Office of Naval Research Technical Progress Report

Contract Number: N00014-93-1-0305
Contract Period: January 1, 1993 - December 31, 1994
Title: Computational Approaches to Human Shape Representation

PI's: Michael J. Tarr (Psychology)
David J. Kriegman (Electrical Engineering)
Institution: Yale University
Address: PO Box 208205
New Haven, CT 06520-8205

tel: (203) 432-4637
fax: (203) 432-7172
email: tarr@cs.yale.edu

Table of Contents

Introduction

The primary goal of this project has been a computational investigation into the underlying representations used in human object recognition. To this end, we began with a relatively new approach to object representation in computer vision, that of aspect graphs (Koenderink & van Doorn, 1979). An aspect graph representation is a complete representation of an object at all image resolutions that relies on a small class of topological invariants in the line drawing of the object. Because these invariants are qualitative configurations of viewpoint-dependent features, becoming visible or occluded with changes in viewpoint relative to the object, the representation is a linked set of characteristic views defined by unique configurations of features (Freeman & Chakravarty, 1980). The aspect graph approach has gained in popularity as computational methods for deriving aspect graphs from three-dimensional models have been developed (e.g., Eggert, 1991; Kriegman & Ponce, 1990). Quite independently, there has been growing interest within psychology in the view-based approach to object representation. In particular, several researchers have demonstrated that object recognition of both novel and familiar objects is often viewpoint dependent (e.g., Bülthoff & Edelman, 1992; Jolicoeur, 1985; Tarr & Pinker, 1989). Such results led to the multiple-views hypothesis that objects are represented in human visual memory as a collection of viewpoint-specific images. In this approach, objects are recognized by normalizing an image of the perceived object to the nearest encoded view. One of the most crucial open questions in the multiple-views approach has been how such representations acquired and organized, and, specifically, what features are used within the representation and to delineate the boundaries between views.

At one level the aspect graph approach offers an attractive method for formally defining what is a view. Indeed, in early work on this project we explored whether human perceivers were sensitive to the qualitative changes in the feature configurations that define the boundaries between views in aspect graphs. We found evidence (Tarr & Kriegman, submitted) that humans are better able to discriminate between images of objects when the they contain qualitatively different configurations of features as defined by the aspect graph. Maxima in performance were always located at qualitative changes in the aspect graph. However, observers were also *insensitive* to some qualitative changes in the aspect graph. This latter result is not surprising — one of the most formidable problems with aspect graphs is the huge number of views per an object if all image scales are considered. Our results suggest that part of the resolution to this problem may lie in ignoring some qualitative changes, in particular, those that occur at scales too small to be relevant to the perceived shape of the object. Another possibility raised by our results was that the boundaries between views are determined primarily by qualitative changes in the silhouette of an object. Thus, regardless of scale, changes occurring in the internal contours of an object may not give rise to additional views.

Below we review some of the work that has been initiated since these original results. While we have used a diverse range of methods, the underlying theme has been an investigation into the image features that are used in long term object representations. Such features are crucial if we are to understand how the human recognition system structures object views, selects preferred views, and efficiently recognizes objects across both exemplar-specific and categorical discriminations.

Viewpoint-dependent features in the recognition of novel objects

Recognition of multi-part objects. One fundamental issue of a multiple-views representation is how to define where one view stops and another begins. While the configurations of features used in aspect graph representations may play an important role in this process, they have been criticized as too unstable, resulting in relatively complex representations (Biederman & Gerhardstein, 1993). As an alternative, Biederman and Gerhardstein suggest that configurations of non-accidental properties defining 3D volumes (geons) are used to construct the multiple-views representation. In their model each characteristic view (or as they refer to them — geon-structural-descriptions) is defined only by the configuration of the three most salient parts. Consequently, most objects will have relatively stable representations in that parts will become visible or occluded only over large changes in viewpoint. While this model is problematic for several reasons, not the least of which is the reliable recoverability of geons from images (Tarr & Bülthoff, in press), it is possible to test this model against a model in which views are defined by configurations of image features rather than 3D parts.

Biederman and Gerhardstein provide some evidence that qualitative changes do mediate recognition judgments across changes in viewpoint. They employed line drawings of the 10 objects depicted in Figure 1 (indeed these rendered images were designed to duplicate their objects in both shape and viewpoint). A sequential matching task (same/different judgment) was used in which an object was displayed for 200 ms, a mask was displayed for 750 ms, a second object (either the same or different from the first) was displayed for 100 ms, followed by a mask for 500 ms. The particular viewpoints were selected so that the middle image for each object is a 45° rotation in depth from each of the flanking images. In each triplet the image to the left has the same parts visible (no-part-change), while the image to the right has different parts visible (part-change). On each trial the center image was shown paired with either itself, one of the two flanking images, or a different object (one of the other 9). Biederman and Gerhardstein found that while both rotations were somewhat slower than the same viewpoint being displayed, the part-change condition was reliably slower than the no-part-change-condition (Figure 2). From this they conclude that view-restricted object representations are delineated by changes in the visible parts. Unfortunately, this experiment contains a serious confound: rotations in depth that resulted in a change in part visibility also resulted in a change in the image structure of qualitative features, such as those used in aspect graphs (in fact this must be the case — however, it is possible for a rotation in depth to maintain the part configuration, but produce changes in the image structure — this is addressed in the experiment following this one).

To test whether image features or 3D parts were mediating the difference in performance found between the part-change and no-part-change conditions we replicated Biederman and Gerhardstein's experiment with the rendered objects in

Figure 1. The design was essentially identical with one exception: the number of trials was doubled and on 50% of the trials the second object was a silhouette rather than a rendered image. This condition is diagnostic because image features in the bounding contour are still available for judging object identity, but sufficient features to recover geons are unavailable. Figure 3 shows the response times and Figure 4 shows the error rates for 30 subjects. There are several notable features to this data:

1) Replication of B&G's results using rendered images.
2) Replication of reliable difference between qualitative-change condition vs. no-qualitative-change condition for the silhouettes.
3) An interaction between the rotation condition and the silhouette/rendered condition whereby there is an advantage for rendered images over silhouettes when the image structure did not change, but no difference when there was a qualitative change.

Three major points may be taken from these results. First, the fact that silhouettes showed the same qualitative-change cost indicates that such changes are *not* mediated by parts as defined in geon theory. Second, because the only image features available were in the bounding contours of the silhouettes, the qualitative features mediating this effect are most likely in silhouette. Third, the interaction raised in (3) indicates that qualitative features are not the only factor in recognition judgments. Here the availability of shared *quantitative* image features (e.g., shading and internal contours) facilitated recognition when the objects were rendered, *so long as the same qualitative features were present*. However, when a change in viewpoint produced different qualitative features, quantitative image features also changed, and recognition was equal between the rendered and silhouette conditions. This supports Tarr and Kriegman's (submitted) proposal that qualitative features delineate view boundaries, but that both quantitative and qualitative features mediate recognition. The contribution of this experiment is two-fold: qualitative changes in the bounding contour may predominate over those found in internal contours, thereby reducing the complexity of the representation by keeping the number of views somewhat compact; quantitative measures may be more important in recognition *within* views and relatively unimportant in recognition *across* qualitatively different views.[1]

A second experiment investigated the degree to which the qualitative effects found in the previous study generalize to more "typical" recognition conditions. Here we have operationalized typical as a context in which the viewpoint of the object is not restricted to a small number of views (three in the previous experiment). The same sequential matching task was used with the inclusion of many more viewpoints. From the initial arbitrarily defined 0° view (the leftmost in Figure 1) new views were generated by rotations of 30°, 45°, 60°, and 90°. Additionally all pairwise combinations of these views were shown to subjects.

---

[1] It is also true that familiar objects may be represented so as to minimize the magnitude of any normalization for recognition (Tarr, 1989; Tarr & Pinker, 1989). In such cases, almost any viewpoint will match to a stored qualitatively similar view. In such view-to-view matching quantative image features will often influence recognition performance.

Results are shown in Figure 5. Each line represents all trials in which a given viewpoint appeared as the closest to 0° — thus, for example, there are 5 points for object pairs separated by 0° and the points for 45° rotations denote the data for 0°→45° and 45°→90° trials. The major result of this experiment is that regardless of initial viewpoint and whether the rotation in depth altered the qualitative image structure between the image pairs, performance was dependent on the magnitude of the rotation. Neither changes in visible parts *nor* features strongly influenced recognition. In particular, the same 3 views used in the previous experiment were embedded in the viewpoints used here (dashed lines). In this somewhat more typical and less restricted context there was *no* reliable effect between the qualitative-change and no-qualitative change conditions. This indicates that under more common recognition conditions, changes in visible parts will not determine whether recognition is viewpoint invariant or viewpoint dependent (Biederman & Gerhardstein, 1993) — rather, recognition is viewpoint dependent even when parts and image structure do not change. How do we reconcile this claim with the conclusions of the previous experiment? One possibility is that subjects are sensitive to qualitative changes, but that these are more likely to mediate the organization of the representation, not the mechanisms used in recognition. Thus, regardless of whether an object is seen in a qualitatively familiar view or in a qualitatively unfamiliar view, normalization mechanisms are used to match this to a stored view. However, when the view is qualitatively familiar, no additionally view learning is likely to occur; in contrast, when the view is qualitatively unfamiliar, it is likely to be instantiated as a new view of the object.

Recognition of single-part objects. A model of qualitative change in image structure rather than geons predicts that single volumes should also reveal effects of qualitative change over viewpoint (as in the first experiment). This experiment tested that prediction using the 3D volumes shown in Figure 6 (adapted from Biederman & Gerhardstein, 1993). Objects were each rendered in three views separated by a total of 90° of rotation in depth; the middle view in each instance was 45° from the other two and in one instance contained the same image structure and in the other instance contained a different image structure (not the views shown in Figure 1). The sequential matching paradigm was used — each trial consisted of the center view and either the same view, the qualitative-change view, or the no-qualitative change view. This experiment provides a stringent test of geon-structural description theory in that it predicts *complete* viewpoint invariance for single parts (because the invariant features are parts, not image features). Here not only are we testing whether such invariance is obtained, but also whether the predictions of the alternative model are confirmed. Specifically, an approach in which object representations are multiple-views organized by qualitative changes in image structure predicts that even simple parts will show qualitative effects across changes in viewpoint. The results of is experiment are straightforward: a reliable difference was found between the qualitative-change and the no-qualitative-change condition (Figure 7). Qualitative change in the image structure across rotations in depth produced significant performance costs that are not predicted by part-based theories, but are predicted by view-based theories, and in particular, by a multiple-views model in which views are delineated by qualitative changes in image structure.

Viewpoint-dependent features in the recognition of familiar common objects

The same tests of qualitative change may be extended to familiar common objects. Specifically, part-based theories predict viewpoint-invariance for small rotations in depth, and, in particular, rotations that do not change the visibility of major parts of the object (Biederman & Gerhardstein, 1993). In contrast, image-based theories predict that small rotations will produce viewpoint-dependent performance (because of sensitivity to quantitative features; Tarr & Kriegman, submitted). Beyond such subtle effects, part-based theories predict larger costs for rotations that result in changes in the visible parts; in contrast, image-based theories predict larger costs for rotations with a different image structure regardless of part visibility. In the experiments presented above we tested this hypothesis, concluding that qualitative features in the silhouette provide the best model of recognition performance. However, these experiments only manipulated adjacent viewpoints so that similar *patterns* of performance were predicted by both theories (no-qualitative-change views were essentially mirror-reflections about the object's symmetry plane). Here we use adjacent and non-adjacent viewpoints in a sequential same/different task so as to dissociate changes in image structure from part changes. As illustrated in Figure 8 (in the experiment, images were gray scale) rotations were selected so that the center view was a 180° depth rotation away from the left view, while the right view was only a 60° rotation from the left view. Crucially, the 60° rotations preserved the visibility of most parts, while the 180° rotation changed almost all visible parts. In contrast, the 60° rotation has a very different silhouette (qualitatively different) from the standard view, while the 180° rotation has nearly the same silhouette (discounting effects of perspective — this may be seen clearly in Figure 10 where the silhouettes are shown). Therefore, a model in which qualitative changes in the silhouette mediate recognition across viewpoint predicts better performance for the more distant rotation; in contrast, a model in which parts mediate recognition predicts better performance for the nearer rotation. Indeed, it is unclear that a parts-based model predicts any change in performance between the same viewpoint being shown and a 60° rotation so long as the same parts are visible (Biederman & Gerhardstein, 1993) — small effects may occur, but only because part visibility is sometimes altered by the rotation. However, a view-based model predicts some costs for a rotation regardless of whether the image structure is qualitatively similar — in this instance the quantitative features will provide some facilitation.[2]

---

[2] Quantitative features are more likely to be shared between the initial view and the 60° rotation; a performance advantage for the 60° view may be found. Consequently, if the 180° view is still found to have an overall advantage, this advantage in terms of qualitative features is likely to be an underestimate. Therefore, this experiment provides a stringent test of whether qualitative features in the silhouette mediate recognition and positive results would indicate that such features can predominate over competing quantitative features.

