REPRESENTATION OF HUMAN VISION IN THE BRAIN: HOW DOES HUMAN PERCEPTION RECOGNIZE IMAGES?

by

Lawrence W. Stark, Michela Azzariti, Ted Blackmon, Dimitri Chernyak, Yeuk Fai Ho, Claudio M. Privitera and Huiyang Yang

Memorandum No. UCB/ERL M99/49

18 October 1999

REPRESENTATION OF HUMAN VISION IN THE BRAIN: HOW DOES HUMAN PERCEPTION RECOGNIZE IMAGES?

by

Lawrence W. Stark, Michela Azzariti, Ted Blackmon, Dimitri Chernyak,
Yeuk Fai Ho, Claudio M. Privitera and Huiyang Yang

Memorandum No. UCB/ERL M99/49

18 October 1999

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Invited paper, accepted for publication in *JEI 2000*, special issue on Electronic Imaging and Human Vision

REPRESENTATION OF HUMAN VISION IN THE BRAIN: HOW DOES HUMAN PERCEPTION RECOGNIZE IMAGES ?

Lawrence W. Stark, Michela Azzariti, Ted Blackmon, Dimitri Chernyak, Yeuk Fai Ho, Claudio M.Privitera, Huiyang Yang

Neurology and Telerobotic Units, School of Optometry
University of California at Berkeley 94720

0. ABSTRACT

The repetitive scanpath eye movement, EM, sequence enabled an approach to the representation of visual images in the human brain. We supposed that there are several levels of "binding" --- semantic or symbolic binding; structural binding for the spatial locations of the regions-of-interest; and sequential binding for the dynamic execution program that yields the sequence of EMs. The scanpath sequences enable experimental evaluation of these various bindings that appear to play independent roles and are likely located in different parts of the modular cortex.

EMs play an essential role in top-down control of the flow of visual information. The scanpath theory proposed that an internal spatial-cognitive model controls perception and the active looking EMs. Evidence supporting the scanpath theory includes experiments with ambiguous figures, visual imagery, and dynamic scenes. It is further explicated in a top-down computer vision tracking scheme for telerobots using design elements from the scanpath procedures. We also introduce procedures --- calibration of EMs, identification of regions-of-interest, and analysis and comparison programs ---- for studying scanpaths. Although philosophers have long speculated that "we see in our mind's eye", yet until the scanpath theory, no strong scientific evidence was available to support these conjectures.

Keywords: top-down vision, scanpath theory, representation, structural binding, sequential binding, read-out mechanisms, computer vision

1. INTRODUCTION

Vision. Human vision is complex. The essential problem is how to match bottom-up, BU, confirmatory signals coming both from the wide peripheral visual field, with only low resolution, but with high sensitivity for moving objects, and from multiple high-resolution glimpses by the centrally located fovea, a small, circa one-degree region. These foveal regions-of-interest, ROIs, are sequentially visited by a string of fixations, shifted by a string of saccades, rapid eye movement, EM, jumps, and are simultaneously matched by top-down, TD, symbolic, spatial and sequential representations or bindings of the hypothesized image.

When the retinal field is mapped onto the visual cortex, there is a considerable geometrical magnification of the signals coming from the fovea, and a consequent reduction of signals coming from the periphery. The log-polar distortion (Figure 1) is a rather good depiction of the geometry of the visual image mapped onto the visual cortex.(Dow et al. 1981; Schwartz 1984) When the high-resolution fovea is fixated on a particular part of the picture, such as the sailboat at the edge of the beach, that ROI is magnified on the visual cortex. Contrariwise, those parts of the image lying on the periphery of the retina are minified, so that only color and textural segmentation of large areas can be appreciated at the low resolution of the periphery. Two such

foveal and peripheral representations (Figure 1) provide an indication of the kind of BU information coming into the visual brain.

FIGURE 1 Log-Polar Distortions of a Picture
Two fixations (left and right panels below) on original picture (upper) show log-polar distortions with high cortical magnification (irregular shapes, lower left and lower right) of successive foveal ROIs (circles), as well as minification of peripheral regions likely captured as textured or colored segments (surrounding small squares, lower).

Scanpath EMs. Observations of repetitive sequences of EMs while a subject looks at a picture led Noton and Stark (Noton and Stark 1971a; Noton and Stark 1971b; Noton and Stark 1971c) to the experimental definition of the scanpath as an idiosyncratic alternation of glimpses (called fixations or foveations) and rapid jumps of eye position (called saccades) to various ROIs, in the viewed scene. (Crosby 1990; Jeannerod, Gerin, and Pernier 1968; Locher and Nodine 1974; Mackworth 1978; Mackworth and Bruner 1970; Mackworth and Morandi 1967; Mandler and Whiteside 1976; Parker 1978; Schifferli 1953; Senders, Fisher, and Monty 1978; Yarbus 1967) EMs and attention shifts are very closely linked; it is only in an unusual laboratory situation that the two can be putatively separated; psychologists generally study attention shifts without measuring EMs, and neurologists study EMs without measuring attention shifts.(Mackeben and Nakayama 1993; McPeek, Maljkovic, and Nakayama 1999; Nakayama and Joseph 1998)

Examples of two such EM scanpaths (Figure 2) are shown for the classical ambiguous figure, "Two Faces or a Vase" (Figure 2, left column). Depending upon the TD internal cognitive model, the subjects "sees" one or another of these two interpretations.(Ellis and Stark 1979; Stark and Ellis 1981) Some control over the current interpretation can be induced by 'priming' the subject with a non-ambiguous distortion of the ambiguous figure (Figure 2, right column). Not only does the subject report on the interpretation, but her EMs show quite different patterns (left column), easily noted to be appropriate for the comparison of two faces in one case (upper row) or viewing the vase in the other case (lower row). Of course, the actual picture viewed after the priming was the same in both cases. This evidence from Ellis and Stark(Ellis and Stark 1979) supports the scanpath theory (Fig. 2 provided by Privitera and Weinberger, 1998).

FIGURE 2 EMs while Looking at an Ambiguous Figure: The Ellis Experiment
Identical ambiguous figures of vase (lower left) and two faces (upper left). EMs superimposed on ambiguous figures as they were actually seen following exposure to priming stimuli (right).

Dynamic Scenes. We introduce our EM recording and analysis methods in the context of an ongoing experimental study of interest to our research group, that is, the nature of scanpath EMs when looking at dynamic displays (Figures 3 and 4). Animations, constructed graphical scenarios (Figure 4), showed a set of moving cars on intersecting roads.(Blackmon et al. 1999) Subjects were clearly interested in the possibility of collisions. Their internal models evidently developed dynamical internal representations that guided their EMs to follow these objects of interest. Snapshots at five-second intervals, of a fifteen second presentation of a reduced dynamical graphic display, are shown (Figure 3) with time along the oblique axis. EMs were recorded throughout, and the heavy dots represent the location of EM fixations and thus, of the fovea during the five seconds preceding each of the snapshots.

FIGURE 3 Dynamic Display with EMs
Animation of dynamical scenarios (illustrated as snapshots every five seconds time proceeding from lower left to upper right). EM positions (black circles) taken every 50ms are integrated over

the preceding 5s, and are superimposed onto snapshot images; they represent the basic data captured for this experiment.

Scanpath Theory. The scanpath theory proposed that an internal spatial-cognitive model controlled both perception and the active-looking EMs of the scanpath sequence (Noton and Stark 1971a; Noton and Stark 1971b; Noton and Stark 1971c) and evidence for this came from new quantitative methods, experiments with ambiguous figures (Figure 2),(Ellis and Stark 1979; Stark and Ellis 1981) and more recently from experiments on visual imagery (Brandt et al. 1989; Brandt and Stark 1997; Kosslyn 1980) and from MRI studies on cooperating human subjects. (Kosslyn 1994) (See Appendix M.) The scanpath theory is illustrated (Figure 4) with the actual EM positions (upper left). After analysis of the EMs into fixations and smooth pursuits (upper right) the foveations and smooth pursuit tracking episodes are numbered in sequence.

Smooth pursuits are EMs that continually track a moving object; in so doing, they place the fovea on top of the moving target.(Weirda and Maring 1993) Thus, a smooth pursuit may appear to the perceiving brain as would a fixation on a stationary target. Of course, the brain is informed by efferent copy of the EM commands of the motion of the eyes as well.(Lawden et al.) We treat these smooth pursuits in a similar fashion to fixations for the purposes of the scanpath theory and apply the quantitative measures we use to analyze the scanpath (see below).

The string of such glimpses, both fixations and smooth pursuits, is shown in a more abstract form (lower right) with string labels, "FABCBE." A non-iconic representation of an operational model capable of generating such a scanpath (lower left) includes both fixations (lettered boxes) and commands to EMs to shift the fovea (circles with arrows). Glimpses and EM commands alternate. These sequences are not deterministic, but rather probabilistic. The solid lines are for the particular experiment illustrated (Figure 4), while the dashed lines represent other experimental scanpaths measured during repeat studies.

FIGURE 4 Scanpath Theory
EM positions (q 50ms) during dynamic display shown as a connected sequence (upper left) while the dynamical ROIs visited form a connected sequence (upper right and lower right). By numbering or letter identification of smooth pursuit or static fixations, this sequential string of visited ROIs could be defined. A non-iconic model of alternating perceptual ROIs (lettered squares) and saccadic EMs (circles with arrows) is shown by solid arrows for the experiment presented. This is the "feature ring" of the scanpath theory. On other presentations of the stimulus, other ROIs and sequences could be formed (dashed arrows).

Binding. How is the internal model distributed and operationally assembled? The concept of binding speaks to the assigning of values for the model and its execution by various parts of the brain. (Stark et al. 1999; Wolfe 1998) We assume that there are several levels of "binding" --- symbolic or semantic binding, spatial binding for the structural locations of the ROIs,(He and Nakayama 1992; Nakayama, He, and Shimojo 1995; Ploner et al. 1999) and sequential binding for the dynamic execution program that yields the sequence of EMs. The EM scanpath approach is complementary to studies being vigorously pursued in other laboratories --- on attention shifts, (Wolfe, Alvarez, and Horowitz in preparation, 2000) without recording concomitant EMs, on functional magnetic resonance imaging, fMRI (see Appendix M), and positron-emission tomography, PET, in man, and on the neuroanatomy and neurophysiology of animals. We thus try to use this current neurological information to localize where these different aspects of the spatial-cognitive model might exist in the brain.

Aim. The aim of this study is to attempt to dissect out different forms of binding and to test their respective contributions to the experimental scanpaths. The use of the scanpath in robotic computer vision both illustrates a successful application, and provides a detailed operational explication of the scanpath theory. Finally, we speculate on the nature of perception and the cortical organization underlying vision.

2. METRICS and ANALYSES

Stimuli. Several sets of stimuli were used -- static and dynamic pictures for the 'looking experiments' and also a series of grid patterns for the 'visual imagery experiments.' A complex of computer workstations were interconnected in our Berkeley laboratory Internet; these included an Indigo SGI for display, the SGI and a PC-586 for collection of EM data, the PC-586 alone for choice, CH, experimental display and data collection, and for analysis either of these work stations. Each of the computers was dedicated to running a part of the complex experiment. Software generated the protocol, displayed the stimuli, recorded the EMs, analyzed the collected data, compared (vector) sequences of fixations for analysis, and displayed the intermediate and final results. Software ranged from special lab programs to Matlab toolboxes for certain functionalities.

2.1. Visual Imagery Scanpath Experiments with Grid Picture

Simple grid with a pattern of alphabetical symbols is viewed by the subject for two seven-second periods (Figure 5, upper row, left two grids). The EMs show somewhat repetitive scanpaths for these 7-second looking periods (middle row, left two grids). Next a blank grid is displayed (upper row, third grid). When asked to engage in visual imagery (Kosslyn 1980; Kosslyn 1994; Kosslyn and Osherson 1995; Singer, Greenberg, and Antrobus 1971) and to imagine the previous pattern, the subject makes a scanpath (middle row, third grid) very similar to those made when looking at the patterns. (Brandt et al. 1989; Brandt and Stark 1997; Stark, Choi, and Yu 1996) However, at this time there is no external pattern, only the subject's memory, that is, the internal cognitive spatial model, to guide the EMs. Then the subject is asked to draw the pattern from memory (right grids); this is useful as an operational instruction to impress upon the subject to localize carefully the components of the pattern to be remembered. Finally, a finite-state automata is derived from the experimental data (bottom row, middle) whose frequency of transitions is indicated by number and strength of connecting arrows; these could also be placed as coefficients in a Markov matrix (bottom row, right). (Brandt et al. 1989; Brandt and Stark 1997) Figure 5 has been modified from Stark, Choi, and Yu, 1996.(Stark and Choi 1996)

FIGURE 5 EMs while Engaged in Visual Imagery: The Brandt Experiment
Scanpath EM sequence is almost the same for the second looking presentation (middle row, second grid) as for the first visual imagery presentation (middle row, third grid). During the visual imagery presentation, no information about the location of the alphabetic symbols, Fs, was available; thus, the remembered representational model must have controlled the scanpath in a TD fashion. Quantitative metrics could be obtained in the analysis procedure (lower row) by creating a finite state automata (middle) for generating the scanpath; then transition probability coefficients could be arranged in a Markov matrix (right) for later statistical analysis.

2.2. EMs

EM Recording. A video camera system for EM tracking was convenient, but as with all EM recording systems, required repeated calibration.(Llewellyn-Thomas 1968) With calibration any drifts or non-linearities could be removed. Fixation identification algorithms separated the raw

EMs into fixations, rapid saccadic EMs, and smooth pursuits (Figures 4, 5, and 6) so that further analysis could be done on the loci and sequences of fixations considered as a string of loci of glimpses, the vector of active looking. Besides recording EMs, we also introduce another experimental technique. Here, the subject is asked to click a mouse cursor onto various ROIs in the picture, or grid pattern, displayed on the computer screen. When presented with an empty grid, the subject must depend only upon memory bindings.

EM Analysis. EM experiments must be carried out carefully. Calibrations are used to linearize the data and to avoid drift. While the trajectories of the EMs may be displayed (Figure 5), the locations of fixations and smooth pursuits and of saccadic Ems are usually automatically identified. This is easier for static stimuli, than for dynamic stimuli (Figure 6), but with classification programs EM identification is quite feasible (Figure 7). The programs in our laboratory have a long history dating back to 1959, at Yale University, going through our stay at M.I.T., the University of Illinois, and our long residence (31 years) at Berkeley. Although it is impossible to list all the names of the persons contributing to these programs, we must mention Robert Payne, Alan Sandburg, John Semmlow, Christian Freksa (Stark et al.) (especially for the scanpath analysis), A. Terry Bahill, Michael Clark, An Nguyen, Yun Choi, Yong Yu, and most recently, Y. F. Ho.

FIGURE 6 EM Trajectories and Classification
Trajectories of EMs displayed as functions of vertical and horizontal angles (solid lines) and time. Location of the dynamic objects shown as dashed and continuous lines (lower), or for one comparison as a dashed line for horizontal angle (upper). Note saccades, S, and smooth pursuits, SP, that show up clearly; these and other types of movements (see text) could be identified and analyzed (lower steps and labels).

FIGURE 7 EM Classification Program
Flow diagram for EM calibration, linearization, differentiation to obtain velocities, and then analysis into various categories of EM types. The program could also resolve conflicts among specific EM identification algorithms (see text).

2.3. Analyses of the Vectors of Looking

Following the analysis of the EMs, the series of fixations is defined as a string. These strings represent actions of a finite-state automata (Figure 5, bottom row). The probabilistic transitions of the finite state automata-model is the basis for further analysis, including the resultant metrics comparing these strings --- Sp, Ss, Y-matrices and parsing diagrams (Figure 8). (Additional explanation of the string-editing algorithm is provided in Appendix A and in Figure 26.)

We compare each pair of strings, to see how many letters they have in common; this matches the locations of the fixations and gives an Sp similarity index for the similarity of loci. We further compare each pair of strings as to the order of the string letters; this provides Ss, the similarity index for sequence strings.(Choi, Mosley, and Stark 1995; Hacisalihzade, Stark, and Allen 1992; Kruskal 1983; Privitera, Krishnan, and Stark 1999; Privitera and Stark 1998; Stark and Ellis 1981; Stark and Choi 1996; Wagner and Fischer)

Simplified scanpaths are compared (Figure 8, upper row). Two different scanpaths (left) with no locational similarity, Sp = 0, and no sequential similarity, Ss = 0, may be compared to two scanpaths with exact similarity, Sp = 1 and Ss = 1 (right), and to a pair of scanpaths with exact locational similarity, Sp = 1, but with no sequential similarity, Ss = 0 (middle). Pairwise comparisons of all scanpaths were assembled in Y-matrices. A Y-matrix is an ordered array of

the string similarities, either Sp or Ss, to enable further sorting and averaging. Two simplified Y-matrices (middle row) show such arrays and their row and column labels (Subject 1, S1, and Subject 2, S2, are two of the five subjects; Picture 1, P1, and Pict 2, P2, are two different scenarios). Finally, sorted and averaged values are collected into parsing diagrams (lower row). The Parsing Diagram enables comparisons among these averaged similarity coefficients: R, for repetitive scanpaths, same subject looking at the same picture at different times; Local = L, different subjects, same picture; Idiosyncratic = I, same subject, different pictures; Global = G, different subjects, different pictures; Random = Ra, random strings compared.

FIGURE 8 Simplified, or Toy, Diagram Illustrating Metrics for Comparing Scanpaths; and Parsing Diagram for the Dynamic Scanpath Experiment
Quantitative methodology diagrammed to show similarity indices, Sp and Ss (upper panel). These pairwise comparisons are organized into Y-matrices (middle panel) and then indices segregated, averaged, and placed into parsing diagrams (lower panel). Note statistical tests indicated by bolding, as well as arrows (see text). These two parsing diagrams summarize the experimental base from a dynamic scanpath study (see text).

2.4. Analyses of Dynamic Scanpath Experiment

To illustrate further how we use our analytic methods, we describe the Ss parsing diagrams for the dynamic scanpath experiment (Figure 8, bottom panel).(Stark et al. 1999) Numbers in parentheses are standard deviations; bolded values represent significant differences (at a p value <0.01) from the Ra, random values of 0.16 ($p < 0.01$). ANOVA analysis provided tests of significance, and arrows represent significant differences with respect to the G, global value; this was considered a 'bottom anchor.'

An important distinction is that between Repetitive similarity, R, (Figure 8, upper left box) and Random similarity, Ra. When using different dynamic scenarios with the same general background, the same subject with the same stimulus showed a repetition value, R, of 0.45. This indicated that 45% of the sequences were congruent, and should be compared with the randomly expected Ss, Ra value of 0.16. However, when scanpaths for the same subject looking at different scenarios were compared (Figure 8, bottom row, left parsing diagram), yielding the Ss-I value, the sequences were only 21% similar. This quantitative comparison documents that the scanpath theory generating the sequential EMs developed quite different sequences for different scenarios, that is, for different patterns of motion of the automobiles in the graphical scene.(Blackmon et al. 1999; Stelmach, Tam, and Hearty 1992) We could conclude that different scenarios were viewed by different scanpath sequences.

A different result was obtained when different viewpoint motions were compared (Figure 8, bottom row, right parsing diagram), for the same scenario. The three viewpoint motions were panning, zooming, and static. Panning, or horizontal scanning, and zooming, or near/far approach of a camera, are standard movie filmic maneuvers; static means the camera point-of-view is at rest. The scenario remained the same, and the idiosyncratic similarity index, I, was 0.38. We could conclude that viewpoint motion did not make the scanpath sequence different for different motions. Statistical analysis, ANOVAs, supported these conclusions. For both sets of experiments, the similarities for different subjects, L, are almost the same as the R values. This may be due to the fact that the sequential motions of the different cars capture the attention of different subjects in a similar way, perhaps in a bottom-up fashion, or in cortical area MT. (Blackmon et al. 1999; Born and Tootell 1992; Culham et al. 1998; Flagg 1978; Tootell et al. 1995a; Tootell et al. 1995b)

2.5. Protocols and Subjects

As indicated above, we could use a second method of 'read-out' -- 'choice,' CH, clicking on a mouse cursor position, instead of measuring EMs. Of course, we studied the similarities and differences between the usual scanpath experiments, classical read-out method 1, allowing subjects to freely gaze at the picture stimuli (e.g.,Figure 9), and the new second method of read-out, CH, by asking subjects to move a cursor over the stimuli pictures and click deliberately on ROIs. (See Results, Section 3.1.)

Another experimental task was to indicate the remembered patterns in the visual imagery experiment. We presented subjects with grids containing patterns of alphabetic symbols, letters, and asked them to image the pattern. The subjects then moved a cursor over blank grids and click deliberately on imagined or remembered loci. In this way, the 'CH' method provided for an objective 'read-out' of the structurally and sequentially bound memory traces. Further, we compared similarities and differences between this "cursor-CH" method and a third method of readout, that is, utilizing a locomotory pattern. In this third method, 'walking,' WK, we asked subjects starting from a fixed initial position to walk over a blank grid marked on the floor and stop sequentially on those grid squares that represented remembered loci of the alphabetical letters.

Subjects were students visiting in our laboratory who participated without pay; according to the rules of the Berkeley Committee for the Protection of Human Subjects they could terminate the experiment at will if they experienced any discomfort. They received oral and written and also 'operational' instruction, viewed a few preliminary pictures or grids, and usually were able to complete an experiment in less than twenty minutes. Operational instructions enforced a pattern of behavior by requiring the subjects to carry out procedures that serve as additional re-enforcement.

3. RESULTS: STRUCTURAL AND SEQUENTIAL BINDING

Our experimental results on binding, explained in detail below, compare the memory similarities between different read-out modes. Two different protocols compare EMs vs choice, CH, while looking at a set of pictures (Figures 9 and 10), and choice, CH, vs walking, WK, while remembering a set of grid patterns (Figures 11 and 12). Different 'read-out' motor behaviors, indicating remembered patterns, were analyzed in the same way, and with the same methodology. Pairwise comparisons between scanpaths were carried out with each read-out mode (see Figures 10 and 12, left and middle parsing diagrams), and then, between all pairs of one-mode-against-the other mode (see Figures 10 and 12, right parsing diagrams).

By studying the within-mode similarities against the across-mode similarities we can assign quantitative numbers to the relative strengths of inherent and of read-out binding. In addition, we enriched the experimental protocols by examining the phenomenon of "consolidation"; by this term is meant the memory coherence within repeated response patterns that may be stronger than the memory persistence from stimulus to response. (For additional results, see Section 4, Results: Symbolic Binding, and also Appendix B, Consolidation).

A moderate stressor to reduce accuracy of memory was the use of interruptions, often with other sets of experimental grids. Another stressor we used was to tilt the grid and require the subject to adjust his display of the remembered pattern to the tilted blank grid. Most often, a training period, whereby the subject made horizontal and vertical lines on tilted pictures, was introduced. Experiments (not illustrated herein, see Stark 1999(Stark et al. 1999)) utilizing this "tilting"

paradigm demonstrated the robustness of the structural binding to adaptation, producing 45-degree rotations of the blank grids and the memory pattern. (We wish to thank Dr. G. M. Gauthier, CNRS, University of the Mediterranean at Marseilles-Luminy, for suggesting the tilting paradigm to us. (Gauthier et al. 1979; Gauthier et al. 1994))

3.1. Choice, CH, Compared with Ems; Picture Viewing

The alteration of the scanpath protocol, substituting mouse-cursor location and clicking for measured EM fixations, has many experimental advantages.(Stark et al. 1999) However, it had to be carefully evaluated by comparisons between EM fixations and choice, CH, loci in a variety of studies. In developing and expanding the protocol, we had subjects look at a number of pictures and then we studied their EMs (Figure 9). The methodological procedures to go from raw EMs (upper left) to identified fixations (upper right) were necessary. Linearized EMs of a subject looking at a cave painting of a horse (upper left) were transformed by a "fixation algorithm" to a sequential string of fixations (upper right, circles), with connecting vectors representing saccades and their sequence (upper right, arrows).

Next, sequential string of CHs, produced by mouse cursor and clicks (lower left, squares) are also connected by vectors (lower left, arrows). Then, these two strings can then be compared as to identity of their loci, within a distance determined by a K-means algorithm, to calculate the Sp and Ss similarity indices between EMs and CH procedures.(Privitera and Stark 1998) A K-means algorithm proceeds by calculating a parameter, such as distance, through each distance value, and then determines the optimum distance in terms of a criterion such as the highest Sp match. As an extra bonus for the reader, consider that this cave painting, and artistic work created 31,000 years ago has perhaps been equaled but not surpassed in the ensuing millenia of human social prehistory and history. The scanpath theory has awakened new interest in the neurology of artistic communication. (Elderfield 1998; Zangemeister, Sherman, and Stark 1995; Zeki and Moutoussis)

FIGURE 9 EMs Compared with Choice, CH: Selection of ROIs Compared
Linearized EMs (upper left) were analyzed into fixations and saccades (upper right) while the subject looked at a cave painting of horses. Loci chosen by mouse clicks (lower left) could then be compared (lower right) with EM fixations (see text).

As explained in the Methods section, many pairwise comparison indices are collected and sorted using the Y-matrices. Averaged results are then organized in the parsing diagram (Figure 10: Sp, upper row; Ss, lower row). EM comparisons (middle column) document that the R values, 0.62 and 0.26 are significantly different from Ra and from G, the two bottom anchors (bolding or heavy arrows indicate $p < 0.01$). Note that while Sp-L has a relatively high value, indicating that different subjects selected similar ROIs, the Ss-L value is lower suggesting that different subjects utilized different sequences for the same picture and similar loci across subjects. The Ss-Ra values throughout are much lower than the Sp-Ra values, since there are many ways to establish sequences among similar loci. Almost identical results are found for CH comparisons (left column), with somewhat higher coherences, perhaps due to the more deliberate TD selection mental process for cursor clicks vs natural EMs.

