CARICATURE GENERATOR

by

SUSAN ELISE BRENNAN

B.A., Cornell University
1975

Submitted in Partial Fulfillment
of the Requirements for the
Degree of
Master of Science in Visual Studies
at the
Massachusetts Institute of Technology
September, 1982

Signature of Author . . . .
Department of Architecture
August 6, 1982

Certified by . . . .
Nicholas Negroponte, Professor
of Computer Graphics
Thesis Supervisor

Accepted by . . . .
Professor Nicholas Negroponte, Chairman
Departmental Committee for Graduate Students

CARICATURE GENERATOR

by

Susan Elise Brennan

Submitted to the Department of Architecture on August 6, 1982 in partial fulfillment of the requirements for the degree of Master of Science in Visual Studies.

ABSTRACT

The human face is a highly significant visual display which we are able to remember and recognize easily despite the fact that we are exposed to thousands of faces which may be metrically very similar. Caricature is a graphical coding of facial features which seeks to be more like the face than the face itself: selected information is exaggerated, noise is reduced, and the processes involved in recognition are exploited. After studying the methods of caricaturists, examining perceptual phenomena regarding individuating features, and surveying automatic and man-machine systems which represent and manipulate the face, some heuristics for caricature are defined. An algorithm is implemented to amplify the nuance of a human face in a computer-generated caricature. This is done by comparing the face to a norm and then distorting the face even further away from that norm. Issues of style, context and animation are discussed. The applications of the caricature generator in the areas of teleconferencing, games, and interactive graphic interfaces are explored.

Thesis Supervisor: Nicholas Negroponte
Title: Professor of Computer Graphics

The work reported herein was supported by the Cybernetics Technology Division of the Defense Advanced Research Projects Agency, under contract number MDA-903-8-C-0097

TABLE OF CONTENTS

1.0 Introduction 5

2.0 The Plasticity of the Face 8
  2.1 The Face as Display 8
  2.2 Face-to-Face Communication 9
  2.3 Research Issues 13

3.0 The Recognition of Faces: Invariants 15
  3.1 Faces as Patterns 16
  3.2 Invariants and the Recognition of Faces 20
  3.3 Invariants in Pictures 23
  3.4 The Information in Caricature.................. 30
  3.5 Empathy 31

4.0 Caricature as Transformation...A Definition 33
  4.1 Evidence 35
  4.2 Three-Dimensionality 37
  4.3 Animals and Other Metaphors 39
  4.4 Other Elements in Caricature 42
  4.5 A Definition 45

5.0 The Heuristics of Caricature 46
  5.1 Some Musings Based on Personal Experience:
      What are the Ways to Caricature? 46
  5.2 Portraiture 47
  5.3 Rules For Drawing Caricatures 51
  5.4 Cartoonists 54

6.0 Automatic and Man-Machine Systems for
      Face Recognition and Representation 59
  6.1 Line Drawings 59
  6.2 Computer Recognition of Faces 63
  6.3 Computer-Warped Faces 66
  6.4 Computer-Synthesized Faces: Two-Dimensional
      Projections 69
  6.5 Computer-Synthesized Faces: Three-Dimensional . 72

7.0 Constraints/Automation Issues 75
  7.1 Primitives 75
  7.2 Input Constraints 78
  7.3 Features 79
  7.4 Stereo Input 81
  7.5 The Point of Departure 81
  7.6 Man-Machine Interaction 83

8.0 Implementation: System Design and Programs 86
  8.1 Input Program 86
  8.2 The Rubber Band Face 90
  8.3 Gumby 92

9.0 The Future of the Caricature Generator 101
9.1 Applications 101
9.2 Simple Animation 103
9.3 Other Transformations 103
9.4 Context 104
9.5 Style 107
9.6 Participatory Caricature 109

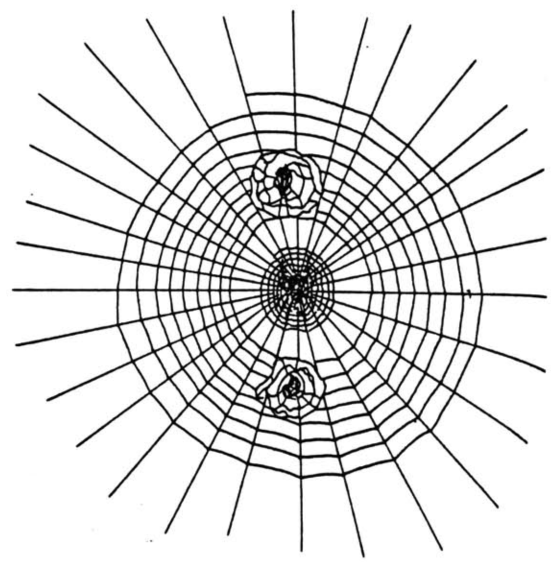
Acknowledgements 110

Bibliography 111

1.0 INTRODUCTION

The art of caricature dates from the end of the sixteenth century, most probably named for the Carracci brothers who popularized a style of overcharged or loaded portraits. The natural aspect of caricature, however, goes back much farther; the representation of a thing by exaggerating its key features is not even solely the activity of human beings.

Certain visual patterns have enhanced perceptual significance to the organism, probably due to some advantage toward survival provided by detailed differentiation among relatively similar patterns.



The spider Cyclose mulmeinensis (right, center of web), increases his survival odds with a form of self-caricature. (Wickler, 1978)

A clue to discovering what are the key features that an animal uses to make important discriminations between threats and non-threats or food and non-food can be unearthed while studying the defensive disguises found in many species. Protective coloration exploits the simple fact that it is more difficult to recognize a figure which shares the same color and texture as its background. Batesian mimicry occurs when a vulnerable "mimic" species shares some feature with a protected (usually, poisonous) "model" species in order to fool its predators into leaving it alone. Survival of the mimic species is actually based on the selection of two sets of key features - the visual features which are mistakenly recognized by the predator as belonging to the model species, and the other features (often olfactory or behavioral) which identify the mimic to its own species so that it can find individuals to mate with.

Wolfgang Wickler commented on another form of mimicry found in the Siamese lantern-fly:

> "This insect mimics two insect antennae with the tips of its hindwings and bears a large dummy eye on each fore-wing. The lantern-flies (Fulgoridae) are good jumpers and jump in a direction opposite to that expected from a view of the dummy head." (Wickler, 1978)

These instances of mimicry are characterized by the transmission of a particular message based on the amplification of certain salient features and the suppression of others, and are dependant for success upon recognition by another organism.

Caricature is a graphical coding of key facial features which seeks to be more like the face than the face itself: selected information is exaggerated, noise is reduced, and the processes involved in face recognition are exploited. This distortion is perceptually significant in that it appeals to one's mental model of what is unique, or individuating about a face, often in such a way as to instigate delight at seeing the features so represented. Traditionally executed with a minimum of lines and loaded with symbols and juxtapositions, caricature is a sophisticated form of bandwidth compression.

2.0 THE PLASTICITY OF THE FACE

2.1 THE FACE AS DISPLAY

Because of its plasticity and its capacity for expression and abstraction, a human face is a display of the highest resolution. It is a channel with the capacity for input as well as output.

The face is a significant carrier of information during face-to-face communication; its detailed musculature enables it to be modulated in all ways anatomically possible. Each face has a dynamic range of its own to which other people must calibrate their observations, before interpretation. While there are certainly universal tendencies in the expressive behavior of faces individuals go to different extremes in using their faces to express a great variety of subtle messages. Noise may be introduced into the communication channel in the form of idiosyncratic facial gestures, nervous tics, hair worn covering the face, stammering, glasses, smoking, etc. (although these phenomena may also provide punctuation for another message, such as personal identity).

The plasticity of the human face is intimately experienced by all of us on a tactile/muscular level, and by artists, actors and mask-makers on an objective visual level as well. The face which begins by being at rest can be modulated in many diferent ways and to greater or lesser degrees. Some individuals consciously participate in the surface plasticity of their own faces by varying color, texture and shading with makeup. The plasticity of the face over time displays the effects of another modification of the face, that of aging. Sometimes the countenances of spouses seem to grow more alike over the years, perhaps from decades of mirroring one another's expresions, or from sharing similar environments and experiences. Years of fleeting facial gestures become inscribed as permanent lines.

2.2 FACE-TO-FACE COMMUNICATION

In an early experiment in teleconferencing members of the Architecture Machine Group explored the notion of "expression space" where a relaxed, attentive face was distorted along expressive axes such as active/passive, agreeable/disagreeable, etc. It was proposed that pictures of faces could be located spatially according to a very primitive classification using any one of the following techniques: subjective rating of faces' emotional states by onlookers, analysis of the degree of muscular activity, analysis of some gross voice parameters such as changes in speech rate or amplitude, or by directing the subject to enact a series of facial expressions ("irritated - angry - furious"). Actors and volunteers were videotaped, and short cycles of moving expressive sequences were mastered onto optical videodisc where they could be subsequently still framed or played backward or forward at any

Example of "Expression Space" from the Transmission of Presence teleconferencing project at the Architecture Machine Group, M.I. T. (1981)

ATTENTIVE *IRRITATED* *ANGRY* *FURIOUS*

speed and in any order. The cycles of facial expressions stored on videodisc were selected according to key words recognized by a connected speech processor. Once the expressive category was identified in this manner and the appropriate videodisc images were cued up, a voice channel was used as an on/off switch to select alternate cycles of talking/ silent faces. The effect was like that of a haphazardly dubbed film, and became known as auto dubbing, or "Zero-Bandwidth Video", since the image sequences were not transmitted but synthesized using the speech channel. The purpose of the experiment was to evaluate the contribution of facial expression and graduated degrees of lip sync to the intelligibility of a message within a teleconferencing context where speech was used to drive and select visual images.

Another approach was to combine real-time moving parts of the face (such as the eyes or mouth) with computer-generated or pre-stored parts of the same face, and to observe the effects of possible "clashes" of expressions in the resulting facial collage. The purpose of this experiment was to investigate the possibility of dynamically allocating reduced video bandwidth to that part of the face which carries the most information. It was discovered that as long as the facial expression was in accord with the tone of voice, a greater discrepancy in lip sync was tolerable. This experiment was known as "semantic bandwidth compression". This research is on the lunatic fringe of teleconferencing.

Semantic Bandwidth Compression, from Transmission of Presence project, Architecture Machine Group, MIT, 1980.

In this videotaped teleconference simulation, the face and background originate from a slow-scan video camera, while the moving mouth is transmitted in real time. Later systems took the background face from optical videodisc, cycling back and forth over several video frames to animate the image. Result: a realistic face that does not transmit a complete facial expression, and is therefore misleading.

2.3 RESEARCH ISSUES

Some of the work described above raised questions regarding the ethics and utility of driving an image with speech. Does that really provide useful information? It does in situations such as conference calls where it is difficult to identify who is speaking at a given moment. A voice switch is used to animate the face of the speaker(s) at a given moment; providing a spatial referent for each participant orchestrates the conversation. On the other hand, if the image is realistic enough that the viewer cannot tell whether it is full-bandwidth video, or semantically compressed, then he does not know how far to "trust" the talking head. The facial expression may not be the same as the one on the speaker's face. (Media experts know that on television there are things that one can say with a smile that one would never say in print). One could argue that it is more informative and more ethical not to try to emulate real talking heads, but to ensure that the viewer knows at all times what context, or degree of literalness, is represented by the display. This principle makes it possible to abstract from the face in a teleconferencing display.

> "...next to the photograph, the human heart trusts the caricature - which is all opinion. We may not agree with the opinions of the caricaturist about that particular subject, but they are frankly exposed, and we know what allowance to make for them."
>
> (critic Charles Marriott in
> THE NATION AND THE ATHENAEUM, after Lynch, 1929)

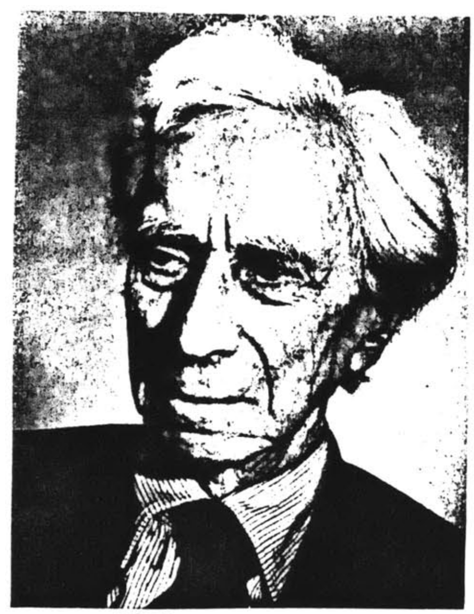
The work in caricature at the Architecture Machine Group grew out of the research on faces and was originally inspired by the investigation of bandwidth reduction in face-to-face communication. Actually caricature is an enduring personal interest of mine; its semantic impact is much greater than that of a mere a reductionist medium. My intention is to develop a theory of computation for caricature by considering definitions of caricature by perceptual psychologists, historians, caricaturists and others; then to construct an algorithm for transforming a face or a line drawing of a face into a caricature; and finally, to define constraints for and to implement the automatic caricature generator.

Thomas Nast's classic caricature of Boss Tweed (Harper's Weekly, 1872)

3.0 THE RECOGNITION OF FACES: INVARIANTS

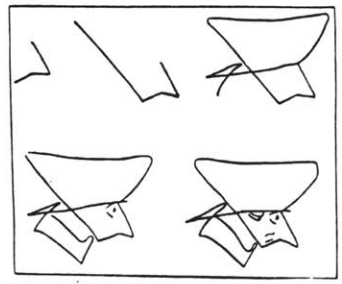
> "...what we learn about caricature will help us understand how faces themselves are perceived."
> (Hochberg, 1972)

Hochberg speculated that the way a caricature is encoded and stored is somehow identical to the way a physiognomy is encoded and stored. Others agree: the relationship of caricature to face perception is a theme in E. J. Gibson (1969) and in Perkins (1975). Therefore let us look at the recognition of faces as a reasonable starting point toward a theory of caricature.



Bertrand Russell (clockwise from top): David Low's caricature (from E. J. Gibson, 1967), as child and adult (from Gombrich, 197), Al Hirschfeld's caricature (1961).

3.1 FACES AS PATTERNS

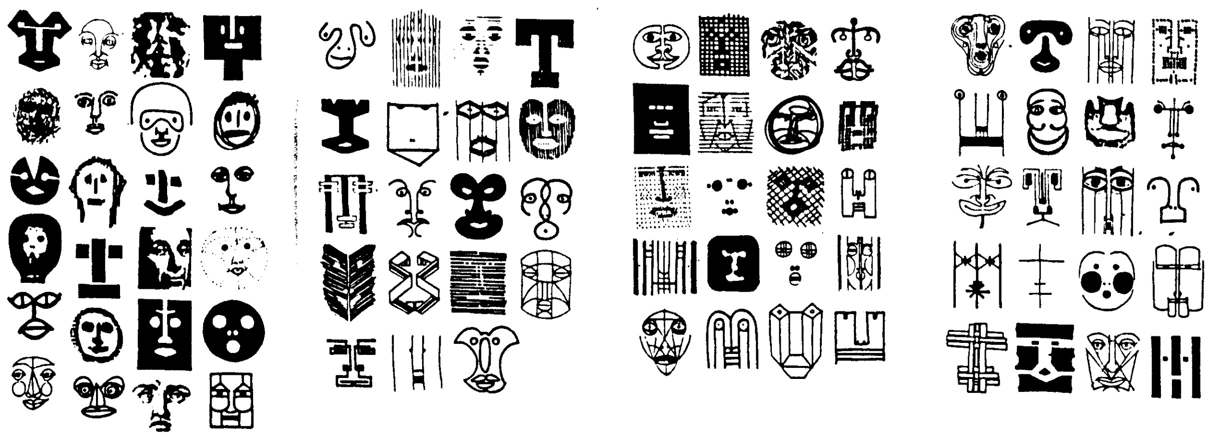


Wilhelm Busch's steps for drawing Frederick the Great

Human faces are such compelling configurations that we see them in trees or in cracks in the wall whenever such an interpretation is even remotely possible. The fundamental human ability to interpret an abstract pattern as a face, or as a specific face, makes possible the selective compression of the facial image into a very few lines, with no loss of essential information. The process of reduction without sacrificing recognition is the beginning of caricature.

Cliff near Beddgelert, North Wales (Michell, 1979)

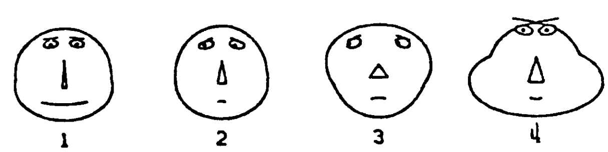
The Sphinx Rock, north-west England



Abstract faces from Munari (1966)

Seen simply as patterns (as a machine might see them), faces contain more similarities to one another than they do differences; but in truth, they are much more than the sum of their parts. We are so sensitive to facial proportions and surface curvature that the slightest change is often enough for us to perceive a wholly different structure or expression within the face.

Chernoff's faces (1973)



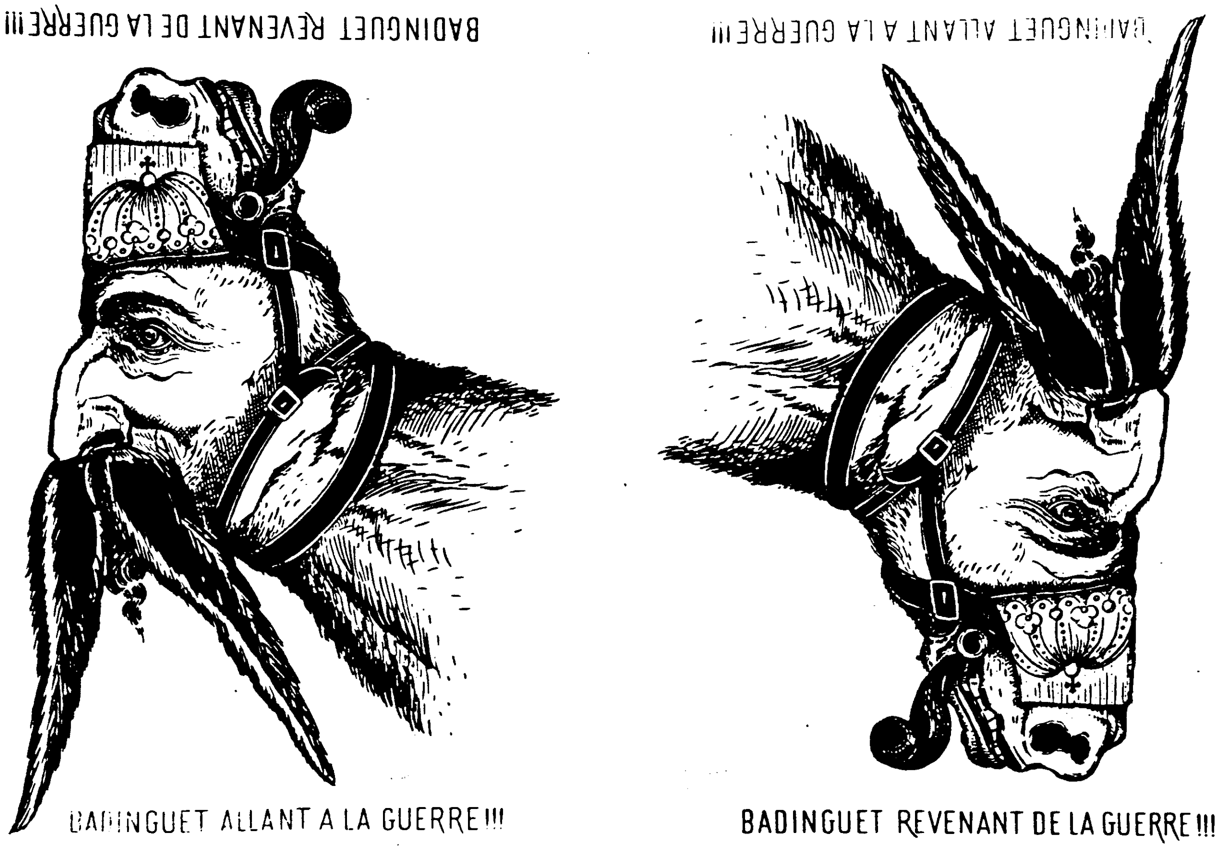
It has been shown that pictures of faces are more difficult to recognize when they are spatially inverted (Yin, 1969; Hochberg and Galper, 1967) or presented in photographic negative (Galper, 1970), even though, as intensity arrays, they contain the same amount of information as a positive, upright photograph.

The Aspen videodisc produced by the Architecture Machine Group included a large sample of pictures of human faces all registered on the eyes. After studying the remarkable variety among these faces displayed one after another on a television monitor, it is obvious that people lack the vocabulary to make all of the subtle differentiations that their memories are prepared to make in the case of faces. There are no absolutes; some facial dimensions such as interocular distance or height of forehead influence the perceived dimensions of adjacent forms, such as the length of the nose and upper lip. Blurred images can be recognized as familiar faces. It is intuitive that faces are significant as Gestalten, understood at a glance.

On the other hand, facial description experiments have been conducted which imply that hierarchies of attributes are checked during face recognition (Bradshaw and Wallace, 1971), or which construct faces out of primitives. Individual features have been manipulated to mask recognition; parts of the face have been presented in isolation and the responses analyzed. Recognition has been approached as a context-variable look-up table of facial features, or as a sequence of template-matching operations.

