

PERFORMANCE ASSESSMENT METHODOLOGY:
TASK DEPENDENT EVALUATION OF DISPLAY SYSTEMS

by

Dean H. Orrell

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Industrial and Systems Engineering

APPROVED:

Paul T. Kemmerling, Co-chair

Robert J. Beaton, Co-chair

C. Patrick Koelling

John K. Burton

Robert D. Dryden

February, 1997

Blacksburg, Virginia

PERFORMANCE ASSESSMENT METHODOLOGY:
TASK DEPENDENT EVALUATION OF DISPLAY SYSTEMS

by

Dean H. Orrell

Committee Co-chair: Paul T. Kemmerling, M.S.

Committee Co-chair: Robert J. Beaton, Ph.D.

(ABSTRACT)

As the focus of this research, a new methodology -- human Performance Assessment Methodology (PAM), is introduced. PAM provides a quantitative basis for evaluating display image quality based on the visual events that occur in a task. The PAM approach identifies the visual events, decisions, and actions for a display system. To support PAM, a theoretical model, the Model of Visual Events (MOVE), is proposed for describing the relationship between visual events, decisions, and actions. MOVE describes four categories of perceptual decisions (i.e., detect, identify, discriminate, and evaluate) associated with visual events. Formal efficiency metrics are introduced in PAM to describe performance at the visual event, task, and network levels. Using PAM, an efficiency model was created for one visual display parameter (i.e., luminance), one decision type (i.e., detection) and one dependent variable (i.e., visual angle).

Two experiments were accomplished to examine the validity of PAM. A two-factor mixed design was employed for both experiments, where decision type was varied between-subjects and visual display parameter (i.e., luminance or sharpness) was varied within-subjects. In the first experiment, luminance was varied across four levels (3.2, 4.5, 8.6, 16.5 cd/m²) for two decision types (detection and identification). In the second experiment, three levels of sharpness (50% spot width - 0.508, 0.711, 0.864 mm) were combined factorially with two decision types (detection and identification). In both experiments, participants visually 'walked down a path' and either detected or identified visual targets presented on the screen. Time-to-target and subjective responses were measured for each study.

The results of the first experiment show that time-to-target and subjective rating significantly change as a function of luminance. For the sharpness variable in the second experiment, a significant difference was found for time-to-target while subjective rating was non-significant. In both studies, participants detected visual targets quickly, but required more time to identify targets.

Using the PAM, functional relationships for luminance and sharpness were determined for detection and identification decisions. When detection data from the current study were contrasted with previous detection data, general agreement was found between the data sets. This research defines PAM and shows its utility for modeling the functional relationships among visual parameters. Further research is needed to validate and refine the PAM approach.

ACKNOWLEDGMENTS

A special tribute to my wife, Jennifer, for all her dedication and inspiration during the last four years. She was a constant source of cheer and often worked tirelessly to support my work. Her grace and understanding during our Blacksburg visit greatly encouraged me throughout the long dissertation journey. She helped me collect the data during the hot summer days and provided insights that improved my research. Thanks especially for putting up with all those late nights at the lab and supporting me while I worked many hours on the “document.” She’s my sweetheart and I love her.

I would like to thank my parents, Dean and Jeanette, for their love and faithfulness through the years. I also think of my grandparents, who would have been proud to know that their grandson was finishing his Ph.D.

My heartfelt thanks go out to both of my co-chairs, Prof. Paul Kemmerling and Dr. Bob Beaton for their sage inputs and their commitment throughout this dissertation process. Dr. Pat Koelling, Dr. John Burton and Dr. Bob Dryden (the other members of my committee) provided valuable comments on the written document. I owe Lovedia Cole (the ISE student records administrator) a sincere thank you for her help with all my scheduling and program of study issues. I also greatly appreciate the 60 participants who gave up an hour of their time to help me complete this research.

Finally, I must thank the awesome Creator, “the Alpha and the Omega”, for being my enduring source of strength. As King David wrote many years ago in the Psalms, “God is my rock.”

TABLE OF CONTENTS

INTRODUCTION.....	1
BACKGROUND.....	4
Task-based models.....	4
Review of task-based models.....	8
PERFORMANCE ASSESSMENT METHODOLOGY.....	9
Model of Visual Events.....	14
PAM efficiency metrics.....	20
Visual event efficiency.....	22
Task efficiency.....	24
Network efficiency.	25
Functional relationships.....	26
Stimulus intensity-response detection relationship.....	27
Visual parameter-performance relationship.....	28
Luminance-detection efficiency relationship.....	29
RESEARCH OBJECTIVE.....	32

METHOD.....	34
Participants.....	34
Experimental design.....	34
Materials.....	36
Equipment and testing environment.....	37
Procedure.....	37
RESULTS.....	41
Factorial analysis.....	42
Experiment 1- Luminance and decision type.....	43
Experiment 2 - Sharpness and decision type.....	47
Efficiency analysis.....	51
Efficiency curve generation.	52
Detection conditions.	54
Identification conditions.	57
DISCUSSION AND CONCLUSIONS.....	61
Effects - Factorial design.....	61
PAM - Application in current study.....	62

Comparison with luminance-detection model.....	66
Utility of PAM.....	70
Future research.....	73
REFERENCES.....	77
APPENDICES.....	82
Appendix A - Creating an efficiency curve using normal scores.....	83
Appendix B - Instructions to participants, Informed consent form.....	88
Introduction to the study.	88
Informed consent form.....	89
Participant instructions - Detection task.	91
Participant instructions - Identification task.	92
Appendix C - Geisser-Greenhouse correction factor.....	93
Appendix D - Detailed statistical analysis - Experiment 1.....	94
Appendix E - Detailed statistical analysis - Experiment 2.....	96
Appendix F - MacCurveFit program.....	98
Appendix G - Hyperbolic tangent function.....	99
Appendix H - Simple effects - Luminance-Detection conditions.....	102

Appendix I - Simple effects - Sharpness-Detection conditions.....	104
Appendix J - Simple effects - Luminance-Identification conditions.....	106
Appendix K - Simple effects - Sharpness-Identification conditions.....	108

LIST OF FIGURES

Figure 1.	Perception-Decision-Action Model (AGARD, 1989).....	9
Figure 2.	Performance Assessment Methodology (PAM) - task level.....	12
Figure 3.	Norman's (1988) Action Cycle and Seven Stages of Action.....	15
Figure 4.	Model for Visual Events (MOVE).....	19
Figure 5.	Performance Assessment Methodology - Event, decision, action level.....	21
Figure 6.	Efficiency ratings in EDA network for a given task.....	25
Figure 7.	Psychometric function - classical threshold theory (Gescheider, 1985).....	28
Figure 8.	The relationship between visual parameters and performance (Bullimore, Howarth, and Fulton, 1995).....	29
Figure 9.	Visual angle as a function of background luminance.....	31
Figure 10.	Visual angle of targets viewed on screen as a function of time-to-target.....	40
Figure 11.	Interaction between luminance and decision type for time-to-target in Experiment 1.....	43
Figure 12.	Main effect of luminance for time-to-target in Experiment 1.....	44
Figure 13.	Main effect of decision type for time-to-target in Experiment 1.....	44
Figure 14.	Interaction between luminance and decision type for rating in Experiment 1.....	45
Figure 15.	Main effect of luminance for rating in Experiment 1.....	46
Figure 16.	Main effect of decision type for rating in Experiment 1.....	46
Figure 17.	Interaction between sharpness and decision type for time-to-target in Experiment 2.....	47
Figure 18.	Main effect of sharpness for time-to-target in Experiment 2.....	48

Figure 19.	Main effect of decision type for time-to-target in Experiment 2.....	48
Figure 20.	Interaction between sharpness and decision type for rating in Experiment 2.....	49
Figure 21.	Main effect of sharpness for rating in Experiment 2.....	50
Figure 22.	Main effect of decision type for rating in Experiment 2.....	50
Figure 23.	Z-scores and associated probabilities for time-to-target in the luminance-detection condition.....	52
Figure 24.	Functional relationship for luminance and time-to-target - Luminance-Detection condition (Efficiency ₁).....	55
Figure 25.	Functional relationship for luminance and subjective rating - Luminance-Detection condition (Efficiency ₂).....	56
Figure 26.	Functional relationship for the sharpness and time-to-target - Sharpness-Detection condition (Efficiency ₃).....	57
Figure 27.	Functional relationship for the luminance and time-to-target - Luminance-Identification condition (Efficiency ₄).....	58
Figure 28.	Functional relationship for luminance and subjective rating - Luminance-Identification condition (Efficiency ₅).....	59
Figure 29.	Functional relationship for sharpness and time-to-target - Sharpness-Identification condition (Efficiency ₆).....	60
Figure 30.	PAM applied to current research.....	62
Figure 31.	Efficiency metrics for current research.....	63
Figure 32.	Luminance as a function of visual angle.....	67
Figure 33.	Sharpness as a function of visual angle.....	67
Figure 34.	Plot of detection data on Moon and Spencer (1944) luminance curve.....	68
Figure 35.	Plot of detection and identification data on Moon and Spencer (1944) luminance curve.....	70
Figure A1.	Unit normal distribution and associated probabilities.....	83

Figure A2.	Discrimination response time as a function of pixel size for aliased images (Booth, Byrden, Cowan, and Plante, 1987).....	85
Figure A3.	Ratings of picture quality as a function of pixel size for aliased images (Booth, Byrden, Cowan, and Plante, 1987).....	85
Figure A4.	Ratings of image quality for color slides as a function of image sharpness (Westerink and Roufs, 1989).....	86
Figure A5.	Identification of military vehicles using thermal imaging techniques (van Meeteren, 1991).....	87
Figure G1.	Hyperbolic tangent function and its parameters.....	100
Figure G2.	Slope of the hyperbolic tangent function.....	101

LIST OF TABLES

Table 1.	Taxonomy of visual events, decisions, and actions.....	10
Table 2.	Taxonomy of actions, decisions, and display parameters.....	11
Table 3.	Experimental design.....	34
Table 4.	Conditions and data points per experiment.....	35
Table 5.	Detection - Subjective rating scale.....	38
Table 6.	Identification - Subjective rating scale.....	38
Table 7.	Factorial analysis.....	41
Table 8.	Efficiency analysis by condition.....	42
Table 9.	Preparatory steps to apply PAM.....	72
Table 10.	Application of PAM.....	73
Table 11.	Criteria of model effectiveness	74
Table D1.	Experiment 1 - Time to target - ANOVA summary table.....	94
Table D2.	Experiment 1 - Time to target - Means, standard deviations and standard errors.....	94
Table D3.	Experiment 1 - Subjective rating - ANOVA summary table.....	95
Table D4.	Experiment 1 - Subjective rating - Means, standard deviations, and standard errors.....	95
Table E1.	Experiment 2 - Time-to-target - ANOVA summary table.....	96
Table E2.	Experiment 2 - Time-to-target - Means, standard deviations, and standard errors.....	96
Table E3.	Experiment 2 - Subjective rating - ANOVA summary table.....	97
Table E4.	Experiment 2 - Subjective rating - Means, standard deviations, and standard errors.....	97

Table H1.	Luminance-Detection condition - Time to target - Univariate F-test.....	102
Table H2.	Luminance-Detection condition - Time to target - Means, standard deviations, and standard errors.....	102
Table H3.	Luminance-Detection condition - Subjective rating - Univariate F-test.....	103
Table H4.	Luminance-Detection condition - Subjective rating - Means, standard deviations, and standard errors.....	103
Table I1.	Sharpness-Detection condition - Time to target - Univariate F-test.....	104
Table I2.	Sharpness-Detection condition - Time to target - Means, standard deviations, and standard errors.....	104
Table I3.	Sharpness-Detection condition - Subjective rating - Univariate F-test.....	105
Table I4.	Sharpness-Detection condition - Subjective rating - Means, standard deviations, and standard errors.....	105
Table J1.	Luminance-Identification condition - Time to target - Univariate F-test.....	106
Table J2.	Luminance-Identification condition - Time to target - Means, standard deviations, and standard errors.....	106
Table J3.	Luminance-Identification condition - Subjective rating - Univariate F-test.....	107
Table J4.	Luminance-Identification condition - Subjective rating - Means, standard deviations, and standard errors.....	107
Table K1.	Sharpness-Identification condition - Time to target - Univariate F-test.....	108
Table K2.	Sharpness-Identification condition - Time to target - Means, standard deviations, and standard errors.....	108
Table K3.	Sharpness-Identification condition - Subjective rating - Univariate F-test.....	109

Table K4.	Sharpness-Identification condition - Subjective rating	
	Means, standard deviations, and standard errors.....	109

INTRODUCTION

Human vision has been studied for centuries and from a variety of perspectives. Snyder (1988) defines three typical areas of visual research: temporal, color, and spatial characteristics. In the temporal domain, visual stimuli change over time. In the chromatic area, stimulus factors such as purity of color and physical wavelength are considered.

In the spatial area, it is common to consider “image quality metrics” as a measure for a given visual display system. These metrics are based on the idea that the “luminance output pattern at the display surface must be quantified precisely in order to define what is meant by image quality for any given observer response” (Snyder, 1973). These image quality techniques evolved from the evaluation of photographic film image quality in the 1920's and 1930's and were later applied to computer monitor design (Snyder, 1988).

Beaton and Farley (1991) describe two important concepts related to image quality -- the Modulation Transfer Function (MTF) and the Contrast Threshold Function (CTF).

Most image quality metrics are based on spatial frequency domain representations of the display device and the human observer. Specifically, the capacity of the display system to transmit a signal is indexed by the MTF. The MTF expresses the relative amount of modulation (i.e., normalized luminance contrast) produced by a display system as a function of the spatial frequency components comprising an image signal. Each value of the MTF is the ratio of output-to-input at a particular spatial frequency value. For all realizable display systems, the MTF decreases with increasing spatial frequency, and, thus, all displays possess a finite capacity (bandwidth) to transmit information signals.

While the MTF expresses the capacity of a display system to transmit visual signals, a similar spatial frequency-domain function can be used to index the capacity of the human observer to perceive those signals. The CTF expresses the amount of modulation required by the human visual system to detect the spatial frequency components of an image signal. Each CTF value represents the minimum amount of modulation required by the human visual system to detect a particular spatial frequency component. In general, the CTF increases with increasing spatial frequency; however, the CTF magnitudes are affected by viewing conditions (pp. 1-2).

Thus, MTF describes a display system by determining the effect of spatial frequency on contrast and is calculated using Fourier analysis, while CTF represents the amount of luminance contrast (modulation) that is capable of being detected by the human eye.

Several image quality metrics are determined by the relationship between MTF and CTF, such as Modulation Transfer Function Area (MTFA; Snyder, 1973), Integrated Contrast Sensitivity (ICS; van Meeteren, 1973) and Square-Root Integral (SQRI; Barten, 1987) methods. For each of these metrics, a function based on MTF and CTF is mathematically integrated between the lower and upper spatial frequency limits. Among these metrics, MTFA gained wide acceptance in the evaluation of display systems while ICS and SQRI were later attempts to improve the MTFA. The ICS was designed to be more sensitive to small changes in the MTF and CTF, while the SQRI uses non-linear scaling across spatial frequencies to account for contrast discrimination patterns found in human subjects.

While helpful for describing display system characteristics and assessing image quality, these metrics provide a limited understanding of the task-relevant

issues for display design and do not address task-specific requirements needed by users. That is, image quality metrics do not account for task variables. For example, when watching television at home, the resolution and refresh rate of the screen usually are adequate for the task. However, for many other tasks (e.g., open-heart surgery), a television monitor may be inadequate. What is the difference in these situations? Since the monitor remains unchanged, the difference in image quality is not based on display technology, but on task demands. Also, image quality metrics do not account for the direct influence of operator performance or proficiency under given visual conditions. For example, if a visual parameter such as luminance changes, the effect on performance must also be considered.

The purpose of this research to describe and empirically validate a new methodology that provides a quantitative basis for evaluating display image quality. This methodology is based on the visual events that occur in a task and focuses on describing the functional relationships for visual display parameters.

BACKGROUND

Task-based models

Task analysis can be defined as “the study of what an operator is required to do, in terms of actions and/or cognitive processes, to achieve a system goal” (Kirwan and Ainsworth, 1992). Task analysis helps to describe, organize, and decompose a task or group of tasks into meaningful components and help provide a framework for how the task fits into the overall system.

While there are many task analytic methods, they seem to fall into two general descriptive categories: “macro” and “micro” methods. At the macro level, tools include Hierarchical Task Analysis (HTA; Kirwan and Ainsworth, 1992) and Integrated Computer Aided Manufacturing (ICAM) definition diagrams (Deuermeyer and Cullinane, 1981). These tools provide a global picture of how a system operates and how tasks inter-relate to each other. At the micro level, tools such as Operational Sequence Diagrams (OSDs), task decomposition, and task network models (e.g., Micro-SAINT) are often used by human factors engineers. The following review describes several task analysis techniques outlined by Kirwan and Ainsworth (1992).

To illustrate a typical task analysis, task decomposition (also known as Tabular Task Analysis; Miller, 1963) starts with a task description and decomposes it into basic elements or components. The procedure breaks down a task into the following categories: description, subtask, cues and initiating action, controls used, decisions, typical errors, response, criterion of acceptable performance, and feedback. The task categories are then listed in a

tabular format for easy reference. This technique allows one to consider systematically the issues that influence the task. However, increased time and effort are needed for decomposition of larger systems.

Hierarchical Task Analysis (HTA; Duncan, 1974) generates a hierarchy of system “operations” or tasks that need to be done and “plans” that determine how tasks will be completed. Kirwan and Ainsworth (1992) provide an extensive description of the technique. In HTA, goals comprise “desired states of systems under control of supervision,” while tasks are methods to attain goals. Plans determine when goals and sub-goals are accomplished.

The hierarchical nature of the technique provides a good frame of reference for how tasks fit together. As a diagnostic tool, HTA provides insight as to how the organization operates and how it is functioning (in terms of processes and procedures). Also, since HTA requires operators to become involved in the data gathering process, they gain a better understanding of how their system works. The technique requires a moderate amount of training to implement, and, thus, it is not a “simple” technique. Also, collaboration and input from managers is necessary, which increases the amount of time and effort to use the technique.

OSDs are defined to be “a detailed description of operator, equipment, and software actions in which each of these elements is sequentially integrated with the other in a time sequence” (Drury, Paramore, Van Cott, Grey, and Corlett, 1987). A typical OSD depicts a task sequence in a flow diagram. OSDs use five symbols to define information flow (i.e., operator decision, action,

transmitted information, received information, and previously stored information). Temporal OSDs add a timeline component, while partitioned OSDs break out task dimensions according to individual responsibilities (e.g., one per person in a multi-person task). Spatial OSDs represent spatial or geographical links among operations and are usually represented by a map or panel diagram.

Using OSDs it is relatively easy to show the many relationships among operations for single and multi-person operators. OSDs also have a high face validity (i.e., when people see them, they feel they are accurate representations of the system). But, for large and complex systems, they can become cluttered and confusing due to the difficulty in following the relationships between diagrams. Also, each OSD tends to be an entity to itself, and, even though multiple dimensions can be represented (e.g., distance, time), it is difficult to represent these on the same chart.

Simulation techniques model features of the operator and the system. Well-known examples of simulation models include: System Analysis of Integrated Networks of Tasks (SAINT; Siegel and Wolf, 1969) and Micro-SAINTE (Laughery and Drews, 1984). In simulation methods, subtasks and their inter-relationships are modeled in the computer software. The user then enters appropriate task parameters and the computer model provides the output.

Once the system has been modeled, users can explore and optimize different system configurations. For computer models, the quality of output and the model's predictions depend on the accuracy of the input data. Also, given

the complexity of large scale systems, the greater the level of accuracy or detail modeled in a system, the lower the ability to generalize the model to other systems. In addition, it is time-intensive to update and maintain the database that represents the system, especially if the system is constantly changing.

Examples of research using simulation models include work by Moscovic (1992) and Green (1995). Moscovic (1992) integrated the MODular Arrangement of Predetermined Time Standards (MODAPTS) methodology (Shinnick and Gerber, 1985) and Micro-SAINT. Participants in the study used Naval electronic warfare equipment designed for anti-ship missile defense purposes. Using a Naval display console, participants performed a complex auditory and visual task. She found that the MODAPTS technique was effective for assigning task completion times in Micro-SAINT.

Green (1995) extended this research and introduced the “Performance Assessment Methodology” (PAM) terminology by developing a prediction model with Micro-SAINT. He chose a “functional recognition” task from the Naval electronic warfare tasks, in which the operators were required to “determine the purpose of the emitting radar.” This task required operators to process both auditory and visual information. He found the correlation between observed and predicted times to be quite high (i.e., 0.958) indicating a high degree of agreement between the Micro-SAINT model and operator performance of the functional recognition task.

Review of task-based models

While previous models are helpful in understanding how tasks can be represented, few specifically address the issues of how vision relates to performance. The limited visual data that have been collected provide a few data points for the Micro-SAINT simulation model. However, a general performance-based framework for understanding vision has not been provided. A new model is needed that goes beyond the vision image quality metrics discussed previously. Although most of the previous task models provide a detailed breakout of system functions, few provide ways to relate these functions to important decisions in the task. A new model could describe how important visual parameters relate to decisions, based on the decomposition nature of extant models.