Now when we compare EMs and CH (right column) we find related distributions of similarity indices; R values are large and significantly different form G and Ra. The R-Sp index is large, indicating similarity of objects across modes; that is, it coheres for both read-out modes. Thus it appears that almost none of the structural binding is related to read-out mode differences. However, because the cross-mode R-Ss value equal to 0.17 is less than the R-Ss values for either

choosing or EMs, we must, in this case, partition the sequential binding between inherent and readout components.

FIGURE 10 Parsing Diagrams; EMs Compared with Choice, CH
Sp (upper) and Ss (lower) parsing diagrams for the choice compared with EM study. Intramodal read-out comparisons (left and middle panels) as well as intermodal read-out comparisons (right panel). Results described in text.

3.2. Choice, CH, Compared with Walking, WK; Grid Viewing

In another set of experiments, we presented grids (see Figures 5 and 11) to be memorized and to be recalled. Again two modes of response were compared: --- using a cursor moving over a blank grid presented on a computer screen, or walking, WK, over a large grid outlined over the laboratory floor (Figure 11). (Stark et al. 1999)

Squares with alphabetical symbols represent grid patterns that the subject could look at for a period of three seconds for each of two presentations (Fig. 11, two left-most columns). Subjects were then asked to move the cursor sequentially to each of the visually imaged locations of the symbols and to click the mouse buttons (this took about ten seconds) to indicate the remembered alphabetically-labeled grid squares (Fig. 11, four right-most columns); thus providing an output string of remembered alphabetically-labeled grid squares. There was a fixed initial position from which the subjects started each time. Experiments were also carried out with subjects instructed to walk freely over a large grid placed on the floor; again, there was a fixed initial position from which they started each time. They were also instructed to stand with two feet in the appropriate grid-spaces for a brief moment, to indicate each labeled locus; in this way, the experimenter could record the sequences of stops. Again, this provided an output string of remembered alphabetically-labeled grid squares. CH and Walking, WK were alternated without additional refreshment (Fig. 11, blank regions, second and fourth rows).

Also CH and WK could be presented with refreshment (Fig. 11, fifth and sixth rows). This refreshment (two leftmost grids, sixth row) allowed a modified, reinitialized, sequential pattern to be developed in the subject's representation (compare fifth and sixth rows). (For additional results, see Appendix B: Consolidation.) (See also 4. Results: Symbolic Binding; compare control experiment with refreshment (Figure 15, upper) with the top anchor experiment without refreshment (Figure 16, upper), and also the summary of Symbolic Binding results (Figure 17).)

FIGURE 11 CH compared with Walking Protocol
Experimental protocol for cross-modal comparison between choice, CH, and walk, WK, read-out modes. Note similarity of patterning when a second display of stimuli patterns was not given (absent grids in both sets of upper panels); note difference in patterning when refreshment of stimulus pattern allowed a new memory schema to be formed (lower panel).

To buttress the qualitative results as shown in Figure 11, we provide quantitative assessments from the similarity indices. The results came from many pair-wise comparisons for four subjects, naive with respect to the purpose of the experiment, but performing quite well in the task; their similarity indices were sorted using the Y-matrices and averaged in the parsing diagrams (Figure 12: Sp, upper row, and Ss, lower row). CH comparisons (Fig. 12, left column) document that the R values, 0.90 and 0.77, are significantly different from Ra and from G, the two bottom anchors. Bold values or heavy arrows indicate $p < 0.01$, that the values differed from Ra (bold) and from G (arrows). Again, the Ss-Ra values throughout are much lower than the Sp-Ra values, since there are many ways to establish sequences among similar loci. Almost identical results, 0.91 and 0.81,

are found for WK comparisons (Fig. 12, middle column) with respect to the R values, both Sp and Ss.

When we alternate CH and WK without refreshment (Fig. 12, right column), we find somewhat modified distributions of similarity indices. R values, 0.87 and 0.60 are large and significantly different from G and Ra. That Sp-R is identical to the values for CH x CH and for WK x WK indicates that structural binding relied upon an inherent component.. That Ss-R is less than for CH x CH and for WK x WK indicates that both the inherent and the readout components were important for sequential binding.

FIGURE 12 Visual imagery: Parsing Diagram for Walking, WK, vs Choice, CH
Sp (upper) and Ss (lower) parsing diagrams for the choice, CH, compared with walking, WK, study. Intramodal read-out comparisons (left and middle panels) as well as intermodal read-out comparisons (right panel). Results described in text.

3.3. Modular Cortical Organization: The New Phrenology

A sketch of the lateral view of the human cortex (Figure 13, upper) is presented to help understand the logic of these different readout experiments. We are trying here to distinguish between inherent sequential binding, likely located in the prefrontal cortex, from variable sequential binding, dependent upon readout mode. The modes we are exploring are EM fixations, as in the classical scanpath experiments, choice, CH, using mouse-cursor positioning and clicking, and locomotion over a grid on the laboratory floor.

Information about localization in the cortex comes from a variety of sources. Classical neuroanatomy and analyses of neurological syndrome have existed for several centuries, and had achieved a considerable degree of sophistication. Experimental ablation and electrical stimulation physiological studies next came to play. Modern methods, ranging from intrusive single-unit neurophysiology, to current PET and fMRI are daily supplementing earlier studies (see Appendix M).(Zeki and Bartels) We have collected in Appendix N a few significant references to the neurophysiology in higher-level functions, that are pertinent to new concepts of modular cortical organization. Note (Figure 13, upper) the 'what' ventral pathway from visual cortex, VC to the temporal cortex, TC, (especially left side) to which we attribute Semantic Binding. Similarly, note the 'where' dorsal pathway from VC to the parietal cortex, PC, (especially on the right side) to which we attribute Structural Binding. Known connections from PC to the pre-frontal cortex, PFC, have been shown to be related to temporal sequencing, and to which we attribute Inherent Sequential Binding. Connections continue to the frontal eye fields, FEF, to which we attribute one form of Read-Out Sequential Binding. (Although a complete review of this fascinating area is beyond the scope of the present paper, a number of articles are referred to in the Discussion and Appendix sections below.)

FIGURE 13 Modular Cortex and Connectivity
Recent studies in neurophysiology and fMRI have established a "new phrenology," the modular cortex (upper), with different functions assigned to specific regions of the cortex (see text for further explanation). Connectivity explored in our experiments on inherent and read-out sequential binding, and as well, on the influence of symbolic binding, is indicated as numbered arrows joining labeled regions (lower). (See text for further explanation.)

3.4. Inherent vs Read-out Sequential Binding

Now, we may consider the connectivity of the modules of the modular cortex (Figure 13, lower), as an aid to define the logic of the results of these experiments.

A summary of coefficients comparing within-mode and across-mode experiments (Sp, left columns, and Ss, right columns) provided the basic similarity coefficients from the parsing diagrams (Figures 10 and 12). These were then normalized (Diff & N % columns) as percentages, by setting the bottom Ra anchor to 0% (e.g., 0.27, Figure 14, upper left panel), and the within-mode R values as 100% (e.g., 0.67, Figure 14, upper left panel). The two experimental protocols, CH x EMs (upper panels) and CH x Walking, WK (lower panels), have yielded reasonably consistent results.

For Sp, we subtract the R values for across-mode from the within-mode values; the resulting normalized percentage numbers are 85% and 94% for the inherent component of structural binding. Since for structural binding, the inherent component dominates, this may be interpreted as putting the parietal lobe structural memory as too early in the process to be disturbed or altered by read-out mode differences.

For Ss, we again subtract the R values for across-mode from the within-mode values; the resulting normalized percentage numbers are 50% and 74% for the inherent component of structural binding. We see that for sequential binding, although the inherent component is larger (2/3rds), the readout component is significant (1/3rd) and thus both components are important in sequential binding. This may be interpreted as allowing the pre-frontal lobe inherent sequential memory to be somewhat altered by readout mode located further back in the frontal lobe; these regions are, of course, different for EM, for hand movement and for locomotion.

To summarize, structural binding is inherent; that is, it is the same independent of readout modes. Sequential binding has strong components for both inherent and for readout binding; that is, the readout mode contributes strongly (about one-third) to the memory of the sequence.

FIGURE 14 Inherent vs Readout Sequential Binding
Sequential read-out experimental findings can be summarized as almost 100% inherent binding, for spatial or structural similarity of patterns (middle column). However, sequential bindings are markedly influenced by read-out mode; only two-thirds of the binding is inherent (right column).

4. RESULTS: SYMBOLIC BINDING

How Can We Experiment on Symbolic Binding? The naming of a pattern, or its symbolic binding, plays an important role in this scanpath memory process. Quantitative experiments were carried out by Yang and Stark (Yang and Stark 2000) to explore this phenomenon. Subjects were asked to remember lettered grids under a variety of conditions. Often, they were presented only with the letter, or symbol, of the pattern, and asked to remember the grid pattern that they previously were able to reconstruct. Interruptions, such as becoming familiar with and reconstructing other grid patterns, were most often interjected between the learning phase and the test-of-memory phase.

Control Experiment: Same Pattern, Same Label with Refreshment. An important control experiment was to test the ability of subjects to carry out pattern reconstruction by memory (Figure 15, upper panel, upper row). Here, subjects were presented for a few seconds with two identical lettered grids, immediately followed by four successive blank grids wherein the subject

attempted to reconstruct the previously seen patterns. Subjects were then interrupted with other tasks. Next, the identical lettered grid was presented (Figure 15, upper panel, lower row), and the subjects attempted again to reconstruct the labeled pattern identified only by label, onto four successive blank grids. Subjects were able to carry out this task very well, with Ss values of 0.71 (s.d. = 0.32), and Sp values of 0.80 (s.d. = 0.24).

FIGURE 15 Control Experiment and Main Experiment
Experimental protocol for the control (upper panel) and main experiment (lower panel) to analyze the influence of symbolic binding. A major result is the influence of dissimilarity of labeling on the dissimilarity of the sequential pattern. Clearly, the spatial loci are the same, and thus, the structural similarity remains high. (Note that refreshment in the form of two additional looking stimuli are presented in both experiments (lower pair of grids in each of the panels).)

Main Experiment: Same Pattern, Different Label with Refreshment. The main experiment again tested whether subjects attempted to re-remember a newly presented pattern identical to an old remembered pattern, with an important, significant difference (Figure 15, lower panel). The new, identical pattern was labeled with a different symbol or letter!

In both of these sequences, the subject was able to perform consistently over the reconstructions in the four blank grids (Figure 15, lower panel, upper row and lower row, four right grids). However, the new symbol encouraged the subject to reinitialize the memory pattern. Thus, the Ss value fell to 0.46 (s.d. = 0.35), when the first and second presentations were compared. (Note differences in sequential patterning (Figure 15, lower panel, upper row, compared to lower row). Of course, the localization of the clicks, Sp =0.76 (s.d. = 0.32) was equally accurate to the control experiment described above. Only the sequence was newly established because of the new label.

Top and Bottom Anchor Experiments. The range of values for the Sp and Ss similarity indices could be established in two more experiments, the top anchor and the bottom anchor (Figure 16).

An experiment quite similar to the control experiment was next performed. Its main difference was that it allowed for no re-presentation or refreshment of the lettered pattern for the second set of the memory test blank grids (Figure 16, upper panel; note the absence of the second presentation of the lettered grids in the lower row, upper panel). Thus, the memory trace remained more or less the same without additional information relating to the pattern or the letter being presented. This gave us the highest values, Ss = 0.86 (s.d. = 0.06) and Sp = 0.94 (s.d. = 0.20). We thus consider this to be the top anchor of the similarity scales.

For the bottom anchor, we used the same letter symbol, but in a completely different grid pattern (Figure 16, lower panel). As might be expected, the two sets of memory tests, with four blank grids each, showed little inter-trial coherence or similarity of their patterns, with Ss = 0.07 (s.d. = 0.04), and Sp = 0.36 (s.d. = 0.10). The structural inter-trial coherence and the sequential coherence are very low, and close to random values; thus, we can use this experiment as a bottom anchor. Although the subjects were "tricked" by having the same label for different patterns, still the structural pattern dominated over the symbolic label.

FIGURE 16: Top Anchor and Bottom Anchor Experiments
Experimental protocol to establish the range of values for Sp and Ss similarity. Top anchor (upper panel) shows high correlation when no refreshment is permitted (two absent grids, lower row, upper panel). Bottom anchor (lower panel) shows absence of structural and sequential similarity when a different pattern is presented with the same label.

Summary of Symbolic Experiment. To summarize the symbolic experiment (Figure 17), we have been able to establish that in the control experiment, the re-presentation of the same stimulus with the same label a second time allows some reinitialization of the memory trace. However, the values remain quite close to the top anchor values, where the absence of a second presentation did not allow for even mild changes of the memory pattern. Of course, with time and interruption, as was carried out in our experimental protocol, some decay of the memory pattern occurs.

The main experimental result is that when the same pattern grid is presented a second time, but with a new label, this new label encourages reintialization of the memory trace. Thus, the second set of responses are quite different (about 50% loss of coherence for the sequence, Ss, but essentially no loss of coherence for the pattern, Sp); recall that the same loci were re-presented with the new label, and thus, Sp should remain quite consistent. The quantitative result, 50% loss of coherence due to a changed label, comes from averages of many experiments done with a variety of subjects. Subjects varied, and even the same subject would produce much higher or lower coherence in different trials. More experiments are necessary to establish if a quasi-switching occurs between coherent and non-coherent results.

This experiment documents the crucial role symbolic labeling plays in memories of spatial patterns.(Tanenhaus et al. 1995; Tempini et al. 1998; Thompson-Schill et al. 1997) It also raises questions and points out suggestive interpretations for connectivity between operations in different parts of the modular cortex (Figure 13, lower panel).

FIGURE 17 Summary of Symbolic Experiment
Symbolic binding experimental findings can be summarized. Symbolic memory has important influence on sequential binding, producing an average 50% loss of coherence (compare 0.46 with 0.71, next to bottom row) when the labeling is changed. Since the same loci were re-presented with a different label, the structural binding, of course, remained the same (compare 0.76 with 0.80, bottom row).

5. PERCEPTION AND CORTICAL REPRESENTATION

5.1. Perception and Sensory Organization

Philosophers have long speculated that we see in our "mind's eye," but until the scanpath theory, little scientific evidence was available to support these conjectures. On the other hand, philosophers from Plato onwards have thought deeply about these matters, and we have tried to summarize their views. Using four terms defined by the philosopher Kant, a five-component visual perceptual schema has been developed to incorporate the relevant concepts of experimental metaphysics. (The senior author is appreciative of early discussions with Professor W.H. Zangemeister, (Stark et al. 1986) that led to an early version of Figure 18.) (Kant 1949; Russell 1945; Stark and Choi 1996) We start (Fig. 18, column one) with the world of **appearances**, the "chaos" of early Greek philosophers; in our terminology it is called "BU stuff." At one time, we used "things" for the so-called 'real' outside world, but an anonymous discussant pointed out that by the time the brain had done figure-ground separation to identify an object as distinguished from background, and applied knowledge about physical coherence of the object, much of the perception of the object had already been accomplished! The next stage (Fig. 18, column two), **sensation**, represents the inflows of energy onto body sense endings. It now appears that the filtering expected by Muller for "specific nerve energies" is actually accomplished by "specific nerve endings," and specific nuclei on which they project.

FIGURE 18 Philosophical Approach to Perception
Five stages of the perceptual process (five columns) are illustrated with icons (upper), also showing BU and TD processes (curved arrows). See discussion in text regarding philosophical and physiological sources of this schema.

We call the next stage (Fig. 18, third column), **sensory organization**, BU physiology, wherein the Kantian internal constructs of space and time are added. The frog's eye, using 'bug' detectors, can calculate the velocity of a small moving spot accurately enough to keep frogs very well in bugs. (Ingle 1971; Lettvin et al. 1959) While it took 350 million years of vertebrate evolution to arrive at the frog's eye, yet another 350 million years was necessary to arrive at the brains of McCulloch and Lettvin, so that they were capable of demonstrating the elegance of this aspect of sensory organization of the frog's eye. Since velocity requires both space and time computation, it is clear that these Kantian internal processes have been captured by evolution. (Itti and Koch 1999; Niebur and Koch 1998)

If we jump ahead (Fig. 18, rightmost column five) to **representation**, the 'ideals' of Plato and the 'notions' of Berkeley, we see that our term, 'TD cognitive models', may perhaps be symbolized with a file drawer icon. We will return to the question of representation in the third subsection below. Such models, acting TD onto the critical stage (**perception per se**, Fig. 18, fourth column) can be seen to be planned, forceful, determined sets of activities. (Pribram 1971; Searle 1983) In our model for perception, the TD active looking scanpath plays its role as the operational phase of perception per se. The set of five columns (Fig. 18), dissecting the overall perceptual process, leads to an important question we can pose for the neurophysiologist, "Where does TD meet BU?" Our conjecture is --- where TD iconic inputs to levels I, II, and III of the visual cortex meet BU iconic visual signal information going to levels IV and V in the retinotopic visual cortex (Figure 19). This is the site of the "iconic matching" process.

5.2. Visual Cortex: Where Does Top-Down, TD, Meet Bottom-Up, BU?

In their famous paper, Pitts and McCulloch conjectured that the inflow information from eye and lateral geniculate would reach the striate cortex (Figure 19). (McCulloch 1965; Pitts and McCulloch 1947) This was a BU theory as was the later frog's eye paper. (Lettvin et al. 1959)

We have now modified this BU approach to add TD perception. The visual cortex has a retino-topical organization that is apt for matching a TD iconic sub-feature representation with incoming BU sensory signal flows. Likely some interactive feedback process could match these two maps, one TD, the other BU, to some criterion of fit (see Microscopic Cortical Processes, below, and Appendix N). This, then, permits the scanpath, if confirmed to this point, to continue to the next ROI or sub-feature of the representation. In this way, the TD model moves, fixates, and foveates the eye, to bring forward successive sub-features for checking. An absence of fitting forces a new model and a revised scanpath. In this way, the scanpath as an operational mechanism plays an active role in the overall perceptual process. (Henderson and Hollingsworth 1999)

FIGURE 19 Micro-Cortical Processes
Six levels or layers of the visual cortex, known from neuroanatomy, are suggested as the iconic matching region, where TD input to the visual cortex, at layers I, II, and III, interact with BU input going to layers IV and V, from retina via geniculate and optic tracts (modified from Pitts and McCulloch, 1947).

5.3. Representation

Our results are interpretable in terms of a set of models or schemata (Figure 20). These models suggest visual patterns of thinking about --- i) procedures for visual perception and recognition, ii) the macroscopic, and iii) microscopic neuro-anatomical underpinnings of these memory processes as they are interpreted according to current neurological knowledge, and iv) quantitative and normalized values for relative strengths to the several components of memory binding (Figure 20). The connectivity of the cortex is vast; studies by Valentino Braitenberg and other neuro-anatomists, from Golgi and Cajal on, have illuminated many aspects of this constrained mesh and link the microscopic and macroscopic views. (See Appendix C, The Braitenberg Cortex.) Since the classical studies of Hubel and Wiesel, a number of approaches have developed have developed to further understand the neuroanatomical and neurophysiological substrates of cortical connectivity. (See Appendix N.)

FIGURE 20 Cortical Representation of Perceptual Processes
Although only the microanatomy of the visual cortex is known well enough to support a graph theoretical model, yet we have suggested a variety of such graphs for structural, sequential, and symbolic binding, with loci as per labels in the modular cortex. Geometrical binding is used in our modeling schema, for syntactical interaction between foveal ROIs and peripheral segments. Different forms of the graphs do not represent any knowledge about feasible or understood properties of the brain, but rather stress our ignorance.

Macroscopic Cortical Processes: Where Does Memory Dwell? Recently, especially with the advent of functional magnetic resonance, fMRI, and its associated imaging technology, there has been an increase in localization studies on awake cooperating human that has led to a new 'phrenology' --- this time hopefully based upon more scientific evidence. (Palmer 1975a; Palmer 1992) It is beyond the scope of the present paper to provide a full review; see, however, Appendix M. (Colby, Duhamel, and Goldberg ; Meystel et al. 1992; Palmer 1975a; Palmer 1975b; Palmer 1992; Palmer 1999; Palmer and Kimchi 1986; Palmer, Neff, and Beck 1997; Rybak, Golovan, and Gusakova 1993; Umeno and Goldberg 1997)

Of particular interest to our own studies are pathways connecting the visual cortex to other cortices. The ventral pathway from visual cortex, VC to the temporal cortex, TC (especially left-side), is the 'what' pathway to which, we attribute Semantic Binding; in similar fashion, the dorsal pathway from VC to the parietal cortex, PC (especially on the right side), the 'where' pathway to which we attribute Structural Binding. Spatial vision and memory and their uses in animal and human behavior are crucial functions that have been widely studied. (Klatzky 1998; Klatzky et al. 1990) There are strong known connections (Pribram's Law) from PC to the pre-frontal cortex, PFC, that are related to temporal sequencing, and to which we attribute Inherent Sequential Binding. Then connections continue to the frontal eye fields, FEF, to which we attribute Read-Out Sequential Binding. Of course, there are different motor areas for different behaviors used to indicate imaged loci and sequences in our experiments and in normal behavior, more generally.(Wolfe, Alvarez, and Horowitz in preparation, 2000) Indeed, our experiments were designed to test some of these physiological-anatomical conjectures and to see if there were differences between inherent and read-out sequential bindings that could be captured in an experiment. In 1970, after a lecture by Professor Bela Julesz on his famous random-dot stereograms, I asked if he thought psycho-anatomical procedures obeyed a transitivity rule. The question illustrated the possible complexity of cortical connectivity; our diagrams are only a simplistic beginning.

We also use as metaphor, a robotic computer vision study (see Section 6 below) that has a complete TD model of the robot working environment, the robot kinematics and dynamics, the pose of the robots and the monitoring cameras (see Section 6, below). Here the model directs and limits the scanning of the video images to known positions of the ROIs in the 2D camera projections of the 3D-operating world. The model may be displayed on a computer screen for the supervisory controller to observe. Now, we ask a hypothetical question, "Where in the computer is the model located?" The answer makes us realize that the model is a collection of non-iconic programs and parameters, widely distributed in active memory, in rotating memories, in registers, caches and pointers of the running programs and most often cannot be definitively located. This metaphor tempers our attempts to fix memory loci in the brain.

Microscopic cortical processes. We know little about microscopic cortical processes.(Desimone 1992; Sillito and Grieve 1991) As Hubel and Wiesel have pointed out, their classical work served to locate processes, rather than to establish how these processes occur. Similarly, the new phrenology substantiating the modular cortex, and as well, f studies, serve to fix anatomical locations. We therefore have used a variety of graphs to express our ignorance of function (Figure 20, multiple graphs for different functions). (Freksa 1992; Freksa 1997; Schill et al. 1999; Stark 1993; Stark 1994; Zangemeister, Stiehl, and Freksa 1996) We do not at all suggest that the differences in these graphic displays represent known functional differences for the macroscopic modules. Pioneering and future studies of microscopic cortical function are and will be an exciting subject (Appendix N). What we emphasize is that the different memory functions, or bindings, in different parts of the modular cortex, must be carried out by some cellular networks, as postulated by McCulloch and Pitts. The cellular anatomical diagram for the visual cortex alone serves to give body to the above discussion.(Sillito and Grieve 1991; Thompson-Schill et al. 1997; Zangemeister, Stiehl, and Freksa 1996)

6. TELEROBOTIC SCHEME: TD SCANPATH MODE FOR COMPUTER VISION

6-1. Telerobotic Control System

Because the scanpath theory rests upon continuing studies of the human brain, we necessarily lack a complete operational model. There are some neural models with BU approaches (Rybak, Privitera), and as well, a general appreciation by the computer vision world of an important future role for "image understanding." (Aloimonos and Herve 1992; Bolle, Aloimonos, and Fermuller 1998; Carpenter, Grossberg, and Lesher 1998; Carson et al. 1997; Crevier and Lepage 1997; Foresti and Pieroni 1998) For some years now, we have developed a vigorous, explicit and functioning model of the TD scanpath scheme to aid our researches into robotic vision. (Blackmon and Stark 1996; Buttolo, Kung, and Hannaford 1995; Ho and Stark 1997; Ho and Stark 1999b; Ho and Stark 1999c; Ho and Stark 2000; Kim et al. 1987; Kim, Takeda, and Stark 1988; Kim, Tendick, and Stark 1987; Liu et al. 1993; Nguyen and Stark 1993; Stark et al. 1988; Sutro and Lerman 1973; Yu and Stark 1995) We now explain this model in some detail.

Quasi-autonomous robotic systems are designed with human supervisory control (Bejczy 1980; Ferrell and Sheridan 1967; Moray et al. 1989; Sheridan 1992; Yoerger and Slotline 1987) restricted to planning and emergency actions (Figure 21). The human operator, H.O., uses the supervisory control interface and path planner, to generate actual sequences of movements for a specified task. Each movement is then segmented in serial order, and each segment sent to the low-level feedback controller as input to the robot plant, whose output is Ya. The robot pose control signal, U, is indicated by a skeletal robot model with circles for the VEs, attached to critical kinematic points of the robot (Fig. 21, upper right inset). Under certain modes of operation, U is used to control the image processing algorithmic procedure, IP Alg. Under other

modes of operation, the previous located VEs, Ym, are used to predict the locations of the VEs in the next image.