The results of this experiment are shown in Figure 9: in both response times and error rates, there was a reliable performance advantage for the 180° rotation condition over the 60° rotation condition. While the effect is not large, it is still surprising in that the images of the 180° condition are dissimilar to the 0° images in terms of both visible parts and internal image contours. What is common to these images is the shape of the bounding contour and, in particular, the configurations of qualitative features in that contour. Additionally, the finding that both conditions revealed reliably poorer performance than the same-viewpoint condition indicates that quantitative features play some role in recognition. However, when an object is rotated in depth the qualitative features in the silhouette are a better predictor of performance than either the visible parts or internal image features.

Active object exploration and recognition

A final direction we have pursued involves measuring perceptual exploratory behavior and object recognition performance under somewhat more ecological conditions. One concern with the experiments reviewed above (as well as almost every experiment in the field of object recognition) is the reliance on static images depicting the appearance of an object from a fixed viewpoint — under normal conditions human observers perceive at least a small range of adjacent viewpoints. To simulate this more natural context this study relied on a novel technique for training: subjects were presented with 6 unfamiliar 3D objects (left panels of Figures 11-16) on a Silicon Graphics IndigoXZ workstation and were told that they had three minutes to learn each object for later recognition. To facilitate learning they were given control of the displayed viewpoint via a Spaceball[3] which afforded control over all three degrees of freedom in rotation space (translation was fixed). During the subject's exploration of each object we monitored viewpoint once per a second. Such data informs us of preferred views (dwell times) and trajectories of transformation (right panels of Figures 11-16). It is expected that such results will provide specific information about the kinds of feature configurations used to acquire both feature-based and view-based object representations.

A second concern in most recognition experiments has been the generalizability of restricted recognition contexts to "normal" recognition. Features that appear to play some role in mediating performance in the context of a small number of novel objects may become far more confusable if the complete recognition set is even a portion of the objects we know about (Tarr & Bülthoff, in press). While this same problem occurs for familiar common objects, such stimuli give rise to an even more difficult issue: because of the possibility that subject have previously encoded multiple-views, apparent viewpoint-invariance may be due to optimally-placed

---

[3] The Spaceball is an input device that permits control over all 6 degrees of freedom in 3D space. The ball is fixed to a post and torque or pressure in any direction determines the rotation or translation direction as well as the magnitude of the transformation. This transformation is applied in real-time to the rendered object displayed on the screen. In this manner, grasping and manipulating the ball corresponds to manipulating the actual object. To ensure that subjects felt comfortable with this mode of interaction, they were given practice prior to the actual experiment using the Spaceball to play a game that involved manipulating an object.

views so as to minimize any normalization (Jolicoeur, 1985; Tarr, 1989; Tarr & Pinker, 1989). Thus, there is an asymmetry in what can be concluded from viewpoint-invariant performance relative to viewpoint-dependent performance: when results are viewpoint dependent, a viewpoint-dependent mechanism must be implicated, but when results are viewpoint invariant, no inference may be made regarding mechanism or representation (Tarr & Bülthoff, in press). For that reason, many researchers have opted to use novel stimuli. However, as mentioned, small sets may lead to reliance on features that do not generalize to richer contexts. To address this problem we have developed the *continuous distractor task* in which subjects learn a small set of novel objects (to control for the possibility of previous learned views), but recognize these objects in the context of hundreds of familiar common objects (all objects were colored with the same material). For each of the novel objects, the subjects' task was to name the object across rotations in depth; for each of the familiar common objects, the subjects' task was to categorize the object as living or non-living. Such a paradigm is used to control for the generalizability of features in typical recognition contexts. Therefore, while the 6 novel objects may yield viewpoint invariance if recognized in isolation, they may not do so when possible distractors include objects drawn from hundreds of real-world categories — in this instance, features that may have supported viewpoint invariance (because they were unique) will no longer be unique and viewpoint-dependent recognition mechanisms may be used. Thus, we are able to generalize performance with novel objects (where multiple-views are unlikely to have been learned) to the more common recognition context in which an object must be discriminated from a large number of other categories. Another point addressed by this paradigm is a comparison between exploration behavior during familiarization and preferred views in recognition as marked by faster response times and lower error rates. It is expected that some non-arbitrary relationship will exist between these variables. Overall, the design of the complete experiment was as follows:

1) Training with the Spaceball input device.
2) Familiarization with 6 novel objects via active exploration.
3) Brief object-name pairing training session.
4) Recognition of 6 objects and categorization of familiar common objects across rotations in depth.

The results of this experiment (to date — this and related studies are still in progress) are shown in Figures 11-16 (preferred views for each object during familiarization) and Figures 17-19 (response time functions for each object). The data are quite complex. For analysis of exploration behavior the dwell times of 25 subjects were combined and then histogrammed over an equal-area tessellation of the viewsphere. Frequency is plotted as hue, with dark purple representing the least frequently observed views and bright yellow representing the most frequently observed views. For each object, the four most preferred views were selected and plotted.[4] One immediate feature to note is that there were preferred views. Because

---

[4] Because the data are represented as points on the viewsphere, these analyses include no information about picture-plane orientation. Therefore, the depicted viewpoints are completely indeterminate with regard to the picture-plane and a particular orientation was arbitrarily selected.

these plots are averaged over 25 subjects, random exploration strategies or individually-varying exploration strategies would yield a nearly uniformly-hued sphere. Even more remarkable is the degree to which preferred views are preferred — they are significantly more frequent than surrounding views. Thus, we have verification of both the methodology and of preferred views in object exploration.[5] Beyond this, several notable patterns emerge from our initial analyses:

1) In almost every case, one of the most preferred views was an oblique side view in which two faces of the object and attached parts may be clearly seen. Such views provide more information as compared to accidental views of any single face.
2) In the oblique side views, the attached parts protrude in the silhouette, thereby defining a qualitative view distinct from a view perpendicular to either face. It is possible that preferred views of objects may be characterized as maximizing the number of qualitative features in the silhouette.
3) In almost every case, another preferred view was a top view orthogonal to the preferred oblique side view. Such views again maximize the number of parts that protrude into the silhouette.
4) Many of the histograms also revealed preferred transition paths from one view to another (moderately lighter purple trails). It may be that such paths maximally preserve the information available in the silhouette.
5) Preferred views do not seem to correspond to different configurations of parts — in many instances, two or more preferred views show essentially the same parts, but different image structure. Consequently, part-based models are unlikely to account for human object exploration behavior.

We are currently developing competence models of qualitative change as defined by the change in features in the silhouette of each object. These will be used to assess the patterns of performance depicted in these view-histograms and to better understand the kinds of features used in organizing object representations.

The second set of results from this experiment concern recognition performance across rotations in depth in the continuous distractor task. Each of the 6 novel objects was presented several times in 12 different views defined by 15° rotations around the vertical axis. For each object we have plotted mean response times and a subsampling of the views shown. The single most important result is that for 5 of the 6 objects, there is a clear pattern of viewpoint dependence. For example, the first object displayed appears to have preferred views in the 0°-15° and 165°-195° ranges. While other objects do not exhibit such well-defined minima, the range of mean response times does vary over a wide range indicating significant variation in preference among views. The only exception is the final object displayed (a teardrop shape). One possible explanation for this viewpoint invariance is that some shape features within this object were unique even in the context of many familiar common objects. What remains to be completed is a comparison of the data from this phase with the view preferences from the exploration phase and with the

---

[5] In some ways this paradigm provides a better measure than would exploration by holding an object in one's hand. In such a case subjects would be biased by natural gravitational orientation (which is likely to play a role in defining canonical views) and by the best grasp points on a given object.

predictions of a competence model of preferred views based on qualitative change. However, the overall conclusion is clear: restricted contexts that yield viewpoint-invariance may be atypical of normal recognition — when novel objects must be recognized in a more natural domain (that of familiar common objects), recognition is viewpoint dependent. Such a result offers strong evidence against part-based theories and for multiple-views theories.

Note that we are continuing to pursue this paradigm, believing it provides a powerful new method for assessing recognition performance. As we refine both our technical methods and our understanding of object learning and representation we expect the continuous distractor task along with active exploration to provide insights into the nature of canonical views, efficiency of representation, and recognition mechanisms in both exemplar-specific discriminations and categorization tasks.

Ongoing Work

Our results with single-part objects (Figure 6) were somewhat surprising. Given the extreme dissimilarity between each volume, one might predict immediate viewpoint invariance in recognition. For small numbers of dissimilar objects such a prediction would hold regardless of whether one assumed unique parts (Biederman & Gerhardstein, 1993) or image features (Tarr & Bülthoff, in press) were used. Indeed, for picture-plane rotations, Eley (1982) demonstrated that unique features within each object support viewpoint-invariant recognition. In contrast, for rotations in depth, we found that viewpoint invariance was restricted to qualitatively similar views as defined by image features (Figure 7). Consequently, the recognition paradigm used to obtain these results, a same/different recognition judgment, may be used to assess where human perceivers delineate views for objects with known aspect graphs (Eggert, 1991). This method provides significant advantages over the previously employed task of judging same or different viewpoint in that viewpoint judgments may rely on features that are not necessarily relevant to the recognition of objects — here we are directly assessing object recognition.

We have begun a series of experiments in which the stimuli employed are all solids of revolution (Figure 20) with known aspect graphs. Unlike our earlier studies, these objects are somewhat more complex (Figure 20 shows only a few of the available objects — Eggert, 1991, provides nearly 100 such objects). Added complexity makes it less likely that recognition performance and computational predictions will correlate solely because of few available features. Moreover, complexity offers more opportunities for investigating which qualitative changes are salient in recognition and which are ignored. Such results have the potential for constraining which image features are used in building view-based representations, thereby keeping the total number of views per an object to a manageable level. Secondly, these studies employ some elements of the active object exploration paradigm discussed above. Unlike earlier studies (Tarr & Kriegman, submitted), familiarization with the objects prior to the recognition tests (most likely sequential matching paradigms) will be active. Subjects will have 3 minutes to explore each object and their dwell times will be recorded for comparison to both the complete aspect graph and to their recognition performance.

We have also initiated several projects with the Bülthoff group at the Max-Planck Institut in Tübingen, Germany. One example involves the development of a multi-part object generator. The idea is to define an object world in which a restricted set of 3D parts may be attached to each other at randomly selected connection points so as to create multi-part objects similar to those shown in Figure 1. Give a set of 30 parts and 5 connection points per an object face, over 100,000,000 different objects may be generated. By adjusting a variety of parameters (e.g., which part is used as the base, the coloring of parts, ...) we can use such objects in a wide range of recognition studies. We are currently planning several recognition memory studies in which we manipulate level of discrimination (subordinate vs. categorical) across changes in viewpoint. The potential for additional studies is quite great with the added appeal that stimulus properties of shape, spatial configuration, color, texture, and illumination may all be precisely controlled.

References

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1162-1182.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci USA, 89,* 60-64.

Eggert, D. W. (1991). *Aspect graphs of solids of revolution.* Unpublished Doctoral Dissertation. Computer Science & Engineering, University of South Florida.

Eley, M. G. (1982). Identifying rotated letter-like symbols. *Memory & Cognition, 10,* 25-32.

Freeman, H., & Chakravarty, I. (1980). The use of characteristic views in the recognition of three-dimensional objects. In E. S. Gelsema & L. N. Kanal (Eds.), *Pattern Recognition in Practice* (pp. 277-288). New York: North-Holland Publishing Company.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13,* 289-303.

Koenderink, J. J., & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics, 32,* 211-216.

Kriegman, D. J., & Ponce, J. (1990). Computing exact aspect graphs of curved objects: Solids of revolution. *International Journal of Computer Vision, 5,* 119-135.