This model should categorize the visual parameters and show the differential contributions of each visual parameter for a given decision. The general procedure would be to relate system parameters to performance (e.g., determine relationship between resolution and performance--task completion time) and weight performance measures to determine influence on a given parameter. This can be done by considering the influence of display system parameters (e.g., brightness, resolution, and contrast) for each visually-based decision in the task sequence. Also, it is important to show inter-relationships between system components and, thus, to provide a hierarchy of decision classes. Finally, the model should provide specific metrics to account for operator performance.

PERFORMANCE ASSESSMENT METHODOLOGY

The Performance Assessment Methodology (PAM) is defined as a framework that shows the relationships among operator tasks, display system parameters (e.g., brightness, resolution, and contrast), and performance measures for these parameters. In the PAM, operator tasks are decomposed into visual events, decisions, and actions. When visual events occur, the operator makes a decision and then responds with an action. This premise is similar to other work (e.g., Perception-Decision-Action model, AGARD, 1989; Green, 1995), where observable perceptions or events lead to decisions and actions (see Figure 1 below). A task might have many events, decisions, and actions associated with it.

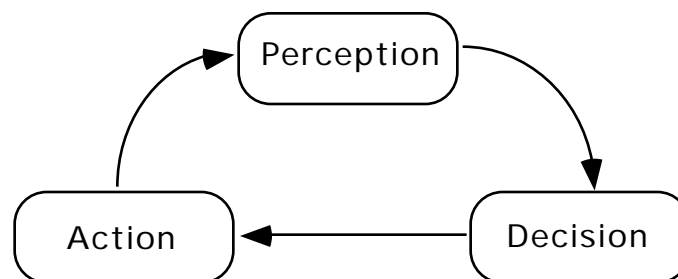


Figure 1. Perception-Decision-Action Model (AGARD, 1989).

PAM provides a framework for describing the relationships between Visual Events, Decisions, and Actions (EDA) and their associated visual parameters (see Table 1). In this taxonomy, a visual parameter may be associated with many visual events, and individual decisions and actions can

be isolated for further evaluation. Thus, a parameter can be related to any level of an EDA, so that Decision 1 and Decision 4 may use the same parameter.

Table 1.

Taxonomy of visual events, decisions, and actions

	Visual parameter
Visual event 1	brightness, resolution
Decision 1	
Action 1	
Action 2	
Decision 2	
Action 3	
Visual event 2	field of view, contrast, resolution
Decision 3	
Decision 4	
Action 4	
Visual event 3	contrast, addressability
Decision 5	
Action 5	
Action 6	

One also may start with the actions and then determine how events and decisions influence these actions (see Table 2). For example, suppose a task has 10 actions associated with it. Several questions help define this task taxonomy. First, what are the efficiency parameters associated with each action? Next, given these parameters, how well does the display system allow us to make the decisions (i.e., detect, identify, recognize, and interpret) associated with these actions? The answers to these questions define the relationship between actions and their associated displays.

Table 2.

Taxonomy of actions, decisions, and display parameters

Action 1

Decision 1- detect

Parameter 1 - luminance

Parameter 2 - resolution

Parameter 3 - field of view

Decision 2 - identify

Parameter 1 - brightness

Action 2

Decision 3 - evaluate

Parameter 1 - addressability

Parameter 2 - resolution

For operator tasks, PAM shows: 1) the inter-relationships between visual events, 2) how display system parameters influence individual events, and 3) how these parameters are related to performance measures. This process allows the designer to consider how visual display parameters such as brightness, resolution, and field of view influence task performance. PAM is based on a network analysis model of time sequential visual events. Figure 2 shows a sequence of EDA that leads to task completion and the influence of display parameters on a selected EDA node.

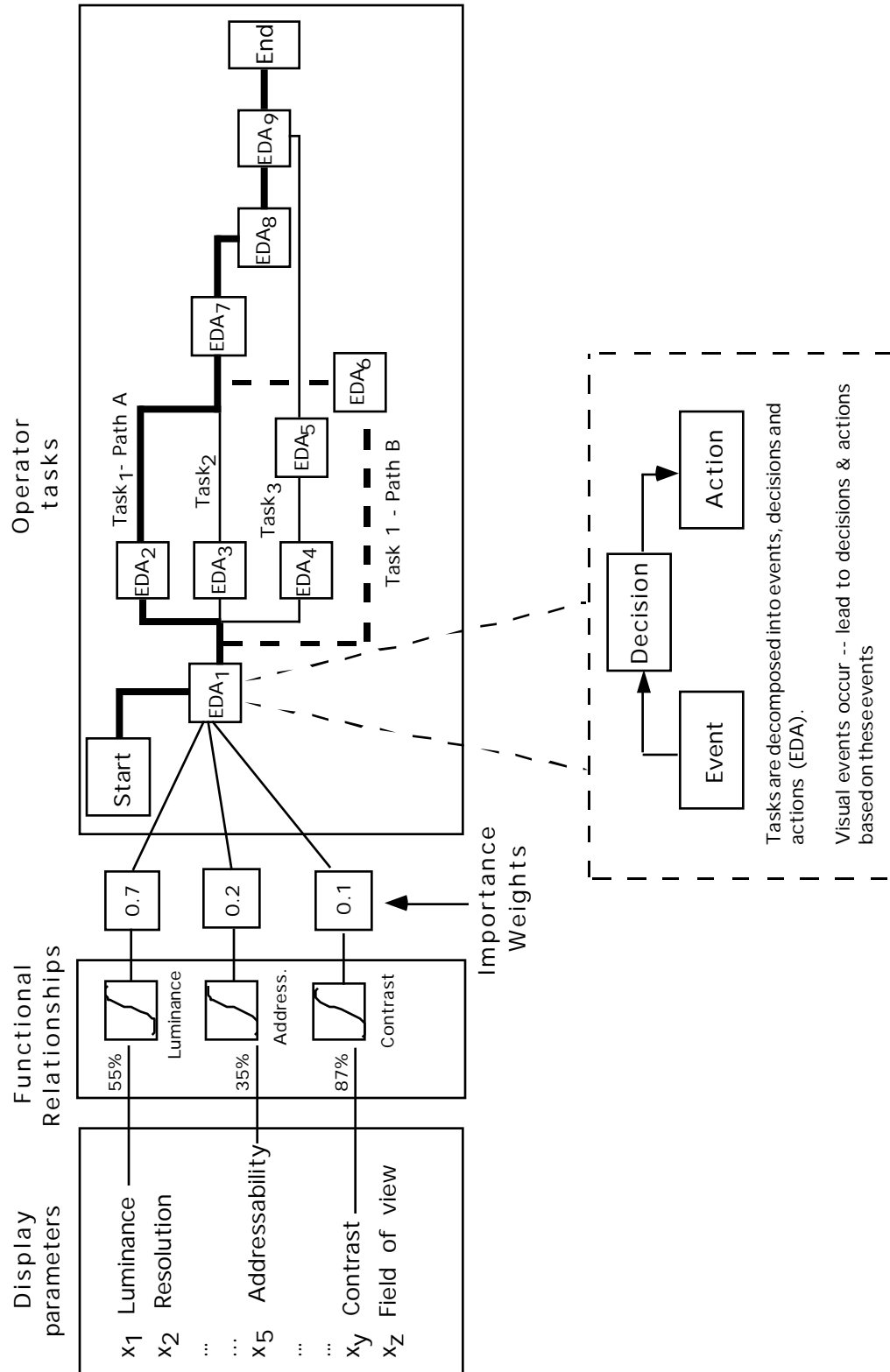


Figure 2. Performance Assessment Methodology (PAM) - task level.

In PAM, the influence of display device characteristics (or parameters) are determined for each task-relevant visual event. Typical parameters considered in computer system design include: resolution, raster modulation/active area, luminance and contrast, polarity, image stability, flicker, color selection and contrast, character design, text spacing, and glare control (Snyder, 1988). Bullimore, Howarth, and Fulton (1995) break these parameters into physical and cognitive components. Simple physical aspects include: illumination, contrast, size, and exposure time; while those involving cognitive components include: image dynamics, pattern recognition and coding, redundancy, use of color, display location, size, and area. PAM allows the consideration of each of these parameters for a task.

Once the display parameters have been selected, functional relationships between performance measures and display parameters must be determined. In PAM, an “efficiency” measure is created by combining and weighting performance scores across visual parameters. Thus, efficiency is an aggregate measure of performance (e.g., time, errors, and workload) for individual display parameters. PAM allows consideration of any human performance measure including accepted measures such as speed (time) and accuracy (errors) (Drury, 1995; Snyder, 1973). Stress or workload defined as “the difference between the perceived demands of the task and a person’s perceived capacity to cope when coping is important” provides an additional measure of performance (Drury, 1995).

In PAM, two components need definition by subject matter experts before a model may be created: 1) task-relevant visual events and 2) display parameters. Obviously many things could be classified as a visual event and display parameters are dependent on the nature of the task. A technique known as “ranking and rating” found in the multiple criterion decision analysis literature is used to determine important visual events and important display parameters. The ranking and rating procedure “has been used by numerous decision makers, has been found to be easy to use, and has led to useful evaluations” (Morris, 1977).

Using this procedure, one ranks in order of importance each factor within a category (ties are permitted) and then assigns a rating to each factor on a ratio scale from 0 to 1.0. Those factors with the highest ratings are then selected for a given task. The cutoff point (between 0-1) for inclusion in a task is also determined by subject matter experts. In PAM, the visual event weights are not included in the formal model calculations, but simply help define important visual events. For display parameters, the actual weights are combined with a performance measure in the PAM model.

Model of Visual Events

To support this research, a model relating visual parameters, decision types, and human performance is proposed herein. This model, called the Model of Visual Events (MOVE), provides a theoretical foundation for representing how one responds to visual information in a given environment. Based on a review of the literature, it identifies four basic decision types for the

PAM framework. MOVE also expands upon the AGARD (1988) event-decision-action model by incorporating the “perception,” “decision-making,” and “response/action sequence” components of Norman's (1988) Action Model and Wickens' (1992) Information Processing Model.

Norman (1988) presents the seven-stage action cycle shown in Figure 3. In this model, one starts out with a goal and accomplishes two basic steps; execution of a task sequence and evaluation of the results in light of the goal. The basic steps of execution and evaluation are further decomposed into the substeps shown in Figure 3.

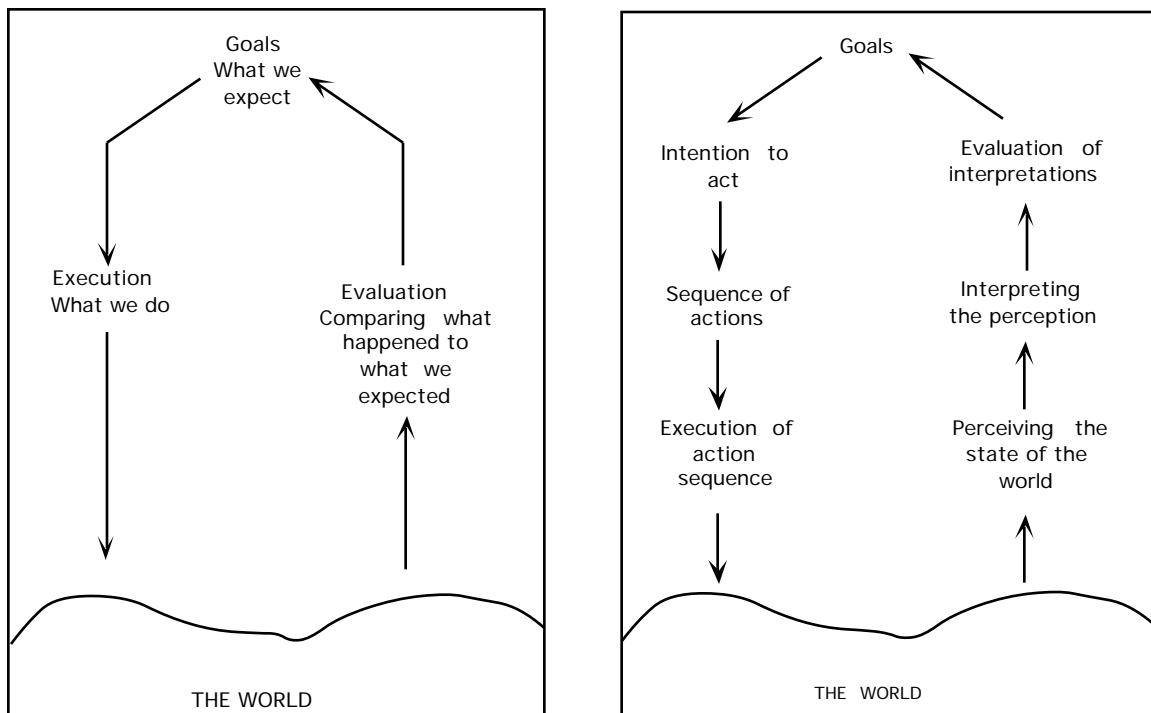


Figure 3. Norman's (1988) Action Cycle and Seven Stages of Action.

During the execution stage, the first step is to transform one's goal into specific statements or actions called *intentions*. In this model, goals and intentions become plans, which, in turn, are put into an action sequence. For

example, if the goal is to load socks into the clothes washer, then the intention is to move the socks from the laundry basket and load them into the washer. The next step, the *action sequence* is the very specific physical movements of the hand and elbow to accomplish this task. Finally, the action sequence is actually *executed*. In this context, the intention is a more global plan and the action sequence is a detailed series of steps or motions required to implement the intention.

During the evaluation mode, one *perceives* the information being sent (e.g., I see a sock on the floor). Next, one *interprets* what this means (e.g., I must have dropped it when loading it into the washer). Finally, the outcome is *evaluated* (e.g., Did my socks successfully make it into the washer on my first attempt? No). So, the goal was not completed, thus the sock is picked up and put in the washer. Norman's (1988) model provides a baseline for understanding how an action is performed and his concepts are used to develop a general model for visual events.

Wickens (1992) presents a general model of the human as an information processor that accounts for the influence of attentional resources and memory. His model provides a framework for understanding how one processes and responds to information. This is a three stage process, in which one *perceives* a stimulus, makes a *decision and response selection*, and develops a *response*. For the perceptual process, he differentiates between *detection* processes and processes of *recognition, identification, and*

categorization. These distinctions provide the basis for a visual model based on these processes.

Coren, Ward, and Enns (1994) break these perceptual elements into four basic components: detection, identification, discrimination (or evaluation), and scaling. In the detection process, a person might say, "Yes, I see something there" and might identify it by "I see it is a tank." These first two steps (detection and identification) are thought to be "quick and automatic processes unless conditions such as weather or other competing stimuli make them difficult." In the discrimination phase, the stimulus is differentiated from others: "It is not an M1 tank" (discriminating it from a friendly tank) and it is determined that "The tank is shooting at me." Finally, scaling questions such as "How big is the tank?" or "How far away is it?" are considered.

Snyder (1973) identifies four perceptual categories when observers search for objects in a visual field: detection, recognition, area recognition, and discrimination.

Detection is said to occur when the observer correctly indicates his decision that an object of interest exists in the field of view.

Recognition is said to occur when the observer correctly indicates to which class of objects the detected object belongs.

Area recognition is said to occur when the observer correctly indicates that the location of the target is in the field of view.

Discrimination is said to occur when the observer correctly indicates that the *singular* target of interest is in the field of view; that is, he is correct in separating the single target of interest from the class of recognized targets.

Bullimore, Howarth, and Fulton (1995) provide additional support for decomposition of perceptual processes. They outline the tasks of detection, recognition, and interpretation. In a detection task, a person might be able to see that a warning light is on. In a recognition task, there must be discrimination between complex stimuli that involves judgment of what a person has detected. Discriminating between a dashboard oil and fuel indicator is an example. Interpretation requires users to determine the meaning or consequence of a visual event. Examples are when dashboard lights indicate some type of engine malfunction.

MOVE provides a framework for these perceptual components as shown in Figure 4. In this model, once an object is perceived, a decision is made relative to the perceptual information. In MOVE, the flow of the model is dependent on the nature of the task. For a simple detection task, one would detect and immediately evaluate the detected image. However, for a more complex task, one detects, identifies or recognizes, discriminates, and then interprets or evaluates.

The seven steps from “detection” to “formation of an action sequence” are part of the perceptual/decision making process outlined in both Norman's and Wickens' models. The detect, identify/recognize, discriminate, interpret/evaluate components of MOVE are used to model visual events, decisions, and actions. These components are the four cognitive decision types for a visual task in PAM. The formalism that describes these components is presented in the next section.

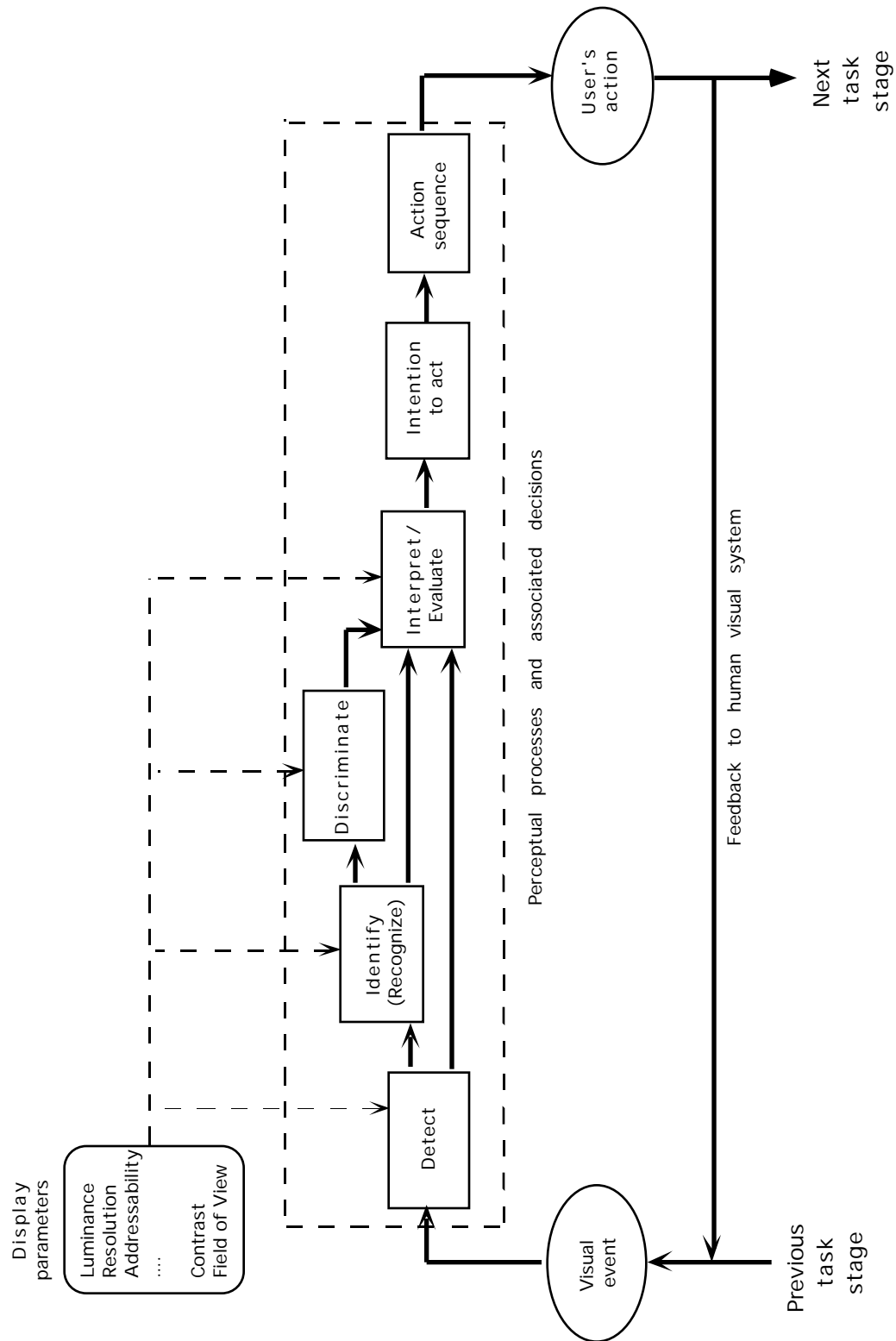


Figure 4. Model for Visual Events (MOVE).

When users first encounter a visual event, they make one of four decisions (i.e., detect, identify, recognize, interpret) based on the current task. For any decision type: the display system parameters influence the decision that is made, the appropriate action sequence is determined, and finally action is taken.

PAM efficiency metrics

In this section, metrics for efficiency are developed that formally relate visual events, decisions, and actions. The formalism for determining efficiency at the individual EDA level, the task level, and network level is provided. A detailed model is developed for a single visual event that relates the display system parameters to performance (see Figure 5). If the user's task is, for example, to identify a graphic on the computer screen, the visual event is the perception of the graphic on the screen. The visual display parameters influencing this visual event include luminance (Parameter 1), sharpness or resolution (Parameter 2), and field of view (Parameter 3). Efficiency for a given decision type involving a graphic image can be determined when the relationship between the parameters and visual perception is specified.

