ABOUT FACE:

COMPUTERGRAPHIC SYNTHESIS AND MANIPULATION OF FACIAL IMAGERY

by

Peggy Weil

B.A., Radcliffe College, Harvard University
1976

Submitted in Partial Fulfillment
of the requirements for the
Degree of
Master of Science in Visual Studies
at the
Massachusetts Institute of Techonolgy
September 1982

© Massachusetts Institute of Technology 1982

⌒

Signature of Author........·/···{···/··························
Department of Architecture
August 6, 1982

Certified by./..{.....................\.................
Andrew Lippman, Assistant
Professor of Media Technology
Thesis Supervisor

Accepted by.............../·············{··················
Nicholas P. Negroponte
Chairman, Departmental Committee on Graduate Students

Rotch

MITLibraries
Document Services

DISCLAIMER NOTICE

The accompanying media item for this thesis is available in the MIT Libraries or Institute Archives.

Thank you.

ABOUT FACE:

COMPUTERGRAPHIC SYNTHESIS AND MANIPULATION OF FACIAL IMAGERY

by

Peggy Weil

ABSTRACT

A technique of pictorially synthesizing facial imagery using
optical videodiscs under computer control is described. Search,
selection and averagin processes are performed on a catalogue
of whole faces and facial features to yield a composite,
expressive, recognizable face. An immediate application of
this techique is the reconstruction of a particular face from
memory for police identification, thus the project is called,
IDENTIDISC. Part I-FACEMAKER describes the production and
implementation of the IDENTIDISC system to produce composite
faces. Part II-EXPRESSIONMAKER describes animation techniques
to add expression and motion to composite faces. Expression
sequences are manipulated to make 'anyface' make any face.

Historical precedents of making facial composites, theories of
facial recognition, classification and expression are also
discussed.

This thesis is accompanied by two copies of FACEMAKER-III, an
optical videodisc produced at the Architecture Machine Group in
1982. The disc can be played on an optical videodisc player.
The length is approximately 15,0000 frames. Frame numbers
are indicated in the text by [ ].

CONTENTS

I. FACEMAKER

FACEMAKER is dedicated to all the faces
that added another point of view.

1.0 Introduction and Objectives

> *Our general impressions are founded upon blended memories.*
> — *Francis Galton,*
> *Inquiries into Human Faculty*

> *The purpose of recall implies at least the possibility of*
> *causing two impressions to appear as quasi-likeness...Without*
> *imagination there would be no resemblance between things.*
> — *Michel Foucoult,*
> *The Order of Things*

FACEMAKER is a technique for pictorially synthesizing facial images by intensity averaging a selection of photographs of registered faces which are stored on optical videodisc. The stated, and most immediate application of this technique is the reconstruction of a particular face from memory, as in the case of police identification of a suspect from the description of an assailant. The objective of this research is to demonstrate a computerized Identification system to generate a target face by averaging photographs to form a composite.

The procedure is to average (either photographically or computationally) together faces which are judged to be similar to the "remembered" face (hereafter referred to as the "target"). Photographic averaging is done by making multiple exposures in the camera, computational averaging is done by averaging intensity values of pixels in successive digitized images. In the averaging process, the features of the component faces which bear a resemblance to the target

6.

tend to become emphasized, as the features which are dissimilar tend to be lost—effectively a pictorial noise cancellation procedure. The result is an approximate likeness to the target; a pictorial list of traits for which the components were chosen. The Identidisc system replaces photographic methods of averaging with electronic techniques to allow for dynamic interaction, breadth and to accommodate classification procedures.

The challenge that FACEMAKER poses to feature-based identification systems is based on the assertion that it is compatable with the way we recognize and remember faces. Chapter Two examines research regarding facial recognition and recall. Chapter Three reviews the historical precedents of making photographic facial composites—a practice that dates back to the early days of photography. Chapter Four considers theories of facial classification—relevant to the task of organizing several thousand photographs of faces. Chapters Five, Six, Seven and Eight describe the IdentiDisc System; they elaborate and illustrate the procedure for producing composite faces.

EXPRESSIONMAKER forms the second part of this thesis. The objective of EXPRESSIONMAKER is to add expression and motion to composite faces by integrating techniques described above with animation. The text is organized similarly to FACEMAKER; and could almost be considered a separate body of work. Chapter Nine is an introduction. Chapter Ten recounts theoretical approaches to the study of facial expression—conjectures as to its cause and origin and attempts at classification. Chapter Ten discusses the portrayal of facial expression in art, theater and animation. Chapter Eleven is an account

of my efforts to animate faces within the mixed contexts of ongoing
research in teleconferencing and facial identification.

2.0 Facial Recognition and Memory

The issues of facial recognition and recall are clearly pertinent to the creation of a facial identification system. There is a broad spectrum of research concerned with memory for faces; ranging from psychological literature investigating the underlying perceptual mechanisms for pictorial stimuli to engineering questions regarding the task of machine recognition of faces from mathematical descriptions. Of interest to this project are studies of the role of photographs to identification procedures, the effect of the length of exposure for recall, and discussions of factors which have been shown to affect recognition.

Following is a brief survey of some of the literature with references to these issues. The ability to reconstruct a face may not be commensurate with the ability to recognize one so these studies must be interpreted as suggestive rather than conclusive within the given context. The difference between experimental laboratory procedures and actual situations where an identification system might be used highlight areas which remain to be studied; specifically, Ellis et.al. (1975) suggests that research be done to determine the affect of emotional excitement (i.e., the trauma of a criminal encounter) on recognition, and on the discrepancies between recognition of photographs and real people.

9.

Our capacity for face recognition is awesome; as yet, no one has demonstrated the limits, if any, of our ability to make any amount of faces familiar, and to subsequently recognize them regardless of exposure to additional faces or the lapse of time (Bahrick, Bahrick & Wittinger, 1975). We can distinguish among particular faces despite distortions due to expression, transformations due to aging, injury or disease, shielding due to glasses or clothing or point of view. Faces present a special problem in the investigation of perception; in contrast to the tremendous perceptual resiliancy cited above, inversion, either spatial (turning a picture upside-down) or tonal (substituting a photographic negative for a positive) greatly impair recognition performance (Rock, 1974; Galper, 1970). Carey, in studies of the development and biological basis of face recognition, interprets this as suggesting our representation of faces may be largely configurational:

> Rather than being identified by particular features, it
> appears that faces are encoded with reference to some
> canonical arrangement of features....Furthermore, indivi-
> dual faces present a more unique distinguishing feature—
> the spatial relations among their parts. For instance,
> the ratio of the distance the hairline is above the
> chin to the distance the bridge of the nose is above the
> center of the upper lip could be a distinctive feature
> which might serve to individuate faces. Represented by
> a number of such spatial relations, each face forms a
> Gestalt as unique as a snowflake. (Carey, 1978).

Harmon's (1973) work with recognition of precisely blurred photographs seems to support this view. His objective was to determine the perceptual threshold for recognition as visual information

was successively reduced. He found that the most severely blurred pictures were still identified with an accuracy approaching sixty percent. He concludes that, in these cases, recognition cannot depend on the identification of individual features and that low-frequency information (shape and rough hairline) may be adequate.

Davies and Christie (1982) have specifically addressed the significance of this issue with regard to the task of police identification from feature-based composites. They found that subjects had difficulty making judgments of similarity between a target and facial features when the features were isolated from a facial context. They attributed this difficulty as evidence supporting a Gestalt position on facial recall, suggesting that existing feature-based identification systems may be incompatible with the way we remember faces. They suggest that alternative methods of facial reconstruction use combinations of whole faces and that features be presented in configurations rather than in isolation—fitting the method of recalling a face to the way we remember one.

Most studies of facial recognition use photographs. At issue here are questions whether the type of photograph is a factor influencing recognition. Laughery et.al. (1971) in a study designed to model the task of a subject attempting to identify a criminal from a set of mug shots found that neither type (black and white or color) nor point of view (frontal, three-quarter, or full profile poses) appreciably affected subjects' ability to recognize faces. These findings were counter to their expectations and they suggested, may be

11.

attributable to the subjects having reached a threshold of information

processing within the context of the experiments. It was thought that

perhaps that memory for different features were emphasized within the

different poses to account for the findings but no mention was made of

using more than one point of view to aid identification.

There do not seem to be studies involving recognition of

faces using stereo photographs, only conjecture:

> As yet, no one has examined the relative effect on
> recognition of showing three-dimensional faces, but
> it is unlikely that much difference would be found,
> unless perhaps the three-dimensional faces are moved
> in the expressive manner normally encountered when
> viewing a real face. Such movements might assist
> later recognition through additional verbal coding
> (e.g., "the one who always smiles"). (Ellis, 1975)

Hoffman (1981) argues that depth information is essential to

build representations of faces in a paper which discusses the visual

interpretation of faces as a computational problem. He cites studies

done by Carey and Diamond (1977, 1981) which suggest that children

need stereo and motion cues in order to build holistic descriptions of

faces. Adults probably use a priori knowledge of the structure of

faces to infer shape from shading in photographs and lines in simple

line drawings. Experiments along these lines would contribute to the

understanding of the value of stereo in face recognition.

There are several issues of exposure: the length of time

the subject is exposed to the target, the interval between that expo-

sure and the attempt at identification, and the interference effect

of viewing large numbers of faces prior to success at identification.

These considerations, as will be seen later, are especially relevant to this thesis and resulted in the creation of an animation program to pre-select a pool of faces at the start of the facemaking process.

Carey (1978) cites studies (Bahrick, Bahrick & Wittenger, 1975) which suggest that neither time nor exposure to large amounts of faces interfere with our memory for faces. These questions have also been studied by Laughery et.al. (1971), Mauldin (1978) and the University of Houston Mugfile Project within the context of testing the ability to search mugfiles to make an accurate identification.

As would be expected, it was found (Laughery et.al., 1971) that the probability of recognition improved as the exposure time (that of the subject to the target) was lengthened. Both Laughery (1974) and Mauldin (1978) conducted tests which showed that the interval between exposure and identification was not a significant factor affecting recognition. Laughery (1971) found the position of the target within the set of faces to be searched to be a significant factor; recognition performance was impaired as the number of pictures the subject had to look through increased.

Mauldin (1978) examines the influence of initial recall of facial characteristics (in order to reduce the search set) upon sub-sequent recognition. He found that "memory rehearsal" involved in generating a composite seemed to facilitate later identification and suggests that the utility of composites may not be in providing good representations of suspects, but rather as a procedure of memory enhancement preparing the witness for the task of identification.
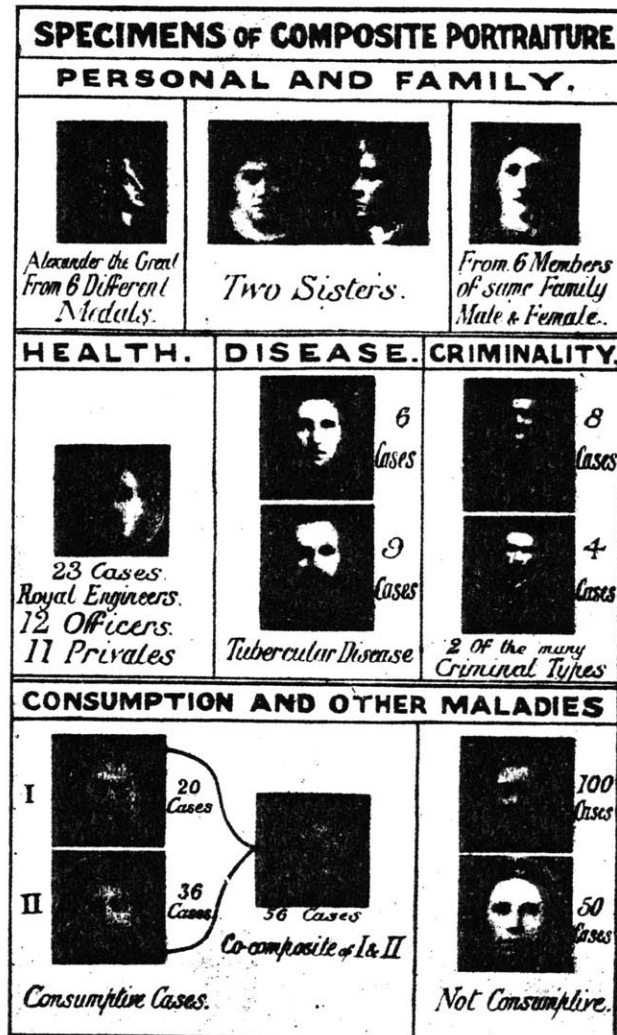
3.0 Facial Composites

*The photographic process of which I there spoke...represents
no man in particular, but portrays an imaginary figure
possessing the average feature of any given group of men.
These ideal faces have surprising air of reality, nobody
who glanced at one of them for the first time would doubt
its being the likeness of a living person, yet, as I have
said, it is no such thing; it is the portrait of a type and
not of an individual.*
                                        —Galton, 1879

The technique of composite portraiture was pioneered, perhaps
invented, in the later part of the nineteenth century by the British
philosopher, Sir Francis Galton. He introduced composite portraiture
as a method of "pictorial statistics"—photographic averages which were
equivalent to statistical tables of physical traits. His examples are
attempts to present a "generic type"; the creation of an imaginary
individual to represent, say, criminality, race, a specific disease,
etc. His "Portrait of Health" is a composite of twenty-three British
Royal Engineers—twelve officers and eleven privates—all, presumably
in good health. Galton hoped to produce a true likeness of Alexander
the Great by making a composite of his portraits on six different
medallions, claiming that his method would compensate for any errors
on the part of the individual artists. [Frames 11652-3]

Galton made his composite portraits photographically by making
multiple exposures of registered faces. Although he prefers photographs,
he cites a letter, forwarded to him by Darwin, which suggests the use

SPECIMENS OF COMPOSITE PORTRAITURE

PERSONAL AND FAMILY.

Alexander the Great From 6 Different Medals.

Two Sisters.

From 6 Members of same Family Male & Female.

HEALTH. DISEASE. CRIMINALITY.

23 Cases. Royal Engineers. 12 Officers. 11 Privates

6 Cases

9 Cases

Tubercular Disease

8 Cases

4 Cases

2 Of the many Criminal Types

CONSUMPTION AND OTHER MALADIES

I   20 Cases

II   36 Cases

56 Cases
Co-composite of I & II

Consumptive Cases.

100 Cases

50 Cases

Not Consumptive.

To face page 8

figure 1.   Galton(1879)

15.

of a stereoscope. Two faces presented separately to each eye will combine to form a perceived composite image.

> I find by taking two ordinary carte-de-visite photos of two
> different persons' faces, the portraits being about the same
> sizes, and looking about the same direction, and placing them
> in a stereoscope, the faces blend into one in a most remark-
> able manner, producing in the case of some ladies' portraits,
> in every instance, a decided improvement in beauty. (A.L.
> Austin cited in Galton, 1879).

Although Galton doesn't seem to have suggested the application of criminal Identification, he anticipates it with his view that a composite portrait is a means of accessing human memory—the blended image is an attempt to make a physical representation of an idea, an aggregate of multiple impressions.

> My argument is, that the generic images that arise before
> the mind's eye, and the general impressions which are faint
> and faulty editions of them, are the analogues of these
> composite pictures which we have the advantage of examining
> at leisure, and whose peculiarities and character we can
> investigate, and from which we may draw conclusions that
> shall throw much light on the nature of certain mental
> processes which are too mobile and evanescent to be directly
> dealt with. (Galton, 1879).