Redundant feedback, Ym, is provided by both GPS sensors (not shown) and by image processing algorithms operating on the camera capture of the actual position of the robot. This feedback, Ym, is provided for closed loop operation, as long as it does not widely violate certain constraints; the measured position, Ym, of the robot is indicated by the skeletal model with Xs marking the measured locations of the VEs (Fig. 21, lower left inset). Additionally, this measured output, Ym, acting outside the main control loop, may modify the current and next segments in the serial output of the higher level controller, and thus, updates the world model for system consistency. Plant and image noise has been added in simulated runs in related studies (Ho et al. 1999; Ho and Stark 1997; Ho and Stark 1999a; Ho and Stark 1999b; Ho and Stark 1999c; Ho and Stark 2000) that have provided estimations of the amount of redundancy necessary to attain robustness during actual operating conditions.

It is important to note that using the scanpath schema, image processing is controlled in a TD fashion by the feed-forward model. The model knows the kinematics, dynamics, and pose of the robot, and its commanded positions at each iteration. It also has a model of the robotic working environment, RWE. (Zelnio 1991)

FIGURE 21 Feedback Model for Supervisory Telerobotic Control
Control systems diagram for telerobotic scheme showing higher level control, with supervisor and path planner. The serializer provides input to the basic feedback control loop, with camera and image processing algorithms, IP Alg, monitoring actual position, Ym, of robot (right inset).

6-2. TD Robotic Vision

The complete model of the robots (Figure 22), consists of compacting vehicles carrying out civil engineering dam building.(Ho and Stark 1999a) A set of VEs made up of prominent lights is easily detected by distant cameras. For the image processing aspect of the scheme, it is important to note that the model of the robot includes the knowledge of the placement of these luminaires (Figure 22, upper left). In addition, the known camera loci, directions, and optical parameters enable prediction of the 2D projection of the scene onto any particular camera image plane.(Ho and Stark 1997; Ho and Stark 1999b; Ho and Stark 1999c; Ho and Stark 2000; Miyata and Stark 1992) The display mode (Figure 22) indicates the model expectation of each luminaire location by showing white boxes outlining ROIs (upper right); expected locations may not be the actual locations and thus may require feedback correction.

As with TD scanpath control, the robot model predicts where the vehicle will move and this provides anticipatory information for locating the ROIs. According to the scanpath model, the image processing algorithms will move in sequence from ROI to ROI in the camera plane to carry out the various image-processing procedures (heavy white arrows, Figure 22, lower left). Thus, the alternation that occurs in the case of computer vision, is similar to the alternation between TD motor-control of EMs for successive foveations and the matching of the BU visual signals to the TD iconic model in the visual cortex in the case of human vision.

After image-processing steps, the display mode indicates the locus of the actual measured VE with a cross (lower right). The vector of actual locations is then passed to the feedback control mode, as discussed above. (Ho and Stark 1997; Ho and Stark 1999a; Nguyen and Stark 1993; Stark et al. 1988; Yu and Stark 1995)

FIGURE 22 TD Scanpath Scheme for Robotic Vision
Four image processing steps showing robot vehicles with VEs (upper left) and model ROI-predicted locations (white squares, upper right). Note scanpath sequence for computer image processing of ROIs (white arrows, lower left) yielding centroid-calculated loci (white crosses, lower right).

6-3. TD Model-Based Image Processing

The cornerstone of the algorithms is the model that represents the TD information the system has about the robotic working environment, RWE.(Ho and Stark 2000) This model consists first of the robot kinematics; in addition, the dynamic component describes how model configuration changes over time. The remote camera component describes the pose and geometry of each camera; finally, the locations and sizes of objects that may interact with the robots are also stroed in the RWE. As discussed above, the VEs that aid in image processing; their geometrical representation is also reflected in the model. Now, given this environment, the visual algorithms perform four main steps in sequence.

TD pre-filtering. The TD model predicts the expected incoming signals, that is the 3D locations of the VEs using the last known pose of the robots, the kinematic model of the robots and the control signal history. Using known camera pose and geometry, the 3D locus prediction is then projected onto a camera frame of reference. An ROI with the resulting location is then assigned. The estimated apparent size of the feature is calculated in a similar manner, and alters the ROI size for that feature. Implicitly, the IPA output of the ROI is logically bound to the feature at this stage. (Bajcsy and Krotkov 1993)

The viewability or detectability of the predicted features is thus aided by the known estimated locations of all objects in the RWE. For example, occlusion or possible overlapping of the VEs by the robot itself or by other known objects in the environment can be predicted; a sampling dependability factor can then be generated. Sensitivity of robot pose to the 2D loci of features viewed from a given camera may be calculated; the jacobian matrix provides static and dynamic weightings. Both sensitivity and dependability factors are used to determine the significance of the sampling in a particular ROI. The dependability factor is of particular importance if the sampling cost of each ROI is high, say the sampling speed is slow with respect to a limited time window for IP and thus a sampling priority has to be assigned to each ROI. The sampling sequence of the ROI, similar to the scanpath sequence, is then generated based upon these factors.

BU Image Processing. The TD model applies an appropriate BU image processing algorithm, IPA, suitable for the ROI and its feature of interest. In the case illustrated, features are spherical light sources and the IPAs utilized are adaptive thresholding followed by a centroid calculation. Video camera images, even under the best condition, are often very noisy (Figure 23, upper). Indeed our design of the luminaires was an engineering attempt to provide adequate signal/noise ratios. By restricting the image processing only to the small ROI area, the amount of noise impinging upon our signal processing is greatly reduced; thus, the adaptive thresholding techniques yield robust results.(Uttal, Baruch, and Allen 1995) The ROIs, are indicated as rectangular vertical boxes (Figure 23, lower), the top border of each is the actual adaptive threshold, utilized in each local area. The VEs can be clearly distinguished as narrow hilltops (Figure 23) above these adaptive thresholds, and contribute to the robustness of the BU image processing procedures controlled in this TD fashion. Indeed, the fovea of the retina and its magnified cortical representation (Figure 1) must also possess local adaptive advantages of a similar sort.

FIGURE 23 Advantages of TD Control of BU Image Processing
Pixel intensity diagram forming a 3D representation of the video image (upper). Note hilltops representing VEs. By predicting ROI loci using TD model (rectangular boxes, lower), it is possible to do adaptive thresholding only in a small localized region, and thus, achieve important signal-to-noise ratio improvements. Clearly, foveal fixation in normal human vision achieves the same functionality.

TD plus BU post-filtering. The next procedure verifies the integrity of the individual centroid measurements in the contest of the overall TD model (Figure 21, lower left inset). For a type of feature, in this instance a spherical VE, the system generates a criterion, such as moment invariance, to test the for possibility of error due to unexpected effects. Thus, significant ellipticity of features in an ROI would be marked as unreliable, and thus, weighted less in the data integration part of the program.

TD Data Integration. The RWE model is next synchronized with the feedback measurements, Ym, so as to produce a consistent updated model. Each ROI locus creates two constraints in the estimate of robotic pose. These, the dependability factor, TD model-pose and occlusion information, and the reliability factor, judged by the fit of the image-prediction to the BU processed image signal, are fed into an optimization routine that finds the optimal robot pose such that constraint violations are minimized. If fitting error is high, indicating a failure in the IPA procedures, re-initialization of the IPA subsystem is performed. (Yu and Stark 1995) Otherwise, the resultant estimated robot pose accepted, updates the model and the next IPA iteration is performed. These stages of the computer vision scheme have not only a sequential structure but also multiple interactions (Figure 24).

FIGURE 24 Block Diagram Explicating Telerobotic Vision Scheme
Flow diagram schema to aid in understanding steps of our telerobotic TD scanpath approach to image processing and to supervisory control.

7. SUMMARY

TD vision. This paper has considered the TD aspects of human vision to be equally (or more) important to vision as a whole than are the usual text book presentations of BU vision with constellations of psychophysical and neurophysiological experimental paradigms. Recall that classical experimental designs themselves prejudice the vision scientist to think in unidirectional input-output terms. We began with the essential role that EMs play, and their function in the TD control of the flow of visual information. The scanpath theory proposed that an internal spatial-cognitive model controls perception and the active looking EMs, of the scanpath sequence. Evidence supporting the scanpath theory (Noton and Stark 1971a; Noton and Stark 1971b; Noton and Stark 1971c) includes experiments with ambiguous figure and visual imagery. (Brandt and Stark 1997; Stark and Ellis 1981) Also application to dynamic scenes, although only beginning, yet has many lessons for further visual studies.(Blackmon et al. 1999) We also have provided an introduction to the experimental procedures including careful calibration of EMs, definition of ROIs, and the analysis and comparison programs for studying scanpaths.

The scanpath research and the recent studies on memory binding(Stark et al. 1999) described in this paper help to understand the dual role played by TD and BU visual processes (Figure 25). The TD representation in the mind's eye, and as elaborated in this paper, throughout the brain, controls not only EMs, but the placement of spatially defined iconic models in the visual cortex. Here they are matched with BU signal information, arriving from the so-called "real" world. These BU signals, are known to have a distorted log-polar iconic form (Figures 1 and 25), up to

and including the primary visual cortex. First note that the picture (Figure 25, top) in the 'mind's eye' is close to the picture (bottom) in the "real" world, or our species, heavily dependent upon vision, would have disappeared. Successive EM fixations produce retinal images (three are shown, just above "real" world picture); from this ensues cortical magnification of foveal regions and cortical minification of peripheral low resolution segments. At the same time, the TD representation sends similar iconic representations (three are also shown here, just below the mind's eye picture) to the visual cortex for matching; TD scanpath EMs predict the spatial loci for these matches. At the visual cortex (Figure 25, middle), iconic matching of TD and BU occurs. (Ballard, Hayhoe, and Pelz 1994; Driels and Acosta 1992; Gould 1967; Groner, Walder, and Groen 1984)

FIGURE 25 Iconic Matching in Visual Cortex TD Representation and BU Signals
BU retinal image is shifted with each EM fixation to provide a centered and magnified foveal projection in the visual cortex. These may be matched by predicted TD iconic representations from the mind's eye image. Continuous periphery is shown broken into segments, also suitable for TD symbolic coding.

Representation. More recently, we have used the repetitive scanpath EM sequence to approach problems of the representation of the visual image in the brain. We suppose that there are several levels of "binding" --- semantic or symbolic binding, structural binding for the spatial locations of the ROIs and sequential binding for the dynamic execution program that yields the sequence of EMs. The scanpath sequence has enabled experimental dissection of these various bindings that appear to play independent roles and are likely located in different parts of the modular brain (see Appendix B for some principles of cortical connectivity, largely abstracted, perhaps with erroneous simplifications, from Professor Valentino Braitenberg). Cortical localization has advanced recently with fMRI studies on cooperating humans (see Appendix M) and supporting animal experiments (see Appendix N); this is not to ignore important sub-cortical loci with likely major functions.

In experiments carefully described in this paper, it is shown that symbolic binding strongly influences sequential binding, but can not overwhelm spatial or structural memory. Sequential bindings themselves appear to be partitioned between inherent and 'read-out' memory. The inherent sequential memory component is closely linked to structural binding, whereas the read-out components are apparently modified by each of the different motor systems we have explored --- EMs, hand control of a cursor on a computer screen, and locomotion over a grid on the laboratory floor.

Philosophy. The background of visual perception has ancient roots in philosophy. Although philosophers have long speculated that "we see in our mind's eye" and that we can have no certain knowledge of the external chaos or classes of appearances in which we find ourselves, yet until the scanpath theory no strong scientific evidence was available to support their conjectures. (The senior author was influenced by the strong TD structures he himself built into so-called 'artificial intelligence' programs so that the remaining 'self-organizing' was largely a matter of optimization, itself influenced by successive modifications.) (Stark 1993; Stark 1994; Stark, Okajima, and Whipple 1962; Stark 1997)

Computer Vision. Use of the TD scanpath for robotic computer vision has proved itself in a series of applications. Clearly TD information can change an ineffective vision system into a robust feedback mechanism for control of telerobots. Another positive aspect has been a complete and explicit demonstration of how a scanpath mechanism works in the artificial system.

This model may also suggest further directions for extending the sparse experimental data about the human brain mechanisms controlling our own vision. (Aloimonos and Herve 1992)

8. ACKNOWLEDGEMENTS

The senior author wants first to thank two of his oldest friends; Professor Valentino Braitenberg, Emeritus Professor of the Max Planck Institute of Biological Cybernetics in Tuebingen, whose studies of cortical connectivity and cerebral and cerebellar cortical models have been an important stimulus; and Professor Elwin Marg, of the School of Optometry at the University of California, Berkeley, whose passionate interest in visual neurophysiology and his pioneering single-unit studies of human brain have likewise been important. We wish to acknowledge support from Drs. Stephen Ellis and Michael Sims, scientific monitors for research grants with NASA-Ames Research Center; also Drs. Kamran Siminou of Neuro-Optics and Ken Kawamura of Fujita Research. Our laboratory colleagues served as subjects during preliminary control experiments, and often contributed discussion comments with bite bars in their mouths. Professors V. V. Krishnan and Gabriel M. Gauthier provided helpful, detailed critiques.

Appendix A: String-Editing Algorithm

The string-editing algorithm is a discrete dynamic programming method.(Bellman and Lee 1984) Using the operations --- insertion, In, deletion, De, and replacement, Re, the algorithm of Wagner (Wagner and Fischer) finds the minimum distance or cost to convert from string2(i) to string1(j); this defines the matrix (Figure 26, upper matrix). The two strings label the rows, string2(i), and columns, string1(j), of the matrix . Insertions result in horizontal shifts, deletions, vertical ones, and replacements produce shifts along the diagonal. Each operation may add to the cost; the coefficients of the matrix are the hypothetical costs to reach that cell. The middle matrix describes the stage just after the insertion of a "B" in the preceding step with an added cost of 1 (circled coefficient); the next step enables the "C's" to match without added cost. At the end (lower matrix) deletion of "A" (note vertical shift) finally matched string 2(i) to string 1(j), at a minimum cost of 2 (circled coefficient). Thus, the normalized distance is equal to 0.4 (2 divided by string length, 5); Ss, the sequential similarity index, is equal to 0.6 (1 – 0.4).

A short "C" program (Figure 26, equations listed to right of matrices) enabled these calculations.
1. Initialize D-matrix to zero. An additional possible step (not herein employed) is to truncate longer string.
2. Distance of first to null – do this by deleting each character in string j one by one; at most this will equal string length. Distance of second to null and first to second also calculated and will also equal at most string length.
3. Using dynamic programming, proceed from row to row and from top to bottom to calculate minimum distances; this fills out the D-matrix.
4. This triple computation, in addition to replacements, takes into account the effect of deletions and insertions in sidewise shifting of sting elements, and thus traveling along a minimum cost discrete path in the D-matrix. Wagner(Wagner and Fischer) proved these operations will find the optimim solution; this extended the discrete dynamic programming algorithm.
5. Lowermost right corner of D-matrix will be the minimum total cost of making string2(i) identical to string1(j).
6.- An additional possible step (not herein employed) is to assign non-unity costs to each operation

Our use of string editing in matching loci and sequences in images is a bit unusual. However, once we have established a finite state automaton and equivalently, a Markov model (see Figure 5, lower row), the sequences are inherently in a form appropriate for application of the string-editing algorithm. The widest use of string-editing algorithms is perhaps in spell-check programs. The use in matching of double-stranded chromosomes and sequences of nucleic acids within them, is an important current application. By using perhaps as yet undiscovered principles of biomechanical mechanisms, it may be possible to assign weightings, or non-unity costs, to such strand-distorting actions such as caused by insertions or deletions. For example, the redundant looping often seen in chromosomes may not be permissible for loops that are too short.

FIGURE 26 String Editing Algorithm
Successive stages of the discrete dynamic programming algorithm (matrices at left) document minimum cost optimization of string editing distance, and thus, accurate measure of string sequential similarity; computational equations at each stage (to right of matrices).

Appendix B: Consolidation

Not presented above, but also studied was Sp-st, that is the similarity of the loci to the stimuli loci (CH x CH, 0.70; WK x WK, 0.69; CH x WK, 0.68) that were less than Sp-coh (CH x CH, 0.90; WK x WK, 0.91; CH x WK, 0.87), the coherence of a set of remembered loci to one another as a result of consolidation. This was quantitative evidence that a remembered pattern was often closely repeated, without however, necessarily being an accurate reflection of the stimulus.

In other experiments, the alternation of CH and WK, was performed with refreshment, such that the second viewing opportunity enabled them to develop another somewhat different short-term memory schema. This in turn led to lower WK x CH and CH x WK similarities, Sp = 0.62 and Ss = 0.25, as compared to values of Sp = 0.87 and Ss = 0.60 above without refreshment. Contrariwise, the Sp-st value with refreshment, 0.72, was almost the same as the value from the above experiment, Sp-st = 0.68. This ability to establish a new memory trace could be enhanced even further by allowing a delay of about ten minutes and also with interjection of other patterns, and of intercalated tilt adaptation experiments.(Stark et al. 1999) (This topic will be further developed below in dealing with the symbolic experiment.)

Appendix C: The Braitenberg Cortex

Some principles of cortical connectivity are listed, largely abstracted, perhaps with erroneous simplifications, from Professor Valentino Braitenberg. (Braitenberg 1977; Braitenberg 1990; Braitenberg 1994; Braitenberg and Schüz 1998)

1.- The idea of columns of cortical cells may have been initiated by Warren McCulloch (1945); then Vernon Mountcastle with the somatosensory cortex, and David Hubel and Torsten Wiesel with the visual cortex, strongly supported this concept in the 1950s.

2.- Six layers or levels of cortical neurons.

3.- Almost all are pyramidal cells; exceptions seem to be truncated or inverted pyramidal cells, with eponymous names.

4.- Ponder the constraint of a pair of cortical cells not to connect more than twice with one another -- one connection in each direction. How then is any pair more connected than any other pair? By virtue of their connectiveness to a common group of cells. Each cell has about 20,000 outputs (axons) and about 20,000 inputs (dendritic knobs or synapses). If two cells share none of their other connections then they are 'unrelated'; if almost all of the connections are to the same cells then they are closely 'related'.

5.- For each pyramidal cell the 20,00 inputs and outputs connect almost exclusively to other cortical cells. On average, only about one output proceeds toward an output motor relay and only about one input arrives from a sensory waystation. Let not the skeptic deny that most of our brain computation is within the cortex and not with multiple stages of processing input and output. Of course, the design of present-day experiments forces this input-output view with paradigms to test responsivity to stimuli and to observe regular output responses to stimulation.

6.- Lateral axons and their web of connectivity show a decreasing density of connections with distance. Apical axons to apical dendrites appear to be relatively independent of distance; this leads to widespread connectivity. Indeed the size and number of a column in a human brain is sufficient for each column to be connected with every other column of the cortex. For the mouse with its much smaller brain, this is still true. The lesser number of cells in a column still suffice for this order of connectivity to the lesser number of columns.

7.- Braitenberg and Almut Shutz showed in hamsters that the tremendous growth of dendritic knobs and axonal connections occurred developmentally before these creatures experienced the outside world. Similarly, development in Coghill's salamander embryos went on apace and functional connectivity reached an appropriate stage even when the salamanders were raised in anesthetic solution until tested. (Coghill 1929)

Appendix M : Function Magnetic Resonance Imaging, fMRI

We are excited about the new experimental paradigms using fMRI and PET as a non-invasive techniques for working with alert, cooperating humans engaged in perceptual and other high-level tasks. For the benefit of readers, we include a few references to current work ongoing in fMRI. In general, the neurological studies and neurophysiological findings regarding the modular cortex have been strengthened and deepened with these ongoing human researches.(Born and Tootell 1992; Coull, Frackowiak, and Frith 1998; Culham et al. 1998; de Jong et al. 1999; Fletcher et al. 1998; Kastner et al. 1998; Kleinschmidt et al. 1998; Kosslyn 1980; Kosslyn 1994; Kosslyn and Osherson 1995; Mackworth 1978; O'Sullivan et al. 1995; Ploner et al. 1999; Posner and Raichle 1994; Rugg et al. 1998; Tempini et al. 1998; Thompson-Schill et al. 1997; Tootell, Malonek, and Grinvald 1994; Tootell et al. 1995a; Tootell et al. 1995b; Zeki and Bartels 1998)

Appendix N: Recent Neurophysiology for Perception and Other Higher Level Functions

Classical neurophysiological techniques, and by have enriched our knowledge of the animal brain, and by analogy, of the human brain; these are reinforced by longstanding and recent neurological studies. Of course, animal experiments are difficult, especially considering that it is impossible to have verbal interaction with the subjects of the experiment, so essential for studying higher-level functions. As mentioned in the paper, the design of neurophysiological experiments has been forced into paradigms that are exclusively input-output studies. That is, a visual stimulus has consequences which can be measured in various locations in the animal brain. Contrariwise, it is all but impossible to measure what the TD functions of the animal brain are signalling to these regions. However, with great ingenuity, a number of inroads have been made.

Most neurophysiology is neuroanatomy; that is, locations of regions that show activity during particular sensory processing. Pioneering and future studies in actual microcortical neurophysiology are expanding with multiple electrode, optical and molecular biological approaches. (Andersen 1995; Colby, Duhamel, and Goldberg 1995; Desimone 1992; Devalois 1960; Galletti et al. 1996; McAllister, Lo, and Katz 1995; Nelson and Katz 1995; Rogers et al. 1998; Sillito and Jones 1996; Sillito, Salt, and Kemp 1985; Sillito and Grieve 1991; Umeno and Goldberg 1997; Zeki and Bartels 1998)

9. FIGURE LIST

0. ABSTRACT Keywords

1. INTRODUCTION

2. METRICS AND ANALYSES

3. RESULTS: STRUCTURAL AND SEQUENTIAL BINDING

4. RESULTS : SYMBOLIC BINDING

5. PHILOSOPHY AND CORTICAL MODULUES

6. TELEROBOTIC SCHEME

8. ACKNOWLEDGEMENTS

10. REFERENCES

Aloimonos, Y., and J.Y. Herve. 1992. Exploratory active vision: theory. Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, at Los Alamitos, CA.

Andersen, R. A. 1995. Coordinate transformations and motor planning in posterior parietal cortex. In *The Cognitive Neurosciences*, edited by M. S. Gazzaniga. Cambridge, MA: MIT Press.

Bajcsy, R., and E. Krotkov. 1993. Active vision for reliable ranging: cooperating focus, stereo, and vergence. *International Journal of Computer Vision* 11:187-203.

Ballard, D.H., M.M. Hayhoe, and J.B. Pelz. 1994. Visual Representations in Natural Tasks. Proceedings of the Workshop on Visual Behaviors, at Los Alamitos, CA.

Bejczy, A. K. 1980. Sensors, controls and man-machine interface for advanced teleoperation. *Science* 208:1327-1335.

Bellman, R., and E.S. Lee. 1984. History and development of dynamic programming. *IEEE Control Systems Magazine* 4 (4):24-28.

Blackmon, T. T., Y. F. Ho, D. Chernyak, L. W. Stark, and M. Azzariti. 1999. Dynamic Scanpaths: Eye Movement Analysis Methods. *IS&T/SPIE's Symposium on Electronic Imaging*.

Blackmon, T. T., and L. W. Stark. 1996. Model-Based Supervisory Control in Telerobotics. *Presence* 5:205-223.

Bolle, R., Y. Aloimonos, and C. Fermuller. 1998. Toward motion picture grammars. Third Asian Conference on Computer Vision. Proceedings, at Hong Kong.

Born, Richard T., and Roger B. Tootell. 1992. Segregation of global and local motion processing in primate middle temporal visual area. *Nature* 357 (6378):497-499.

Braitenberg, Valentino. 1977. *On the texture of brains: an introduction to neuroanatomy for the cybernetically minded.* Edited by T. Elisabeth Hanna Braitenberg. New York: Springer-Verlag.

Braitenberg, Valentino. 1990. The cerebral cortex as site of associative memory. *Sistemi Intelligenti* August 2 (2):213-227.

Braitenberg, Valentino. 1994. *Vehicles, experiments in synthetic psychology.* Cambridge, MA: MIT Press.

Braitenberg, Valentino, and Almut Schüz. 1998. *Cortex : statistics and geometry of neuronal connectivity.* 2nd ed. Berlin: Springer.

Brandt, S., L. W. Stark, S. Hacisalihzade, J. Allen, and G. Tharp. 1989. Experimental Evidence for Scanpath Eye Movements During Visual Imagery. Proc. 11th IEEE/EMBS, at Seattle, WA.

Brandt, Stephan A., and Lawrence W. Stark. 1997. Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience* 9 (1):27-38.

Buttolo, P., D. Kung, and B. Hannaford. 1995. Manipulation in Real, Virtual, and Remote Environments. *IEEE Systems, Man, Cybernetics* SMC-5:4656-4661.

Carpenter, G.A., S. Grossberg, and G.W. Lesher. 1998. The what-and-where filter. A spatial mapping neural network for object recognition and image understanding. *Computer Vision and Image Understanding* 69 (1):1-22.

Carson, C., S. Belongie, H. Greenspan, and J. Malik. 1997. Region-based image querying. Proceedings, IEEE Workshop on Content-Based Access of Image and Video Libraries, at San Juan, Puerto Rico.

Choi, Yun, Anthony Mosley, and L. W. Stark. 1995. String Editing Analysis of Human Visual Search. *Optometry and Vision Science* 72 (7):439-451.

Coghill, G.E. 1929. *Anatomy and the Study of Behavior.* Cambridge: Cambridge University Press.

Colby, CL, JR Duhamel, and ME. Goldberg. 1995. Oculocentric spatial representation in parietal cortex. *Cerebral Cortex* 5 (5):470-481.

