990

Issues in Facial Animation

Catherine Pelachaud
University of Pennsylvania

Norman I. Badler
University of Pennsylvania, badler@seas.upenn.edu

Mark Steedman
University of Pennsylvania

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-90-88.

Issues in Facial Animation

Abstract

Our goal is to build a system of 3-D animation of facial expressions of emotion correlated with the intonation of the voice. Up till now, the existing systems did not take into account the link between these two features. 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. We will look at the rules that control these relations (*intonation/emotions and facial expressions/emotions*) as well as the coordination of these various modes of expressions. Given an utterance, we consider how the message (what is *new/old* information in the given context) transmitted through the choice of accents and their placement, are conveyed through the face. The facial model integrates the action of each muscle or group of muscles as well as the propagation of the muscles' movement. It is also adapted to the FACS notation (Facial Action Coding System) created by P. Ekman and W. Friesen to describe facial expressions. Our first step will be to enumerate and to differentiate facial movements linked to emotions from the ones linked to conversation. Then, we will examine what the rules are that drive them and how their different actions interact.

Keywords

facial animation, emotion, intonation, coarticulation, conversational signals

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-90-88.

Issues in Facial Animation

MS-CIS-90-88
GRAPHICS LAB 36

Catherine Pelachaud
Norman I. Badler
Mark Steedman

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

1990

Issues In Facial Animation

MS-CIS-90-88
GRAPHICS LAB 36

Catherine Pelachaud
Norman I. Badler
Mark Steedman

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

November 1990

Issues in Facial Animation

Catherine Pelachaud, Norman I. Badler, Mark Steedman
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104

Abstract

Our goal is to build a system of 3D animation of facial expressions of emotion correlated with the intonation of the voice. Up till now, the existing systems did not take into account the link between these two features. 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. We will look at the rules that control these relations (*intonation/emotions* and *facial expressions/emotions*) as well as the coordination of these various modes of expressions. Given an utterance, we consider how the messages (what is *new/old* information in the given context) transmitted through the choice of accents and their placement, are conveyed through the face. The facial model integrates the action of each muscle or group of muscles as well as the propagation of the muscles' movement. It is also adapted to the FACS notation (Facial Action Coding System) created by P. Ekman and W. Friesen to describe facial expressions. Our first step will be to enumerate and to differentiate facial movements linked to emotions from the ones linked to conversation. Then, we will examine what the rules are that drive them and how their different actions interact.

Key words: facial animation, emotion, intonation, coarticulation, conversational signals

1 Introduction

Emotions are part of our daily life. They are one of the most important human motivations. They are mainly expressed with the face and the voice.

The face is an important communication channel. While talking, a person is rarely still. The face changes expressions constantly, following the flow of speech. Faces have their own language where each expression is not only related to emotions, but is also linked to the intonation and the content of speech. Thus, in order to ameliorate facial animation systems, understanding such a language and its interaction with intonation is one of the more important steps.

There already exist some facial animation systems which incorporate speech and facial expression. F. Parke (**[PAR75]**, **[PAR90]**) was the first one to propose a facial model whose animation was based on parameters which effect not only the structure of the model but also its expressions (opening the mouth, raising the eyebrows). Various animations such as "Tony de Peltrie" (**[BER85]**) and "Sextone for President" (**[KLE88]**) were made from a library of digitized expressions. Other systems used parameterized models. N. Magnenat-Thalmann and D. Thalmann, for their movie "Rendez-vous à Montreal" (**[MAG87]**) differentiated two levels of facial expressions (one for the shape of the mouth for phonemes and the second for emotions), and they control their model through parameters. M. Nahas, H. Huitric and M. Saintourens defined a B-spline model (**[NAH88]**) where the animation is done using given control points. K. Waters (**[WAT87]**, **[WAT90]**) simulated muscle motion with a 2-D deformation function. Lip shapes for each visible vowel and consonant were defined in the same way. He recorded real actors and manually matched lip positions and phonemes. D. Hill, A. Pearce, and B. Wyvill, using F. Parke's parametric model (**[HIL88]**, **[WYV90]**) presented an automatic process for synchronizing speech and lip movements. Their approach is to synthesize both the speech sounds and the corresponding facial expressions by a set of rules where every element (of the sound and of the face) can be modified interactively through the use of parameters. However, no system up to now has considered the link between intonation and emotion to drive their system.

Our work will resolve the difficulty of manipulating the action of each muscle by offering to the user a higher level of animation by lip synchronization and automatic computation of the facial expressions related to the patterns of the voice. We will differentiate facial expressions linked to emotion from nonexpressive ones. We will elaborate a repertory of such movements.

After defining what *emotion* is for present purposes, we introduce our facial model. We also present the intonational system we are using and the characteristics of the voice for each of the emotions we are looking at.

Finally, we will characterize one by one the various types of facial expressions. Specifically, we will look at how we solve lip synchronization and coarticulation problems.

2 Emotion

Emotion arises as a change in neural stimulation due to internal/external events. Each emotion modifies in a particular way the physiology of a being. The variations of physical organs affect the vocal track while the variations of muscle actions affect the facial expressions.

Emotion is viewed, rather than as a particular state of an organism, as a process (**[SCH86]**, **[EKM89]**), with various components such as physiological responses (visceral and muscular states), autonomic nervous system and brain responses, verbal responses (vocalizations), memories, feelings and facial expressions.

For the emotions we are studying (anger, disgust, happiness, fear, sadness and surprise), there are three main areas in the face where changes occur: the upper part of the face with the brows and forehead, the eyes, and the lower part of the face with the mouth (**[EKM75]**). Each emotion is characterized by specific changes: fear is recognized by the raised and drawn together eyebrows and lips stretched tense; sadness is characterized by the inner side of the brows drawn up, the upper eyelid inner corner raised and the corners of the lips down (Figure 8).

3 Other Facial Expressions and their Rules

All facial expressions do not necessarily correspond to emotion. The raising eyebrows can punctuate a discourse and not be a signal of surprise. P. Ekman ([EKM89]) characterizes facial expressions into the following groups:

emblems correspond to movements whose meaning is very well-known and culturally dependent. They are produced to replace common verbal expressions. For example, instead of saying 'sure' or 'I agree' one can nod.

emotional emblems (also called referential expressions or mock expressions) are made to convey signals about emotions. A person uses them to mention an emotion: he does not feel the emotion at the time of the facial action. He only refers to them. It is quite common, when talking about a disgusting thing, to wrinkle one's nose. Such movements are part of the emotional ones (wrinkling the nose is part of the facial expression of disgust).

conversational signals (also called illustrators) are made to punctuate a speech, to emphasize it. Raising the eyebrows often accompanies an accented vowel.

punctuators are movements occurring at pauses.

regulators are movements that help the interaction between speaker/listener. They control the speaking turn in a conversation.

manipulators correspond to the biological needs of the face, like blinking the eyes to keep them wet, and wetting the lips.

affect displays are the facial expressions of emotion.

We have to include all these movements to obtain a more complete facial animation. A face can make many more movements such as grimacing, contorting, lip-bitting, twitching, and so on, but we are not considering them. They are not related to emotion or speech (or at least not directly). Also, the consideration of emblems and emotional emblems is out of scope of this study since they imply the voluntary participation of the speaker.

3.1 Synchronism

An important property linking intonation and facial expression (in fact, it is extended to body movement) is the existence of synchrony between them ([CON67], [CON71]). Synchrony implies that changes occurring in speech and in body movements should appear at the same time.

Synchrony occurs at all levels of speech. That is, it occurs at the level of phoneme, syllable (these two are defined by how their patterns are articulated), word, phrase or long utterance. Some body and facial motions are isomorphic to these groups. Some of them are more adapted to the phoneme level (like an eye blink), some others at the word level (like a frown) or even at the phrase level (like an hand gesture).

The main point is that there is no part of speech or body motion that is not grouped together in some sort of cluster. This is the basic rule we are using to compute animation in relation to speech.

4 Facial Model

4.1 Facial Action Coding System

Facial Action Coding System or **FACS** is a notational system developed by P. Ekman and W. Friesen ([**EKM78**]). It describes all visible facial movements that are either emotional signals or conversational signals. **FACS** is derived from an analysis of the anatomical basis of facial movements. Because every facial movement is the result of muscular action, a system could be obtained based on how each muscle of the face acts to change visible appearance.

One of the constraints of **FACS** is that it deals only with movements and what is visible on the face (no other perturbations, like blushing or tears, are considered). They also introduce what they call an Action Unit (**AU**). It is an action produced by one or more muscles.

4.2 Structural Model

Our model of the face was developed by Steve Platt (Figure 7); it is a hierarchically structured, regionally defined object ([**PLA85**]). The face is decomposed into regions and subregions. A particular region corresponds to one muscle or group of muscles. Each of them is simulated by specifying the precise location of their attachment to the surface structure. These regions can, under the action of a muscle, either contract or be affected by a propagated movement of an adjacent region. A region can contain 3 types of information:

- physical information (what is displayed on the screen): a set of 3D points.
- functional information (where it is now): how an **AU** will modify the region.

- connective information: which regions its movements should be propagated to bring secondary movement.

We use **FACS** here to encode any basic action. Concurrent actions can occur. In such a case the final position of a region is the summation of movements (or propagation) of all applied **AUs**. The entire animation system is script oriented. This model integrates the elasticity of the muscle and the skin. It uses **FACS** and simulates the muscle propagation, taking into account secondary motions. We choose this model for its muscle structures, movement simulation and its hierarchical definition of the face.

5 Intonation

Intonation is defined as the melodic feature of an utterance and can be decomposed into three components linked to: the syntax of an utterance (such as interrogative, declarative), the attitudes of the speaker (what the speaker wants to explicitly show to the listener: for example, politeness, irony) and finally the emotions (involuntary aspects of the speaker's speech) ([**SCH84**]). In our current research, we are not considering the second feature. The third feature, also called *paralanguage* ([**CRY75**]), is differentiated mainly by the pitch (while frequency is a physical property of sound, pitch is a subjective one), loudness (the perceived intensity of a sound), pitch contour (the global envelope of the pitch), tempo (rate of speech) and pause. For example, anger is characterized by a high pitch level, wide pitch range and large pitch variations. Its intensity has a very high mean, a high range and also high fluctuations. Its articulation is precise and its speech rate fast. Sadness, however, is characterized by a low pitch level, a narrow pitch range and very small pitch variations. Its intensity is soft, its mean low, its range narrow and its fluctuations small. Its speech rate is slow with the highest number of pauses of the longest duration ([**CAH89**], [**LAD85**], [**WIL81**]).

To define the syntactic structure of intonation, we are using Janet Pierrehumbert's notation ([**HIR86**]). Under this definition, intonation consists of a linear sequence of accents. Utterances are decomposed into *intonational* and *intermediate* phrases. Both of them consist of *pitch accent(s)*, a *phrase accent*; intonational phrase are terminated by a *boundary tone*. Different intonational "tunes" composed of these elements are used to convey various discourse related distinctions of "focus". That is givenness or newness of informations, contrast and propositional attitude. Thus they serve to indicate the status of the current phrase related to the next one, for example, the continuation of the same topic or the introduction of a new one.

We can represent the decomposition of an utterance into intonational (or intermediate) phrases by brackets (see below). The appropriate use of intonational bracketing is determined by the context in which the utterance is produced and on the meaning of the utterance (i.e. what the speaker wants to focus on, what he considers as new information versus old). This bracketing is (partially) reflected in intonation.

Consider the sentence "Julia prefers popcorn" (the example is related to one discussed in ([**STE90**])).

The possible intonational bracketings reflect the distinction between an utterance which is about *Who prefers popcorn* or about *What Julia prefers*:

- (Julia)(prefers popcorn)
- (Julia prefers)(popcorn)

These bracketings can de imposed by intonationa tones.

For example, in the following context, we will have the following tune:

```
Question: Well, what about JUlia? What does SHE prefer?
Answer: (JUlia prefers) (pOpcorn).
Accent: (L+H*       LH%) ( H*  LL%)
```

(**H** and **L** denote high and low tones which combine in the various pitch accents and boundary tones. **L+H*** and **H*** are different kinds of pitch accent, and **LH%**, **LL%** and **L** below are boundaries.)

By contrast, in the following context, we will have a different bracketing, imposed by a different set of intonational tunes:

```
Question: Well, what about the pOpcorn? Who prefers IT?
Answer: (JUlia) (prefers pOpcorn).
Accent: (H*  L) (         L+H* LH%)
```

These two examples show different intonational patterns. They emphasize different information (in the first context, the new message is 'popcorn' versus 'Julia' in the second one). The bracketing of the sentence, the placement of pauses and the type of accents vary also. Consequently the facial conversational signals and punctuators related to the first utterance will differ from those of the second one.

We assume that the input is an utterance already decomposed and written in its phonetic representation with its accents marked in its bracketed elements. For the moment, we are using recorded natural speech to guide our animation. After recording a sentence, we extract from its spectrogram the timing of each phoneme and pause. We would like later on to use analysis-and-resynthesis methods to automate the determination of paralanguage parameters and phoneme timing ([CHA89], [HAM98]) driven by a representation like the above.

6 Steps for Computing Facial Expressions

First, we compute the list of **AUs** for the given emotion. We add to this list the **AUs** needed for the mouth shape synchronized with each phoneme. Finally, using a set of rules, we compute the conversational signals, the punctuators, head and eye movements, and eyeblinks.

Our first step for the animation is lip synchronization. Speechreading techniques offer the possibility to define a lip shape for each cluster of phonemes.

6.1 Speechreading

J. Jeffers and M. Barley ([**JEF71**]) define speechreading as "the gross process of looking at, perceiving, and interpreting spoken symbols". This method is designed for hearing-impaired. These people learn to read speech from lip movements and facial expressions. Unfortunately, there exists a lot of homophonous words; that is words that look alike on the face, even if they differ in spelling and meaning. These words cannot be differentiated by their lip, jaw or tongue movements. For example, 'b', 'p', and 'm' involve the same facial movements. Moreover, most of the speech sounds are highly, if not completely invisible (they might involve only an obscure tongue movement, for example).

In our case, we are only interested in visible movements. Vowels and consonants 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 group, the least deformable one, while 's', 'n' are very context dependent). We should notice that such clustering depends on the speech rate and visual conditions. These are defined by the visual accuracy the listener has of the speaker (such as light on the speaker, and physical distance between speaker and hearer). A person who articulates each word carefully shows more speech movements, of course. The faster a person speaks, the less effort he makes and fewer movements will be produced. With a fast speech rate or under poor visual conditions, the number of clusters diminishes. In addition, the lip shapes of most groups lose their well pronounced characteristics (lips drawn backward for the 'i', lips puckered for the 'o') and tend to have a more neutral position (moderate opening of the mouth).

Intonation of an utterance is the enunciation of a sequence of accented and non-accented phonemes. An accented vowel is differentiated acoustically from the remaining part of the utterance by its longer duration and increased loudness; visually, the jaw dropping motion is a characteristic of accented or emphasized segments.

This phonemic notation, however, does not tell us how to deal with the difficult problem of coarticulation. In the next section, we introduce a first attempt to solve this problem.