Tarr, M. J. (1989). *Orientation dependence in three-dimensional object recognition.* Unpublished Doctoral Dissertation. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233-282.

Publications and conference presentations supported by the contract

Tarr, M. J., & Kriegman, D. J. Toward understanding human object recognition: Aspect graphs and view-based representations. Submitted to *Psychological Review.*

Tarr, M. J., & Bülthoff, H. H. (in press). Conditions for viewpoint dependence and viewpoint invariance: What mechanisms are used to recognize an object? *Journal of Experiment Psychology: Human Perception and Performance.*

Tarr, M. J. Allocation of views: Behavioural evidence for a theory of human object recognition. The 17th Annual Meeting of the European Neuroscience Association, Vienna, Austria, September 4-8, 1994.

Tarr, M. J., Hayward, W. G., Gauthier, I., Williams, P. Geon recognition is viewpoint dependent. The 35th Annual Meeting of the Psychonomic Society, St. Louis, November, 1994.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (in press). How are three-dimensional objects represented in the brain? *Cerebral Cortex.*

Hayward, W. G., & Tarr, M. J. Viewpoint effects in the recognition of natural stimuli. Annual Meeting of the American Psychological Society, Washington, DC, June 30-July 3, 1994.

Williams, P., & Tarr, M. J. 3D possibility of both studied and tested objects affects object decision performance. Annual Meeting of the American Psychological Society, Washington, DC, June 30-July 3, 1994.

Other activities

The contract has supported the initiation of research collaborations with Heinrich H. Bülthoff at the Max-Planck-Institut für biologische Kybernetik, Tübingen, Germany and with Niko Logothetis at the Baylor College of Medicine, Houston, TX.

Fig 1

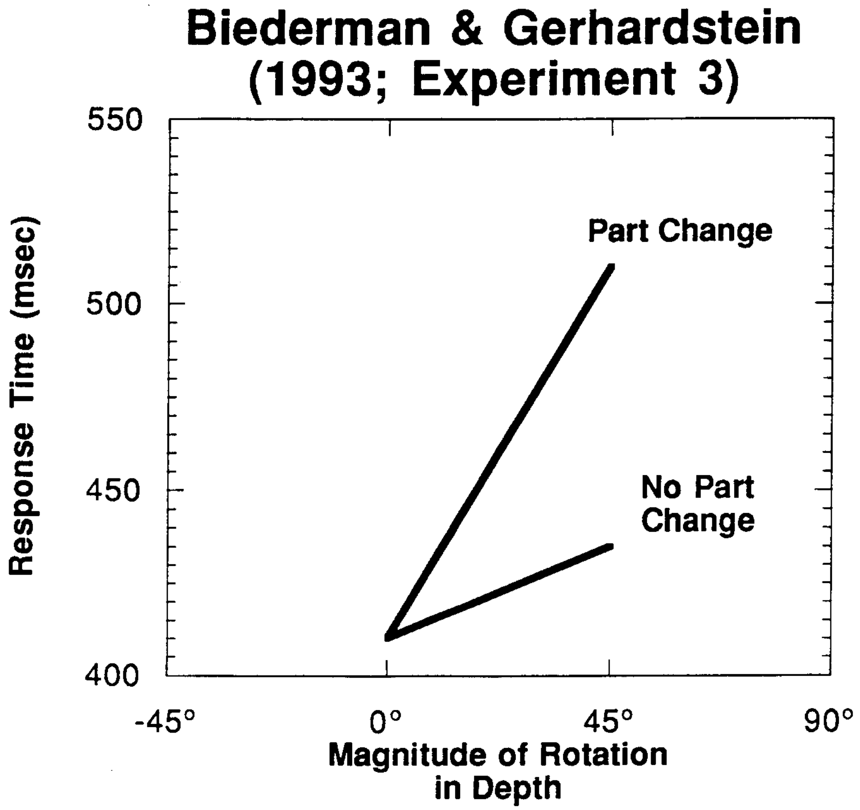
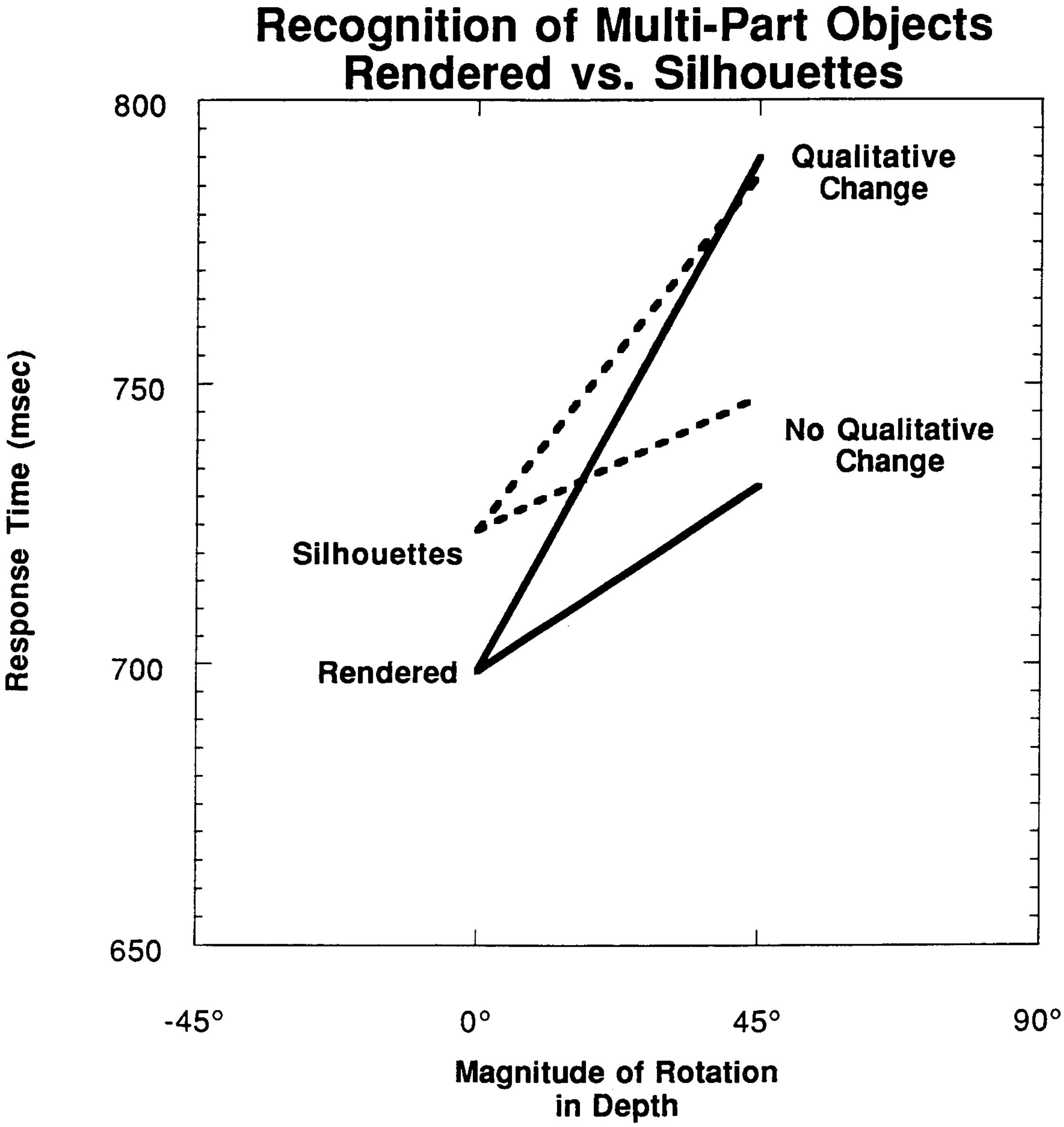

Biederman & Gerhardstein
(1993; Experiment 3)
550
500
450
400
Response Time (msec)
Part Change
No Part Change
-45°
0°
45°
90°
Magnitude of Rotation in Depth

Fig 2


Recognition of Multi-Part Objects
Rendered vs. Silhouettes

30 subjects

Fig 3

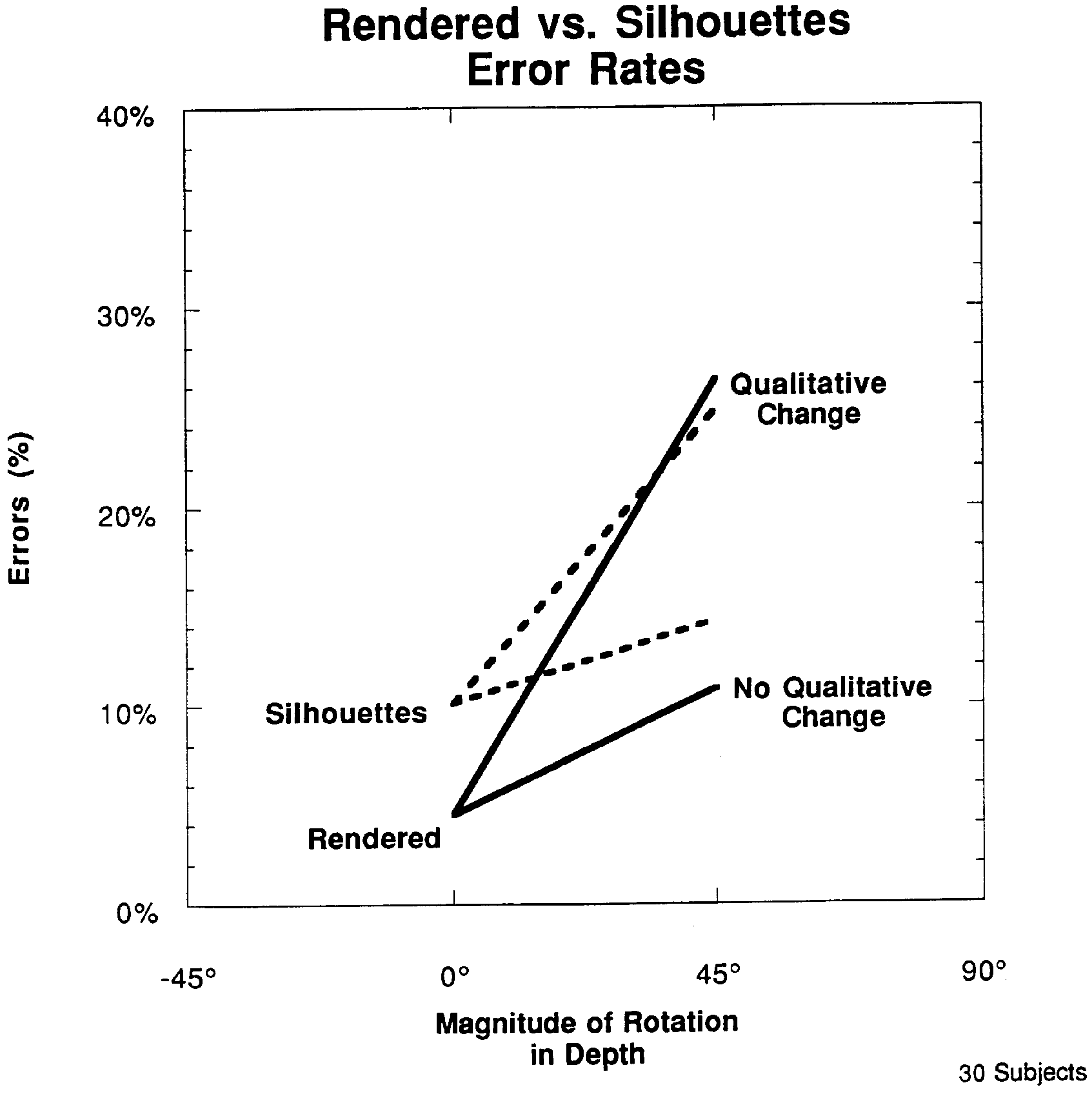
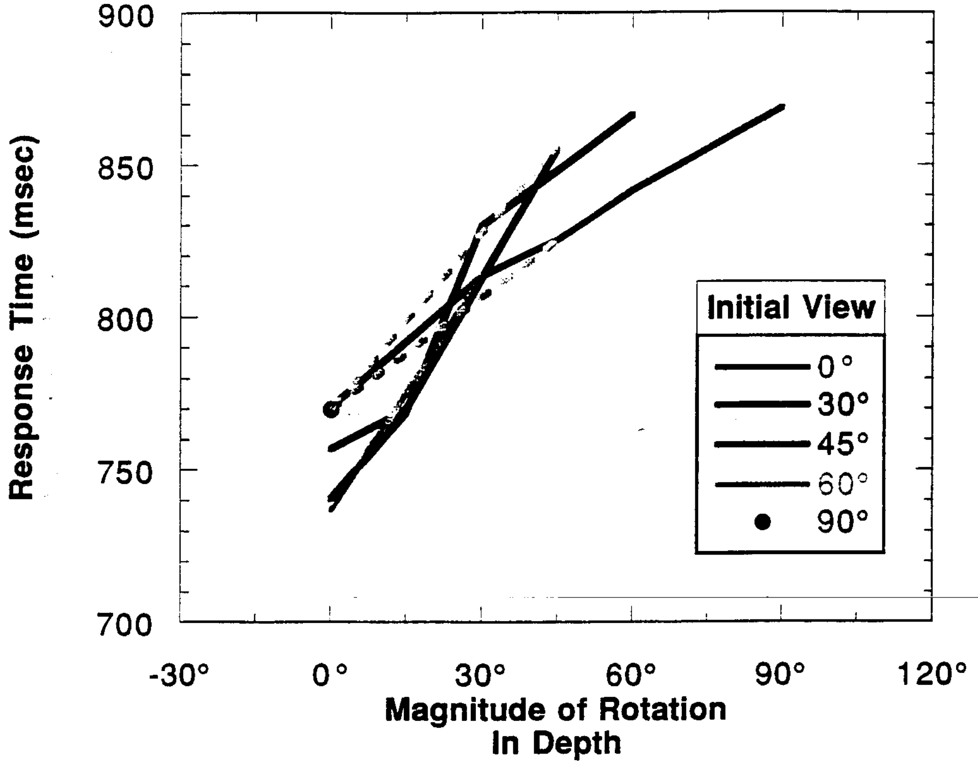
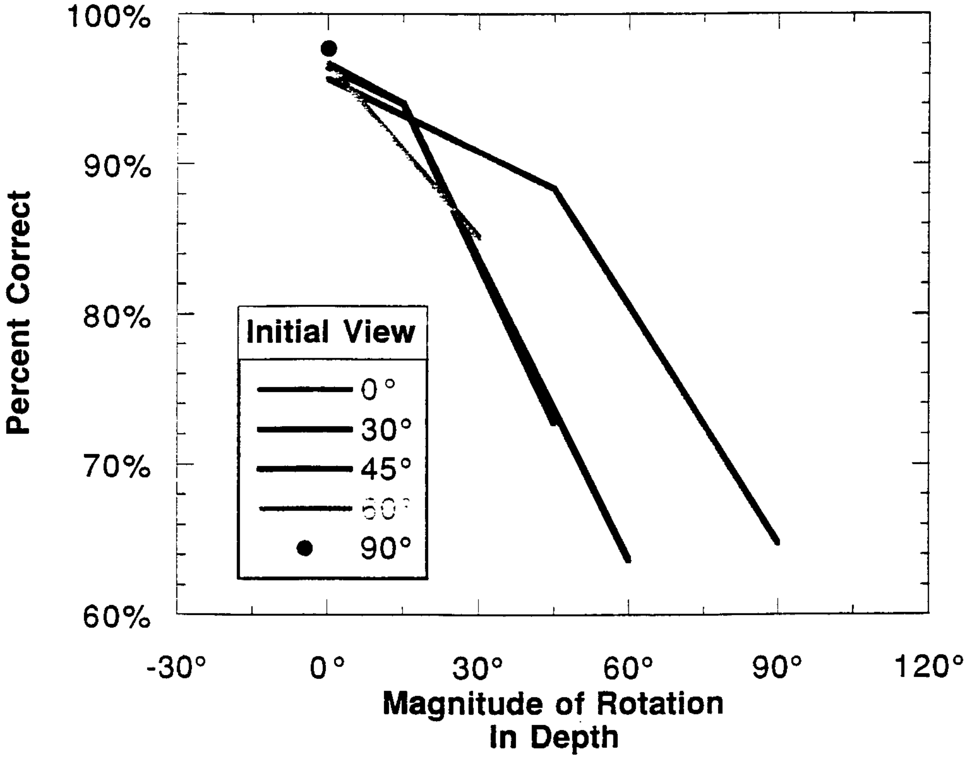