Coull, J.T., R.S. Frackowiak, and C.D. Frith. 1998. Monitoring for Target Objects: Activation of Right Frontal and Parietal Cortices with Increasing Time on Task. *Neuropsychologia* 36 (12):1325-1334.

Crevier, D., and R. Lepage. 1997. Knowledge-based image understanding systems: a survey. *Computer Vision and Image Understanding* 67 (2):161-185.

Crosby, M. E. 1990. How do we read algorithms? *Computer* 23:25-35.

Culham, Jody C., Stephan A. Brandt, Patrick Cavanagh, Nancy G. Kanwisher, Anders M. Dale, and Roger B. H. Tootell. 1998. Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology* 80 (5):2657-2670.

de Jong, B.M., R.S. Frackowiak, A.T. Willemsen, and A.M. Paans. 1999. The Distribution of Cerebral Activity Related to Visuomotor Coordination Indicating Perceptual and Executional Specialization. *Cognitive Brain Research* 8 (1):45-59.

Desimone, R. 1992. Neural Circuits for Visual Attention in Primate Brain. In *Neural Networks for Vision and Image Processing*, edited by G. A. Carpenter, and S. Grossberg. Cambridge, MA: MIT Press.

Devalois, Russell L. 1960. Color Vision Mechanisms in the Monkey. *Journal of General Physiology* 43 (Supplement):115-128.

Dow, B. M., A. Z. Snyder, R. G. Vautin, and R. Bauer. 1981. Magnification factor and receptive field size in foveal striate cortex of monkey. *Experimental Brain Research* 44:213-228.

Driels, M., and J. Acosta. 1992. The duality of haptic and visual search for object recognition. Proceedings of the IEEE International Symposium on Intelligent Control, at New York.

Elderfield, John. 1998. Seeing Bonnard. In *Bonnard*. New York: Museum of Modern Art.

Ellis, Stephen R., and Lawrence W. Stark. 1979. Reply to Piggins. *Perception* 8 (6):721-722.

Ferrell, W. R., and T. B. Sheridan. 1967. Supervisory Control of Remote Manipulation. *IEEE Spectrum* 4:81-88.

Flagg, B. N. 1978. Children and Television: Effects of Stimulus Repetition on Eye Activity. In *Eye Movements and the Higher Psychological Functions*, edited by J. W. Senders, D.F. Fisher, and R. A. Monty. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fletcher, P.C., T. Shallice, C.D. Frith, R.S. Frackowiak, and R.J. Dolan. 1998. The functional roles of prefrontal cortex in episodic memory. II. Retrieval. *Brain* 121 (7):1249-1256.

Foresti, G. L., and G. Pieroni. 1998. Exploiting neural trees in range image understanding. *Pattern Recognition Letters* 19 (9):869-878.

Freksa, C. 1992. Using orientation information for qualitative spatial reasoning. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. International Conference GIS - From Space to Territory, 21-23 September 1992, at Pisa, Italy.

Freksa, C. 1997. Foundations of computer science: Potential theory - cognition. In *Spatial and temporal structures in cognitive processes: Foundations of computer science.*, edited by C. Freksa, M. Jantzen, and R. Valk. Berlin: Springer-Verlag.

Galletti, C., P. Fattori, P.P. Battaglini, S. Shipp, and S. Zeki. 1996. Functional demarcation of a border between areas V6 and V6A in the superior parietal gyrus of the macaque monkey. *European Journal of Neuroscience* 8 (1):30-52.

Gauthier, G., J.-M. Hofferer, W.F. Hoyt, and L. Stark. 1979. Visual-Motor Adaptation: Quantitative Demonstration in Patients with Posterior Fossa Involvement. *Archives of Neurology* 36:155-160.

Gauthier, Gabriel M., Jean-Louis Vercher, David S. Zee, and . 1994 Oct. 34 (19): p. 2613-2627. 1994. Changes in ocular alignment and pointing accuracy after sustained passive rotation of one eye. *Vision Research* 34 (19):2613-2627.

Gould, J.D. 1967. Pattern recognition and eye-movement parameters. *Perception and Psychophysics* 2:399-407.

Groner, R., F. Walder, and M. Groen. 1984. Looking at faces: Local and global aspects of scanpaths. In *Theoretical and Applied Aspects of Eye Movement Research*, edited by A. G. Gale, and F. Johnson. North Holland, Amsterdam.

Hacisalihzade, S.S., L.W. Stark, and J.S. Allen. 1992. Visual perception and sequences of eye movement fixations: A stochastic modeling approach. *IEEE Transactions on Systems, Man, and Cybernetics* 22:474-481.

He, Zijiang J., and Ken Nakayama. 1992. Surfaces versus features in visual search. *Nature* 359 (6392):231-233.

Henderson, and Hollingsworth. 1999. Higher level scene perception. *Annual reviews of psychology* 50:243-271.

Ho, Yeuk F., Hideyoshi Masuda, Hiroshi Oda, and Lawrence Stark. 1999. Distributed Control for Tele-Operations. *Journal of Fujita Technical Research Institute* 10:1-7.

Ho, Yeuk F., and L. W. Stark. 1997. Top-Down Image Processing and Supervisory Control Limitations in Robotics: A Simulation Study. 8th Intl. Conf. on Advanced Robotics (ICAR '97), July 7-9, at Monterey, CA.

Ho, Y. F., and L. W. Stark. 1999a. Design Simulation of a Web-Based Supervisory Control System. International Conference on Web-Based Modeling and Simulation, January.

Ho, Y.F., and L.W. Stark. 1999b. Model-based visual detection and verification system. SPIE: Electronic Imaging, at San Jose.

Ho, Y. F., and L. W. Stark. 1999c. Visual Tracking of Tele-Operated Robots Using Model-Based Algorithms. IS&T/SPIE's Symposium on Electronic Imaging, January 24-29, at San Jose, CA.

Ho, Yeuk Fai, and L. W. Stark. 2000. Scanpath-Based Model for Visual Tracking of Tele-Robots. IS&T/SPIE's Symposium on Electronic Imaging, at San Jose, CA.

Ingle, D. 1971. Prey-catching behavior of Anurans toward moving and stationary objects. *Vision Research* Supplement 3:447-456.

Itti, L., and C. Koch. 1999. Comparison of feature combination strategies for saliency-based visual attention systems. SPIE: Electronic Imaging, at San Jose.

Jeannerod, M., P. Gerin, and J. Pernier. 1968. Deplacements et fixation du regard dans l'exploration libre d'une scene visuelle [French]. *Vision Research* 8:81-97.

Kant, I. 1949. *Prolegomena to Any Future Metaphysics*. New York: Bobbs-Merrill Company, Inc.

Kastner, Sabine, Peter De Weerd, Robert Desimone, and Leslie G. Ungerleider. 1998. Mechanisms of Directed Attention in the Human Extrastriate Cortex as Revealed by Functional MRI. *Science* October 2:108-111.

Kim, Won Soo, Stephen R. Ellis, Mitchell E. Tyler, Blake Hannaford, and L. W. Stark. 1987. Quantitative Evaluation of Perspective and Stereoscopic Displays in Three-Axis Manual Tracking Tasks. *IEEE Systems, Man and Cybernetics* 16:61-72.

Kim, W. S., M. Takeda, and L. Stark. 1988. On-the-Screen Visual Enhancements for a Telerobotics Vision System. Proceedings of the IEEE International Conference of Systems, Man, & Cybernetics, at Beijing.

Kim, Won Soo, Frank Tendick, and L. W. Stark. 1987. Visual Enhancement in Pick-and-Place Tasks: Human Operators Controlling a Simulated Cylindrical Manipulator. *IEEE J. of Robotics and Automation* 3:418-425.

Klatzky, R.L. 1998. Allocentric and egocentric spatial representations: definitions, distinctions, and interconnections. In *Spatial Cognition. An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, edited by C. Freksa, C. Habel, and K. F. Wender. Berlin: Springer-Verlag.

Klatzky, Roberta L., Jack M. Loomis, Reginald G. Golledge, Joseph G. Cicinelli, and et al. 1990. Acquisition of route and survey knowledge in the absence of vision. *Journal of Motor Behavior* March 22 (1):19-43.

Kleinschmidt, A., C. Buchel, S. Zeki, and R.S. Frackowiak. 1998. Human Brain Activity During Spontaneously Reversing Perception of Ambiguous Figures. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265 (1413):2427-2433.

Kosslyn, S. M. 1980. *Image and Mind.* Cambridge, MA: Harvard University Press.

Kosslyn, Stephen Michael. 1994. *Image and Brain: The Resolution of the Imagery Debate.* Cambridge, MA: MIT Press.

Kosslyn, Stephen M., and Daniel N. Osherson, eds. 1995. *Visual cognition.* 2nd ed. Cambridge, MA: MIT Press.

Kruskal, J.B. 1983. An overview of sequence comparison: Time warps, string edits, and macromolecules. *SIAM Review* 25:201-237.

Lawden, M.C., H. Bagelmann, T.J. Crawford, T.D. Matthews, and C. Kennard. 1995. An effect of structured backgrounds on smooth pursuit eye movements in patients with cerebral lesions. *Brain* 118 (1):37-48.

Lettvin, J.Y., H.R. Maturana, W.S. McCulloch, and W.H. Pitts. 1959. What the Frog's Eye Tells the Frog's Brain. *Proceedings of the IREE* 47:1940-1959.

Liu, Andrew, Gregory Tharp, Stephen Lai, Lloyd French, and Lawrence W. Stark. 1993. Some of What One Needs to Know about Using Head-Mounted Displays to Improve Teleoperator Performance. *IEEE Transactions on Robotics and Automation* 9 (5):638-48.

Llewellyn-Thomas, E. 1968. Movements of the Eye. *Scientific American* 219:88-95.

Locher, P.J., and C.F. Nodine. 1974. The Role of Scanpaths in the Recognition of Random Shapes. *Perception and Psychophysics* 15:308-314.

Mackeben, Manfred, and Ken Nakayama. 1993. Express attentional shifts. *Vision Research* 33 (1):85-90.

Mackworth, A. K. 1978. How to See a Simple World: An exegesis of some computer programs for scene analysis. *Machine Intelligence* 8:510-537.

Mackworth, N. H., and J. S. Bruner. 1970. How adults and children search and recognize pictures. *Human Development* 13:149-177.

Mackworth, N.H., and A.J. Morandi. 1967. The gaze selects informative details within picture. *Perception and Psychophysics* 2:547-552.

Mandler, M.B., and J.A. Whiteside. 1976. The role of scanpaths in recognition of random dot patterns. *Journal of Undergraduate Psychology Research* 3 (84-90).

McAllister, A.K., D.C. Lo, and L.C. Katz. 1995. Neurotrophins regulate dendritic growth in developing visual cortex. *Neuron* 15 (791-803).

McCulloch, Warren. 1965. *Embodiments of Mind.* Cambridge, MA: MIT Press.

McPeek, Robert M., Vera Maljkovic, and Ken Nakayama. 1999. Saccades require focal attention and are facilitated by a short-term memory system. *Vision Research* 39 (8):1555-1566.

Meystel, A.M., I.A. Rybak, S. Bhasin, and M.A. Meystel. 1992. Multiresolution stroke sketch adaptive representation and neural network processing system for gray-level image recognition. *Proceedings of SPIE: Intelligent Robots and Computer Vision XI: Biological, Neural Net, and 3-D Methods* 1826:261-278.

Miyata, K., and L. Stark. 1992. Active camera control: seeing around obstacles. Power Electronics and Motion Control: Proceedings of the International Conference on Industrial Electronics, Control, Instrumentation, and Automation.

Moray, N., R. Ferrell, H. G. Stassen, and D. R. et al. Yoerger. 1989. Supervisory Control: 30 Years and Counting. Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.

Nakayama, Ken, Zijiang J. He, and Shinsuke Shimojo. 1995. Visual surface representation: A critical link between lower-level and higher-level vision. In *Visual Cognition: An*

Invitation to Cognitive Science, edited by S. M. Kosslyn, Daniel N. Osherson, et al. Cambridge, MA: MIT Press.

Nakayama, Ken, and Julian S. Joseph. 1998. Attention, pattern recognition, and pop-out visual search. In *The Attentive Brain*, edited by R. Parasuraman, et al. Cambridge, MA: MIT Presss.

Nelson, D.A., and L.C. Katz. 1995. Emergence of functional circuits in ferret visual cortex visualized by optical imaging. *Neuron* 15:23-34.

Nguyen, A. H., and L. W. Stark. 1993. Model Control of Image Processing: Pupillometry. *Computerized Medical Imaging and Graphics* 17 (1):21-33.

Niebur, Ernst, and Christof Koch. 1998. Computational architectures for attention. In *The Attentive Brain*, edited by e. a. Raja Parasuraman. Cambridge, MA: MIT Press.

Noton, David, and Lawrence W. Stark. 1971a. Eye Movements and Visual Perception. *Scientific American* 224 (6):34-43.

Noton, David, and Lawrence W. Stark. 1971b. Scanpaths in Eye Movements During Pattern Perception. *Science* 171 (3968):308-311.

Noton, D., and L. W. Stark. 1971c. Scanpaths in Saccadic Eye Movements while Viewing and Recognizing Patterns. *Vision Research* 11 (9):929-42.

O'Sullivan, E.P., I.H. Jenkins, L. Henderson, C. Kennard, and D.J. Brooks. 1995. The functional anatomy of remembered saccades: a PET study. *Neuroreport* 6 (16):2141-2144.

Palmer, S. 1975a. Visual Perception and World Knowledge: Notes on a model of sensory-cognitive interaction. In *Explorations in Cognition*, edited by D. A. Norman, and D. E. Rumelhart. San Francisco: Freeman.

Palmer, Stephen E. 1975b. The Effects of Contextual Scenes on the Identification of Objects. *Memory & Cognition* 3 (5):519-526.

Palmer, Stephen E. 1992. Reference frames in the perception of spatial structure. In *Cognition, information processing, and psychophysics: Basic issues*, edited by H.-G. Geissler, Stephen W. Link, et al. Hillsdale, NJ: Lawrence Erlbaum Associates.

Palmer, Stephen E. 1999. *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

Palmer, Stephen E., and Ruth Kimchi. 1986. The information processing approach to cognition. In *Approaches to cognition: Contrasts and controversies*, edited by T. J. Knapp, Lynn C. Robertson, et al. Hillsdale, NJ: Lawrence Erlbaum Associates.

Palmer, Stephen E., Jonathan Neff, and Diane Grouping Beck. 1997. Amodal Completion. In *Indirect Perception*, edited by I. Rock. Cambridge, MA: MIT Press.

Parker, R.E. 1978. Picture processing during recognition. *Journal of Experimental Psychology: Human Perception and Performance* 4:284-293.

Pitts, Walter, and W. S. McCulloch. 1947. How We Know Universals: The Perception of Auditory and Visual Forms. *Bulletin of Mathematical Physics* 9:127-147.

Ploner, Christoph J., Bertrand M. Gaymard, Nathalie Ehrle, Sophie Rivaud-Pechoux, Michel Baulac, Stephan A. Brandt, Stephane Clemenceau, Severine Samson, and Charles Pierrot-Deseilligny. 1999. Spatial memory deficits in patients with lesions affecting the medial temporal neocortex. *Annals of Neurology* 45 (3):312-319.

Posner, M.I., and M.E. Raichle. 1994. *Images of Mind*. New York: Scientific American Library.

Pribram, K.H. 1971. *Languages of the Brain*. Englewood Cliffs, NJ: Prentice-Hall.

Privitera, C., N. Krishnan, and L.W. Stark. 1999. Clustering algorithms to obtain regions-of-interest (ROIs). SPIE: Electronic Imaging, at San Jose.

Privitera, C. M., and L. W. Stark. 1998. Evaluating Image Processing Algorithms that Predict Regions of Interest. *Pattern Recognition Letters* 19:1037-1043.

Rogers, R.D., B.J. Sahakian, J.R. Hodges, C.E. Polkey, C. Kennard, and T.W. Robbins. 1998. Dissociating executive mechanisms of task control following frontal lobe damage and Parkinson's disease. *Brain* 121 (5):815-842.

Rugg, M.D., P.C. Fletcher, K. Allan, C.D. Frith, R.S. Frackowiak, and R.J. Dolan. 1998. Neural correlates of memory retrieval during recognition memory and cued recall. *Neuroimage* 8 (3):262-273.

Russell, B. 1945. *A History of Western Philosophy*. New York: Simon and Schuster.

Rybak, I.A., A.V. Golovan, and V.I. Gusakova. 1993. Behavioral model of visual perception and recognition. SPIE Proceedings: Human Vision, Visual Processing, and Digital Display IV.

Schifferli, P. 1953. Etude par enregistrement photographique de le motricite oculaire dans l'exploration, dans la reconnaissance et dans la representation visuelles [French]. *Monatschrift fur Psychiatrie und Neurologie* 126:65-118.

Schill, K., E. Umkehrer, S. Beinlich, G. Krieger, and C. Zetzsche. 1999. Knowledge-based scene analysis with saccadic eye movements. SPIE: Electronic Imaging, at San Jose.

Schwartz, E. L. 1984. Anatomical and physiological correlates of visual computation from striate to infero-temporal cortex. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-14: 2.

Searle, J.R. 1983. *Intentionality, an essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Senders, J.W., D.F. Fisher, and R.A. Monty, eds. 1978. *Eye movements and the higher psychological processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sheridan, T. B. 1992. *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: MIT Press.

Sillito, A.M., and H.E. Jones. 1996. Context-dependent interactions and visual processing in V1. *Journal de Physiologie* 90 (3-4):205-209.

Sillito, A.M., T.E. Salt, and J.A. Kemp. 1985. Modulatory and inhibitory processes in the visual cortex. *Vision Research* 25 (3):375-381.

Sillito, A.M.J., and K.L. Grieve. 1991. A re-appraisal of the role of layer VI of the visual cortex in the generation of cortical end inhibition. *Experimental Brain Research* 87:521-529.

Singer, J., S. Greenberg, and J. Antrobus. 1971. Looking at the mind's.eye: Experimental studies of ocular motility during day dreaming. Transactions of the New York Academy of Science.

Stark, L. 1993. Neural nets, random design and reverse engineering. Proceedings of the IEEE International Conference on Neural Networks, at San Francisco, CA.

Stark, L. 1994. ANNs and MAMFs: Transparency or Opacity? Proceedings of the European Neural Network Society, at Sorrento.

Stark, L., and Stephen Ellis. 1981. Scanpaths Revisited: Cognitive Models Direct Active Looking. :193-226.

Stark, L., Stephen Ellis, Hiromitsu Inoue, Christian Freksa, Zipora Portnoy, and Joshua Zeevi. 1979. Cognitive Models Direct Scanpath Eye Movements: Evidence obtained by Means of Computer Processing of Perceptual Eye Movements. XII International Conference on Medical and Biological Engineering, August 1979, at Jerusalem, Israel.

Stark, L., Barbara Mills, An Nguyen, and Huy X. Ngo. 1988. Instrumentation and Robotic Image Processing Using Top-down Model Control. In *Robotics and Manufacturing*, edited by J. e. al. New York: ASME.

Stark, L., Mitsuharu Okajima, and Gerald H. Whipple. 1962. Computer Pattern Recognition Techniques: Electrocardiographic Diagnosis. *Communications of the Association for Computing Machinery* 5:527-532.

Stark, L., W. Zangemeister, B. Hannaford, and K. Kunze. 1986. Use of Models in Brainstem Reflexes for Clinical Research. In *Clinical problems of brainstem disorders*. New York: Thieme.

Stark, L. W. 1997. Top-Down and Bottom-Up Image Processing. 1997 International Conference on Neural Networks.

Stark, Lawrence W., Yun Choi, and Yong Yu. 1996. Visual Imagery and Virtual Reality: New Evidence Supporting the Scanpath Theory Explains the Illusion of Completeness and Clarity. *Visual Science: Papers in Honor of J. Enoch.*

Stark, L. W., and Yun S. Choi. 1996. Experimental Metaphysics: The Scanpath as an Epistemological Mechanism. In *Visual Attention and Cognition*, edited by W. H. Zangemeister, H. S. Stiehl and C. Freksa. Amsterdam: Elsevier.

Stark, L. W., C. M. Privitera, H. Yang, Y. F. Ho, M. Azzariti, A. Chan, C. Krischer, and A. Weinberger. 1999. Scanpath Memory Binding: Multiple Read-Out Experiments. IS&T/SPIE's Symposium on Electronic Imagining, Jan., at San Jose, CA.

Stelmach, L., W. Tam, and P. Hearty. 1992. Static and dynamic spatial resolution in image coding: An investigation of eye movements. Proceedings, SPIE, at San Jose, CA.

Sutro, L.L., and J.B. Lerman. 1973. Robot Vision. First National Conference on Remote Manned Systems, at Pasadena.

Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268 (5217):1632-1634.

Tempini, M.L., C.J. Price, O. Josephs, R. Vandenberghe, S.F. Cappa, N. Kapur, and R.S. Frackowiak. 1998. The neural systems sustaining face and proper-name processing. *Brain* 121 (11):2103-2118.

Thompson-Schill, S. L., M. D'Esposito, G. K. Aguirre, and M. J. Farah. 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Science* 94 (26):14792-14797.

Tootell, R.B., D. Malonek, and A. Grinvald. 1994. Optical imaging reveals the functional architecture of neurons' processing, shape, and motion in owl monkey area MT. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 258:109-119.

Tootell, Roger B. H., John B. Reppas, Anders M. Dale, Rodney B. Look, and et al. 1995a. Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature* 375 (6527):139-141.

Tootell, Roger B. H., John B. Reppas, Kenneth K. Kwong, Rafael Malach, and et al. 1995b. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience* 15 (4):3215-3230.

Umeno, MM, and ME. Goldberg. 1997. Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *Journal of Neurophysiology* 78 (3):1373-1383.

Uttal, William R., Todd Baruch, and Linda Allen. 1995. The effect of combinations of image degradations in a discrimination task. *Perception and Psychophysics* 57 (5):668-681.

Wagner, R.A., and M.J. Fischer. 1974. The string-to-string correction problem. *Journal of the Association for Computing Machinery* 21 (1):168-173.

Weirda, M., and W. Maring. 1993. Interpreting eye movements of traffic participants. In *Visual Search 2 - Proceedings of the Second International Conference on Visual Search*, edited by D. Brogan, A. Gale and K. Carr. London: Taylor and Francis.

Wolfe, Jeremy M. 1998. Visual memory: What do you know about what you saw? *Current Biology* 8 (9):R303-R304.

Wolfe, Jeremy M., George A. Alvarez, and Todd S. Horowitz. in preparation, 2000. Attention is fast but volition is slow.

Yang, Huiyang, and L. W. Stark. 2000. How Do We Recognize Images? IS&T/SPIE's Symposium on Electronic Imaging, at San Jose, CA.

Yarbus, A. L. 1967. *Eye Movements and Vision*. New York: Plenum Press.

Yoerger, D. R., and J. R. Slotline. 1987. Supervisory Control Architecture for Underwater Teleoperation. *Proceedings of the IEEE International Conference on Robotics and Automation* RA-3:2068-2073.

Yu, Yong, and Lawrence W. Stark. 1995. An Active Model-Based Algorithm for Correspondence and Estimation of Pose Parameters of Objects. IEEE International Conference on Systems, Man, and Cybernetics: Intelligent Systems for the 21st Century, October 22-25, at Vancouver, British Columbia.

Zangemeister, W. H., K. Sherman, and L. W. Stark. 1995. Evidence for a Global Scanpath Strategy in Viewing Abstract Compared with Realistic Image. *Neuropsychologia* 33 (8):1009-1025.

Zangemeister, W.H., H.S. Stiehl, and C. Freksa, eds. 1996. *Visual Attention and Cognition.* Amsterdam: Elsevier.

Zeki, S., and A. Bartels. 1998. The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 353 (1377):1911-1914.

Zeki, S., and K. Moutoussis. 1997. Temporal hierarchy of the visual perceptive systems in the Mondrian world. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 264 (1387):1415-1419.

Zelnio, E. G. 1991. Air Traffic Control Paradigm; Comparison with Emphasis on Model-based Vision. *SPIE Proceedings* 1609:2-15.

FIGURES

FIGURE 1 Log-Polar Distortions of a Picture
Two fixations (left and right panels below) on original picture (upper) show log-polar distortions with high cortical magnification (irregular shapes, lower left and lower right) of successive foveal ROIs (circles), as well as minification of peripheral regions likely captured as textured or colored segments (surrounding small squares, lower).

FIGURE 2 EMs while Looking at an Ambiguous Figure: The Ellis Experiment
Identical ambiguous figures of vase (lower left) and two faces (upper left). EMs superimposed on ambiguous figures as they were actually seen following exposure to priming stimuli (right).

FIGURE 3 Dynamic Display with EMs
Animation of dynamical scenarios (illustrated as snapshots every five seconds time proceeding from lower left to upper right). EM positions (black circles) taken every 50ms are integrated over the preceding 5s, and are superimposed onto snapshot images; they represent the basic data captured for this experiment.

FIGURE 4 Scanpath Theory
EM positions (q 50ms) during dynamic display shown as a connected sequence (upper left) while the dynamical ROIs visited form a connected sequence (upper right and lower right). By numbering or letter identification of smooth pursuit or static fixations, this sequential string of visited ROIs could be defined. A non-iconic model of alternating perceptual ROIs (lettered squares) and saccadic EMs (circles with arrows) is shown by solid arrows for the experiment presented. This is the "feature ring" of the scanpath theory. On other presentations of the stimulus, other ROIs and sequences could be formed (dashed arrows).