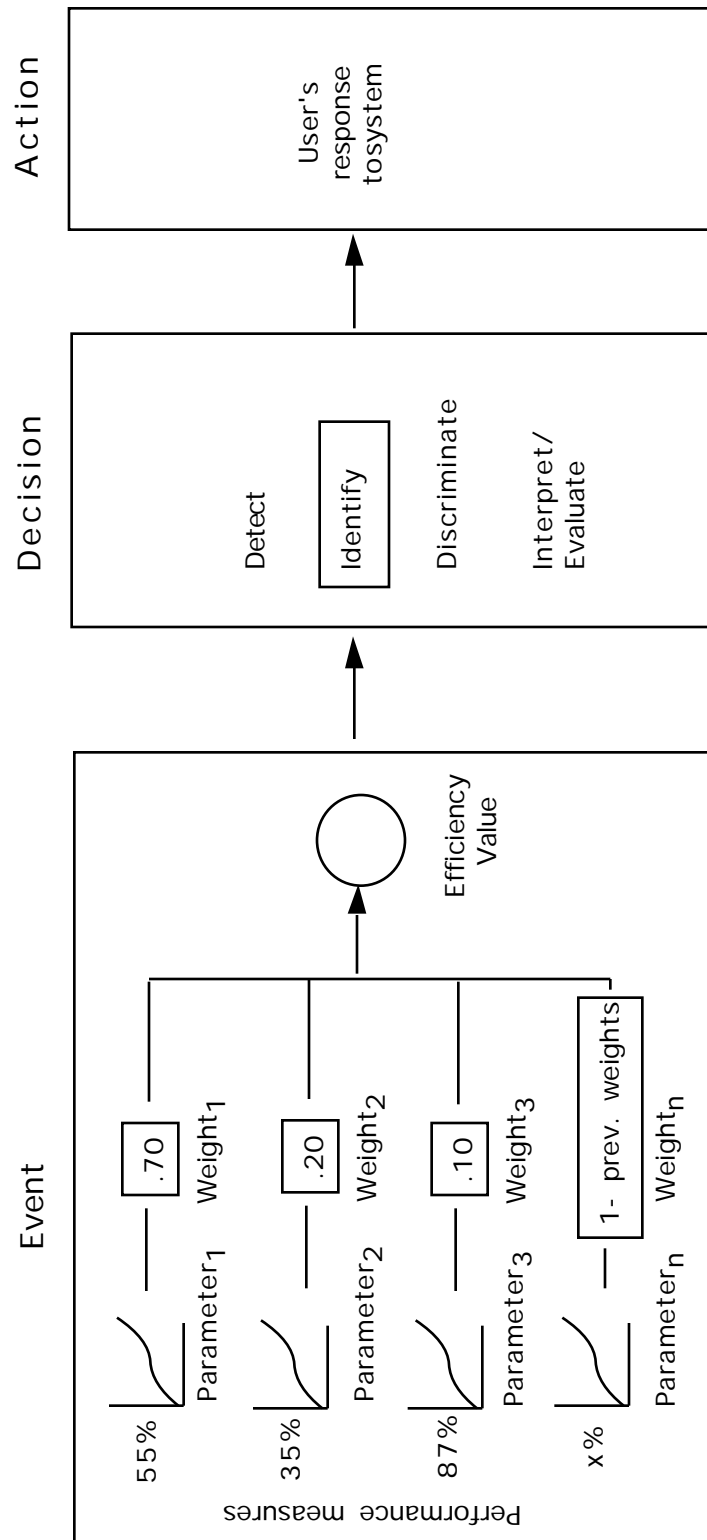


Figure 5. Performance Assessment Methodology - Event, decision, action level.

In this model, a one-to-one mapping for each visual event is posited (e.g., for each visual event, one decision is made). For the visual event shown in Figure 5, the general relationship for a given MOVE decision class is shown as:

$$\text{Decision class}_i = \{a_1P_1, a_2P_2, a_3P_3, \dots, a_nP_n\}$$

where

Decision class_i = a unique MOVE decision class

(i.e., detect, identify/recognize, discriminate, interpret)

P₁ = visual display parameter in a given decision class

a₁ = weighting value associated with a given parameter

n = number of visual parameters

Further, each class of decisions (i.e., detect, identify/recognize, discriminate, interpret/evaluate) is thought to be associated with a unique set of display system parameters. Example parameters (P) for each MOVE decision class are shown below:

$$\text{Decision class}_{\text{Detect}} = \{a_1P_1, a_2P_2, a_8P_8\}$$

$$\text{Decision class}_{\text{Identify/Recognize}} = \{a_2P_2, a_4P_4, a_{10}P_{10}\}$$

$$\text{Decision class}_{\text{Discriminate}} = \{a_3P_3, a_4P_4, a_5P_5\}$$

$$\text{Decision class}_{\text{Interpret/Evaluate}} = \{a_2P_2, a_9P_9, a_{11}P_{11}\}$$

Note that individual parameters may belong to several decision classes.

Visual event efficiency.

Since several parameters (e.g., luminance, sharpness, field of view, etc.) may influence a particular *visual event*, each must be considered. The

relationship between efficiency measures for visual parameters is modeled as additive, where the effects of several parameters are averaged across a visual event. For a visual event, the efficiency values for physical parameters are weighted and summed and a single efficiency value assigned. So that:

Efficiency =

$[(\text{visual parameter}_1)(\text{weight}_1) + (\text{visual parameter}_2)(\text{weight}_2)$
 $\dots + (\text{visual parameter}_n)(\text{weight}_n)]$; where individual weights
 sum to 1.

Thus, for a selected visual event, associated decision, and action (EDA):

$$EDA_{\text{efficiency}} = \sum_{j=1}^n (a_j P_j) \quad \text{Eq. 1}$$

where

P_j = efficiency value for an individual visual parameter (based on
 the functional relationship between the visual parameter
 and performance)

a_j = weighting factor associated with a parameter, where $\sum_{j=1}^n a_j = 1$

n = number of visual parameters

The individual weight (a_j) assigned to each parameter is determined using the “ranking and rating” procedure previously discussed (Morris, 1977).

For example, consider the parameters associated with the following visual

event. If efficiency is 0.75 for luminance of 9 cd/m² and 0.70 for sharpness of 0.445 mm and both have equal weighting factors (i.e., 0.5), then the combined efficiency value for the visual event is:

$$[(0.5)*(0.75) + (0.5)*(0.70)] = 0.725$$

This value represents the efficiency for a given visual event, decision, action point and accounts for the influence of individual parameters.

Task efficiency.

Efficiency values for a *task* are calculated as a function of the individual EDA proficiencies.

$$Task_{efficiency} = \sum_{k=1}^s (b_k EDA_k) \quad \text{Eq. 2}$$

where

EDA_k = efficiency value for an visual event, decision, action

b_k = weighting factor for a given EDA node, where $\sum_{k=1}^e b_k = 1$

s = number of EDA nodes

Thus, for a task, individual EDA proficiencies are added together to provide an overall estimate of efficiency for a selected group of EDAs or unique completion paths as shown in Figure 6.

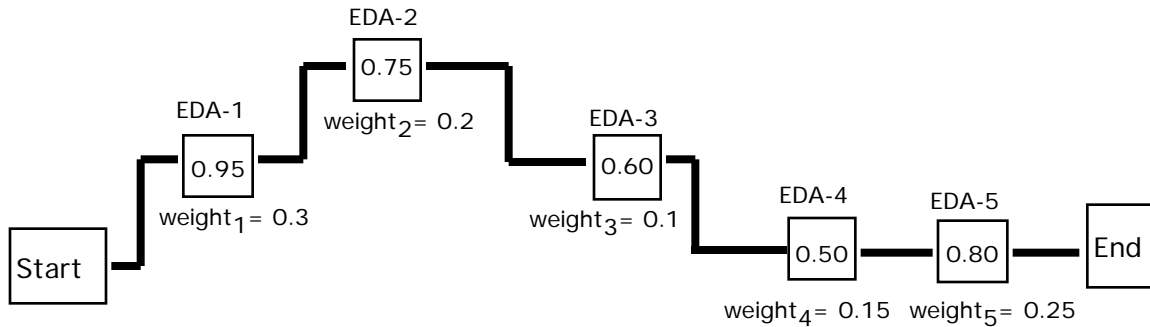


Figure 6. Efficiency ratings in EDA network for a given task.

For this example, task efficiency

$$\begin{aligned}
 &= [(0.95)(0.3) + (0.75)(0.2) + (0.60)(0.1) + (0.5)(0.15) + (0.80)(0.25)] \\
 &= [0.285 + 0.15 + 0.06 + 0.075 + 0.2] = 0.77
 \end{aligned}$$

Network efficiency.

Finally, an efficiency value for the overall *network* is calculated as:

$$Network_{Efficiency} = \sum_{l=1}^t (c_l Task_l) \quad \text{Eq. 3}$$

where

Task_l = efficiency value for a task

c_l = weighting factor for a selected task, where $\sum_{l=1}^t c_l = 1$

t = number of tasks

Network efficiency is computed across the multiple completion paths for a task. In the case where any one of three paths may be used to complete a task, the network efficiency is simply averaged across the paths.

Functional relationships

In PAM, functional relationships between visual display parameters and performance measures are created. Thus, the relationship between a given parameter and performance must first be determined. This information may be drawn from previous studies or derived empirically. However, the best performance curves are created by aggregating data across many research studies. This provides a comprehensive view of how data fit together for a given display variable.

Since PAM seeks to combine performance scores for several parameters and provide a single measure of efficiency, scores from individual parameters must be combined. For example, in Figure 5, the performance scores for three parameters are combined to provide a unitary measure of efficiency. A prevalent way to compare scores across studies is by using “standard scores” or “z scores” (Kantowitz, Roediger, and Elmes, 1991). Using this technique, performance scores are standardized and combined. Thus, to empirically determine functional relationships for each parameter, a cumulative probability function based on the normal distribution is determined. An efficiency curve is created by plotting the standardized performance measure versus visual parameter (see Appendix A for an overview of how efficiency curves are created and example efficiency curves).

Three examples of functional relationships are provided below. The first is based on a detection task from classical threshold theory (Gescheider, 1985) and is based on performance data. The second example is a theoretical model

of visual performance for the parameters of illumination, contrast, size, and exposure time (Bullimore, Howarth, and Fulton, 1995). The third is a luminance-detection model which will provide a basis of comparison for the current research data.

Stimulus intensity-response detection relationship.

For a basic detection decision, classical threshold theory shows that the relationship between stimulus intensity and responses follows an ogive function. In classical threshold theory, one's "momentary threshold" for a particular stimulus is affected by parameters that fluctuate from moment to moment. Accordingly, if the variation of momentary thresholds is normally distributed, an ogive function for stimulus intensity is expected as shown in Figure 7 (Gescheider, 1985). In this relationship, the threshold is defined to be the point at which 50% of the responses are "yes".

The graph shows the proportion of "yes" responses as the dependent measure. Thus, for a detection task with a dichotomous dependent variable (i.e., "yes"/ "no" responses), an ogive function is predicted. Obviously, the shape of the curve is dependent on the relationship between variables. In the case of the detection task, an ogive function is created since the variables are directly related to one another. If the "yes" responses decreased as a function of stimulus intensity, the variables would be inversely related and the function would represent an exponentially decreasing curve.

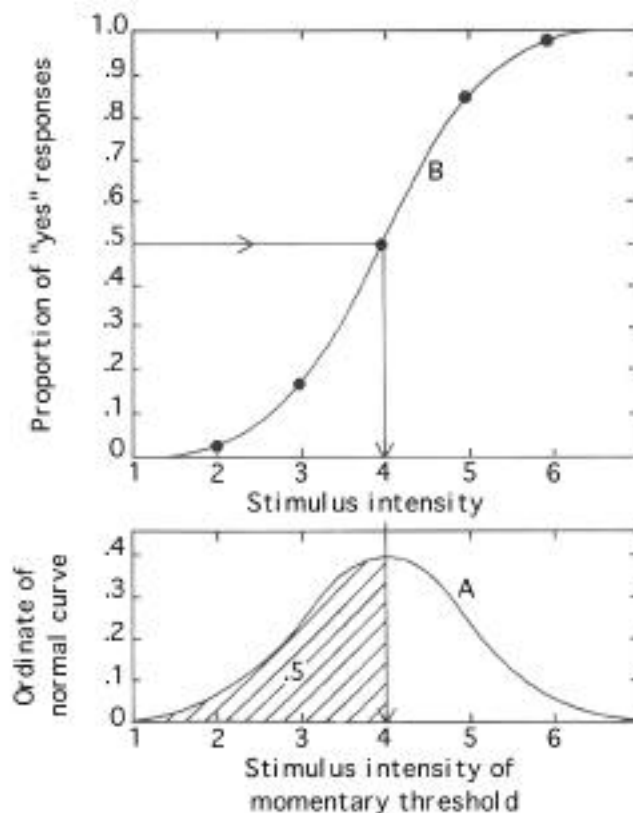


Figure 7. Psychometric function - classical threshold theory (Gescheider, 1985).

Visual parameter-performance relationship.

Bullimore, Howarth, and Fulton (1995) provide a model showing the general relationship between task performance and four visual parameters (size, contrast, luminance, and exposure time) as shown in Figure 8. Bullimore, Howarth, and Fulton describe this relationship as follows:

As a general rule, visual performance is better the brighter the ambient lighting, the greater the contrast between object and background, the larger the object, and the longer the viewing time. If the parameter value is too low (e.g., the size is too small), then the task will be below the threshold. As the parameter increases threshold is reached and subsequent increases will improve performance until the optimum level is reached. In some cases, if the parameter continues to increase performance will eventually decline (p. 818).

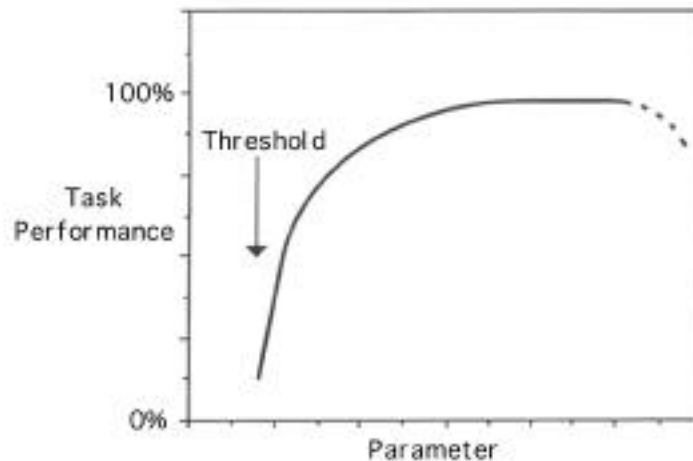


Figure 8. The relationship between visual parameters and performance (Bullimore, Howarth, and Fulton, 1995).

The Bullimore, et al. (1995) model is a general theoretical one, and it is not based on any one data set. It is based on the idea that for a given visual parameter, a threshold does exist. Above threshold performance tends to level off and finally at some point performance decreases as the parameter continues to increase. For example, if luminance or contrast is too high performance falls off because of glare (Howarth, personal communication, 1996). Note that if values below threshold were plotted, the shape of the function would be similar to the ogival function from classical threshold theory.

Luminance-detection efficiency relationship.

Efficiency curves can be developed for many possible dependent measures (e.g., response time, errors, workload, visual angle), for the four MOVE decision types (i.e., detect, identify, discriminate, interpret), and for many visual parameters (e.g., luminance, sharpness, field of view, contrast, etc.).

Indeed, a family of curves may be generated based on the decision type, the

parameters that influence a given decision type, and the dependent variables of interest.

A theoretical efficiency curve for one visual parameter (**luminance**), one specific decision type (**detection**), and one dependent measure (**visual angle**) will be developed. This curve will later be contrasted with experimental data. Visual angle has been selected as the dependent measure since it allows for comparison across several studies. It is a measure of performance, where an increase in visual angle generally improves one's ability to detect or identify an object. It is usually reported in minutes of arc (min arc).

Moon and Spencer (1944) examined detection data from six studies (Cobb and Moss, 1928; Lythgoe, 1932; Conner and Ganoung, 1935; Shlaer, 1937; Shlaer, Smith, and Chase, 1942; and MacAllister, unpublished data) to show the relationship between visual angle and background luminance. In each of these studies, participants performed a binocular visual acuity task under varying luminance levels. Moon and Spencer originally reported visual angle in microradians and luminance in log blondels, while Chapanis (1949) used minutes of visual angle and log millilamberts for the same data. In Figure 9, the Chapanis data was converted from millilamberts to nits (i.e., 3.183 millilamberts = 1 cd/m^2) and then plotted on a logarithmic scale.

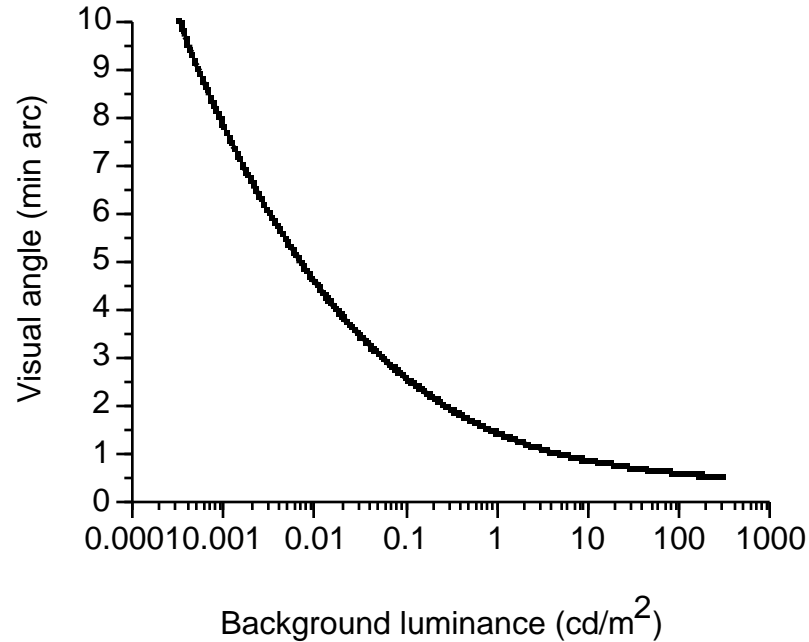


Figure 9. Visual angle as a function of background luminance.

This luminance-detection model, based on the Moon and Spencer (1944) composite data show the effects of luminance across a wide range of values. Some of the current research data will be contrasted with this model to provide a context for interpretation. This comparison provides a basis for understanding the luminance-detection data. It is anticipated that there will be general agreement between the data sets. However, the utility of PAM modeling is not dependent on any one particular data set. The strength of the PAM approach lies in its ability to describe functional relationships for visual display parameters and to combine performance measures for display parameters by creating efficiency metrics.

RESEARCH OBJECTIVE

The goal of this research was to develop and validate the proposed PAM approach. To validate this approach, the functional relationships between the display parameters and performance were established with a visually guided operator task. When designing a visual system, a number of display parameters are important to consider for ensuring adequate image quality. Clapp (1985) identifies these as resolution, scene detail, brightness (or display luminance), contrast, color, scene dynamics, and field of view. Two of these variables -- display luminance and sharpness (or resolution) were chosen for this study. Display or peak luminance is defined as the "maximum screen light intensity," while sharpness is the "size and shape of picture elements" (Beaton and Farley, 1991).

To determine these functional relationships, the research focused on a single task comprised of one visual event (repeated several times). The relationships between two display parameters (i.e., luminance and sharpness) and two dependent measures (i.e., time-to-target and subjective rating) were determined for two MOVE decision types (i.e., detection and identification).

To create these functional relationships, the data were analyzed using both factorial and efficiency analyses. In the factorial analysis, a two factor Analysis of Variance (ANOVA) was performed for each dependent variable. In the efficiency analysis, simple effects were calculated using individual F-tests for each of the four visual conditions (i.e., luminance-detection, luminance-identification, sharpness-detection, sharpness-identification) for both time-to-

target and rating. The efficiency analysis isolates the effects in the four conditions so that functional relationships may be developed.

After the functional relationships were determined, efficiency values were calculated for the single visual event for both decision types (i.e., detection, identification). Although the PAM model allows for weighting of individual display parameters, these weighting factors are determined by subject matter experts and are not the focus of the present study. For purposes of this research, the weights for all display parameters were assigned equivalent values of 0.5.

METHOD

Participants

Sixty participants (44 males and 16 females) were solicited through local Internet bulletin boards and from Industrial Engineering classes at Virginia Polytechnic Institute and State University. All participants had 20/20 or 20/20 corrected vision and they were paid \$5.00 per hour. Based on the research objective, performance differences for gender were not anticipated nor were they analyzed.

Experimental design

Two experiments were conducted to determine the effects of luminance and sharpness on decision type (i.e., detection and identification). A mixed design was used in each experiment as shown in Table 3.

Table 3.

Experimental design

Experiment 1 - Luminance - 30 participants

IVs Luminance (4 levels: 3.2, 4.5, 8.6, 16.5 cd/m²) - within-subject factor
Decision type (2 levels: detection, identification) - between-subject factor

DVs Time-to-target
Subjective rating (ease of detecting/identifying target on a 9-point scale)

Experiment 2 - Sharpness - 30 participants

IVs Sharpness (3 levels: 0.508, 0.711, 0.864 mm) - within-subject factor
Decision type (2 levels: detection, identification) - between-subject factor

DVs Time-to-target
Subjective rating (ease of detecting/identifying target on a 9-point scale)

Participants were placed randomly in one of four possible conditions: luminance-detection, luminance-identification, sharpness-detection, or sharpness-identification. In the first experiment, four levels of luminance (3.2, 4.5, 8.6, 16.5 cd/m²) were used for both luminance-detection and luminance-identification conditions. In the second experiment, for the sharpness-detection and sharpness-identification conditions, three sharpness levels (0.508, 0.711, 0.864 mm) were used. For each experiment, presentation order was randomized across trials, while conditions were varied according to Table 4.