Rendered vs. Silhouettes
Error Rates
40%
30%
20%
10%
0%
Errors (%)
Qualitative
Change
No Qualitative
Change
Silhouettes
Rendered
-45°
0°
45°
90°
Magnitude of Rotation
in Depth
30 Subjects

Fig 4

Recognition of Multi-Part Objects





40 subjects

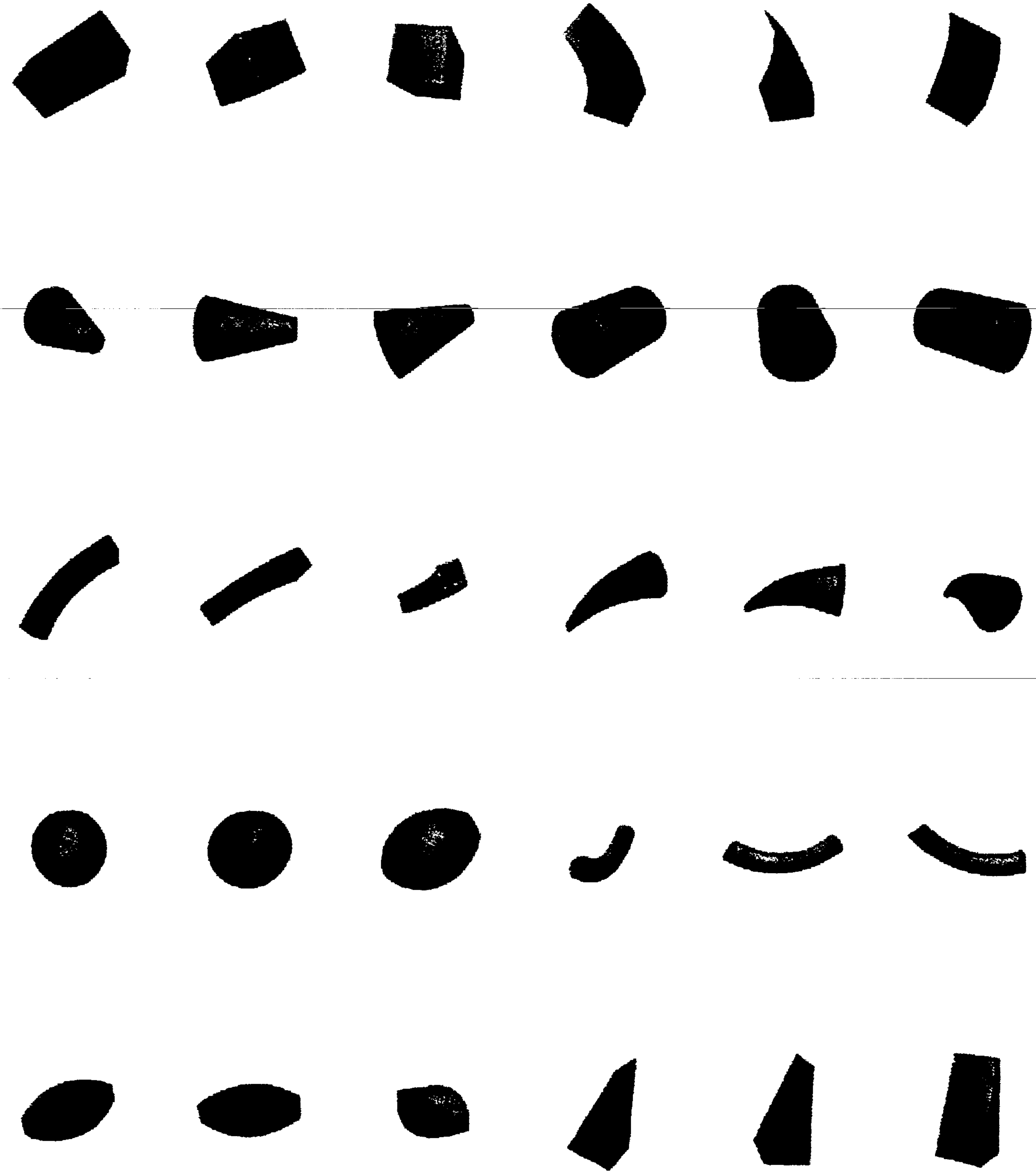
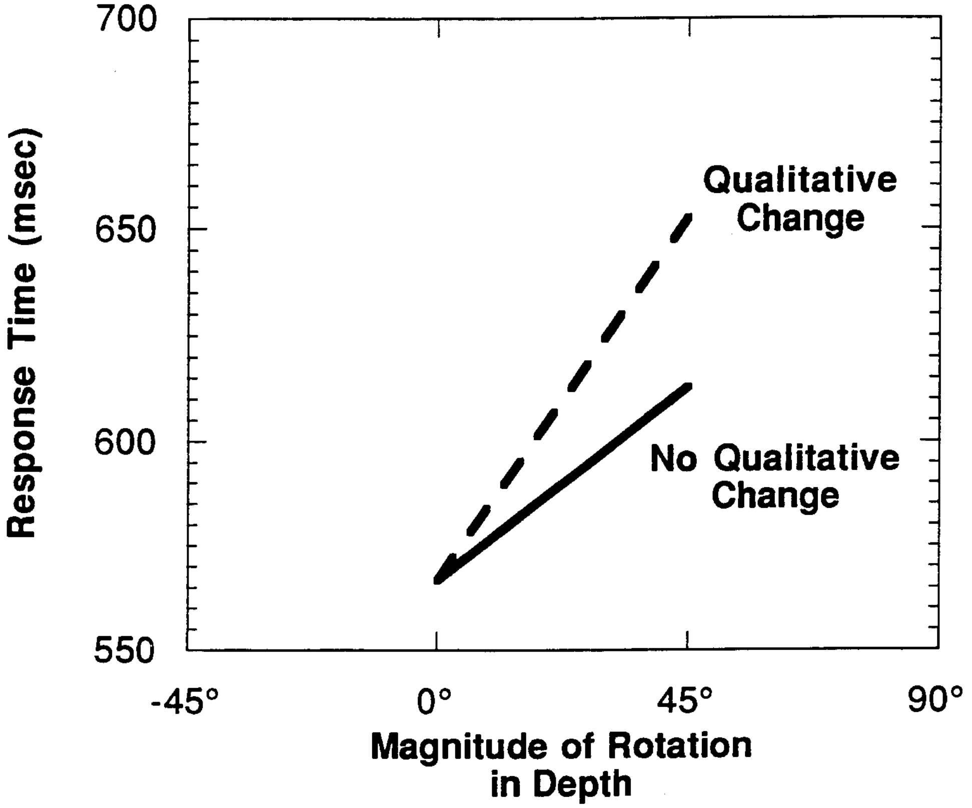
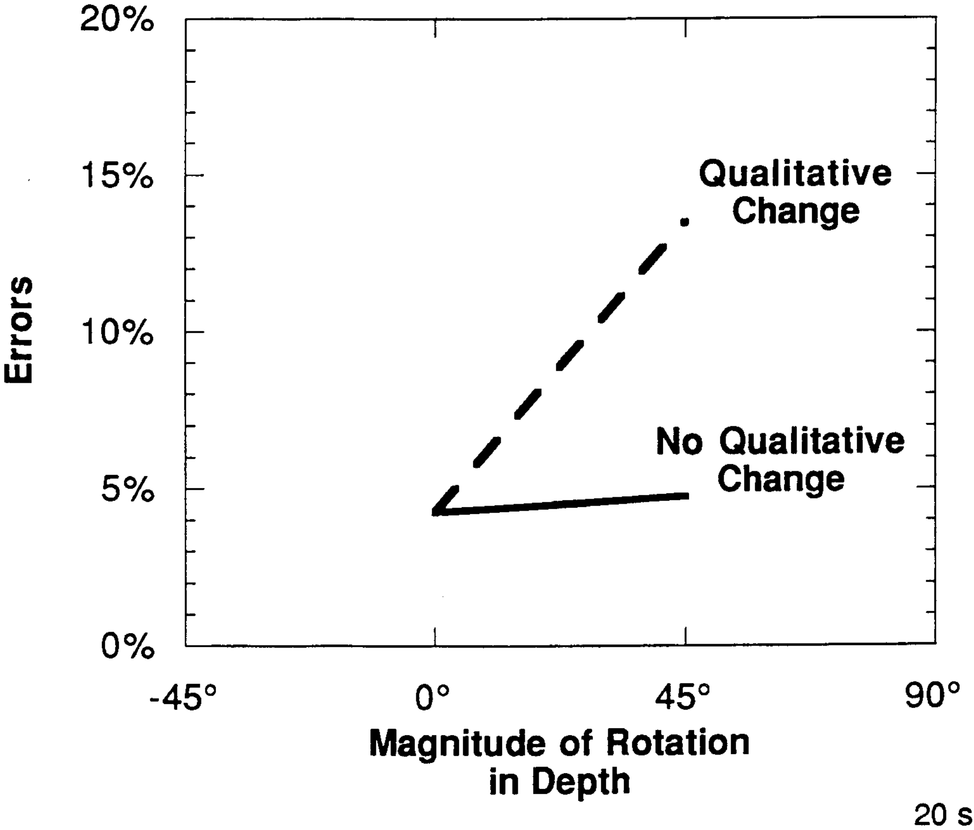