6.2 Coarticulation

Coarticulation means "articulatory movements associated with one phonetic segment overlap with the movements for surrounding segments." ([KENT77]) if one does not consider the problem of coarticulation, incorrect mouth positions can occur.

Speech has been decomposed into a sequence of discrete units such as syllables and phonemes. However, speech production does not follow such constructions. There is an overlap between units during their production, thus the boundaries among them are blurred.

A simple solution to the problem of coarticulation will be to look at the previous, the present, and the next phonemes to determine the mouth positions ([**WAT87**]). But in some cases this is not enough, since the correct position can depend on a phoneme up to five positions before or after the current one ([**KENT77**]).

Forward coarticulation is defined when "an articulatory adjustment for one phonetic segment is anticipated during an earlier segment in the phonetic string" ([**KENT77**]) while backward coarticulation is defined when "an articulatory adjustment for one segment appears to have been carried over to a later segment in the phonetic string" ([**KENT77**]). For example, forward coarticulation arises in a sequence of consonants (not belonging to the highly visible clusters such as 'f', 'v', ...) followed by a vowel, since the lips show the influence of the vowel on the first consonant of the sequence. In the sequence of phonemes 'istrstry' (example cited in ([**KENT77**])) the influence of the 'y' is shown on the first 's' (forward rule).

We have implemented these two coarticulation rules. Even though they do not consider every case of coarticulation problems, there does not yet exist a complete set of such rules. To solve particular problems which cannot be solved by these two rules, we consider a two-step algorithm. On the first pass, coarticulation rules are applied to all clusters which have been defined as context-dependent. The next pass is to consider relaxation and contraction time of a muscle and to look at the way two consecutive actions are performed. Therefore, the speech context is considered.

After the first computation, we check that each action (**AU**) has time to contract after the previous phoneme (or, respectively, to relax before the next one). 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 same manner, if the time between two consecutive phonemes is smaller than the relaxation time, the current phoneme will influence the next phoneme when relaxing.

Finally, we take into account the geometric relationship between successive actions. Indeed, the closure of the lips is more easily performed from a slightly parted position than from a puckered position. The intensity of an action is rescaled depending on its surroundings context.

At the end of these passes, we obtain a list of **AU**s for each phoneme.

7 Conversational Signals

A stressed segment is often accompanied not by a particular movement but by an accumulation of rapid movements (such as more pronounced mouth motion, blinks, or rapid head movements).

Conversational signals may occur on an accented item within a word, or, it may stretch out over a syntactic portion of the sentence (corresponding to an emphatic movement).

Most of the time these signals involve actions of the eyebrows. P. Ekman ([**EKM89**]) found that **AU1+2** (the eyebrows raised of surprise) and **AU4** (the frown of anger) are commonly used. Raised eyebrows can occur to signal a question, especially when it is not syntactically defined. Head and eye motions can illustrate a word. An accented word is often accompanied by a rapid head movement ([**HAD84**]). A blink can also occur on a stressed vowel ([**CON71**]).

Each emotion does not activate the same number of facial movements. An angry or happy person will have more facial motions than a sad person. Also, emotion intensity or strength affects the amount and type of facial movements ([**COL85**]). Thus we will select the occurrence of conversational signals depending on the emotion and intensity.

8 Punctuators

Punctuators can appear at a pause (due to hesitation) or to signal punctuation marks (such as comma or exclamation marks). The number of pauses affect the speech rate (a sad person has a slow speech rate due in part to a large number of long pauses, while a frightened person's speech shows very few pauses, and these of short duration) ([**CAH89**]). Thus the occurrence of punctuators and their type (i.e. their corresponding facial expressions) are emotion-dependent. A happy person has the tendency to punctuate his speech by smiling. Certain types of head movements occur during pauses. A boundary point (between intermediate phrases, for example) will be underlined by an ordinary movement and a final pause will coincide with stillness ([**HAD84**]). Eyeblinks can occur also during pauses ([**CON71**]).

9 Regulators

Regulators correspond to how people take turns speaking in a conversation, or any ritual meeting. We are still in the process of implementing this section. Much study has been given to speaking-turn taking. S. Duncan ([**DUN74**]) enumerates them:

- Speaker-Turn-Signal: is emitted when the speaker wants to give his turn of speaking to the auditor. It is composed of several clues in his intonation, paralanguage, body movements and syntax.
- Speaker-State-Signal: is displayed at the beginning of a speaking turn. It is composed, at the least, of the speaker turning his head away from the listener and the starting of a speaker gesticulation (arms and so on).
- Speaker-Within-Turn: is used when the speaker wants to keep his speaking turn, and assures himself that the listener is following. It occurs at the completion of a grammatical clause; the speaker turns

his head toward the listener.

- Speaker-Continuation-Signal: frequently follows a Speaker-Within-Turn. In such case, the speaker turns his head (and eyes) away from the listener.

10 Manipulators

Blinking is the only phenomena we are taking into account in this category. The eye blinks occur quite frequently. They serve not only to accentuate speech but also to address a physical need (to keep the eyes wet). There is at least one eye blink per utterance.

The internal structure of an eye blink, i.e., when it is closed and when open, is synchronized with the articulation ([**CON71**]). The eye in blinking might close over one syllable and start opening again over another word/syllable. Blink occurrence is also emotion dependent. During fear, tension, and anger, excitement and lying, the amount of blinking increases: it decreases during attention or concentrated thought ([**COL85**]).

We first compute all the blinks occurring as conversational signals or punctuators. Then, since eyeblinks should occur periodically we add any necessary ones. The period of occurrence is emotion-dependent. This time will be shorter for fear and longer for sadness.

11 Pupil dilation and constriction

The pupil constricts in bright light, while it dilates in weak light. Moreover, a person with light eye color will have the tendency to have larger pupils and will show larger pupil dilation. Pupil changes also occur during emotional experiences. Pupil dilation is followed by pupil constriction during happiness and anger and remains dilated during fear and sadness (**HES75**]).

12 Animation

We should note that every facial action (except those involved in the lip synchronization) will have three parameters: onset, apex, and offset. Apex corresponds to the time the action is occurring. Onset and offset define the manner of appearance and disappearance of the action;they are emotion dependent. For example, surprise has the shortest onset time while sadness has the longest offset ([**EKM84**]).

Having computed the list of **AU**s for each phoneme, the animation can then be performed. We apply the heuristic that quick abrupt changes for a particular portion of the face cannot occur in too short a time. The regions of the face are organized into three sets; one with high movement (like the lips, the brows), one with medium movement (forehead, cheeks), and one with low movement (outer part of the face) (Figure 1). We use rational cubic Bsplines with associated weights for each group of regions to take this fact into account.

13 Example

Let us consider the example introduced in a previous section: 'Julia prefers popcorn'. We record this utterance; we find its intonational pattern; we decompose it into a sequence of phonemes (we use Dectalk's notation); and we extract from its spectrogram the timing of each phoneme.

We consider only the computation of the lip shapes. For every phoneme, we find the group (as defined in [**JEF71**]) in which it belongs. To highly influencable phonemes (such as 'n', 't'), we apply the forward and backward coarticulation rules. In Figures 2 and 3, the graphs depict the intensity of **AU18** (lip pucker) and **AU25** (opening of the mouth) over time. The phoneme /LL/ in the word 'Julia' receives the same list of **AUs** with lower intensity as its preceding vowel (/UW/ belongs to a less influencable cluster than /YY/; therefore the backward rule is applied for /LL/). At this stage of the computation the red line on the graphs correspond to the intensity over time.

Our next step is to consider the environment of each phoneme and its relaxation and contraction times. The blue lines on the graphs in Figures 2 and 3 indicate the final values of intensity for the **AUs**. For the phoneme /YY/ in 'Julia', we can notice the apparition of **AU18** from the phonemes /UW/ and /LL/. The lip shapes for /LL/ do not have enough time to relax completely from their puckered position to their extended lip shapes: Some puckered effect remains, so we have applied a control over time. On the other hand, the pucker position of the item /AO/ from the syllable 'pop' is altered due to its surrounding lip closures for the two /PP/s, so we applied a control over space.

The red lines in the graphs in Figures 4 and 5 correspond to the position in space (horizontal (x-axis) for Figure 4 and vertical (y-axis) for Figure 5) over time of specific nodes (the right corner of the lip for Figure 4, and the center upper point of the lower lip for Figure 5) after the first computation. The blue lines show the final position of the nodes. There are less sharp transitions in space between adjacent points (Figure 6). The lips do not extend or open as much as it did over time. An overall smoother animation is obtained.

These constraints between adjacent **AUs** are defined by a constant and are easily changed as is relaxation/contraction simulation. Moreover, lip shapes associated with each phoneme are determined by rules and are also easily modified. This provides a tool for phoneticians to study coarticulation problems.

14 Summary

We have presented here a tool which enhances facial animation. Our method is based on finding the link between the spoken intonation, the transmitted information in the given context, and the facial movements. First, we presented our facial model. We also enumerated and differentiated facial movements due to emotion or due to conversation. We look more particularly on the coarticulation problem where we examine how the action of a muscle is affected by temporal and spatial context. Currently we are working on the rules that coordinate these various facial motions with the intonation. Indeed, while a substantial number of these relations have been studied and described, many more remains to be investigated. We offer a tool to analyze, manipulate and integrate these different channels of communication and to facilitat the further research of human communicative faculties via animation.

15 Acknowledgements

I would like to thank F. Azuola and M. Edwards to have developed the Rational B-spline program for the animation software. This research is partially supported by Lockheed Engineering and Management Services (NASA Johnson Space Center), NASA Ames Grant NAG-2-426, NASA Goddard through University of Iowa UICR, FMC Corporation, Martin-Marietta Denver Aerospace, Deere and Company, Siemens Research, NSF CISE Grant CDA88-22719, and ARO Grant DAAL03-89-C-0031 including participation by the U.S. Army Human Engineering Laboratory and the U.S. Army Natick Laboratory.

16 Bibliography

[ARG76] M. Argyle, M. Cook, Gaze and Mutual Gaze, *Cambridge University Press*, 1976.

[BER85] P. Bergeron, P. Lachapelle, "Controlling Facial Expressions and Body Movements in the Computer Generated Animated Short 'Tony de Peltrie"', *ACM SIGGRAPH'85 Tutorial Notes, Advanced Computer Animation Course*, 1985.

[BOL86] D. Bolinger, Intonation and its part, *Stanford University Press*, 1986.

[BOU73] G.H. Bourne, The Structure and Function of Muscle, vol. III, Physiology and Biochemistry, *Academic Press*, 1973, Second Edition.

[BUL85] P. Bull, G. Connelly, "Body Movement and Emphasis in Speech", *Journal of Nonverbal Behavior*, vol. 9, n. 3, 1985.

[CAH89] J. Cahn, "Generating expression in synthesized speech", *Masters Thesis, M.I.T.*, 1989.

[CHA89] F. Charpentier, E. Moulines, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones", *Proceedings EUROSPEECH' 89*, vol. 2, 1989.

[COL85] G. Collier, Emotional expression, *Lawrence Erlbaum Associates*, 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.

[CRY75] D. Crystal, "Paralinguistics", in *The body as a medium of expression*, J. Benthall, T. Polhemus ed., 1975.

[DIT74] A.T. Dittman, "The body movement-speech rhythm relationship as a cue to speech encoding", in *Nonverbal Communication*, Weitz ed., 1974.

[DUN74] S. Duncan, "Some signals and rules for taking speaking turns in conversations", in *Nonverbal Communication*, ed. Weitz, 1974.

[EKM75] P. Ekman, W. Friesen, Unmasking the Face: A guide to recognizing emotions from facial clues, *Prentice-Hall*, 1975.

[EKM78] P. Ekman, W. Friesen, Facial Action Coding System, *Consulting Psychologists Press*, 1978.

[EKM79] P. Ekman, "About brows: emotional and conversational signals", in *Human ethology*, M. von Cranach, K. Foppa, W. Lepenies, D. Ploog ed., 1979.

[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 edition, *Academic Press*, 1990.

[FRI75] J. Friedman, "Computer exploration of fast speech rules", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, v. 23, n.1, 1975.

[FRO64] V. Fromkin, "Lip positions in American english vowels", *Language and Speech*, v. 7, n. 3, 1964.

[GRA69] E.C. Grant, "Human facial expression", *Man*, n. 4, 1969.

[HAM89] C. Hamon et al., "A diphone synthesis system based on time-domain prosodic modifications of speech", *ICASSP' 89*, 1989.

[HES75] E.H. Hess, "The role of the pupil size in communication", *Scientific American*, Nov. 1975.

[HIL88] D.R. Hill, A. Pearce, B. Wyvill, "Animating speech: an automated approach using speech synthesised by rules", *The Visual Computer*, v. 3, 1988.

[HIR86] J. Hirschberg, J. Pierrehumbert "The intonational structuring of discourse", *24th Annual Meeting of the Association for Computational Linguistics*, 1986.

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

[KEN67] A. Kendon, "Some Functions of Gaze-Direction in Social Interaction", *Acta Psychologica*, vol. 26, 1967.

[KEN72] A. Kendon, "Some Relationships Between Body Motion and Speech", *Studies in Dyadic Communication*, ed. A.W. Siegman, B. Pope, 1972.

[KENT77] R.D. Kent, F.D. Minifie, "Coarticulation in recent speech production models", *Journal of Phonetics*, n. 5, 1977.

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

[LAD85] D.R. Ladd, K. Silverman, F. Tolkmitt, G. Bergmann, K. Scherer, "Evidence for the independent function of intonation, contour type, voice quality and F0 range in signaling speaker affect.", *Journal of Acoustical Society of America*, n. 78, November 1985.

[LEW87] J.P. Lewis, F.I. Parke, "Automated Lip-Synch and Speech Synthesis for Character Animation", *CHI + GI*, 1987.

[MAG87] N. Magnenat-Thalmann, D. Thalmann, "The direction of synthetic actors in the film *Rendez-vous à Montrèal*", *IEEE Computer Graphics and Applications*, Dec. 1987.

[NAH88] M. Nahas, H. Huitric, M. Saintourens, "Animation of B-spline figure", *The Visual Computer*, v. 3, 1988.

[PAR82] F. Parkes, "Parameterized Models for Facial Animation", *Computer Graphics and Applications*, Nov. 1982.

[PAR90] F.I. Parke, "Parameterized facial animation - Revisited", *ACM SIGGRAPH'90 Course Notes, State in the Art in Facial Animation*, 1990.

[PEA86] A. Pearce, B. Wyvill, D.R. Hill, "Speech and expression: a computer solution to face animation", *Graphics Interface'86, Vision Interface'86*, 1986.

[PLA81] S.M. Platt, N.O. Badler, "Animating facial expressions", *Computer Graphics*, v. 15, n. 3, Aug. 1981.

[PLA85] S.M. Platt, A Structural Model of the Human Face, Ph. D. thesis, Computer and Information Science Department, University of Pennsylvania, 1985.

[SCH84] K. Scherer, D.R. Ladd, K. Silverman, "Vocal cues to speaker affect: testing two models", *Journal of Acoustical Society of America*, number 76, November 1984.