FIGURE 5 EMs while Engaged in Visual Imagery: The Brandt Experiment
Scanpath EM sequence is almost the same for the second looking presentation (middle row, second grid) as for the first visual imagery presentation (middle row, third grid). During the visual imagery presentation, no information about the location of the alphabetic symbols, Fs, was available; thus, the remembered representational model must have controlled the scanpath in a TD fashion. Quantitative metrics could be obtained in the analysis procedure (lower row) by creating a finite state automata (middle) for generating the scanpath; then transition probability coefficients could be arranged in a Markov matrix (right) for later statistical analysis.

FIGURE 6 EM Trajectories and Classification
Trajectories of EMs displayed as functions of vertical and horizontal angles (solid lines) and time. Location of the dynamic objects shown as dashed and continuous lines (lower), or for one comparison as a dashed line for horizontal angle (upper). Note saccades, S, and smooth pursuits, SP, that show up clearly; these and other types of movements (see text) could be identified and analyzed (lower steps and labels).

FIGURE 7 EM Classification Program
Flow diagram for EM calibration, linearization, differentiation to obtain velocities, and then analysis into various categories of EM types. The program could also resolve conflicts among specific EM identification algorithms (see text).

FIGURE 8 Simplified, or Toy, Diagram Illustrating Metrics for Comparing Scanpaths; and Parsing Diagram for the Dynamic Scanpath Experiment
Quantitative methodology diagrammed to show similarity indices, Sp and Ss (upper panel). These pairwise comparisons are organized into Y-matrices (middle panel) and then indices segregated, averaged, and placed into parsing diagrams (lower panel). Note statistical tests indicated by bolding, as well as arrows (see text). These two parsing diagrams summarize the experimental base from a dynamic scanpath study (see text).

FIGURE 9 EMs Compared with Choice, CH: Selection of ROIs Compared
Linearized EMs (upper left) were analyzed into fixations and saccades (upper right) while the subject looked at a cave painting of horses. Loci chosen by mouse clicks (lower left) could then be compared (lower right) with EM fixations (see text).

FIGURE 10 Parsing Diagrams; EMs Compared with Choice, CH
Sp (upper) and Ss (lower) parsing diagrams for the choice compared with EM study. Intramodal read-out comparisons (left and middle panels) as well as intermodal read-out comparisons (right panel). Results described in text.

FIGURE 11 CH compared with Walking Protocol
Experimental protocol for cross-modal comparison between choice, CH, and walk, WK, read-out modes. Note similarity of patterning when a second display of stimuli patterns was not given (absent grids in both sets of upper panels); note difference in patterning when refreshment of stimulus pattern allowed a new memory schema to be formed (lower panel).

FIGURE 12 Visual imagery: Parsing Diagram for Walking, WK, vs Choice, CH
Sp (upper) and Ss (lower) parsing diagrams for the choice, CH, compared with walking, WK, study. Intramodal read-out comparisons (left and middle panels) as well as intermodal read-out comparisons (right panel). Results described in text.

FIGURE 13 Modular Cortex and Connectivity
Recent studies in neurophysiology and fMRI have established a "new phrenology," the modular cortex (upper), with different functions assigned to specific regions of the cortex (see text for further explanation). Connectivity explored in our experiments on inherent and read-out sequential binding, and as well, on the influence of symbolic binding, is indicated as numbered arrows joining labelled regions (lower). (See text for further explanation.)

FIGURE 14 Inherent vs Readout Sequential Binding
Sequential read-out experimental findings can be summarized as almost 100% inherent binding, for spatial or structural similarity of patterns (middle column). However, sequential bindings are markedly influenced by read-out mode; only two-thirds of the binding is inherent (right column).

FIGURE 15 Control Experiment and Main Experiment
Experimental protocol for the control (upper panel) and main experiment (lower panel) to analyze the influence of symbolic binding. A major result is the influence of dissimilarity of labelling on the dissimilarity of the sequential pattern. Clearly, the spatial loci are the same, and thus, the structural similarity remains high. (Note that refreshment in the form of two additional looking stimuli are presented in both experiments (lower pair of grids in each of the panels).)

FIGURE 16: Top Anchor and Bottom Anchor Experiments
Experimental protocol to establish the range of values for Sp and Ss similarity. Top anchor (upper panel) shows high correlation when no refreshment is permitted (two absent grids, lower row, upper panel). Bottom anchor (lower panel) shows absence of structural and sequential similarity when a different pattern is presented with the same label.

FIGURE 17 Summary of Symbolic Experiment
Symbolic binding experimental findings can be summarized. Symbolic memory has important influence on sequential binding, producing an average 50% loss of coherence (compare 0.46 with 0.71, next to bottom row) when the labelling is changed. Since the same loci were re-presented with a different label, the structural binding, of course, remained the same (compare 0.76 with 0.80, bottom row).

FIGURE 18 Philosophical Approach to Perception
Five stages of the perceptual process (five columns) are illustrated with icons (upper), also showing BU and TD processes (curved arrows). See discussion in text regarding philosophical and physiological sources of this schema.

FIGURE 19 Micro-Cortical Processes
Six levels or layers of the visual cortex, known from neuroanatomy, are suggested as the iconic matching region, where TD input to the visual cortex, at layers I, II, and III, interact with BU input going to layers IV and V, from retina via geniculate and optic tracts (modified from Pitts and McCulloch, 1947).

FIGURE 20 Cortical Representation of Perceptual Processes
Although only the microanatomy of the visual cortex is known well enough to support a graph theoretical model, yet we have suggested a variety of such graphs for structural, sequential, and symbolic binding, with loci as per labels in the modular cortex. Geometrical binding is used in our modeling schema, for syntactical interaction between foveal ROIs and peripheral segments. Different forms of the graphs do not represent any knowledge about feasible or understood properties of the brain, but rather stress our ignorance.

FIGURE 21 Feedback Model for Supervisory Telerobotic Control
Control systems diagram for telerobotic scheme showing higher level control, with supervisor and path planner. The serializer provides input to the basic feedback control loop, with camera and image processing algorithms, IP Alg, monitoring actual position, Ym, of robot (right inset).

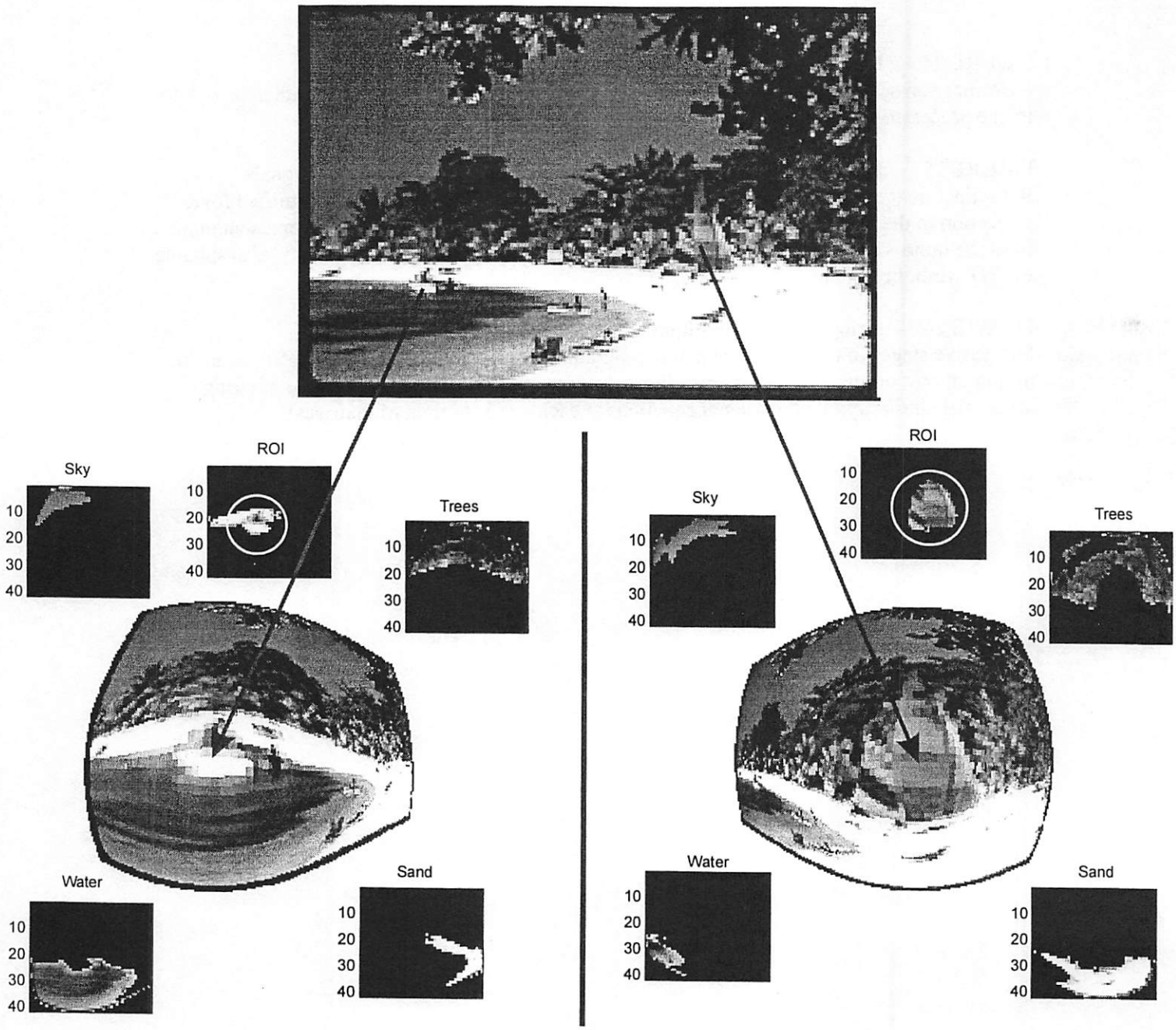
FIGURE 22 TD Scanpath Scheme for Robotic Vision
Four image processing steps showing robot vehicles with VEs (upper left) and model ROI-predicted locations (white squares, upper right). Note scanpath sequence for computer image processing of ROIs (white arrows, lower left) yielding centroid-calculated loci (white crosses, lower right).

FIGURE 23 Advantages of TD Control of BU Image Processing
Pixel intensity diagram forming a 3D representation of the video image (upper). Note hilltops representing VEs. By predicting ROI loci using TD model (rectangular boxes, lower), it is possible to do adaptive thresholding only in a small localized region, and thus, achieve important signal-to-noise ratio improvements. Clearly, foveal fixation in normal human vision achieves the same functionality.

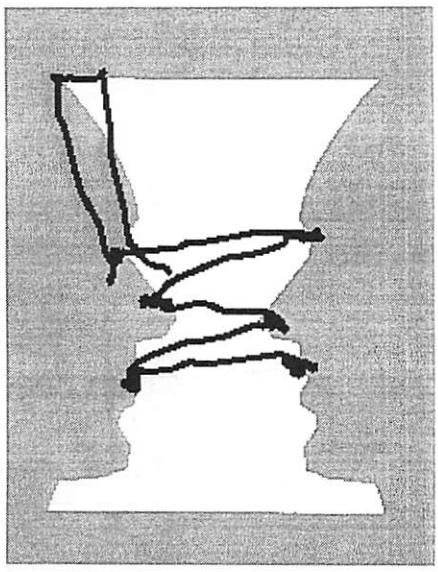
FIGURE 24 Block Diagram Explicating Telerobotic Vision Scheme
Flow diagram schema to aid in understanding steps of our telerobotic TD scanpath approach to image processing and to supervisory control.

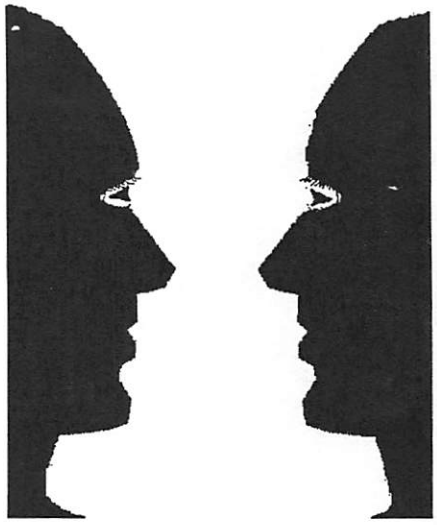
FIGURE 25 Iconic Matching in Visual Cortex TD Representation and BU Signals
BU retinal image is shifted with each EM fixation to provide a centered and magnified foveal projection in the visual cortex. These may be matched by predicted TD iconic representations from the mind's eye image. Continuous periphery is shown broken into segments, also suitable for TD symbolic coding.

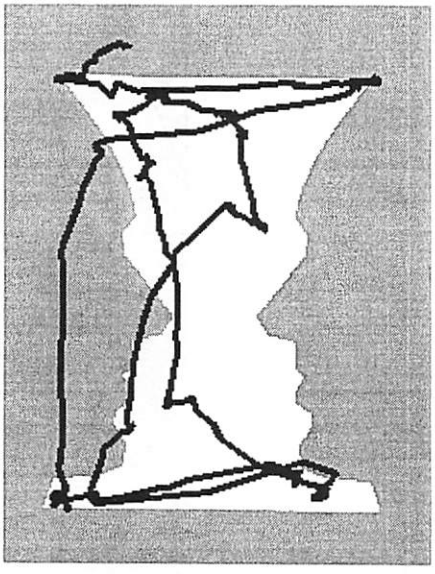
FIGURE 26 String Editing Algorithm
Successive stages of the discrete dynamic programming algorithm (matrices at left) document minimum cost optimization of string editing distance, and thus, accurate measure of string sequential similarity; computational equations at each stage (to right of matrices).