This position was later taken up by Psotka (1977) who discussed composite photographs (he averaged objects as well as faces) as approximations of a concept in memory. He suggests the application of composite portraiture to criminal Identification and conducted experiments to demonstrate its potential value. Subjects were asked to remember a target face and then attempt to reconstruct its likeness by selecting similar faces out of a larger collection and then making an averaged composite. Psotka, like Galton, made his averages

photographically by making multiple exposures or projecting super-

imposed images:

> Presumably, each of the selected individual faces differed
> from the appropriate target in fewer ways than it was
> similar to the target. Compositing these faces enhanced
> the points of similarity and diminished the areas of
> difference. (Psotka, 1977).

He found that another group of subjects were able to success-

fully match the composite photographs made by the first group to the

appropriate targets and concludes that the procedure could be devel-

oped for criminal identification.

4.0 Classification of Faces

Several thousand faces stored on optical videodisc form the raw material for making faces with this system. Decisions regarding classification affected the actual sequencing of the faces on the disc, presentation and selection schemes. The construction of a data-base referencing all the faces on the disc has yet to be accomplished. It is difficult to establish a precise system to describe a face, the configurations which may be necessary to trigger a memory are subtle and (perhaps) subjective. Existing classification systems vary according to the research which motivated their inception. Their example informs the development of a useful data-base for FACEMAKER.

Harmon (1971) makes the point that it is less important (and more difficult) to determine whether and how humans use features in facial identification than it is of interest to determine a set of features which may be useful building a system to accurately identify faces. Procedures used in such a system may or may not be the same as those used in normal perceptual process. He devised a system of classification with twenty-one features in order to program a computer to sort a population of faces and to assist in the task of face identification. Successful sorting was illustrated by the computers' convincing choice of the most similar pair and the most dissimilar pair within a set of 256 faces. In order to isolate a given picture from a set Harmon outlines a "20-question" procedure; lists of features were made to describe target faces in order of decreasing extremeness (most

18.

extreme feature listed first), the computer then made a binary sort of
the complete set of faces for each feature listed. He found that the
search-set was reduced to less than 4% 99% of the time and concludes
that the system may be of value in criminal identification.

Another approach to classification is suggested by Ekman (1978).
He has isolated and catalogued the primary signs which can be reliably
communicated solely on the basis of appearance. He lists eighteen
types of information (identity, kin, race, gender, temperament, person-
ality, beauty, sexual attractiveness, intelligence, disease, emotion,
mood, emblems, adaptors, illustrators, regulators, age, previous
emotional life) which may be conveyed by varied combinations of four
types of signs: static (skeletal, skin pigmentation), slow (skin
changes, hair, etc.) and rapid (movements, color, etc.).

Some combination of Harmon's and Ekman's classification schemes
is probably a good place to start in order to establish a more elaborate
data-base. In the organization of the disc itself, we anticipated that
the general categories of age, sex and race would be helpful although
we freely mixed them to create new faces.

To restore something of the conversational nature of descrip-
tion, a data-base could be designed to gradually incorparate natural
language. A dynamic data-base could be implemented to take the features
determined by Harmon (see below) as a point of departure, but also
allow for the inclusion of subjective descriptors having specific mean-
ing for the individual user. Qualitative adjectives such as "handsome,"
or "homely" and cultural references such as "preppy," "looks like a

movie-star" or "typical nurd" carry considerable meaning and could be introduced into the data-base as they are used.

It is premature to determine an optimal presentation scheme as the many variables which may affect recognition have yet to be tested in this context. Exposure to too many faces or the wrong faces may be confusing, restriction to general types may be misleading, saturation could conflict with or even impair the mental image. Lacking clear evidence regarding the fragility of mental imagery, we make the assumption that memory may be bolstered by the act of "making a match." Ideally, a dynamic data-base would be synthronously modified by a concurrent background process capable of detecting favored attributes. Faces with similar qualities could be retrieved and displayed to the user to aid selection.

5.0 IDENTIDISC

When I was a child, my father was charged with taking my cousins
and me to the circus. One of the cousins, an identical twin,
got lost. Ever resourceful and with an unfailing sense of
humor, Dad marched the other twin to the LOST & FOUND and said,
"I need to find someone who looks just like this one."

5.1 Premise

IDENTIDISC is a prototype facial identification system based on
the techniques described above of pictorially synthesizing faces. The
composite face from IDENTIDISC is the result of averaging the intensity
values of several digitized photographs of whole faces. This is a
hypothesized replacement for feature-based systems such as "IdentiKit"
or "Photofit" which require composite assembly from a library of isolated
features.

A facial identification system based on averaging whole faces
as opposed to the compilation of individual features may be demonstrated
to have several distinct advantages:

— The first claim is that the system is better from the point of
view of input; the subject while choosing faces has only to make a
binary choice: "Yes"—it resembles the target (include it in the
average) or "No"—it doesn't look anything like the target (discard it).
A feature based system requires analysis ("Were his eyebrows close
together or far apart?") at every juncture. The user does not have to
consider segregated components; it may be simpler to respond positively
or negatively to a picture than it is to specify particular traits.

—   The second claim is that the system is better from the point of
view of output. The result of averaging photographis is a photograph
which looks like a person whereas the output of a feature-based system
is a caricature and requires synthesis on the part of the viewer. While
caricatures may be evocative in the case of a known personality, it is
not clear if they are adequate to represent an unfamiliar face for the
purposes of identification (see Perkins, 1975).

—   The third claim is that the slightly blurred equality of a compo-
site image is advantageous. It is inherently and automatically ambiguous;
it looks like a person, but is actually a scaled representation of the
memorable characteristics of a particular individual. Sharpness in the
final composite is a function of certainty; if the subject is "sure"
about a particular feature, that feature would be emphasized and distinct.
Similarly, blurriness would be indicative of uncertainty. Exclusing an
actual photograph of the target, a completely sharp (or sharpened) pic-
ture could be misleadingly accurate; contraindication is avoided in
favor of informed ambiguity.

—   Included on the IDENTIDISC are several thousand stereo photo-
graphs (see below for a description of implementation). Although we
have been long accustomed to two-dimensional representation of (at
least) three-dimensional phenomena, faces are not flat and are charac-
terized as much by the space they occupy as by the relative position
of features in a planar configuration. The addition of volumetric in-
formation augments the image, literally, by another dimension.

—   The use of optical videodiscs interfaced with computer graphics
allows the possibility of a wide range of post-processing of the image.

It may be possible to generate multiple viewpoints, expression (see
EXPRESSIONMAKER) below for an elaboration of the addition of expres-
sion), and movement, all of which may contribute to the evocativeness
of the image.

5.2 Implementation

*(1) I collected photographic portraits of different persons,
all of whom had been photographed in the same aspect (say full
face), and under the same conditions of light and shade (say
with the light coming from the right side).*

*(2) I reduced their portraits photographically to the same
size, being guided as to scale by the distance between any two
convenient points of reference in the features; for example,
by the vertical distance between two parallel lines, one of
which passed through the middle of the pupils of the eyes and
the other between the lips.*

*(3) I superimposed the portraints like the successive leaves
of a book so that the features of each portrait lay as exactly
as the case admitted, in front of those of the one behind it,
eye in front of eye and mouth in front of mouth. This I did
by holding them successively to the light and adjusting them,
then by fastening each to the preceding one with a strip of
gummed paper along one of the edges. Thus I obtained a book,
each page of which contained a separate portrait, and all the
portraits lay exactly in front of one another.*

*(4) I fastened the book against the wall in such a way that
I could turn over the pages in succession, leaving in turn
each portrait flat and fully exposed.*

*(5) I focused my camera on the book, fixed it firmly, and put
a sensitive plate inside it.*

*(6) I began photographing, taking one page after the other in
succession without moving the camera, but putting on the cap
whilst I was turning over the pages, so that an image of
each of the portraits in succession was thrown on the same
part of the sensitised plate.*
*   —Galton, 1879.*

5.2.1 Production and Contents of IdentiDisc

IDENTIDISC is a unique application of optical videodisc technology. A century after Galton painstakingly made a "book" of faces, we've utilized single frame storage on optical videodiscs to make a similar "book without pages" (Negroponte, 1979). We've photographed several thousand faces and stored one face per frame which, like the pages of a book, can be accessed randomly according to the particular interests of the user. If the disc is interfaced with a computer, it is possible to include a data-base classification system, providing an "index" of features to aid in constructing particular faces. If the data-base were to be embodied in the organization of the disc itself, i.e., if the images were carefully ordered at the time of mastering, a computer could also control a camera. As we are averaging computationally instead of optically, the computer is used to search for the designated faces and perform the averaging process.

At the time of this writing there are three generations of IDENTIDISC: the first was a series of faces taken by Bob Mohl as part of the Movie Map in 1979. He photographed nine hundred citizens of Aspen, Colorado, for inclusion on the Aspen Movie Map—all frontal views and registered at the eyes. Two years later this group of faces was used to attempt a composite as a preliminary study for IDENTIDISC. The result was MrAspen (see Figure 2), a composite of six faces chosen for their similarity to another, arbitrarily chosen face on the disc. On the strength of that, we decided to produce a disc specifically for making faces. FACEMAKER-II is the basis of much of the work described below. FACEMAKER-III is the latest version and includes all
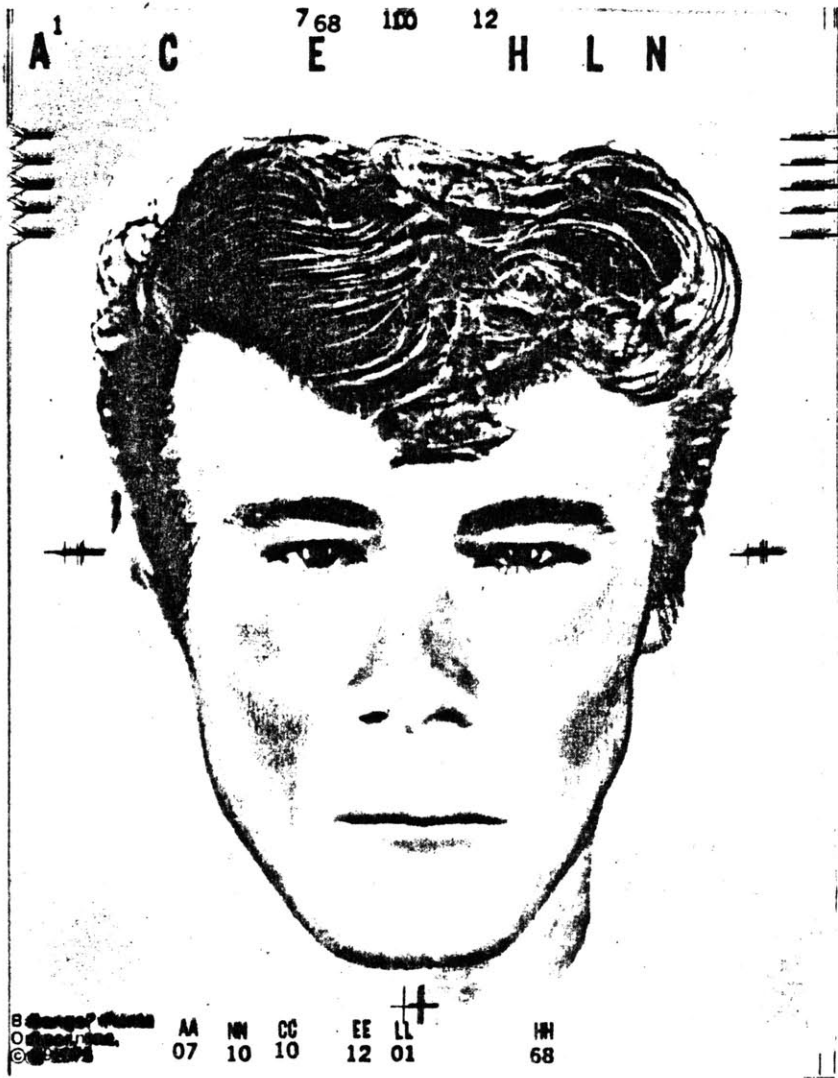
24.

figure 2a.   composite from IdentiKit

figure 2b. MrAspen - composite from IdentiDisc

of FACEMAKER-II plus substantially more faces, with an emphasis on two thousand stereo views.

This work is exploratory so almost every parameter shows experimentation; photographic and registration procedures, viewpoint, and classification. Also included on the discs is a variety of historical data; studies of faces and facial expression from the fields of psychology, physiology and art. (See appendix for index to the material included on the disc.) Following is an account of the production of FACEMAKER-II and III and a discussion of some of the decisions we made while working.

We followed roughly the same procedure as that outlined by Galton. We photographed several thousand people with a mobile 35 mm camera rig—our major consideration was to obtain a set of portraits that were consistent with respect to size, illumination, direction, and position within the frame. The pictures required post-registration which was done optically as the slides were transferred to half-frame 35 mm movie film for eventual transfer to video for the disc mastering process.

Registration was the most difficult problem. We chose to register the faces on the eyes for several reasons: the pupils of the eyes, and more specifically, the reflections of light on the eyeballs, provided an exact point on the face for post-registration. Maintaining a constant interpupilary horizontal axis also allowed us to correct for tilt. The consequence of keeping the inter-ocular distance constant is that the size of the head varies; the alternative would be to make the head size constant and sacrifice the lineup of the individual features. Another

consequence of using the eyes for registration is that the eyes become the most distinct portion of the averaged image. The potential disadvantage of this as a distraction is outweighed by the visual effect; within a slightly blurry composite face, sharp eyes pull the disparate images together. We tend to overlook shadowy multiple ears or a muddy hairline if the eyes are clear whereas a face with several eyes is an unusually distressing image.

To maintain constant illumination and take advantage of very precise reflections on the eyeball for registration, we used a flash mounted slightly off to one side. For the most part, we were shooting frontal views and put cross-hairs in the viewfinder in an attempt to frame each face consistently.

There are six degrees of freedom of rotation for the face within the frame and we could correct for four of them: tilting and panning the camera on a tripod compensated for horizontal (x) and vertical (y) displacement and tilt (rotation about z). We used a zoom lens to adjust for apparent distance (z) in order to maintain a constant inter-ocular measurement. We couldn't correct for nodding (rotation about x) or looking to the side (rotation about y). Initially, we simply asked people to look straight into the lens which wasn't really adequate. The only other solution would be to build a physical head registration mount to hold subjects' heads as we photographed them—an idea we discarded as being too intrusive. For the Aspen series of faces Bob experimented with a system which required the subjects to position themselves by shifting until they could view a light at the end of two tubes with both eyes. This was discarded as it required too much

practice and time on the part of the viewer. The disc includes a certain number of faces with slight degree of rotation; we chose not to reject them for the sake of quantity and diversity.

The majority of the faces on the disc are frontal views. Also included are double views (a frontal view and a profile), triple views (frontal, three-quarter and profile), full rotations and several thousand stereo views. The doubles and triples were taken simultaneously by using mirrors, the rotations were done successively and the stereo views were taken with a double-camera rig (see illustration).

Post-registration was done photographically with an optical printer. Each slide was manually re-photographed mounted in a precision stand which could be adjusted to exactly match the two points of light reflected in the eye from the flash with fiducial marks in the viewfinder.

In order to produce a 3-D image on a videodisc the right and left photographs are interplaced onto alternate fields in the disc mastering process. To view the images there are PLZT glasses; piezo-ceramic lenses are synchronized with the video signal so that each eye alternatively sees just one field—an electronic version of the stereo-scope used at the turn of the century.