[SCH88] K. Scherer, Facets of emotion: recent research, *Lawrence Erlbaum Associates Publishers*, 1988.

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

[WAL83] M.B. Walker, C. Trimboli, "The Expressive Function of the Eye Flash", *Journal of Nonverbal Behavior*, vol. 8, n. 1, 1983.

[WAT87] K. Waters, "A muscle model for animating three-dimensional facial expression", *Computer Graphics*, v. 21, n. 4, July 1987.

[WAT90] K. Waters, "Modeling 3D facial expressions", *ACM SIGGRAPH'89 Course Notes, State in the Art in Facial Animation*, 1990.

[WIL81] C. Williams, K. Stevens, "Vocal correlates of emotional states", *Speech evaluation in psychiatry*, ed. Darby, 1981.

[WYV90] B. Wyvill, D.R. Hill, "Expression control using synthetic speech", *ACM SIGGRAPH'90 Course Notes, State in the Art in Facial Animation*, 1990.

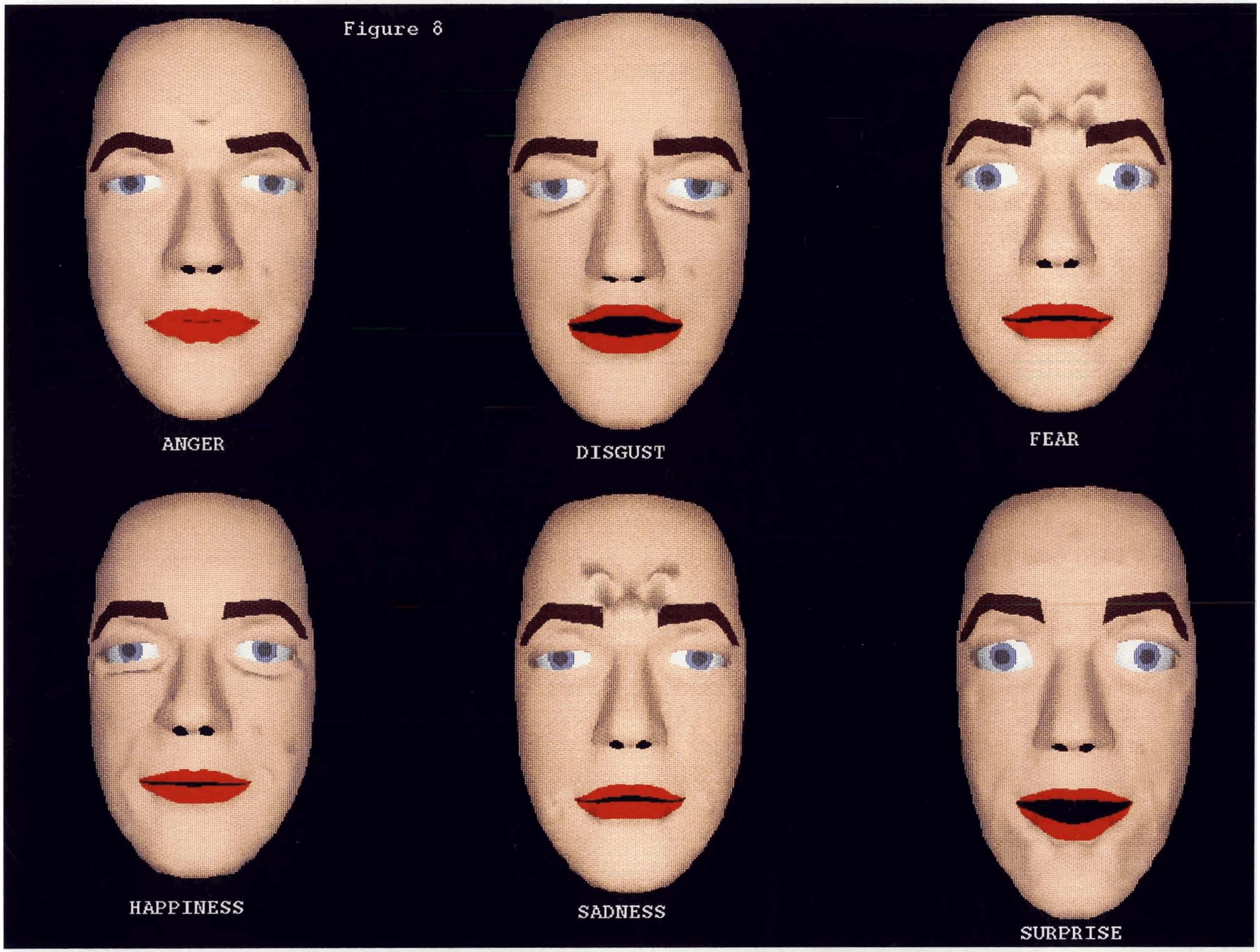


Figure 8
ANGER
DISGUST
FEAR
HAPPINESS
SADNESS
SURPRISE

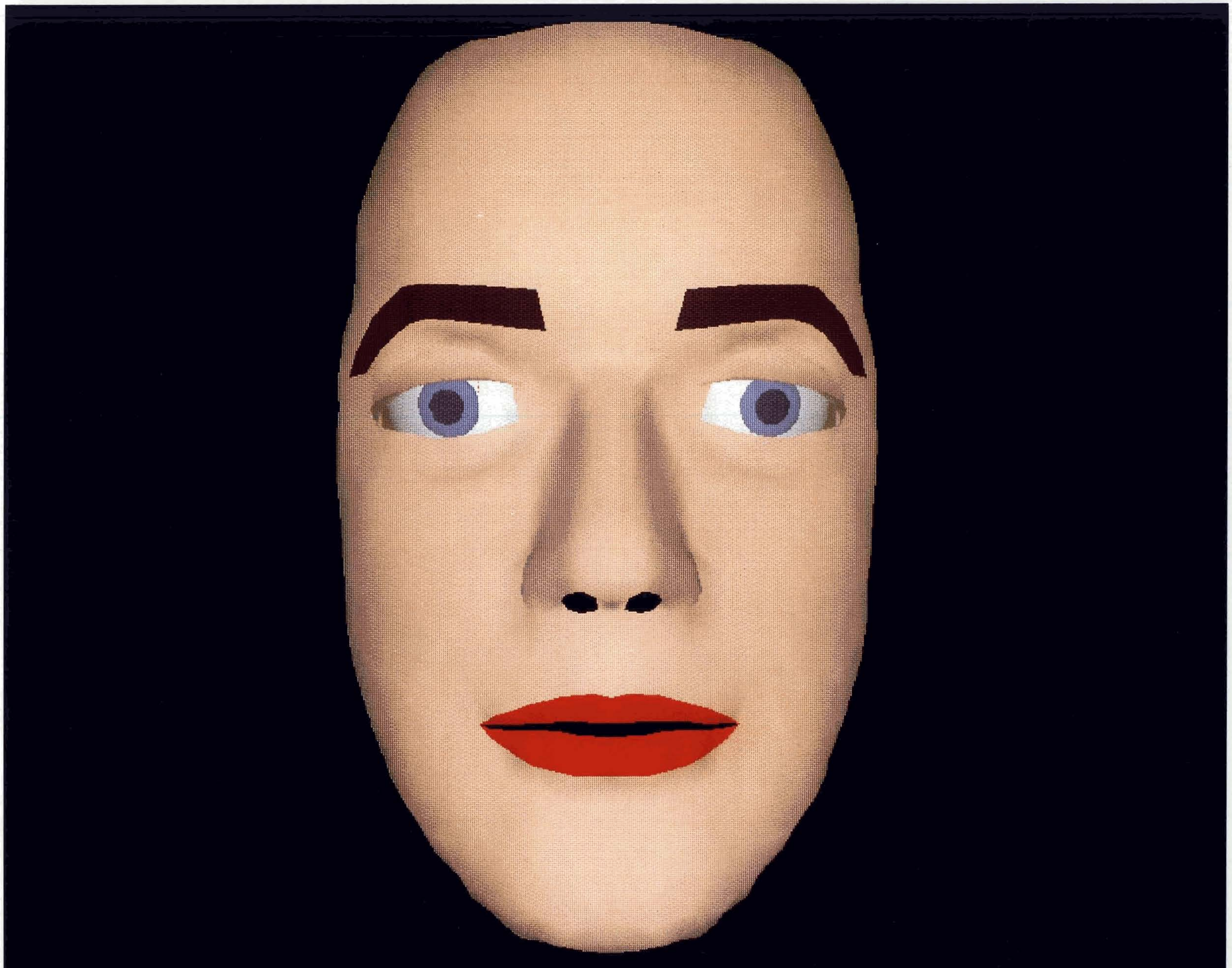


Figure 7

Neutral Face

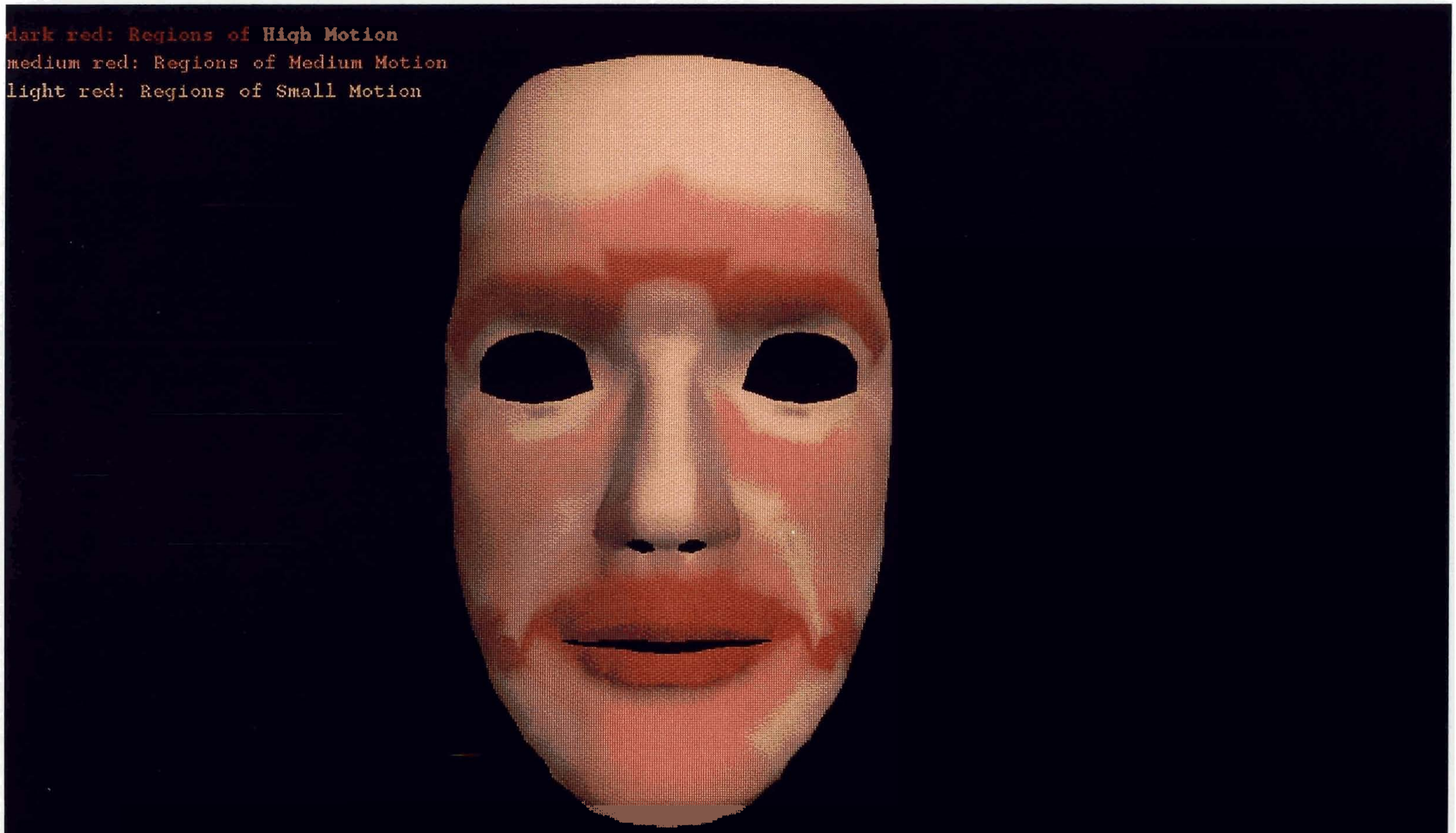
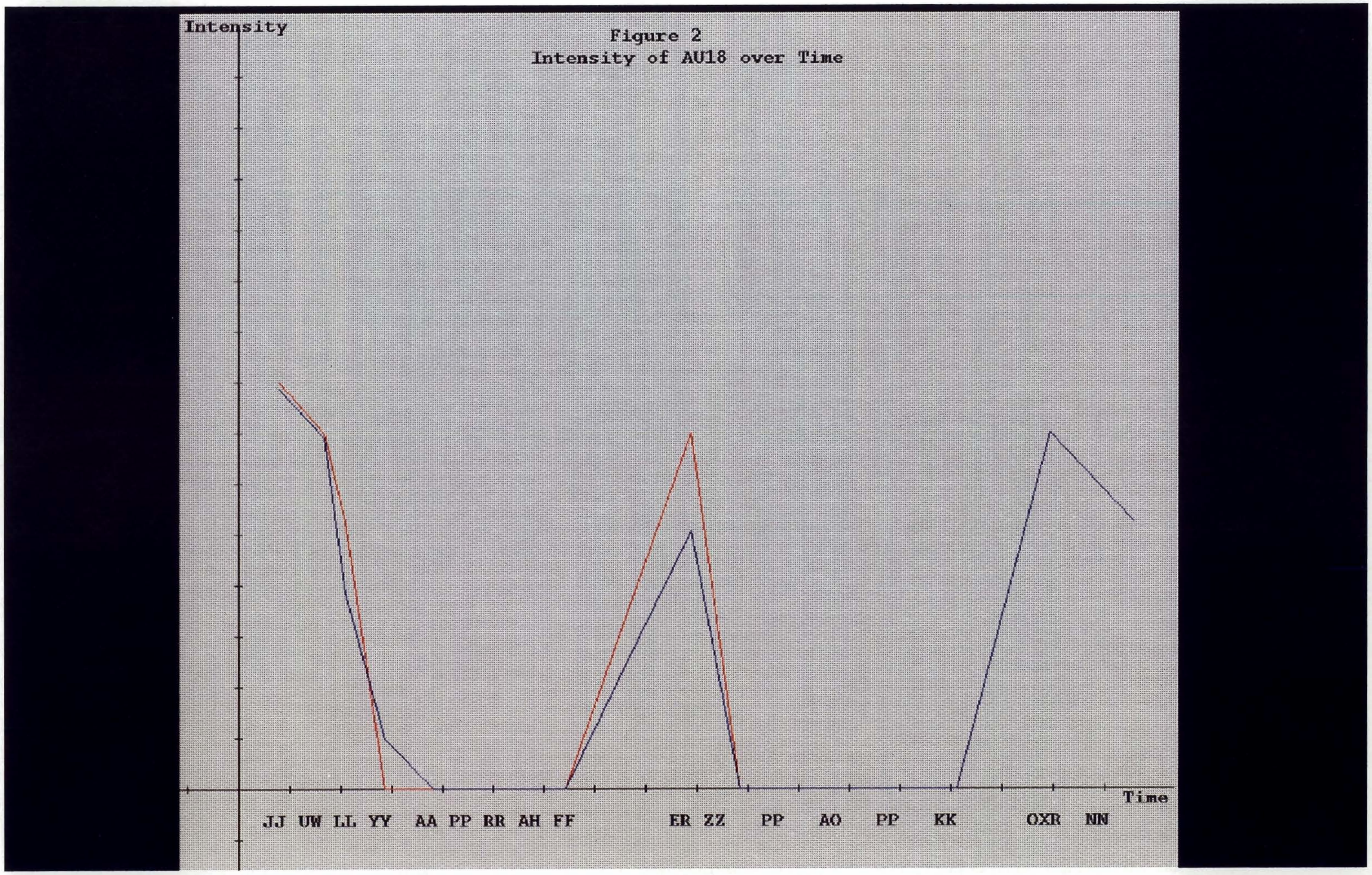