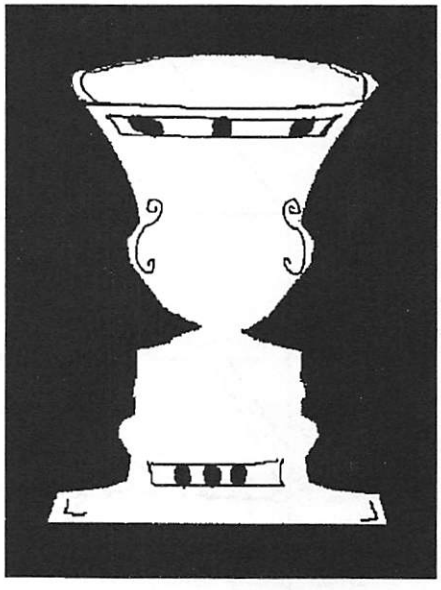


F01

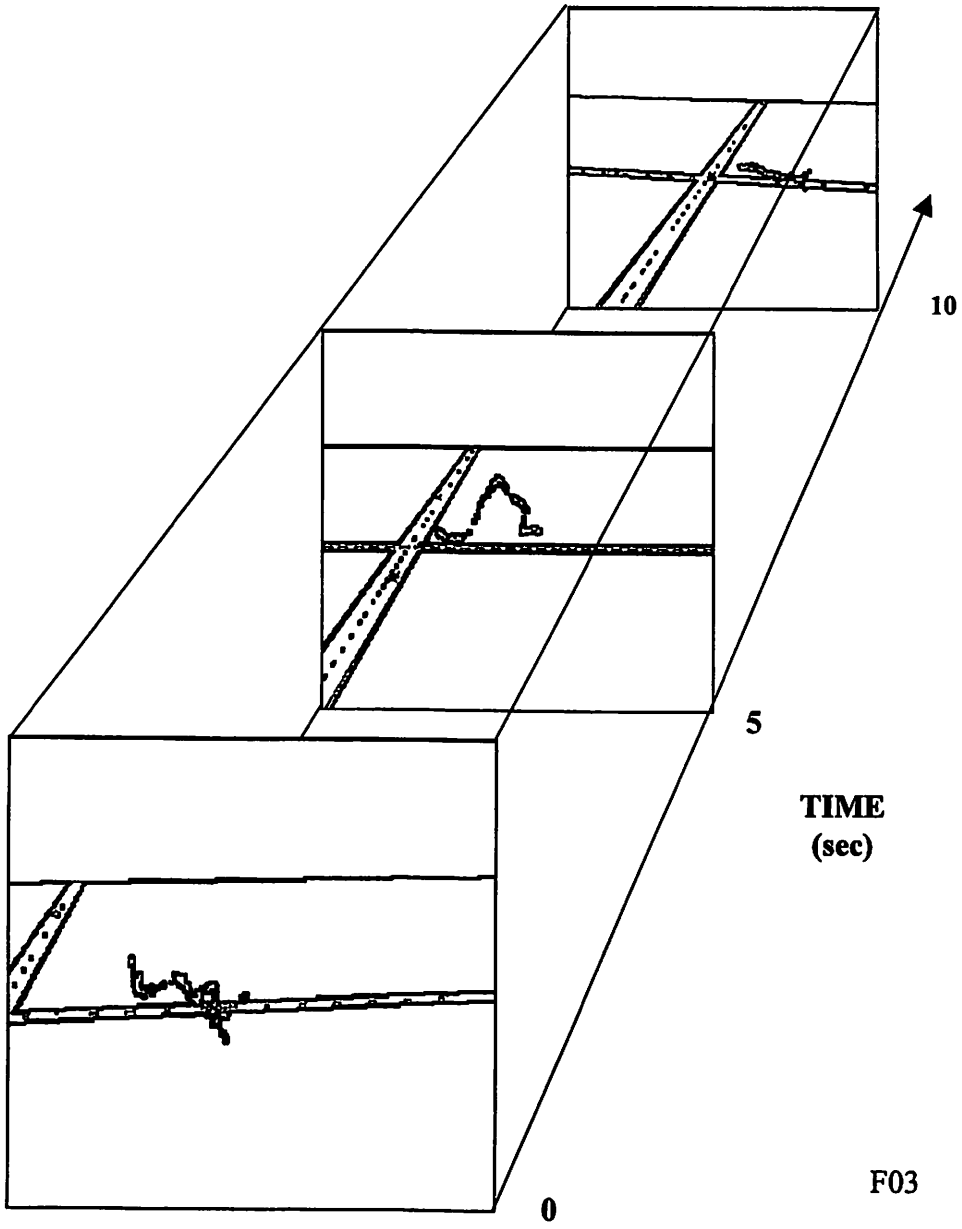




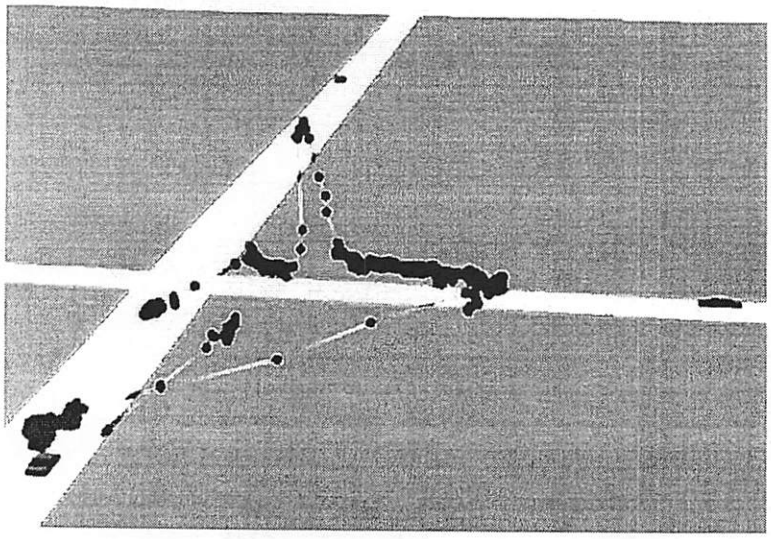


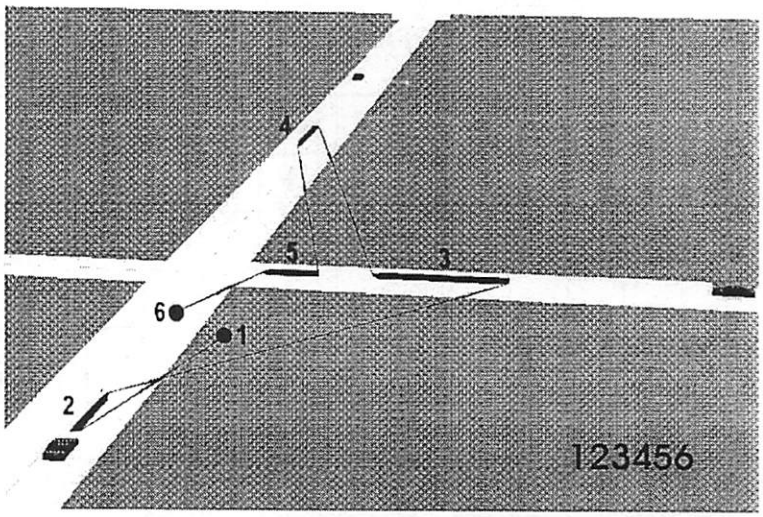


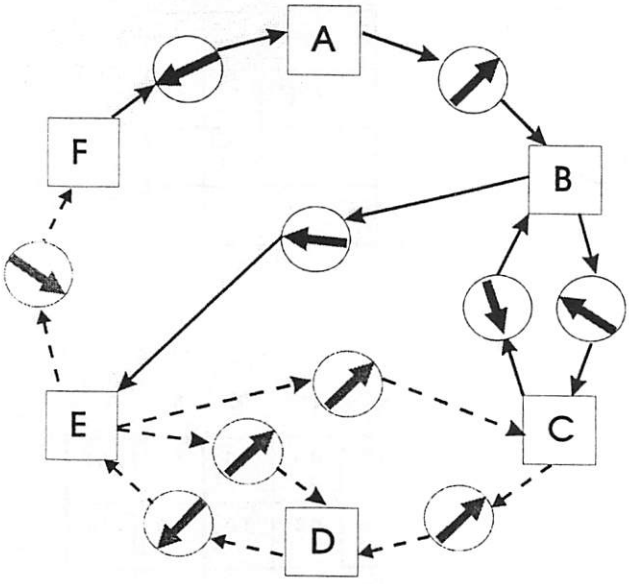
F02

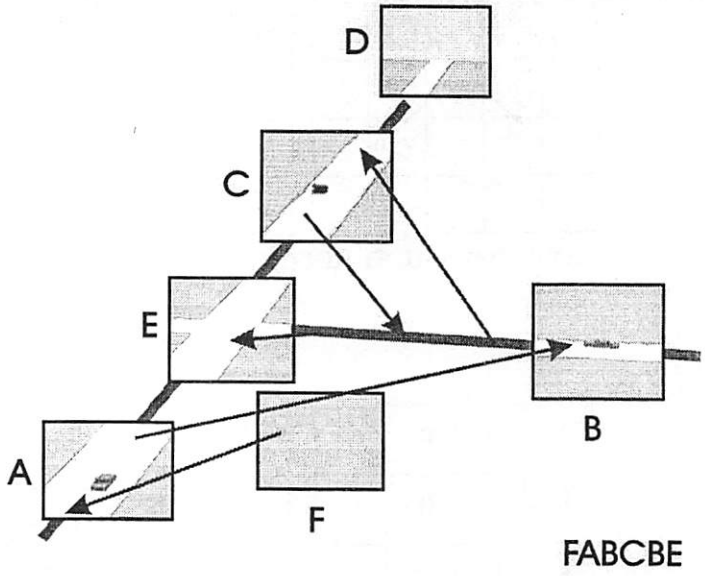


F03



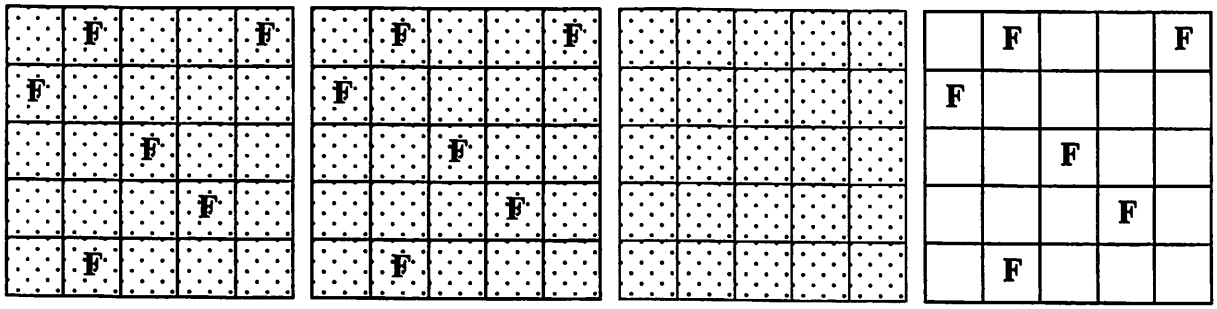




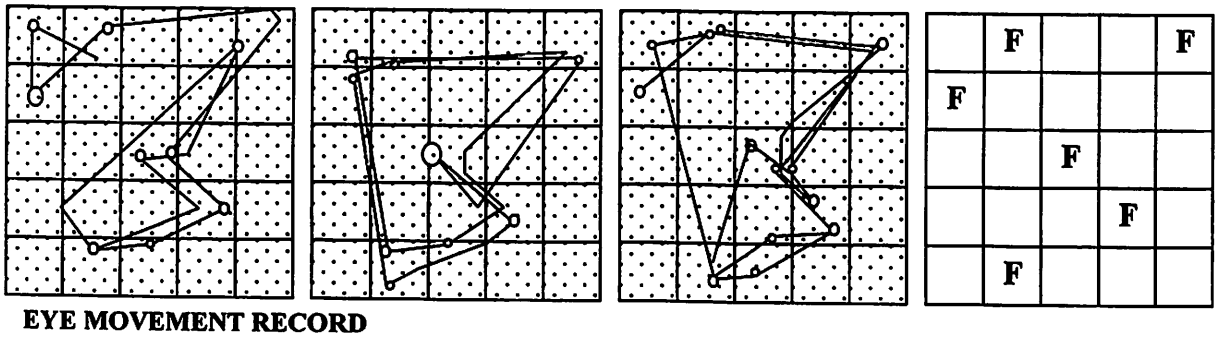


F04

IMAGERY EXPERIMENT

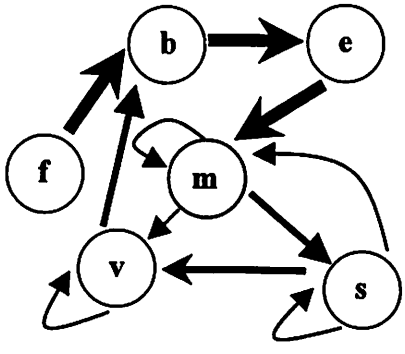


EXPERIMENT PROTOCOL



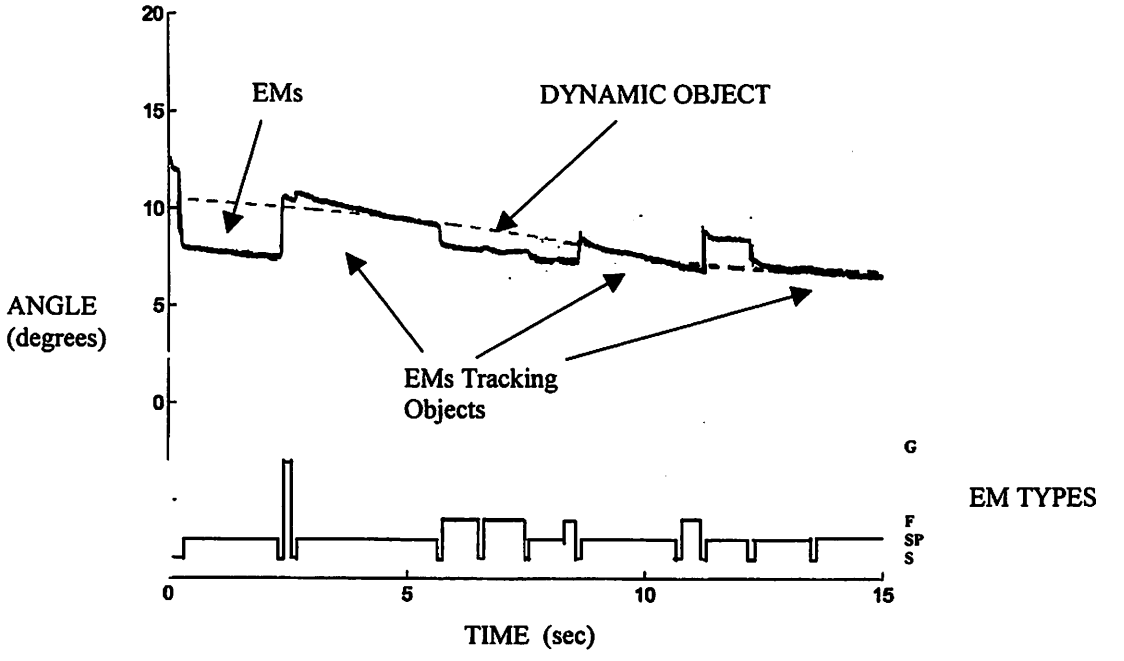
EYE MOVEMENT RECORD

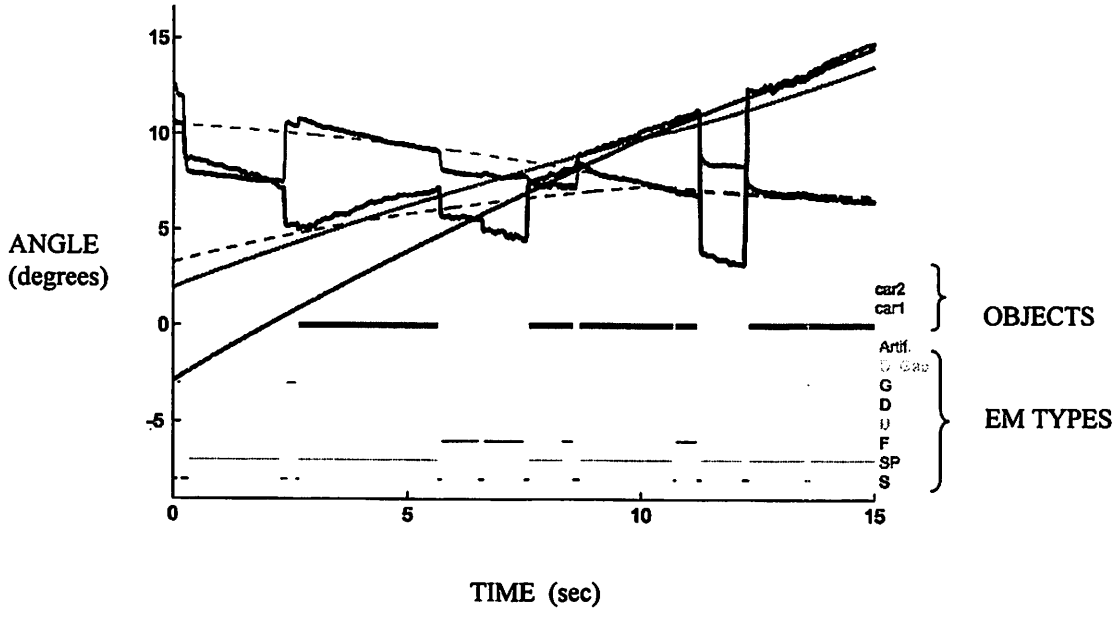
a	b	c	d	e
f	g	h	i	j
k	l	m	n	o
p	q	r	s	t
u	v	w	x	y



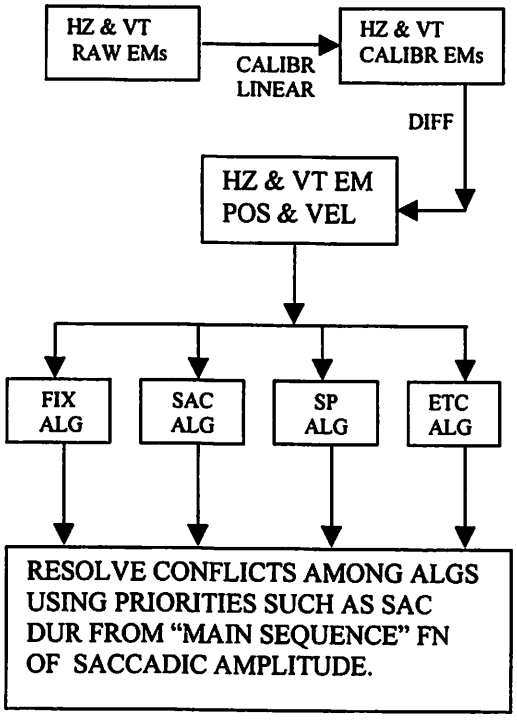
	b	e	f	m	s	v
b	0.0	1.0	0.0	0.0	0.0	0.0
e	0.0	0.0	0.0	1.0	0.0	0.0
f	1.0	0.0	0.0	0.0	0.0	0.0
m	0.0	0.0	0.0	0.2	0.6	0.2
s	0.0	0.0	0.0	0.3	0.3	0.4
v	0.5	0.0	0.0	0.0	0.0	0.5

ANALYSIS PROCEDURE


20
15
10
5
0
EMs
DYNAMIC OBJECT
ANGLE
(degrees)
EMs Tracking
Objects
G
F
SP
S
EM TYPES
0
5
10
15
TIME (sec)

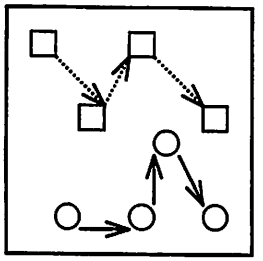

15
10
5
0
-5
ANGLE
(degrees)
car2
car1
OBJECTS
Artif.
G
D
F
SP
S
EM TYPES
0
5
10
15
TIME (sec)

F06

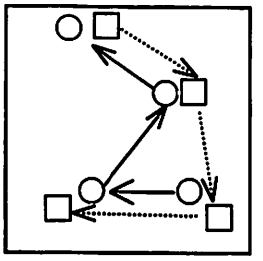

HZ & VT
RAW EMs
CALIBR
LINEAR
HZ & VT
CALIBR EMs
DIFF
HZ & VT EM
POS & VEL
FIX
ALG
SAC
ALG
SP
ALG
ETC
ALG
RESOLVE CONFLICTS AMONG ALGS
USING PRIORITIES SUCH AS SAC
DUR FROM "MAIN SEQUENCE" FN
OF SACCADIC AMPLITUDE.

F07

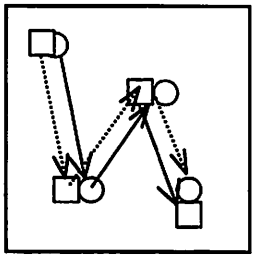
INDICES



Sp = 0; Ss = 0



Sp = 1; Ss = 0



Sp = 1; Ss = 1

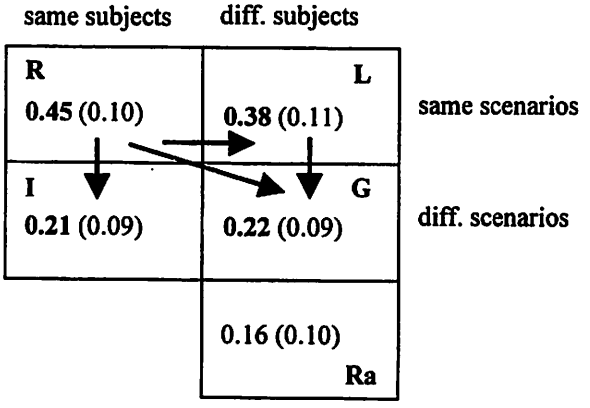
Y-MATRICES

Sp	Subject 1		Subject 2	
	Picture 1	Pict 2	Picture 1	Pict 2
S1 P1	0.65 R	0.38 I	0.54 L	0.18 G
S1 P2		0.60 R	0.31 G	0.47 L
S2 P1			0.69 R	0.33 I
S2 P2				0.58 R

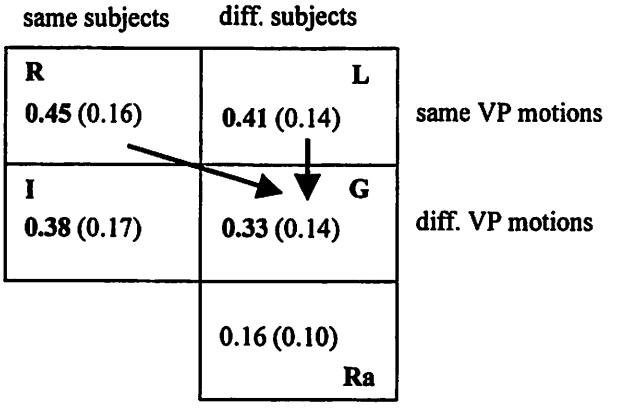
Ss	Subject 1		Subject 2	
	Picture 1	Pict 2	Picture 1	Pict 2
S1 P1	0.40 R	0.24 I	0.31 L	0.08 G
S1 P2		0.39 R	0.13 G	0.19 L
S2 P1			0.43 R	0.21 I
S2 P2				0.24 R

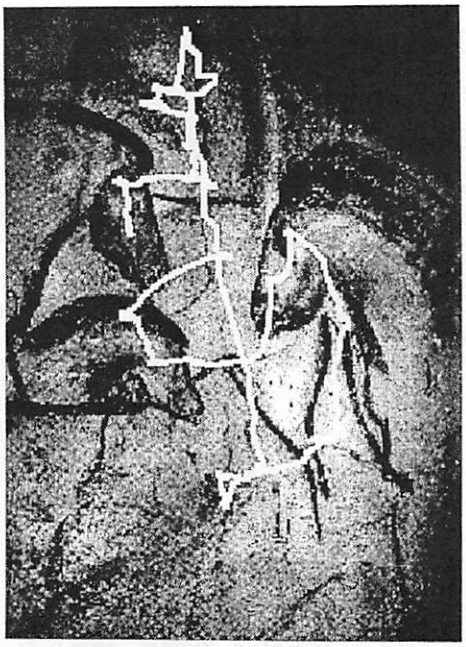
Ss PARSING DIAGRAMS

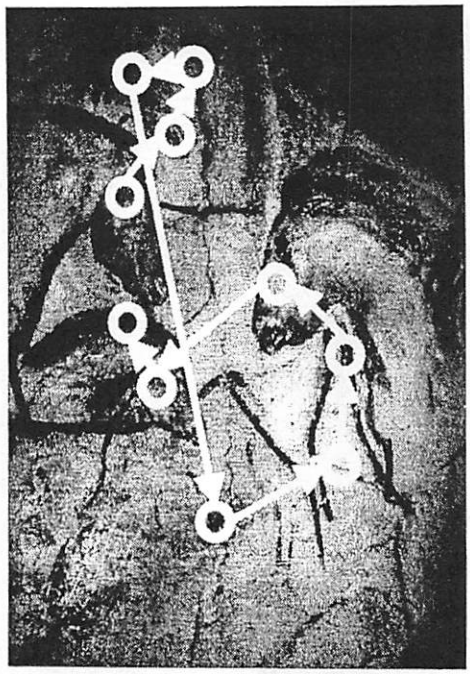
DIFFERENT SCENARIOS

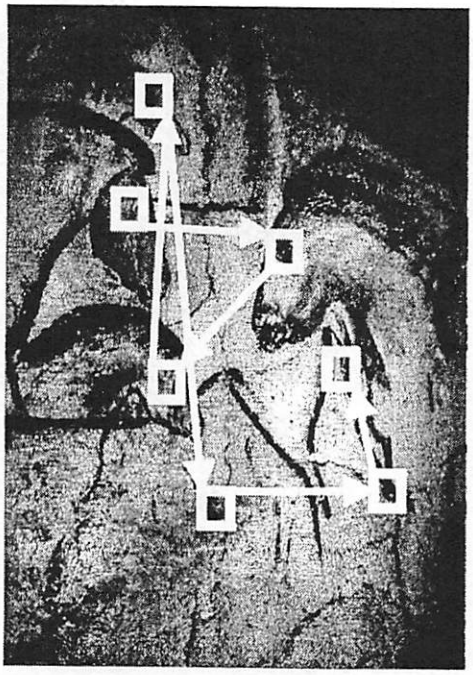


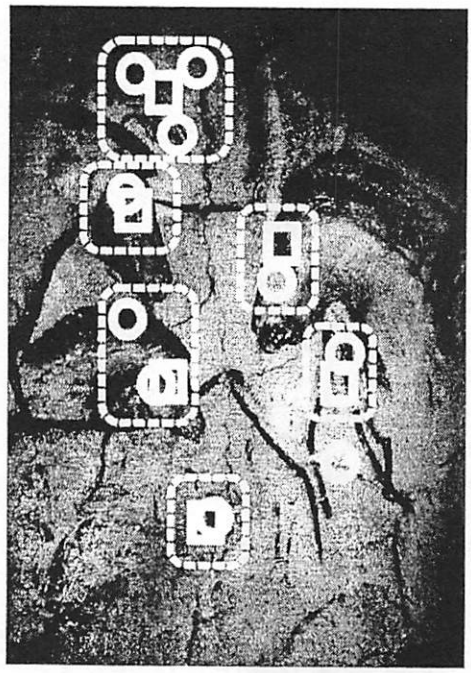
THREE VIEWPOINT MOTIONS



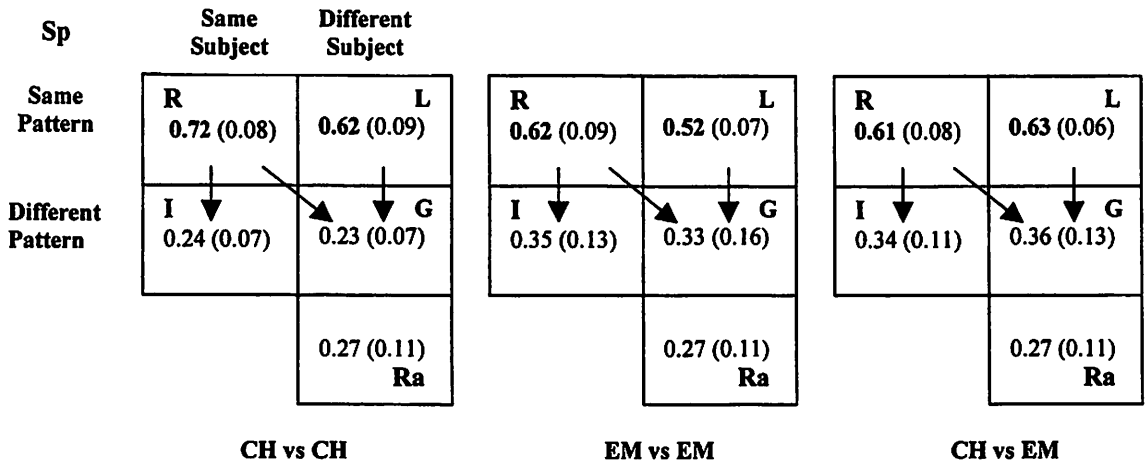


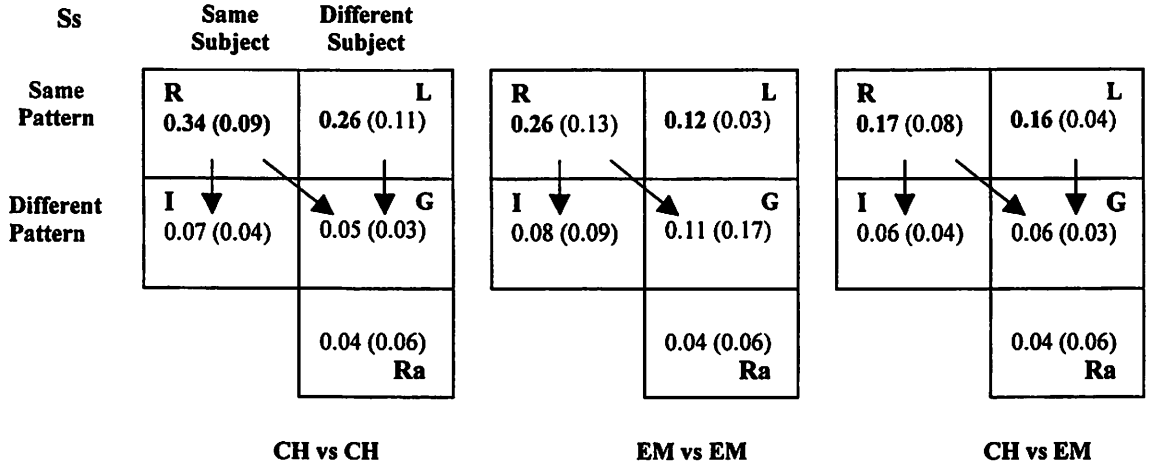


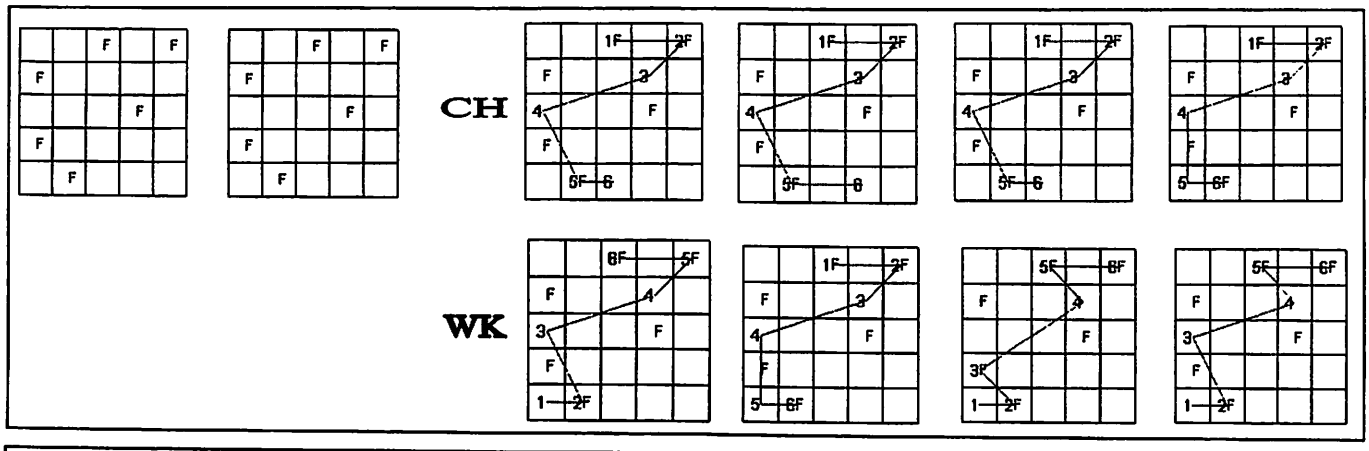


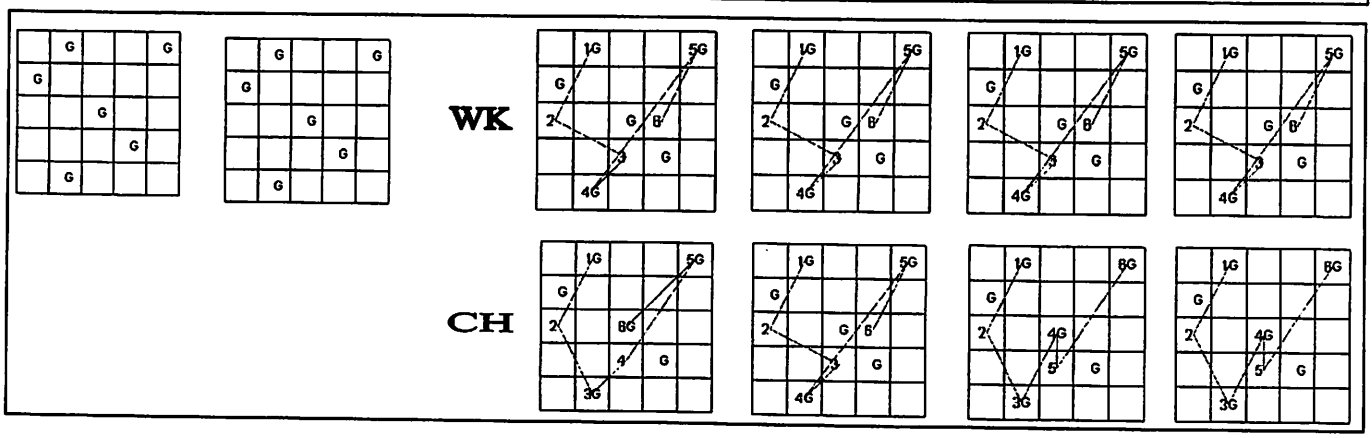


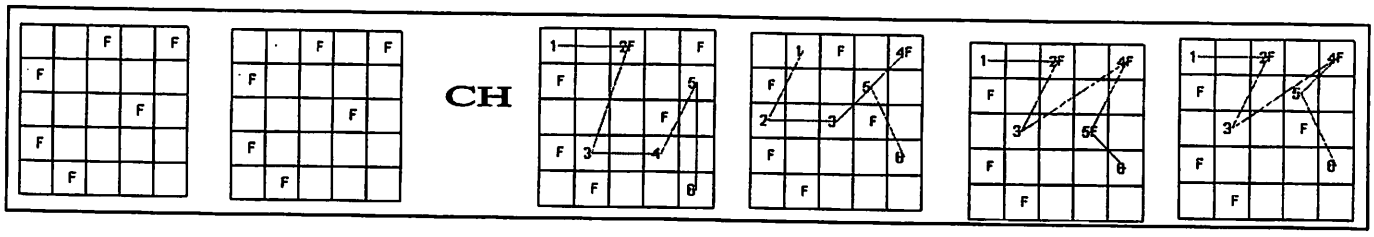
F09

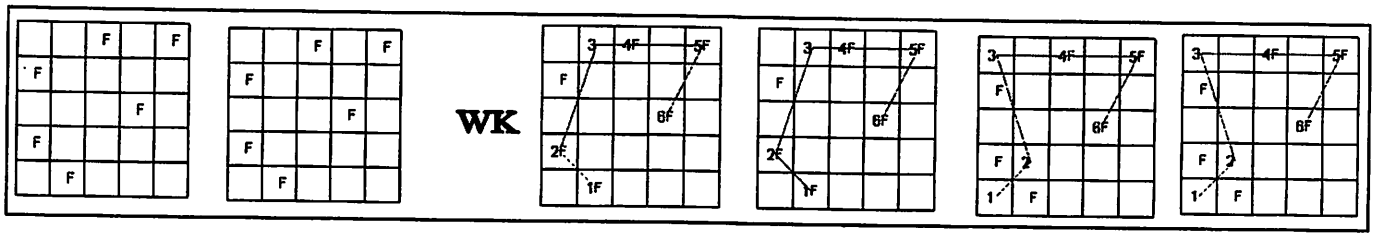

Sp
Same Subject
Different Subject
Same Pattern
Different Pattern
R
0.72 (0.08)
L
0.62 (0.09)
I
0.24 (0.07)
G
0.23 (0.07)
0.27 (0.11)
Ra
CH vs CH
R
0.62 (0.09)
L
0.52 (0.07)
I
0.35 (0.13)
G
0.33 (0.16)
0.27 (0.11)
Ra
EM vs EM
R
0.61 (0.08)
L
0.63 (0.06)
I
0.34 (0.11)
G
0.36 (0.13)
0.27 (0.11)
Ra
CH vs EM