Fig 6

Recognition of Single 3D Volumes



Error Rates



20 subjects

Fig 7

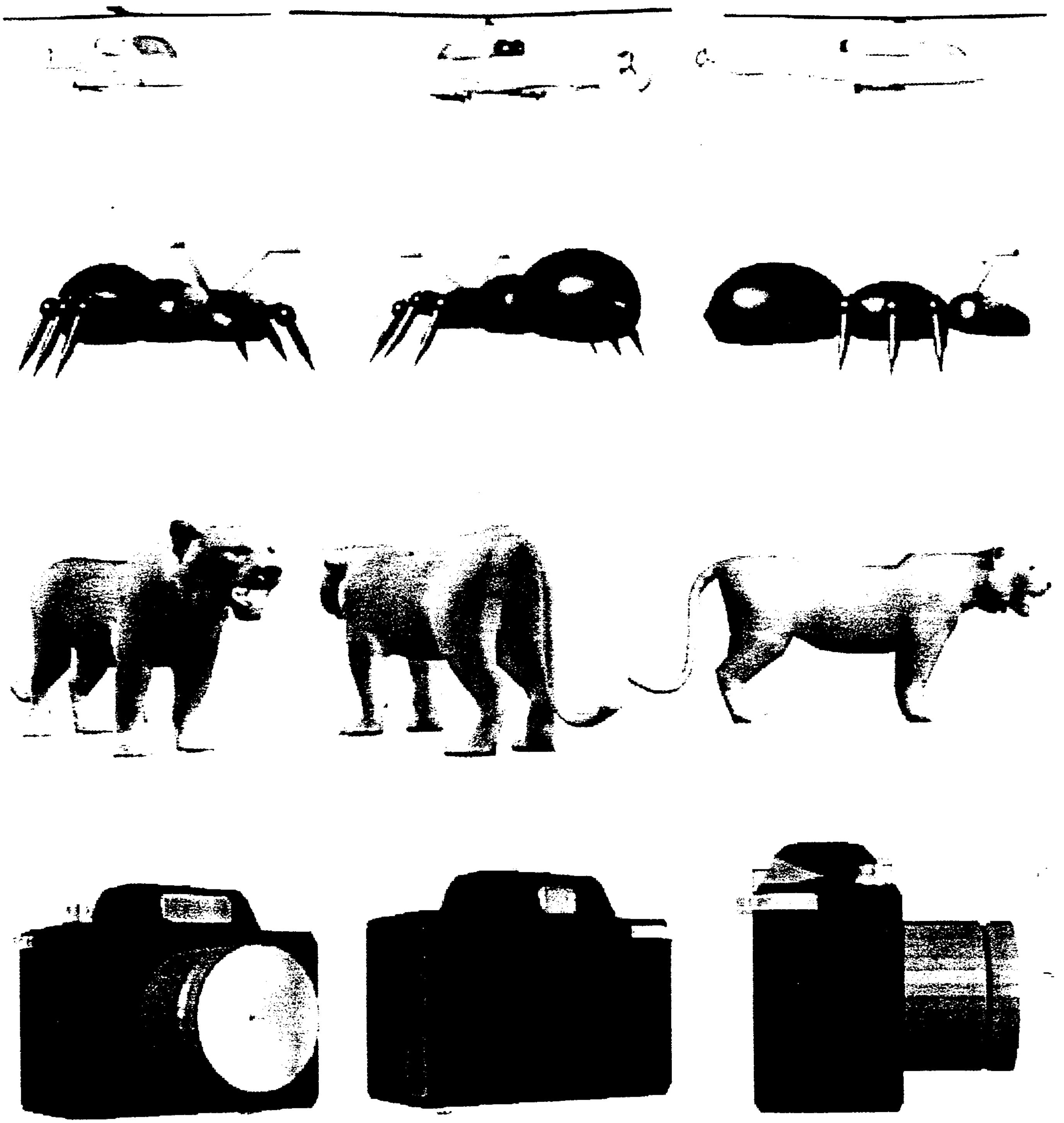
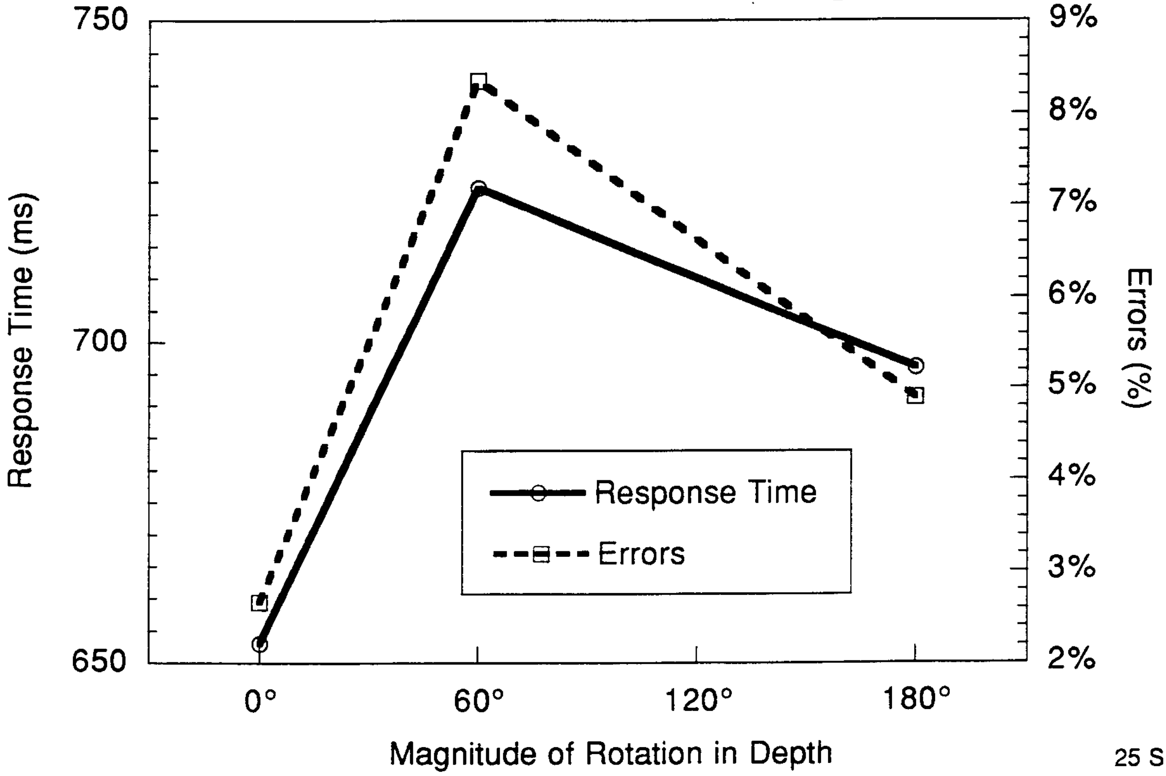