Clearly we wanted as wide a range of faces as possible. Many of the photographs were taken at M.I.T., several hundred are from Aspen. We also photographed people at elementary schools, YMCA, and Lodge Beano (bingo) games, a party at a Rock Club in New York, and a major photo trade show while maintaining a set-up at the lab for visitors and experiments with transformations and expressions.
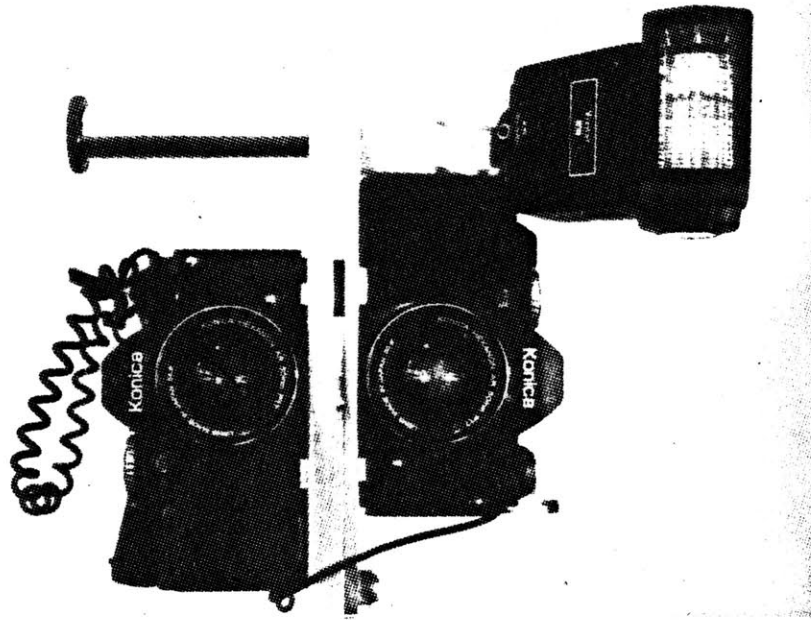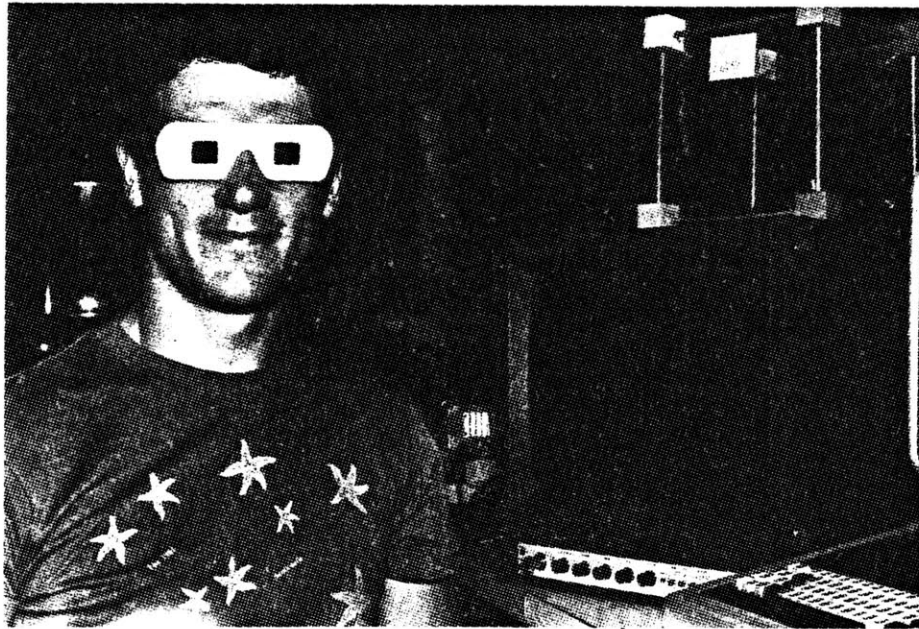
figure 3.   stereo camera rig

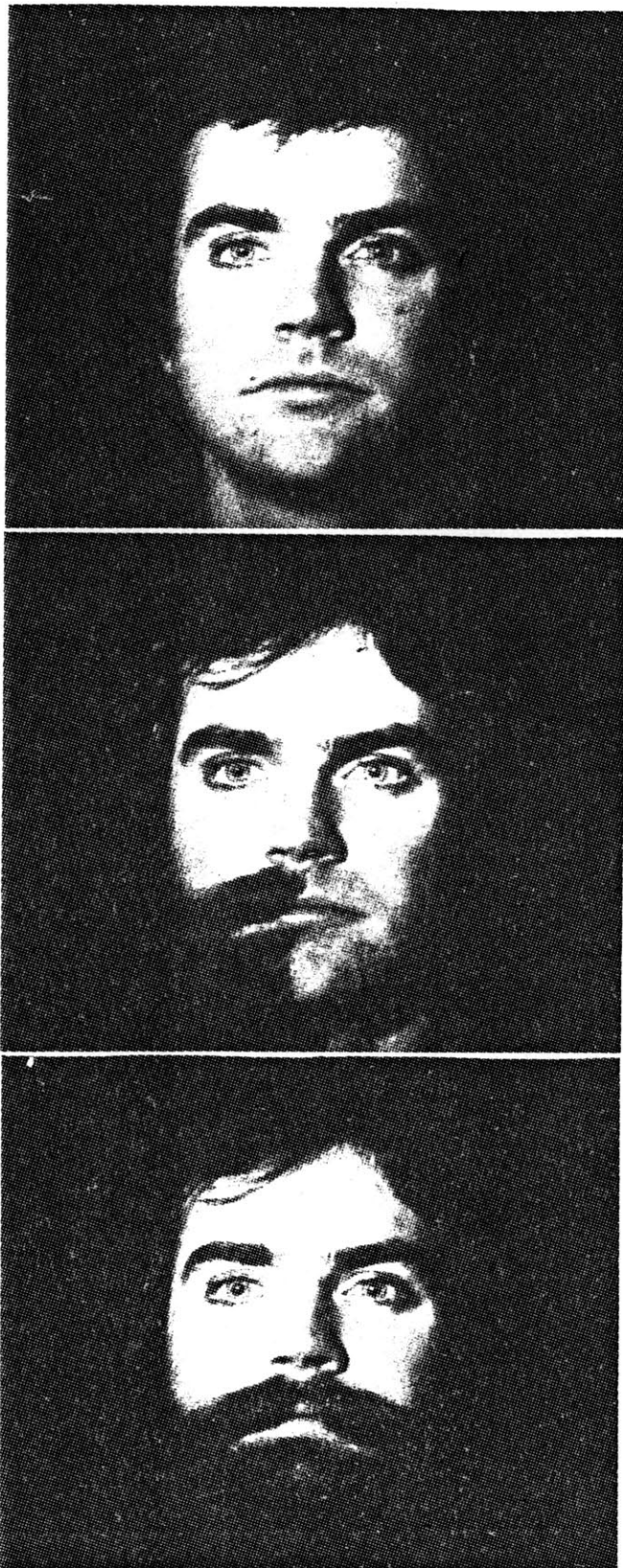figure 4. piezo-ceramic viewing glasses

figure 5a. doubles, triples

31.

figure 5b. transformations
32.

A complete disclog appears in the Appendix, a general census of the population follows; there is a predominance of faces between the ages of twenty and thirty-five; roughly two thousand of them. Children, teen-agers, ages thirty-five to forty-five and over forty-five are represented by about five hundred faces each. Approximately fifty-five percent of the faces are men, the remaining forty-five percent are women. Eighty percent of the population is White, the remaining twenty percent include Orientals, Indians, and Blacks. These figures are a reflection of our opportunities to photograph and the sites we chose to photograph and are not necessarily representative of an ideal sample for this application.

6.0 Facemaking System

An optical videodisc player, television monitor and a camera
capable of recording a video image is all that is needed in order to
electronically reproduce the setup used by Galton and Psotka. In use,
images are sequentially retrieved from the disc, accepted or rejected
by the viewer, and photographically averaged by the recording camera.
The immediate advantages of an electronic system are that it allows for
publication and access to a substantially greater number of images.
This configuration, without a classification system, does not include
a computer.

Classification of general facial types and physical groupings
is possible by the organization of the videodisc itself. The faces on
FACEMAKER-III are organized by age, sex and race. The space on the disc
allows for multiple groupings; many of the same faces are grouped else-
where randomly. Supplied with a categorical "index," the viewer need
peruse only the relevant sections of the disc. Utilizing the disc
itself as an ordering mechanism, while labor intensive at the time of
disc production, is optimal as it allows for extremely simple image
recovery.

The actual research environment included a minicomputer, a
digitizing framestore, and a second disc player in addition to the con-
figuration described above. The addition of a second disc player
permits a background process to average the faces as they are selected—
eliminating a delay after each choice. Two copies of the FACEMAKER

videodisc are played on two computer-controlled players. One disc is designed for search and selection of faces while the other is the one used for digitization. This also allows for the juxtaposition of non-contiguous images for comparison and reference.

The advantage of computationally producing the composite with a digitizing framestore is that the image is flexible and dynamically alterable; intermediate stages of the composite can be displayed as they are produced, parts of the face can be rejected, emphasized or animated. A minicomputer, as well as controlling the framestore and disc players, can contain and control a classification data-base greatly improving the ease of the selection system.

These technical improvements provide a degree of immediacy and interactiveness ot the system. While the direct utility of these user interface issues of facial recognition may at first seem orthogonal to the basic task, and unnecessary when the user is a highly motivated victim of a recent crime, they reflect attitudes to system design that are critical to the general success of this and similar computational systems (see Negroponte, Lippman, etc.). As a system gets more complex technically it should get more simple to use; the overall function must be apparent and easy to learn, the essential tools must be obvious and accessible, and the feedback should be continuous and informative. This work aims to satisfy the final goal of obtaining a likeness by being attentive to the visual parameters as well as to the creative procedure.

7.0 Facemaking Programs

The "first order" software environment was used as a basic feasibility test and as a means of generating many test composites. It employs the computational hardware environment described above and was designed to be simple and supportive so that any unmotivated member of the research laboratory would participate in the investigative process by synthesizing images of faces.

The object is for subjects to approximate a "target" (either from memory or for purposes of demonstration any face arbitrarily chosen from the disc) by perusing through the registered faces on the disc and designating those that resembled the target. The software is designed to enable the user to rapidly locate a promising neighborhood of faces, easily reference the entire library for the purposes of comparison or revision and continually assess the emerging composite.

Three main procedures direct the facemaking scenario: "FACESPACE" provides an animated classification scheme to introduce the available images on the disc for pre-selection, "PERUSE" controls the disc players, guiding the user through a selection of faces while signaling a concurrent background process to perform the averaging on the designated faces. Several programs exist for the manipulation of data files for reference and revision. "EMPHASIZE" is a system for adding expression or emphasizing a feature by weighting the average within a face (see below,

EXPRESSIONMAKER). A complete catalogue of programs is included in the Appendix.

The present organization of the disc is based on apparent resemblance, with the general population growing progressively older. Position according to age was made on the basis of appearance only; within an age group the faces are sequenced roughly according to varying categories of physical likeness; face shape, hair color and style, expression. There is a section of faces with glasses, hats, wigs, and make-up, another with families which showed obvious resemblance. [Frames 5866-6003.]

"FACESPACE" attempts to improve the presentation of a large data-base of faces. Several hundred representative faces from the disc form a grid with age forming one axis (x) and sex and race forming another (y). The user can "travel" (pan) across FACESPACE and choose a portion of the disc to view. A third axis (z) represents the additional information on the disc "behind" each face. This could be a repetoire of expressions or movement, or it could represent an extensive data-base of facial types.
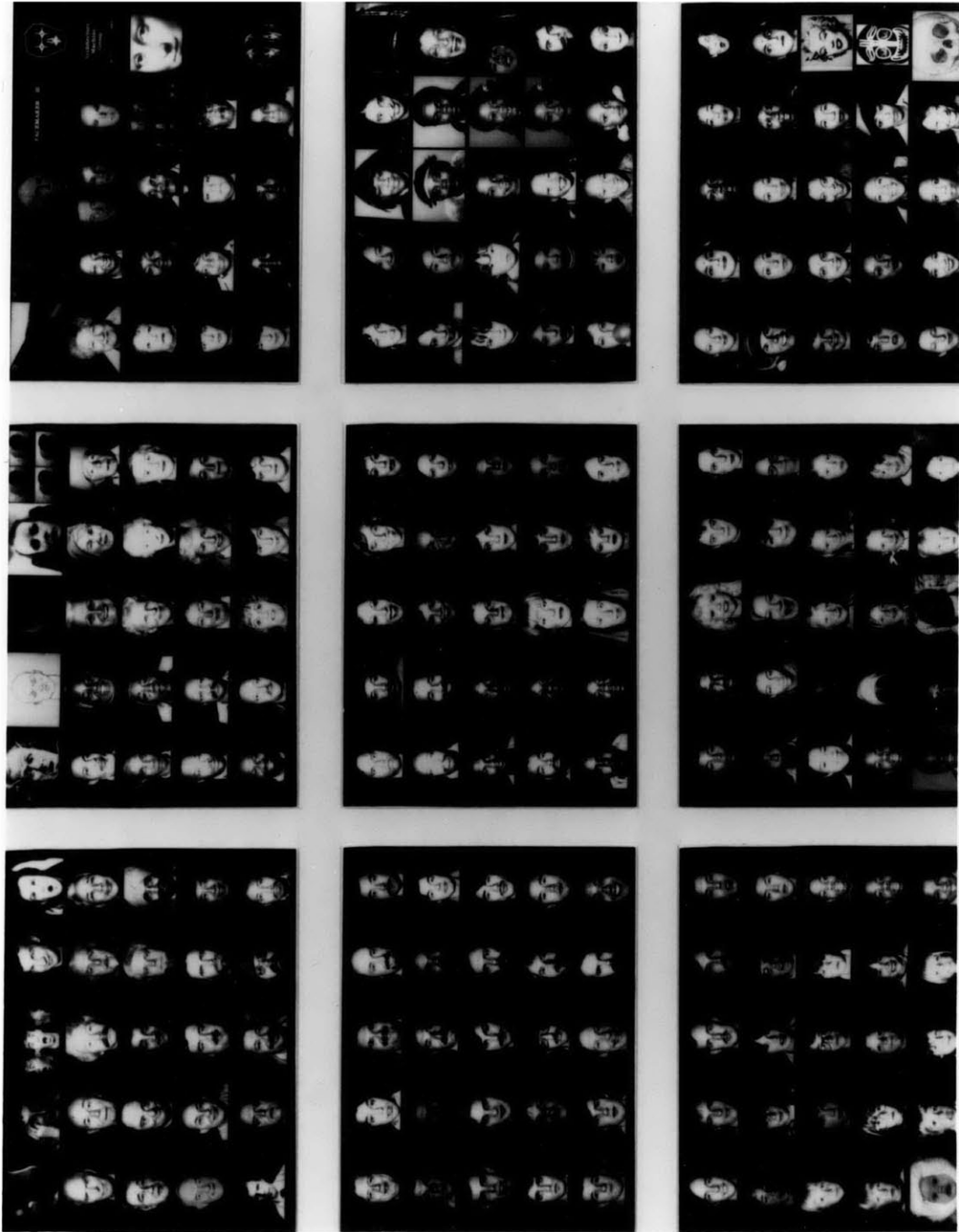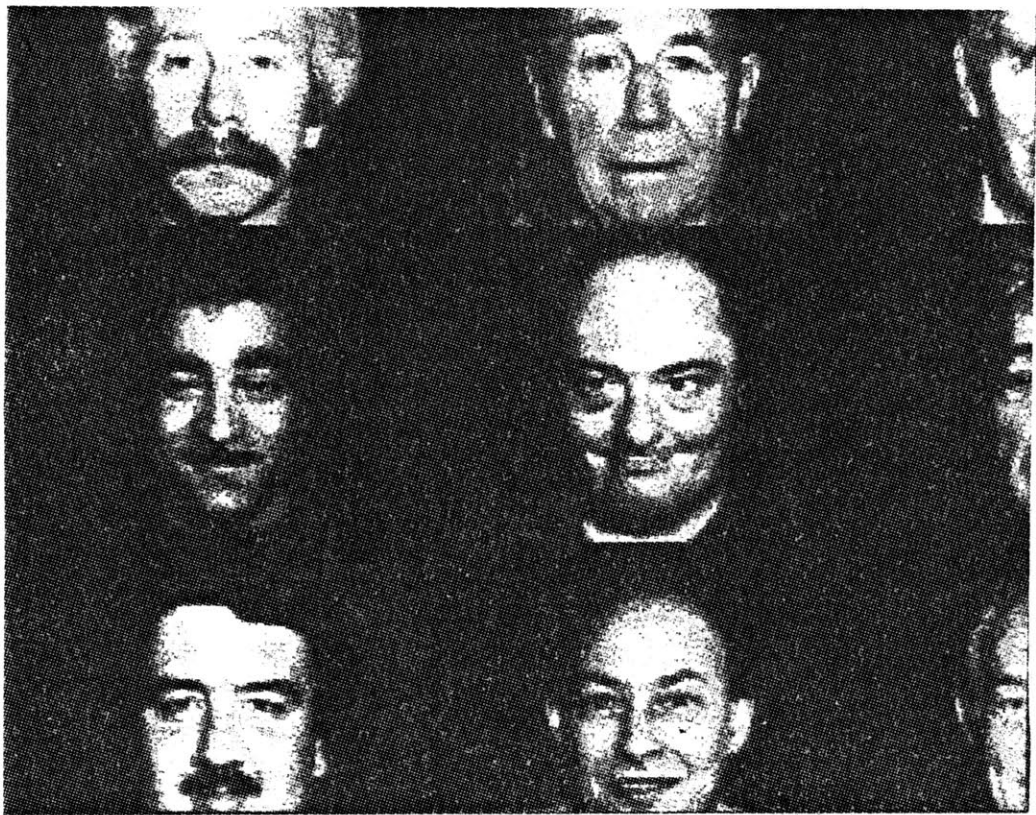
figure 6.   FACESPACE