Figure 1
Partition of the Face for the Computation of the Weights


Intensity
Figure 2
Intensity of AU18 over Time
Time
JJ UW LL YY AA PP RR AH FF ER ZZ PP AO PP KK OXR NN

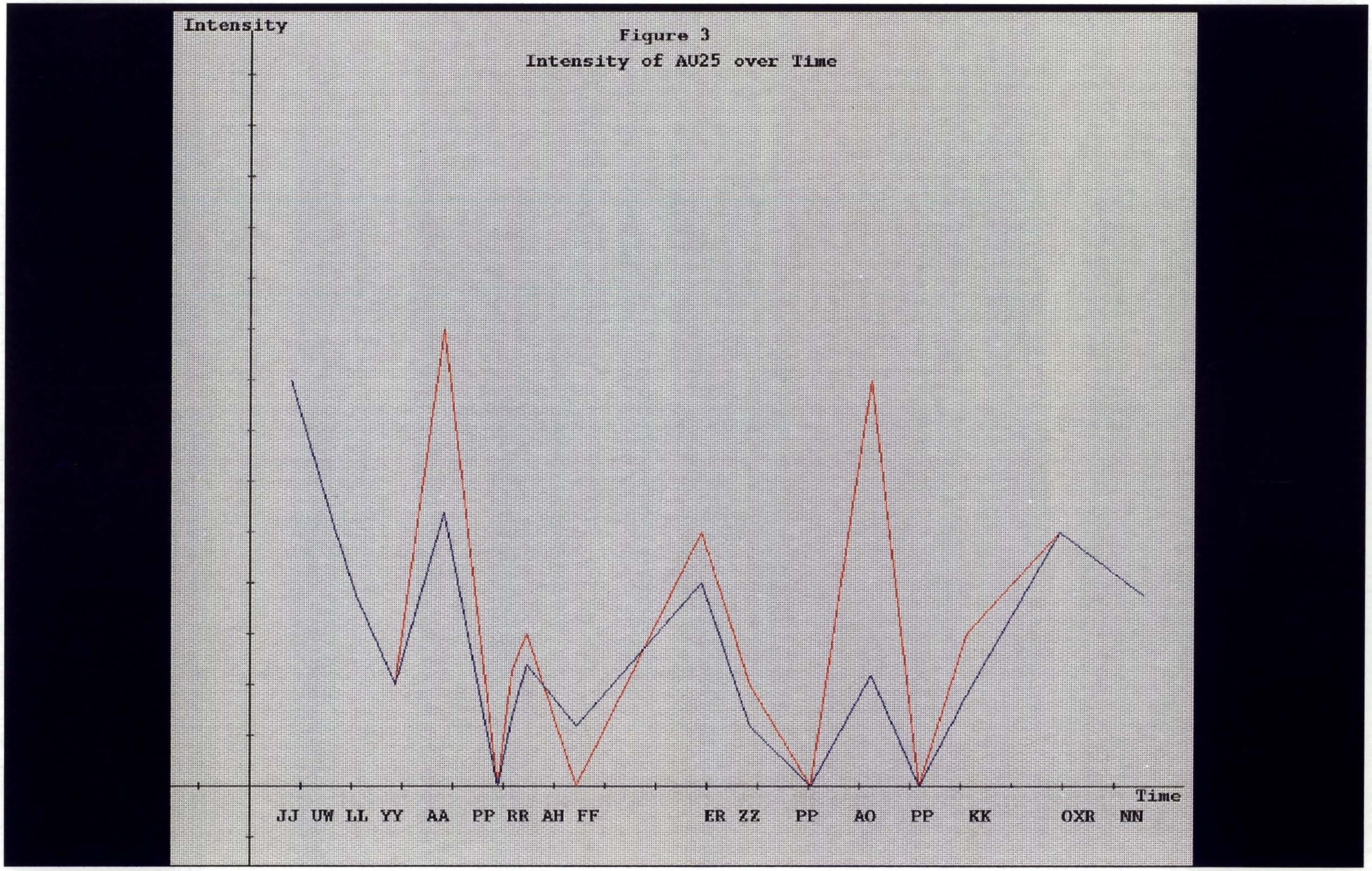


Figure 3
Intensity of AU25 over Time
Intensity
Time
JJ UW LL YY AA PP RR AH FF ER ZZ PP AO PP KK OXR NN

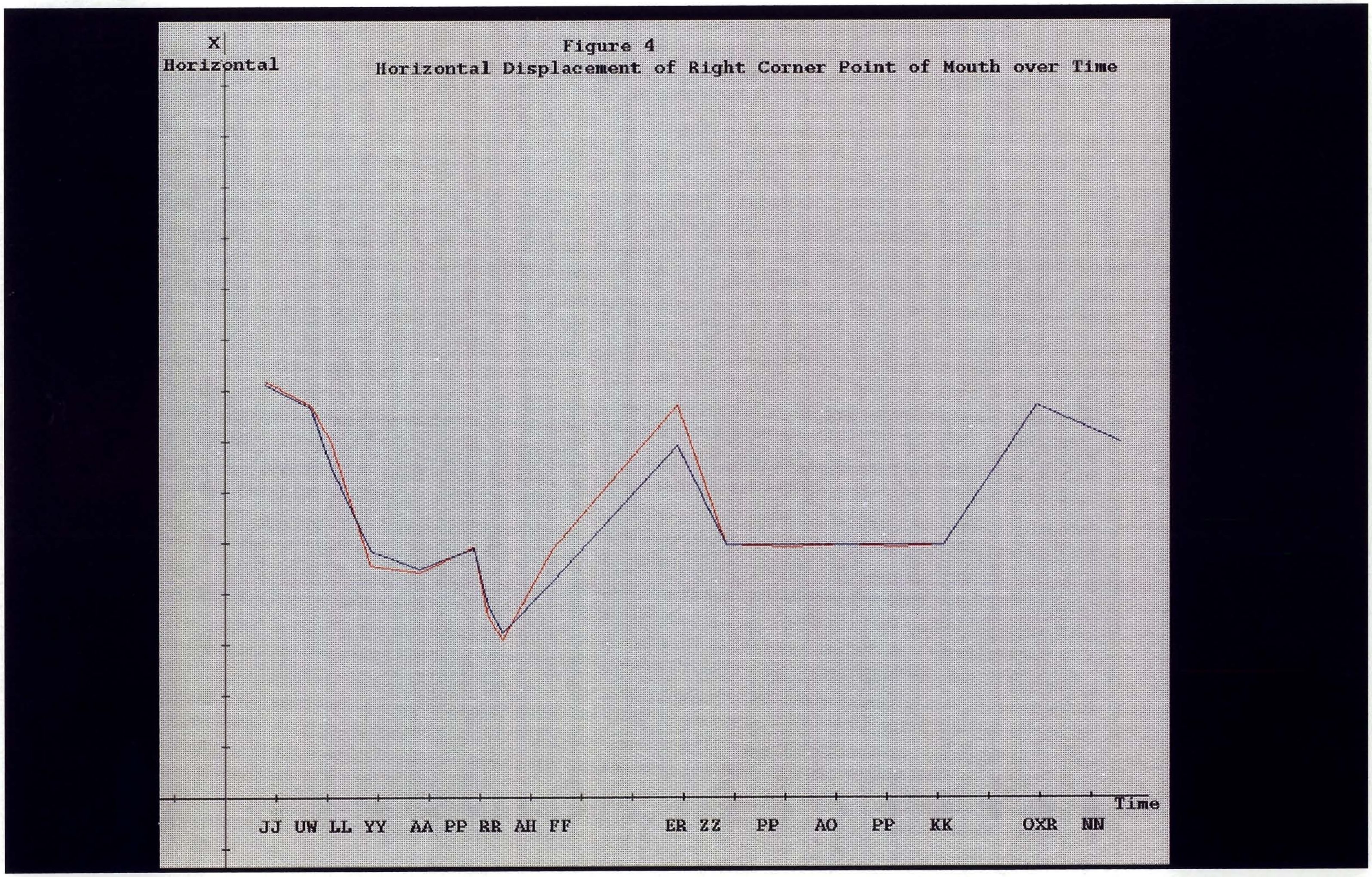


Figure 4
Horizontal Displacement of Right Corner Point of Mouth over Time
X
Horizontal
Time
JJ UW LL YY AA PP RR AH FF ER ZZ PP AO PP KK OXR NN