Vision researcher D. D. Hoffman investigated dividing up the face into patches, each with its own coordinate system to represent surface curvature by canonical parameterization. He has taken a generalized approach to the perception of facial geometry but allows the possibility of a subsequent "'facial model' that represents important facts about the face in a coordinate system constructed by rules that are quite face-specific." (Hoffman, 1981)

It is not the object of the caricature research to fan the controversy about the mechanisms of the visual perception of faces. It seems likely that the visual system recognizes a face using a flexible process which operates simultaneously using both top-down and bottom-up information, constantly revised by context. It is possible that the process of face perception is not solely a vision process.



Reversible head of Napoleon III, anon., c. 1870 (from Lucie-Smith, 1981)

3.2 INVARIANTS AND THE RECOGNITION OF FACES

- likeness and unlikeness

As with the recognition of many other objects, recognition of faces is not inhibited by most changes in viewing angle or lighting conditions. The literature on face recognition implies that the ability to recognize and differentiate among a vast population of faces is a learned ability, acquired very early in life. There is much evidence that faces are a very special case in cognitive coding. The human face is a highly significant visual display which we are able to remember and recognize easily despite the fact that we are exposed to thousands of faces which may be metrically very similar.

How do people accomplish this miracle of coding and decoding faces? E. H. Gombrich comments:

> "...it is not really the perception of likeness for which we are originally programmed, but the noticing of unlikeness, the departure from the norm which stands out and sticks in the mind. " (Gombrich, 1972)

Far right: Theodore Roosevelt; near right: Harper's Weekly cartoon by E. W. Kemble, 1912 (from Gombrich, 1963)

Gombrich cites the masking effect that a strong feature has upon the perception of more individuating but subtle features. For example, during initial exposure to an unfamiliar face, unexpected dress, coiffure, scars or racial identity may inhibit the viewer from attending to structural features while causing him to remember the face according to that one superficial characteristic. This coding works only as long as that characteristic always belongs only to that one face in the population of faces to be recognized. Galper (1973) illustrated what most of us know intuitively, that indeed, we have a great deal of difficulty recognizing faces which belong to a different racial group than the one(s) with which we are most familiar; the overwhelming predominance in perception of the racial difference masks the real individuating characteristics. She found that with enough training and familiarity, however, people can learn to recognize members of other races easily.

In face recognition we pick out the invariants. Expression, aging, hairstyle, and props are perceived as transient elements superimposed upon the permanent facial structure. Many of us have had the experience of running into an associate and noticing that there is something inexplicably different about his appearance. We may exclaim "You've lost weight!" when actually the person has merely shaved off his beard. We notice that some superficial element has changed, but may be hard pressed to identify exactly what that element is, since our model of that person's face is based on its more permanent features. This ability to code the invariants of faces enables us to distill personal identity from the ravages of time and fashion and to identify as individuals known to us even such essentially strange faces as those we encounter at high school reunions. Carey (1977) discusses the early development of this ability to discriminate in her studies of face recognition by children who were able to filter out the effects of superficial cues such as the same hat or expression appearing on different faces by the age of ten years.

People are adept at separating the permanent structure of the face from the temporary interplay of expressive musculature which takes place due to emotion and speech. This ability to differentiate is acquired because we are normally exposed to real faces which are not frozen in any one configuration which might provide merely a unique viewpoint of that face. (I use the word "unique" to mean a particular viewpoint which implies a relationship among lines or objects which is not corroborated by most other points of view, and is therefore misleading. A machine vision system cannot analyze a scene properly if all it has available happens to be a unique view). Because of experiences with binocular vision and seeing motion, our own inner database about human faces is one based on volumes and activity, not on flat, still images; therefore, unless the distortion is to be only superficial, a caricature should be an orthographic projection of three-dimensional information compiled from multiple viewpoints, and then exaggerated.

Recognition of faces from a single photograph is therefore an unreliable proposition; unless a photograph of a face represents a characteristic view of a familiar person, or unless any distortion is extreme or obvious, it is often difficult to separate the superficial characteristics from the invariants. When a permanent structural element is misinterpreted as a superficial modulation of the face, such as one caused by an expression of emotion, one's face may be perceived as always looking worried, astonished, etc. The more viewpoints, the better one can model the invariants of that face.

3.3 INVARIANTS IN PICTURES

Psychologist J. J. Gibson (1971) rejects what he sees as two prevalent theories of the perception of pictures (theories which seem, after all, overly simplistic): that pictures are composed of symbols and are read as such; or that they simulate exactly the same pattern of light rays falling on the retina as the thing that they represent. He proposes instead that the important information in a picture is the same whether it is a caricature or a photo - or a retinal image. His somewhat controversial theory of "formless invariants" is an attempt to separate visual sensation from cognition in the perception of a thing despite any change of context or continual transformations which may be applied to it:

> "There is no form' left in a continuous transformation. It has vanished and all that remains is the invariants."
> (Gibson, 1973)

Perhaps when one is first exposured to a new face one makes a mental caricature and commits it to memory, subsequently revising it slightly as more information is added to the mental model. As far as recognition is concerned, a caricature may be preferable to a more literal representation:

> "In observing a caricature or a political cartoon one often does not notice the lines as such...but only the information they convey about the distinctive features of the person caricatured. The caricature may be a poor projection of his face but good information about it. The form of the face is distorted but not the essential features of the face." (Gibson, 1973)

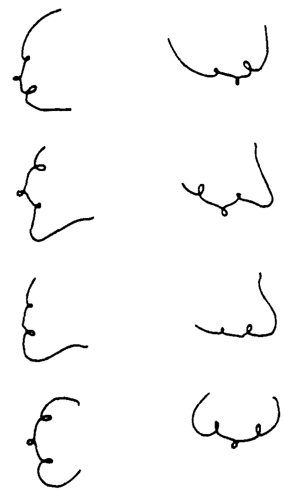
It is important for us to add to this concept that caricature not only reduces noise by eliminating non-essential information (which for recognition purposes may only be providing distraction), but also amplifies the essential features of the individual's face by exaggeration of the distinctive features. The distortion is not random, but deliberate and essential; it ensures recognition only if done intelligently. The reason that the caricature is an inaccurate projection of accurate information is that it often compresses information from several viewpoints into one two-dimensional projection of a face.

Rudolph Arnheim (1979) objects to Gibson's notion of invariants:

> "By no means is the term invariant' applicable if it implies, as it does in Gibson's usage, that the variables due to changing station points drop out of perception as irrelevant accidentals. This objection is especially valid when Gibson proposes to speak of invariants in relation to perception as a grasping of structural features. No invariant distinction between what is essence and what is accident can be made for any target of perception or of representation, because the level of abstractness of a percept or picture varies from instance to instance. If someone looks at the autumn colors of trees, are the colors of the essence of the trees or are they accidental and therefore not subject to perception?" (Arnheim,1979)

This criticism overlooks the fact that the color of a tree can be treated as an invariant that has two states, being either green or any one of a range of possible autumnal colors. The level of abstraction of a picture is something that the observer immediately assesses when he looks at it, and therefore the list of invariants applicable to that picture is calibrated to the ground rules of the picture itself. For example, naive observers often report confusion when presented with new artistic styles which are meaningful to other more experienced observers. Gombrich notes the following regarding different viewer expectations determined by the degree of realism vs. abstraction in pictorial style:

> "the artist who uses such an abbreviatory style can always rely on the beholder to supplement what he omits. In a skilled and complete painting, any gap will be disturbing.."
>
> (Gombrich, 1960)



Scrawls presented as faces (far left); the same scrawls rotated 90° and presented as writing (left). E. J. Gibson, 1967)

The invariants depend upon the context. The particular characteristics that become distinctive features of an image depend in part upon their ability to be contrasted with the distinctive features of other images. Eleanor J. Gibson cites a study devised by J. J. Gibson testing this hypothesis. He presented the same set of scrawls to two groups of subjects. One group was told that they would see faces, and were shown the scrawls in an upright position; the other was told they would see secret writing, and were presented with the scrawls rotated 90 degrees. Subjects had significantly more success remembering the patterns as faces than as writing. This is an interesting result, given that people are also skilled in recognizing writing with all its variations in typeface or penmanship. Eleanor J. Gibson concludes that stimuli that specify distinctive features in one class are irrelevant in another.

J. J. Gibson mentions some information-bearing features such as straightness vs. curvature, perpendicularity, parallelity vs. convergence, intersections, closures and symmetry. These features appear not only in faces but in all classes of images. It would be helpful if J. J. Gibson had been more specific as to how invariants are derived from visual stimuli. His theory does not help us to explain the special case of face recognition, as opposed to any other pattern recognition.

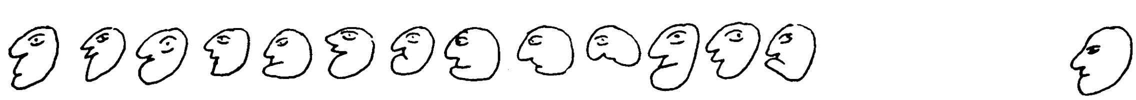
E. H. Gombrich (1960), on the other hand, makes a special case for faces:

> "The recognition of the human face...is not wholly learned. It is based on some kind of inborn disposition."

Rudolph Arnheim (1966) attacks this notion, saying:

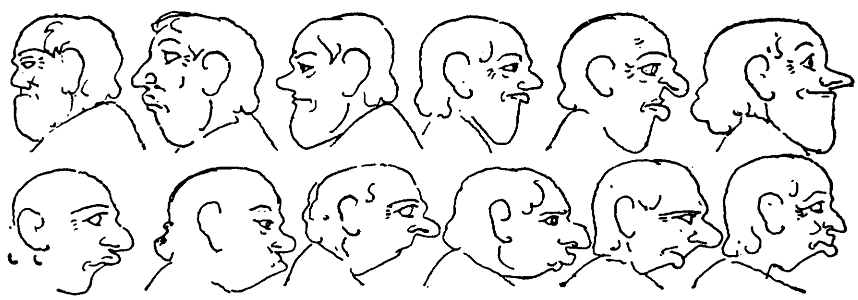
> "Gombrich would assume that the recognition of the human face is based on some kind of inborn disposition.'"

Although Gombrich is not specific, this inborn disposition could be the result of tactile as well as visual coding of the human face. Arnheim says that Gombrich implies that the ability to caricature originates without mediation from the outside world. Probably Gombrich is merely commenting on the fact that one can learn how to distort the human face and render all the possible expressions by experimental doodling:



"The practical physiognomics needed for a picture story could be learned by a recluse who never sets eyes on any human being. All he needs is drawing material and some perseverance. For any drawing of a human face, however inept, however childish, possesses, by the very fact that it has been drawn, a character and an expression. This being so, and being quite independent of knowledge and art, anybody who wants to try should be able to find out the traits in which this expression resides. All he must do is to vary his scrawl systematically. (Gombrich, 1960)

(previous page and below) Topffer's doodles from the "Essay du Physiognomie", (1845)



Note that Gombrich is not describing the origin of caricature in this passage, but the origin of the picture story or comic strip. Random or playful distortion of faces in general is not caricature, but cartooning. Caricature is the deliberate distortion of a particular face. Strictly speaking, gargoyles and the grotesque heads of Leonardo da Vinci are not caricatures, unless the definition is broadened to include portraits of a specific "type".

Da Vinci's study of a young and old man

3.4 THE INFORMATION IN CARICATURE

> "More realism is not necessarily a good rule of thumb for more informativeness in pictures." (Perkins, 1980)

A caricature may contain more information than a photograph in that it amplifies what is remarkable about the image; the photograph does not care. Ryan and Schwartz (1956) exposed subjects to four types of pictures for brief moments, lengthening the intervals until the subjects could just recognize what they saw. For pictures of hands, the subjects were asked to respond by positioning their hands in the same way. The cartoon hands were recognized correctly after less exposure than the more "realistic" unshaded line drawings. For pictures of machines and switches, cartoons were the most effective mode of representation. Photos and shaded drawings were next. The metrically accurate outline drawing needed the longest exposure to be perceived correctly. Overall, cartoons proved to be the most recognizable pictures.

These results are examined by Hochberg (1972) who concludes that those pictures which emphasize spatial characteristics of hands are more informative than some of the more accurate but linear two-dimensional projections of hands. He points out that in the cartoon the smoothness of the contours has been simplified and exaggerated, intersecting contours have been drawn so that they meet at right angles, and the relative separation of boundaries between fingers has been increased. Thus the most informative images of the hand manage to convey volume through shading or exaggeration, while the outline drawing is very vulnerable to confusion due to noise or slight metric distortion.

Evidence of this sort corroborates the usefulness of caricature as a mode of representation in some cases over the photograph, and in all cases over the outline drawing. The following hypothesis will be investigated using the caricature generator: a literal line drawing, whether it is rotoscoped, traced, or derived using edge filters on a single picture of a face, is extremely vulnerable to metric distortion. Caricature is a promising form of representation that is not hurt by metric distortion.

3.5 EMPATHY

There is a strong tendency to experience the faces of others not only in a visual but in a tactile way, by identifying one's own facial muscles with the tension and relaxation perceived in the face of another. Both the physical and the visual contribute to a language of gesture. Hochberg mentions the role of empathy in the example above as a physiological response to the cartoon representation of the hands. Gombrich comments on this reaction by identification and its effect on the coding of facial distortion:

> "The role of our own bodily reaction in the experience of equivalence may also help to account for the outstanding features of caricature, its tendency to distortion and exxageration: for our inner sense of dimensions differs radically from our visual perception of proportion. The inner sense always exaggerates. Try to move the tip of your nose downward and you will feel you have acquired a very different nose while the actual movement you achieved was probably no more than a fraction of an inch."
>
> Gombrich (1972)

Therefore, our amazing ability to make minute distinctions among the thousands of similar patterns formed by the faces we see may be due to the intimate tactile experiencing of our own faces, amplified by the muscular response which codes the contortions we ourselves would have to go through to BECOME that face. This idea implies an individual point of reference in both face recognition and caricature.

> "You should also understand beforehand that every caricature is also a bit of a self-portrait of the cartoonist and that no two caricaturists will do exactly the same kind of thing with the same subject."
>
> (Richardson, 1977)

4.0 CARICATURE AS TRANSFORMATION...A DEFINITION

Caricature exploits facial plasticity to the most radical degree in that it encompasses not only all possible anatomical dimensions of expressive facial change, but also all imaginable distortions of human faces. The expression which grotesquely but briefly distorts the face is but a gesture in the direction of the structural distortion that is caricature.

Louis Leopold Boilly's "Les Grimaces", 1823 (from Lucie-Smith, 1981)

> "However regular we may imagine a face to be, however harmonious its lines and supple its movements, their adjustment is never altogether perfect: there will always be discovering the signs of some impending bias, the vague suggestion of a possible grimace, in short, some favourite distortion towards which nature seems to be particularly inclined. The art of the caricaturist consists in detecting this, at times, imperceptible tendency, and in rendering it visible to all eyes by magnifying it. He makes his models grimace, as they would do themselves if they went to the end of their tether."
>
> (Henri Bergson,1956)

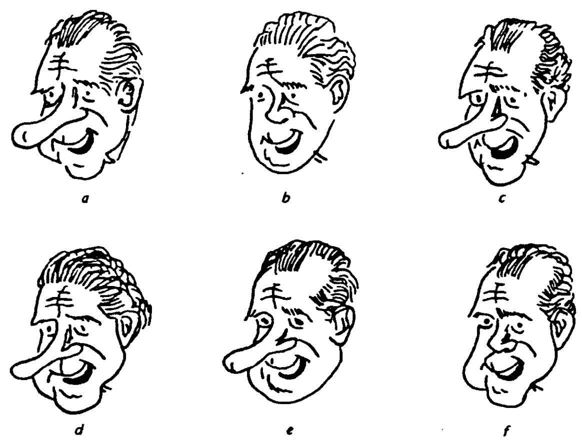
In order to be perceived as a distortion, a caricature must have at its origin some archetype which provides a reference point or context. As we will see later, this archetype may consist of an "ideal" face, a statistical average, or a mental model based on the caricaturist's visualization of his own face. It is with respect to this reference point that caricature amplifies a face. Therefore the amusing common description of a caricature as being more like the face than the face itself is not made with tongue in cheek.

In section 5.0 I will examine the activity of caricaturing, in order to implement this process in a simplified form on a computer. But first it is necessary to construct a definition of caricature.

4.1 EVIDENCE

Perkins and Hagen (1980) evaluated a range of hypotheses about the links between face recognition and caricature. In the experiment testing the transfer value of caricatures and photographs, recognition was much better than chance, but not dependable: subjects correctly identified caricatures only 1/3 to 1/2 of the time after seeing a photograph, and transferred somewhat less accurately to photographs from caricatures. This evidence tends to refute the theory that a caricature is a superportrait, and the theory that people code and recognize caricatures in exactly the same way they do faces.

More promising is the "selection" theory in which the recognition of faces requires attention to key attributes as well as exact metric detail, whereas the recognition of caricature requires attention to key attributes while ignoring the negation of exact attributes. Perkins provided more evidence to support this theory with his (1975) study of Nixon caricatures. He found: a) that almost all successful caricatures of the hapless president contained the same four exaggerated features: distinctive nose, jowls, hairline and chin (to these I would add Nixon's heavy, peaked eyebrows), b) that the CONTRAINDICATION (illustration, following) of any of these features would inhibit recognition (while the LACK of any one of these features would not) and c) that the caricatures could be recognized despite great metric distortion in the rest of the face. This evidence is also consistent with the phenomenon of apprehending a likeness to a friend's face in the face of a stranger, while not actually mistaking his identity.



The contraindication of features, from Perkins, 1975 (left); my Nixon, from the Cornell Daily Sun (below)

Perkins, in his definition of caricature (1975), described two essential elements: individuation and exaggeration. These two elements are fundamental in our approach to the caricature generator. In that the grotesque heads drawn by artists such as Leonardo da Vinci are either pictures of "types" or experiments in the plasticity of the human face, they do not fit the first criterion. As for the second criterion, Perkins differentiated exaggeration from mere distortion in that the transformation must be done in relation to some norm:

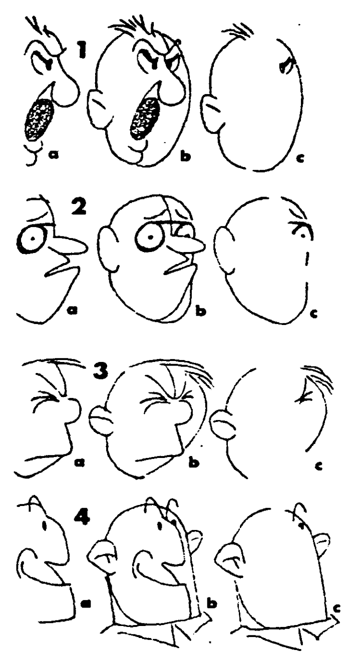
> "Idealization seems intuitively the very contrary of caricature. Both depart from faithful portraiture, but somehow in opposite directions."
>
> (Perkins, 1975)

This notion of exaggeration from a norm is one which I have implemented in the caricature generator.

4.2 THREE-DIMENSIONALITY

A good caricature amplifies what is remarkable about the face. This information is derived from comparisons with other faces we have seen; because of experiences with binocular vision and seeing motion, our own inner database about human faces is one based on volumes and activity, not on flat, still images; therefore, unless the distortion is to be only superficial, a caricature should represent an orthographic projection of either real three-dimensional volumes or of limited three-dimensional volumes compiled from multiple viewpoints, exaggerated. A good caricature may contain more information than a photograph in that it takes visual information from several points of view or from a three-dimensional model of the face and codes it into a single two-dimensional image.

Perkins comments that such shapes as the nose are hard to represent in the frontal view, and that many cartoonists resort to placing a 3/4-view nose on the frontal caricature of a face (Perkins, 1975). This is a common convention. Hamm suggests actually drawing the profile outline within the frontal view.



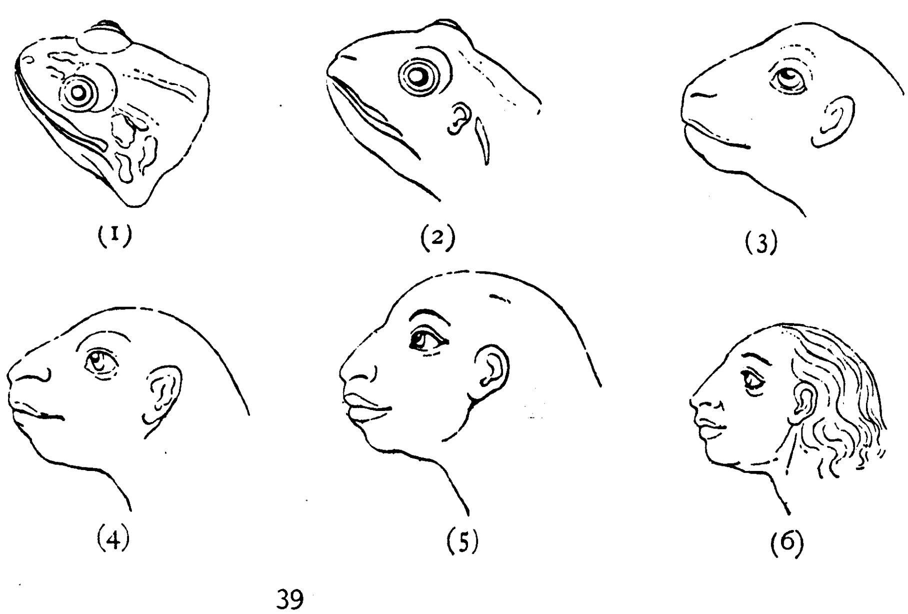
Hamm, 1967

Consequently, a caricature generator which compares a face to a norm of some sort would benefit from three-dimensional input. In a study such as Perkins' on the transference between caricature and other modes of representation, it would have been informative to expose subjects to the actual people (either live, or with head motion on videotape) who were caricatured, to see if recognition is aided when the viewer has a three-dimensional model of the face.