Table 4.

Conditions and data points per experiment

Experiment 1 - Luminance and decision type

Practice 6 targets/per trial x 4 conditions = 24 targets
 1 rating/per trial x 4 conditions = 4 ratings

Experiment 6 targets/per trial x 4 conditions x 2 trials x 30 participants = 1440
 1 rating/per trial x 4 conditions x 2 trials x 30 participants = 240

Experiment 2 - Sharpness and decision type

Practice 6 targets/per trial x 3 conditions = 18 targets
 1 rating/per trial x 3 conditions = 3 ratings

Experiment 6 targets/per trial x 3 conditions x 2 trials x 30 participants = 1080
 1 rating/per trial x 3 conditions x 2 trials x 30 participants = 180

Materials

For both experiments, four video tapes approximately 130 seconds in length were created. These videos showed what a person might see if they were “walking down the center of a path” while looking straight ahead. The tapes were filmed outdoors during normal daylight conditions and consisted of simple visual targets (circles, squares, and triangles) placed on each side of a 300 meter (m) pathway. The target location or side of the path (left or right) and type of target (circle, square, triangle) were selected randomly for each video. The distance between targets was 50 m.

The targets were created from red/brick-colored poster board and the physical extent of each target was 11.43 cm (four and 1/2 inches). This color provided clearly visible targets in the bright sunlight conditions. The size of the target was determined by placing targets of varying sizes on the path and experimentally determining which size was most visible at a distance of 50 m. Each target was affixed onto thin wooden dowels approximately 15.24 cm (six inches) down from the top of the dowel. The target was placed approximately 15.24 cm (six inches) off the path and 30.48 cm (12”) off the ground in an upright position with the front of the target facing forward.

The target background consisted of the natural habitat (i.e., grass and weeds) along the side of the pathway. The background was selected to allow for a “medium level of visual complexity” since this provides optimal detection and identification in most visual settings (Beaton, 1996, personal communication).

Equipment and testing environment

A 81.28 cm (32") diagonal television set (Model: Sony Trinitron KU-32510) with a 66.04 cm (width) x 48.26 cm (height) (26" x 19") viewing area was used to present the video material. All participants sat on a small couch 2.84 m (112") from the monitor. This viewing distance falls within the normal 6:1 rule of thumb for a television monitor (Beaton, 1996, personal communication). This rule suggests that the optimal viewing distance for any monitor is approximately six times the physical height (48.26 cm or 19") of the monitor. The room lights were turned off while observers watched the video to eliminate glare effects from ambient room lighting. A photometer (Model: Minolta, CS 100) was used to measure peak luminance on the screen while sharpness (or resolution) was determined with a loupe (Model: SDS, 25X5D).

Procedure

Participants were screened with a standard test for far vision (i.e., identification of Landolt rings) to ensure they had 20/20 vision. During this test, they correctly identified 10 visual targets and were only tested to determine if they met the 20/20 standard (i.e., no testing was done to assess better than 20/20 acuity). Participants were given instructions and signed an informed consent form before they participated (see Appendix B).

Before each experiment, participants were given practice trials to allow them to become familiar with the visual conditions, the tasks, and rating scale. For each trial, participants rated ease of detection or identification of targets depending on their given condition (see Tables 5 and 6). A simple rating scale

was used (as described by Sinclair, 1995). In their ratings, participants were asked to select a whole number or one of the mid-points between numbers.

Table 5.

Detection - Subjective rating scale

Please rate the overall trial:

hard to detect	-4	^	-3	^	-2	^	-1	^	0	^	1	^	2	^	3	^	4	easy to detect

Table 6.

Identification - Subjective rating scale

Please rate the overall trial:

hard to identify	-4	^	-3	^	-2	^	-1	^	0	^	1	^	2	^	3	^	4	easy to identify

The participant's task was to detect or identify the visual objects in the scene as they watched a video on the television monitor. During each trial, participants were required to detect or identify six targets while the experimenter recorded the elapsed time for each response with a stop watch. In the detection conditions, participants were required to detect (i.e., "I can see it") simple geometric shapes (i.e., circle, square, triangle), while in the identification task they distinguished between various shapes (i.e., "It is a circle").

At the beginning of each trial, the experimenter pressed the “start” button to start the stopwatch and when a target was identified or detected, the participants responded verbally (e.g., “circle”) and the experimenter pressed the “lap” button and denoted the elapsed time from the start position. Six times were recorded for each trial. Following each trial, participants rated ease of detection or identification of targets depending on their given condition.

Accuracy was stressed to the observers during task completion so that differences in performance could be measured based on time-to-target. During both experiments, as visual angle increased, time-to-target decreased, so that as participants got closer to the target, the objects appeared larger on the screen (see Figure 10). Note that the time-to-target scale ranged from 30 to 0 for the true experimental condition. The distance from the observer to the screen was fixed at 2.84 meters. Viewing angle for the targets was calculated using the height of the target on the screen and the time-to-target data points shown in Figure 10. The time-to-target data points represent averaged values across the four experimental video tapes.

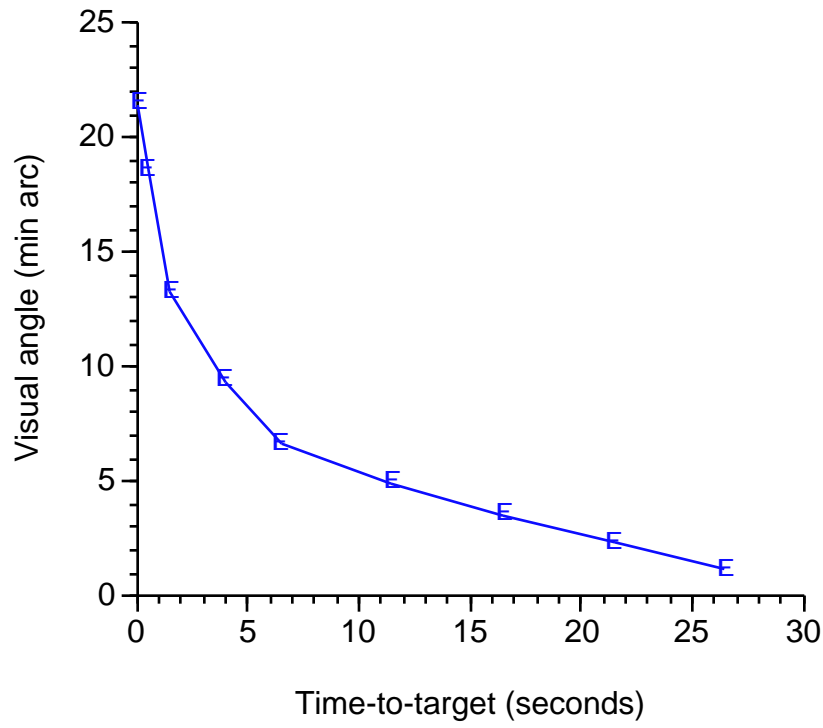


Figure 10. Visual angle of targets viewed on screen as a function of time-to-target.

RESULTS

The data collected in the study were subjected to analyses of variance (ANOVA). In Experiment 1, the effect of luminance and decision type on time-to-target and subjective rating is determined while the effect of sharpness and decision type on time-to-target and subjective rating is calculated for Experiment 2 (see Table 7).

Table 7.

Factorial analysis

Experiment 1 Luminance and decision type	Experiment 2 Sharpness and decision type
ANOVA 1 Time-to-target Luminance - 4 levels Decision type - 2 levels	ANOVA 3 Time-to-target Sharpness - 3 levels Decision type - 2 levels
ANOVA 2 Subjective rating Luminance - 4 levels Decision type - 2 levels	ANOVA 4 Subjective rating Sharpness - 3 levels Decision type - 2 levels

However, to create an efficiency model of vision, simple effects are later calculated with separate F-tests for each of the visual conditions (see Table 8). In these planned F-tests, the effect of luminance or sharpness is independently considered for each decision type. To ensure that these effects are independent, individual error terms (i.e., not pooled error terms) are used for each condition.

Table 8.

Efficiency analysis by condition

Detection conditions	Identification conditions
1. Luminance-Detection (time-to-target) 2. Luminance-Detection (rating)	5. Luminance-Identification (time-to-target) 6. Luminance-Identification (rating)
3. Sharpness-Detection (time-to-target) 4. Sharpness-Detection (rating)	7. Sharpness-Identification (time-to-target) 8. Sharpness-Identification (rating)

Factorial analysis

The effects of visual parameters on performance measures were evaluated in two separate experiments. In both experiments, two dependent measures (i.e., time-to-target and subjective rating) were taken; however, separate ANOVAs were completed since an unequal number of observations was recorded for each measure. Since the main effects and interactions were not the focal point of the study, these are only briefly discussed. The results of each ANOVA (see Table 7) were compared to Geisser-Greenhouse critical F-values as suggested by Keppel (1991) (A detailed discussion of Geisser-Greenhouse F-values is provided in Appendix C).

For subjective rating, the scores were converted from a scale of -4 to 4 to a scale of 1-9 by adding 5 to each score, so that statistical analysis could be performed. The error bars shown in the figures denote ± 1 standard error from the mean.

Experiment 1- Luminance and decision type.

In the first experiment, four levels of luminance (3.2, 4.5, 8.6, and 16.5 cd/m^2) were combined factorially with two levels of decision type (detection or identification), where luminance was the within-subjects factor and decision type was manipulated between-subjects.

For time-to-target, the luminance x detection interaction ($F(3, 1381) = 24.1, p < 0.001$) and the main effects of luminance ($F(3, 1381) = 102.39, p < 0.001$) and decision type ($F(1, 28) = 388.6, p < 0.001$) were significant (see Figures 11-13). Summary ANOVA tables and means are shown in Appendix D.

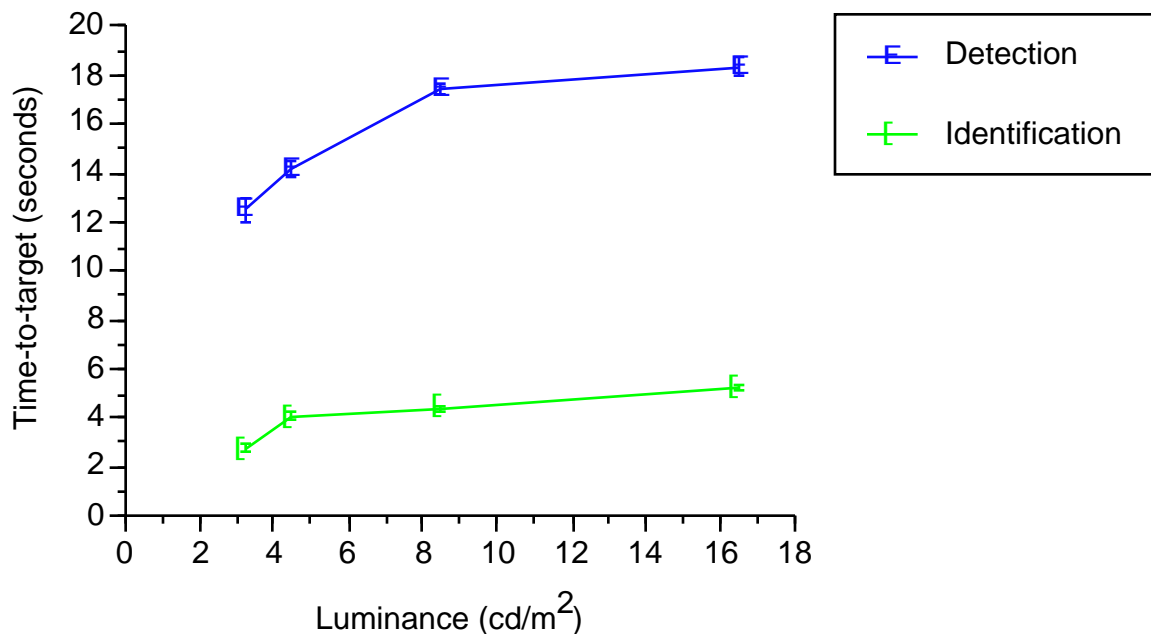


Figure 11. Interaction between luminance and decision type for time-to-target in Experiment 1 (Error bars represent ± 1 Standard Error of the Mean (SEM)).

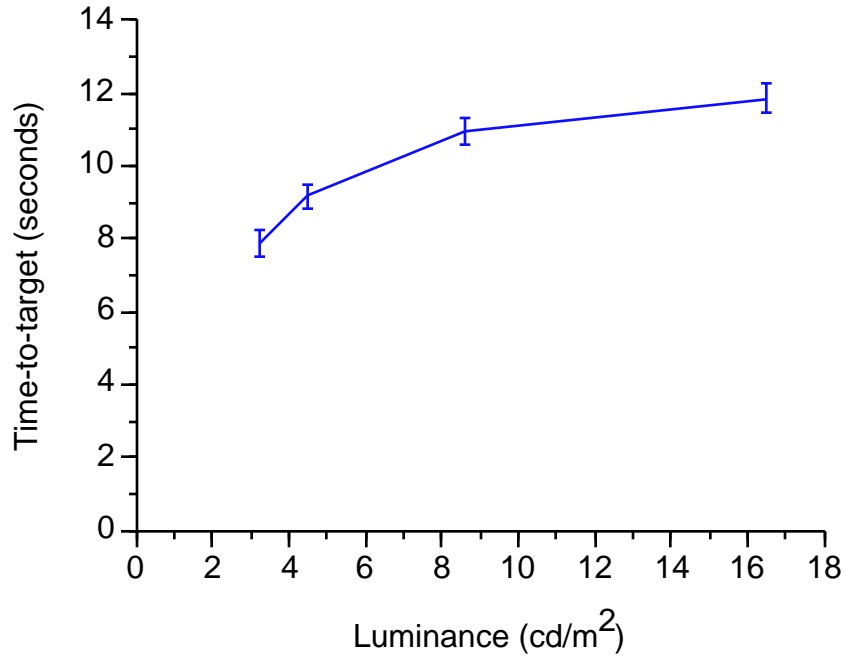


Figure 12. Main effect of luminance for time-to-target in Experiment 1 (Error bars represent ± 1 SEM).

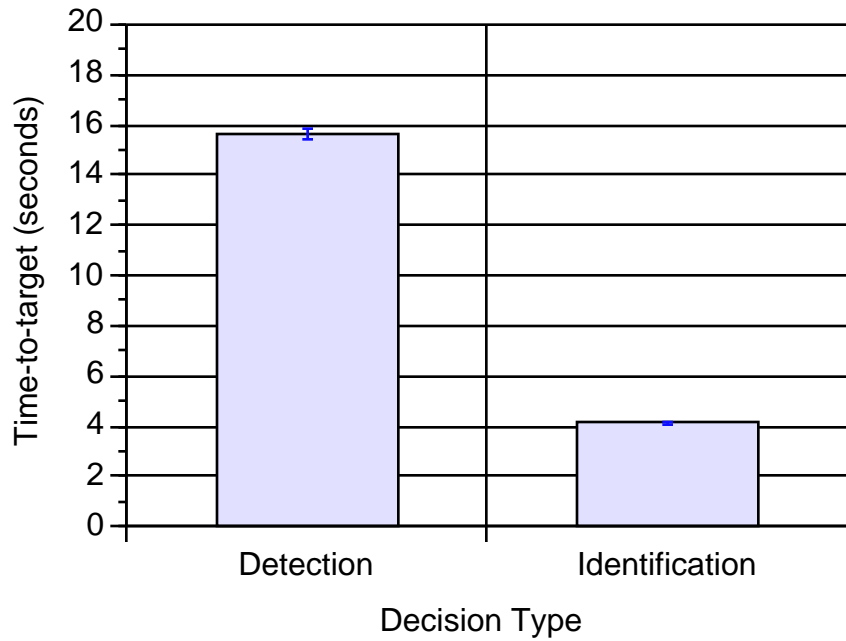


Figure 13. Main effect of decision type for time-to-target in Experiment 1 (Error bars represent ± 1 SEM).

For subjective ratings, the luminance x detection interaction ($F(3, 204) = 1.82, p = 0.145$) was not significant, but the main effects of luminance ($F(3, 204) = 189.19, p < .001$) and decision type ($F(1, 28) = 5.56, p = 0.026$) were significant (Figures 14-16).

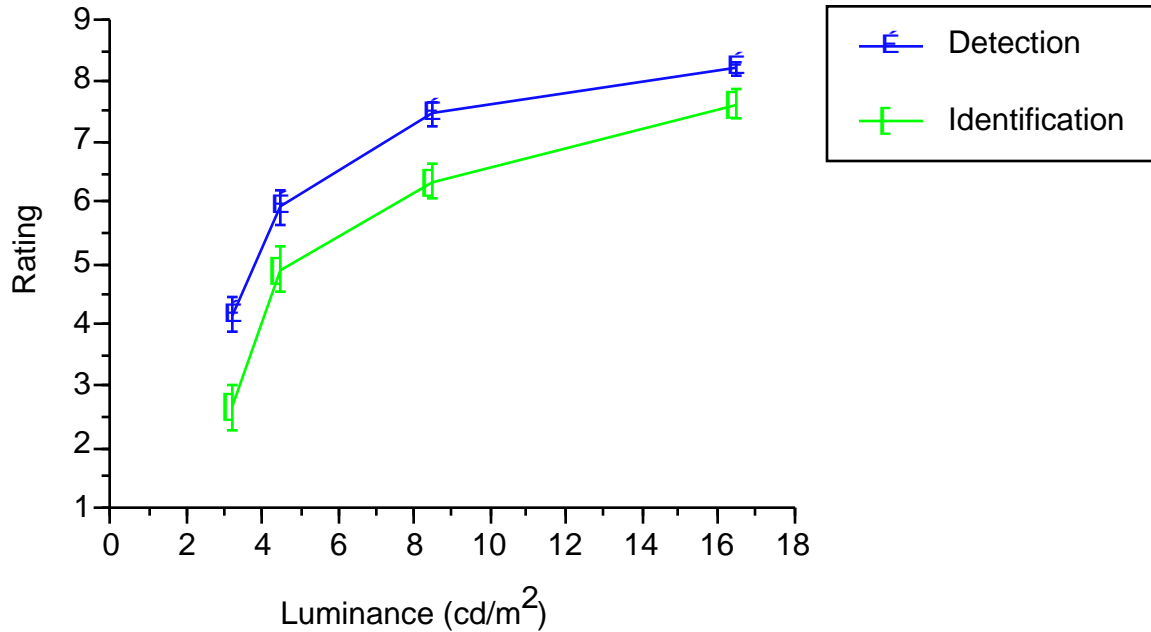


Figure 14. Interaction between luminance and decision type for rating in Experiment 1 (Error bars represent ± 1 SEM).

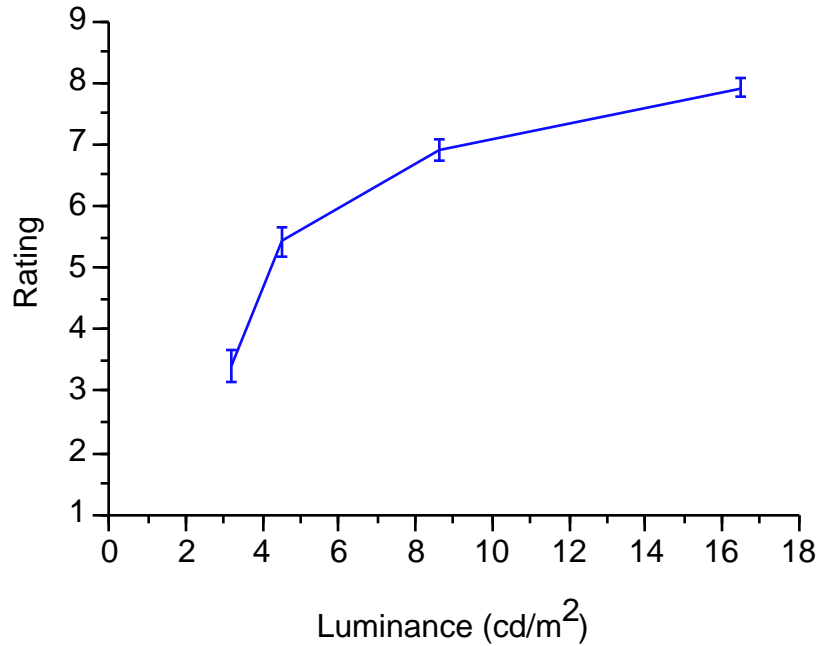


Figure 15. Main effect of luminance for rating in Experiment 1 (Error bars represent ± 1 SEM).