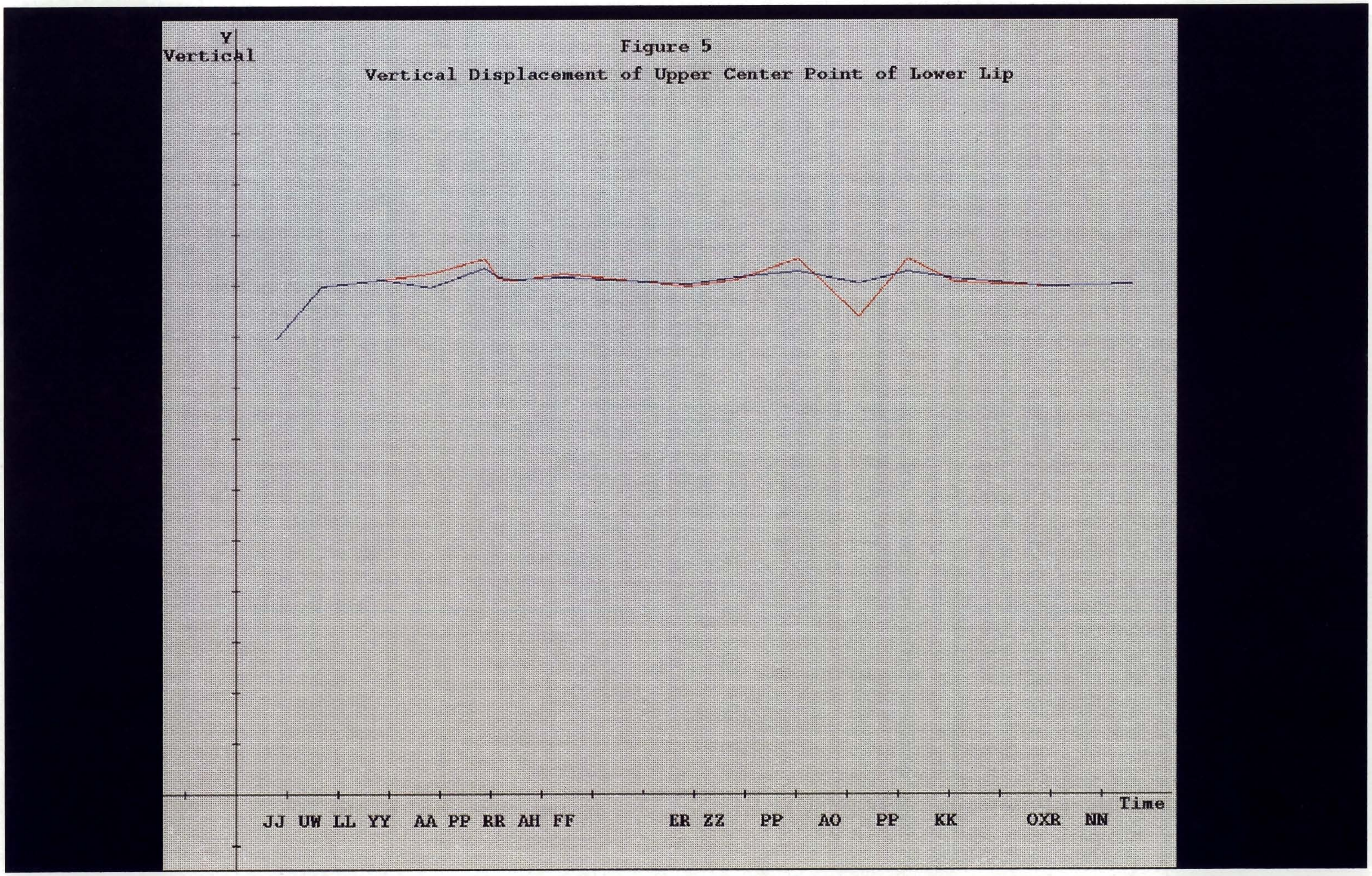


Figure 5
Vertical Displacement of Upper Center Point of Lower Lip

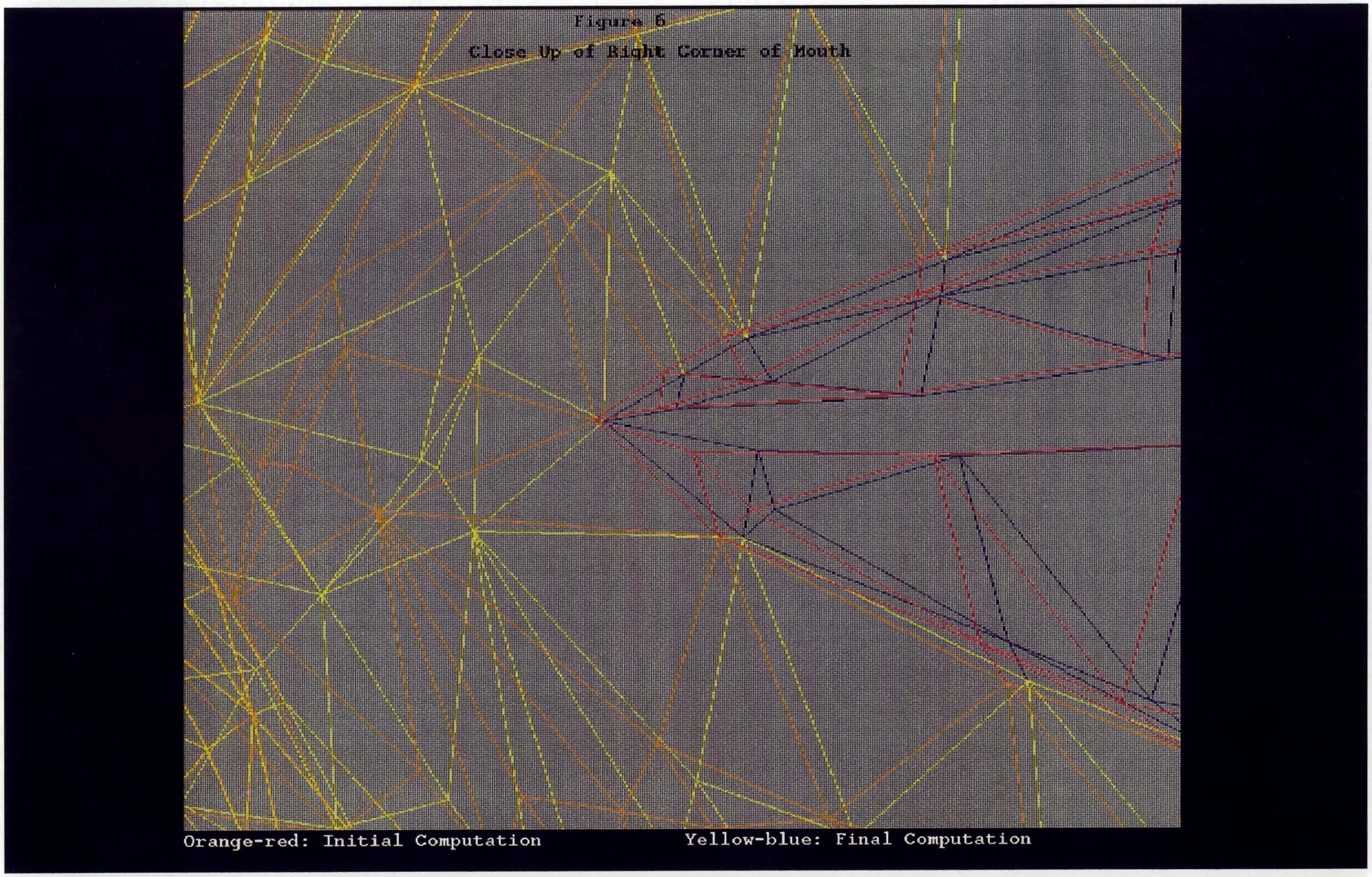
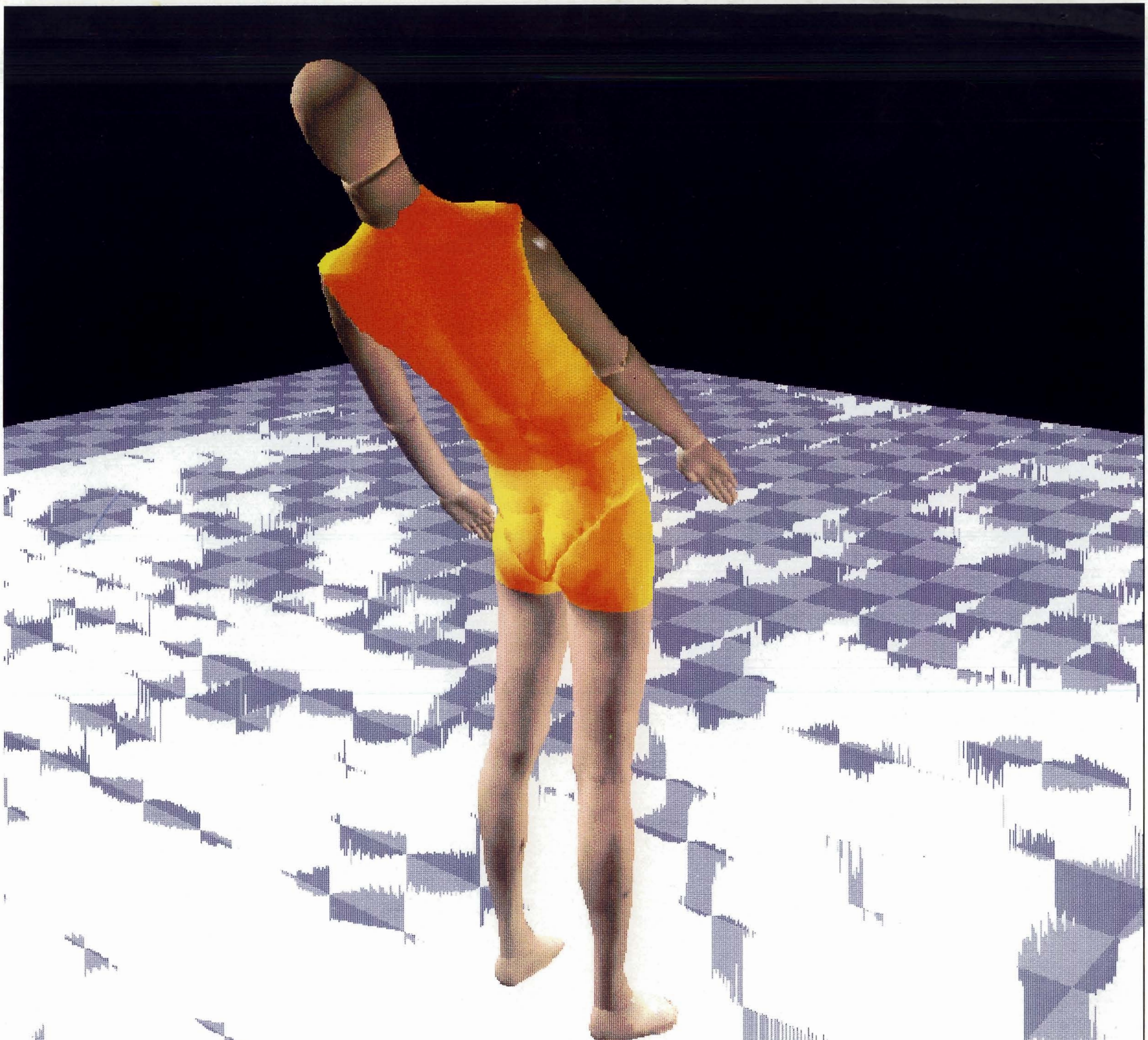


Figure 6

Close Up of Right Corner of Mouth



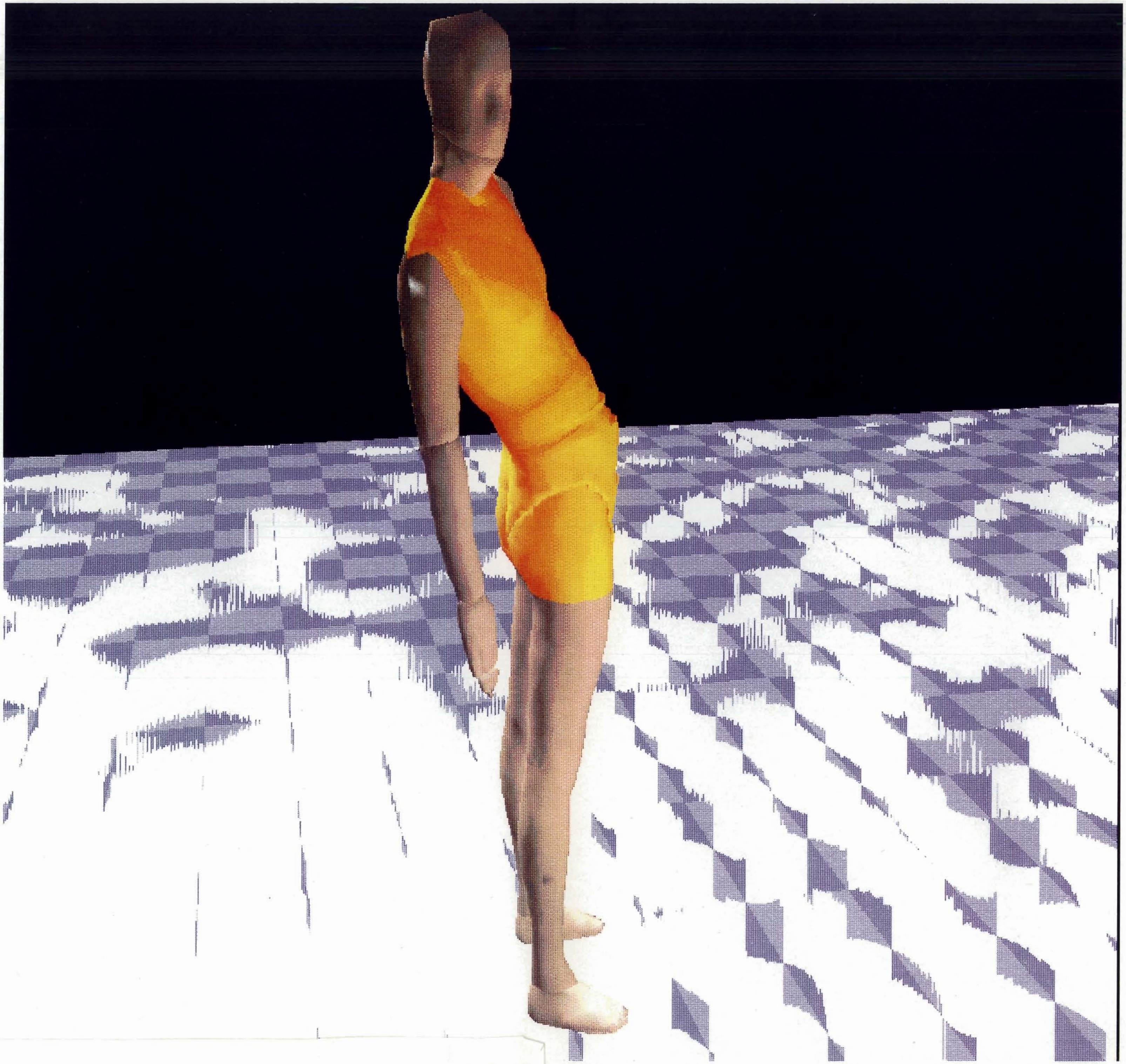


Figure 2.

$(flexion, axial, lateral) = (-1\,\text{deg}, -60\,\text{deg}, -6\,\text{deg})$

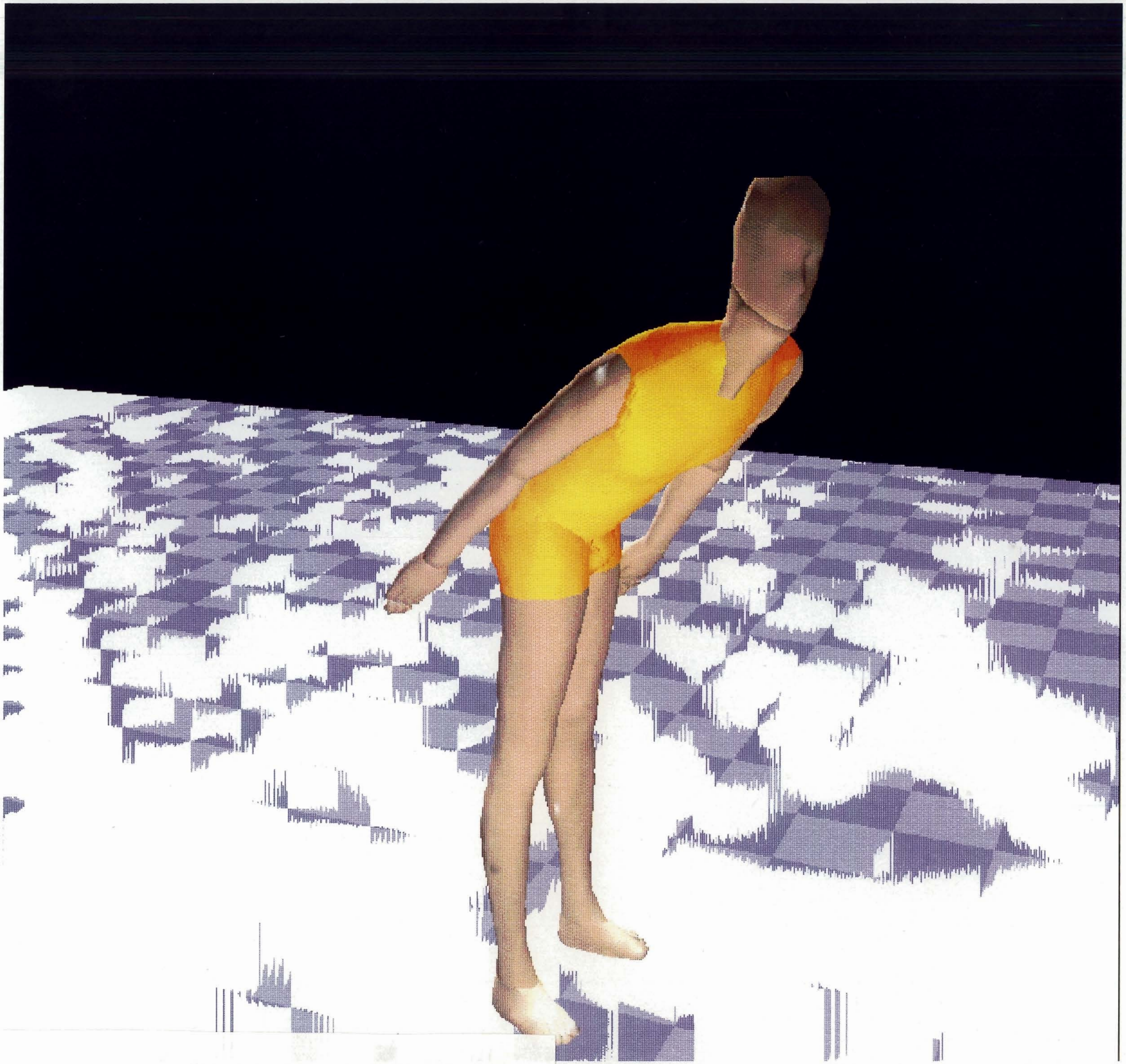


Figure 3.