38.

figure 7.   FACESPACE (detail)

39.

8.0 Examples of Composite Faces

Figures 8 through 10 are examples of facial composites made with the IdentiDisc System. Figure 9 shows a progression of an evolving averaged face. The insert in this series is the target face. The collages in Figures 9 through 10 illustrate the components of each synthesized face. The two larger faces are the target and the final composite. The smaller faces below are the individual faces which were used to make the composite. Figure 10 is an example of using a paint system to retouch the final image. In this case, clean-shaven faces were selected to best approximate the overall shape and features of the target and the image was later retouched to add a moustache and retouch the hairline.

It is difficult to evaluate the effectiveness of the composites as aids to identification. There are no objective criteria for judging resemblance among faces; similarly, possession of a photographic like-ness does not guarantee subsequent recognition. Among the most convinc-ing composites were those done in stereo. A three-dimensional image conveys the volumetric information lacking in a photograph. Faces are, after all, three-dimensional; the addition of depth contributes sub-stantially to the evocativeness of image as a likeness.

FACEMAKER is an effective system for synthesizing images of faces, whether from memory, model or imagination. Computational averag-ing techniques for manipulating photographs allow for wide experimentation
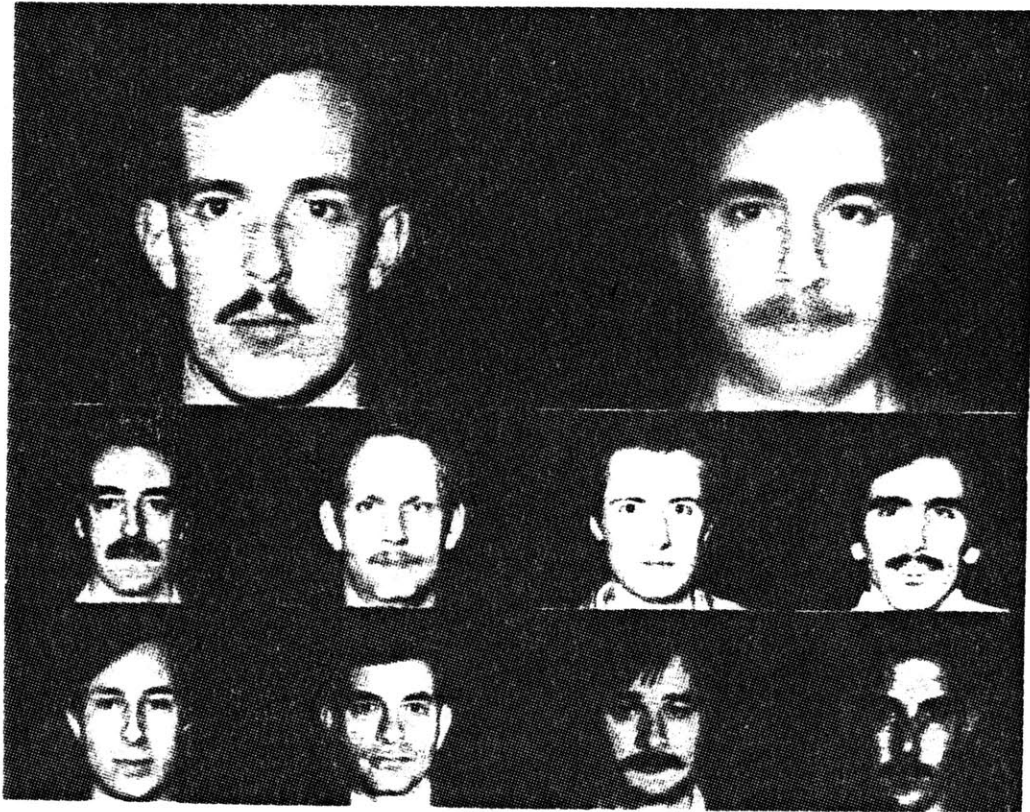
figure 8.
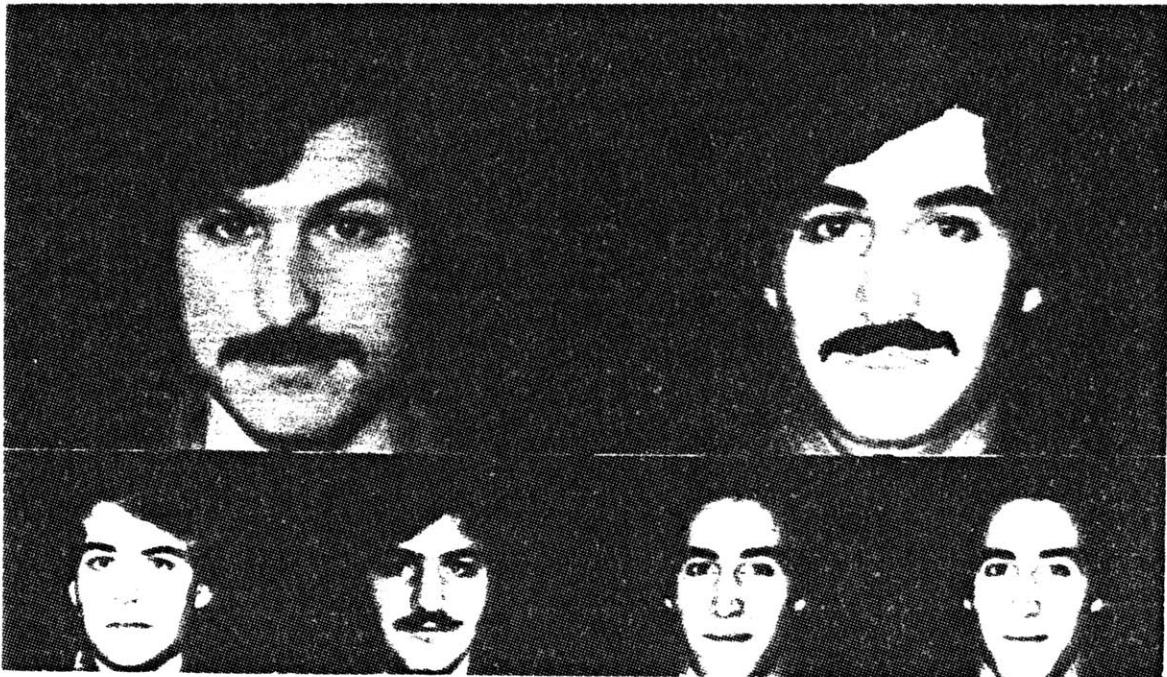
41.

figure 9.

figure 10.

variety and flexibility. EXPRESSIONMAKER will discuss the addition of expression and motion to a composite face.

II. EXPRESSIONMAKER

EXPRESSIONMAKER is dedicated to the first
machine that can make a gesture.

*...when suddenly there came into view a countenance—a
countenance which at once arrested and absorbed my whole
attention, on account of the absolute idiosyncrasy of its
expression. Anything even remotely resembling that expres-
sion I had never seen before. As I endeavored, during the
brief minute of my original survey to form some analysis of
the meaning conveyed, there arose confusedly and paradoxically
within my mind, the ideas of vast and mental power, of
caution, of peruriousness, of avarice, of coolness, of malice,
of bloodthirstiness, of triumph, of merriment, of excessive
terror, or intense—of supreme despair. I felt singularly
aroused, startled, fascinated. "How wild a history," I
said to myself, "is written within that bosom."*
        —Edgar Allen Poe,
        The Man of the Crowd

9.0 Objective

EXPRESSIONMAKER is a method of synthesizing facial expressions
utilizing the random access capability of computergraphic animation.
Expression sequences (paths and cycles through photographs stored on
disc) are blended onto the synthesized faces described above to make
"anyface" make any face.

The major task is to distill a finite set of photographs which
will generate a wide variety of facial expression. Within this "face-
deck" the expressive information is contained in the space between the
frames, their juxtaposition in time. Included in this section is a
survey of research on the topic of facial expression from the fields of
psychology, ethnology, drama, and art.

Expression may be characteristic as are physical features for the
purpose of recognition. In the context of the IdentiDisc project

EXPRESSIONMAKER would be used to more closely approximate the image of a face in memory; the addition of just the "right" smile or tic could be enough to make a convincing image. A serious obstacle to efforts to advance machine recognition of faces are the drastic distortions caused by facial expressions (Bledsoe, 1964). The systematic study and illustration of facial expression should contribute to these efforts as well as to ongoing psychological research. Additional applications in the field of communication include bandwidth compression for teleconferening and improved telecommunication for the deaf. Techniques to simulate facial expression with computergraphics find applications for animators working in entertainment, portraiture, and theater.

10.0 Facial Expression

*In the cradle of civilization men and women habitually sat
face-to-face. They sat on stone benches or on the ground...
An abrupt about face simply did not possess the impact it has
today. No one was deadly wounded by it and therefore it was
not used as a weapon.*
                                —*Walter Abish,*
                                  "Read Only Memory"

Despite the lack of any cohesive theory of facial expression,
the human face is of undisputed significance as the oldest form of
public display. Over the ages men have scrutinized their partners'
(or opponents') faces for indications of their emotional state in order
to gauge their next move. Early man recognized the potency of facial
expressions as evidenced by the wide use of masks and make-up to embody
powerful ideas in ancient ceremonial and theatrical traditions.

Mammals are the only animals which can be said to possess the
capability of facial expression. The origins of the human face are
found in the muscles of the neck region of the lower vertebrates. These
muscles gradually evolved until they began to group around the sensory
organs and gain mobility. The frozen masks of the invertebrates
became supple and mobile. Primary groupings were around the ear, eye,
snout and mouth. Gradually, the face came to indicate identity as well
as expression in man.

The following section presents a brief survey of the study of
facial expression and its status in current literature. Even a descrip-
tion of the field is difficult; the generally accepted term is "facial

expression of emotion," however, some avoid the phrase contending that it implies a particular theory of facial expression. Ekman and Friesen (1970) suggest "face and emotion" or "facial behavior" and Izard (1969) uses "facial patterning." It is almost a neglected field being researched at cross-purposes resulting in widely divergent interpretations and opinions. Emotion and science seem to be a particularly volatile combination and the observation of emotional display nearly defies the scientific method.

The following historical references informed the construction of a system to portray facial expression. Much of the research in the field took place in the context of a debate regarding the cause and origin of facial expression. Investigators have tried to determine to what extent facial expression can be considered innate, and thus, a universal phenomena which can effectively communicate across cultures or learned behavior, and thus, culture specific. The research ranges from broad transcultural surveys to inquiries into the causal relationship of neural-muscular activity to emotional expression. EXPRESSIONMAKER does not require a conclusive answer to these questions in order to pictorially generate facial expressions.

10.1 Theories of Facial Expression

Refuse to express a passion and it dies...
                    —*William James (1980)*

The free expression by outward signs of an emotion intensifies it. On the other hand, the repression, as far as possible of outward signs, softens our emotions...Even the simulation of an emotion tends to arouse it in our minds.
                    —*Charles Darwin (1872)*

There are several approaches to the study of facial expression. In attempting to explain its cause and development, some stress its functional importance within a biological and(or) evolutionary context. Others (Piderit, Wundt, Peiper) postulate facial expression to be a direct consequence of sensory impressions; i.e., disgust is characterized by the motions made in response to a bad taste, etc. Yet another group (Darwin, Andrew) stress biological inheritance with some attention to neuro-muscular activity. Some hold that electro-mechanical laws of the nervous system are entirely responsible for characteristic muscle patterns which came to be associated with specific emotions (Spencer, Dumas). Current theories tend to draw from all the forementioned traditional hypotheses. In reassessing the theoretical groundwork of facial expression, contemporary researchers (Ekman et.al., Izard, Tomkins) have contributed significantly towards building a unified theory.

An analysis of the anatomical evolution of the face is the subject of a major work by Huber (1931). Huber traces the development of the facial muscles from the lower vertebrates to man and from infancy to childhood. He dates the first anatomical studies of the facial musculature to the 16th century, but notes that little progress was made until the 19th century when anatomists, notably Bell (1816) attempted to integrate the action of the nervous system. The work includes a detailed analysis of primate facial musculature as well as a study of comparative racial anatomy.

The classic theoretical work in the field is Darwin's <u>The Expression of Emotion in Men and Animals</u>. Darwin's contribution was to include facial expression within the scheme of evolutionary theory.

Written after The Origin of the Species and The Descent of Man, The
Expression of Emotions in Man and Animals presents Darwin's position
that it is evolution that determines our emotional expression. Darwin
proposes that the principles which correlate a given expression with an
emotion evolved with the species and facial expression is therefore an
innate phenomenon. His work is based upon anecdotal observation fur-
nished by colleagues in response to a questionnaire regarding the appear-
ance of certain gestures occurring in emotional situations. Darwin
distributed his questionnaire among missionaries and medical men living
in isolated cultures including Australia, India, Africa, India and
Malaysia. Out of these reports he presents a theory which presents three
principles of emotional expression: the principle of servicable asso-
ciated habits, antithesis, direct action of the nervous system.