4.3 ANIMALS AND OTHER METAPHORS

There is frequently an element of double reference in caricature, which amounts to a visual pun. Frequently this double reference takes the form of an animal. The use of animals in caricature and cartooning can be inspired by visual likeness, or it can symbolize some personality characteristic or political role of the subject. There are many amusing cartoons where animals are personified; these do not qualify as caricatures in the strictest sense.

In the 1600's, Charles Lebrun illustrated theoretical and extreme likenesses between animals and men (see page following). Later physiognomist Charles Lavater anticipated Darwin's theory of evolution when he published transformations of animal faces to human in 1775.

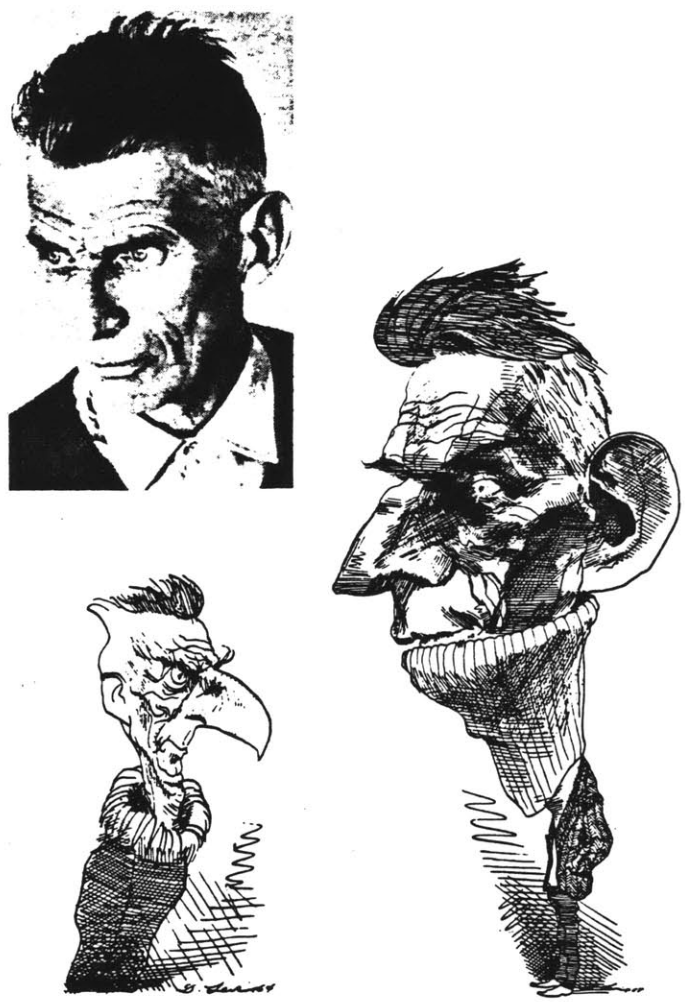


a transformation by Charles Lavater in 1775 (from Hillier, 1970)

Charles LeBrun's heads (from Sorel, 1980)

Perkins provides as an example of an animal form used as an individuating mechanism and as a personality symbol these caricatures of Samuel Beckett done by David Levine. Beckett is recognizable in spite of the fact that he is drawn as a buzzard. Beckett's eyes and nose seem to lend themselves beautifully to Levine's buzzardish interpretation, although the double entendre is more a superimposition of certain personality characteristics of Beckett as reflected in his work. This is a two step process; first Beckett's features are amplified, and then he is transformed into a buzzard.

Illustrations from Perkins (1975): clockwise from top: photograph of Beckett, Levine's portrait caricature of Beckett, combination of caricature and buzzard



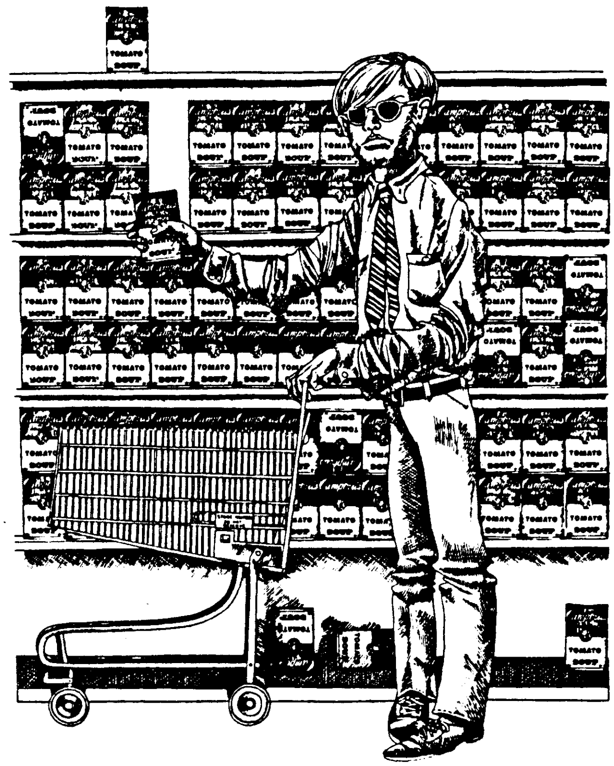
Caricature mapped onto animal (below): Robert Gaillard's caricature of Napoleon III (from Lucie-Smith, 1981)

4.4 OTHER ELEMENTS IN CARICATURE

> "caricature...integrates various forms again, reduces their differentiated detail to a few strong lights and linear signs, one of whose qualities is their ambiguity. From this ambiguity caricature derives many of its surprise effects: a form that can be read or interpreted in many different ways is the appropriate medium for a joke." (Hofmann, 1957)

Because caricature leaves out details, it is more symbolic, less representational than a portrait or photograph - and therefore it is open to a wide variety of interpretations and comparisons with objects. Obviously, most human caricaturists do much more than simply distort faces. Political cartoonists supply a highly subjective context for their drawings.

The fact that many caricatures are funny underscores the effects of the visual ambiguity created by caricature. Freud (1905) refers to the comic in caricature as bringing out hidden features by exaggeration, creating bizarre juxtapositions which would otherwise be unnoticed by the senses or repressed by the mind. Minsky (1980) illuminates the comic with his notion of frames. Previously acquired description structures interpret perceptions by representing them as stereotypes. There is a network of cognitive connections among frames, and abrupt shifting within this network can surface as nonsense, inspiration, or humour. Frame shifts can provide the cartoonist with endless variations in context, and the viewer with delight; caricatures frequently employ many other levels of representation superimposed upon that of the exaggerated portrait. These levels include allegory, comment on social class, political symbolism, surrealism, substitution of the anatomical parts of animals and objects, contradictory scales, verbal puns, etc.

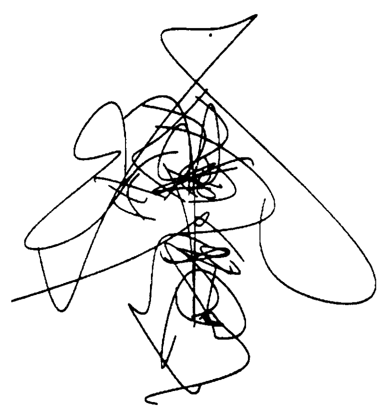


Mr. Andy Warhol Fetches a Work of Art Through a Metaphysical Barrier (Kenner,1973)

The idea that a face may be described to the memory in terms of its departure from a norm is closely related to the fact that things are defined by their contexts; when the frame is changed, the object has a different meaning. The transformation applied to the face which yields a caricature, yields the norm when applied in reverse. Depending on the world around one at the time, one is giant or midget, humane or bestial. Jonathan

Swift (1726) used shifting ideals which he personified as Lilliputians, Brobdingnagians, and Houyhnhnms to amplify the features of his adventurer Gulliver.

The degree to which a machine would ever be able to serve the ends of politically sophisticated caricature is obviously limited by the degree to which an intelligent machine will ever have an opinion of its own. This statement obviously holds different meaning for different people. Ultimately, the intelligent caricaturing machine would be able to draw analogies between facial forms and the shapes of other objects, and to selectively transform one into the other. This idea of transformation implies accomplishment in stages or degrees; just as Leonardo da Vinci drew individual faces knowing that their shapes fell somewhere within the full range of possibile human faces, caricaturists are aware of the various degrees to which they may distort their subjects' faces before they become unrecognizable.



FACELESSNESS:
Our interactive caricature generator has no limits governing the extent to which it will distort a face with respect to a given norm. Here is an extremely exaggerated case.

4.5 A DEFINITION

In summary, the essential points about caricature with which we will concern ourselves in making a computer graphics caricature generator are as follows:

1. A caricature refers to a specific individual.

2. A caricature is a loaded portrait, one which amplifies certain features which are key to recognizing the face.

3. Caricatures, unlike line drawings, are impervious to slight metric distortion.

4. Caricaturing is done with respect to some ideal, whether conscious or not.

5. A caricature exaggerates the face by comparing it to this ideal and then distorting it in the opposite direction.

6. A caricature is more likely to be successful if it can incorporate several points of view, thus including three-dimensional information in a two-dimensional projection.

The following section looks at the process of caricaturing.

5.0 THE HEURISTICS OF CARICATURE

It is the stated goal to develop a theory of caricature with the eventual end of implementing the associated algorithms in a caricature generator. With the caveat in mind that what individuals do and what they say they do are often two entirely different activities (a distinction comparable to that between the sacred and the profane), I have undertaken to use artists and cartoonists as informants in examining the processes involved in caricaturing. Some of these personal algorithms will be evaluated and implemented in the final sections of this paper.

5.1 SOME MUSINGS BASED ON PERSONAL EXPERIENCE:

What are the ways to caricature ?

from memory, with eyes half-closed, after four beers, during hysterical laughter, furiously, furtively. This way always yields the best results.

from photographs, under a deadline, to please someone. This way results in an adequately recognizable drawing.

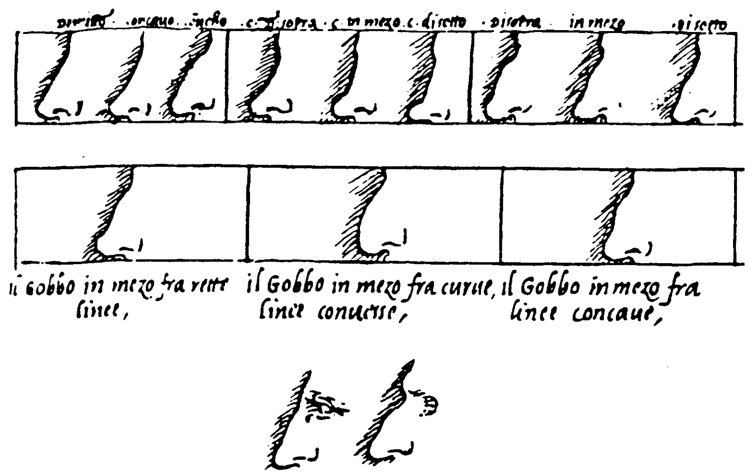
from real life. Strangely, it is sometimes difficult to caricature someone in person without making a sketch or portrait instead of a caricature. If it is someone one does not know, one tries to record too many things at once. This way may be unsuccessful because it circumvents one's own personal filter, the memory of the subject's visual presence that develops after seeing him or his image or his movement.

5.2 PORTRAITURE

> "If Nature had a fixed model for the proportions of the face everyone would look alike and it would be impossible to tell them apart; but she has varied the pattern in such a way that although there is an all but universal standard as to size, one clearly distinguishes one face from another. "
> (Leonardo da Vinci's Advice To Artists, Kelen, 1974)

Artists such as Leonardo da Vinci and Albrecht Durer were obsessed by ideals of ugliness as well as by beauty, and sketched many variations of the human face. There is no evidence that these sketches were meant to be portraits of individuals, so in the strictest sense, they are not caricatures. However they were conscious of the fact that there are certain human variations possible with respect to some ideal. Leonardo advised artists to observe and remember four principle variations in the profile: the nose, mouth, chin, and forehead. He describes some possible variations:

> "First the nose: there are three different sorts, straight, concave, and convex. Of the straight there are but four variations, short or long, high at the end or low. Of the concave type there are three sorts, some with the concavity above, some in the middle and some at the tip. The convex noses also vary in three ways, some projecting in the upper part, some in the middle and others at the bottom. Nature delights in infinite variety, and gives again three changes to those noses which project in the middle, for some are straight, some concave and some convex.
> (from Kelen, 1974)



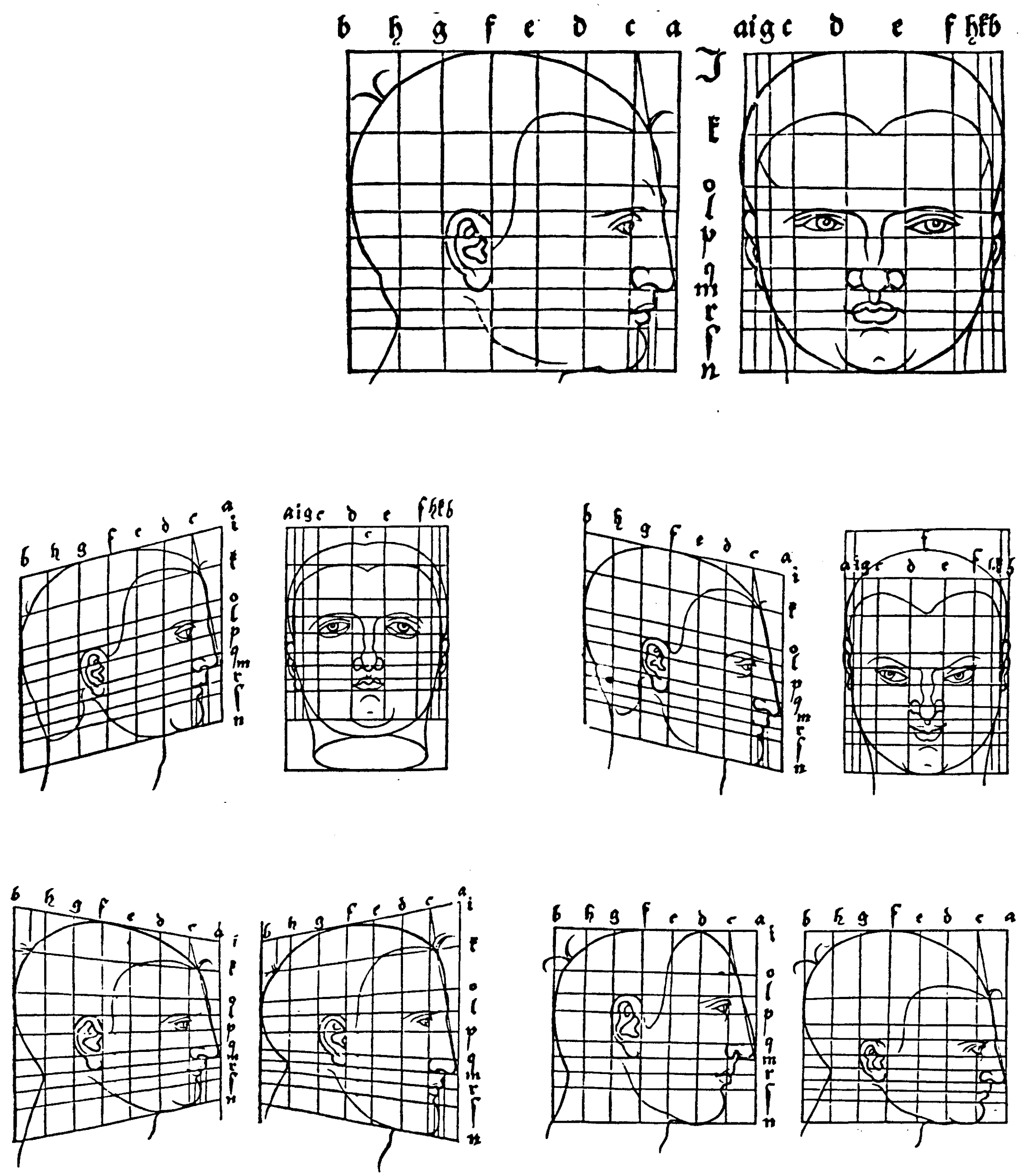
Leonardo da Vinci's noses

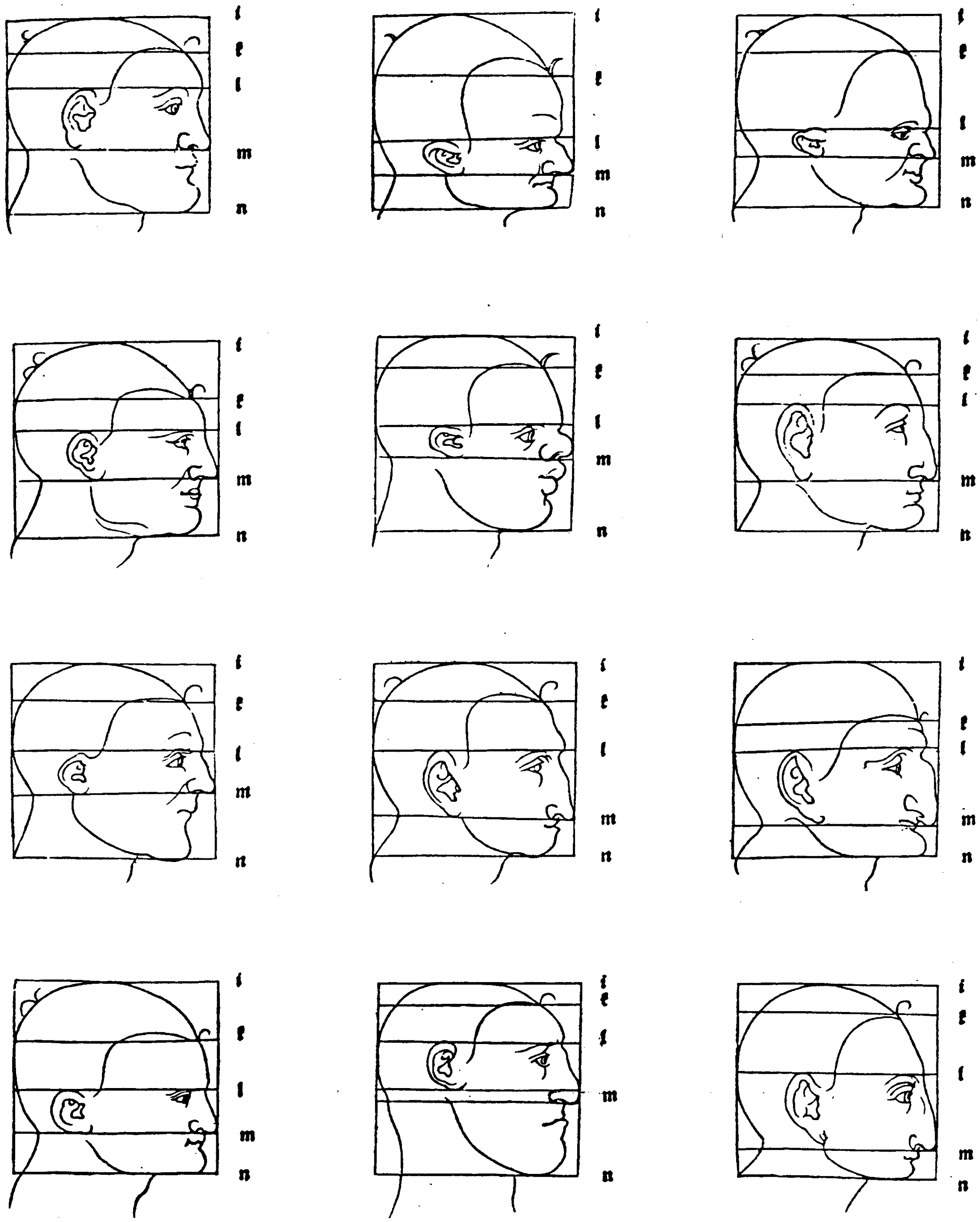
Leonardo went into additional detail for each of the types of noses mentioned above, and for the rest of the face as well. His classification scheme for analyzing and remembering faces is perfectly suited to the sort of branching decision structure that can be used by a computer in constructing a face out of primitives. Later we will evaluate the potential of algorithms like this one for synthesizing faces in computer graphics.

Other artists such as Albrecht Durer, Oskar Schlemmer, and William Rimmer have described schemes for learning the proportions of the human form and face by studying what were considered ideals of their time. It is an interesting exercise to compare these ideals, some of which are fairly realistic and others of which are highly stylized. The Schlemmer ideal contains a mouth which few human faces could approximate. However, in classical painting and sculpture these forms are accepted as the reigning ideals of beauty.