Fig. 8

Recognition of Common Objects Same/Different Paradigm



25 Subjects

Fig 9

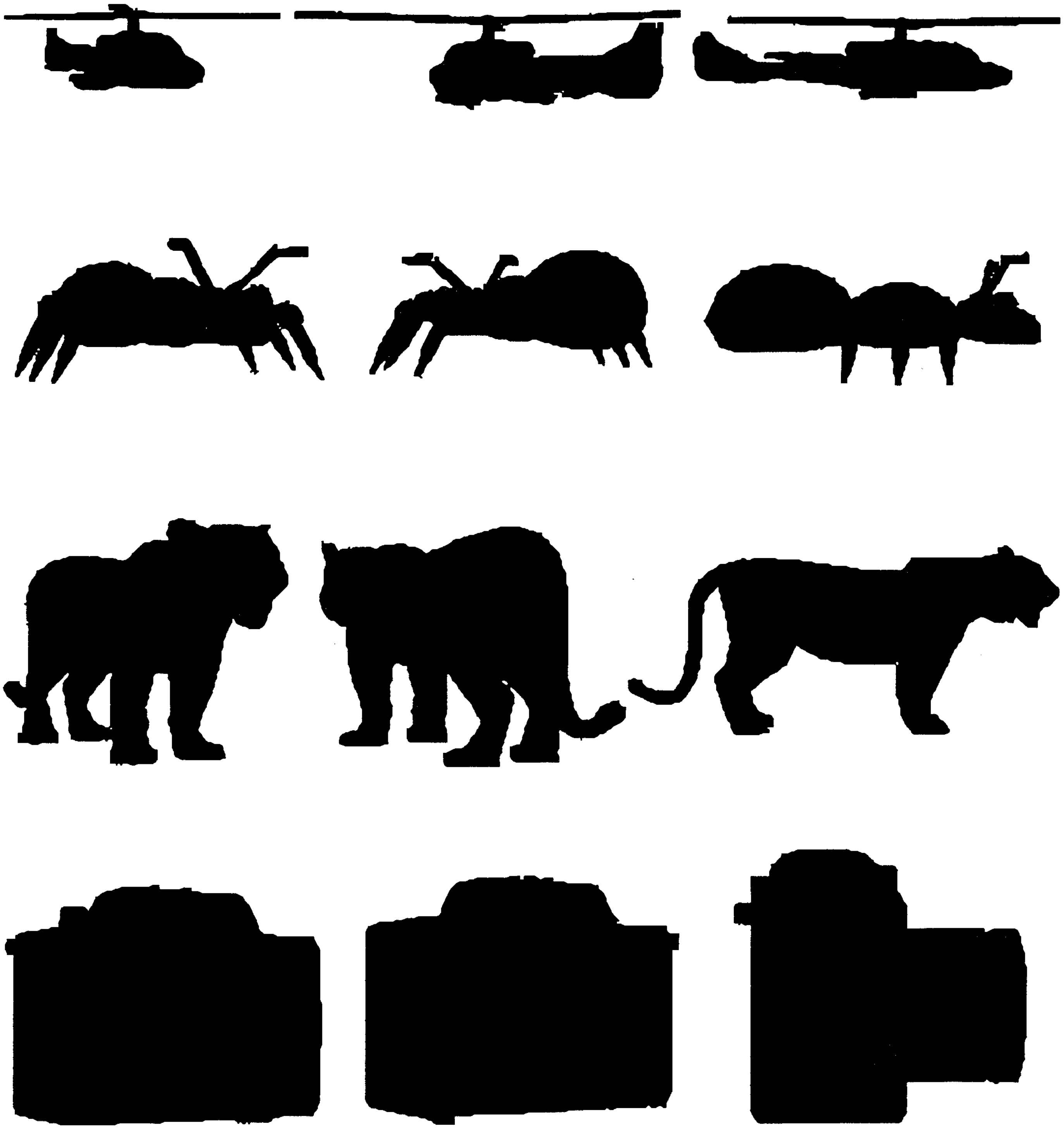


FIG 10

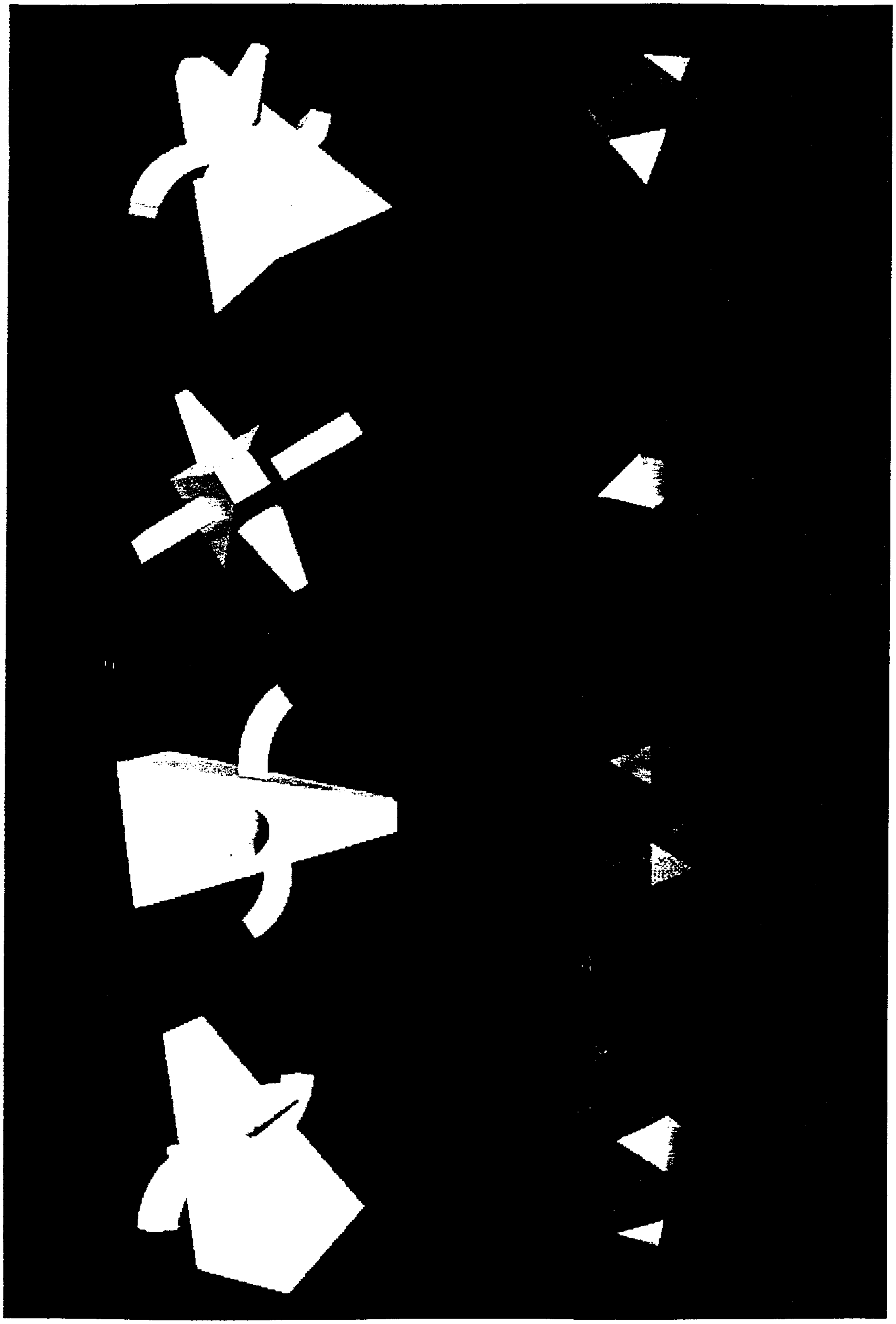


Fig 11

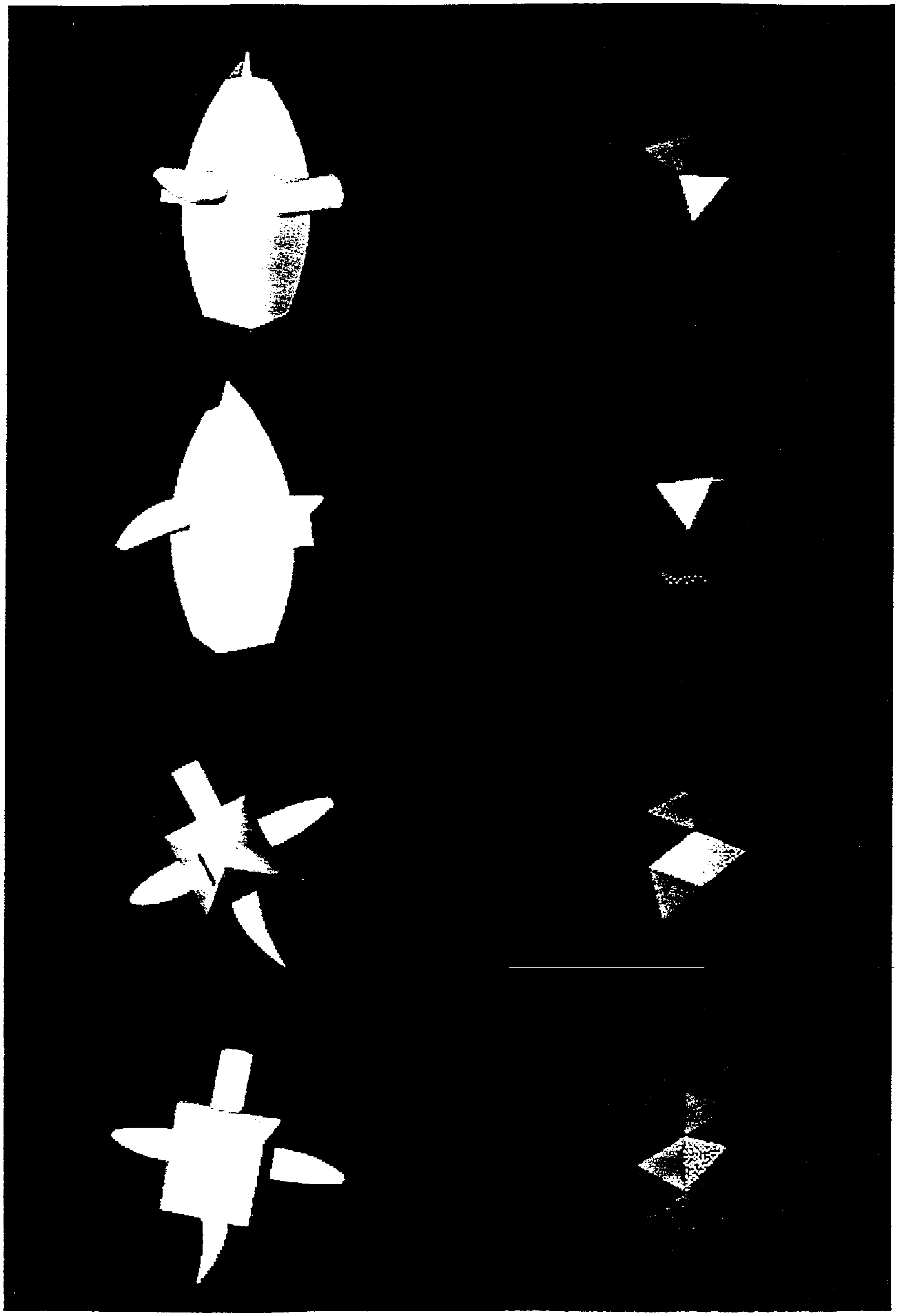


Fig 12.