Ss
Same Subject
Different Subject
Same Pattern
Different Pattern
R
0.34 (0.09)
L
0.26 (0.11)
I
0.07 (0.04)
G
0.05 (0.03)
0.04 (0.06)
Ra
CH vs CH
R
0.26 (0.13)
L
0.12 (0.03)
I
0.08 (0.09)
G
0.11 (0.17)
0.04 (0.06)
Ra
EM vs EM
R
0.17 (0.08)
L
0.16 (0.04)
I
0.06 (0.04)
G
0.06 (0.03)
0.04 (0.06)
Ra
CH vs EM

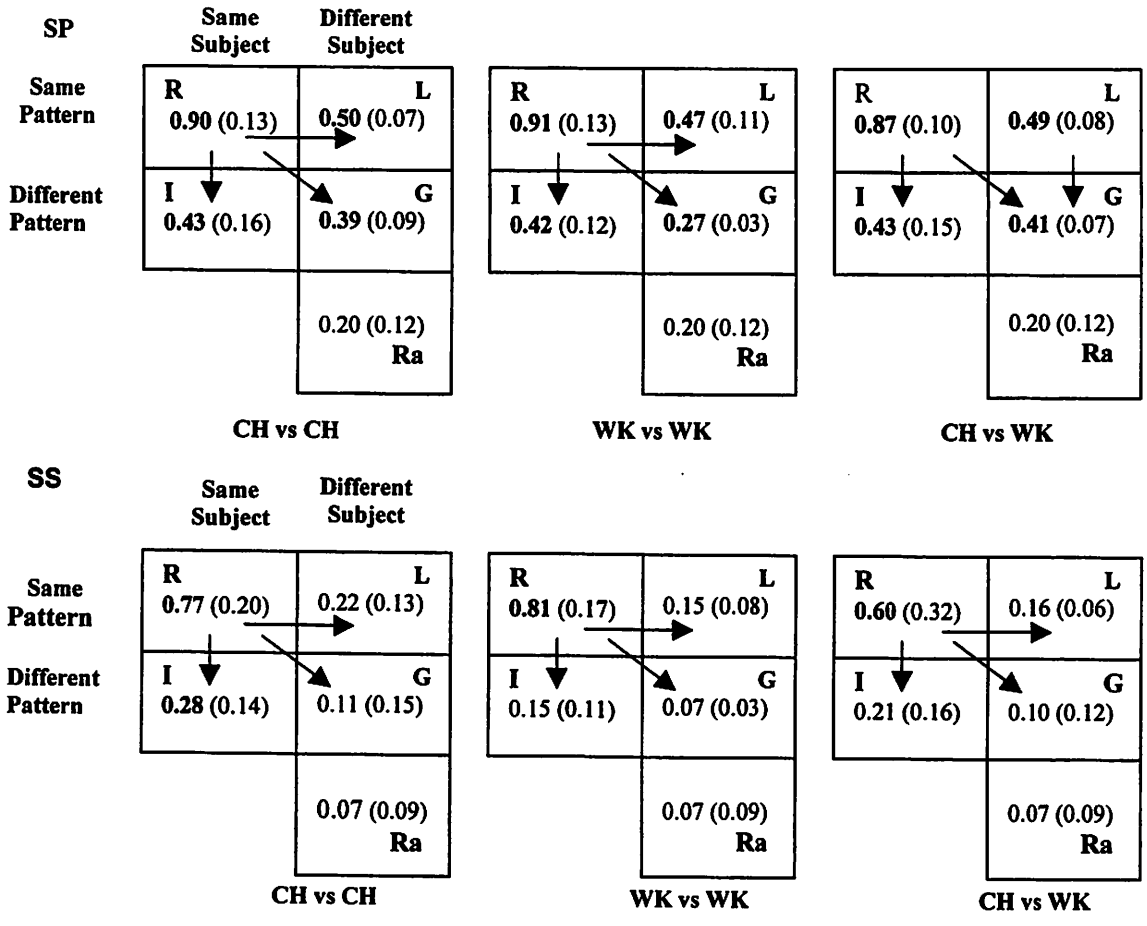

CH
WK


WK
CH


CH

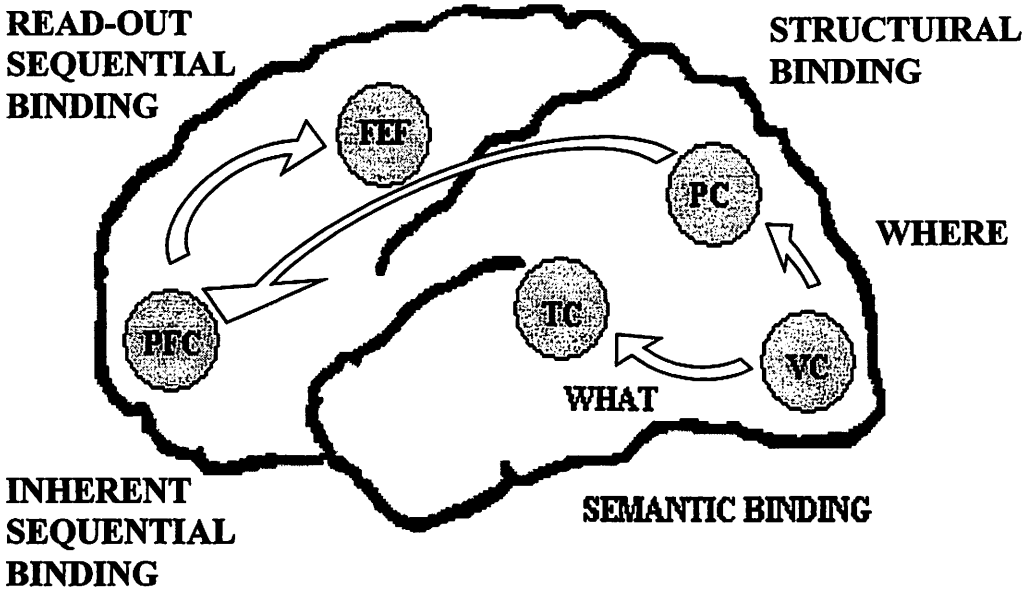

WK

PARSING DIAGRAMS for CHOICE vs WALKING

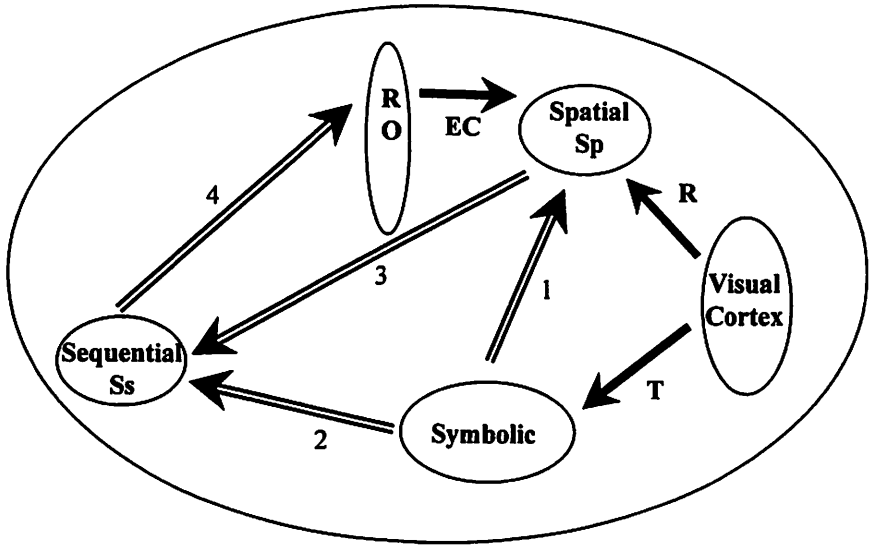


e.g., (*e.g.*,

MODULAR CORTEX



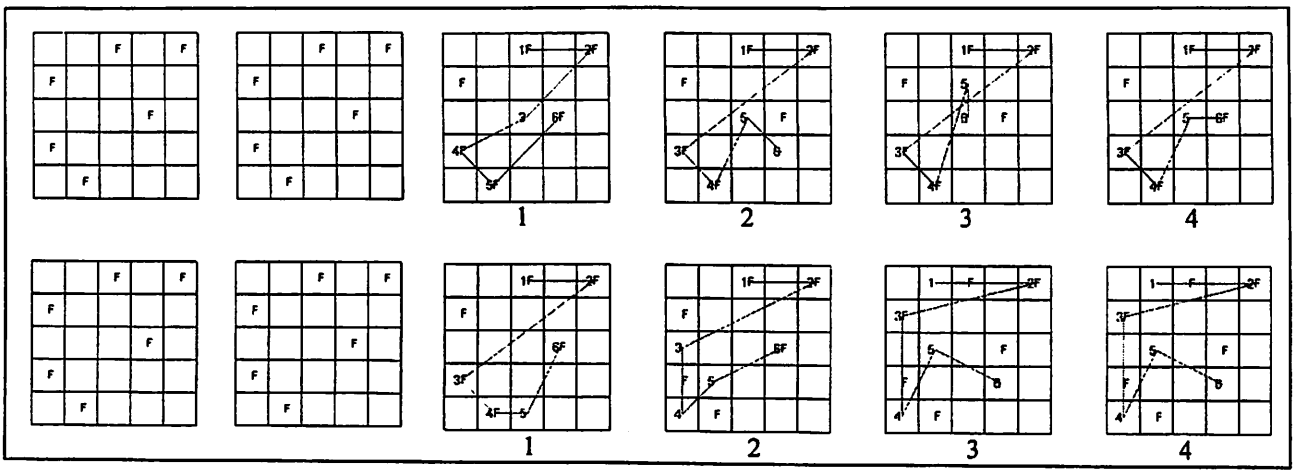
CORTICAL CONNECTIVITY



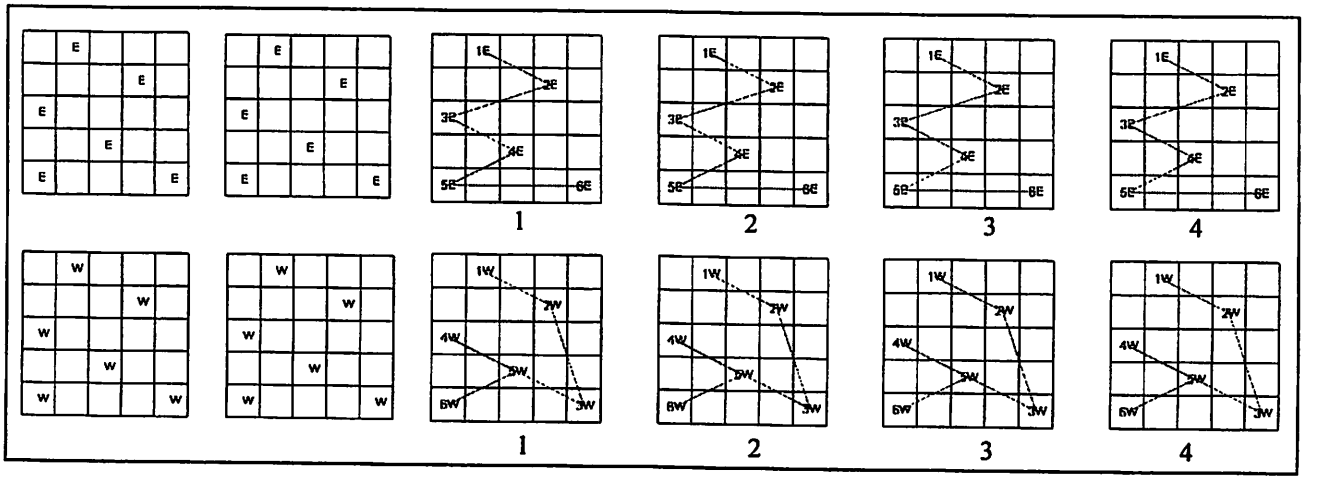
1	Symbolic to Spatial	R	Visual Cortex to Spatial
2	Symbolic to Sequential	T	Visual Cortex to Symbolic
3	Spatial to Sequential		
4	Sequential to Read-Out (RO)	EC	Efferent Copy to Spatial

Expermental type	Sp	Diff & N %	Binding	Ss	Diff & N %	Binding
CH X CH (pictures)	0.72			0.34		
EM X EM (pictures)	0.62			0.26		
average	0.67			0.30		
		0.06 = 15 %	Read-Out		0.13 = 50 %	Read-Out
CH X EM (pictures)	0.61			0.17		
		0.34 = 85 %	Inherent		0.13 = 50 %	Inherent
Random	0.27			0.04		
CH X CH (imagery)	0.90			0.77		
W X W (imagery)	0.91			0.81		
average	0.91			0.79		
		0.04 = 6 %	Read-Out		0.19 = 26 %	Read-Out
CH X W (imagery)	0.87			0.60		
		0.67 = 94 %	Inherent		0.53 = 74 %	Inherent
Random	0.20			0.07		

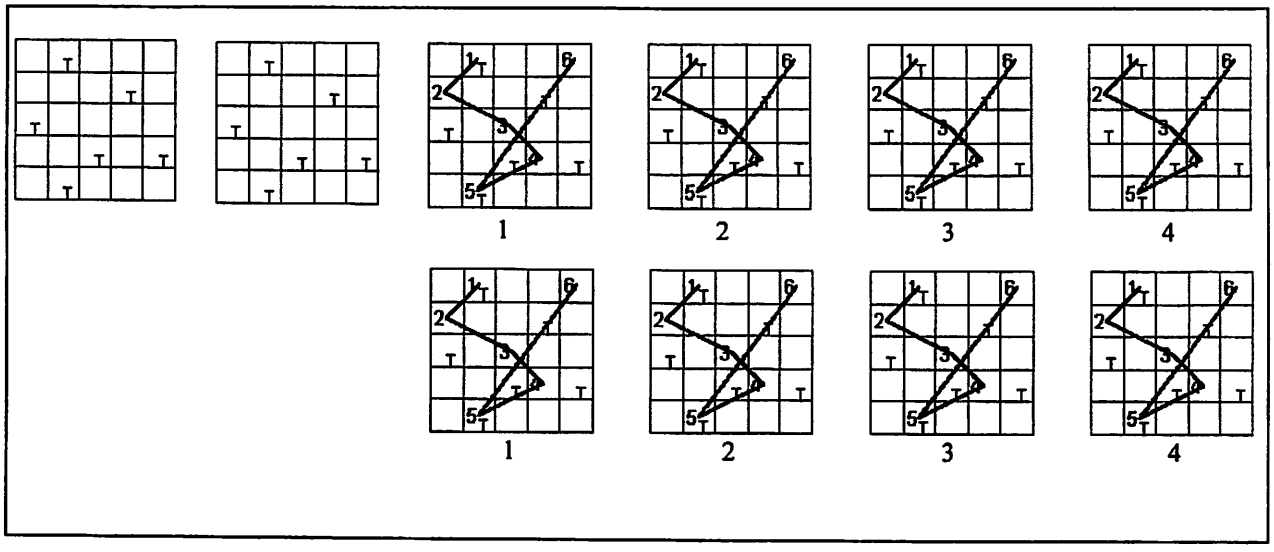
CONTROL EXPERIMENT: SAME PATTERN, SAME LABEL WITH REFRESHMENT



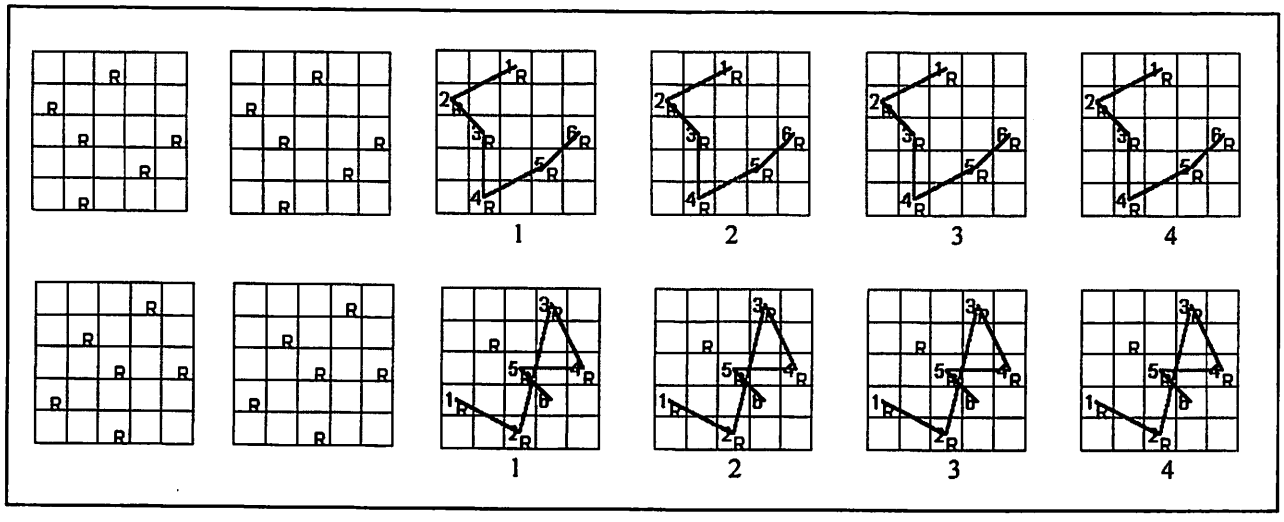
MAIN EXPERIMENT: SAME PATTERN, DIFFERENT LABEL WITH REFRESHMENT



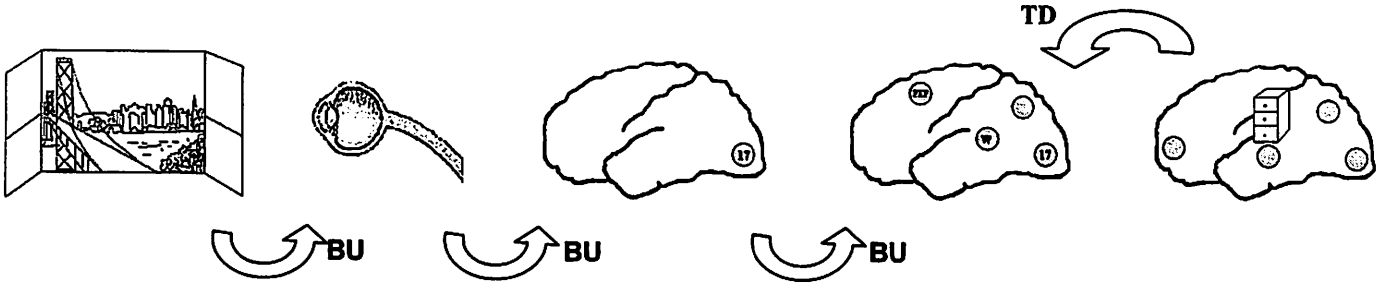
TOP ANCHOR: SAME PATTERN, NO NEW LABEL



BOTTOM ANCHOR: DIFFERENT PATTERN, SAME LABEL



	TOP ANCHOR	CONTROL EXPERIMENT	MAIN EXPERIMENT	BOTTOM ANCHOR
SECOND STIMULUS	None	Same	Same	Different
SECOND LABEL	None	Same	Different	Same
OTHER INFORMATION	Interrupt	Interrupt	New Label	Diff 2nd stimulus
EFFECT		Allows re-initialization	Encourages re-initialization	Spatial dominates over Semantic
SECOND RESPONSE	Same	Same Some differences	Different Some Simularities	Different
SS (COHERENT)	0.86	0.71	0.46	0.07
SP (COHERENT)	0.94	0.80	0.76	0.36

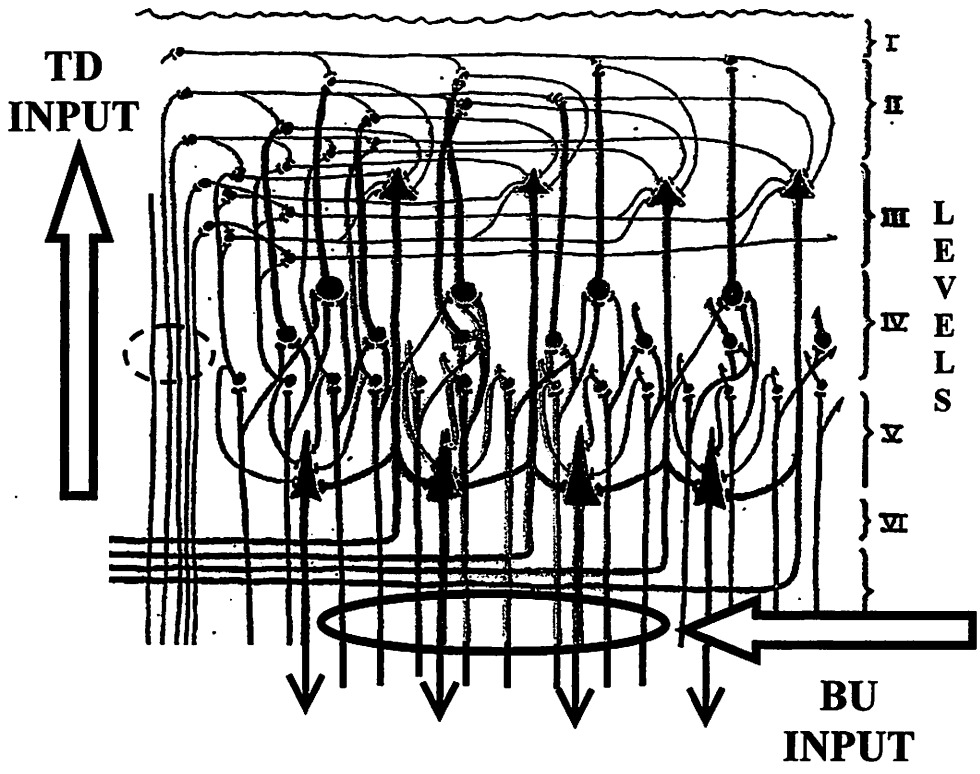


APPEARANCE	SENSATION	SENSORY ORGANIZATION	PERCEPTION	REPRESENTATION
Chaos Class of appearances	Impression Doctrine of specific nerve 'endings'	CNS processing for velocity: $v = dx/dt$ Pre-attentive psychophysics	A planned forceful, determined activity Perception Intuition	Noumenon Symbolic Ideal Notion
Bottom up stuff (not 'things'!)	Bottom up physiology without space and time	Bottom up neurophysiology with space/time computation	Top down active looking scanpath as operational phase of perception per se	Top down cognitive model

Where does TD meet BU?

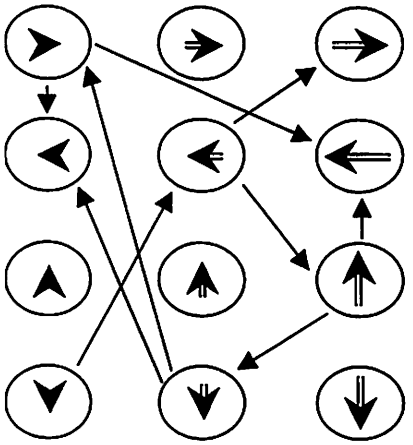
Levels I, II, and III meet level IV in the retinotopic visual cortex.