Nineteenth-century artist William Rimmer listed some of the

Artist Albrecht Durer described faces by transforming the coordinates of an ideal face. Reprinted from PROPORTIONSLEHRE (1528), on this page and the next.





proportions of the average male head:

> "Head an egg-shaped oval. Dividing the head transversely into equal parts, the eyes should be above the median lines: there should be the width of an eye between the eye and the eyebrow; eyebrow highest and widest over the outer third of the eye; front forehead terminating in the outer rim of the orbital circle; the length of an eye between the eyes; the length of an eye across the wings of the nose. The mouth one half wider than the nose; the under lip thicker than the upper lip..." (Rimmer,1877)

Taking into account the idea that there is more than one "ideal", we are faced with a range of choices in constructing a caricature generator which exaggerates with respect to some norm. We can accept any of the ideals from a given artist or school of painting or even that from a given culture; we can construct a stistical norm from a population of heads to which our subject belongs; we can personalize the model by using our own faces as points of departure.

> "Take care to use the best features of faces whose beauty is established by popular agreement rather than by your own particular taste, otherwise you might end up painting over and over again faces that resemble your own - since it is a fact that such similarity pleases us. Then, if you were ugly, you would not be selecting beautiful faces but ugly ones - and that is true of many painters whose types resemble their master."
>
> (Leonardo, from Kelen, 1974)

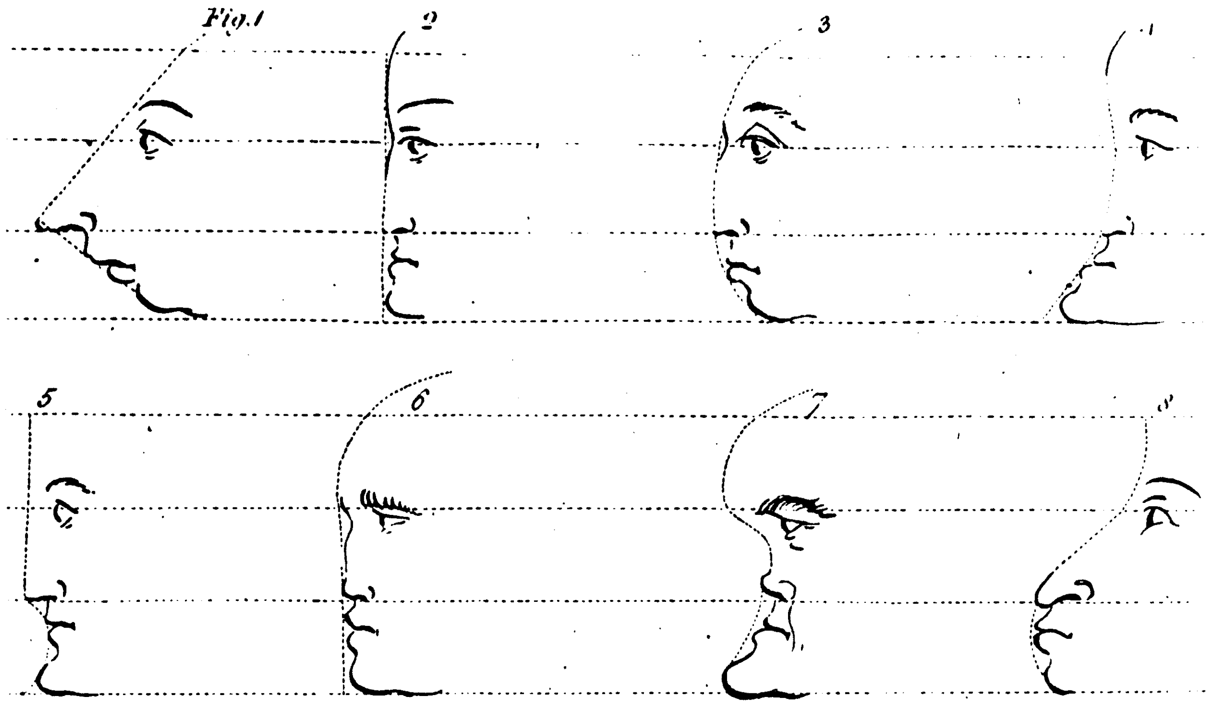
5.3 RULES FOR DRAWING CARICATURES

In 1791 William Hogarth's THE ANALYSIS OF BEAUTY was published which included Francis Grose's RULES FOR DRAWING CARICATURAS.

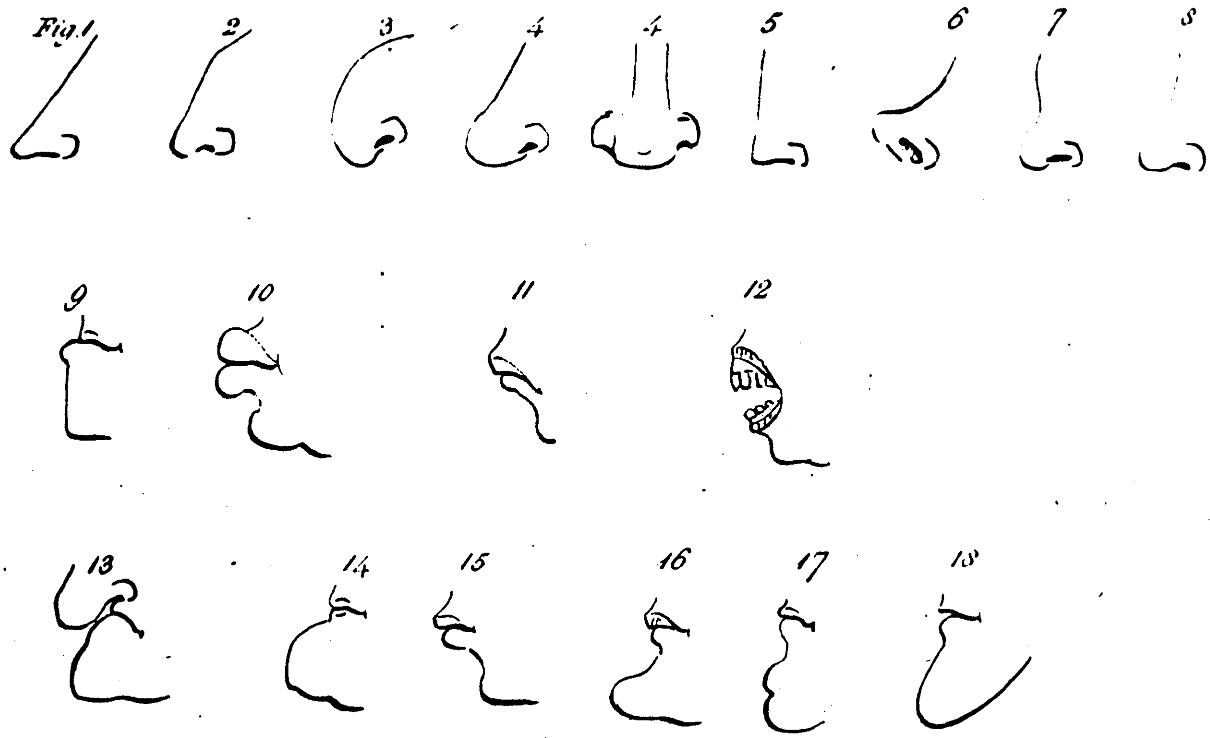
Grose favored a piecemeal way of analyzing the face, and suggested that the caricaturist classify and memorize the elements of the face as a student memorizes Latin. He provided examples which were to be considered as "mathematical diagrams". The following is exerpted from his treatise:

> "The sculptors of ancient Greece seem to have diligently observed the forms and proportions constituting the European ideas of beauty; and upon them to have formed their statues. These measures are to be met with in many drawing books; a slight deviation from them, by the predominancy of any feature, constitutes what is called Character, and serves to discriminate the owner thereof, and to fix the idea of identity. This deviation or peculiarity, aggravated, forms Caricatura...
> On a slight investigation it would seem almost impossible, considering the small number of features composing the human face, and their general similarity, to furnish a sufficient number of characterising distinctions to discriminate one man from another; but when it is seen what an amazing alteration is produced by enlarging one feature, diminishing another, encreasing or lessening their distance, of by any ways varying their proportion, the power of combination will appear infinite.
> Caricaturists should be careful not to overcharge the peculiarities of their subjects, as they would thereby become hideous instead of ridiculous, and instead of laugher excite horror. It is therefore always best to keep within the bounds of probability. Ugliness, according to our local idea, may be divided into genteel and vulgar. The difference between these kinds of ugliness seems to be, that the former is positive or redundant, the latter wanting or negative. Convex faces, prominent features, and large aquiline noses, though differing much from beauty, still give an air of dignity to their owners; whereas concave faces, flat, snub, or broken noses, alays stamp a meanness and vulgarity. The one seems to have passed through the limits of beauty, the other never to have arrived at them.."
>
> (Francis Grose, 1791)

Reprinted from Grose, 1791



Rules for drawing Caricaturas Pl.I



5.4 CARTOONISTS

Conventionally, the distinction between caricaturists and cartoonists is a blurry one in that the same artist may do both portrait caricature and humorous cartoons. The following heuristics are those of cartoonists trying to capture the likeness of a particular subject.

Many cartoonists keep a file of photographs of public figures from which to work. Spencer (1949) suggests doing caricatures from multiple pictures of the subject "talking, laughing, frowning, and front, right, and left profiles..." rather than from memory. This dictum would seem to be justified if the caricaturist has but a passing familiarity with the face of the subject, or if he has seen only one viewpoint of the face from which he is unable to form an impression of its volumes. Nelson (1975) suggests as an alternative to drawing a caricature from a photo that one study the subject, paying attention to the shape of the head, hair, and any outstanding facial features, and then draw from memory. This method makes sense in that it uses the cartoonist's natural ability to encode faces and prevents his making a literal sketch of the face which might turn out not to be a caricature at all in that it may contain extraneous details which do not represent the amplification of key features, but merely noise.

Cartoonists as well as portrait artists have frequently used generalized anatomical models for drawing facial proportions. Richardson's "how-to" approach to caricaturing (1977) begins with the normal proportions of the head (from classical portraiture) and deviates from this norm. He describes an average coordinate system for faces, where the eyes are halfway down the face, the ear extends from the tip of the nose to the eyebrow or so, the eyes are one eyelength apart, the pupils fall directly over the corners of the mouth, and the edges of the nostrils correspond to the inner eye corners. Variables are the length of the nose and the positions of the lips.

> "We are so used to customary scale relationships in faces that even slight changes will produce marked effects."
> (Richardson,1977)

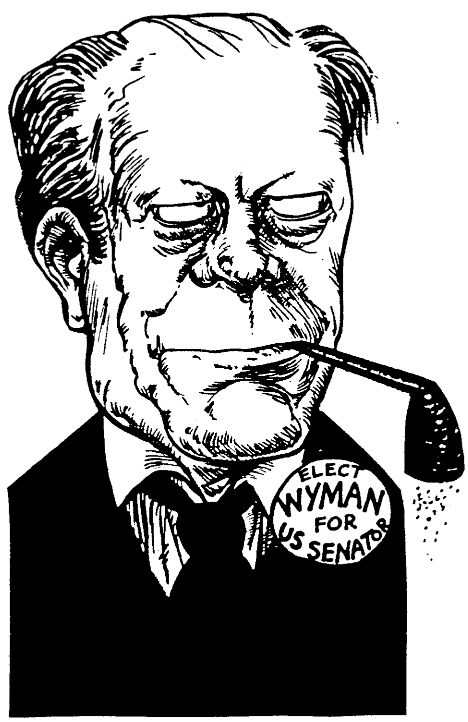
Hamm (1967) outlines the following proportional rules: the average head is approx five eyes wide (with a little trimmed off the first and fifth segments), and that the eyes and mouth fit into an equilateral triangle. His personal algorithm supports our approach to a caricature generator:

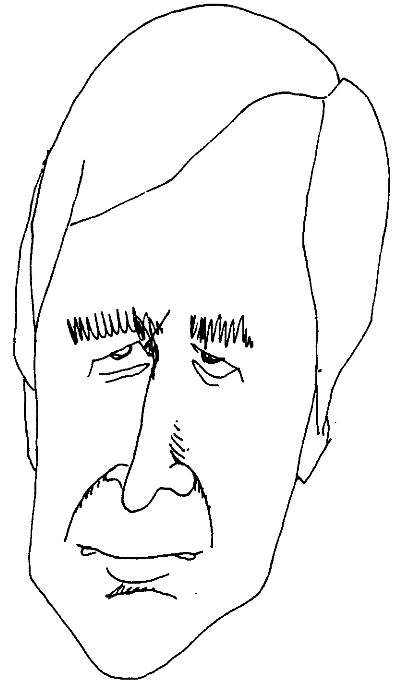
1. Obtain good likenesses of the subject.
2. Decide on the unusual aspects of the face.
3. Play these up; at the same time minimize or omit the rest.

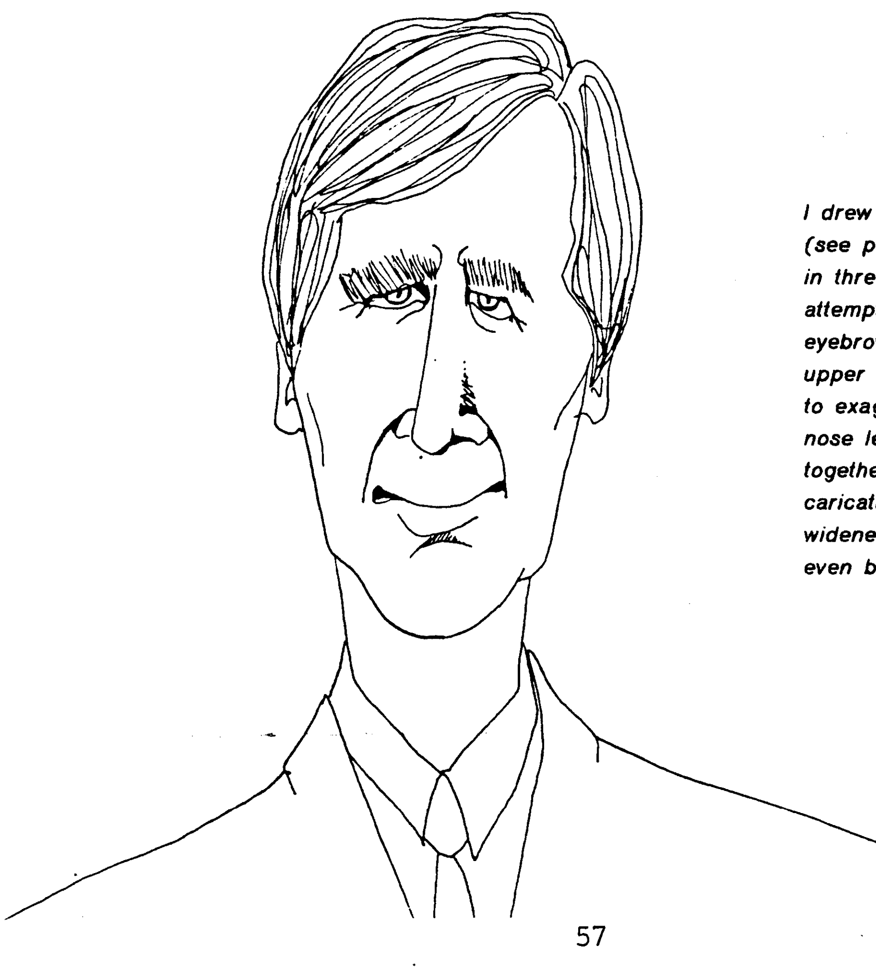
> "A perfectly 'normal or regular' face is difficult to caricature."
> (Hamm, 1967)

Many political cartoonists had a great deal of difficulty shortly after Gerald Ford's inauguration as President because it seemed at first that he had no facial characteristics that, when exaggerated, would make a drawing immediately recognizable to the American public (one could argue that the caricaturists had grown lazy, at a point when they did not have Richard Nixon to kick around any more). The symbols of Ford's office and of his exaggerated clumsiness were exploited. Eventually most syndicated cartoonists settled on certain large, rounded facial volumes such as his forehead, upper lip and chin, and on his deep-set eyes as requirements in a Gerald Ford cartoon.

left: Boston Globe cartoonist Paul Szep's caricature of Ford; right: my Ford, done from six newspaper photos shortly after his inauguration.

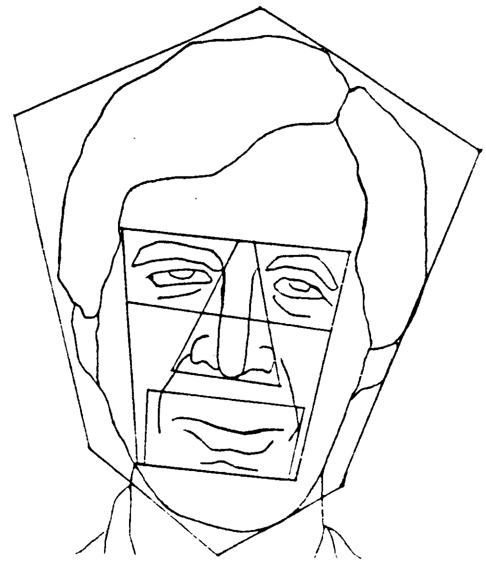






SOME
HEURISTICS

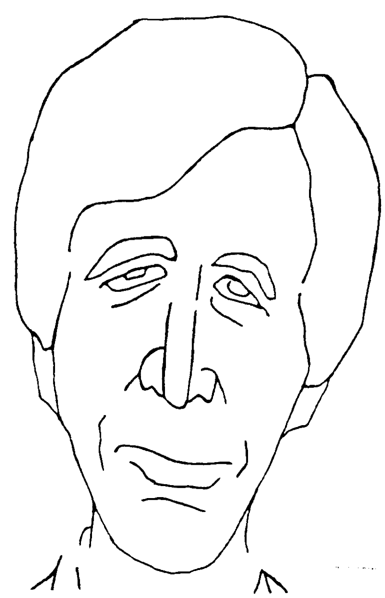
I drew this caricature by hand (see photo of subject, page 95) in three steps. Upper left: an attempt to capture the bushy eyebrows, tilted eyes, and long upper lip. Upper right: attempt to exaggerate the shape of head, nose lengthened, eyes closer together. Left: the finished caricature - the forehead widened, jaw narrowed, eyebrows even bushier, eyes even closer.



Above: traced line drawing of subject on page 95. Note that this is a fairly bad line drawing, not particularly recognizable. By placing a face-shaped grid over the drawing and distorting it very simply, step by step, the "caricature" at right was produced in just five steps using a collaging program.

1. *lengthen face*
2. *lengthen whole head*
3. *rotate eyes*
4. *widen mouth*
5. *widen forehead and narrow jaw*

The automated approach to caricature is based on having the computer make similar judgements about facial measurements and then distort the face along those dimensions where it differs the most.



6.0 AUTOMATIC AND MAN-MACHINE SYSTEMS FOR FACE RECOGNITION AND REPRESENTATION

The following section is a brief survey of systems which have used computers to recognize, distort, or animate the human face.

The problem of automatic feature-finding within faces in an unconstrained environment is not a trivial one - in fact, even finding the face is not easy. Assuming that the image at hand is a face, the next step is to come up with a description that is complete enough for the caricature generator to use in exaggerating the face, a description that is reasonably consistent (although not nearly so powerful as) the primal sketch (see Marr) made by the visual system through local processing mediated by top-down information about such things as knowledge about structure and context.

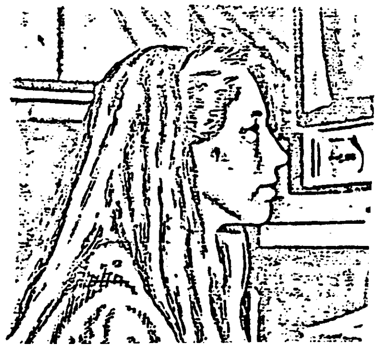
The automated systems which have located and manipulated points on pictures of faces have sometimes relied on combinations of template-matching and environmental uniformity with respect to such factors as scale, location and lighting, or have used man-machine interaction in which an operator selects features or points on the face.

6.1 LINE DRAWINGS

Before the face can be recognized, manipulated or warped it must be consigned to a data structure. Some of the systems which have attempted face recognition have reduced the image to a line drawing and then used some kind of algorithm to locate points on the face. Deriving connected, line drawings from a noisy image is in itself not an easy task.

This can be done computationally by passing a filter over the image which convolves it with an (edge-finding) operator. Marr and Hildreth (1980) developed an optimized edge filter which finds intensity changes at different scales using a Laplacian operator. In this system a CCD camera (which is less subject to distortion than a vidicon) is used to digitize a two-dimensional intensity array off a mirror which scans the face in front of it vertically. Convolution hardware between camera and the digitizer yields images such as the following ones, so that the raw image which is "grabbed" in this way is actually a rough line drawing of sorts consisting of zero-crossings.