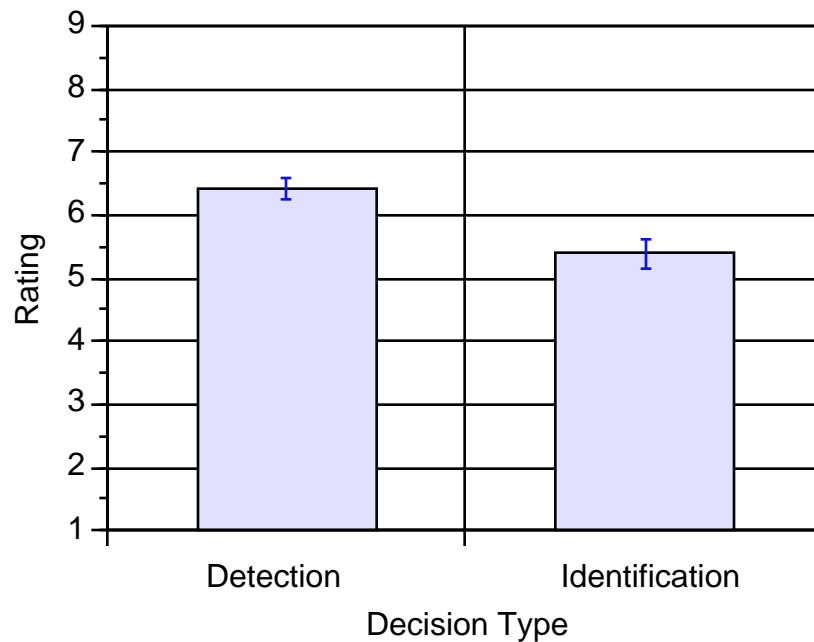


Figure 16. Main effect of decision type for rating in Experiment 1 (Error bars represent ± 1 SEM).

Experiment 2 - Sharpness and decision type.

In the second experiment, three sharpness levels (0.864, 0.711, 0.508 mm) were combined factorially with two decision types (detection or identification). Sharpness was varied within-subjects, while detection was the between-subjects variable.

For time-to-target, the sharpness x decision type interaction was non-significant ($F(2, 1046) = 1.5, p = 0.224$). But, the main effects of sharpness ($F(2, 1046) = 9.15, p < 0.001$) and decision type ($F(1, 28) = 437.56, p < 0.001$) were significant (see Figures 17-19). Summary ANOVA tables and means are shown in Appendix E.

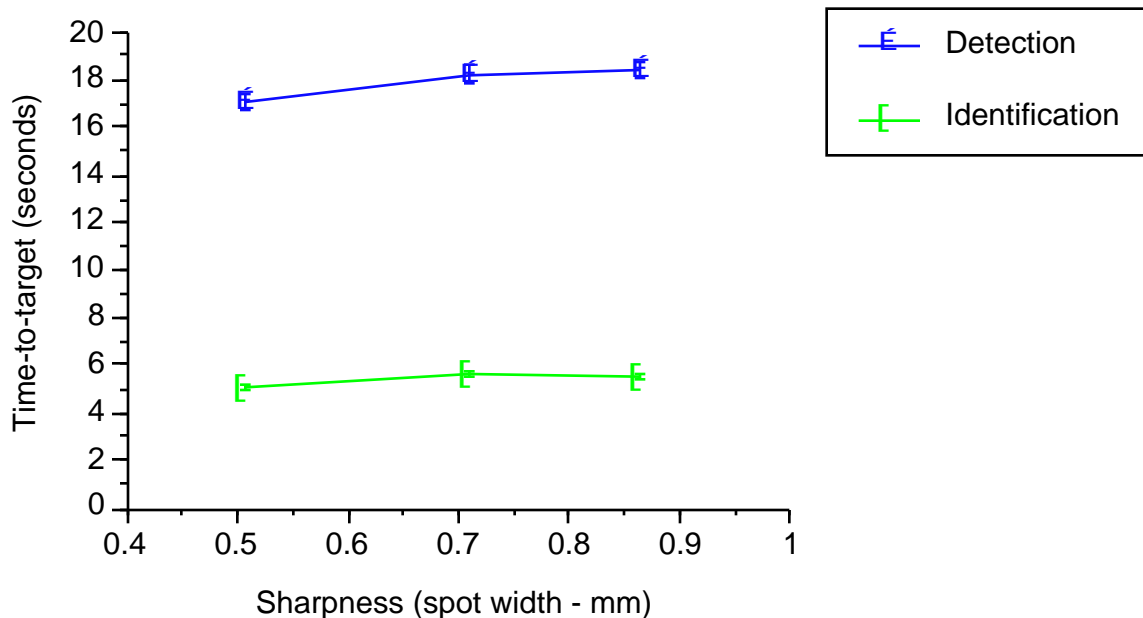


Figure 17. Interaction between sharpness and decision type for time-to-target in Experiment 2 (Error bars represent ± 1 SEM).

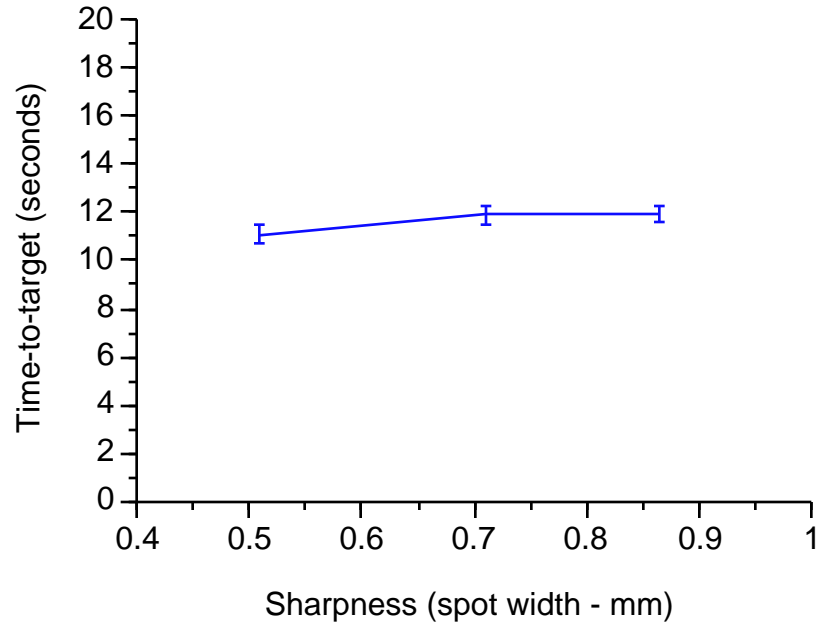


Figure 18. Main effect of sharpness for time-to-target in Experiment 2 (Error bars represent ± 1 SEM).

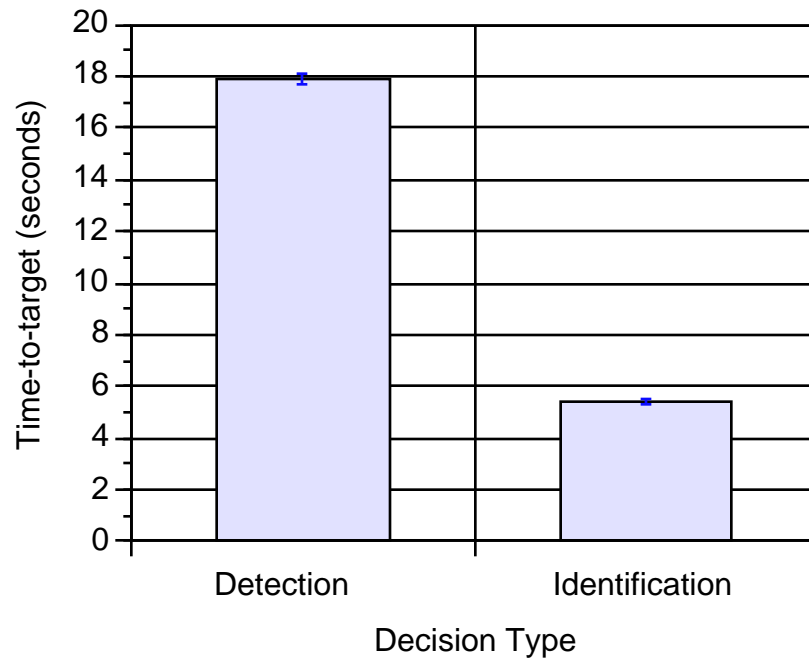


Figure 19. Main effect of decision type for time-to-target in Experiment 2 (Error bars represent ± 1 SEM).

For subjective rating, sharpness x decision type ($F(2, 146) = 0.05, p = 0.952$), sharpness ($F(2, 146) = 1.94, p = 0.147$), and decision type ($F(1, 28) = 0.19, p = 0.67$), were all non-significant (see Figures 20-22).

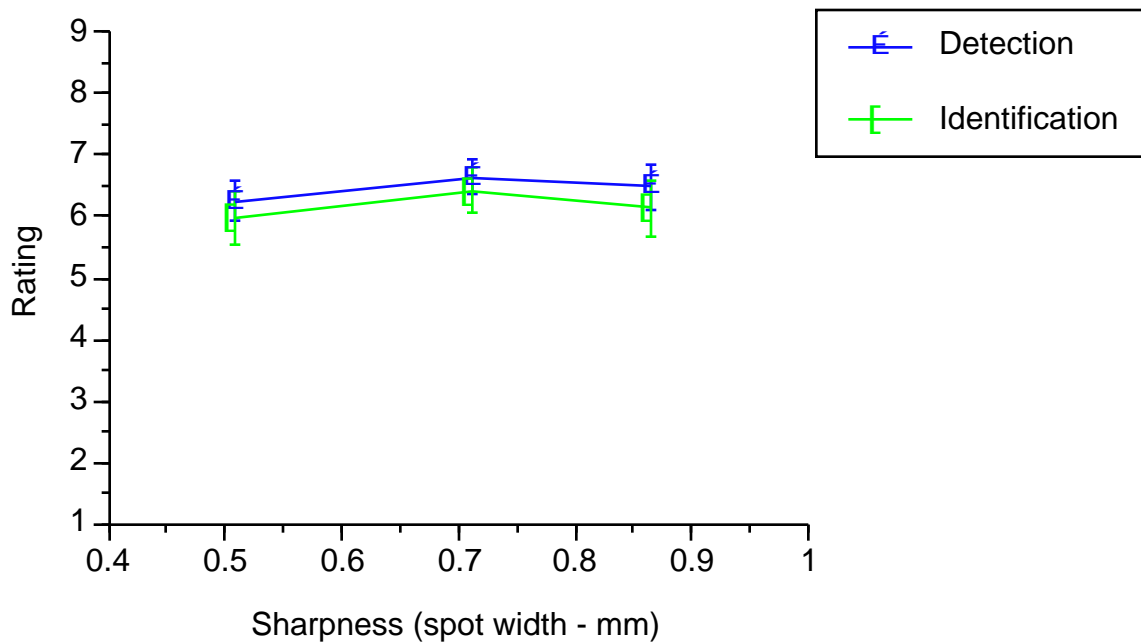


Figure 20. Interaction between sharpness and decision type for rating in Experiment 2 (Error bars represent ± 1 SEM).

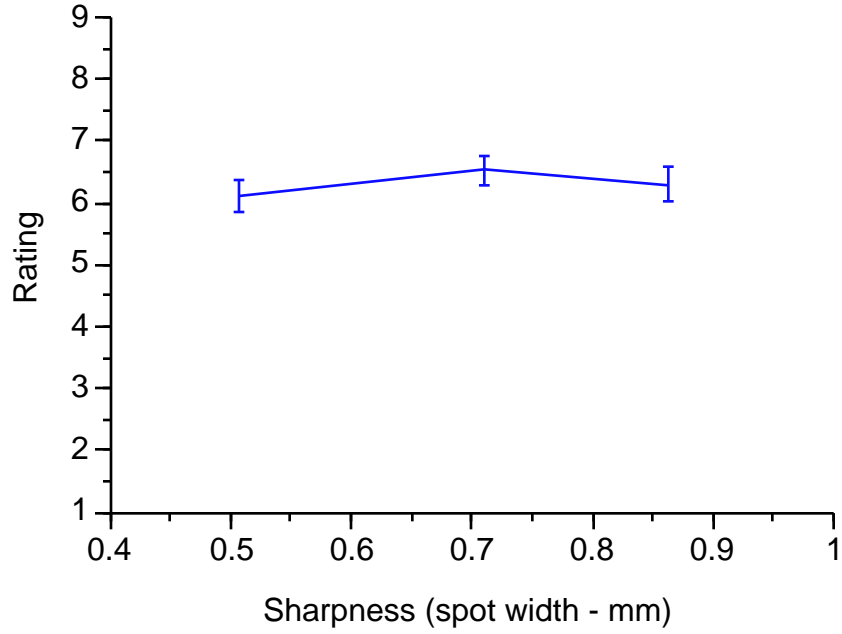


Figure 21. Main effect of sharpness for rating in Experiment 2 (Error bars represent ± 1 SEM).

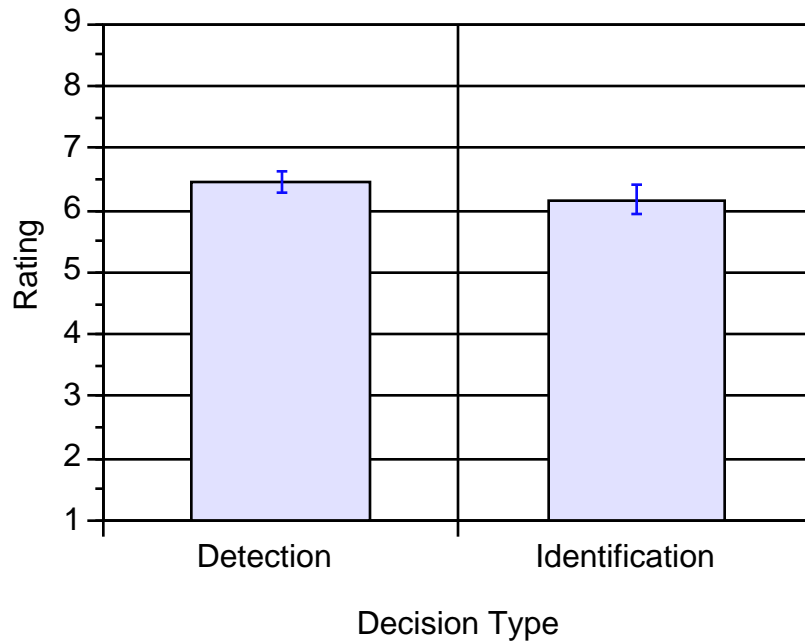


Figure 22. Main effect of decision type for rating in Experiment 2 (Error bars represent ± 1 SEM).

Efficiency analysis

In efficiency analysis, relationships between display parameters (e.g., luminance, sharpness) and their associated dependent variables (e.g., time-to-target, subjective rating) were developed for the decision types (e.g., detection, identification). However, before relationships can be determined, it must be shown that the dependent variable significantly changes as a function of the independent variable (i.e., monotonically increase or decrease). To determine this, F-tests were performed for each of the experimental conditions (i.e., luminance-detection, luminance-identification, sharpness-detection, and sharpness-identification) for time-to-target and rating. Efficiency curves were generated when statistically significant differences were found within a condition.

To empirically derive efficiency scores from experimental data, one starts by averaging the performance scores (e.g., time-to-target) for each level of the independent variable (e.g., luminance) to develop a distribution of sample means. To create a functional relationship for the data, the differences between these sample means must be statistically different (not due to chance) for at least two levels of the independent variable. If significant differences do not exist for at least two levels of the independent variable, then the means or scores on the dependent variable are essentially equivalent. This creates a straight-line, zero-slope function showing no relationship between the independent variable (e.g., luminance) and dependent variable (e.g., time-to-

target). Thus, before efficiency curves are created empirically, an elementary determination is required for an individual parameter.

Efficiency curve generation.

First, a distribution of sample means is created for each condition of interest. For example, in the luminance-detection condition the means are 12.52, 14.14, 17.40, and 18.35. This small distribution of sample means has a mean of 15.60 with a standard deviation of 2.74. A cumulative probability distribution is created based on the mean and standard deviation for this distribution of sample means. The probabilities associated with these raw scores are 0.130, 0.296, 0.744, and 0.842 (see Figure 23).

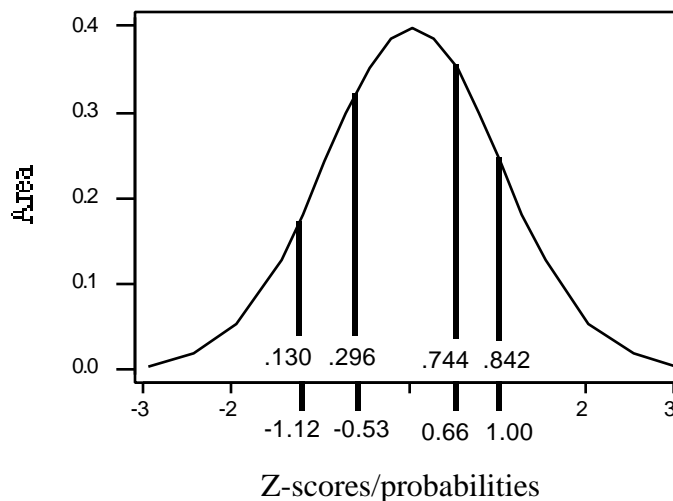


Figure 23. Z-scores and associated probabilities for time-to-target in the luminance-detection condition.

Next the data are fitted to an ogive function of the form previously described. The luminance and sharpness data were fit to a hyperbolic tangent

function using the quasi-Newton method (MacCurveFit software program, Raner, 1996). The function is shown below:

$$f(x) = a \frac{e^{b(x+c)} - 1}{e^{b(x+c)} + 1} + d \quad (\text{Eq. 4})$$

in which, a , b , c , and d are constants determined by the curve-fitting equation.

The MacCurveFit software computes a minimal sum of squares error (SSE) and R^2 to find the best curve fit for the data (see Appendix F for discussion of how SSE and R^2 are computed).

The hyperbolic tangent function was selected since it approximates the stimulus-response psychometric function previously described; that is, it approximates an ogive curve. While the standardized data are thought to follow a general ogive function, other functions also may fit the data. In Eq 4., a is $1/2$ the range of the y - value, b is proportional to the steepness of the middle of the sigmoid curve, c determines the horizontal placement of the sigmoid curve and d determines the vertical placement of the sigmoid curve (for a detailed description of this function - see Appendix G). To calculate an efficiency curve, the value of a and d are set to be equivalent. For an ideal response function, $a = d = 0.5$, that is, the function varies from 0-1. However, for the experimental data, the response function had a limited range for luminance (i.e., $y = 0-0.9$) and sharpness (i.e., $y = 0-0.8$). Thus, the values for a and d are set at 0.45 for the luminance conditions and 0.40 for the sharpness conditions.

Detection conditions.

In the luminance-detection condition, the effect of luminance on time-to-target ($F(3, 42) = 96.08, p < 0.001$) and rating condition ($F(3, 42) = 83.28, p < 0.001$) was significant. Univariate F tables and means are shown in Appendix H. In the sharpness-detection condition, the effect of sharpness on time-to-target was significant ($F(2, 28) = 4.35, p = 0.023$), but was non-significant for rating ($F(2, 28) = 0.82, p = 0.450$). Thus, efficiency curves were not generated for subjective rating under the sharpness-detection condition. Univariate F tables and means are shown in Appendix I.

The luminance-detection curves are shown in Figures 24 and 25, while the sharpness-detection curve is shown in Figure 26. To fit the sharpness data to the ogive function, the scale for sharpness was inverted so that higher numbers indicate greater degrees of visual blur (see Figures 26, 29). The R^2 , SSE, and regression weights for a given condition are shown at the top of each figure.

Luminance-detection efficiency curves.

$R^2 = 0.988$ $SSE = 0.004$ $a=d = 0.45$ $b = 0.590$ $c = -5.871$

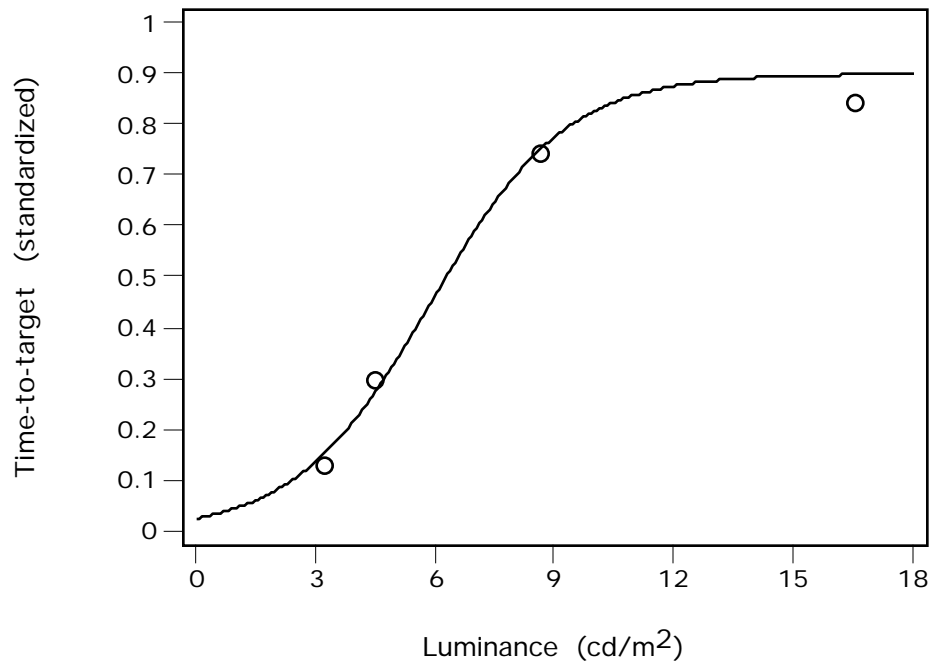


Figure 24. Functional relationship for luminance and time-to-target - Luminance-Detection condition (Efficiency ₁).