$(flexion, axial, lateral) = (3\,\text{deg}, -42\,\text{deg}, 23\,\text{deg})$

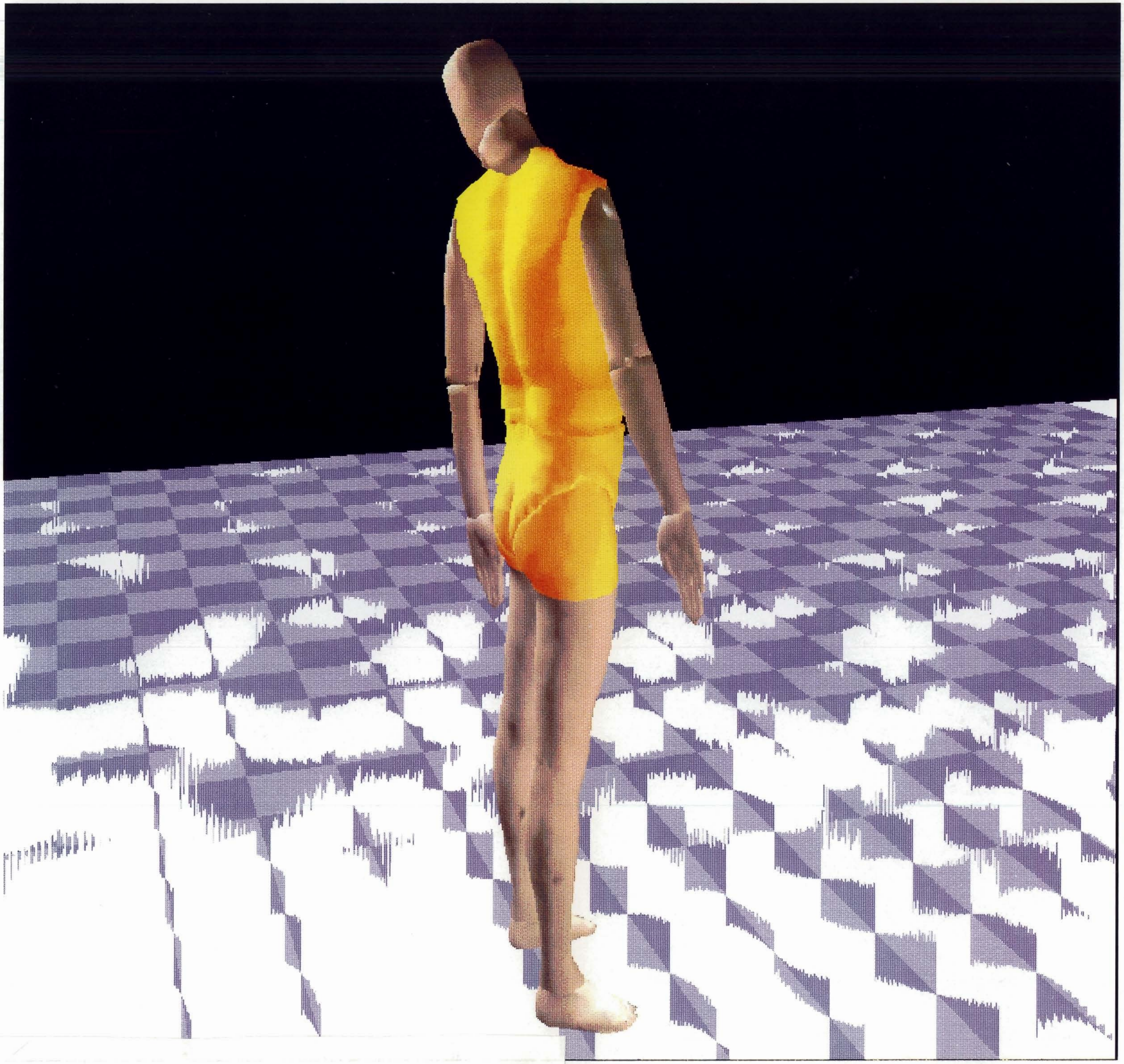


Figure 4.

$(flexion, axial, lateral) = (-10\,\text{deg}, 108\,\text{deg}, -45\,\text{deg})$

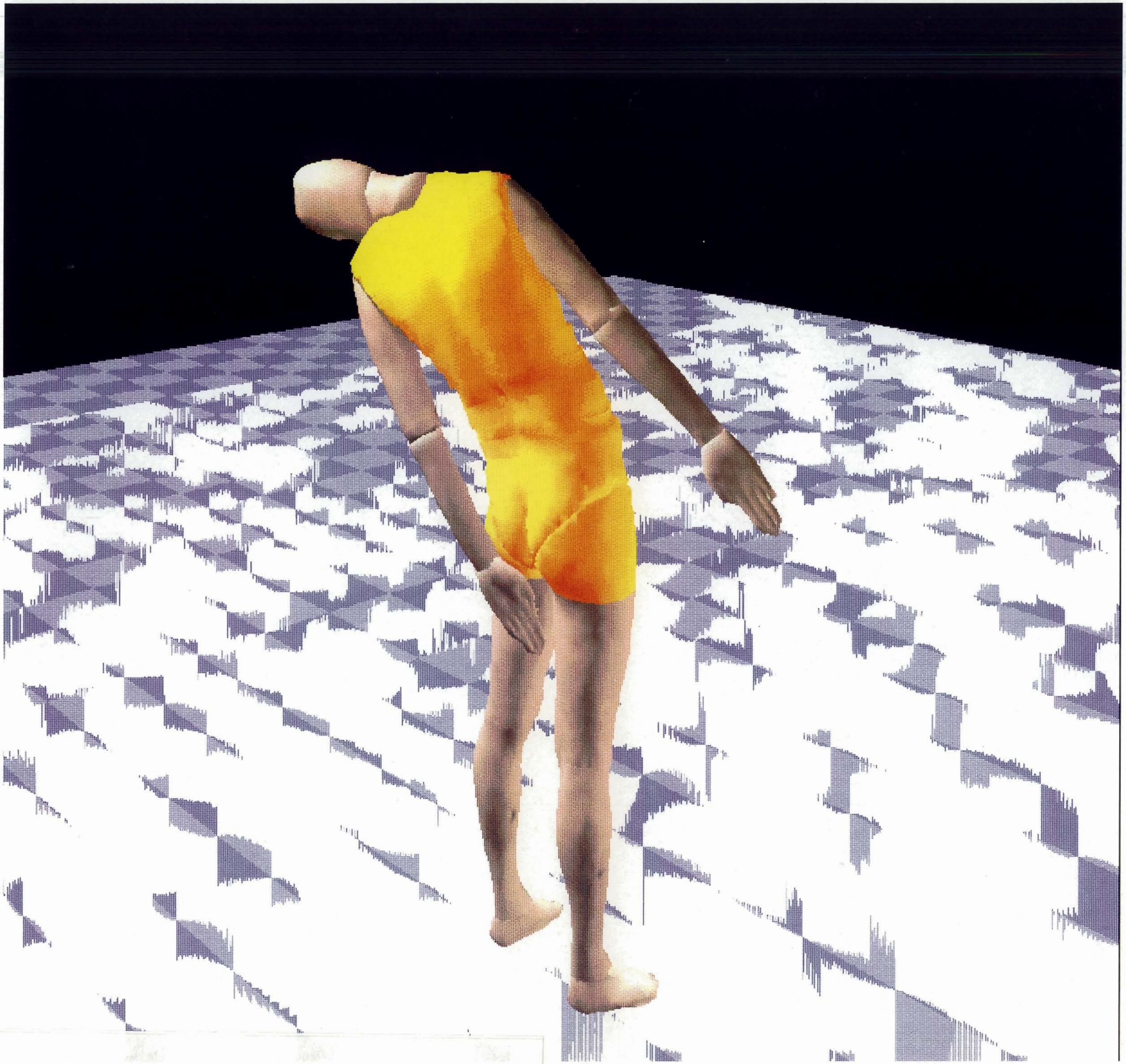


Figure 5.

$(flexion, axial, lateral) = (18\,\text{deg}, 108\,\text{deg}, -68\,\text{deg})$

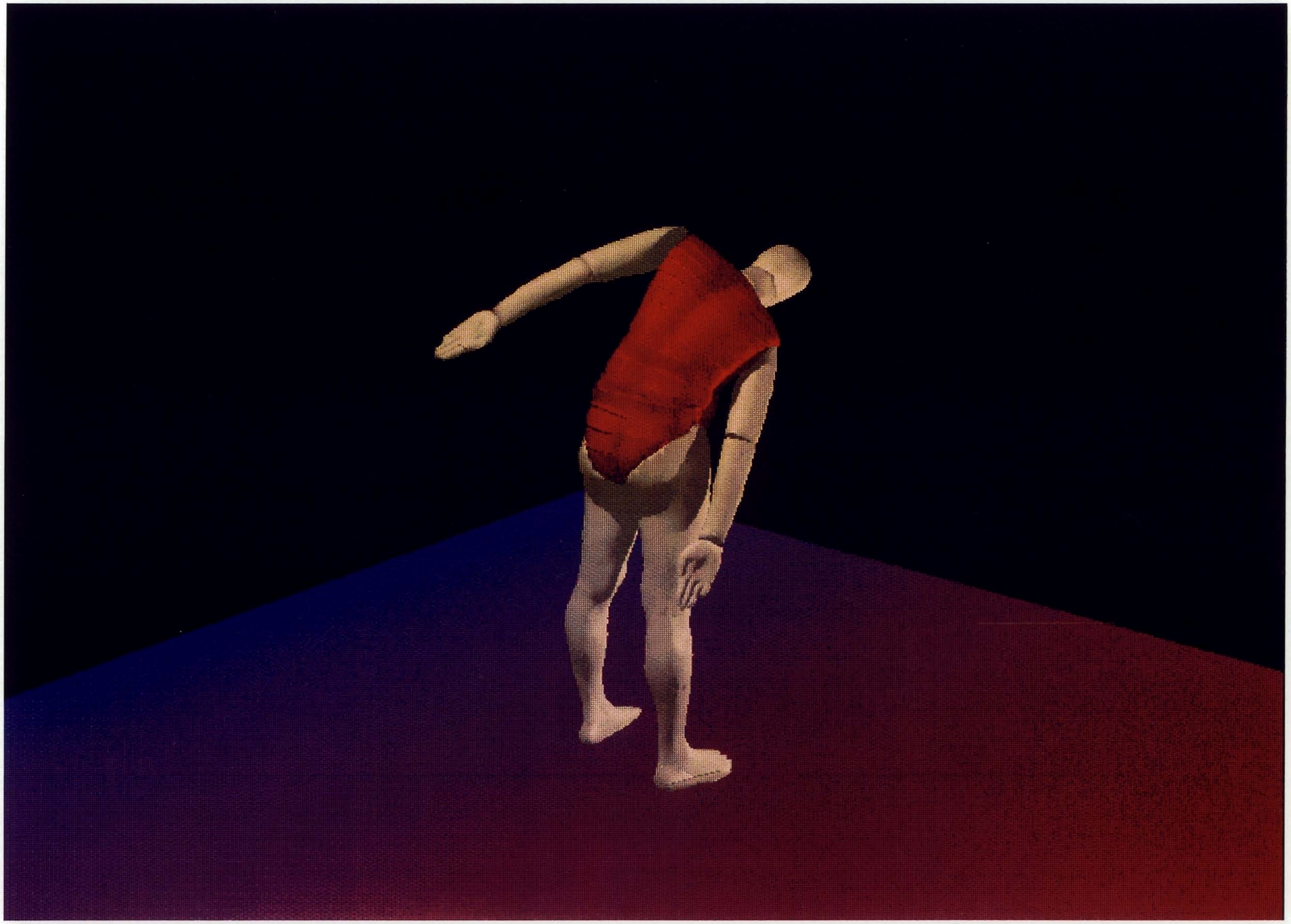


Figure 6.

$$(flexion, axial, lateral) = (-33\,\text{deg}, 39\,\text{deg}, 6\,\text{deg})$$

Figures 7-12. Wireframe polygonal body with visible spine moves through a rolling sequence. Zero interpolation flag is set to "no".