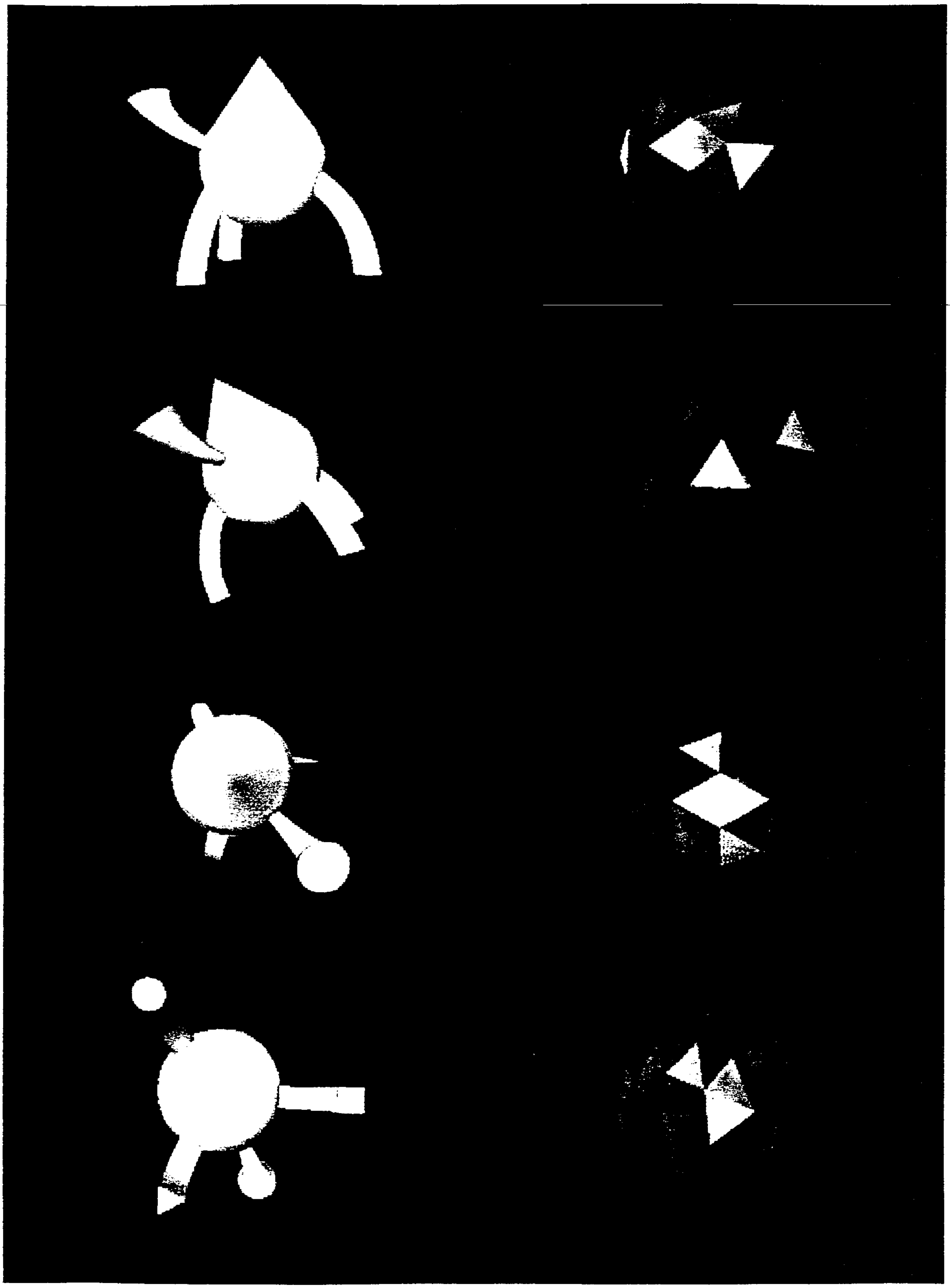


Fig 12

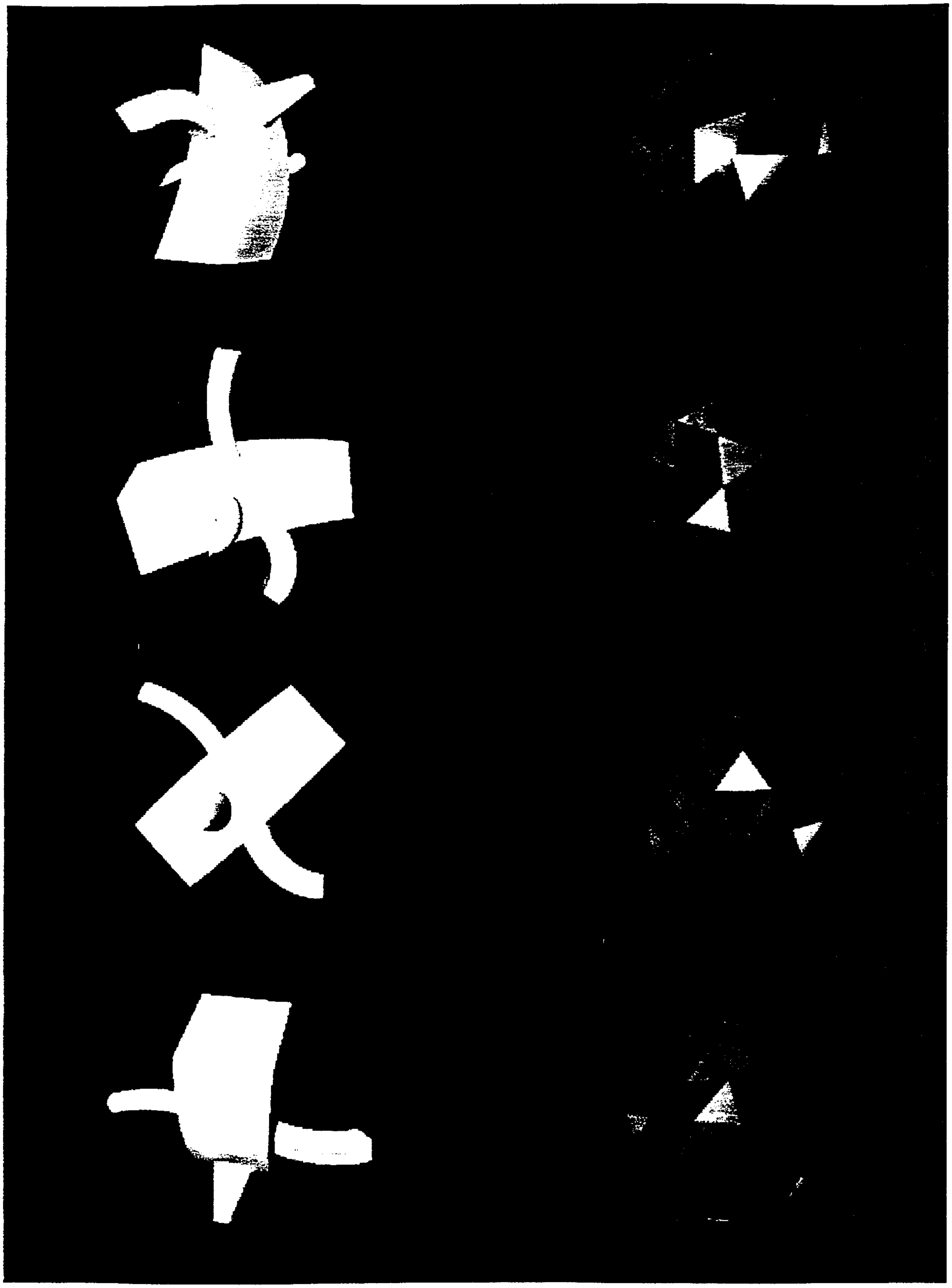


Fig 14

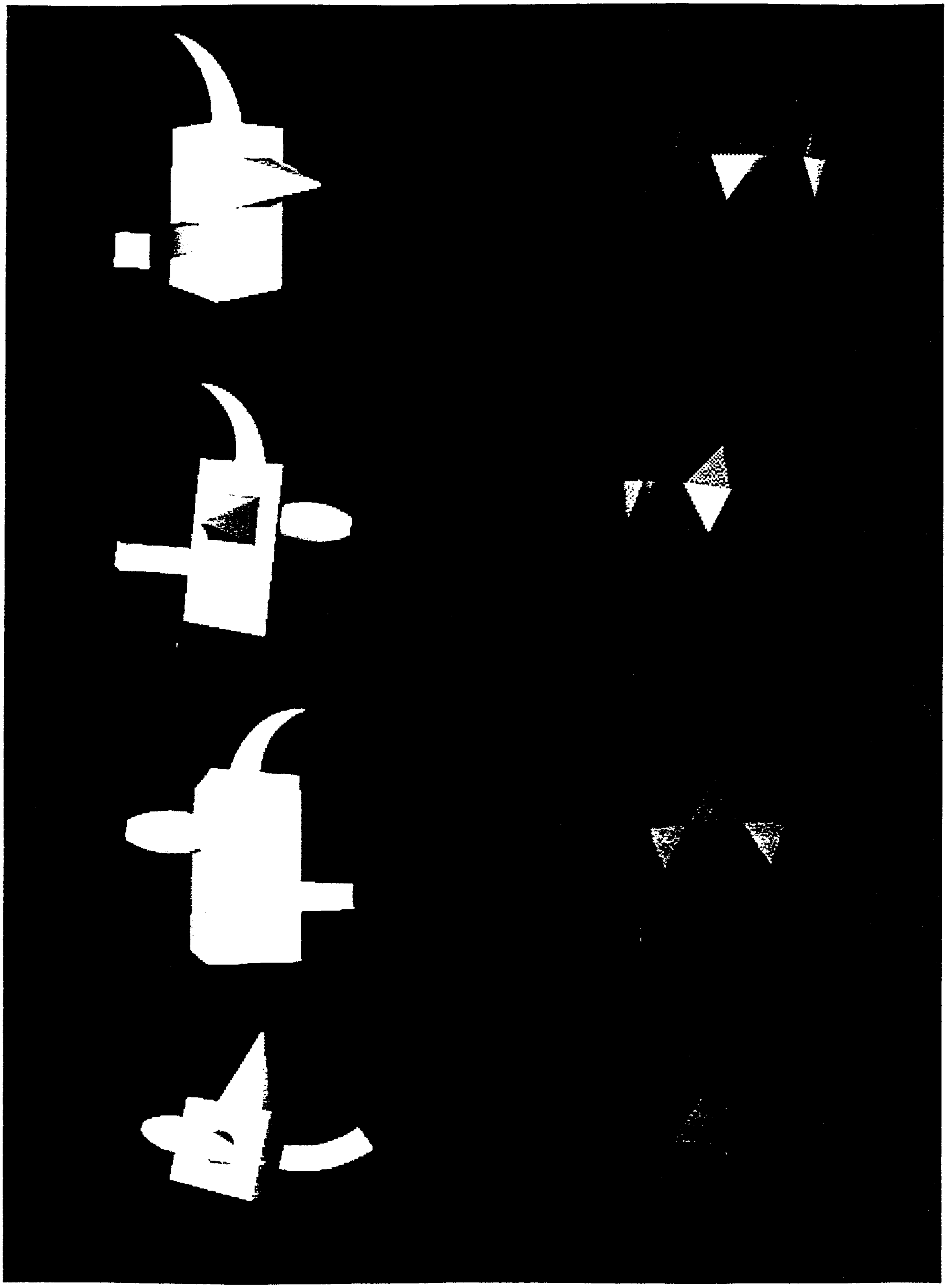


Fig 15

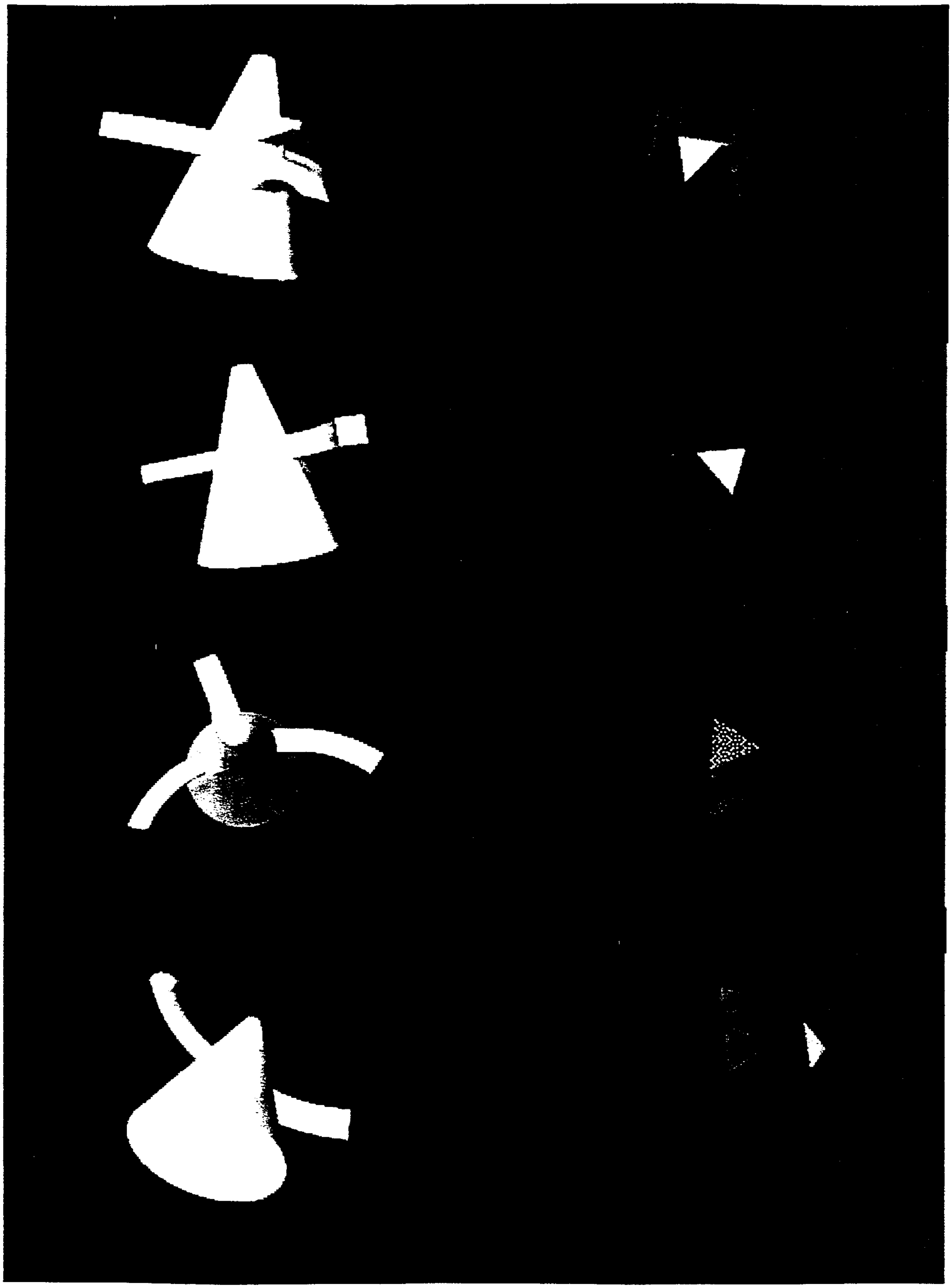


Fig 16

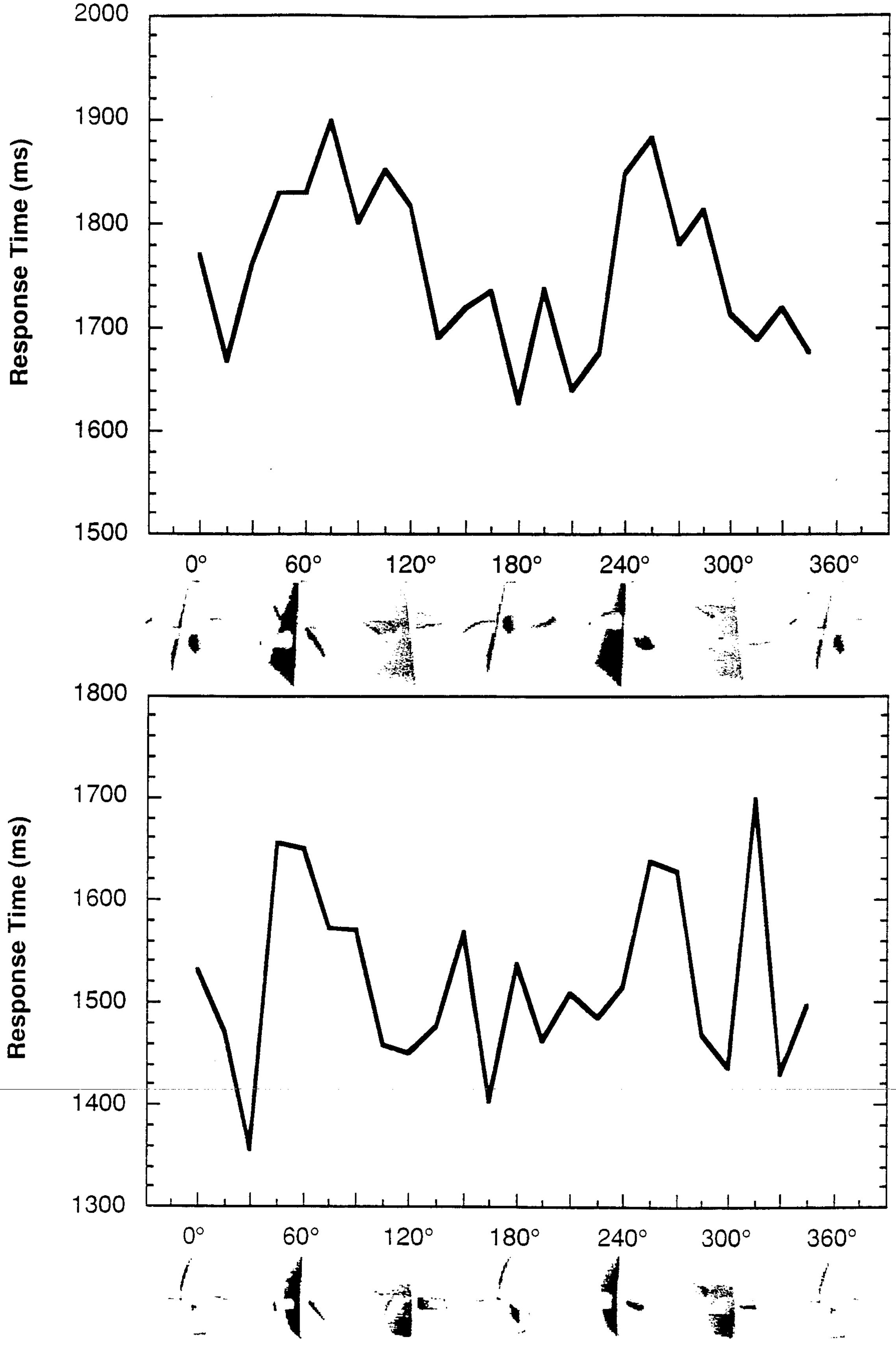

2000
1900
1800
1700
1600
1500
Response Time (ms)
0°
60°
120°
180°
240°
300°
360°
1800
1700
1600
1500
1400
1300
Response Time (ms)
0°
60°
120°
180°
240°
300°
360°