Digitized face convolved at two scales (thanks to Keith Nishihara and the M.I.T. A.I. Lab)



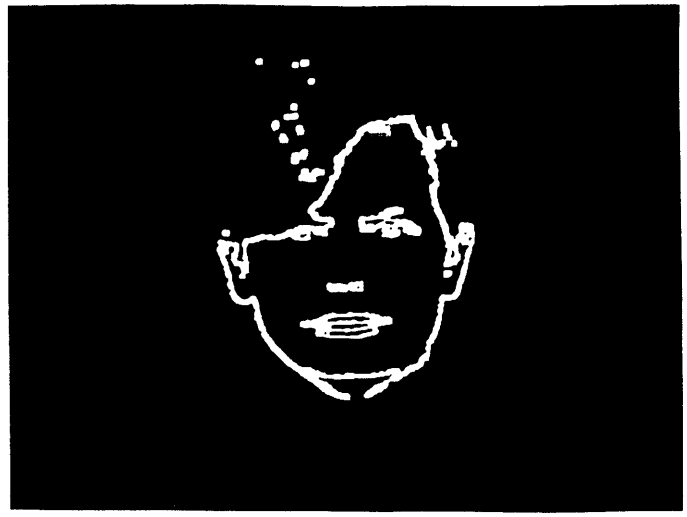
Combination of the two previous convolutions (left)

The image is then digitized as a grayscale image and a difference of Gaussian filter is applied. The effect of the optimized filter is to emphasize edges. Another program segments the lines, and then the image is thresholded. The following line drawing is an example of the result:

Above and right illustrations, courtesy of Keith Nishihara

A more ad hoc method of computationally deriving line drawings by filtering images uses a succession of 3 x 3 filters to emphasize edges. Horizontal and vertical passes are combined into one image, and the lines are normalized with regard to thickness. Thresholding is done, and some of the noise is thrown out. This method is not spatially accurate, and the result depends on the orientation of the edge with respect to the filter.

Ad hoc line drawing, obtained after thresholding by eye and manually combining filtering passes (thanks to Steve Gano)



If there are no constraints limiting the degree of man-machine interaction during the input stage to the caricature generator, one can use the method of rotoscoping, a technique used frequently by character animators such as those in the Disney studios to capture certain realistic aspects of structure or motion for use within a cartoon image. The user traces over a picture of the face by drawing on a graphics tablet while watching a cursor on a screen which displays the digitized picture, or if the screen is a touch-sensitive device, he can trace directly on the screen with his finger. This method is more time-consuming and is vulnerable to slight metric distortions which make the line drawing an inaccurate projection of the face. These distortions are probably random (caused by tracing error) but it is possible that they are somehow stylized, representing the selective distortion of certain information (in other words, perhaps the user makes an unconscious caricature). Based on the theory of caricature developed herein, the caricature generator should not care about metric distortion as long as it is not great enough to destroy key features.

6.2 COMPUTER RECOGNITION OF FACES

Since the mid-sixties, computer recognition of human faces has been investigated with such applications in mind as matching suspects to photograph files in law-enforcement agencies, or providing automatic individual access in security systems. What human beings are able to do routinely has proved to be a sophisticated problem in picture processing and pattern analysis. W.W. Bledsoe (1964,1966) described the problems encountered in automatically analyzing three-dimensional faces using such variables as orientation, direction and degree of illumination, facial expression, age, and picture quality. He avoided the problem of automatic feature-finding by having human operators locate 46 points on each face using a tablet, and then computed 22 Euclidian distances from these points. Variations in image quality and illumination were taken into account. A three-dimensional model of the face consisting of vectors was used to geometrically transform those input photographs of faces that were translated, rotated, tilted or scaled, so that they could be compared with frontal sets of facial points. All measurements were normalized on the subject's interocular (interpupillary) distance. He concluded that, once the face points were accurately designated and geometrically transformed, machines were superior to human beings in recognition across large age differences (1968). He suggested that it may be possible to automate the location of the points (1966), that a system taking stereo photographs would be useful (1966), and that information about facial contours would prove valuable (1968). He acknowledged a dual strategy in face recognition:

> "The procedures used by humans to identify facial photographs are far from clear, but it appears that a combination of global and local cue-matching is often employed. The global match might check such things as the aspect ratio of the face, while the local match checks the detailed structure of the features."
>
> (Bledsoe and Hart, 1968)

In another attempt to identify people by computer, Michael D. Kelly (1970) automated the feature-finding process using pictures of whole bodies in a more controlled and uniform photographic environment. He emphasized goal-directed picture processing and the use of top-down knowledge in finding features - ie, knowing in advance what one is looking for, and reducing search time and error by looking only in those areas where the feature most probably exists. He used dynamic threshold setting to eliminate noise while maximizing such patterns as the light-to-dark transition that characterizes the white and pupil areas of the eye. His system incorporated template matching and edge detection operators. Using a human operator to set input contrast ratios, he succeeded in automatically identifying members of a small population.

The most promising attempt at automatically finding points on the face was that of Kanade (1973), who used a combination of local and global processing with the goal of automatic face recognition. He digitized pairs of images of subjects, all taken from the same point of view (frontal). Binary pictures which resemble rough line drawings were derived using a Laplacian operator. These faces were then analyzed using a clever template-matching scheme which was successful because it knew where to look. The algorithm searched for a prescribed list of points on the face in the following way:

1) first it passed a horizontal rectangular slit down over the top of the image, taking the integral projection (histogram of intensity values) until it yielded a peak which represented the top of the head.
2) Then, using this location to predict where to look next, it took another, larger histogram across the face, shifting until it located the pattern of peaks and valleys which indicated the locations of the cheeks and bridge of the nose.
3) Continuing to predict, search, detect, and evaluate according to the template at any given point, it proceeded to move histograms around to locate points until it failed in a given step; at this point, it went back to a previous step, relaxed the parameters guiding that step, and repeated it. Hopefully, this recursion enables an accurate prediction to be made, so that the failed step can be performed correctly.

The feature points Kanade located were as follows: (1) top of head, (2) cheeks and sides of face, (3) nose, mouth and chin, (4) chin contour, (5) face-side lines, (6) nose lines, (7) eye, and (8) face axis. The program managed to estimate point information that was incomplete due to broken or missing lines by exploiting the symmetryical property of the human face. From the location of the points, measurements were taken and faces compared.

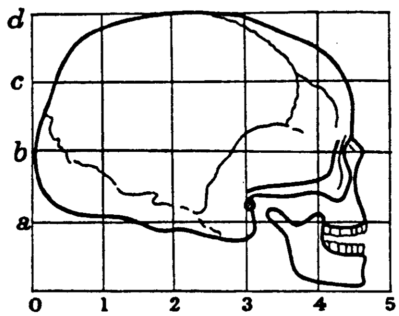
The fact that this program was successful 75% of the time in automatically recognizing faces from a population of 20 people is encouraging.

Harmon's system (1973) used man-machine interaction to classify faces according to numerical judgements of 21 features, which were recorded as the computer's statistical model of that face. He found that these features were sufficient to differentiate among the sample population of 256 white males. The two most similar faces illustrated in his article would not be mistaken for one another, however. Obviously people use more features than the ones used here by the computer to recognize faces. He presents evidence that people recognize extremely blurry pictures and block portraits as spatially significant patterns.

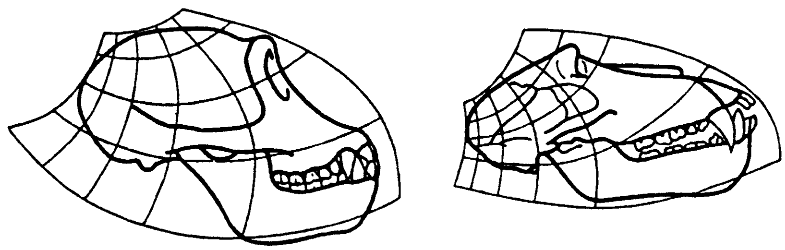
6.3 COMPUTER-WARPED FACES

Simply as a two-dimensional projection, a face can easily be warped from one coordinate system to another. Pittinger, Shaw and Mark (1979) depicted aging as transforming the invariants of the face as a result of cardiodal strain. For example, children have relatively large foreheads, eyes set far apart, and small chins while adults have smaller foreheads and larger, protruding chins. In this study they digitized profiles and warped these accordingly to generate pictures of structural changes that take place in the human profile when years are added or subtracted. The pictures predicting the results of aging are quite plausible. They then applied the same transformation to cartoon pictures of birds, dogs, and even volkswagons, and showed these pictures to subjects who were asked to judge the age of the thing pictured. From the point of view of perceiving an invariant, the evidence from these experiments supports their claim that a visual transformation which resembles that of cardiodal strain is related to the perception of the relative age of the object.

However, that cardiodal strain and particularly the force of gravity is actually responsible for the structural transformation of the profile during aging (Todd and Mark, 1981) seems highly unlikely; if that were true, the profiles of people who were chronically bedridden in their growth years (and thus subject to the forces of gravity in different directions than other people) would be more childlike. It is far more likely that the structural change brought about by aging in human faces is something we learn to recognize through experience; the widespread human tendency to imbue animals and objects with human personae accounts for the perception of age in these non-human representations. The fact that cardiodal strain might cause similar changes is probably coincidental.



Coordinate system of human skull, warped onto chimpanzee (lower left) and baboon (lower right). From D'Arcy Thompson, first published in 1917.



Another system which transforms the human face by warping a coordinate system is a collage program devised by Paul Heckbert at New York Institute of Technology. The user picks up points on a rectangular grid using a tablet and reorients them. The computer maps the affected areas of the picture back onto the new grid. By trial and error it is possible to create an amusingly distorted photographic caricature, although the process is not automatic and depends on the skill and imagination of the user.

A computer program which I call WARP places a face-shaped coordinate system over the image of a face by having the user touch points on a touch-sensitive display. This program subsequently can warp the corresponding areas of any other face to fit into the coordinate system of the first face. For the purpose of caricature, it can also be used to distort patches of the face and reposition them on the same face.

6.4 COMPUTER-SYNTHESIZED FACES - TWO-DIMENSIONAL PROJECTIONS

There have been a number of attempts to synthesize faces in computer graphics. Some have used digitized projections of faces or digitized primitives representing a limited feature set, while others have used sophisticated databases.

Gillenson and Chandrasekaran (1975) used stored primitives in a computer graphics system to help artistically untrained users to produce a recognizable image of any white Caucasian male while looking at a photograph. The program, known as "Whatsisface", began by displaying a statistically average, white male face, mathematically calculated from Harmon's population of faces in his 1973 study of recognition. The user then proceeded to modify this face while looking at photograph. The stored primitives consisted of 17 features, all of which could be interactively scaled, rotated, and located on the screen using a hierarchical manipulation scheme. The resulting recognition rates between photographs and face drawings created by people with a wide range of artistic skill were 62% for free-hand drawings and 87% for computer-aided images.

This study is interesting for two reasons: it used a hierarchical method reminiscent of Leonardo's mneumonic method for faces, mentioned in section 5, and it created faces by departing from a norm. The unfortunate thing about this study is that it did not test the subjects' ability to create faces from memory. It tested merely the ability to manipulate a set of primitives to resemble a photograph. It would be more interesting to test the ability to create faces from memory using such a system, and to determine if that were possible, beginning with a visible norm. It would also be interesting to compare the metric dimensions of the computer-drawn faces to those of the photographs, to see if the users tended to exaggerate or caricature any features, rather than copy them.

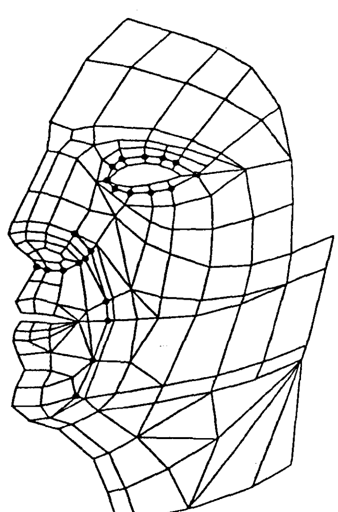
The construction of a remembered face from sets of primitives does not seem to be a particularly promising activity unless the user is allowed to manipulate the facial pattern as a whole. Although some face recognition studies support a hierarchical attribute-checking strategy in that subjects distinguish faces with a greater number of different features more easily, there is neverthelsee evidence that something "wrong" about a face can inhibit recognition altogether. Therefore it may be even more inhibiting to begin with a visible norm.

The "Identidisc" project in progress at the Architecture Machine Group (P. Weil) rejects conventional police artist methods of constructing faces from sets of primitives, and utilizes a more impressionistic method where a witness can view many faces registered on the eyes and stored on optical videodisc and select those faces which in some way resemble a remembered face, without having to analyze them or verbally describe the resemblance. The selected faces are then digitized by a computer and averaged together according to intensity values at any given point, a process similar to the optical process of multiple exposure. The result is a somewhat blurry image which is at least as accurate as the hard line drawing of a face that an Identikit would provide, and which is much less likely to mislead by asserting the shapes of those facial forms which have escaped the memory. Identidisc is less likely to contraindicate information than it is to leave it out. Therefore recognition is more likely.

An earlier computer graphics piece by Nancy Burson at the Architecture Machine Group performed an aging transformation upon digitized faces by scaling, registering and distorting these faces over pictures of elderly faces, intensity-averaging them in such a way as to add wrinkles to the result, and then re-warping the face to its original coordinate system. The intriguing results of this program are not due to any anatomical model for predicting how aging will affect an individual face, but are due to the skill of the artist.

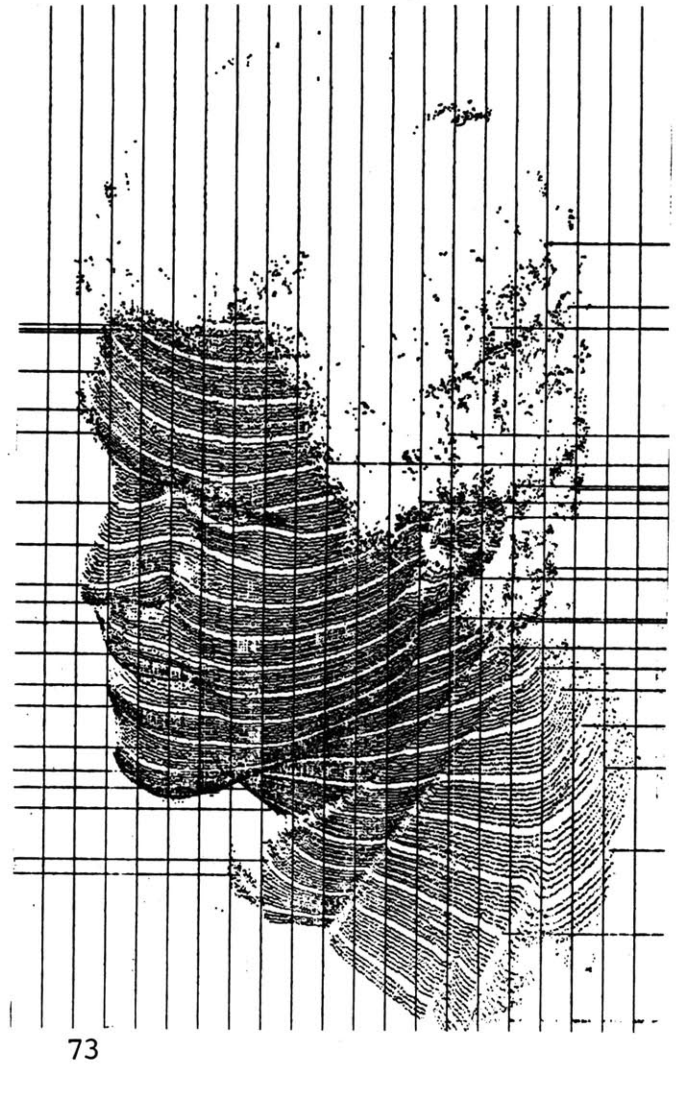
6.5 COMPUTER-SYNTHESIZED FACES - THREE-DIMENSIONAL

Parke (1972) developed a three-dimensional database for creating faces out of shaded polygons and animating them in a way that was sympathetic to the underlying anatomical structure of the face. He developed a parametric model to vary multiple animation transformations over time. His program generated moving, talking heads, a frame at a time, the database of which could also be modified within certain limits to make the head assume the facial structure of a particular individual. Lip sync and feature distortion were accomplished by interpolating from one database to another. His model included on the order of 25 feature nodes, built out of polygons, which were affected by parameters such as width, position, aspect ration, and shape. An elegant part of the work was the way in which the dynamic parts of the face were represented and assembled. This work is currently being continued at New York Institute of Technology, for the purpose of character animation for motion pictures.



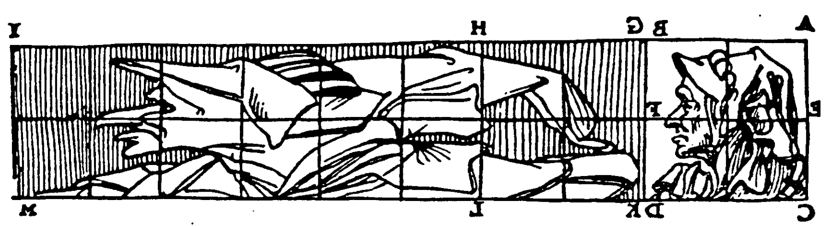
Parke's polygonal face

A company called Solid Photography used an industrial scanner which first digitizes a head from many points of view and then furnishes a database of cartesian coordinates which can then be used to display the head or machine a three-dimensional bust. The resolution of the points is quite good, providing more than 1000 X 1000 points, and more than 300 horizontal cross-sections of the head. This database is fine for display and could be used for making measurements; it would be much more difficult to animate then the Parke head, as the model has no anatomically-based structural knowledge or coherence as a face (in other words, it might as well be a machined part). On the



Three-dimensional model from Solid Photography system (Rongo, 1982) other hand, two of these sets of points could be aligned either by hand or by automatically registering upon the eyes, and the caricature generator algorithm could then perform exaggeration on the three-dimensional face.

In the next section constraints for the caricature generator will be developed using the perspective provided above and the theory of caricature developed earlier.



drawing, 1583 (from Hof man, 1957)

7.0 CONSTRAINTS / AUTOMATION ISSUES

With definition of caricature in hand, it is time to translate these ideas into heuristics that a computer can handle. In this section I will discuss the representations, assumptions, computation theory, and input and output requirements for the caricature generator.

7.1 PRIMITIVES

The implementation of machine-generated caricature is guided by the selection of graphical primitives out of which to construct the image. Computer graphic caricature can take any one of several forms. The image could be a 9-bit color photograph; it could be a grayscale image, the resolution of which can be reduced to fewer bits; it could be a line drawing stored as a list of points, or it can be merely a very few control points from which line segments are subsequently computed, using a cubic B-spline algorithm.

> "One of the most remarkable phenomena of vision is our ability to recognize an outline drawing. Clearly an outline drawing of, say, the face of a man, has very little resemblance to the face itself in color, or in the massing of light and shade; yet it may be a most recognizable portrait of its subject. The most plausible explanation of this is that, somewhere in the visual process, outlines are emphasized and some other aspects of an image are minimized in importance. The beginning of these processes is in the eye itself. Like all senses, the retina is subject to accomodation: that is, the constant maintenence of a stimulus reduces its ability to receive and to transmit

that stimulus....It is quite different on the boundary of two contrasting regions. Here these fluctuations produce an alternation between one stimulus and another, and this alternation, as we see in the phenomenon of after-images, not only does not tend to exhaust the visual mechanism by accomodation, but even tends to enhance its sensitivity. This is true whether the contrast between the two adjacent regions is one of light-intensity or of color....We thus find that the eye receives its most intense impression at boundaries, and that every visual image in fact has something of the nature of a line drawing."

(Norbert Wiener, 1948)

As an initial assumption I have chosen to represent the caricatures as line drawings because:

1) As a way to represent a visual image, a line drawing is consistent with theories about how the human visual system operates when it makes a primal sketch. (for more information, see Marr & Hildreth, 1980).
2) most traditional caricatures are executed using line drawings.
3) as a form of bandwidth compression, line-caricature provides a natural opportunity to explore getting across the most information with the greatest economy of line.
4) lines, especially when stored as control points, are easily distorted computationally. With the goal of keeping bandwidth requirements low, I am working primarily with 3-bit images containing enough gray values so that my black and white caricatures can be composed of de-jaggied (anti-aliased, non-staircased) lines.

In a drawing of a face, lines can be used to represent material changes, occluding edges, and other intensity gradients (such as wrinkles, shadows, shading, etc.) upon the surface.

Given the pitfalls of line-finding, it would be useful if the line drawings did not have to be smooth and regular enough to be themselves transformed into the caricature, but were used only to display the primitives from which the caricature is constructed. A way to avoid software convolutions described in the last section would be to implement a version of the Kanade program so that by knowing where to look (and by not caring about the noise and spurious lines), the input program for the caricature generator could find the locations of certain significant points which lie on intensity gradients, by using integral projections.

Once a small set of reference points is determined from the input image, the points are connected with lines using a cubic B-spline algorithm (common in computer-aided design) which employs a series of polynomial blending functions to compute a curved line from a series of control points. Note that the control points are NOT located directly on the lines they draw, but are calculated from the reference points, which are. The algorithm assigns only local effects to the control points, so that moving one reference point affects the lines spanning two points on either side of it. This decision to represent the parts of the caricature as line segments defined by the reference points generated by a cubic B-spline algorithm enables us to get line drawings from photos while side-stepping the need to make a coherent line drawing. Slight distortion due to the fact that relatively few reference points are used and the resulting splines do not follow the sketch of the face precisely, should be acceptable given our definition of caricature as impervious to slight metric distortion. The method described partially automates the input stage of the caricature generator, requiring user interaction to a much smaller degree than rotoscoping does.