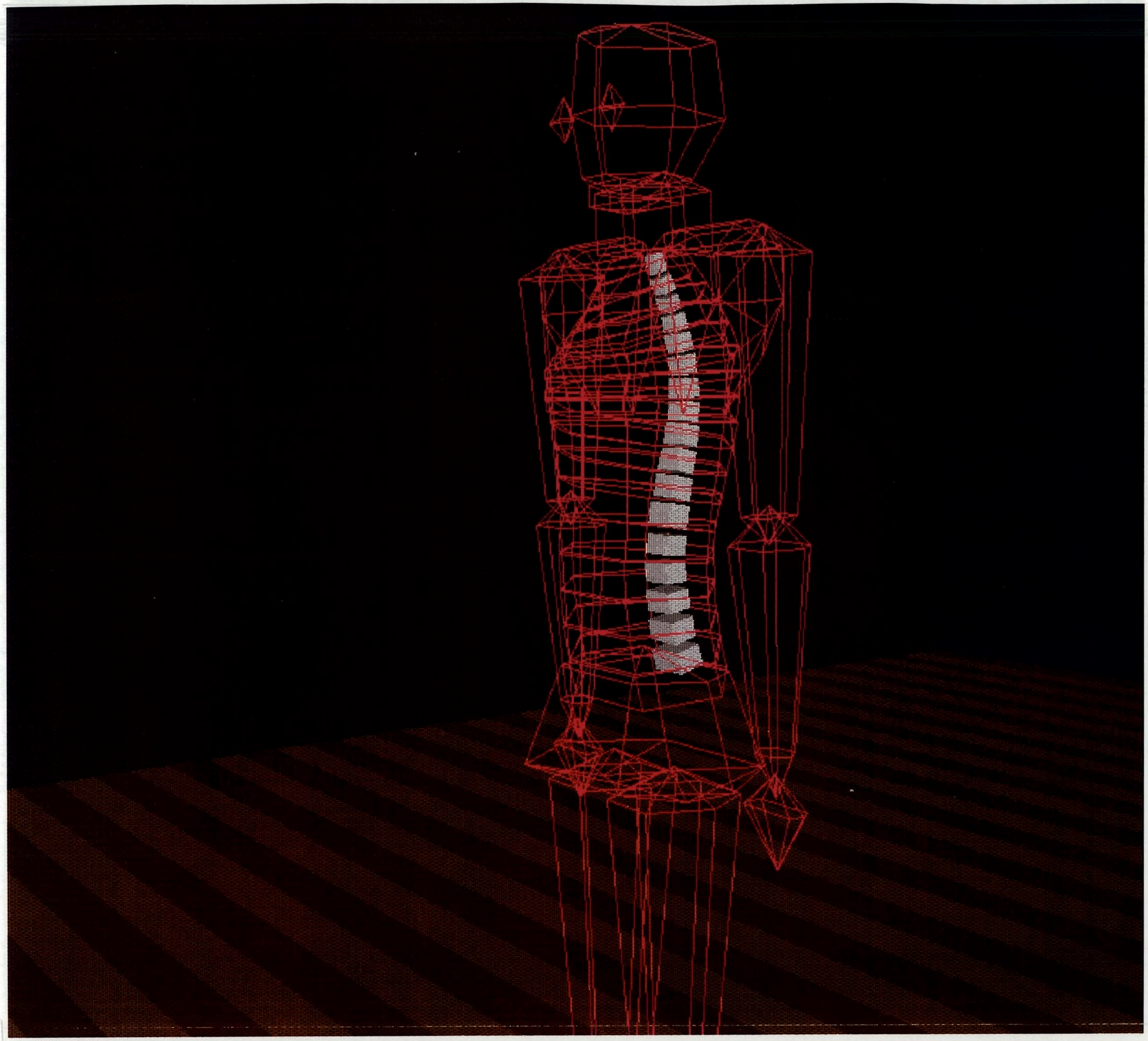


Figure 7

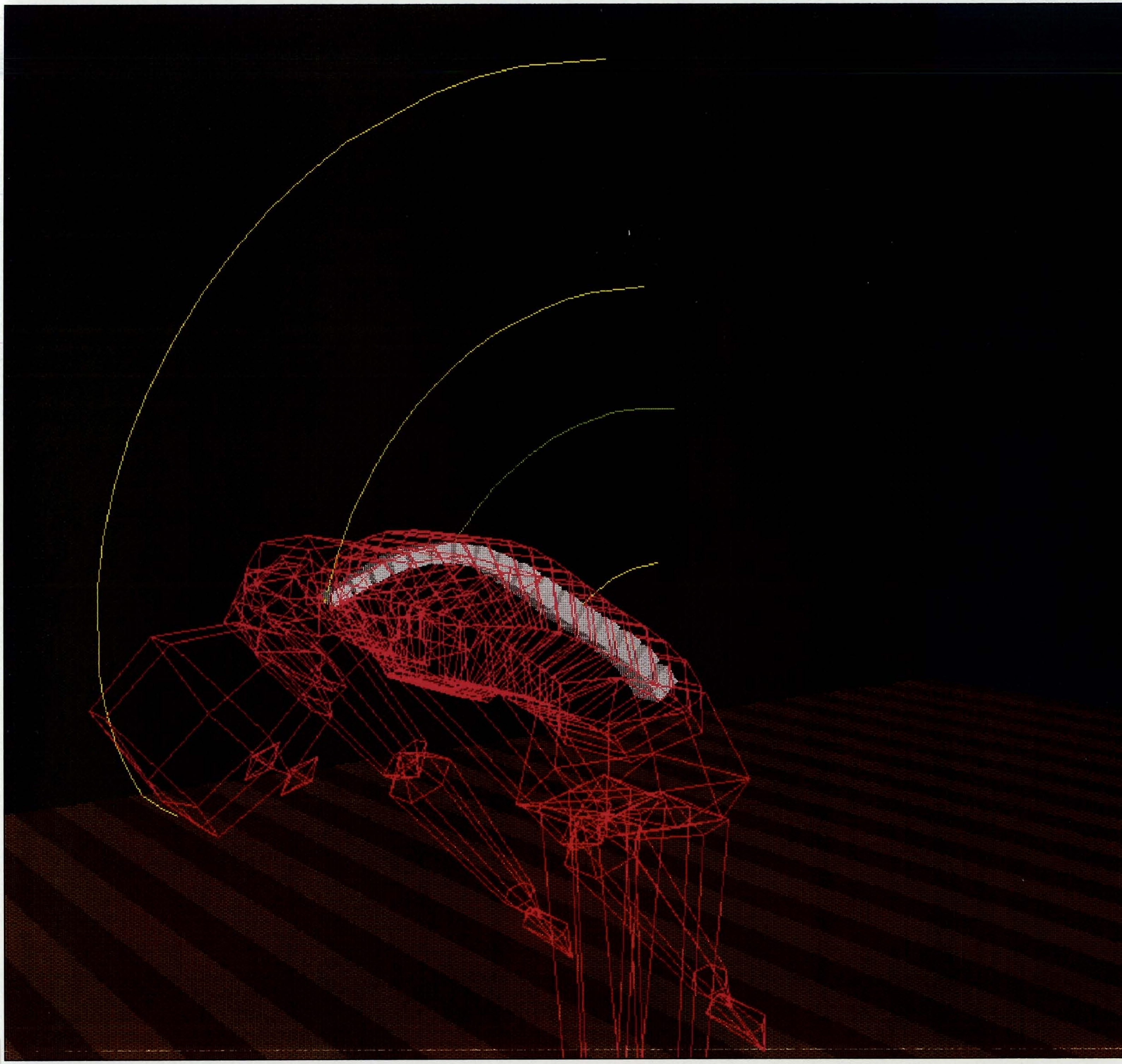


Figure 8

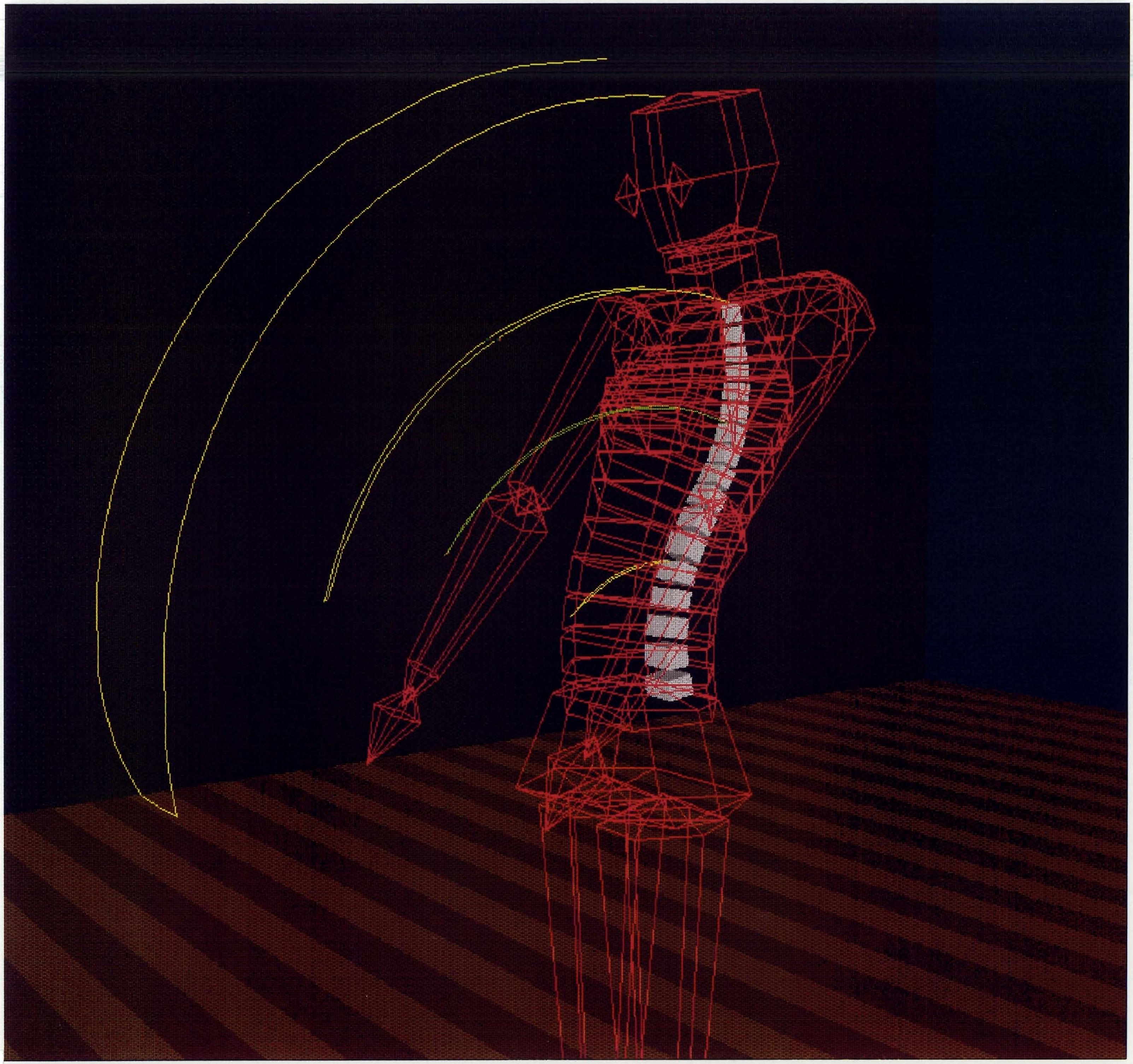


Figure 9

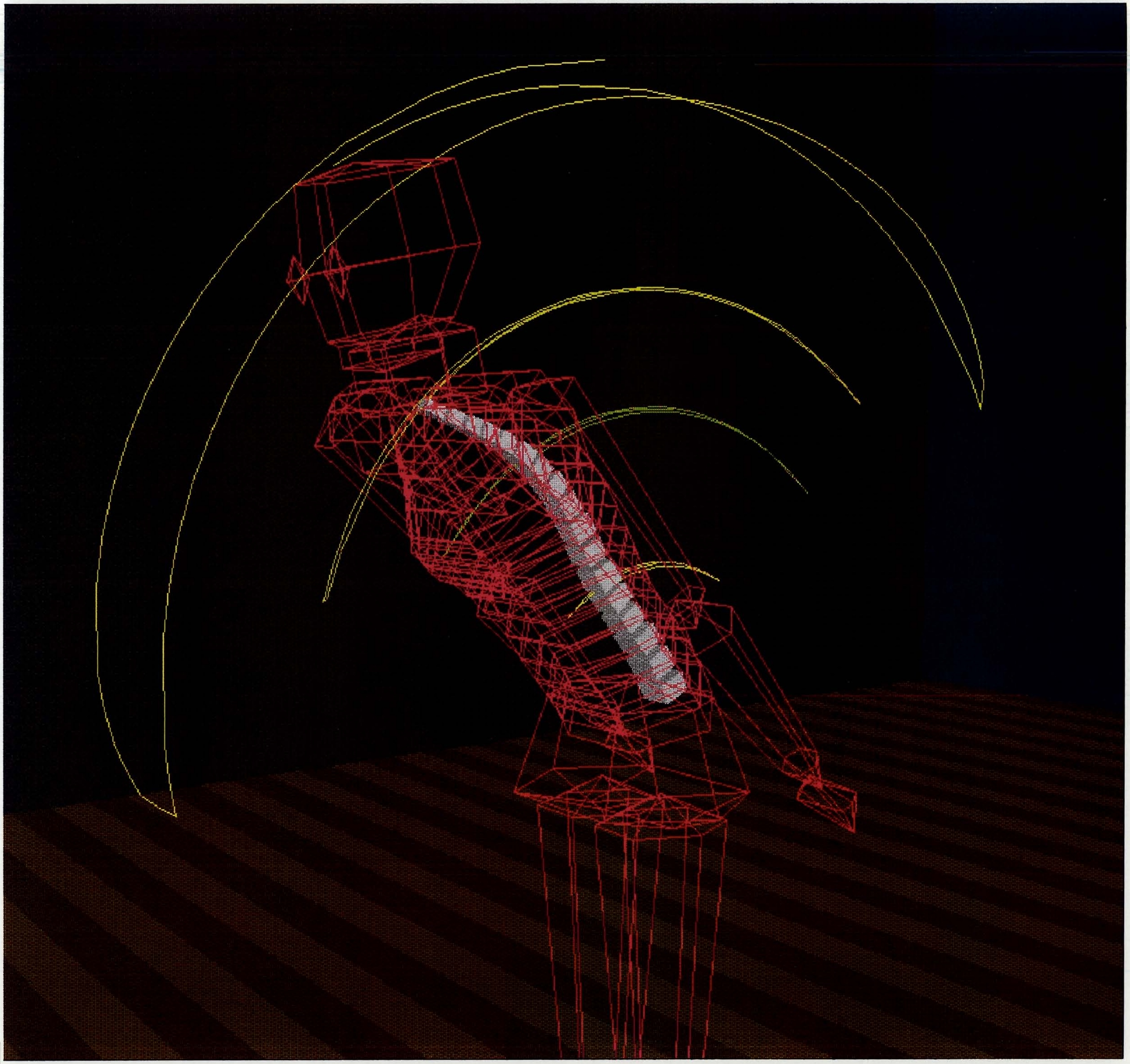


Figure 10

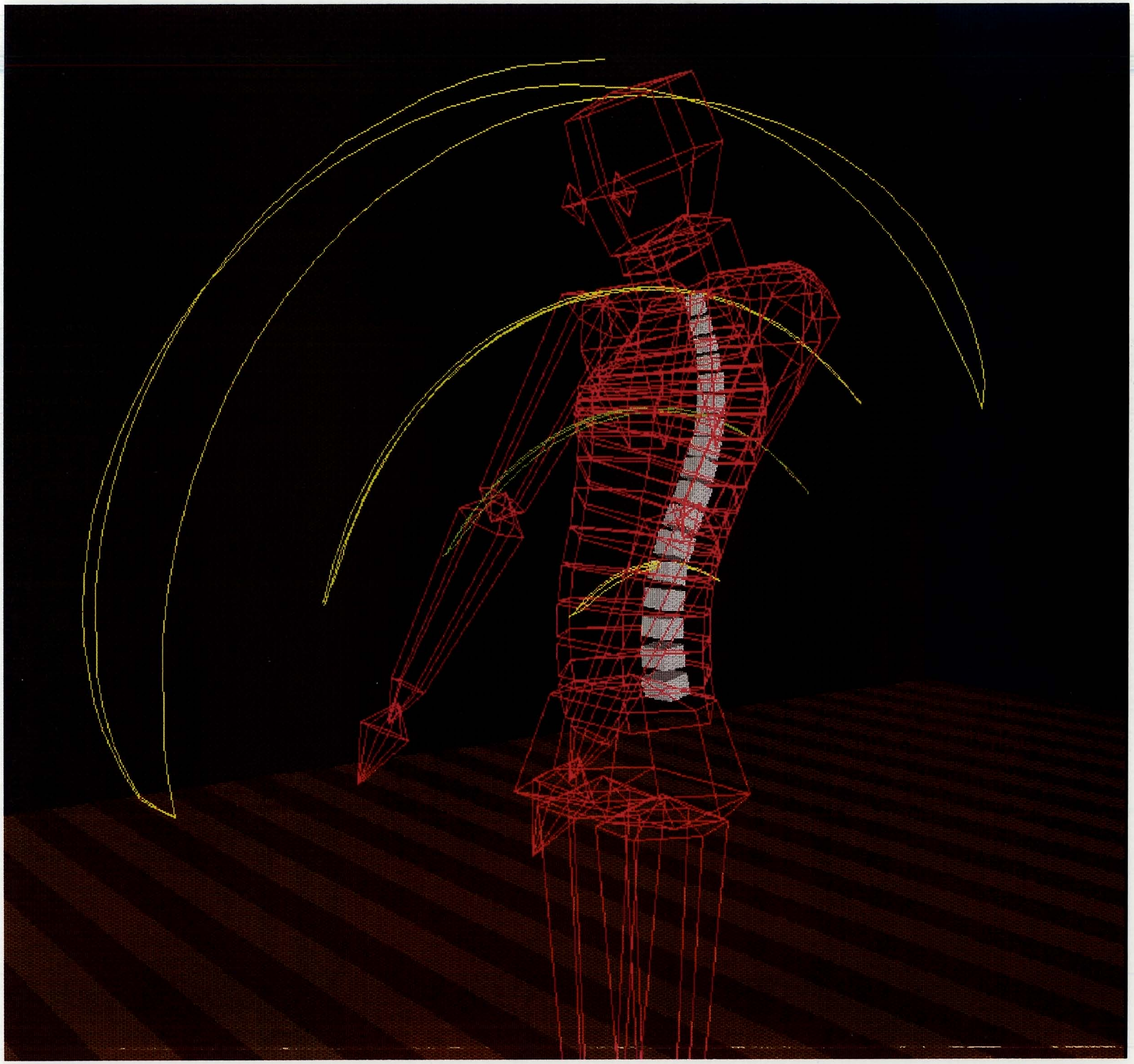


Figure 11

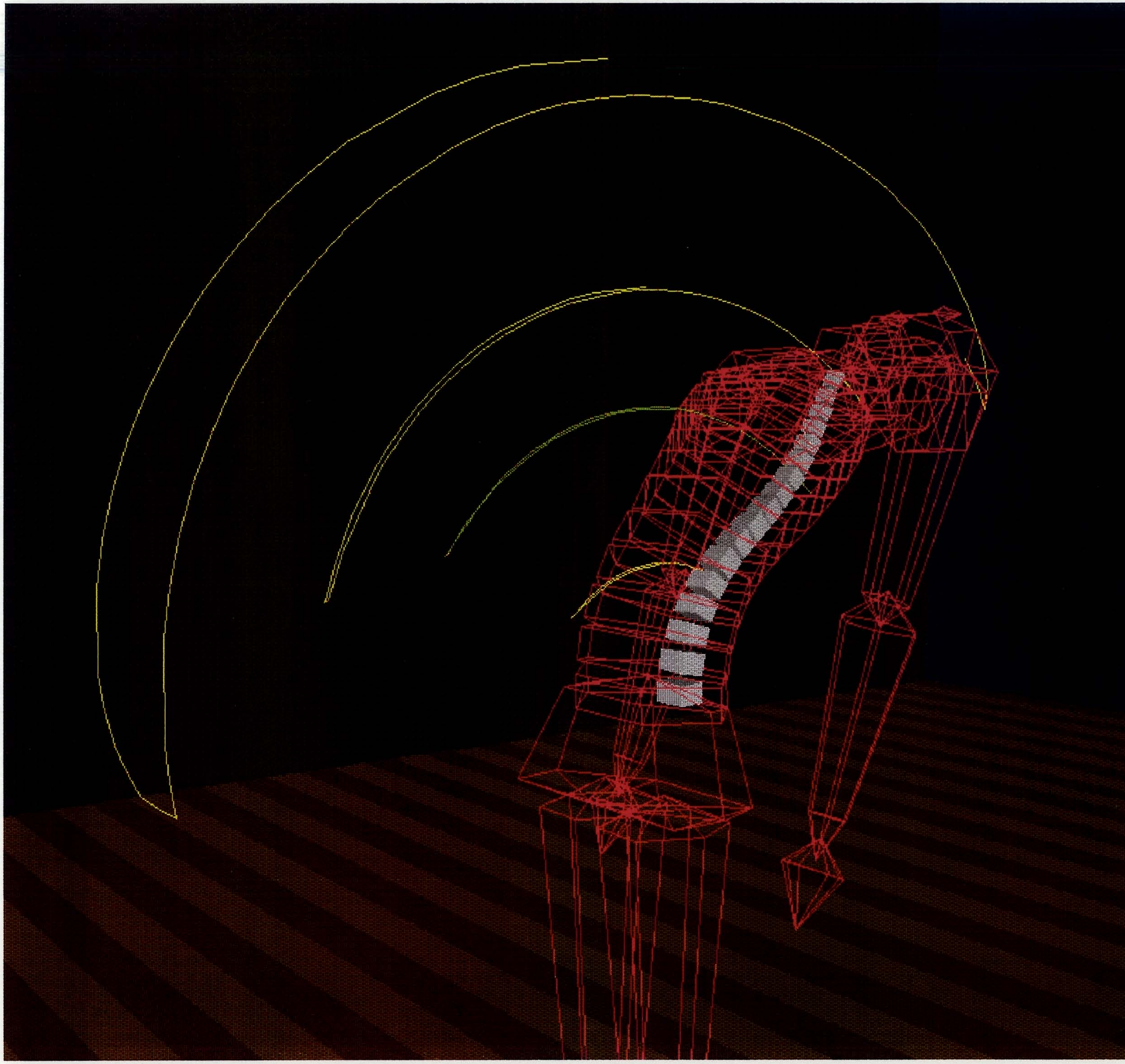


Figure 12