Fig. 17

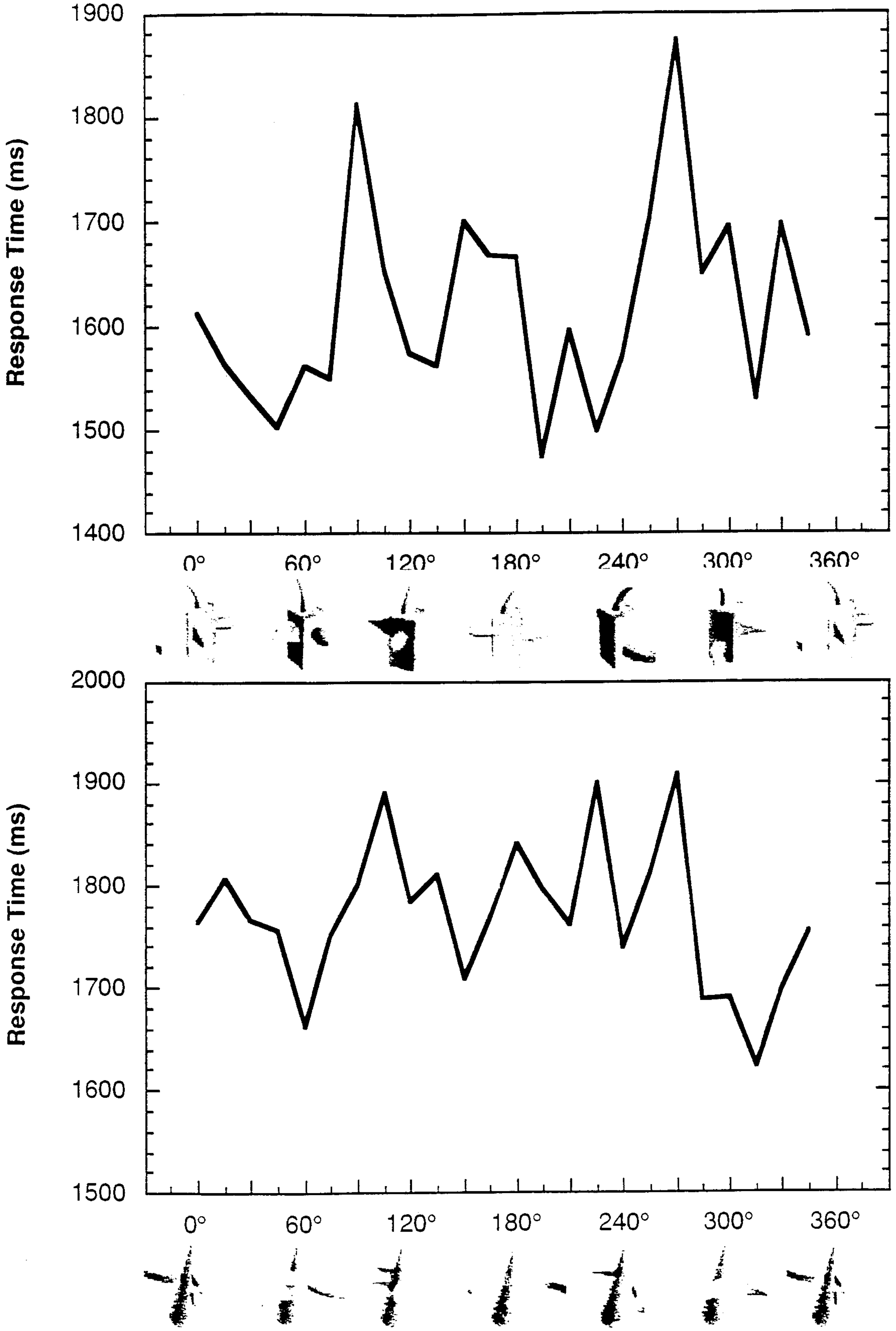

Response Time (ms)
1900
1800
1700
1600
1500
1400
0°
60°
120°
180°
240°
300°
360°
Response Time (ms)
2000
1900
1800
1700
1600
1500
0°
60°
120°
180°
240°
300°
360°

Fig 19

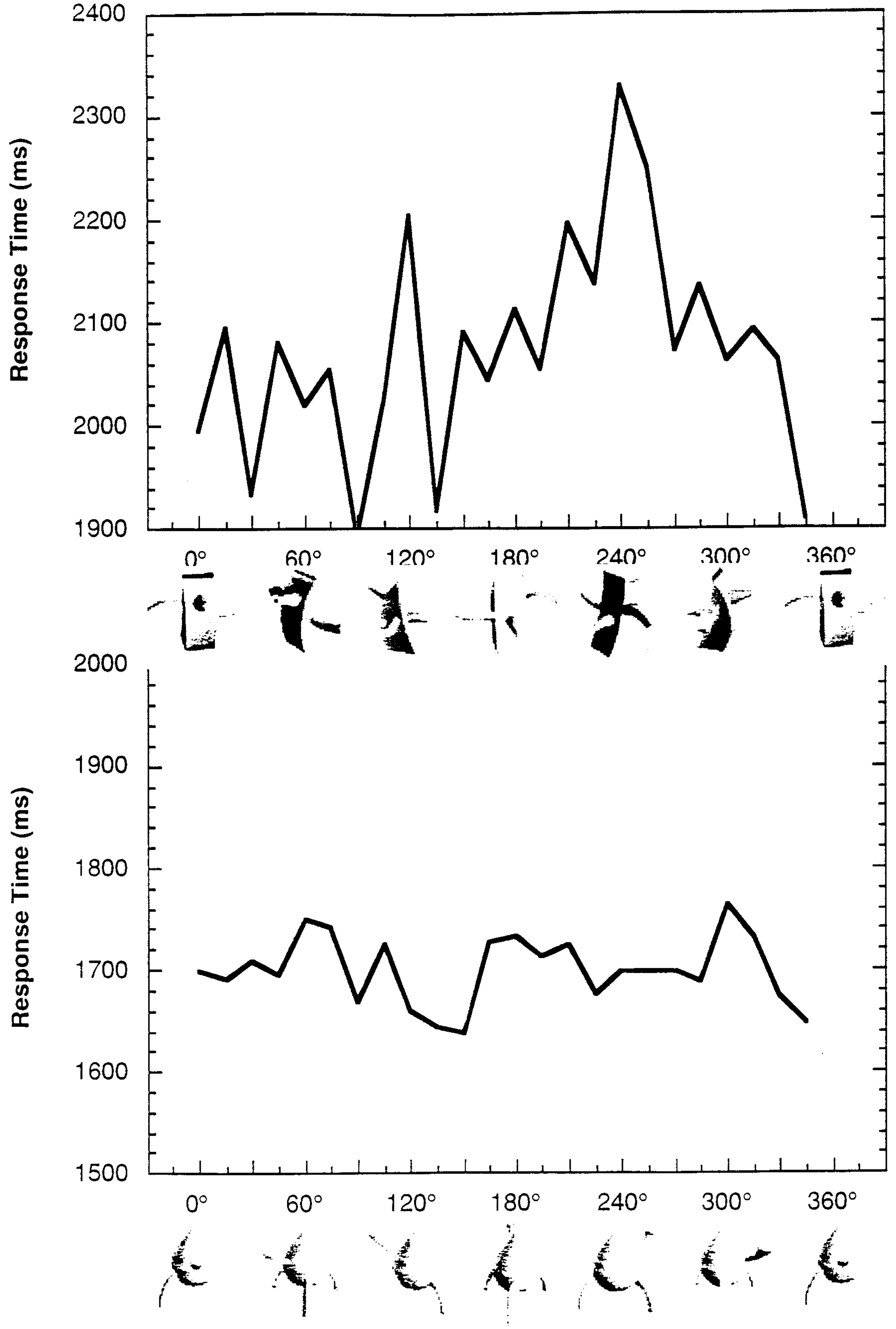
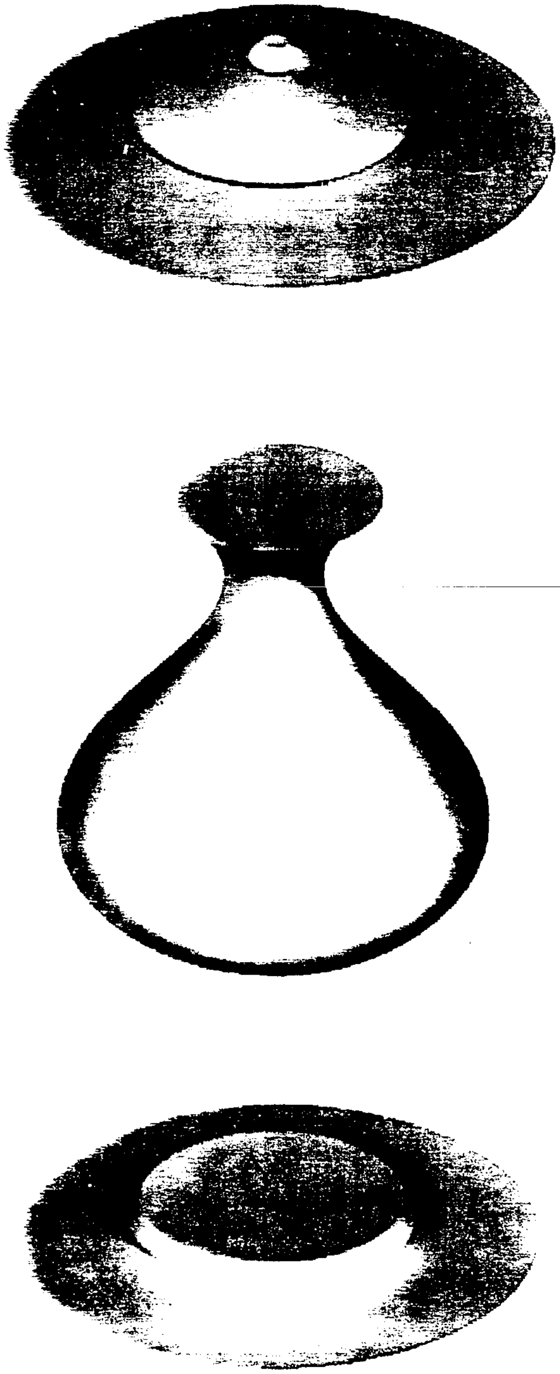

Response Time (ms)
2400
2300
2200
2100
2000
1900
0°
60°
120°
180°
240°
300°
360°
Response Time (ms)
2000
1900
1800
1700
1600
1500
0°
60°
120°
180°
240°
300°
360°

Fig. 10



Adapted from Eggert (1991)

Fig. 10