Darwin's three principles emphasize the role of biological
heredity of habitual responses. The first states that voluntary actions,
initially biologically "servicable," become habitual responses which are
associated with the original causal state and are performed regardless of
their subsequent utility. By antithesis, he means that the effects of
responses to specific states are "reversed" by arousal of the opposite
state (e.g., relaxed, friendly gestures are the antithesis of a fighting
posture). The third incorporates the influence of the neuro-muscular
system.

The anecdotal nature of the correspondence on which he based his
arguments was the source of much later criticism of Darwin's work. Al-
though the book was a best-seller in its day, his position has been
widely debated in the last century. Investigators from diverse fields

have studied infants, the blind, the insane, primates, and people from across many cultures attempting to establish evidence for universality. The transcultural work of Izard (1971) and Ekman (1973) offers evidence which supports Darwin's views. Ekman concludes that while many facial patterns are recognizable across cultures, expression may differ in two major ways: first, the cause of a particular emotion often varies across cultures and individuals. The same stimuli might elicit happiness for one and fear for the other. Second, rules for showing emotion may vary considerably—one may feel a specific emotion but be restrained by cultural convention to hide or mask it.

In the last decade, Izard and Ekman have advanced and expanded research on the human face. Izard (1971), following the work of Tomkins (1962), theorizes that "facial patterning" is a necessary component of emotional experience, thus attributing increased significance to facial expression. Ekman contributed a comprehensive classification system for facial movement (see below). FACS (Facial Action Coding System) isolates and catalogues the primary signs which can be accurately communicated solely on the basis of appearance (see above FACEMAKER, 4.0 Facial Classification).

10.2 Classification

> No one has ever set about charting the exact boundary lines of
> a shiver, the territories of a kiss, the homeland of desire.
> Rough surveys are all that man has mapped out of his personal
> experience. Perhaps someday some learned anatomists will parcel
> the human body among themselves to study the meandering
> itineraries of excitements; they will find this project as
> worthy of their full attention as any other. They will
> publish their conclusions in atlases...
>   —Louis Aragon,
>   Nightwalker

Whatever the origin, the cause or the function of facial expressions, their classification is an awesome task. In part, the field is disunified due to the lack of a common system of classification. The small, extremely mobile facial muscles can articulate thousands of possible combinations. Many of the following systems developed as exhaustive catalogues or as tools to support particular psychological research. I needed to find a system from which I could derive a set of facial movements to test EXPRESSIONMAKER. Many of the series cited below illustrate the FACEMAKER-III videodisc; frame numbers are indicated within [ ].

One of the earliest recorded attempts to classify gesture has been attributed to the Babylonians, as evidenced in preserved fragments dating from Nebuchanezer. The Babylonians apparently gave involuntary facial movements significance as omens. The fragments, known as "Twitching Books," lists facial movements and fortunes; the system is far-fetched from a modern perspective, but represents an early investigation into the meaning and consequences of facial expression.

Before the mid-nineteenth century most of the recorded speculations on meaning conveyed by the face were concerned with static features as opposed to dynamic expression. Physiognomists produced exhaustive catalogues of facial types and largely ignored the changes which took place on an individual face. Lavater (1806) makes an exception for Henry IV of France and includes a plate which illustrates that the royal countenance can maintain an aura of dignity through eight passions [frames 11422-11429].

Contemporary approaches to classify facial expression reflect three principle approaches. Interest in the causes and origins of facial expression has motivated the development of anatomically based notation systems—exhaustive catalogues of the correlations between the visible effects of muscular actions and specific emotions. Psychological research of the perception of facial expression has prompted the production of demonstrational models which are collections of drawings or photographs judged to represent a broad range of human expression. Ethnologists and anthropologists have devised structures to record field observations of human behavior.

The techniques for producing "standard" representations for different emotions vary considerably, ranging from simple observation to electrically stimulating specific muscles. Other methods have included traditional acting techniques, simulation (i.e., shooting a gun to obtain a startled expression, etc.), hypnotism, and hidden camera systems. Many of the investigators posed themselves as models: Felekey (1914), Ruckmick (1921), Hjorstjo (1969), and Ekman (1970). Others used actors, co-workers, students, and even prisoners as subjects.

The investigator's method of categorizing emotions determines the shape of any classification system. There is no standard taxonomy of emotion and the range of classification systems reflect the categorization chosen by the investigators. While some schemes have over a hundred different emotions, most have chosen to work with a hierarchy of six to ten "primary" emotions. None offer solutions for integrating dynamic sequences of emotional expression. In an animation system, the factor "which-picture-might-come-next" is as significant as the

"representative-image-of-expression." A potential approach is suggested by the work of Scholsberg (1954) who attempted to map emotions along dimensional axis forming an "expression space" roughly corresponding systems to graph color.

10.21 Anatomical Classification Systems

The first anatomically comprehensive classification of facial expression was attempted by the French physiologist Duchenne. He sought to isolate the effect of each muscle by "galvinization"—applying electrical current to individual muscles and photographing the resulting musculature contraction. He performed this work with French inmates (some accounts claim that the procedure was painless) and produced an extraordinary set of photographic plates which he publised in 1862. The work, "Mechanism de la Physionomie Humaine" (Duchenne, 1876, first ed., 1862), is a catalogue of facial expression classified according to muscle groups responsible for the appearance of a given emotion. It is organized by emotion. A series of places illustrates each chapter: the muscle of sorrow, grief, pain, happiness, etc. [frames 12355-12426].

For obvious reasons, Duchenne's efforts have not been duplicated so its unique findings are often cited in studies of facial expression. The photographs figures prominently in Darwin's work on expression. The contortions produced by electrical stimulation quite successfully mimics the appearance of different expressions but there are subtle differences from natural expression. Darwin addressed this issue without making any conclusions:
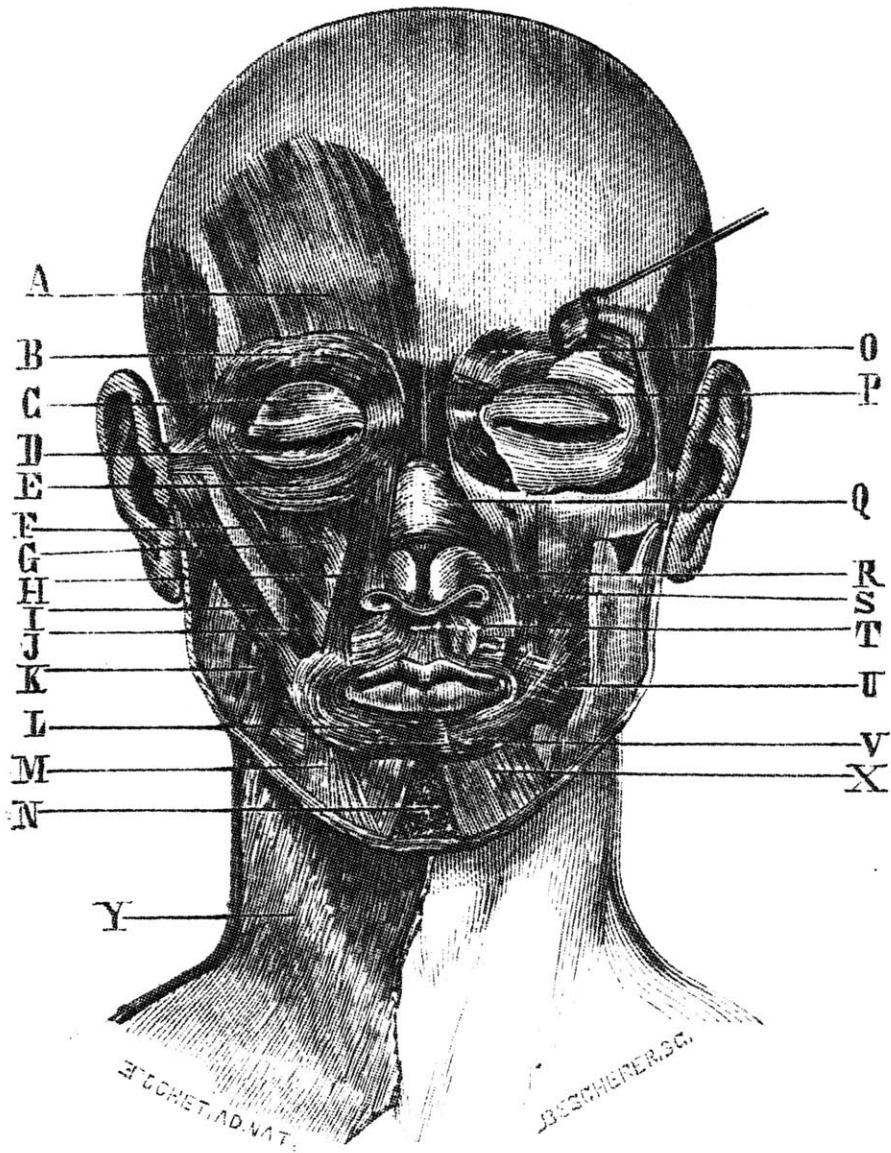
figure 11.   Duchenne(1862)

57.

FIG. 1.

figure 12. Duchenne(1862)

58.

```
A.    Frontal, muscle de l'attention.
B.    Orbiculaire palpebral superieur,
         muscle de la reflexion.
C,D.  Palpebraux superieur et inferieur,
         muscle du mepris et complementaire du pleurer.
E.    Orbiculaire palpebral inferieur,
         muscle de la bienveillance et complementaire
         de la joie franche.
F.    Petit zygomatique,
         muscle du pleurer modere et du chagrin.
G.    Elevateur propre de la levre superieure,
         muscle du pleurer.
H.    Elevateur commun de la levre superieur et
      de l'aile du nez,
         muscle du pleurnicher.
I.    Grand sygomatique,
         muscle de la joie.
K.    Masseter.
L.    Orbiculaire des levres.
M.    Triangulaire des levres,
         muscle de la tristesse et complementaire des
         passions agressives.
N.    Houppe de menton.
O.    Sourcilier,
         muscle de la douleur.
P.    Pyramidal du nez,
         muscle de l'agression.
Q.    Transverse du nez,
         muscle de la lascivete, de la lubricite.
R.    Dilatateur des ailes du nez,
         muscle complementaire des expressions passionnees.
U.    Buccinateur,
         muscle de l'ironie.
V.    Fibres profondes de l'orbiculaire des levres se
      continuant avec le buccinateur.
X.    Carre du menton,
         muscle complementaire de l'ironie et des passions
         aggressives.
Y.    Peaucier,
         muscle de la frayer, de l'effroi et complementaire
         de la colere.


                    -Duchenne, Legende-Preparation anatomique
                       des muscles de la face (1862).
```

In the two photographs given by Duchenne of the same old
man, almost everyone recognized that the one represented
a true, and the other a false smile; but I have found it
very difficult to decide in what the whole amount of dif-
ference consists. (Darwin, 1872).

Of issue in the pictorial simulation of facial expression is
the degree to which artificially induced expression, while anatomically
"correct" appears to be genuine. An animation scheme based on an
anatomical model of the face may have to be combined with other insights
to convincingly evoke emotional expression.

A Swedish anatomist, Hjortsjo, made a systematic investigation
of facial expression. In this study, "Man's Face and Mimic Language"
(Hjortsjo, 1969), he formulates an "alphabet" of expression terming
muscular movements "mimic letters" and facial expressions "mimic words."
His system consists of twenty-four basic emotions, each illustrated in
a clear schematic drawing accompanied by a detailed analysis of the
muscular action.

The anatomically based system of psychologists Ekman and Friesen
is probably the most comprehensive analysis of facial movement patterns.
The Facial Action Coding System or FACS (Ekman et.al., 1978) is an
elaborate notation system designed to differentiate between all visually
distinguishable facial patterns by measuring movement. It is based on
units of muscular movement called action units (AUs) — each AU represent-
ing a combination of muscles responsible for a specific visually discern-
ible change in facial appearance. This system was adapted for a computer-
graphic model of the face in the work of Platt, et. al. (1981) and may
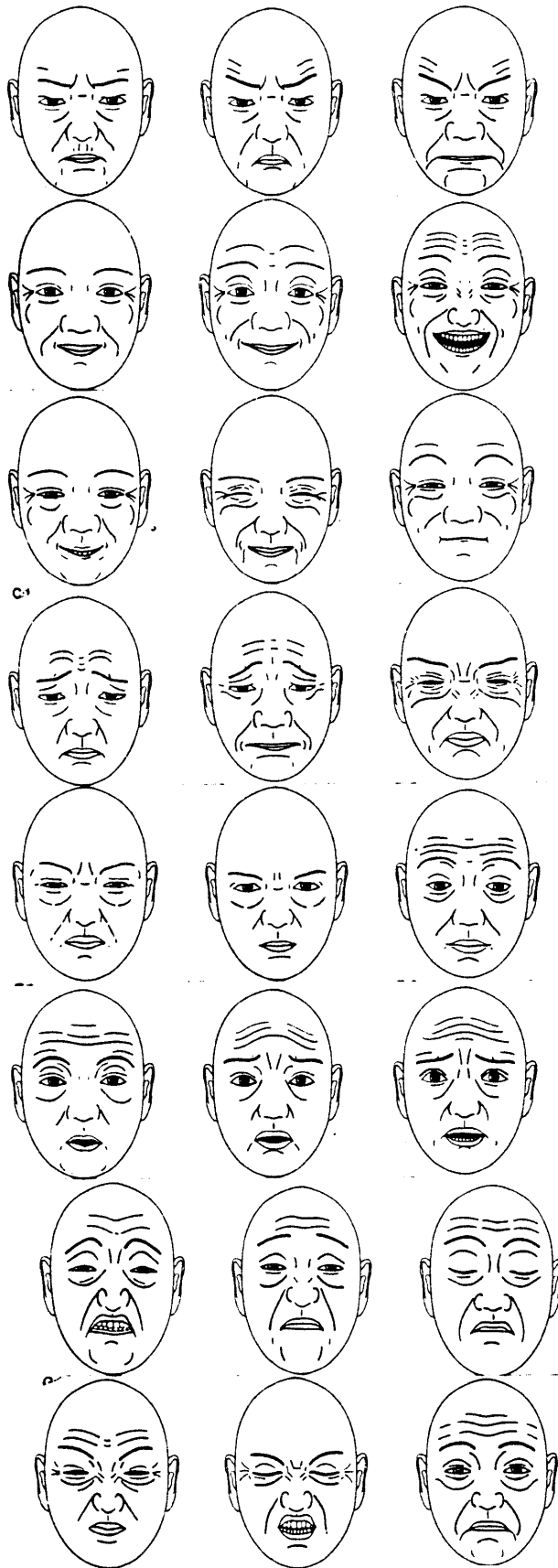well become the standard reference in the field. The correlation of

figure 13.   Hjortsjo(1969)
61.

particular AU configurations with emotional expression is part of a larger effort to produce a human facial "atlas."

10.22 Psychological Demonstration Models

The field of psychology has contributed several "demonstrational models" of facial expression. These collections of drawings or photographs represent human expression and are designed to demonstrate and test various theses in psychology; ability to judge facial expression, cultural concensus of portrayed emotional meaning, saliency of different features, etc. Photographic collections are valuable to this work as reference: "dynamic" systems—sets of features to be viewed in various combinations—are of special interest in that they attempted to derive a set of primitives to convey expression as well as a code for combining them.

One of the first dynamic models was Piderit's "Geometry of Expression" (Piderit, 1885), a series of drawings with interchangeable features which he suggests be transferred to plaster of Paris heads. [frames 11434-11485] Boring and Titchner (1923) adapted these drawings to make a sort of jigsaw puzzle which they termed an "articulated profile." The set includes nine mouths, five eyes, four brows, and two noses which can be combined into 360 combinations. Guilford and Wilke (cited in Ruckmick, 1936) produced a frontal version based on the work of Rudolph and Ruckmick. This includes six foreheads, nine eyes, four noses and twelve mouths. A coding system allows a user to produce a wide range of meaningful combinations out of the several thousand possibilities.