$R^2 = 0.947$ $SSE = 0.017$ $a=d = 0.45$ $b = 0.534$ $c = -5.716$

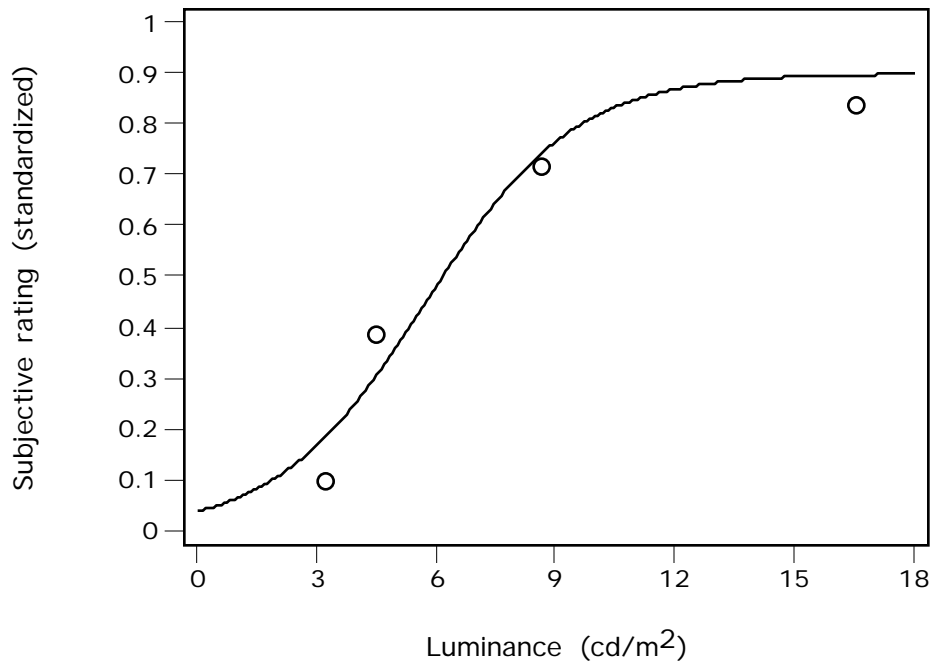


Figure 25. Functional relationship for luminance and subjective rating - Luminance-Detection condition (Efficiency ₂).

Sharpness-detection efficiency curve.

$$R^2 = 0.931 \quad SSE = 0.016 \quad a=d = 0.40 \quad b = 18.964 \quad c = -0.526$$

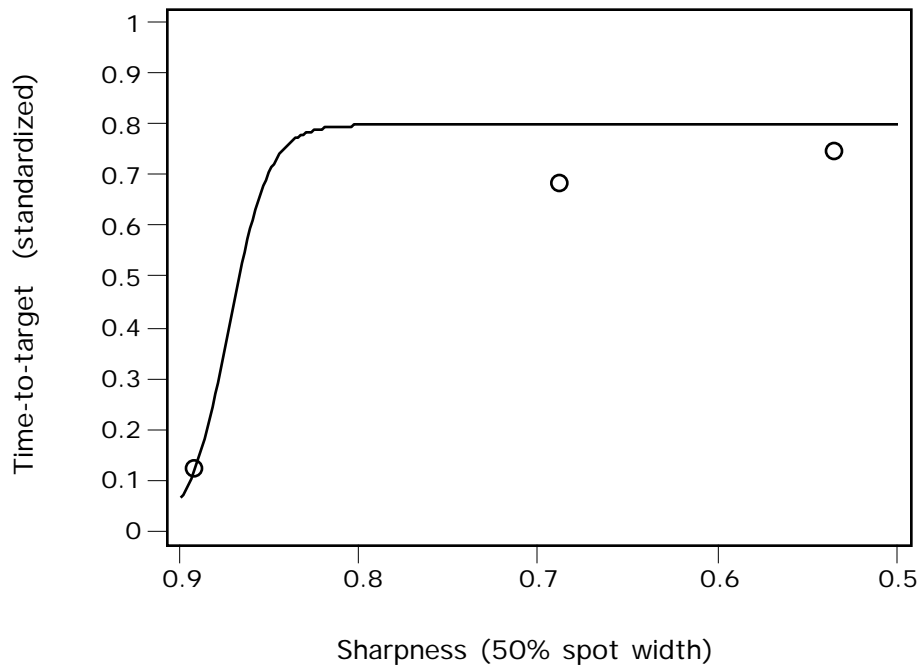


Figure 26. Functional relationship for the sharpness and time-to-target - Sharpness-Detection condition (Efficiency η_3).

Identification conditions .

In the luminance-identification condition, the effect of luminance on time-to-target ($F(3, 42) = 77.82, p < 0.001$) and rating ($F(3, 42) = 64.60, p < 0.001$) was significant. Univariate F tables and means are shown in Appendix J. In the sharpness-identification condition, the effect of sharpness on time-to-target was significant ($F(2, 28) = 3.64, p = 0.039$), but was non-significant for rating ($F(2, 28) = 1.44, p = 0.254$). As a result, efficiency curves are not generated for rating in the sharpness-identification condition. Univariate F tables and means are shown in Appendix K. Figures 27 and 28 show the luminance-identification curves while the Figure 29 shows the sharpness-identification curve.

Luminance-identification efficiency curves.

$$R^2 = 0.852 \quad SSE = 0.045 \quad a=d = 0.45 \quad b = 0.342 \quad c = -6.182$$

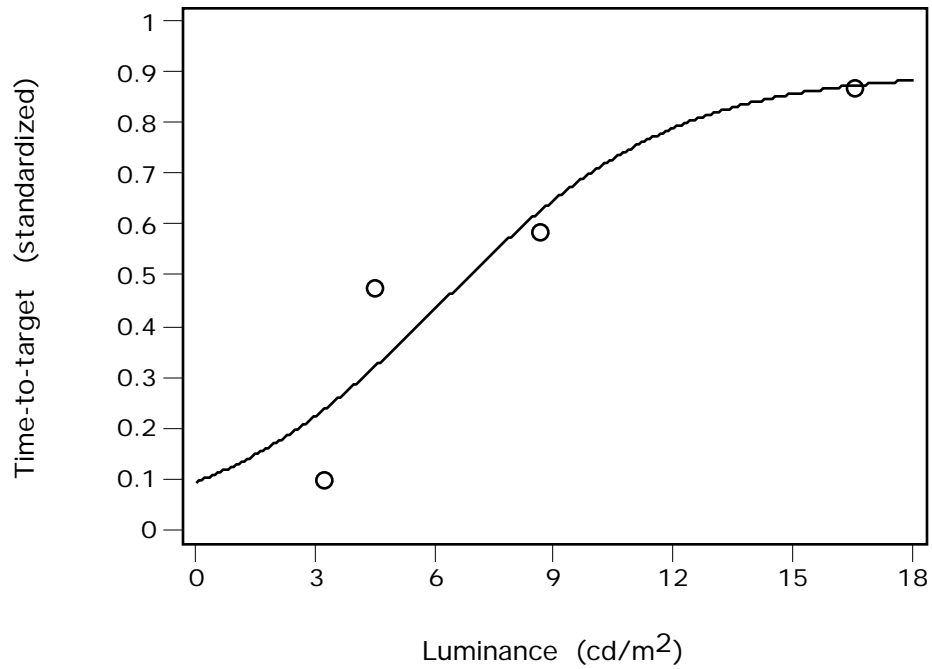


Figure 27. Functional relationship for the luminance and time-to-target - Luminance-Identification condition (Efficiency ₄).

$R^2 = 0.927$ $SSE = 0.023$ $a=d = 0.45$ $b = .458$ $c = -5.887$

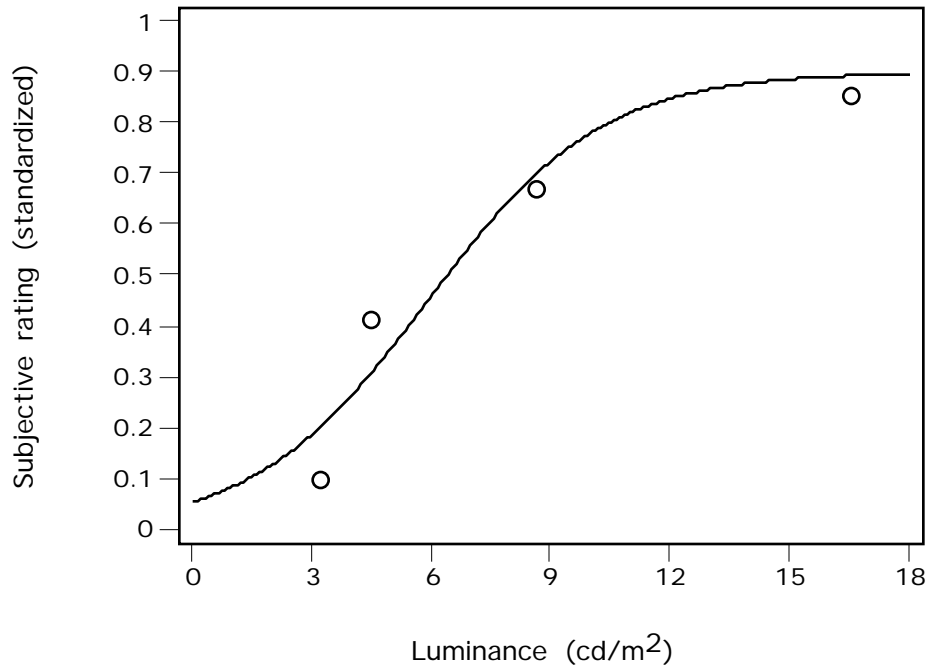


Figure 28. Functional relationship for luminance and subjective rating - Luminance-Identification condition (Efficiency ₅).

Sharpness-identification efficiency curve.

$$R^2 = 0.910 \quad SSE = 0.021 \quad a=d = 0.40 \quad b = 66.135 \quad c = -0.533$$

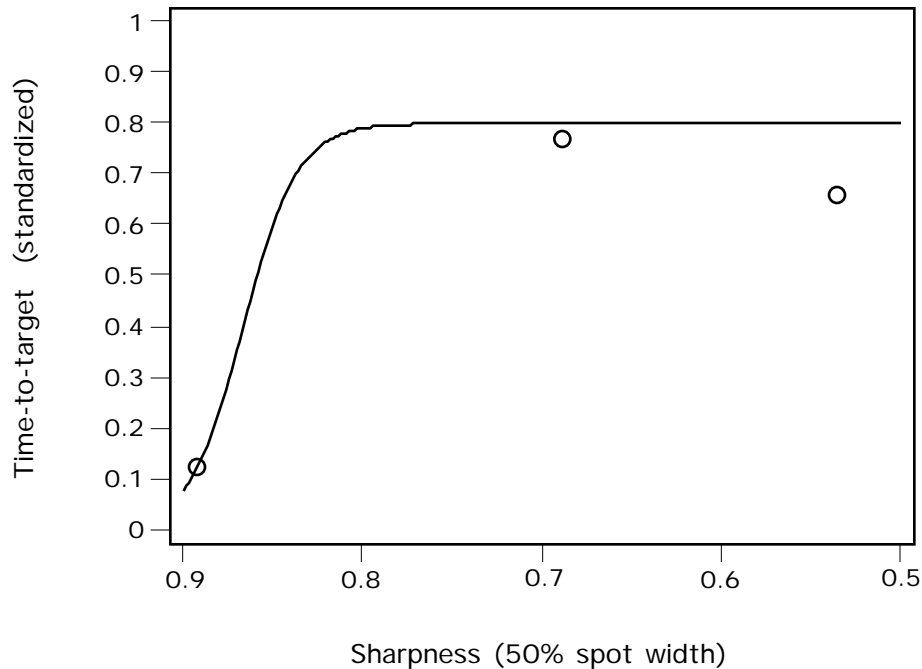


Figure 29. Functional relationship for sharpness and time-to-target - Sharpness-Identification condition (Efficiency ϵ_6)

To summarize, for the luminance-detection conditions, efficiency ϵ_1 represents time-to-target (Figure 24) while efficiency ϵ_2 shows subjective rating (Figure 25). For the sharpness-detection condition, efficiency ϵ_3 is calculated for time-to-target (Figure 26). For the luminance-identification conditions, efficiency ϵ_4 depicts time-to-target (Figure 27) and efficiency ϵ_5 is calculated for subjective rating (Figure 28). Finally, for the sharpness-identification conditions, efficiency ϵ_6 represents time-to-target (Figure 29).

DISCUSSION AND CONCLUSIONS

Effects - Factorial design

The results of this research show that time-to-target and subjective ratings change as a function of luminance. Although significant differences were observed for decision type, these are expected for the time-to-target variable. Participants detected a target quickly, but then they took much longer to identify a target (i.e., circle, square, or triangle). Participants also rated the identification task as harder than the detection task (5.39 vs. 6.44); where lower scores indicate greater difficulty in detecting or identifying. The ratings increased as a function of luminance varying from a low of 3.43 to a high of 7.91, which was expected since it is easier to detect or identify objects when they are more visible.

Even though the main effect of sharpness was statistically significant, the changes in performance for the sharpness variable were relatively small in magnitude. One reason for this result was the restriction of range for sharpness. An “off-the-shelf” monitor with a limited range of adjustability was used in this work. The sharpness adjustment on this monitor is unlikely to influence ease of detection or identification for users. However, the adjustment is probably useful for fine-tuning an image on the screen, which is more in line with its intended purpose.

PAM - Application in current study

The PAM technique was applied to the current experimental task as shown in Figure 30. Remember that the experimental video had one unique visual event, which was repeated six times. Thus, only one EDA block is shown in the model.

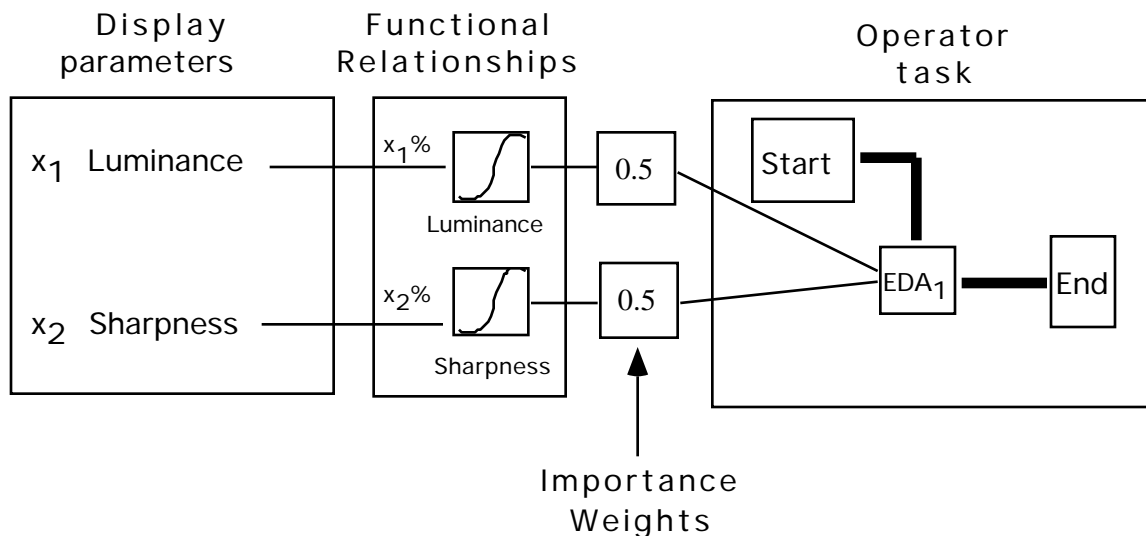
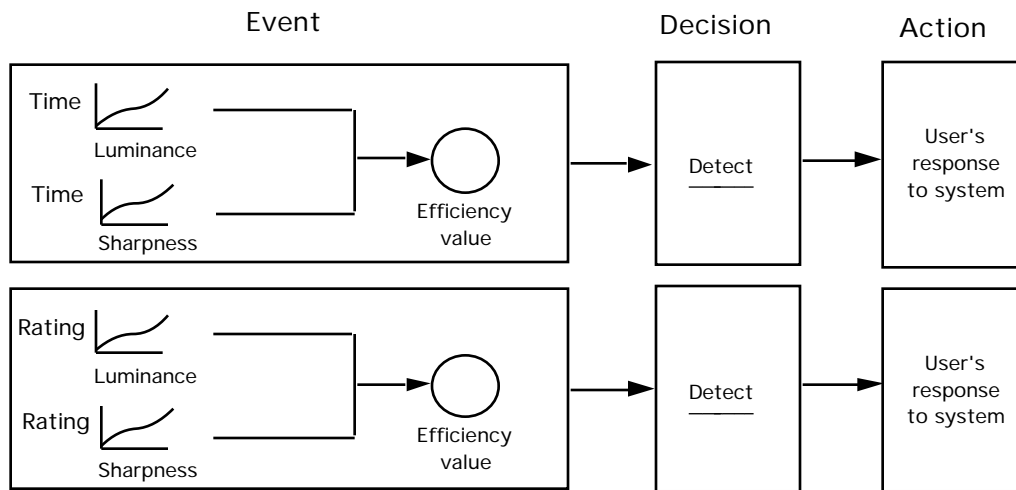


Figure 30. PAM applied to current research.

PAM separates out efficiency scores based on decision types and dependent measures. In theory, these measures are additive across parameters within a selected decision type and dependent measure. In the case of the experimental data, the expectation was to demonstrate this concept using luminance-detection and sharpness-detection values for each of the dependent measures (i.e., time-to-target and subjective rating). However, since efficiency analysis under the sharpness conditions was not possible for subjective rating (i.e., due to non-significant differences between means), it will not be discussed for these conditions.

Based on this research, efficiency metrics may be derived. For each MOVE decision type (i.e., detect and identify), a metric is derived for time-to-target and subjective rating (see Figure 31). For each occurrence of the visual event, time-to-target was recorded. However, the subjective rating was determined at the end of the video or across six replications of the visual event.

Decision type: Detect



Decision type: Identify

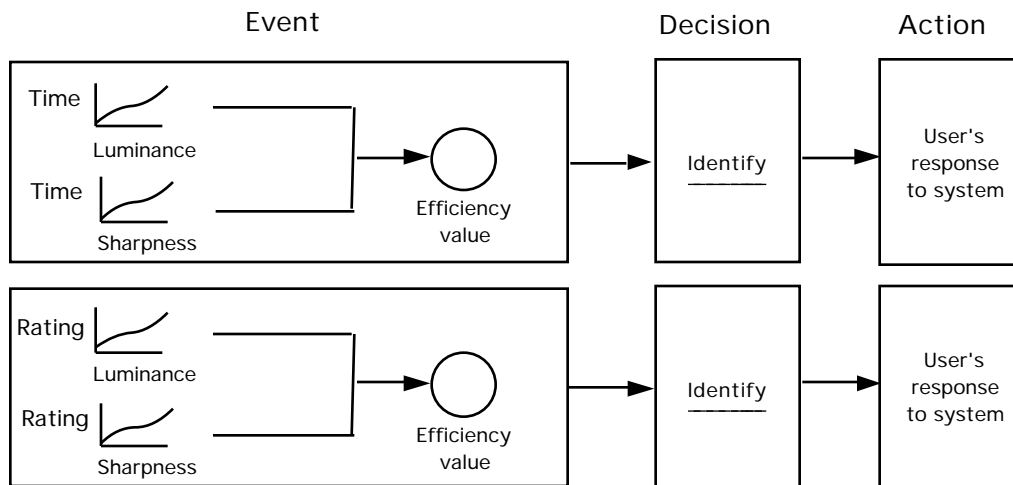


Figure 31. Efficiency metrics for current research

The efficiency curves generated for luminance-detection, sharpness-detection, luminance-identification and sharpness-identification (see Figures 24-29) show that for a given visual parameter, performance can be represented by a normal probability distribution. In these graphs, an efficiency score distribution was created from the raw means for a given level of a visual parameter. In general, the regression equation based on the hyperbolic tangent function provided a good fit as shown by the R^2 and the sum of squares error values for each curve.

The data from these curves can be combined according to the previous efficiency equations to determine the combined effect of sharpness and luminance for each type of visual condition. For a visual parameter, a value of interest is selected and the corresponding efficiency score then is determined using PAM efficiency metrics. For the visual event in the experiment, suppose that a luminance value of 9 cd/m^2 and a spot width of 0.7 mm is desired, where each parameter has a weighting factor = 0.5. For a detection decision, the combined effects of luminance and sharpness would be calculated using Figures 24 and 26. In Figure 24, 9 cd/m^2 corresponds to a 75% performance level, while 0.7 mm corresponds to a 80% performance level (see Figure 26). The efficiency computations are:

$$EDA_{\text{efficiency}} = [(.75)(0.5) + (.80)(.05)] = 0.775$$

For an identification decision, the effects of combined effects of luminance and sharpness are calculated in a similar manner. These computations

demonstrate the utility of the PAM for a selected decision type across different parameters.

A couple of issues arise in the creation of these models. First, the proficiency curves created have a limited number of data points as the luminance curves are based on four sample means while the sharpness curves are based on three. This may not provide an accurate picture of the true functional relationship. To empirically derive well-defined functional relationships, a researcher would need to use 6-8 levels of an independent variable (e.g., two extreme points and five points in-between) to develop a reasonable curve.

A related issue is the range across which the sample means are gathered. For example, in the case of luminance, two extreme points were clearly determined; however, in the case of sharpness, it was not as clear. From a performance view, a significant difference was discovered among sharpness conditions, but no difference was found in the subjective data. Defining these extremes requires a thorough understanding of the visual parameter. For example, performance seems to level off for some visual parameters (Beaton, 1997, personal communication). That is, for certain parameters, performance plateaus at a given level (e.g., 90%). For this example, the true range of variation is only from 0-90 and thus 90 is equated with the maximum value for standardized performance.