7.2 INPUT CONSTRAINTS

Given that the caricature generator will use digitized input consisting of two-dimensional projections of faces, the following constraints can be made: input will consist of photographs taken under relatively uniform lighting conditions, and from fairly standardized viewpoints. By constraining the input images in this way we sacrifice the generality which may be desirable in another type of image-processing system (such as one which is supposed to recognize people in a security system) but which is not essential for our caricature generator. The caricature generator can compensate for small variations in scale and location of the head within the frame by scaling, translating, and rotating the reference points of the standard face to the input face. Convenient points of reference to use for these operations are the centers of the pupils. Our system registers the pupils of the norm (i.e., whatever face is being used for comparison) by scaling and placing them directly over those of the input image. The software within the caricature generator performs this transformation so that subsequent exaggerations will be made within the coordinate system that corresponds to that of the individual input image.

7.3 FEATURES

Once a line drawing is found, there needs to be a be a way for the caricature generator to represent and manipulate this information in some meaningful way. When members of our culture verbally characterize a face, "features" such as eyes, nose and mouth and head shape are commonly described. But these features are partly the conventions of language and partly due to the hierarchy of the senses; as far as the spatial arrangement of face is concerned, they are arbitrary. It is possible that a true perceptual feature for face recognition would be a relationship like the ratio of the upper lip to the height of the lower half of the face. In any case, if the implementation of automatic caricature is to be based on the key feature idea identified in section 4, several problems arise:

1) A set of possible key features should be rigorously determined through visual perception experiments where metric distances and ratios within line drawings of faces are systematically varied, and the most significant of these relationships determined.

2) There is no evidence that the set of key features, determined above, would be the same for each face.

Since the above research is not feasible in this investigation, and since the caricature generator uses lines, it makes sense for our purposes to abandon traditional notions of coherent facial features. Instead, I have defined a small structure of points sufficient to determine the most predominant lines on the face. These have been simplified in that the first pass of the caricature generator will entirely ignore those odd lines and wrinkles that are common to some faces and not to others. Obviously, these lines may turn out to be critical to individuating the face, and should be ultimately included.

An early program in the caricature generator grouped curves into traditional features, within which it connected certain endpoints. Another program simply compared the two face structures reference point by reference point, exaggerated one face, and drew it. While the latter program reached facelessness after it distorted the face by one to one-and-one-half times the amount that it differed from the norm, it seemed to be more successful as a caricature. The sketchy, unconnected lines are not particularly objectionable, and can be interpreted as an analogue of style. For extreme exaggeration a limiting/clipping scheme could be employed to prevent individual lines from crossing each other, and from totally decomposing the image. It is interesting to play with individual faces and norms to see when facelessness occurs.

7.4 STEREO INPUT

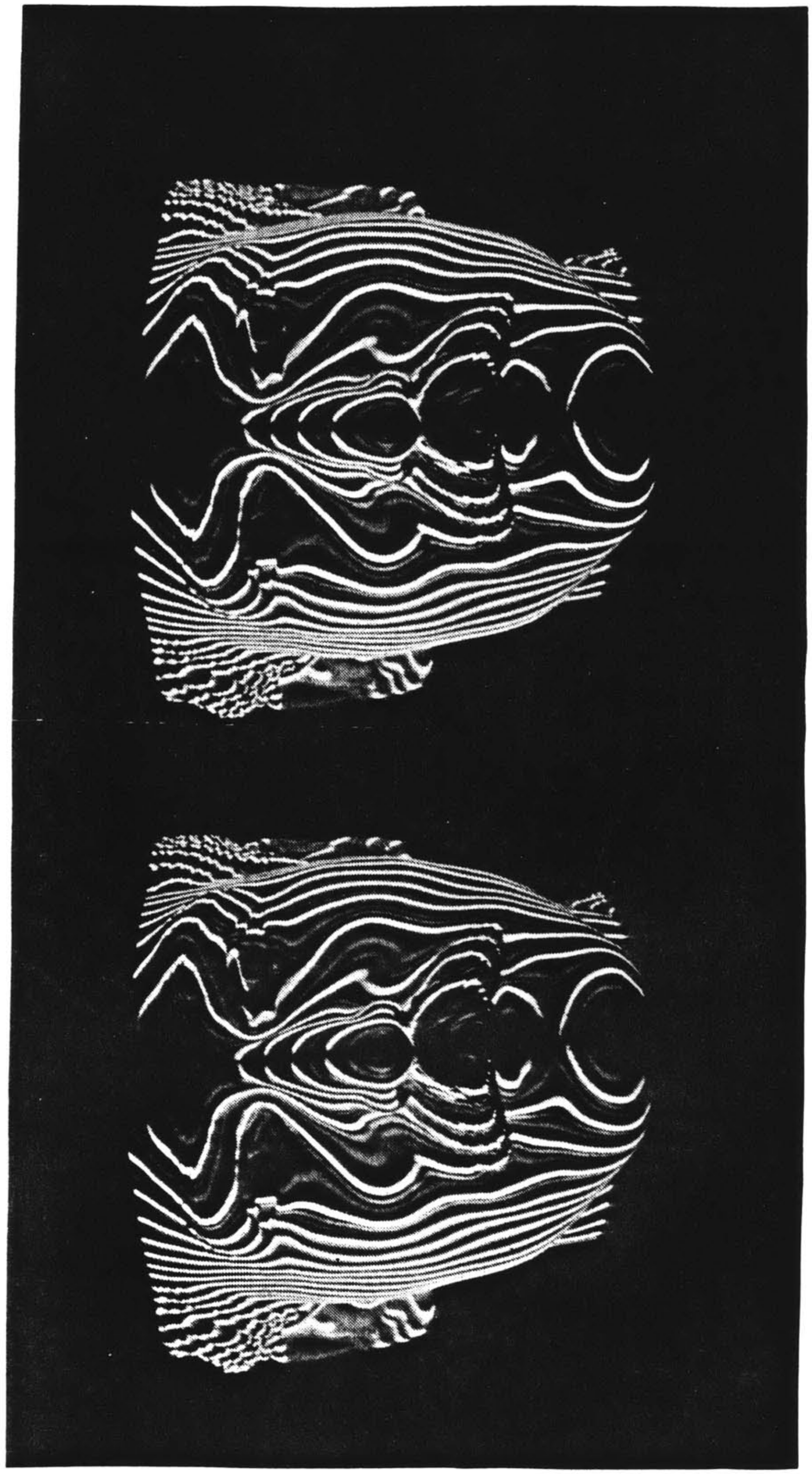
Our theory says that multiple inputs are preferable to one. If stereo input is available, it would be an advantage to increase the disparity between the two viewpoints, as some of the critical depth information provided by a stereo pair will be destroyed by metric distortion of the line drawing stage. After the practice of cartoonists who place 3/4 or even profile noses on frontal faces, and after the wanted posters in post offices, the caricature generator would ideally have access to a frontal and a profile photograph.

The illustration on the following page is a stereo pair of identical grids projected at right angles on a bust. It was made using the normal interocular distance.

7.5 THE POINT OF DEPARTURE

There are several approaches possible in creating norms with which to compare the caricatured face. One could choose some statistical average of faces taken from a population similar to the one from which the subject to be caricatured arises, or similarly, to the one to which the prospective viewers of the

This image by S. Brennan and S. Fisher uses projected lines to emphasize the contours of the face; this effect is heightened by the stereo pair.



caricature belong. In order to yield the best individuating caricature, the norm used should correspond to the sex, race, and perhaps the age of the subject, since these characteristics can be considered invariants, subordinate to the person's identity (unless, of course, a quality such as age is being used to distinguish the subject).

Any number of methods could be used to choose one of the norms stored in the caricature generator, based on the machine's model of its subject according to his age, etc. or his preferences regarding the relative attractiveness of the norms, or by determining which norm is the most like or unlike the subject's own face. Later we will look at some of the results of machine-generated caricatures done with respect to different models.

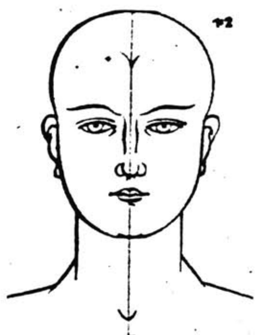
7.6 MAN-MACHINE INTERACTION

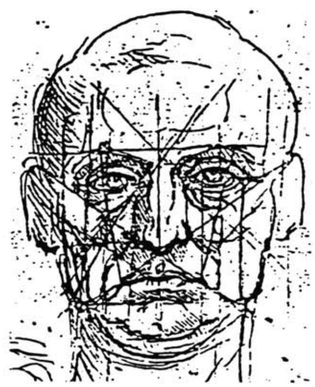
An important question in constructing a caricature generator is - How automatic should such a system be? To what extent is automation desirable - or possible? To some extent constraints will change according to the projected uses of the system. If the system is to be used by students of human cognitive and visual processes in face or caricature recognition, or by artists as a means of self-expression, or by users as an interactive game or graphics program, then parts of the system such as the identification/description of facial features and control points should be accomplished through man-machine interaction. If the goal is to construct a bandwidth-limited, transmission system for faces incorporating high level noise reduction and intelligent amplification of key features, then a more automatic system is called for.

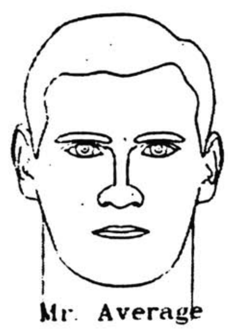
A natural way to configure the caricature generator is with an amplitude knob. The bystander is then able to adjust the degree of distortion in small steps during the transition from sketch to caricature. If a rough analogy is made to mixing music, one would be able to turn up the volume while leaving out the noisy tracks. The volume control could be made to span not only the range from line drawing to caricature, but also the range from line drawing to norm (for some cartoon plastic surgery).

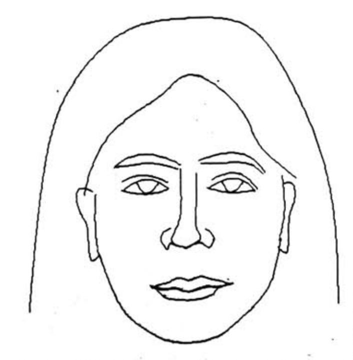
To actually construct a fully automated caricature generator is out of the range of this thesis. However, our investigation yields certain insights concerning how such a system could be developed. The automatic system would have a digitizing camera to grab frames, and the subject would be instructed to look directly at the camera, and then to turn 90 degrees and look at a target. Convolution hardware would approximate zero crossings, and the resulting sketches would be searched using a variation of Kanade's hierarchy of histograms. Reference points would be located automatically and stored in a standardized structure. As a check, the points could be compared to another file for excessive deviations, and the input program repeated if necessary. Then the points are connected by splines, and the image becomes a rubber band face.

Once a norm is chosen, either by the machine or by the subject, the difference between each point and its corresponding point on the norm is a vector. Distortion is performed in the opposite direction to this vector on the frontal drawing of the face, according to a scale determined by the initial distance apart. Then, using data from the profile as well on such protruberances as the nose and chin, the relevent parts of the face are redrawn. In particular, the two long vertical nose lines are redrawn as the average line between the frontal nose and the profile nose, which has been exaggerated with respect to a profile norm in the same manner.









Some norms used by the caricature generator (left to right): averaged faces by P. Weil, an ideal from Oskar Schlemmer, an ideal from da Vinci, the average face from Harmon's sample (1973), my face.

8.0 IMPLEMENTATION: SYSTEM DESIGN AND PROGRAMS

This section briefly describes and illustrates the package of programs which make up the caricature generator. Implementation was on a 32-bit minicomputer to which are interfaced a graphics tablet and a touch-sensitive screen. The display used is a frame buffer consisting of a raster of 640 X 480 picture elements, 9 bits deep. The color matrix was aligned in bit planes. Slots 0-7 contained a gray scale to handle the input picture and the final version of the caricature which consists of anti-aliased black lines. Slots 8-15 contained red, 16-32 contained blue, and 32-64 contained cyan; these colors were used to provide contrast and feedback to the user during the input and exaggeration programs.

8.1 INPUT program

A computer program called INPUT prompts the user with the description of a curve on the face and the number of points needed to define it. The user chooses points on the face by touching the touch sensitive screen or by using a tablet. These reference points lie directly on significant boundaries, intensity gradients and occluding edges. INPUT provides an intuitive way to enter information about curved lines. From these virtual control points the actual control points("out in space") are computed, and from these the curved line segments are computed according the the blending functions used in the cubic B-spline algorithm.

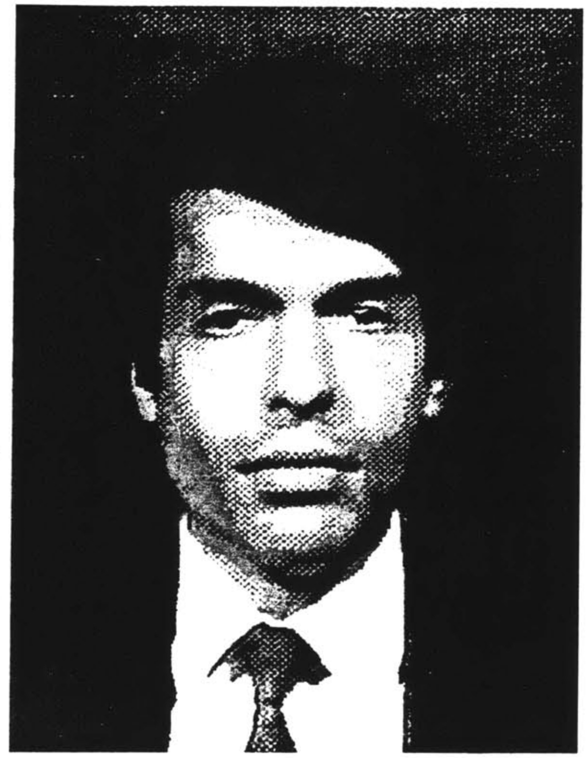
This program is interactive in that one can grab any point in the curve being entered at the time and edit it, like stretching a rubber band. Color is used as feedback; when a point is grabbed it becomes activated - i.e. movable - and changes color, through the manipulation of bit planes.

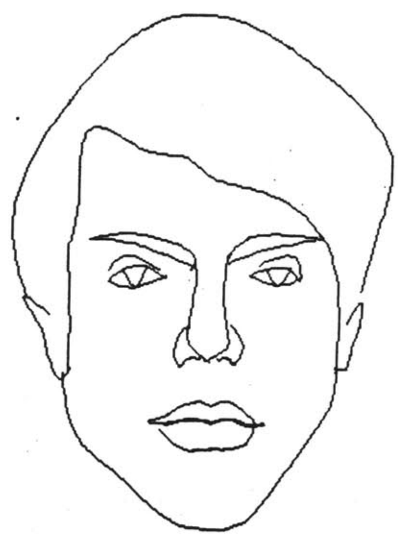
A structure was chosen to contain control points in a prescribed way so that the points located on a variety of faces could be easily compared to one another. A constraint in defining this structure was to use the fewest number of control points that could reasonably define the curvature of the line segments. This makes the input stage easier on the user. The more points used, the less metric distortion in the line drawing.

This method yields a reasonably accurate line drawing, accomplishing through man-machine interaction that which would be computationally slow and unreliable in a fully automated input stage such as the ones described in section 6.

INPUT PROGRAM

The program prompts the user for points which may be entered using a tablet or a touch sensitive screen. The digitized face of the person being caricatured is displayed during this process, and as each set of points is entered, a spline curve is drawn connecting them. The user may edit any line segment by "grabbing" and repositioning one of its points. The points are stored as a description of the face. Ideally, this description would include structures representing other viewpoints of the face, such as the profile.

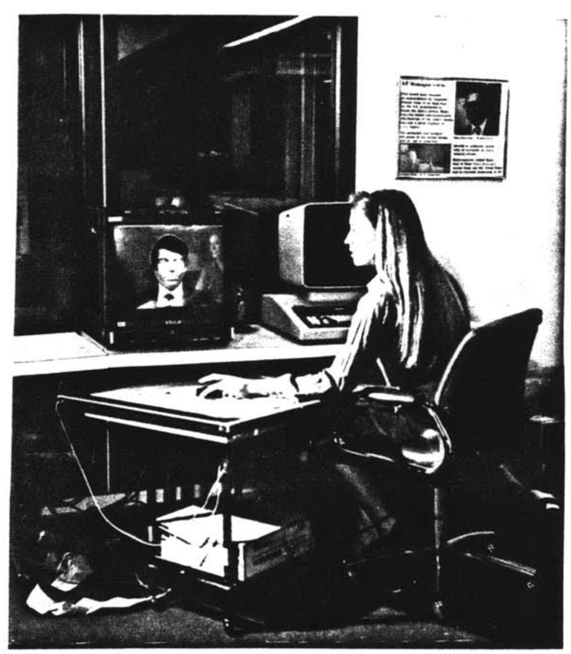




The following structure is used to store a description of the face. This particular structure was obtained through trial and error, and represents a minimal line drawing. The next step is to expand the caricature generator by using a larger structure to represent other facial lines, contours, etc.

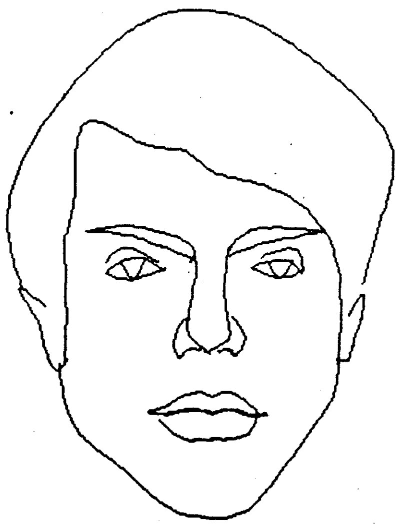
TOTAL: 29 line segments, each consisting of one to nine points. In all, the description consists of 135 points.

SEGMENT:	# OF POINTS:
right pupil	1
left pupil	1
right iris	3
right lid	3
right lower lid	3
right upper lid	3
left iris	3
left lid	3
left lower lid	3
left upper lid	3
right side of nose	5
left side of nose	5
right nostril	5
left nostril	5
bottom of upper lip	7
top of lower lip	7
top of upper lip	7
bottom of lower lip	7
top of right eyebrow	4
bottom of right eyebrow	4
top of left eyebrow	4
bottom of left eyebrow	4
top of head	9
jawline	9
hairline	9
right side of face	3
left side of face	3
right ear	6
left ear	6



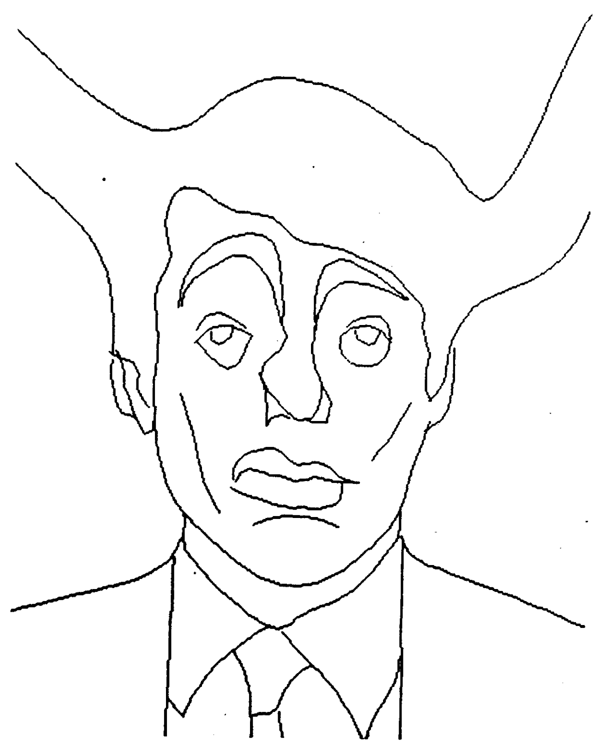
8.2 THE RUBBER BAND FACE

A program called UGH draws a selected face on the screen that has been INPUT using relatively few reference points. The user then selects any point on the face using the tablet. The point changes color and is activated. The position of the puck is followed by a cursor, and the user can put the point back down anywhere on the screen. As soon as the point is relocated, its curve is erased and the new curve is drawn. This happens quickly, and the drawing has an amusing fluidity. The effect is that of a face made out of rubber bands. This program was designed to be used by people with varying artistic skill to create caricatures and fantastic drawings.



RUBBER BAND FACE

The two cartoons below were made by grabbing points on the face at left and stretching the lines. This drawing program, called UGH, is intuitive and amusing to use; because the lines are only one pixel wide, they are redrawn very quickly.



8.3 GUMBY

This is the set of routines that forms the core of the caricature generator. It first calls for the name of the face to be caricatured, and the name of the norm with which it is to be compared. Several norms currently exist on the system. For the following illustrations these average faces were used: a Leonardo da Vinci drawing of facial proportions, an ideal face from the work of Oskar Schlemmer, a composite picture of ten white males from Aspen, Colorado (made from multiple exposures), a line drawing derived from Harmon's study and formerly used as the point of departure in the "Whatsisface" system described in section 7, and a line drawing input from a digitized version of my own face.

This last norm is not entirely feasible for purposes of comparison since it was input from a line drawing that was derived using the automatic method of finding lines by filtering, while the other norms were input over digitized photographs. In addition, most of the caricatures were of males, and our theory assumes that subject and norm should share certain general characteristics. The drawing of my face was included because it represents a step taken in the direction of eventually automating the whole caricature generator, by similarly filtering a digitized projection of the face in hardware or software (see section 6), and then applying the Kanade program which automatically locates the reference points.