Dunlap (1927) produced a set of photographs for making (puzzle-like) composites to emphasize the relative importance of the eyes and mouth to the recognition of expression. He cut the photos in half horizonatlly and made different combinations of eyes and mouths to test for the appearance of different expressions. Originally intending to combine pictures discrete subject's features, he found that faces formed by combining features from different subjects could effectively portray expression. Dunlap makes the point that a still photograph is but a cross-section of any expression naturally occurring in time; the actual expression is a sequence of changes which motion pictures could best represent.

Frois-Wittman (1930) produced two collections portraying facial expression for his studies of the ability to judge emotional expression from features in isolation as well as in whole faces. The first were a series of 46 photographs of himself and the second were drawings (based on the photographs) sliced horizontally in thirds for producing composites. He termed the drawings a compromise between the Piderit Model and photographic images (citing technical difficulties of combining photographs). The drawings, showing a 3/4 view, have more detail than the Piderit schematics, and are coded according to muscular involvement in units terms MIs.

A number of photographic collections demonstrating facial expression exist. One of the earliest in this country was Feleckey's work which attempted a concensus in the judgment of the meaning conveyed in facial expressions. She posed for the photographs, the final set of eighty-six portrays diverse emotions [frames 11487-11608]. The

collections of Rudolph (1903), Ruckmick (1936), and Langfield (1918) are also notable.

Ekman's collection, "Facial Expressions of Emotion," (described in Ekman et.al., 1975) is a careful analysis of the appearance of emotional expression. It illustrates the appearance of six primary emotions (surprise, fear, anger, happiness, sadness, disgust) and "blends" formed by combining primary components (puzzle fashion) in composite photos. The pictures are divided following the contours of the brows, eyes, nose, and mouth. The systematic description of facial behavior for six primary emotions as they appear in four physical zones is concise, flexible and representative of diverse expression.

10.23 Recording by Ethnologists

Anthropological field work demands careful observation and accurate descriptions of human behavior including human facial patterning. In the last century, it has become common to record aspects of culture photographically. FACEMAKER-III includes a brief excerpt from "Deep Hearts," a film shot in 1980 by Robert Gardner. The subject of the scene is a "facial dance" performed by men of the Bororo People of Niger Republic [frames 15558-16437]. Film is the obvious medium for recording facial expression, however, there remains a need for verbal and graphical analytical tools.

Sociologist Erving Goffman has published several essays on human interaction, including a detailed analysis of face-to-face behavior and expression games. His work could provide the groundwork for building a structure to direct the choice of an appropriate expression within a given situation.

Birdwhistell (1970) has developed a symbolic notional system within his study of Kinesics to analyze the basic communicative elements of human body motion. He notes that, while motion is continuous and film can reveal subtleties of position, humans "select" portions from this range to define interaction. In his system, a "kine" is defined to be the "least particle of isolatable body-motion." The system has a linguistic structure; "kines" are modified by "allokines," and can be combined to form "kinemorphs," etc. "Kinemes" form the basic units of movement judged to have social meaning. Regional catalogues of "kinemes" have distilled the primitives of meaningful facial behavior into a manageable set:

> Physiologists have estimated that the facial musculature
> is such that over twenty thousand facial expressions are
> somatically possible. At the present stage of investiga-
> tion, we have been able to isolate thirty-two kinemes in
> the face and head area. (Birdwhistle, 1970).

There are ten kinemes involving the whole head (nods, lateral sweeps, tilts, and "cocks"), four of brow behavior (lifted, lowered, knit and single), four of lid closure (overopen, slit, closed and squeezed), four of the nose (wrinkled, compressed and flares), seven of the mouth (lip positions, snarl and open mouth), two of the chin (anterior and lateral thrust), and two of the cheeks ("puffed" and "sucked"). These units of facial movements are adequate to record a wide range of interactive facial behavior. Kinesic analysis may pro- vide a convenient formal structure for systems to animate facial expression.

11.0 Portrayal of Facial Expression

Five Ways a Man Can Smile:

One: A man can smile with the corners of his mouth turned up, shyly, in an expression that is more grin, really, than smile, and with his eyes slid over to the side watching your reaction to him.

Two: A man can smile full-face, openly, with teeth showing in an attitute of confidence, certain of your approval. His gaze then is level and winning.

Three: A man can smile coquettishly, eyes cast down, lips barely parted to reveal teeth. Are they pretty teeth?

Four: A man can smile a wistful smile if he is wise or old or has a knack for imitating wiseness or oldness or if he is an actor. This same smile, which is barely noticeable, can also accompany the memory of hurt.

Five: A man can smile quickly, passing through a grin into a sort of gasp, mouth widened, teeth showing, eyes wrinkled at the centers, gay.

—Manuel Puig

Many early inquiries and classification schemes for facial expression arose out of a concern for their effective portrayal in the arts. Anatomists (Camper, 1792 and Bell, 1844) wrote treatises on the appearance of facial expression for the painter; painters (LeBrun, 1667 and Topffer, 1845) compiled reference books for rendering "the design of passions" in portraits; and masks, marionettes, puppets, and even the ventriloquist's dummy were all carefully constructed with attention to facial details in order to convey particular sentiments. Theater has contributed numerous methods directing the actor to achieve effective communication on the stage.

The grand masters of portraiture were above all masters of the portrayal of facial expression. The art historian, Gombrich (1956),
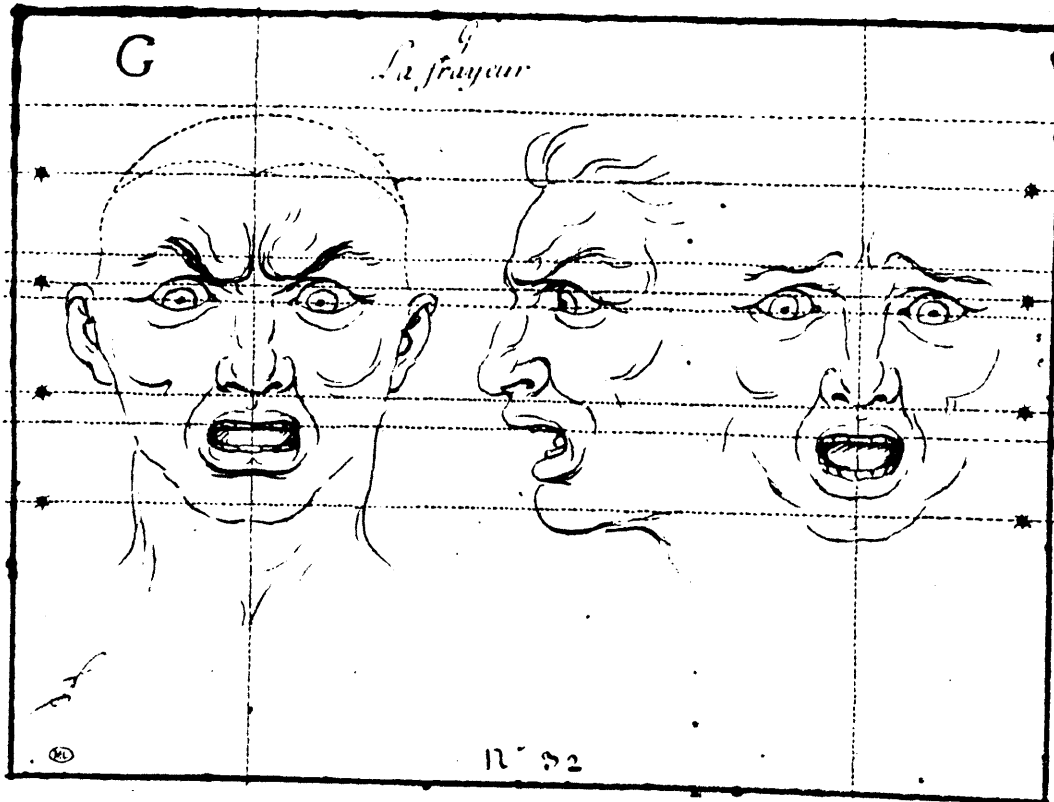
figure 14. Le Frayeur, LeBrun(1696)

points out that "mastery of physiognomic expression" is an almost neglected achievement in the annals of art history, never chronicled as such and taking a second place to the heralded achievements of perspective and technique. The perfect rendering of an expression is delicate and essential to achieving a successful likeness. This subtlety has been exploited in banknote designs—portraits are difficult to forge because it is hard to reproduce the nuance of the original expression. Included on the disc are a selection of portraits beginning with Mona Lisa [frames 11663-11929].

Certain classical theatrical traditions evolved a whole system of gestures coded as "ideograms"—particular gestures evolved to portray a specific idea, mood, or object. The movements from medieval Hindu theater are catalogued in the "Abhinaya-Darpanam: A manual of Gestures and Postures Used in Hindu Dance and Drama," an eleventh century Sanskrit text. Abhinaya is roughly translated from the Sanskrit as "histrionic art" and is defined as movements for suggesting rasa (sentiment) and bhava (mood). Gods, various relations and different castes are represented in gestures. The book delineates specific movements of the parts of the body including nine gestures of the head and eight gestures of the eyes and four gestures of the neck. Indian classical dance emphasizes the precise representation of passion [frames 6004-6018].

In an elaborate scheme, the method of DelSarte (1882) classifies the emotions for the purpose of oratory. An influential text, he methodically catalogues the different combinations of vocal expression and gesture within a strict formula based on three divisions: all posturing is either normal (neutral), eccentric (extroverted) or

figure 15.

69.

concentric (introverted). Grotowski (1968), drawing from classical and non-western traditions as well as from DelSarte, makes a plea for the actor to develop an intricate command of the facial muscles. His widely practiced system of exercises, Plastiques, include"exercises of the facial mask"—isolation exercises designed to permit the actor to construct masks created solely by the facial muscles [frames 11609-11612].

There are two very different approaches to the animation of facial expression. Cartoonists isolate the essential perceptual cues of facial expression by exaggerating characteristic distortions and motion. It is a process of distillation, anatomical accuracy is sacrificed for vitality. Disney pioneered the art of animating facial expression. His animators developed principles for animated emotion based on emphasizing rhythm and gross changes of shape. A computer model of exact facial features simulates the appearance of facial expressions. Platt and Badler (1981) and Parke (1972) contributed two particularly outstanding examples of simulated facial expression. However, while the accuracy of these models is impressive, they fail to effectively convey emotional expression.

12.0 Expressionmaker

Can you make it say, "Hello Breshnev"?
—Visitor to demonstration
of LIPSYNC

FACEMAKER and EXPRESSIONMAKER were first suggested in an Architecture Machine Group proposal, Transmission of Presence (Negroponte, Lippman & Bolt, 1980). The proposal, subtitled, "Advanced Machine Recognition and Image Processing Applied to Teleconferencing" described configurations of interactive media technology to improve telephone-linked teleconferencing with the addition of synthesized moving faces. The projects developed under the auspices of this proposal laid the groundwork for the concerns addressed in EXPRESSIONMAKER.

The objective of "Transmission of Presence" was to imitate a televised (video-linked) teleconference over phone lines. The challenge of transmitting "nuance"or at least an appropriate moving picture of a speaker's face without actually sending a picture was explored via the notion of "bandwidth compression"; getting more out of less. The technique is to use less information (that which can be transmitted over a phone line) more intelligently by taking advantage of images stored on the receiving end. These compressed video images are concatenated according to auto-correlation procedures to produce moving, synthetic faces. "Semantic Bandwidth Compression" encompassed several projects that implemented teleconferencing techniques by trading local intelligence for reduced bandwidth: LIPSYNC, the AUTO-DUBBER, TALKING HEADS, and two discs, SPEECHMAKER I, II, & III.

71.

LIPSYNC, originally implemented by the author and Howard Eglowstein, produces a talking picture with lips moving in syncrony with speech. The first version was driven by synthesized speech, a later version used pre-recorded processed speech and the current version responds to live processed speech. The AUTO-DUBBER, implemented by Susan Brennan and Steve Gano, directed "emotion sequences" stored on videodisc according to auditory signals and key-word cues of a transmitted voice. SPEECHMAKER I, II, & III are videodiscs storing images of registered talking heads in a wide range of positions and expressions for later retrieval in a proposed teleconferencing station.

SCANNER demonstrates the potential of interactive computergraphics to animate or simulate eye movement. Fifteen pictures portraying a range of eye positions were digitized and displayed so that the face would "scan" in a direction indicated by either a joystick or a headmounted eye-tracking device. The latter related to teleconferencing concerns; a relatively small amount of information concerning eye position could be transmitted over a phone line and used to concatenate images of the other end so that the picture mimicked the eye position of the sender, appearing to look where the sender was looking. Similarly, head position could be mimicked by LOOK, with the sender wearing a headmounted body tracking device.