Comparison with luminance-detection model

Although efficiency may be determined for any dependent variable, such as time-to-target or rating, it is desirable to understand the results of a single study within a larger context of established research. This allows greater understanding of a variable's effect on performance but it requires common dependent measures (e.g., visual angle) across studies to be compared. For this reason, the luminance and sharpness data from this research are presented in terms of visual angle. The luminance data are contrasted with the Moon and Spencer (1944) luminance-detection model, while the sharpness data highlight the differences between the luminance and sharpness conditions.

The data in Figure 32 represent luminance as a function of visual angle. The mean time-to-target data for the four luminance levels (3.2, 4.5, 8.6, and 16.5 cd/m^2) were converted to visual angles using the curve in Figure 10. Figure 32 shows the relationship between visual angle and time-to-target as calculated for experiments 1 and 2. Separate curves are shown for luminance-detection and luminance-identification conditions. These data show that target size must be larger for participants to be able to accomplish the identification task. Also, as luminance is increased, visual angle decreases for both tasks. This indicates that more luminance is required to detect a smaller target.

Plots of visual angle as a function of sharpness (see Figure 33) show differences in magnitude for decision type (although the differences between means for identification or detection are not significant).

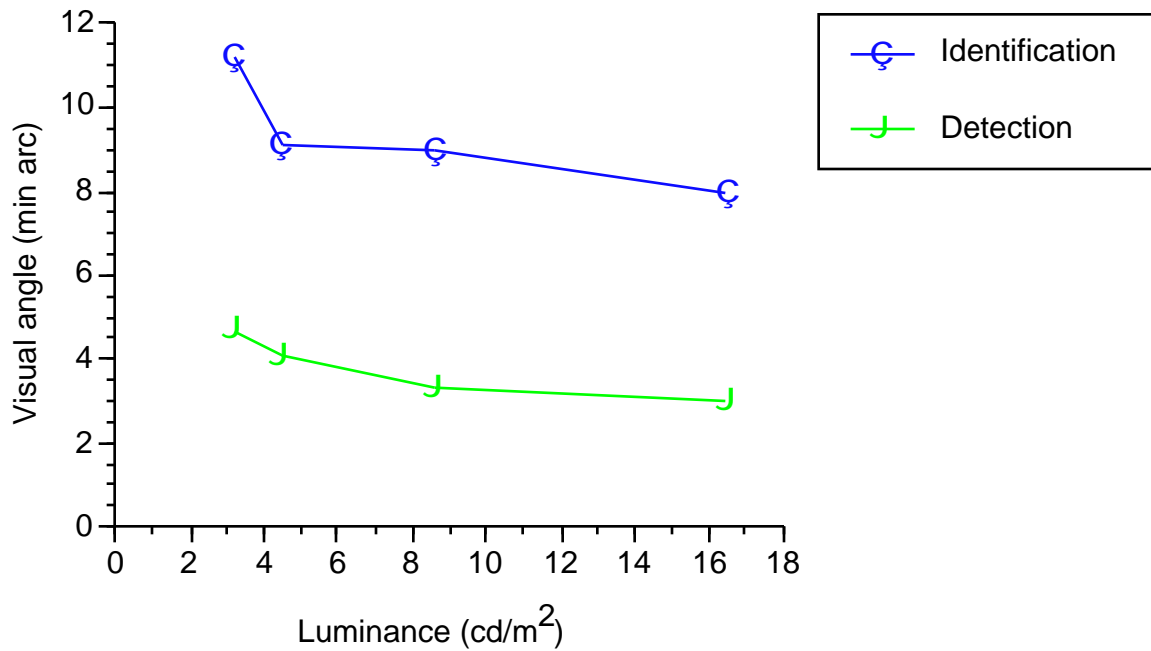


Figure 32. Luminance as a function of visual angle.

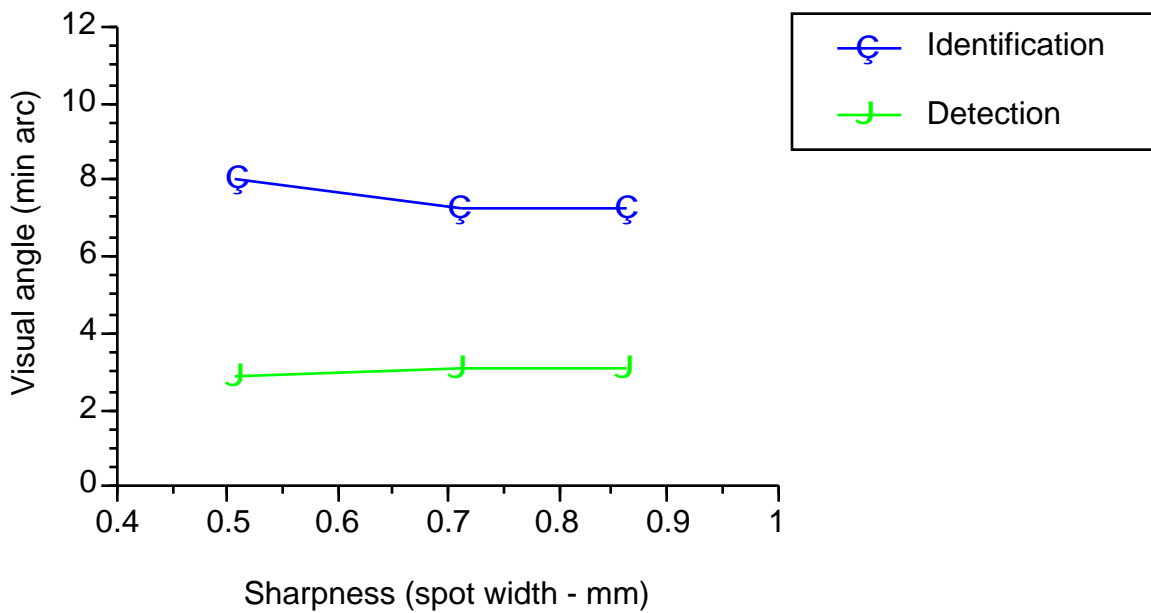


Figure 33. Sharpness as a function of visual angle.

The magnitude of differences between decision types was similar in both the sharpness and luminance figures.

The luminance points for the detection task are plotted with the luminance-detection model generated previously (see Figure 34). The experimental data and the previous luminance-detection model (Moon and Spencer, 1944; Chapanis, 1949) appear to have similar underlying functions. The differences between these curves seem to be a function of magnitude and not direction.

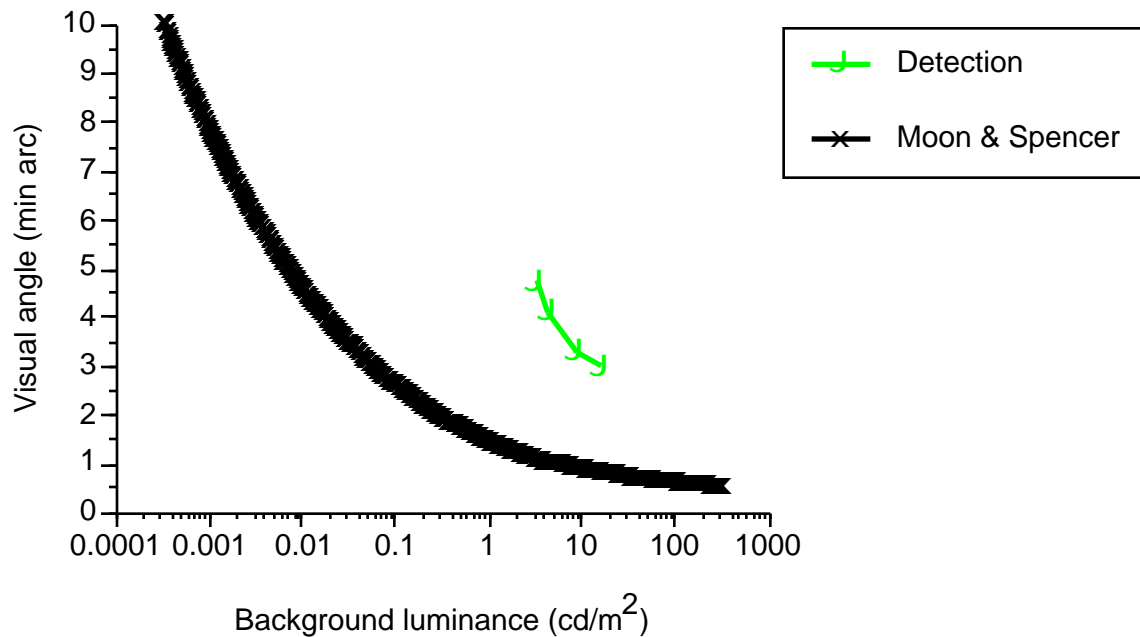


Figure 34. Plot of detection data on Moon and Spencer (1944) luminance curve.

The Moon and Spencer (1944) data were for visual acuity tasks, in which observers were required to distinguish fine detail across a wide range of luminances. The detection and identification tasks in the present study were similar to the visual acuity tasks reported by Moon and Spencer. Detection and identification, by definition, required participants to resolve fine details on the

screen. Also, luminance was varied across the full range of visibility for both data sets. In the current study, participants had great difficulty detecting and identifying the targets at the lower level of the luminance scale (i.e., 3.2 cd/m²) and had virtually no difficulty seeing the targets at the higher luminance levels (i.e., 16.5 cd/m²).

However, clear methodological differences are found between the current study and those reported by Moon and Spencer (1944). In the previous studies, the data reflect a minimum threshold under ideal lighting conditions using still images projected onto a screen. The current study used lower fidelity video images from a Cathode Ray Tube (CRT) with flicker and lower sharpness. These results are not surprising since CRT images have lower image quality when compared to photographs and projected images. These factors account for a shift in the minimum perceptible threshold or vertical offset in the current study as shown in Figure 34.

This task effect is illustrated when the identification curve is added to the detection and Moon and Spencer (1944) data (see Figure 35). There appears to be a family of curves dependent on visual angle, each representing a unique task. The Moon and Spencer data represent the visual angle needed for fine visual acuity tasks under ideal visual conditions. The visual angle increases for a CRT video detection task and an even greater visual angle is needed for a CRT video identification task.

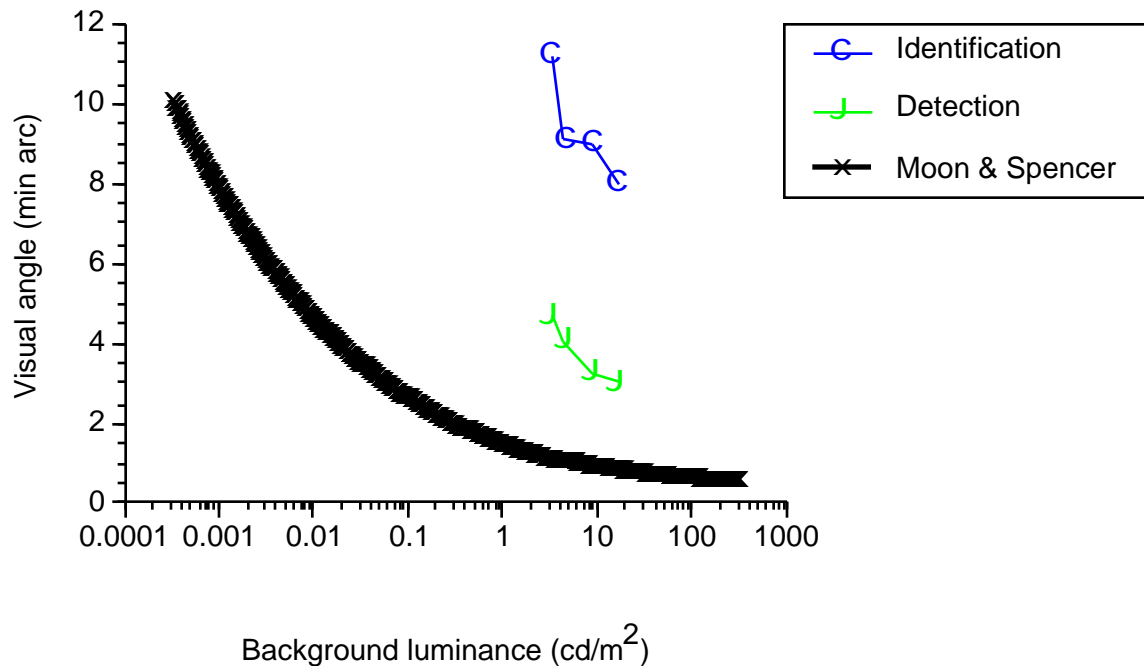


Figure 35. Plot of detection and identification data on Moon and Spencer (1944) luminance curve.

Utility of PAM

The PAM goes beyond “image quality” metrics and provides a framework for understanding visual parameters and their influence on performance. The advantages of using such a methodology allow one to assess several visual parameters for a specific task. The MOVE provides the theoretical underpinnings for decision types used in PAM. It provides a starting point for discussion on how these decision types relate to visual events and their associated actions. In the literature, a variety of terms and operational definitions are used for the various decision types. More work is needed to develop MOVE, especially to formalize the terminology and definitions.

This research has shown that PAM can be used to create functional relationships between display system parameters and associated dependent variables. However, the nature of the relationships can be difficult to ascertain. It is clear that each visual parameter must be analyzed to determine its performance effect. The functions derived are dependent, to a great extent, upon the number of data points for each curve. Also, the range of values must be considered for each dependent variable.

The nature of the task should also be considered when developing functional relationships. The data from this study were clearly different from earlier work by Moon and Spencer (1944). However, it appears that there is agreement between the data sets, where the current data is offset from the Moon and Spencer data due to methodological differences. In light of this, “global” functional relationships aggregated across many different tasks may be unrealistic and researchers might instead work on developing “unique” functional relationships based on a selected task.

As a result of this research, a general procedure is suggested for preparing and applying the PAM (see Table 9). This procedure provides a systems perspective for including visual parameters in the design process. In the preparatory stage one starts with a general task analysis and decomposes tasks into visual events, associated decisions, and actions. Then, detailed visual event analysis determines the visual parameters, the performance measures associated with a visual event, and the functional relationship between these two components.

Table 9.

Preparatory steps to apply PAM

1. General task analysis

- a. Decompose tasks into basic visual events, decisions, and actions.
- b. Identify visual events that are logically grouped together or are required to be completed in a particular order or sequence (e.g., task completion strategies).
- c. Identify parallel completion paths (when appropriate).

2. Visual event analysis

- a. For a given decision/action sequence, determine visual parameters that influence it. Examples include: luminance, field of view, resolution, addressability, and contrast.
 - b. For each visual parameter, determine performance measures (e.g., readability, alignment on screen).
 - c. Determine relationship between the visual parameter and performance (e.g., relationship between resolution and visual angle) for each dependent measure. This is determined by creating a cumulative probability distribution based on the mean and standard deviation of the distribution of sample means.
-

To formally apply PAM, one must sum efficiency ratings for individual visual events (within a decision type), for a task completion pathway (or for groups of visual events), and for groups of tasks to arrive at an overall network measure of efficiency (see Table 10).

Table 10.

Application of PAM

1. Calculate **visual event efficiency** across individual display parameters (using the same dependent measure).

$$EDA_1 = [(weight_1)(display\ parameter_1) + (weight_2)(display\ parameter_2) \dots + (weight_j)(parameter\ parameter_j)]$$

2. Calculate **task efficiency** (for each task completion pathway)

$$Task_1 = [(weight_1)(EDA_1) + (weight_2)(EDA_2) + \dots (weight_k)(EDA_k)]$$

3. Calculate **network efficiency**

$$Network = [(weight_1)(Task_1) + (weight_2)(Task_2) + \dots (weight_i)(Task_i)]$$

Future research

Several issues need to be considered to improve the methodology. PAM focuses on the task level for a display system and does not yet provide the formalization for how tasks are interrelated. Additional development of the model is needed to represent the relationship between tasks. For individual parameters, the relationship has been modeled as additive. However, further research is needed to verify the nature of this relationship. For example, it is possible that a multiplicative model better represents how individual parameters may be combined. The ranking and rating system suggested for determining the weighting parameters needs to be verified and validated in a field test. Particular emphasis should be given to how this procedure works for PAM and what modifications should be made.

Future work is needed to evaluate PAM against formal criteria. A preliminary model for one visual parameter, one decision type, and one dependent variable was evaluated. The PAM process needs further refinement and validation. The criteria (see Table 11) provided by Meister (1995) are good candidates for consideration. Each of these criteria represents a practical concern for application. Operational measures for each of these criteria need to be developed for PAM. Field tests with operational tasks would provide valuable feedback on the PAM model.

Table 11.

Criteria of model effectiveness (Meister, 1995)

<u>Criterion</u>	<u>Definition</u>
Validity	Agreement of model outputs with actual system performance
Utility	The model's ability to accomplish the objectives for which it was developed
Reliability	The ability of various users to apply the model with reasonable consistency and to achieve comparable results when applied to similar systems
Comprehensiveness	Applicability to various types of systems, to various kinds of system devices and to various stages of design
Objectivity	Requires as few as subjective judgments as possible
Structure	Explicitly defined and described in detail
Ease of use	Ease with which the analyst can readily prepare data, apply, and extract understandable results
Cost of development/use	Includes both time and money
Richness of output	Number and type of output variables and forms of presentation

Finally, the PAM is designed to show how individual display parameters may be related to performance. It provides a framework for understanding how these parameters relate to a task. But developing the relationships between display parameters and performance is often difficult. Snyder (1973) wrote:

During the past two decades several hundred laboratory and analytical studies have been performed to assess the relationship between variation in line-scan display image parameters and observer performance. Conclusions drawn from critical reviews of these studies (Snyder, et al, 1967; Harifield, 1970) have indicated that cross-study comparisons are virtually impossible. Variation in specific system design parameters or in the manner by which display image quality is synthetically manipulated is often incompletely controlled, so that concomitant variation in the several contributing sources results (p. 95).

PAM suggests that performance measures for individual parameters can be combined for a given task. However, before this can happen, further studies are needed to validate the underlying relationships for individual parameters such as luminance or sharpness. In the current study, luminance was varied to both extremes for both the identification and detection tasks (i.e., from very difficult to see to very easy to see). Thus, it is possible that the luminance efficiency curves (see Figures 28-29, Figures 31-32) are adequate models for performance of these tasks. However, more data points are needed to fully validate these curves.

More research is needed to develop families of efficiency curves that meaningfully relate to display system parameters. It appears that data readily available from the literature may not be directly combined and translated to global efficiency curves. What is the best way to create these curves? This is a difficult question that needs more research. Given the difficulty of synthesizing

results in this field, hopefully the application of the PAM process will provide researchers with an improved ability for cross-study comparison.

REFERENCES

- Advisory Group for Aerospace Research and Development (1989). *Human Performance Assessment Methods*. AGARD-AG-308. NATO Advisory group for Aerospace Research and Development.
- Barten, P.G.J. (1987). The SQRI method: A new method for the evaluation of visible resolution on a display. *Proceedings of the Society for Information Display*, 28(3), 253-262.
- Beaton, R.J. (1997). *Description of the Hyperbolic Tangent (Sigmoid) Function*. Unpublished manuscript. Blacksburg, VA: Virginia Polytechnic Institute and State University.
- Beaton, R.J. and Farley, W.W. (1991). *Comparative study of the MTF, ICS and SQRI image quality metrics for visual display systems*. Report Number AL-TR-1992-0001. Armstrong Laboratory, Air Force Systems Command, Wright-Patterson Air Force Base, OH.
- Booth, K.S., Byrden, W.B., Cowan, M.F., and Plante, B.L. (1987). On the parameters of human visual performance: An investigation of the benefits of anti-aliasing. *Computer-Human Interaction Conference*, 13-19.
- Bullimore, M.A. Howarth, P.A., and Fulton, E.J. (1995). Assessment of visual performance. In J.R. Wilson and E.N. Corlett (Eds.) *Evaluation of human work: A practical ergonomics methodology*. (2nd ed.). Taylor and Francis.

- Chapanis, A. (1949). How we see: A summary of basic principles. In *Human Factors in Undersea Warfare*. Committee on Undersea Warfare. Washington, DC: National Research Council.
- Clapp, R.E. (1985). Problems of the visual display in flight simulation. In *Image Quality: An Overview*. Granger, E.M. and Baker, L.R. (Eds.) Society of Photo-Optical Instrumentation Engineers, 64-70.
- Cobb, P.W. and Moss, F.K. (1928). The four variables of the visual threshold. *The Journal of the Franklin Institute*, 205(6), 831-848.
- Conner, J.P. and Ganoung, R.E. (1935). An experimental determination of the visual thresholds at low values of illumination. *Journal of the Optical Society of America*, 25(9), 287-294.
- Coren, S., Ward, L.M. and Enns, J.T. (1994). *Sensation and perception*. (4th ed.). New York: Harcourt Brace College Publishers.
- Deuermeyer, B. and Cullinane, T. (1981). *IDEF0-Function Modeling*. Integrated Computer-Aided Manufacturing Program Office. Wright-Patterson Air Force Base, OH.
- Drury, C.G. (1995). Designing ergonomics studies and experiments. In J.R. Wilson and E.N. Corlett (Eds.) *Evaluation of human work: A practical ergonomics methodology*. (2nd ed.). Taylor and Francis.
- Drury, C.G., Paramore, B., Van Cott, H.P., Grey, S.M., and Corlett, E.N. (1987). In Salvendy, G. (Ed.). *Handbook of human factors*. John Wiley and Sons.