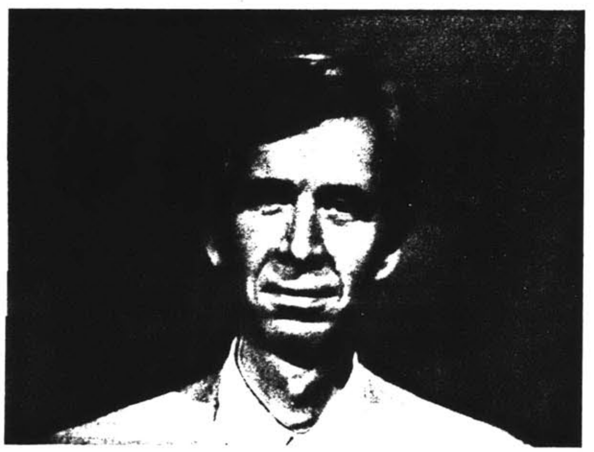
Once the subject face and basis for comparison are chosen by the user, a routine called INTEROCULAR determines an absolute scaling value by normalizing the distance between the pupils of the average face to that distance on the face to be caricatured.

The program COMPARE scales, translates, and if necessary, rotates the norm so that it is spatially aligned with the face. it provides the option of drawing the norm on top of the subject so that the user can visualize what will happen. COMPARE then determines the difference between each point on the face and its corresponding point on the norm (a measurement for each point hereafter referred to as the exaggeration value). Each curve is ranked according to greatest average differences between its points and those of the norm. This ranking could be used to warp the curves in a particular order, to warp only certain curves, or to warp curves according to different scales. GUMBY warps curves based on the exaggeration value of each point and input by the user from the amplitude control.

The amplitude control appears as a rectangular slider to the side of the screen and is dynamic; it takes touch input and redraws a horizontal bar as the user changes his mind. It is set up on an exponential scale. The most extreme fourth of this range of distortion is actually visually off the upper end of the scale, because the selection of amplitudes in this range usually causes the image to become totally unrecognizable as a face.

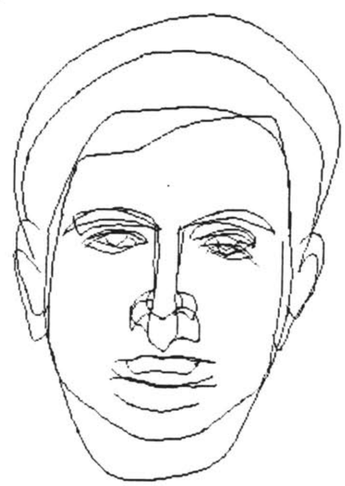
DISTORT is the routine which calculates the new position of each point and then defines the new curve with cubic B-splines. Subsequently, the old curve is erased by manipulating the color values in bit planes, and the new curve is drawn. There is a great deal more variation in some curves than in others. Because the transformation depends on the exaggeraton value for each point, each curve changes at a unique rate and in a different direction In the case of the hairline, the curve may be extremely variable; it varies in location from near the eyebrows to near the top of the head. Also, more than any other curve the hairline may be arbitrarily determined by the way the hair falls across the forehead at a given moment. Therefore, the caricature generator is programmed to minimize the exaggeration performed on the hairline, since there is little point in comparing to curves that may have nothing in common. For all other curves except the hairline, the caricature generator exaggerates in exactly the same way, as outlined above.

Following are some images made with the caricature generator.



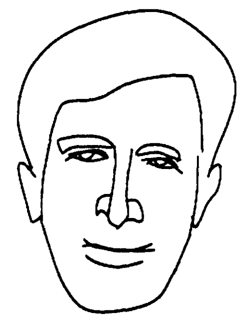
Above left: digitized photograph of the subject.

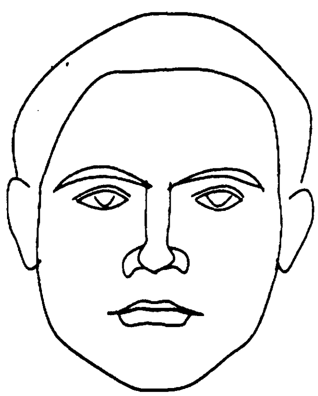
Above right: the line drawing description of the face.

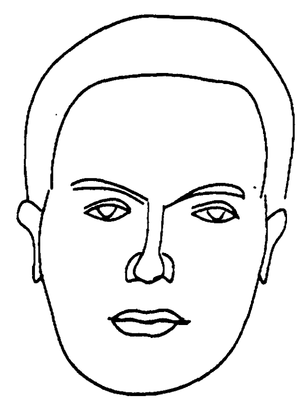


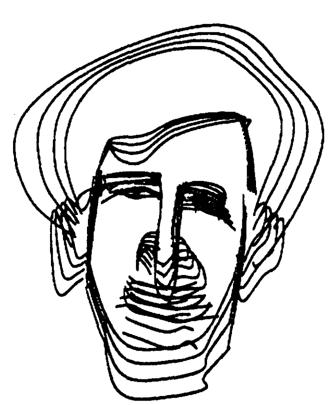
Right: INTEROCULAR scales and aligns the face with the norm. COMPARE finds an exaggeration vector (the distance between the norm and the subject) for each of 135 points.

DISTORT exaggerates the face of the subject (right) away from the norm chosen. Below left: the Leonardo norm, and bottom left: the results, given several amplitudes of distortion. Below right: the averaged Aspen norm, and bottom right: the result. Note that while the two results differ substantially in diplacement of the lines, some of the same characteristics have been exaggerated by both norms - ie, the long face, wide mouth, etc. These caricatures are displayed on the following page.

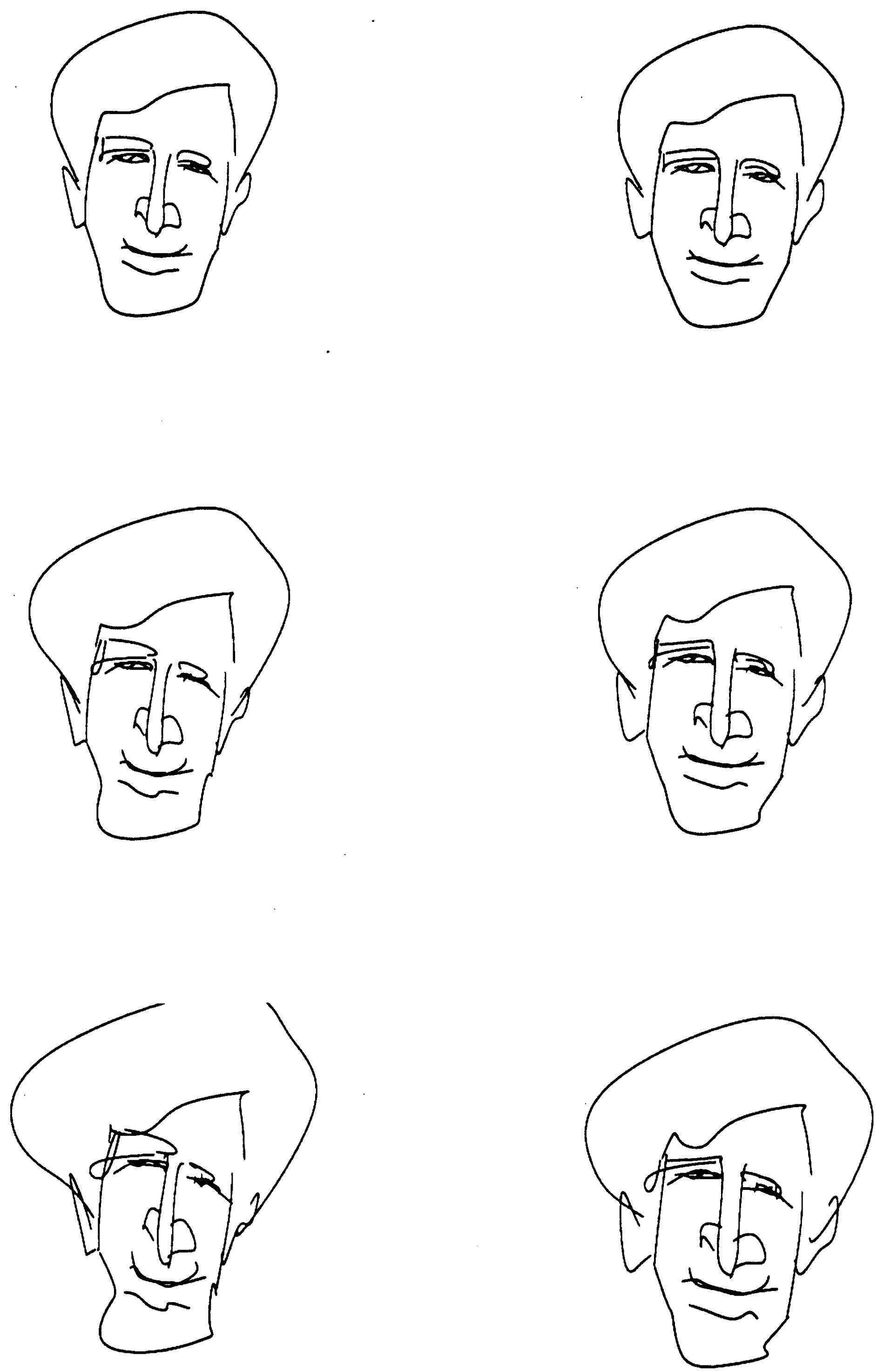


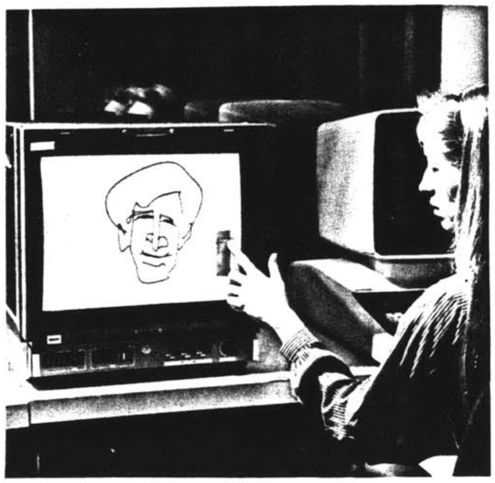






Comparison of the results of two different norms, same subject. Left: three successive distortions using the Leonardo norm. Right: three successive distortions using the Aspen norm.

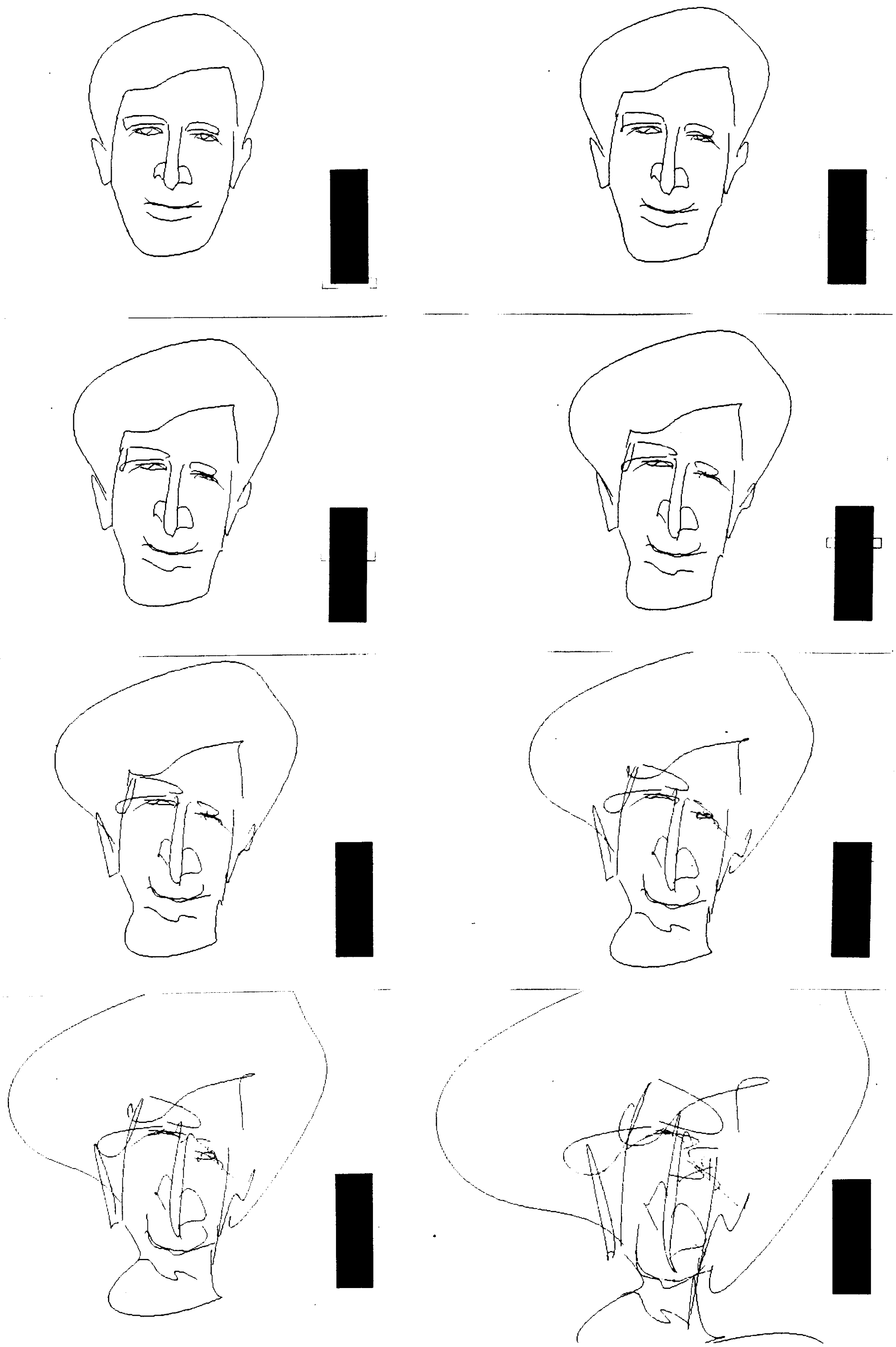




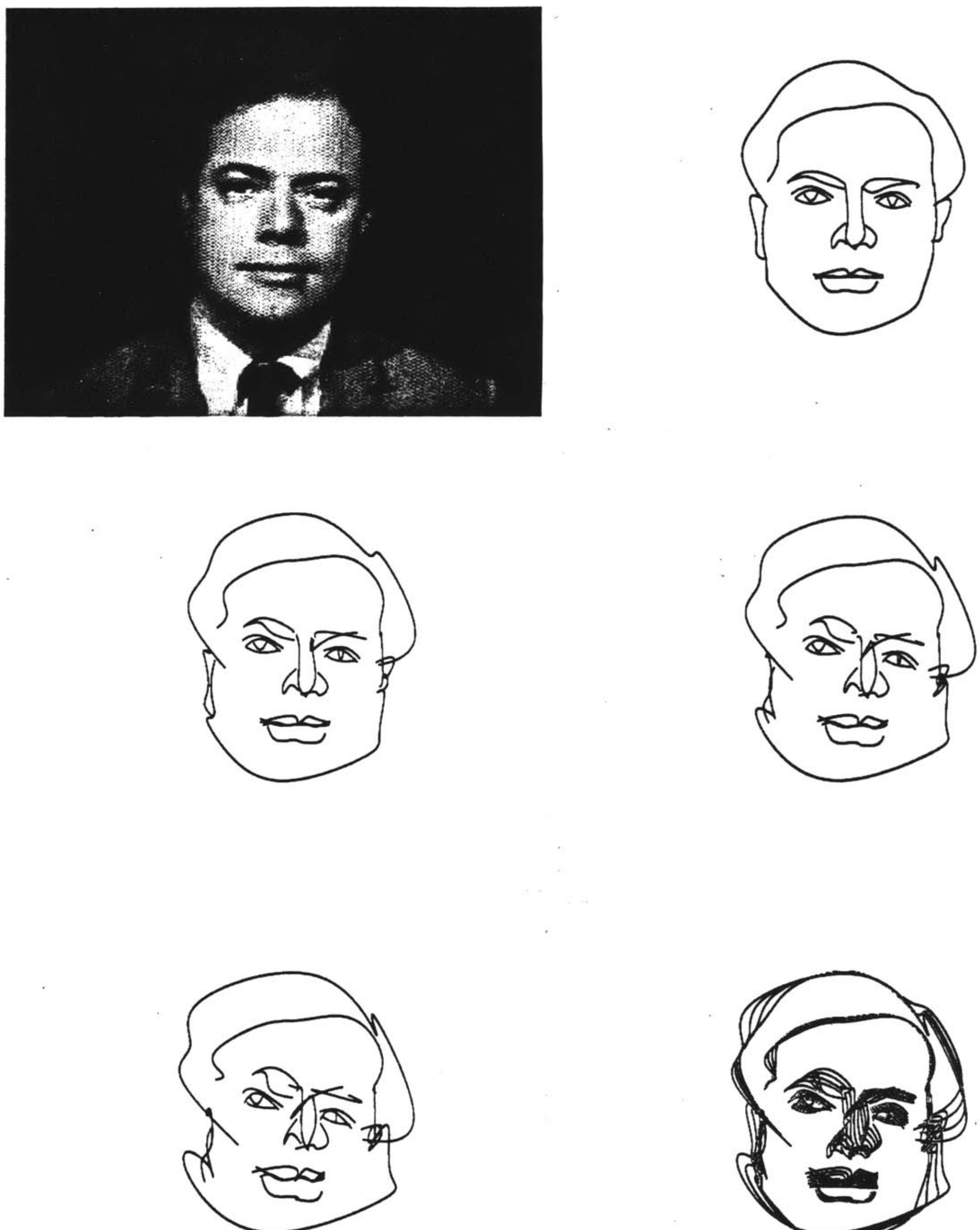
INTERACTIVITY

The rectangle at the right of the caricature represents the volume control. The horizontal bar bisecting it is analogous to a slider. By touching the screen in this region the user can move the slider bar up and down, thereby selecting the amplitude of distortion. The lower end of the control represents no exaggeration at all as in the original line drawing, and the upper end represents extreme exaggeration.

Sequence: exaggeration with respect to the Aspen norm



This subject was exaggerated with respect to the same Aspen norm used earlier. Note that the caricature generator makes his face even shorter, his lips thicker, etc.



9.0 THE FUTURE OF THE CARICATURE GENERATOR

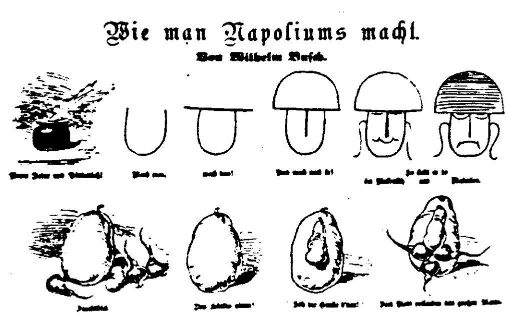
In conclusion: a note on the potential of this interactive program.

9.1 APPLICATIONS: CARICATURE FOR EVERYONE

If there is any application of caricature to teleconferencing, it lies (a) in the amplification of the identity of the face, and (b) in the congruence between the type of face representation and the information it bears. In terms of recognition, the advantage of caricature over line drawing was discussed earlier. During the teleconferencing experiments described in section 4, a system was built to distill vowel phonemes (these being the steady state parts of the speech signal) from the first and second energy formants in speech. This information was used to select the likely lip position from which a particular sound issued. The result: reasonably accurate, totally automatic lip sync. Subsequently, comparisons were made between two representations, one animated from a series of photographs and another animated from a series of caricatures. When displayed side by side, the faces driven by speech supported (at least to some) the fact that lip synchrony need not be as exact in a talking caricature as in a talking head. Surprisingly, users in this teleconferencing project ("Transmission of Presence") preferred the caricature to the animated photograph, since it was more obviously an abstract representation, while managing to convey the visual essence of an individual's face. So in an environment where bandwidth is at a premium, caricature is the ultimate form of semantic bandwidth reduction and is a consistent representation.

The other application, which seems more obvious, is to use the caricature generator as a playful introduction to interactive graphics. Showing the drawing process, or having a picture draw itself, has traditionally been used by animators as a paradoxical and whimsical type of animation, because of its self-consciousness about its own frame. Almost everyone has some recollection of some sort of drawing or molding toy which taught him as much about spatial relationships as it provided pleasure or creative satisfaction. In some sense, Etch-a-Sketch, Silly Putty, Gumby, and Mr. Potatohead are all predecessors of the caricature generator.

The fact that this toy is made out of a computer makes it potentially a powerful tool to learn about the mathematical and heuristic aspects of distortions that the imagination, and subsequently the caricature generator, can apply to an image.