At this stage, "transmission of presence" had effectively demonstrated "transmission of position;" movement of the head, lips, and eyes (presumably the most important features, pictorially, for face-to-face communication) could be successfully simulated despite low bandwidth requirements. The portrayal of facial expression using similar methods
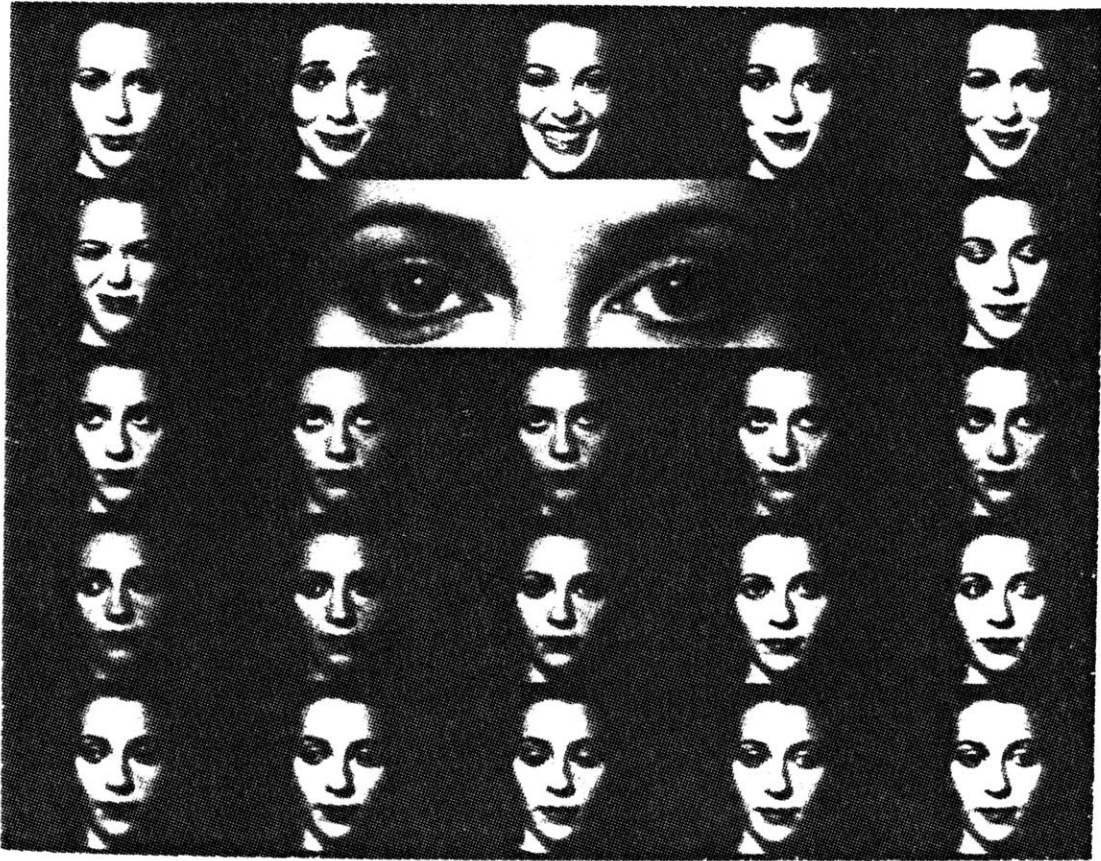
figure 16. LIPSYNC
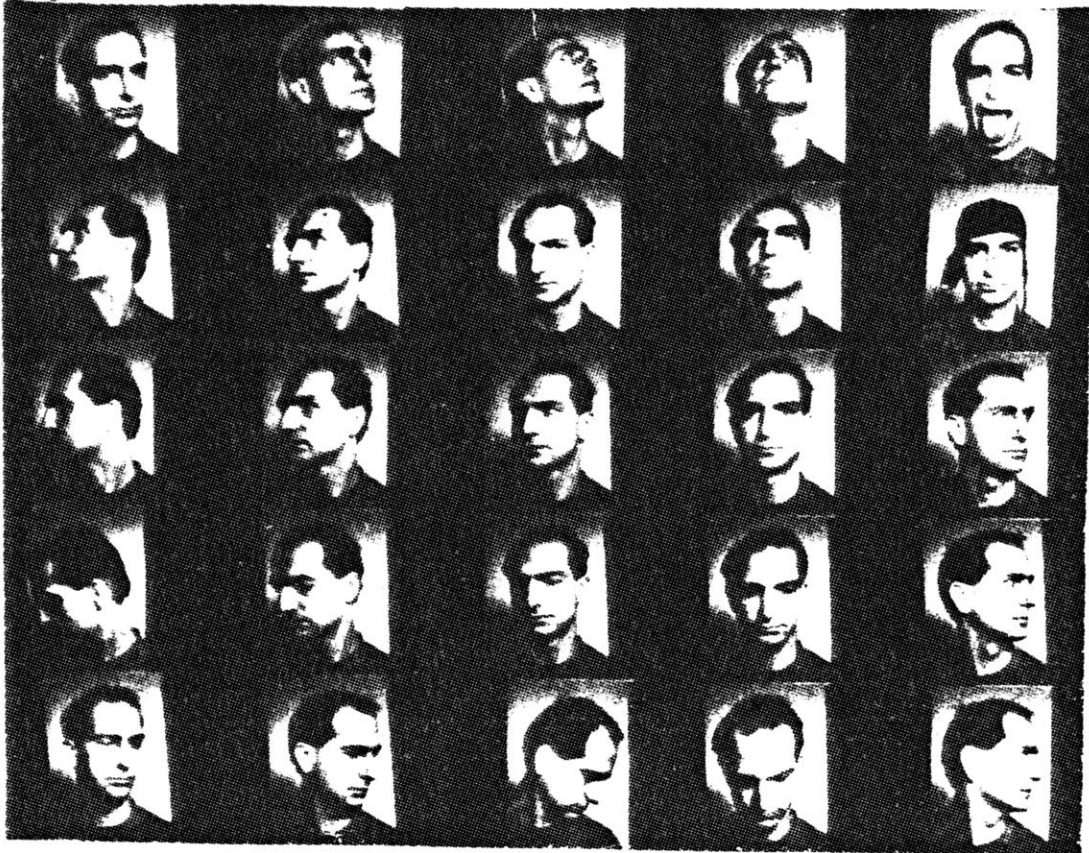
figure 17. SCANNER(above), LOOK(below)

figure 18. EXPRESS

would contribute towards the addition of nuance and intent, necessary adjuncts of "presence."

The notion, borrowed from the traditions of puppetry and animation, that a set of primitives could be sufficient to portray facial expression, as well as lip, head, and eye position, forms the basis for EXPRESSIONMAKER. SLOTFACE was designed as a tool to quickly display faces made from isolated features; either from one face in a variety of expressions or from several different faces. EXPRESS was the first attempt to isolate a set of primitives to portray facial expression. Twenty-five pictures of an actress were selected and digitized from sequences on the SPEECHMAKER-II videodisc. I found that it was possible to generate seven different facial expressions from this small set of images; five for a yawn, five for surprise, three to raise the eyebrows, five to "scrunch" and four to bare teeth. Many expressions could be generated with only two frames; a wink, a blink, and sticking out the tongue. This represents a considerable reduction from thirty pictures a second resulting from realtime video recording.

"Talking Mr. Aspen" was the first attempt to add expression to an averaged face. Two programs call "emphasize" and "smile" allow the user to designate a portion of a face (say, a smile) to be grafted onto another face in an averaging process. Four copies of "Mr. Aspen", an averaged face (see above), were "emphasized" with four different mouth positions from another set of faces on the disc. Cycling through the four resulting frames produced the illusion that the face was talking or smiling. EXPRESSIONMAKER was projected to be a combination of EXPRESS and Talking Mr. Aspen—a bank of expression sequences for
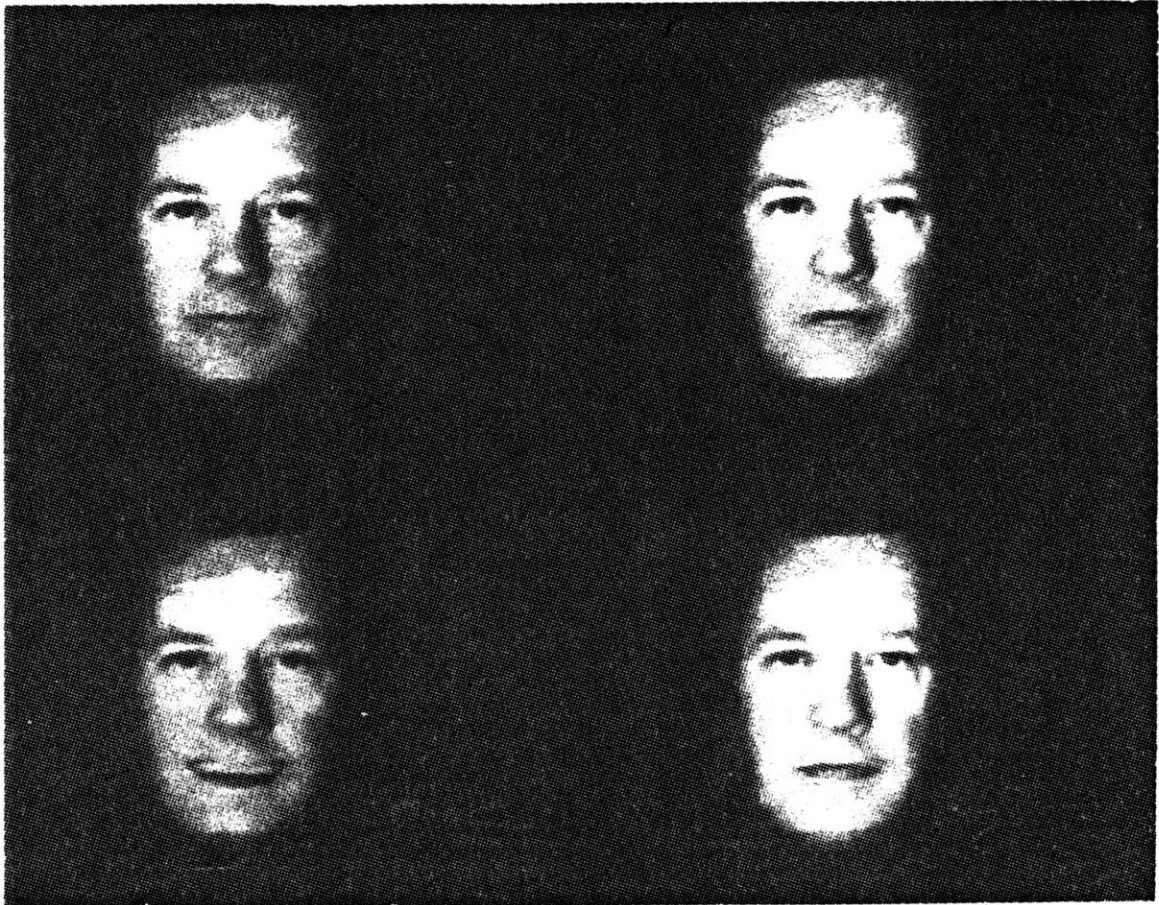
figure 19. Talking MrAspen

several facial types and an effective averaging technique to add expression to averaged faces.

SQUINT is a version of "Talking Mr. Aspen" with expression as well as movement. Facial expressions from one of the models on FACE-MAKER-II were "emphasized" onto an averaged face and animated with an optical video disc recorder. Images were sequenced to portray expression, attention and attitude.
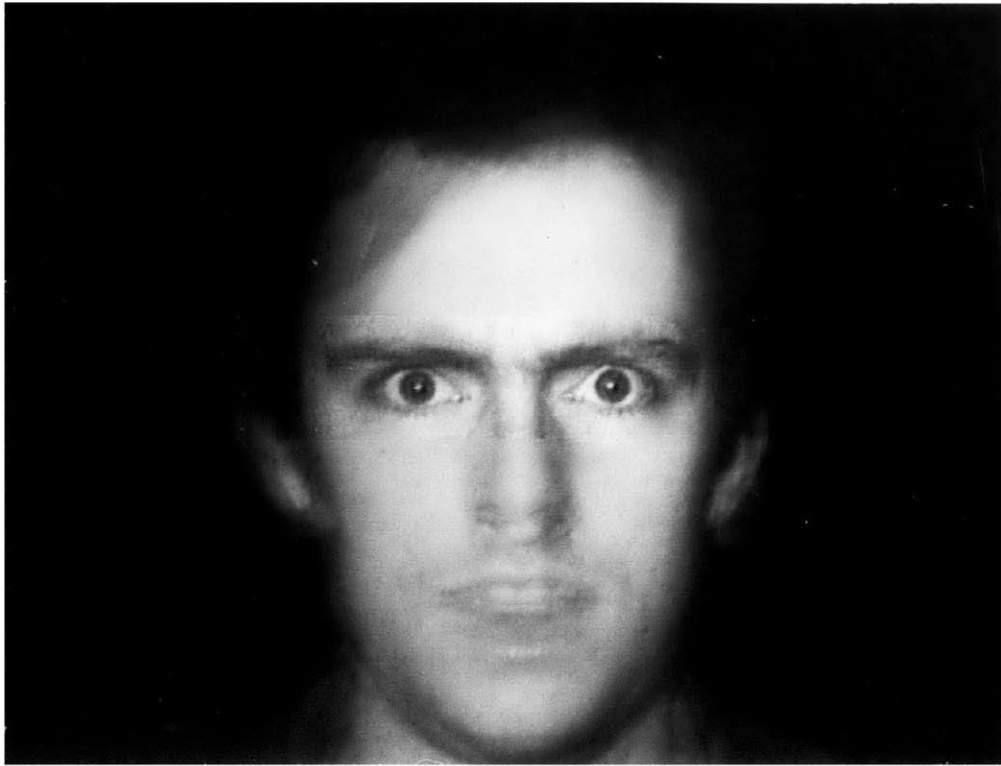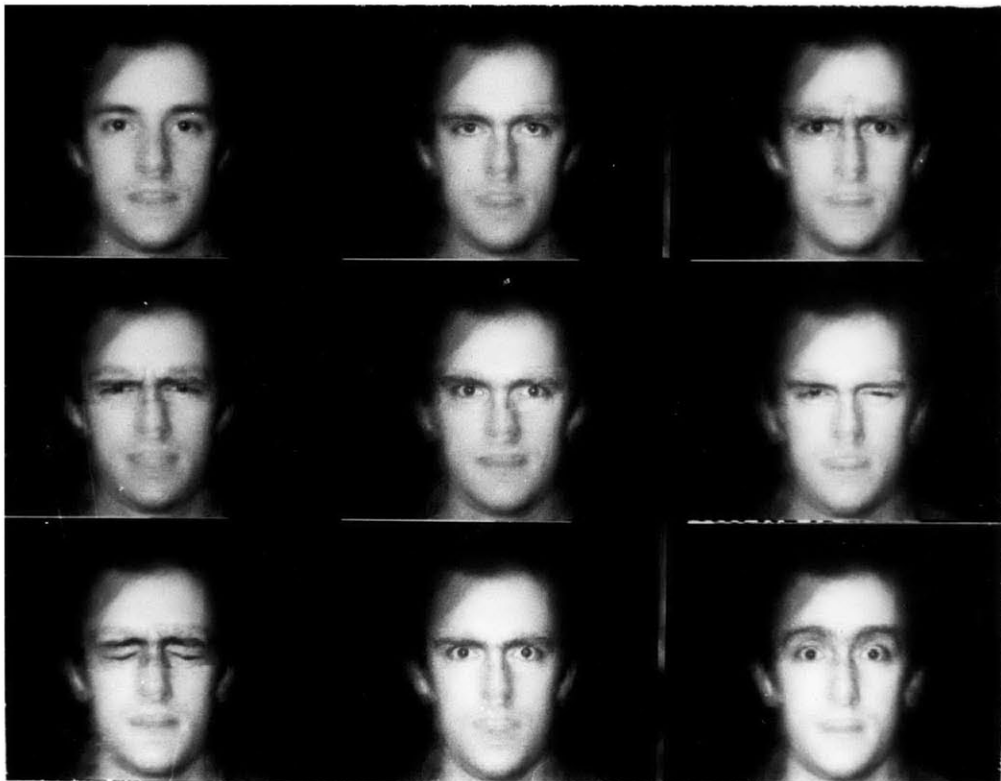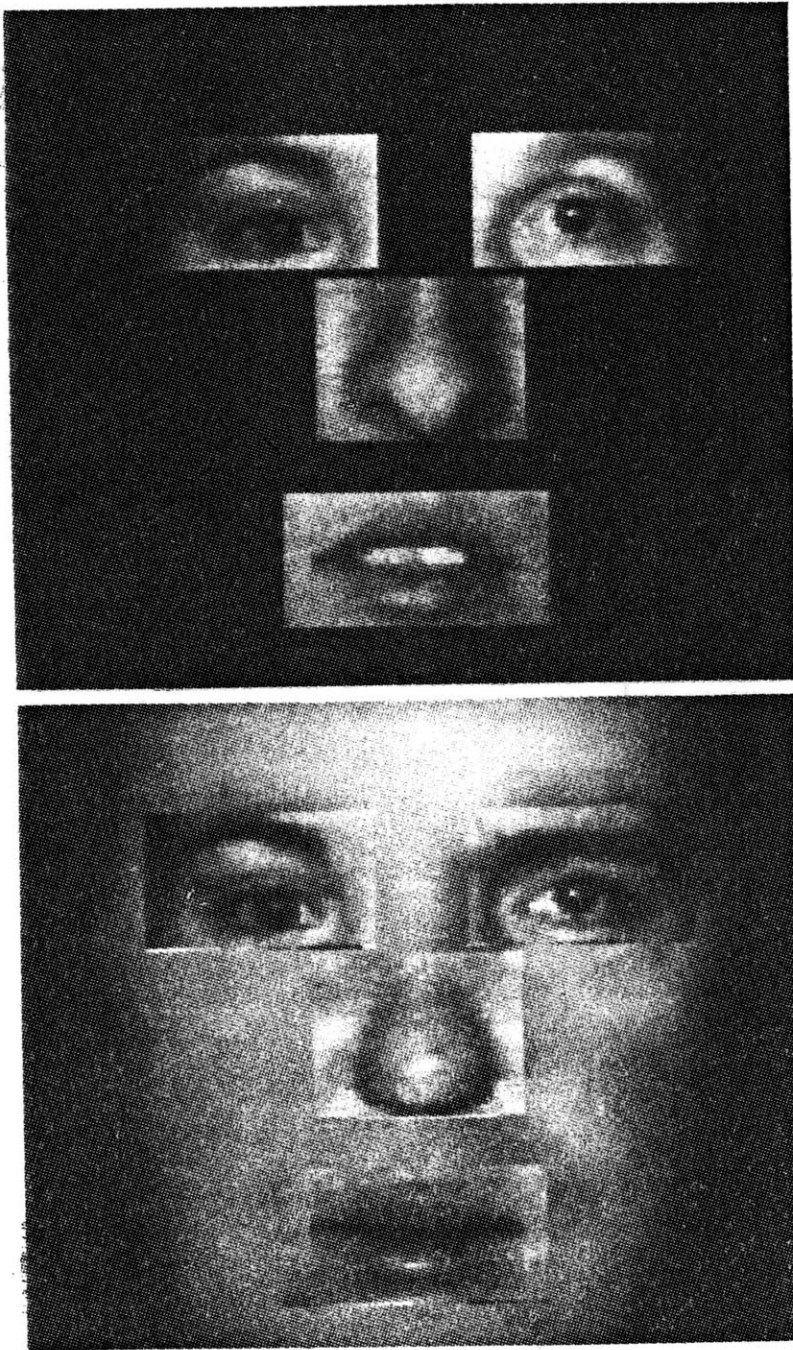
figure 20. SQUINT

figure 21.   SLOTFACE

80.