VISUAL CORTEX

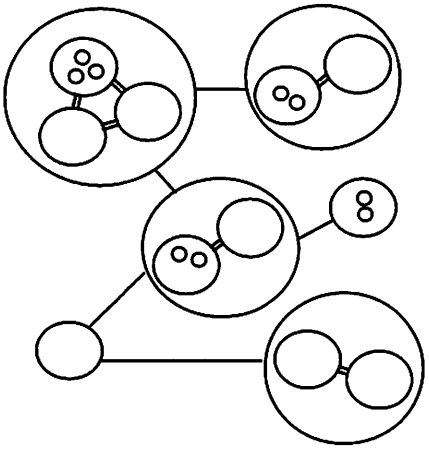


F19

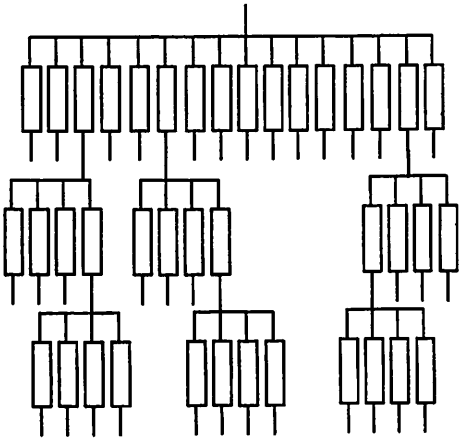
SEQUENTIAL BINDING



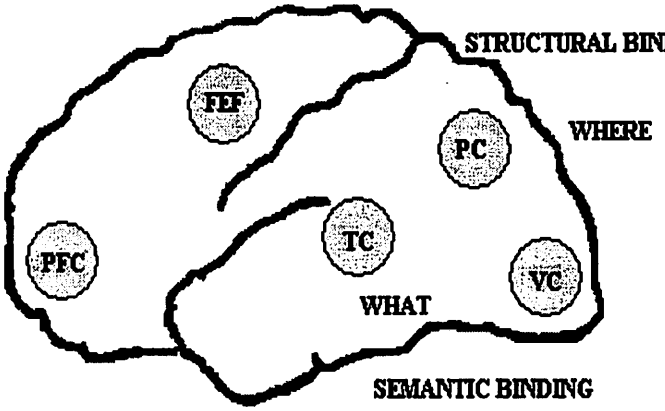
GEOMETRIC BINDING



STRUCTURAL BINDING

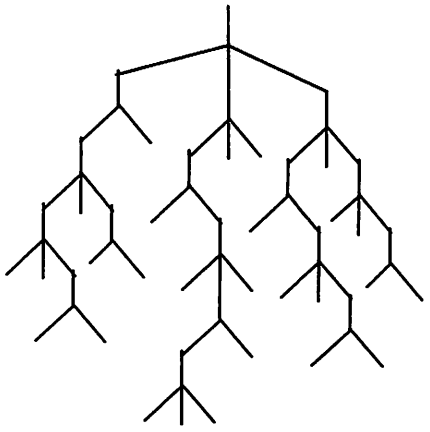


READ-OUT SEQUENTIAL BINDING

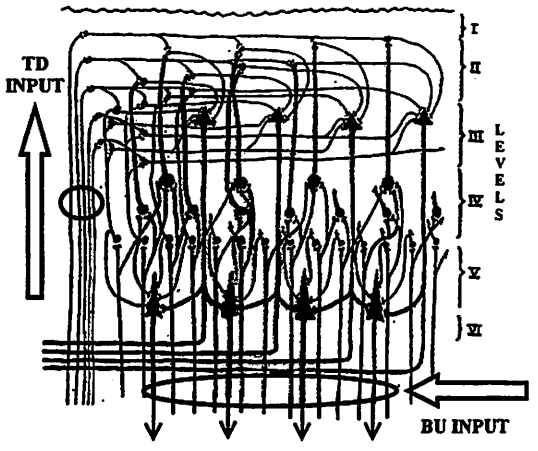


INHERENT SEQUENTIAL BINDING

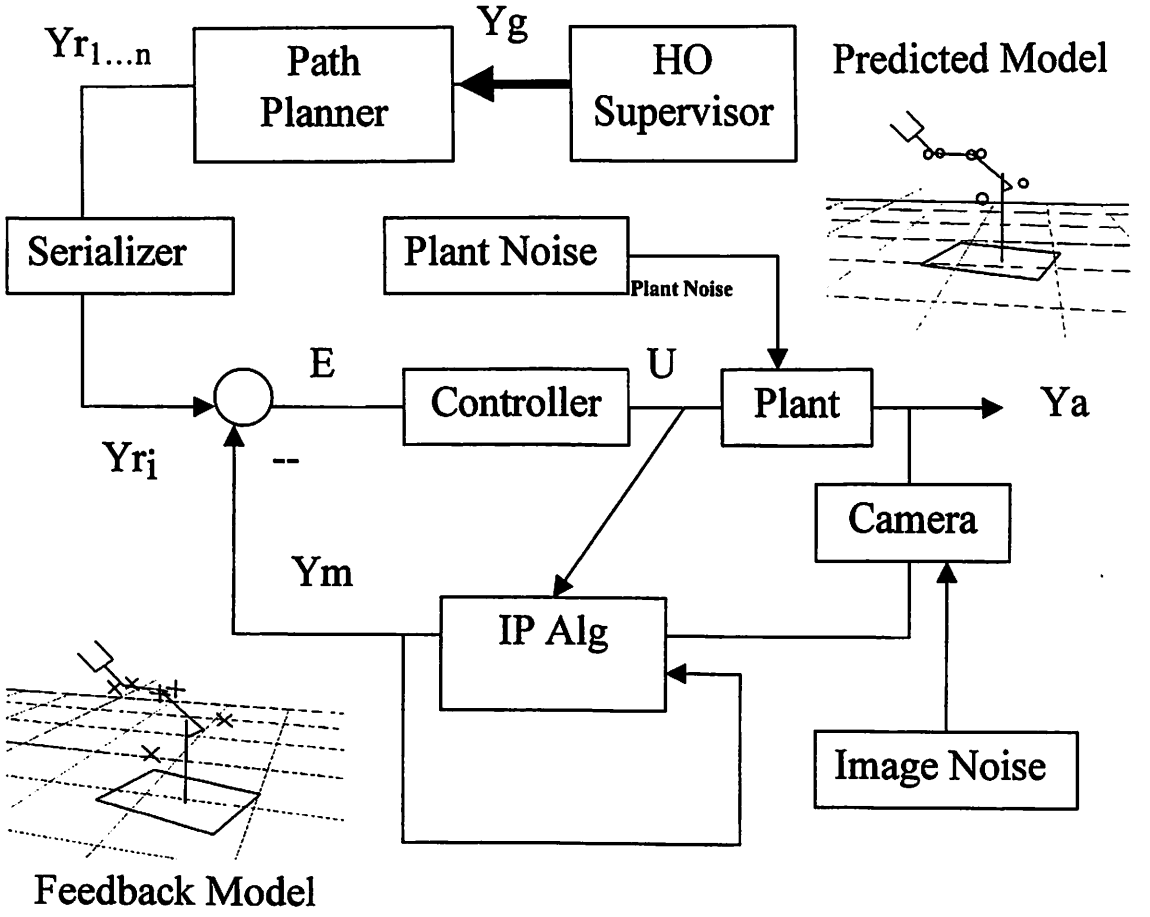
SYMBOLIC BINDING



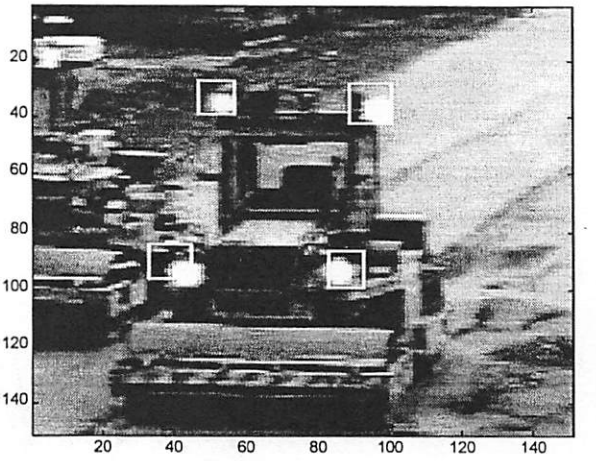
VISUAL CORTEX

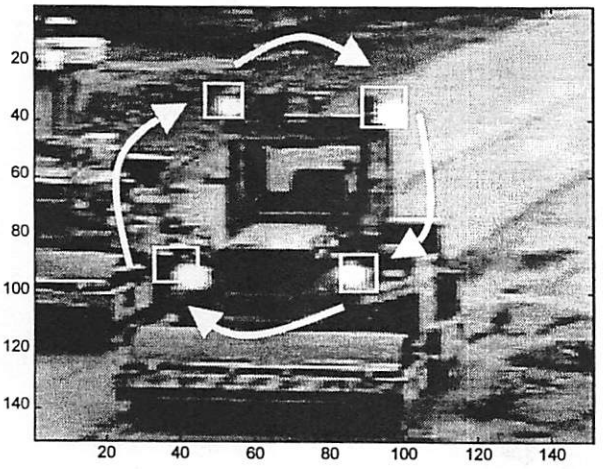


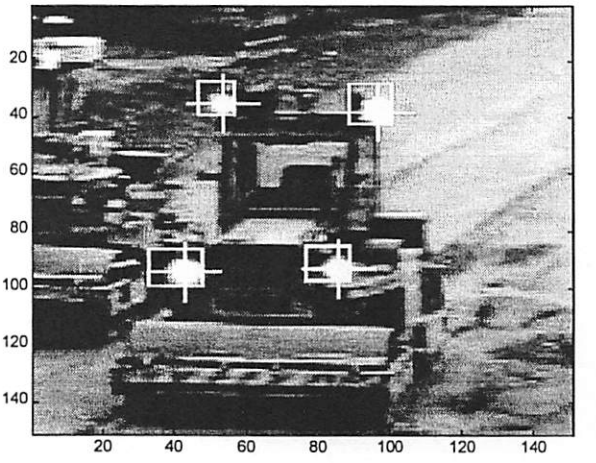
F20


Yr1...n
Path Planner
Yg
HO Supervisor
Predicted Model
Serializer
Plant Noise
Plant Noise
E
U
Controller
Plant
Ya
Yri
--
Camera
Ym
IP Alg
Image Noise
Feedback Model

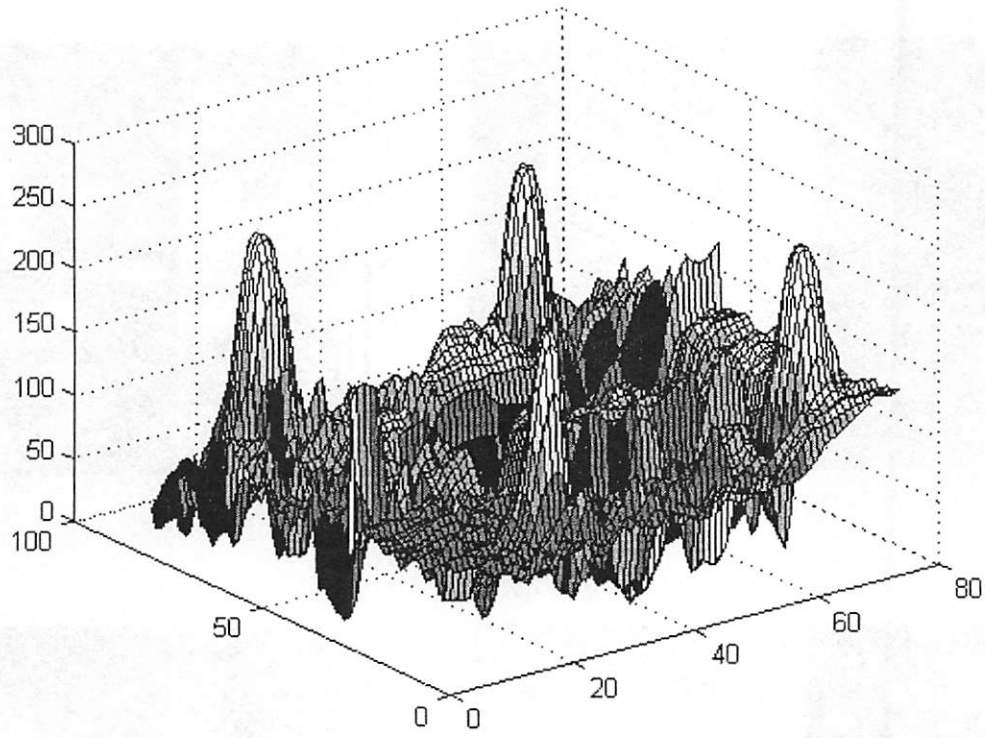
F21

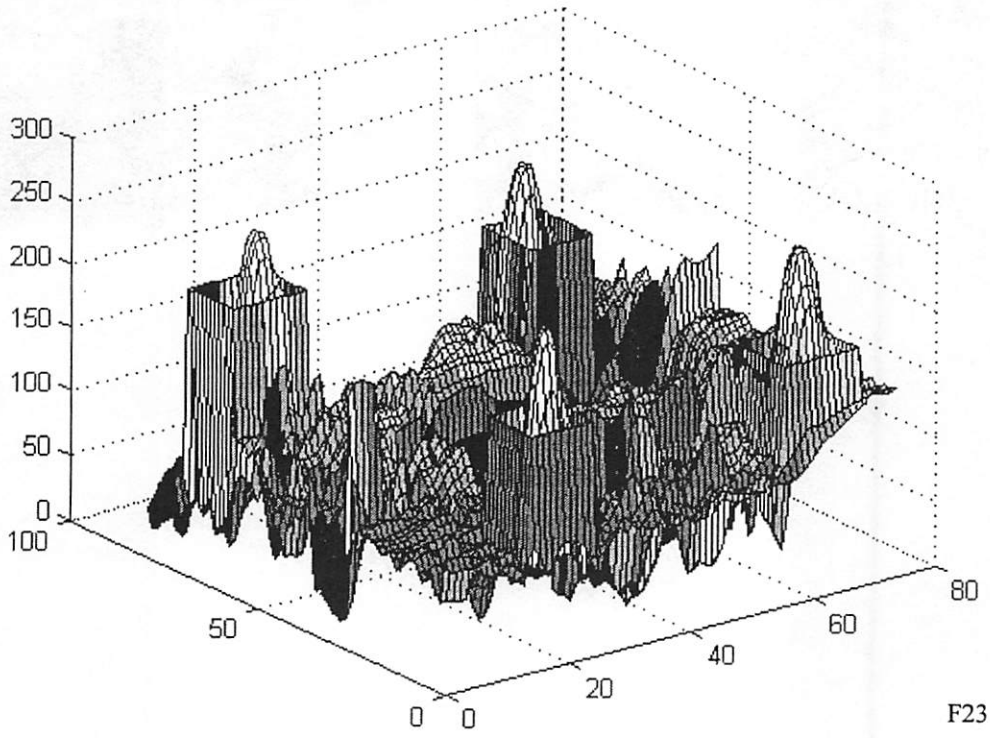

20
40
60
80
100
120
140
20
40
60
80
100
120
140


20
40
60
80
100
120
140
20
40
60
80
100
120
140

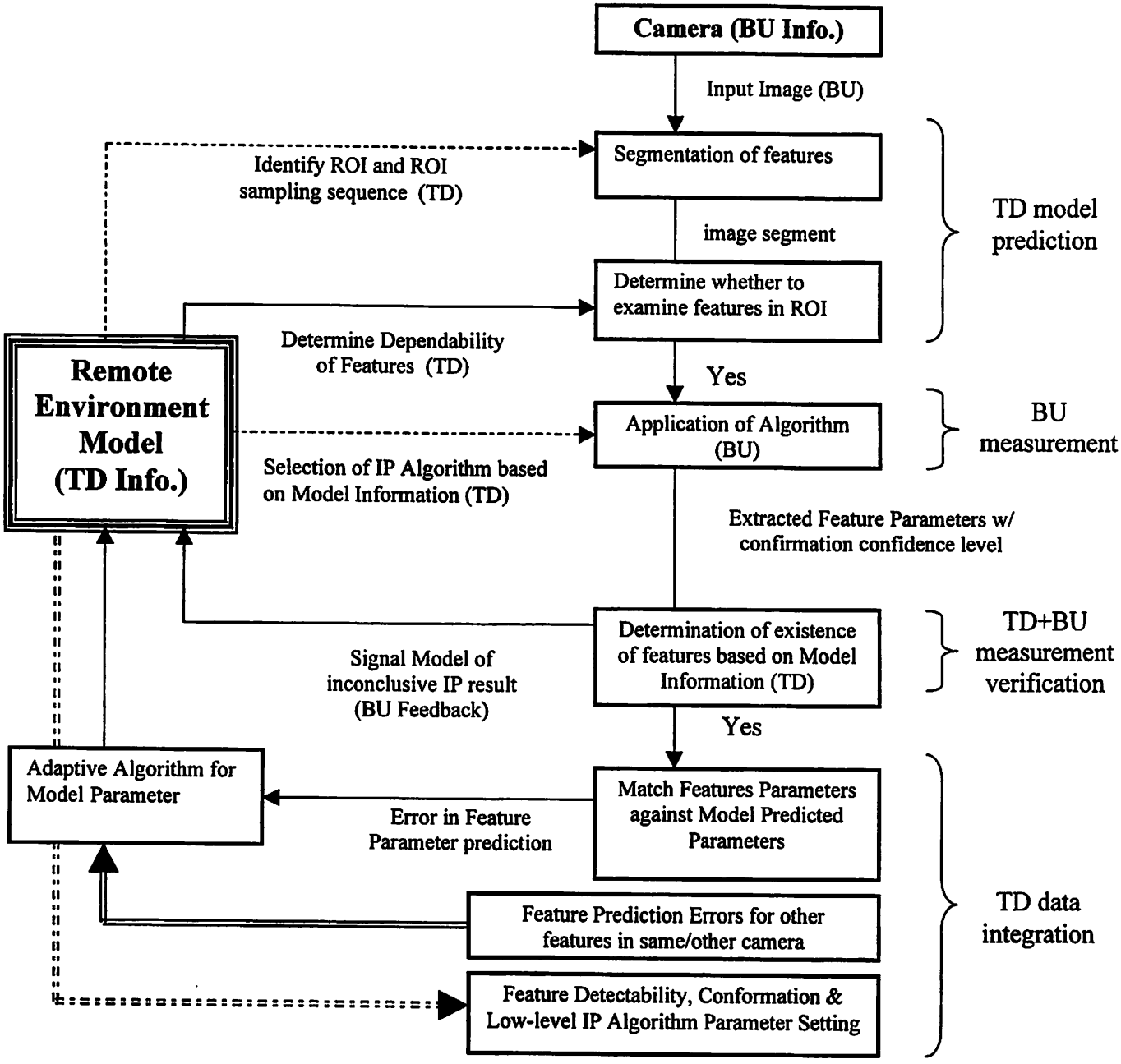

20
40
60
80
100
120
140
20
40
60
80
100
120
140

F22

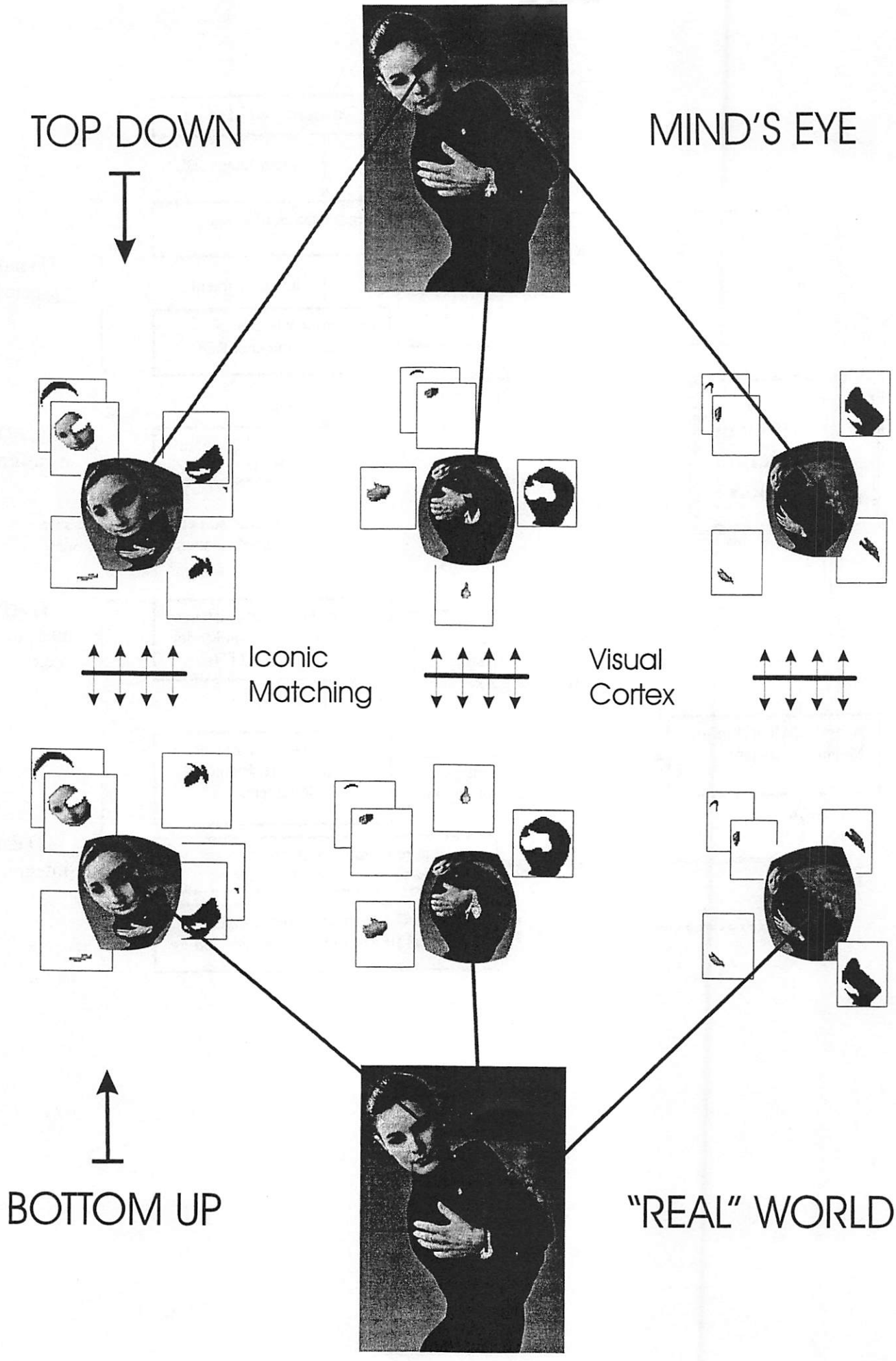

300
250
200
150
100
50
0
100
50
0
0
20
40
60
80


300
250
200
150
100
50
0
100
50
0
0
20
40
60
80

F23

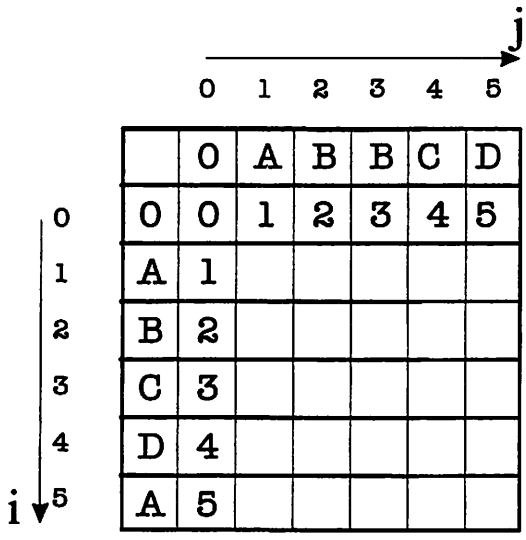

Camera (BU Info.)
Input Image (BU)
Segmentation of features
Identify ROI and ROI sampling sequence (TD)
image segment
Determine whether to examine features in ROI
TD model prediction
Remote Environment Model (TD Info.)
Determine Dependability of Features (TD)
Yes
Application of Algorithm (BU)
BU measurement
Selection of IP Algorithm based on Model Information (TD)
Extracted Feature Parameters w/ confirmation confidence level
Determination of existence of features based on Model Information (TD)
TD+BU measurement verification
Signal Model of inconclusive IP result (BU Feedback)
Yes
Adaptive Algorithm for Model Parameter
Match Features Parameters against Model Predicted Parameters
Error in Feature Parameter prediction
Feature Prediction Errors for other features in same/other camera
TD data integration
Feature Detectability, Conformation & Low-level IP Algorithm Parameter Setting

F24


TOP DOWN
MIND'S EYE
Iconic
Matching
Visual
Cortex
BOTTOM UP
"REAL" WORLD

D-MATRIX FOR STRING COMPARISON.

$S1_j$ = ABBCD $S2_i$ = ABCDA

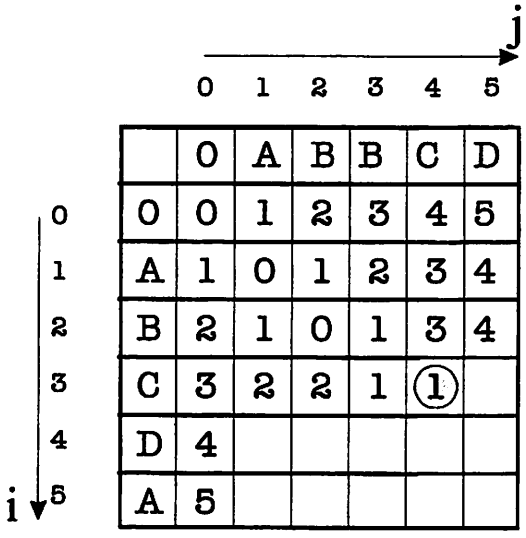


	0	A	B	B	C	D
0	0	1	2	3	4	5
A	1					
B	2					
C	3					
D	4					
A	5					

D-matrix initialization:

D[0][0] = 0
D[i][0] = i
D[0][j] = j

Re(diag); De(vert); In (horz)

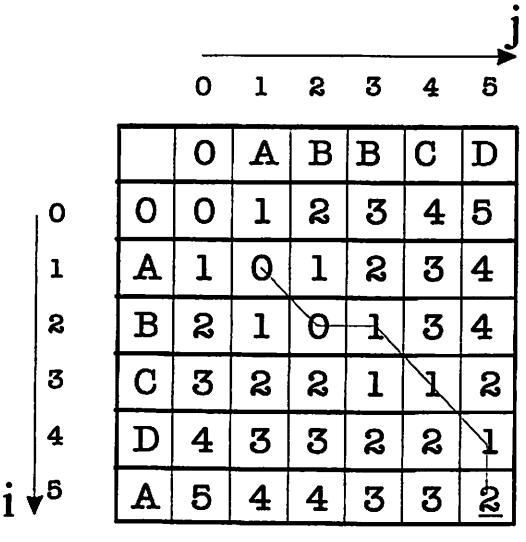


	0	A	B	B	C	D
0	0	1	2	3	4	5
A	1	0	1	2	3	4
B	2	1	0	1	3	4
C	3	2	2	1	(1)	
D	4					
A	5					

D-matrix computation:

Re = D[i-1][j-1]+$hcst_{12}$[i][j]
De = D[i-1][j]+$hcst_{10}$[i][j]
In = D[i][j-1]+$hcst_{02}$[i][j]

D[i][j] = min(Re,De,In);



	0	A	B	B	C	D
0	0	1	2	3	4	5
A	1	0	1	2	3	4
B	2	1	0	1	3	4
C	3	2	2	1	1	2
D	4	3	3	2	2	1
A	5	4	4	3	3	2

D-matrix completion:

D[n][n] = min_cost