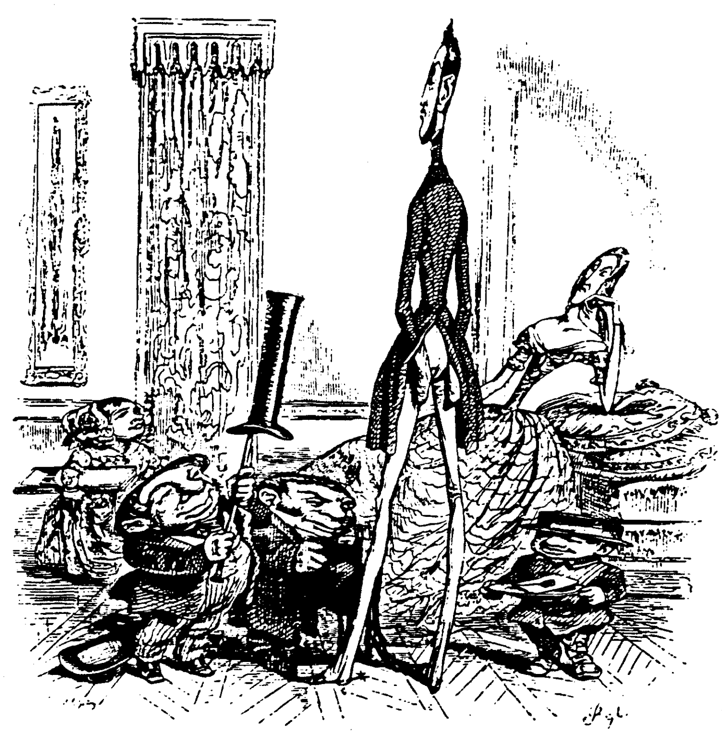
Wilhelm Busch's steps for drawing Napoleon

9.2 SIMPLE ANIMATION

The reference points used by the caricature generator can be used to animate the mouth or form certain elementary expressions. By moving some of the points of the mouth up and down in pre-programmed ways, the images needed for lip sync can be generated. Similarly, a few points can be moved to make the face assume an expression. This same simple transformation could be applied to any face, since the structures for points are all identical.

9.3 OTHER TRANSFORMATIONS

Grandville's caricature as social satire

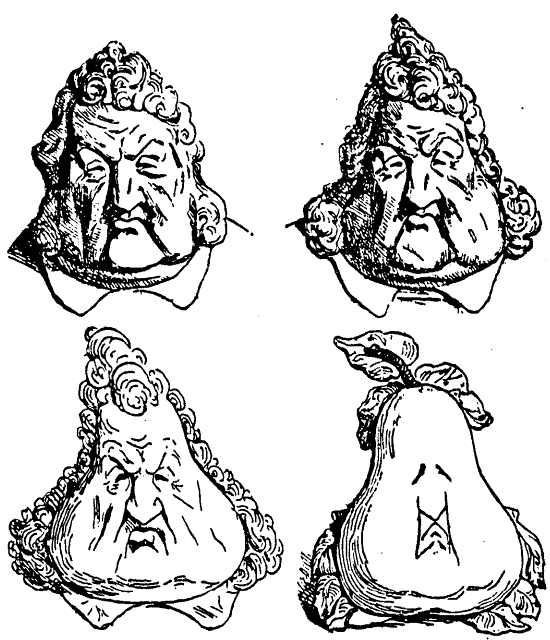


Some of the transformations traditionally used by caricaturists and cartoonists can be incorporated into the caricature generator. Grandville, in the 1840's, used simple horizontal and vertical distortions of the human physique to amplify the aspect of social standing in his characters. A degree of distortion is applied in two directions simultaneously but in reciprocal amounts; inferior social status is invested in closeness to the ground. If one wishes to establish a language of caricature, one need only agree on the conventions to use; this particular convention could be incorporated into the caricature generator programs.

9.4 CONTEXT

It is a straightforward process to digitize a bank of images for the system to use as forgiving templates to be matched automatically to the most distorted facial features. When an association is made the computer can then incorporate the object represented by the template as a replacement for the feature in the caricature. In this way the caricature generator can aspire to supplying context. This playful associative process can of course be made either entirely a random response to the perceived similarity of form, or it could become a process based on an intelligent model of the subject and the context of his/her personality, voice, etc. One can forsee the ultimate electronic Mr.Potato-head.

Another way to use machine imagery is to transform the whole face (pre- or post-distortion) into an appropriate object. Techniques of automated "in-betweening" have been routinely used in computerized animation for several years now. The animation system is supplied with several key frames from which it synthesizes intermediate drawings. This can be done by linearly interpolating positions for objects which move from key frame to key frame, by transforming one coordinate system to another, by optically dissolving between two images, or by following any other likely set of rules. Applied to our automated system, in-betweening techniques could be used to make a caricature resemble any man- or machine-made association. This is what Charles Philipon did by hand when prosecuted for depicting his sovereign Louis-Philippe as a poivre (fathead or pear). His defense was to publish the following set of drawings, and to argue that none of these transformations alone was enough to incriminate him.



Charles Philipon's poivre

Daumier came up with another way to implement the same transformation (1834):

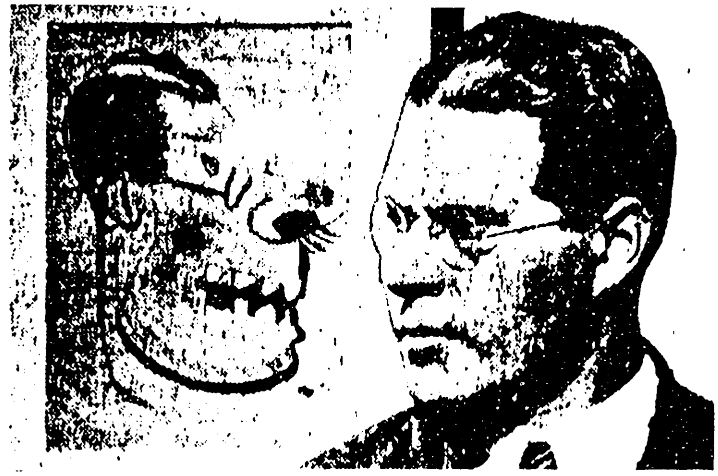
Borrowing Daumier's technique, the computer could treat the face as a surface texture and simply map it onto a projection of a three-dimensional object.

One could automatically generate a recognizable caricature of, say, one's employer. Then one could see how he or she looked as an animal, then as a vegetable, leaving the computer the responsibility of selecting the most likely species. More experienced users could insert the double-entendre earlier in the distortion process.

9.5 STYLE

Coded in every caricature are not only the identity of the subject and of some (often vague) notion of an ideal, but also the information about WHO made the caricature. When one sees a popular caricature one recognizes immediately both the subject and the artist. This fact brings up the existence of yet another invariant superimposed on every caricature - that of style. The caricature generator will evolve to the point where it will be entitled to sign its own work; that is, it could conceivably learn to ape the line quality or degree of angularity or other graphic mannerisms of well-known caricaturists. If there is a user, he could impose his choices in such a way that his experiences with the caricature generator bear a personalized stamp.

Style is perhaps the most elusive quality to pin down for a computer's purposes, but one can begin to speculate on a few of its analogues. I noticed suddenly after trying a variety of structures of reference points that characteristics of the B-spline algorithm had a slightly different overall effect depending upon how many points I used to create the initial line drawing. For example, when a very few number of points were used, the inpression was a bit like a Hirschfeld cartoon; using more control points was a little more like a George Price cartoon. Users should develop a preference for a characteristic number of points with which to work.



Moholy-Nagy and caricature by Gyorgy Kepes

Style may also be related to the choice of the norm. As hinted earlier, the point of departure for the caricature generator may be different for each user (or for the computer's model of each user). Since it is not at all clear except by trial and error which norm is optimal, perhaps that is because the typical norm is mediated by the caricaturist's self image.

Certainly such stylistic things as line thickness, degree of cross-hatching, and whether the computer connects the endpoints of the lines or not, can potentially all be added to this language of caricature. These elements can be consciously chosen by the user, or they can be chosen by the computer, based on some information ABOUT the user himself, implied by his style of interaction with the caricature generator. For example, if the user aggressively punches the amplitude control, the computer can respond by making the image out of big, black lines. If the user responds frequently and quickly (or has a short attention span ?) the lines can be sketchy, unconnected and more abstract. Perhaps the personal information from technologies such as body tracking or touch can be translated into style. The freeness of style of many caricaturists attests to the fact that caricature, more than some other types of imaging, contains the gesture of the artist within the line.

9.6 PARTICIPATORY CARICATURE

> "Caricature indeed also tries to produce an effect, not, however, on' the person caricatured, but on the spectator, who is influenced to accomplish a particular effort of imagination." (Kris, 1952)

It is particularly appropriate that the form and implementation of a style of imaging as whimsical as caricature should finally be free and interactive. The development of friendlier interfaces and graphics software, along with the definition of an intelligent and tangible model that can be manipulated by a computer, makes this possible. As a mode of representation that semantically compresses facial bandwidth, caricaturing is an appropriate algorithm for displaying a face. As a display, a face is an intuitive and versatile choice. As an imaginative form of self-expression, the caricature generator could be used to provide immediate visual gratification to the novice user, encouraging an early commitment to working with the computer. The animated, more intelligent caricature generator could act as as a surrogate personality, as an agent, or as the ultimate friendly interface to a world of information. Caricature, traditionally a spectator sport, could be participatory as well.

ACKNOWLEDGEMENTS

I wish to acknowledge the unique environment at the Architecture Machine Group, where people of eclectic professional, technical and creative backgrounds have the opportunity to work together under the most hectic but supportive conditions imaginable. Thanks also to:

Scott Fisher, for patience, imaginative discussions and aesthetic insight,

Dr. Alan Kay, for inspiration and for providing me with a context in which to implement these ideas,

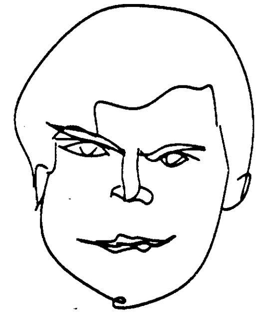
Gyorgy Kepes, my friend and mentor during the past year,

Dr. Jerome Wiesner and Dr. Seymour Papert, for their interest and advice,

Andrew Lippman, director of the lab, for a lively environment,

the hackers in the terminal garden, for the answers to all my programming questions,

And finally, thanks to Nicholas Negroponte, whose vision started the whole thing and whose imaginative guidance fuels it still.



REFERENCES:

Applebaum, Stanley, ed., BIZARRIES AND FANTASIES OF GRANDVILLE, New York, Dover Publications, Inc., 1974.

Arnheim, Rudolph, TOWARD A PSYCHOLOGY OF ART, Berkeley and Los Angeles, University of California Press,1966.

Arnheim, Rudolph, "Some comments on J. J. Gibson's Approach to Picture Perception", LEONARDO, vol. 12, pp. 212-2, 1979.

Bergson, Henri, "Laughter", from COMEDY, ed. by Wylie Sypher, Baltimore and London, John Hopkins University Press, 1956. (first published in 1900)

Bledsoe, W. W., "The Model Method of Facial Recognition", Panoramic Research Inc., Publication PRI-15, August, 1964.

Bledsoe, W. W., "Man-Machine Facial Recognition", Panoramic Research Inc., Publication PRI-22, August, 1966.

Bledsoe, W. W., and Hart, P. E., "Semiautomatic Facial Recognition", Stanford Research Institute Technical Report 6693, Sept. 1968.

Bradshaw, J. L., and Wallace, G., "Models for the Processing and Identification of Faces", PERCEPTION AND PSYCHOPHYSICS, 9(5):443-448, 1971.

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

CARICATURE AND ITS ROLE IN GRAPHIC SATIRE, exhibition, Brown University Department of Art, Providence, Rhode Island, April 7 - May 9, 1971.

Chernoff, H., "The Use of Faces to Represent Points in K-Dimensional Space Graphically", JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, vol. 361, p. 361., 1973.

Durer, Albrecht, Schriftlicher Nachlass - Eine Auswahl, Union Verlag, Berlin, 1963.

Durer, Albrecht, PROPORTIONSLEHRE, Verlag Walter Uhl, Unterschneidheim, 1969. (reprinted from 1528)

Feaver, William, MASTERS OF CARICATURE, New York, Alfred A. Knopf, 1981.

Freud, Sigmund, JOKES AND THEIR RELATION TO THE UNCONSCIOUS, translated and edited by James Strachey, New York, W. W. Norton & Company, 1960. (first published in 1905)

Galper, R. E., "Recogition of Faces in Photographic Negative", PSYCHONOMIC SCIENCE, 19:207-208, 1970.

Galper, R. E., " Functional Race Membership' and Recognition of Faces", Percept. Mot. Skills, 37:455-462, 1973.

Gibson, Eleanor J., PRINCIPLES OF PERCEPTUAL LEARNING AND DEVELOPMENT, New York, Appleton-Century-Crofts, 1967.

Gibson, J. J., THE PERCEPTION OF THE VISUAL WORLD, Boston, Houghton Mifflin Company, 1950.

Gibson, J. J., "The Information Available in Pictures", LEONARDO, vol.4(27), 1971.

Gibson, J. J., "On The Concept of Formless Invariants' in Visual Perception", LEONARDO, vol 6(43), 1973.

Gillenson, Mark L., Chandrasekaran, B., "Whatsisface: Human Facial Composition by Computer Graphics", Proceedings of the ACM Annual Conference, vol. 9, no. 1, Spring, 1975.

Gillenson, Mark L., Chandrasekaran, B., Csuri, Charles, Schwartz, Robert, "Computer-Assisted Facial Sketching", Great Britain, LEONARDO, vol. 9, Pergamon Press, pp. 126-129, 1976.

Gombrich, E. H., ART AND ILLUSION, Princeton, New Jersey, Princeton University Press, 1960 (1972).

Gombrich, E. H., MEDITATIONS ON A HOBBY HORSE AND OTHER ESSAYS ON THE THEORY OF ART, London, Phaidon Press, 1963.

Gombrich, E. H., "The Mask and the Face", from ART, PERCEPTION AND REALITY, Baltimore and London, The John Hopkins University Press, 1970.

Gordon, Louise, HOW TO DRAW THE HUMAN HEAD, New York, The Viking Press, 1977.

Grose, Francis, "Rules for Drawing Caricaturas", bound with William Hogarth's THE ANALYSIS OF BEAUTY, printed in the Strand, for Samuel Bagster, 1791.

Hamm, Jack, CARTOONING THE HEAD & FIGURE, New York, Grosset & Dunlap, 1967.

Harmon, Leon D., "The Recognition of Faces", SCIENTIFIC AMERICAN, pp. 70-84, November, 1973,.

Hillier, Bevis, CARTOONS AND CARICATURES, Studio Vista/Dutton Paperback, 1970.

Hirschfeld, Al, HIRSCHFELD'S WORLD, New York, Harry N. Abrams, Inc., Publishers, 1981.

Hochberg, Julian, "The Representation of Things and People", from ART, PERCEPTION AND REALITY, Baltimore, The John Hopkins University Press, 1972.

Hockberg, J., and Galper, R. E., "Recognition of Faces: I. An Exploratory Study", PSYCHONOMIC SCIENCE, 9:619-620, 1967.

Hoffman, D. D., "Representing Three-Dimensional Shapes for Recognition: Computational Problems Raised by Faces", Massachusetts Institute of Technology, A. I. Memo # 630, 1981.

Hofmann, Werner, CARICATURE FROM LEONARDO TO PICASSO, New York, Crown Publishers, Inc., 1957.

Huard, Pierre, ed., LEONARDO DA VINCI: DESSINS ANATOMIQUES, Paris, Les Editions Roger Dacosta, 1961.

Hulteen, Eric, "The Human Interface to Personal Computers", unpublished master's thesis, Architecture Machine Group, Massachusetts Institute of Technology, January, 1982.

Kanade, Takeo, "Picture Processing System by Computer Complex and Recognition of Human Faces", Department of Information Science, Kyoto University, November, 1973.

Kay, Alan C., "Microelectronics and the Personal Computer", SCIENTIFIC AMERICAN, vol. 237, no. 3, pp 230-244, September, 1977.

Kelen, Emery, ed., LEONARDO DA VINCI'S ADVICE TO ARTISTS, Nashville, Tennessee, Thomas Nelson Inc. Publishers, 1974.

Kelly, Michael D., "Visual Identification of People by Computer", Stanford Artificial Intelligence Project Memo AI-130, Computer Science Department, Stanford University 1970.

Kenner, Hugh, THE COUNTERFEITERS, New York, Anchor Book, Garden City, 1973.

Kris, Ernst, PSYCHOANALYTIC EXPLORATIONS IN ART, New York, Schocken Books, 1952.

Larkin, David, ed., FANTASTIC ART, New York, Ballantine Books, 1973.

Lucie-Smith, Edward, THE ART OF CARICATURE, Ithaca, New York, Cornell University Press, 1981.

Lynch, Bolun, A HISTORY OF CARICATURE, Little, Brown and Company, Boston, 1927.

Marr, D., and Nishihara, H. K., "Representation and recognition of the spatial Organization of Three-dimensional Shapes", PROC. R. SOC. LOND., B 200, pp. 269-294, 1978.

Marr, D., and Hildreth, E., "Theory of Edge Detection", Proc. R. Soc. Lond., B 207, pp. 187-217, 1980.

Minsky, Marvin, "Jokes and the Logic of the Cognitive Unconscious", Massachusetts Institute of Technology A.I. Memo # 603, November, 1980.

Michell, John, NATURAL LIKENESS: FACES AND FIGURES IN NATURE, E. P. Dutton, New York, 1979.

Munari, Bruno, DESIGN AS ART, England, Penguin Books, 1971.

Negroponte, N., THE ARCHITECTURE MACHINE, Cambridge, MIT Press, 1970.

Negroponte, N., Lippman, A., and Bolt, R., "Transmission of Presence", MIT Architecture Machine Group proposal to the Cybernetics Technology Division of the Defense Advanced Research Projects Agency, 1980.

Nelson, Roy Paul, CARTOONING, Chicago, Contemporary Books, Inc., 1975.

Newman, William M., and Sproull, Robert F., PRINCIPLES OF INTERACTIVE COMPUTER GRAPHICS, New York, McGraw-Hill Book Company, 1973.

Nishihara, H. K., "Intensity, Visible-Surface, and Volumetric Representations", ARTIFICIAL INTELLIGENCE, 17, pp. 265-284, 1981.

Nishihara, H. K., and Larson, N. G., "Towards a Real-Time Implementation of the Marr and Poggio Stereo Matcher", IMAGE UNDERSTANDING, Washington, D. C., Defense Advanced Research Projects Agency, 1981.

Panofsky, Erwin, ALBRECHT DURER, vol. 2, Princeton, N. J., Princeton University Press, 1945.

Papert, Seymour, MINDSTORMS, New York, Basic Books, Inc., 1980.

Parke, Frederick I., "Computer Generated Animation of Faces", PROCEEDINGS OF THE ACM ANNUAL CONFERENCE, vol. 1, 1972.

Perkins, David N., and Hagen, Margaret A., "Convention, Context and Caricature", THE PERCEPTION OF PICTURES, vol. 1, Academic Press, Inc., 1980.

Perkins, D., "Caricature and Recognition", STUDIES IN THE ANTHPOLOGY OF VISUAL COMMUNICATION, 2(1), spring, 1975.

Perkins, D. N., "Pictures and the Real Thing", Cambridge, Ma., Project Zero, Harvard University, 1979.

Pittenger, John B., Shaw, Robert E., and Mark, Leonard S., "Perceptual Information for the Age Level of Faces as a Higher Order Invariant of Growth", JOURNAL OF EXPERIMENTAL PSYCHOLOGY: HUMAN PERCEPTION AND PERFORMANCE, vol. 5, no. 3, pp. 478-493, 1979.

Popham, A. E., THE DRAWINGS OF LEONARDO DA VINCI, New York, Harcourt, Brace & World, Inc., 1945.

Ramus, Charles F., ed., DAUMIER: 120 GREAT LITHOGRAPHS, New York, Dover Publications, 1978.

Richardson, John Adkins, THE COMPLETE BOOK OF CARTOONING, Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1977.

Rimmer, William, ART ANATOMY, New York, Dover Publications, Inc., 1962 (reprinted from Little, Brown, and Company, 1877).

Rongo, Robert, ROBOTIC VISION SYSTEMS, INC. DATA-BASE USERS MANUAL, Melville, New York, Robotic Vision Systems (formerly Solid Photography), January, 1982.

Schlemmer, Oskar, MAN, Cambridge, Mass., The M.I.T. Press, 1971.

Sorel, Edward, ed., RESEMBLANCES: AMAZING FACES BY CHARLES LE BRUN, New York, Harlan Quist, 1980.

Spencer, Dick, EDITORIAL CARTOONING, Ames, Iowa, The Iowa State College Press, 1949.

St. Hill, Thomas Nast, THOMAS NAST, New York, Dover Publications, Inc., 1974.

Szep, Paul, THE HARDER THEY FALL: BOSTON GLOBE CARTOONS BY PAUL SZEP, Boston, Globe Newspaper Company, 1975.

Thompson, D'Arcy, ON GROWTH AND FORM, ed. by John Tyler Bonner, Cambridge, Cambridge University Press, 1961. (first published in 1917)

Todd, James T., and Mark, Leonard S., "Issues Related To The Prediction of Craniofacial Growth", AMERICAN JOURNAL OF ORTHODONTICS, Vol. 79, no. 1, pp. 63-80, 1981.

Weil, P., "About Face", unpublished master's thesis, Architecture Machine Group, Massachusetts Institute of Technology, August, 1982.

Wickler, Wolfgang, MIMICRY IN PLANTS AND ANIMALS, New York, McGraw-Hill Book Company, 1968.

Wiener, Norbert, CYBERNETICS, OR CONTROL AND COMMUNICATION IN THE ANIMAL AND THE MACHINE, New York, The Technology Press, John Wiley & Sons, Inc., 1948.

Yin, R., "Looking at Upside-Down Faces", JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 81:141-145, 1969.