AFTERWORD

It's not the eyes, but the glance—not the lips, but the smile...

—Saint-Exupery,
Wind, Sand and Stars

The projects described in EXPRESSIONMAKER are sketches for an animation system to portray facial expression by manipulating photographs. Facial expression is a significant aspect of human communication—it conveys signals of emotion, thought, action and power. Effective portrayal of facial expression contributes to the quality of telecommunications, portraiture, theater and film animation.

Techniques to generate and manipulate facial imagery are valuable to both pre-recorded and realtime animation systems. A representation of facial behavior based on the catalogues and classification systems cited above could be used in animated film production. Sequences that convey emotional expression could be "scripted" similarly to the way existing animation systems manipulate geometric components and motion (see Reynolds, 1982). Such a dynamic graphic tool would allow the animator to rehearse, pre-visualize and edit sequences which specify emotion within an animated film.

An interactive system to portray expression in realtime could be used in teleconferencing systems, broadcast media, and live theatrical performance. A responsive, interactive "portrait" or "character" could be considered the ultimate electronic puppet; a

81.

participatory, responsive environment is fitting to portray a

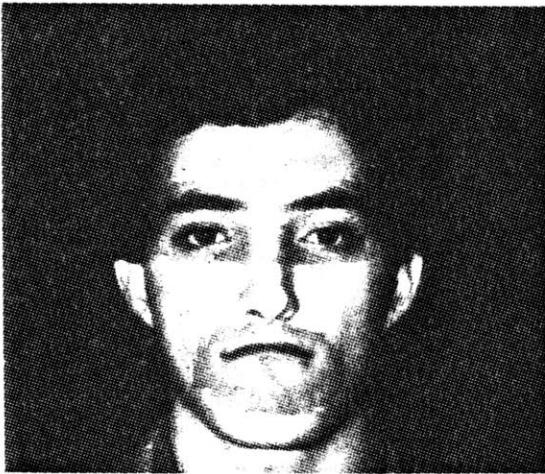phenomenon as spontaneous and personal as facial expression.

figure 22. emphasize

APPENDIX I: FACEMAKER-III DISCLOG

Titles	1200-1206
SLOTFACE	1207-1221
Children:	
single view	1222-1569
stereo left	4220-4238
	4570-4679
stereo right	6914-7023
interlaced stereo	9258-9367
Teen:	
single view	1560-1884
stereo left	4680-4779
stereo right	9024-7123
interlaced stereo	9368-9467
20-35:	
single view	1885-2944
stereo left	4268-4414
	4780-5143
stereo right	6612-6758
	7124-7487
interlaced stereo	8956-9102
	9468-11446
35-45:	
single view	2945-3064
stereo left	4430-4523
	5144-5719
stereo right	6774-6867
interlaced stereo	9118-13734
45+:	
single view	3965-3328
stereo left	4524-4569
	5720-5865
stereo right	6868-6913
	9212-9257
interlaced stereo	11556-11601
Faces from FACEMAKER-II	12427-14407

Expressions:
    single view   3376-4013
    (pw,jane,andy,ned,chao,parker,alex,
    francis,adrian,diego,nancy,laurie,
    becky,jane,pw,wolpe,moseley,peter,
    julia,catherine)

    stereo left   6004-6363
    (neena,bhapa,gloria,mc,bruce,aaron,
    pat,ellie,ed,misc.arcmac,wigs)

    stereo right   8348-8707
    interlaced stereo   10692-11051

Mohl's Beard   4013-4071

Richard's Beard:
    stereo left   4103-4220
    stereo right   6447-6564
    interlaced stereo   8793-8910

Illustrations   11138-12024
    (masks,MilMascaras,Huber,X-ray,
    evolution,anatomy,evolution,
    Lavater,LeBrun,Piderit,Felekey,
    Growtowski,Roosevelt 1882-1943,
    Galton,Wegman,portraits,photos,
    models,snaps,documentation)

Mugshots   12056-12354

Duchenne   12355-12426

Doubles   14438-14941

Triples   14942-15122

Excerpts from "Deep Hearts"   15558-16437
    (film by Robert Gardner shot
    in Niger Republic,1980)

APPENDIX II: FACEMAKER & EXPRESSIONMAKER PROCEDURES

Selection and averaging faces:

        FACESPACE: animation to pan across FACESPACE and
                   choose 'neighborhood' of faces to
                   look through on the disc
        peruse: allows stepping and exploring through
                   the disc, generates a file of designated
                   faces
        PERUSE: backround process during selection -
                   performs averaging and displays averaged
                   face for evaluation

Database classification(not yet implemented):

        DESCRIBE   dynamic database - augments basic data
                   with descriptions that are meaningful
                   to the individual user
        REMIND   evaluates chosen faces against a list of
                   favored features and displays new faces
                   with similar attributes

Evaluation and modification of files:

        editface
        dispface   (programs to display and edit the
        remakeface   component faces of an average)
        collage_face

Addion and emphasis of facial features:

        emphasize: designate feature by drawing on touch
                   sensitive screen
        smile: performs weighted average

Animating facial features:

        LIPSYNC: simulate lipsync with 16 lip positions,
                   voice input by synthesized, recorded or
                   live speech
        SCANNER: simulate eye movement with 17 eye positions,
                   input via joystick or headmounted eye-tracking
                   device
        LOOK: simulate head movement with 23 head positions,
                   input via joystick or body-tracking device
        EXPRESS: simulate expression with 25 pictures
        SQUINT: animated averaged face
        MOVE: animated averaged face
        SLOTFACE: generate faces from isolated features

REFERENCES

FACEMAKER:

Bahrick, H.P., Bahrick, P.O. & Wittlinger, R.P. "Fifty Years of Memory for names and faces: A cross-sectional approach", Journal of Experimental Psychology 1975, v.104, p.54-75.

Bradshaw, J.L. & Wallace, G., "Models for the Processing and Identification of Faces", Perception and Psychophysics, 1971, v.9, p.443-448.

Carey, S. & Diamond, R. "From Piecemeal to Configurational Representation of Faces", Science, 1977, v.195, p.312-314.

Carey, S. A Case Study: Face Recognition. (from Walker, E.(ed.) Explorations in the Biology of Language, Bradford Books, 1978.

Davies, G.M. & Christie, D. "Face Recall: An Examination of Some Factors Limiting Composite Production Accuracy". Journal of Applied Psychology, 1982, 67, 1, 103-109.

Ellis, H.D."Recognizing Faces". British Journal of Psychology, 1975, 66, 4, p.409-426.

Galton, F., INQUIRIES INTO HUMAN FACULTY & ITS DEVELOPMENT. London: J.M. Dent & Sons

Galper, R.E. "Recognition of Faces in Photographic Negative", Psychonomic Science, 1970, v.19, p.207-208

Harmon, L. & Goldstein, A.J. & Lesk, A.,"Identification of Human Faces", IEEE Proceedings, 1971.

Harmon, L., "The Recognition of Faces", Scientific American, November, 1973, 229(5) p.70-84.

Hoffman, D.D., THE VISUAL INTERPRETATION OF FACES. Unpublished MIT Thesis Proposal, 1981.

Laughery, K., Alexander, J. & Laine, A., "Recognition of Human Faces: Effects of target exposure time, target position and type of photgraph", Journal of Applied Psychology, 1971, v.51, p.477-483

Negroponte, N., BOOKS WITHOUT PAGES. IEEE Proceedings,1979.

Negroponte, N., Lippman, A. & Bolt, R. TRANSMISSION OF PRESENCE: ADVANCED MACHINE RECOGNITION AND IMAGE PROCESSING APPLIED TO TELECONFERENCING. Unpublished proposal to Cybenetics Technology Division, DARPA. Architecture Machine Group, M.I.T., 1980.

Negroponte, N., THE IMPACT OF OPTICAL VIDEODISCS ON FILMMAKING. Cambridge: MIT Architecture Machine Group report, 1980.

Mauldin, M.A. "The Effects of Composite Production, Time Delay, and Exposure Duration on Face Recognition", Unpublished Thesis, University of Houston, 1978.

Mohl, R. COGNITIVE SPACE IN THE INTERACTIVE MOVIE MAP: AN INVESTIGATION OF LEARNING IN VIRTUAL ENVIRONMENTS. Unpublished Phd. Thesis, Architecture Machine Group, M.I.T., 1982.

Perkins, D. "A Definition of Caricature and Caricature and Recognition", Studies in the Anthropology of Visual Communication, Spring 1975, 2(1), p.1-24.

Psotka, J. RECOGNIZING GALTON'S COMPOSITE FACES. Unpublished paper, 1977.

Rock, I. "The Perception of Disoriented Figures", Scientific American, 1974, v.230, p.78-85.

Sullivan, T.F. & Slater, J.A. "Minolta Montage synthesizer as Used in the Rockland County", B.C.I. Fingerprint and Identification Magazine, April 1975

EXPRESSIONMAKER:

Bassili, J. "Emotion Recognition: The Role of Facial Movement & The Relative Importance of Upper and Lower Areas of the Face", Journal of Personality & Social Psychology, 1979, v.37, p.2049-58

Bassili,J. "Facial Motion in the Perception of Faces & of Emotional Expression", Journal of Experimental Psychology: Human Perception & Performance, 1978, v.4, 373-79

Birdwhistell, R. KINESICS AND CONTEXT, ESSAYS ON BODY MOTION COMMUNICATION. Philadelphia: University of Pennsylvania Press, 1970.

Boring, E.G. & Titchner, E.B., "A Model for the Demonstration of Facial Expression", The American Journal of Psychology, 1923, 34, p.471-485.

Darwin, Charles. The Expression of Emotions in Man & Animals. Chicago: University of Chicago Press, 1965(First ed.,1872)

Davitz, J. et al. The Communication of Emotional Meaning. New York: McGraw Hill, 1964

Duchenne, G.B. MECHANISM DE LA PHYSIONOMIE HUMAIN. Paris: Bailliere, 1876

Dunlap, K. "The role of Eye Muscles and Mouth Muscles in the Expression of the Emotions", Genetic Psychological Monogrraph, 1927, v.2, p.199-233.

Ekman, P., Friesen, W.V. & Ellsworth, P. EMOTION IN THE HUMAN FACE. Elmsford, N.Y.: Pergamon Publishing Co., 1972

Ekman, P.(ed.), DARWIN AND FACIAL EXPRESSION: A CENTURY OF RESEARCH IN REVIEW. New York: Academic Press, 1973

Ekman, P. & Friesen, W.V., UNMASKING THE FACE. New Jersey: Prentice Hall, Inc. 1975

Ekman, P. & Friesen, W.V., FACIAL ACTION CODING SYSTEM MANUAL. Palo Alto: Consulting Psychologists Press, Inc.: 1978

Felekey, A.M., FEELINGS AND EMOTIONS. New York: Pioneer Press, 1924

Frois-Wittman, J., "The Judgement of Facial Expression", Journal of Experimental Psychology, v.XIII, 1930, 113-151

Ghosh, M. Nandikesvara's Abhinaya-Darpanam: A Manual of Gestures and Postures Used in Hindu Dance and Drama. Calcutta: Metropolitan Printing & Publishing House, 1934

Gillenson, M.L. $ Chandrasekaren, B. WHATISFACE: Human Facial Composition by Computergraphics, Proceedings of the ACM Annual Conference, 1975, v.9(1).

Goffman, E. INTERACTION RITUAL, ESSAYS IN FACE-TO-FACE BEHAVIOR. Chicago: Aldine Publishing, 1967

Gombrich, E.H., ART AND ILLUSION. Princeton: Princeton Unversity Press, 1960.

Gregory, W.K. OUR FACE FROM FISH TO MAN. New York: Capricorn, 1965.

Grotowski, I.J. TOWARDS A POOR THEATRE. New York: Simon & Schuster,1968

Huber, E. EVOLUTION OF FACIAL MUSCULATURE AND FACIAL EXPRESSION. Baltimore: The Johns Hopkins Press, 1931.

89.

Hjortsjo, C.H. MAN'S FACE AND MIMIC LANGUAGE. Sweden: Studentlitterature, 1969.

Izard, C. THE FACE OF EMOTION. New York: Appleton, 1971.

Lavater, G. (ed.M.Moreau), L'Art de connaitre Les Hommes. Paris: 1820.

Lynn, J., "An Apparatus and method for stimulating, recording and measuring facial expression," Journal of Experimental Psychology, 1940, v.27, p.81-88.

Malraux, A., CHARLES LE BRUN, 1619-1690: PEINTRE ET DESSINATEUR. Exposition Chateau de Versailles, 1963.

Meissner, B. "Ein neubabylonisches Zuckungsbuch", Sitzungberichte der Preussischen Akademie Der Wissenschaften. Berlin, 1921.

Parke, F. L., "Computer Generated Animation of Faces", Proceedings of the ACM Annual Conference, 1972, v.1.

Platt, S.M. & Badler, N.I. "Animating Facial Expressions", Proceedings of the ACM Annual Conference, 1981, v.15(3).

Piderit, T. Mimik und Physiognomik. Detmold: 1885(first ed: 1867).

Reynolds, C.W., COMPUTER ANIMATION WITH SCRIPTS AND ACTORS. Proceedings of the ACM Annual Conference, 1982, v.16(3).

Thomas, F. & Johnston, G. DISNEY ANIMATION, THE ILLUSION OF LIFE. New York, Abbefille Press 1981.

Ruckmick, C.A. THE PYSCHOLOGY OF FEELING AND EMOTION. New York: McGrawHill, 1936.

Schlosberg, H. "Three Dimensions of Emotion", Psychological Review, 1954, v.61.

Werner, E.S. DELSARTE SYSTEM OF ORATORY. New York: Werner, 1893

Yates, F.A., THE ART OF MEMORY. Chicago: University of Chicago Press, 1966

ACKNOWLEDGEMENTS