Figures 13-14. Wireframe polygonal body performs circular motion.

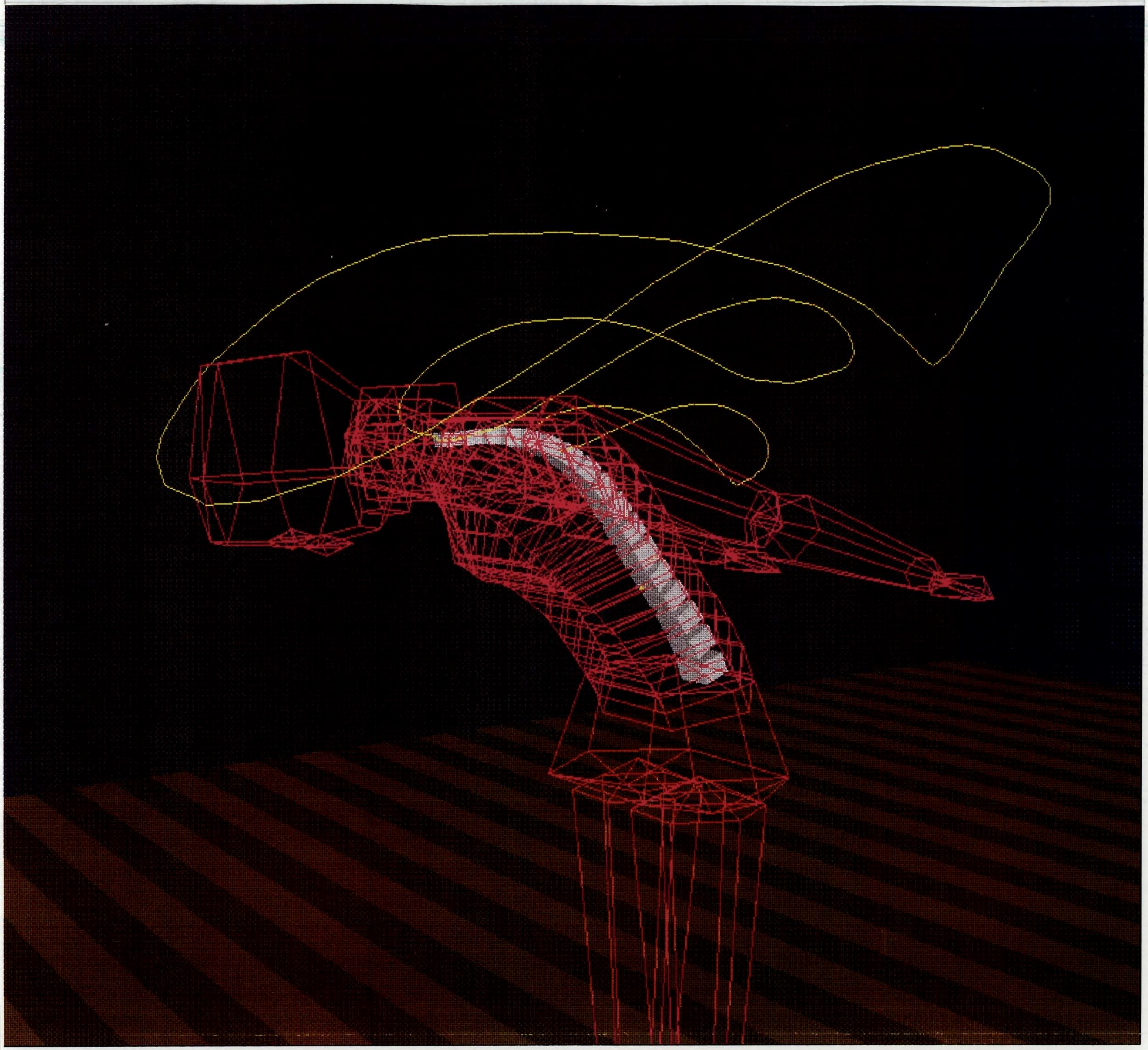


Figure 13

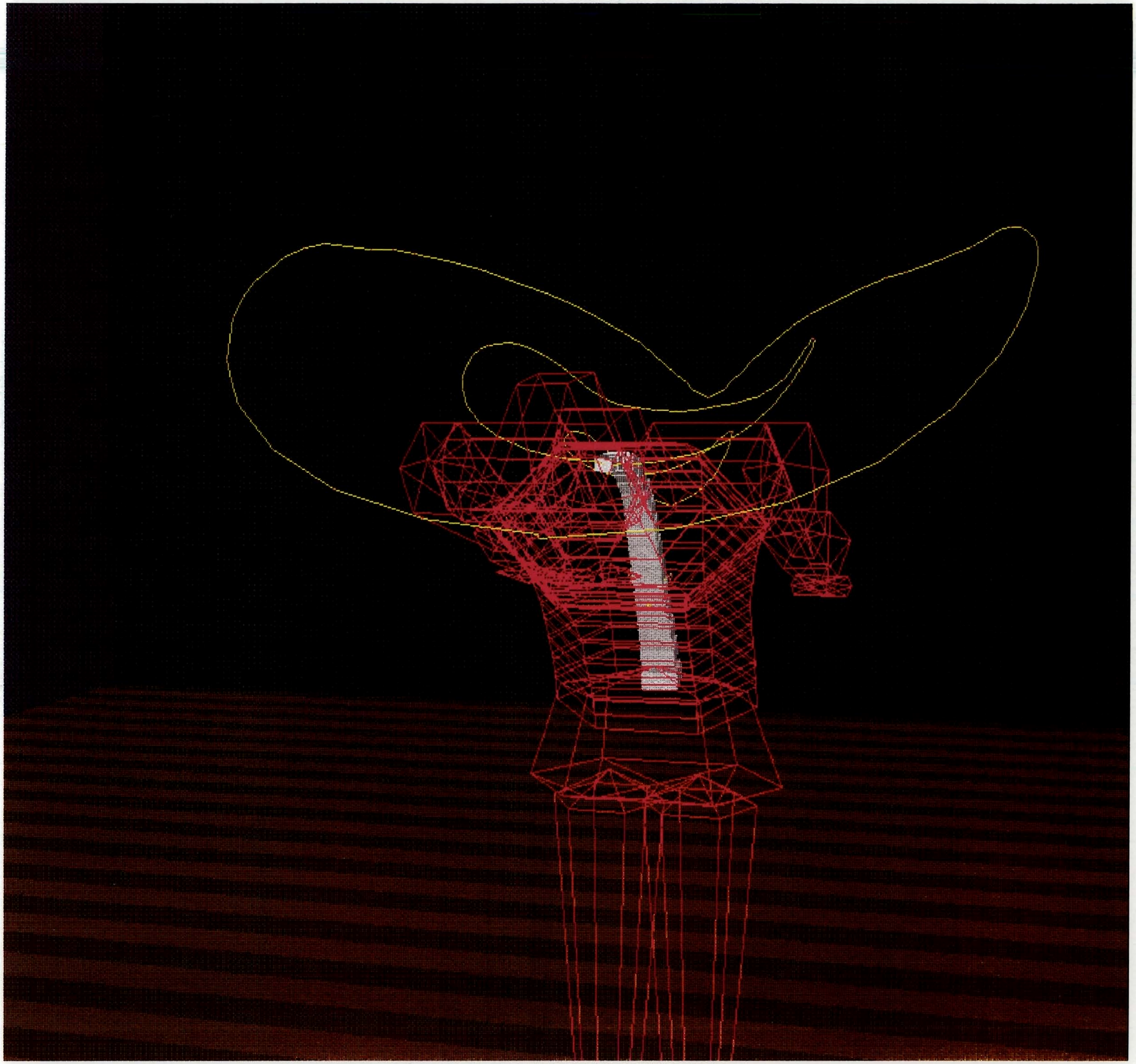


Figure 14

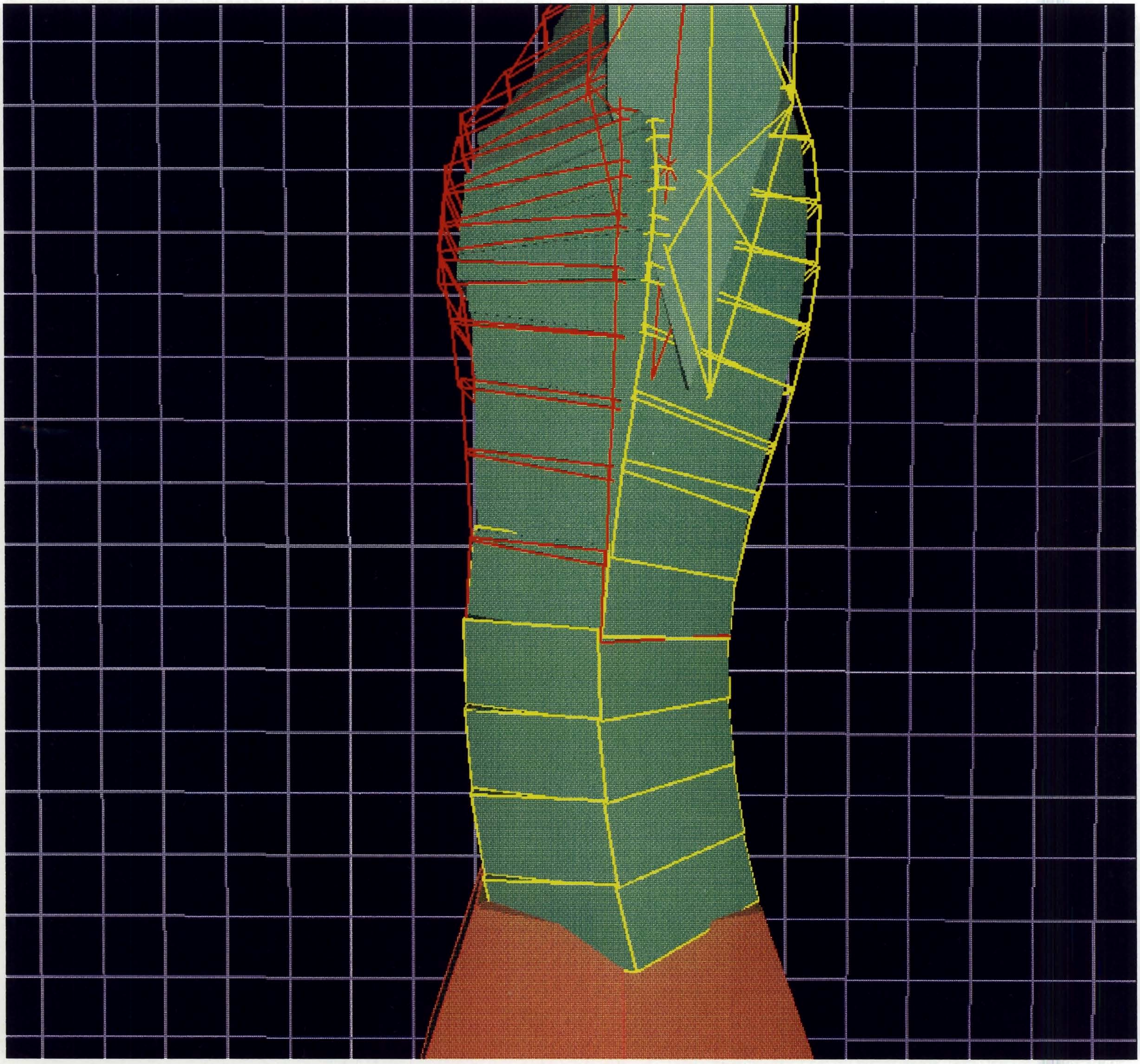


Figure 15. Breathing torso indicates movement over a 5.0 sec breath cycle. Green - 0.0 sec, red - 1.6 sec, yellow - 4.0 sec.