- Duncan, K.D. (1974). Analytical techniques in training design. In Edwards, E. and Lees, F.P. (Eds.). *The human operator in process control*. London: Taylor and Francis.
- Gescheider, G.A. (1985). *Psychophysics: Method, theory and application*. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Green, C.A. (1995). *Development and validation of a methodology for comprehensive performance assessment of complex tasks*. (Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University).
- Howell, D.C. (1992). *Statistical methods for psychology*. (3rd ed.). Belmont, CA: Duxbury Press.
- Kantowitz, B.H., Roediger, H.L. and Elmes, D.G. (1991). *Experimental psychology: Understanding psychological research*. West Publishing Company.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kirwan, B. and Ainsworth, L.K. (1992). *A guide to task analysis*. Taylor and Francis.
- Laughery, K.R., Jr. and Drews, C. (1984). Micro-Saint: A computer simulation designed for human factors engineers. *In proceedings of the Human Factors Society 28th Annual Meeting*. Santa Monica, CA: Human Factors Society.

- Lythgoe, R.J. (1932). The measurement of visual acuity. *Great Britain Medical Research Council*. Special Report Series, No. 173. London, England.
- Meister, D. (1995). Simulation and modelling. In J.R. Wilson and E.N. Corlett (Eds.) *Evaluation of human work: A practical ergonomics methodology*. (2nd ed.). Taylor and Francis.
- Miller, R.B. (1963). Task description and analysis. In R.M. Gagne (Ed). *Psychological principles in system development*. Holt, Rinehart, and Winston.
- Moon P. and Spencer, D.D. (1944). Visual data applied to lighting design. *Journal of the Optical Society of America*, 34(10), 605-617.
- Morris, W.T. (1977). *Decision analysis*. Columbus, OH: Grid, Inc.
- Moscovic, S.A. (1992). *Development of time and workload methodologies for MicroSAINT models of visual display and control systems*. (Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University).
- Norman, D.A. (1988). *The psychology of everyday things*. Basic Books.
- Raner, K. (1996). *MacCurveFit Docs* (software documentation). Kevin Raner Software. Mt. Waverly, Victoria, Australia.
- Shaler, S., Smith, E.L., and Chase, A.M. (1942). Visual acuity and illumination in different spectral regions. *Journal of General Physiology*, 25(4), 553-570.
- Shaler, S. (1937). The relation between visual acuity and illumination. *Journal of General Physiology*, 21(2), 165-188.

- Shinnick, M.D. and Gerber, D.L. (1985). A common language for analyzing work. *Journal of Systems Management*, 36(4), 8-13.
- Siegel, A.I. and Wolf, J.J. (1969). *Man-machine simulation models*. John Wiley and Sons.
- Sinclair, M.A. (1995). Subjective assessment. In J.R. Wilson and E.N. Corlett (Eds.) *Evaluation of human work: A practical ergonomics methodology*. (2nd ed.). Taylor and Francis.
- Snyder, H. L. (1988). Image quality. In M. Helander. (Ed.) *Handbook of human-computer interaction*. (pp. 437-474). North Holland.
- Snyder, H. L. (1973). Image quality and observer performance, In L. M. Biberian (ed.), *Perception of displayed information*. New York: Plenum.
- Valeton, J.M.. and van Meeteren, A. (1990). Target acquisition modeling based on human visual system performance. *Society for Photographic Instrumentation Engineers (SPIE)*, 1486, 68-84.
- van Meeteren, A. (1973). Visual aspects of image intensification. (Unpublished doctoral dissertation, University of Utrecht, The Netherlands).
- van Meeteren, A. (1991). Characterization of task performance with viewing instruments. *Journal of the Optical Society of America*, 7(10), 2106-2023.
- Westerink, J.H. and Roufs, J.A.J. (1989). Subjective image quality as a function of viewing distance, resolution, and picture size. *Society of Motion Picture and Television Engineers(SMPTE) Journal*, 98(2), 113-119.
- Wickens, C.D. (1992). *Engineering psychology and human performance* (2nd ed.). Harper Collins.

APPENDICES

Appendix A - Creating an efficiency curve using normal scores

In the PAM, efficiency curves are created for each parameter of interest. The variation within a given parameter is assumed to be normal, so that efficiency values are represented by probability values in a distribution. First, the mean and standard deviation for each level of the independent variable is determined. These means are really a distribution of sample means. Next, these sample means are transformed into a cumulative probability function based on the normal distribution. For a normal curve, the probabilities of 0.023, 0.159, 0.500, 0.841 and 0.977 are associated with z-scores of -2, -1, 0, 1, and 2 respectively (see Figure A1 below).

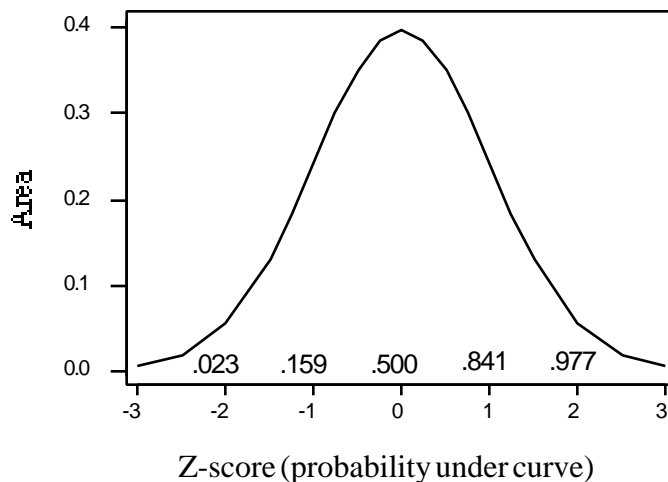


Figure A1. Unit normal distribution and associated probabilities

Several examples of efficiency curves are shown. In these studies, pixel size (resolution), sharpness and scanline density are varied and the performance measure for each study has been converted to a standard normal distribution (i.e., standardized). These examples show the wide variety of functions possible when performance data are standardized. In particular, for the limited data provided in these examples, standardized performance does not always follow an ogive function. The shape of the function depends on whether the function is directly or inversely related to the independent variable. It is also likely that the shape will be influenced by task type, since type of task has been shown to influence performance in a variety of settings. Note also, that for illustration, a variety of dependent variables have been selected. To compare standardized results across studies, a single dependent measure would be selected.

Booth, Byrden, Cowan, and Plante (1987) investigated the effect of screen resolution for a visual recognition task. Participants were presented with three-dimensional objects comprised of geometric cubes and they determined (as rapidly as possible) if the shape was comprised of 9 or 10 blocks. Pixel resolution was varied via pixel size, where size of 1 was composed of a 512 x 512 pixel matrix while size of 8 was a smaller 8 x 8 pixel matrix. The standardized results of their data are shown in Figures A2 and A3.

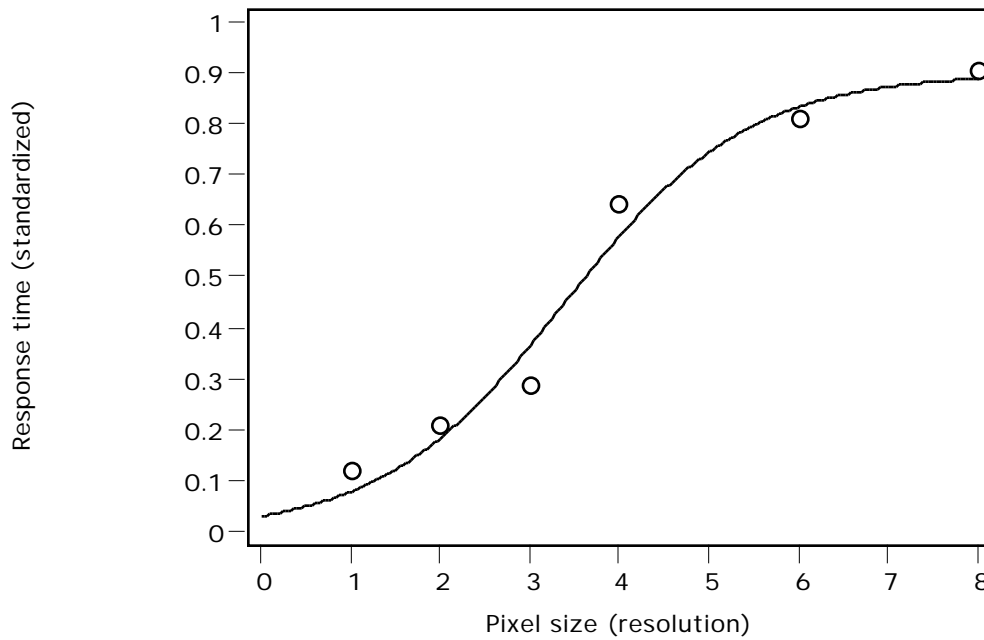


Figure A2. Discrimination response time as a function of pixel size for aliased images (Booth, Byrden, Cowan, and Plante, 1987).

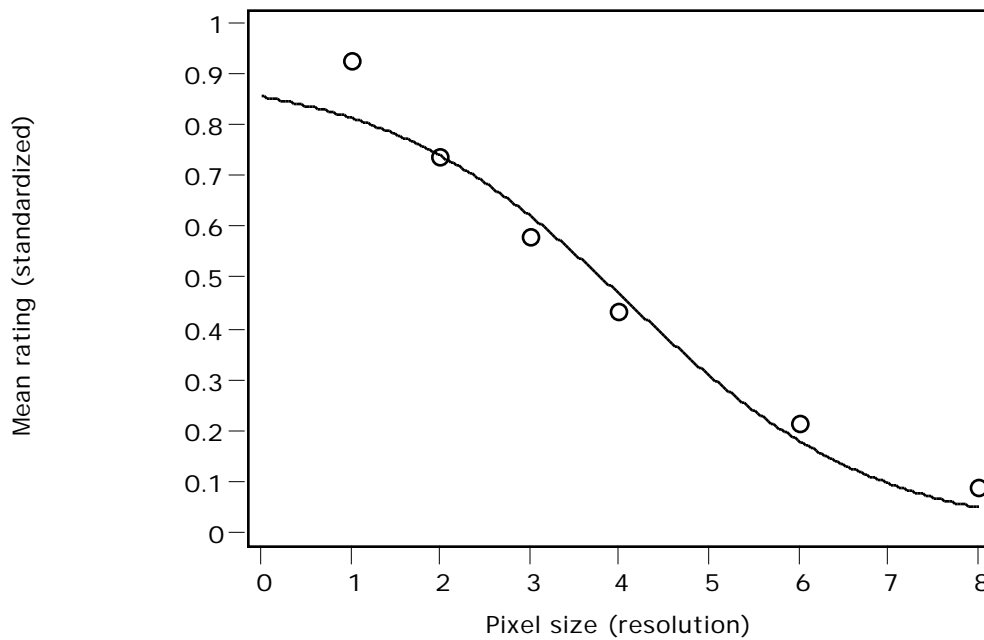


Figure A3. Ratings of picture quality as a function of pixel size for aliased images (Booth, Byrden, Cowan, and Plante, 1987).

Beaton and Farley (1991) calculated the 50% luminance distribution of a pixel (known as a 50% line spread function (LSF)) for various resolution levels used by Westerink and Roufs (1989). Subjective image quality judgments were standardized and plotted for these LSF values as shown in Figure A4.

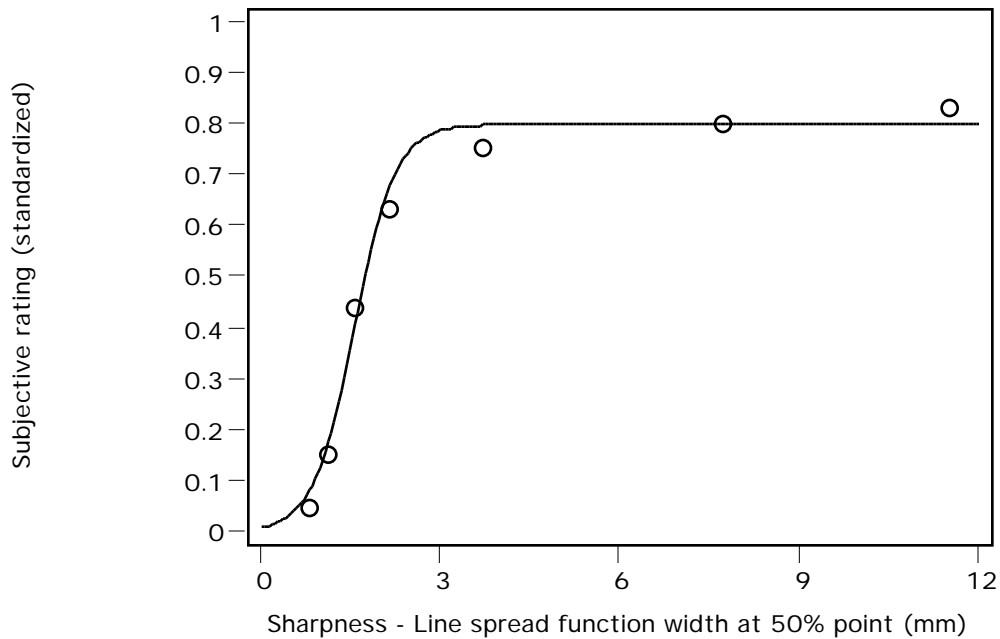


Figure A4. Ratings of image quality for color slides as a function of image sharpness (Westerink and Roufs, 1989).

Valeton and van Meeteren (1990) and van Meeteren (1991) used thermal intensifiers to simulate the shapes of military vehicles. They simulated the effect of distance by scanning in thermographs and varying the scanline density (number of lines per meter) under good visual conditions (without visual noise and with full contrast). They reported the percent correct identification of thermal side views of military vehicles as a function of scanline density. The results of their study has been standardized and plotted in Figure A5.

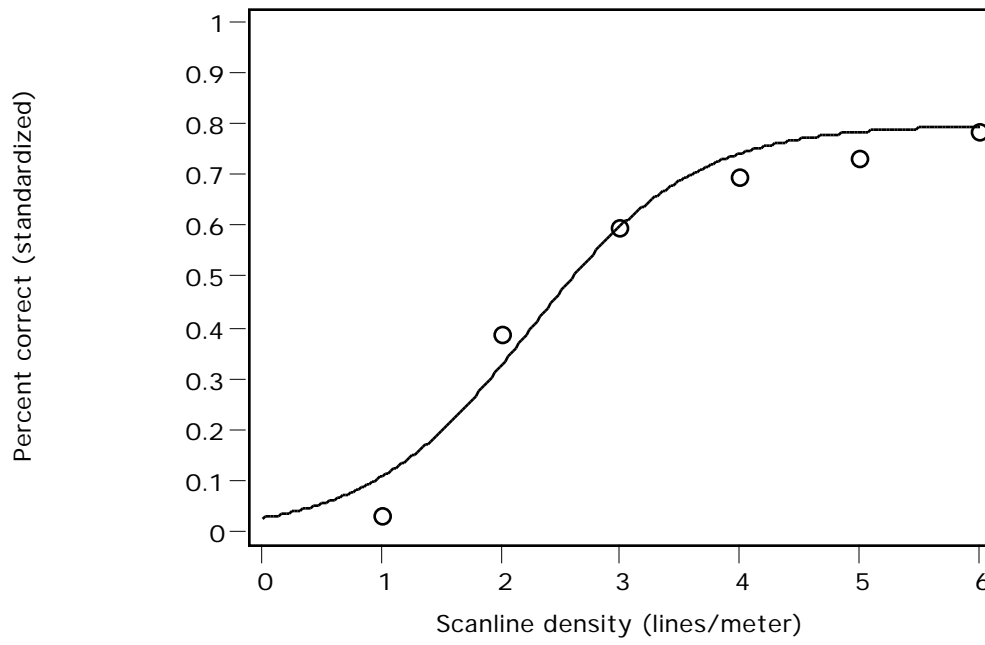


Figure A5. Identification of military vehicles using thermal imaging techniques (van Meeteren, 1991).

Appendix B - Instructions to participants, Informed consent formIntroduction to the study.

The purpose of this research to describe and empirically validate a new methodology that provides a quantitative basis for evaluating display image quality. The study is being conducted in the Displays and Controls Laboratory (DCL), Department of Industrial and Systems Engineering (ISE) at Virginia Tech. The principal investigators are Dean H. Orrell, a graduate student in ISE and Dr. Robert J. Beaton, director of the DCL.

You are being asked to spend approximately 1 hour in the DCL participating in this experiment. In this study you will be asked to perform the visual tasks of detecting and identifying an object under a variety of visual conditions. Throughout the study, please note that the design of the visual system is being evaluated, not you. Please do not be nervous about your performance on any of the tasks, just follow instructions and proceed in a manner that is comfortable for you.

After reading this introduction, you will be asked to fill out an informed consent form. If you agree to participate, your vision will be tested to ensure that you have normal visual acuity (i.e., 20/20 corrected vision). If your vision is not 20/20 corrected, then you will not be able to participate in this study. Before the actual experimental session begins, you will be given further instructions about the tasks and what is expected of you as a participant. You will be given practice on the tasks to be performed and on the rating of these tasks. Once you have practiced and are familiar with all of the tasks, you will be asked to perform them several times. The visual conditions will be altered in some way each time.

If you pass the vision screening and participate in the study, you will be paid \$5 per hour for your participation. During the experiment, if for any reason you decide not to continue, you will be paid for the time that you actually participate. Similarly, if the experiment is interrupted and must be terminated because of an equipment failure (as occasionally happens), you will be paid for the time spent up to the point of termination.

If you are interested in participating in the study, please read and sign the informed consent form. Thank you for your participation.

Informed consent form.

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Informed Consent for Participants of Investigative Projects

Title of Project: Performance Assessment Methodology: Task dependent evaluation of display systems

Principal Investigator: Dean H. Orrell

I. THE PURPOSE OF THIS RESEARCH AND ITS PROCEDURES

- You are invited to participate in this study whose purpose, description, and procedures are contained in the Introduction to the Study document, which you have already read. This study involves 60 participants, including yourself. Remember that you are free to withdraw from the study at any time.

II. BENEFITS OF THIS PROJECT

- While there are no direct benefits to you from participating in this study (other than payment), you may find the experiment interesting.
- No guarantee of benefits has been made to encourage you to participate.
- You may receive a synopsis or summary of this research when it is completed. Please leave a self-addressed envelope or your electronic or regular mailing address with the experimenter if this is what you wish.

III. EXTENT OF ANONYMITY AND CONFIDENTIALITY

- The results of this study will be kept confidential. The information you provide will have your name removed and only a subject number will identify you during analyses and any written reports of the research.

IV. COMPENSATION

- For participation in the project you will receive \$5 per hour. Payment will be made immediately following completion of the experiment.

V. FREEDOM TO WITHDRAW

- You are free to withdraw from this study at any time without penalty. If you choose to withdraw, compensation will be prorated and you will be paid for the time you spent participating in the study. This will also be the case if the investigator terminates the experiment because of equipment failure.

VI. APPROVAL OF RESEARCH

- This research project has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University.

VII. SUBJECT'S RESPONSIBILITIES AND PERMISSION

- I know of no reason I cannot participate in this study.
- I have read and understand the informed consent and conditions of this project. I have had all of my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.
- If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

Signature _____ Date _____

VIII. PARTICIPANT'S CONTACTS

Should I have any questions about this research or its conduct, I will contact:

Dean H. Orrell Investigator	Phone: 382-4743
Dr. Robert J. Beaton Faculty Advisor	Phone: 231-5936
Mr. Tom Hurd Chair, IRB Research Division	Phone: 231-9359

Participant instructions - Detection task.

Overview. In this experiment, you will be asked to perform a set of visual detection tasks that will be repeated a number of times. You will also be asked to rate the ease of detection for each series of detection tasks. First, you will be given a series of several practice trials to allow you to become familiar with the task. Later, you will be given a series of experimental trials. These trials will be presented to you on a standard television screen (similar to one you might have at home).

Experimental Sessions. During the practice and experimental trials, a visual scene will be presented and you will see what a person might see if they were “walking down a path” while looking straight ahead. Your task will be to detect red objects as you walk down the path. You may or may not see the color, but you should see the object. You should respond verbally as soon as you see the object. For example, if an object is detected on the right, you would respond by saying “Right.”

Please be sure you can see the object before you respond -- don't guess.

If you cannot detect an object, just say so. But don't give up too early. Even though we want you to be accurate, it is important to remember that you are not being “tested.” Instead you are helping us evaluate a system and we want you to report only what you see on the television monitor. So, relax and enjoy your visual journey.

At the end of each trial, you will be asked to rate how detectable the objects were during the trial. It is important to stress that there is not a right or wrong answer for your rating. Be honest about what you feel. You will use the scale shown below.

Please rate the overall trial:

easy	-4	-3	-2	-1	0	1	2	3	4	hard
to	^	^	^	^	^	^	^	^	^	to
detect	-----								detect	

Please feel free to ask the experimenter any questions you have now or at any time during the practice trials. Let the experimenter know when you are ready to begin.

Participant instructions - Identification task.

Overview. In this experiment, you will be asked to perform a set of visual identification tasks that will be repeated a number of times. You will also be asked to rate the ease of identification for each series of identification tasks. First, you will be given a series of several practice trials to allow you to become familiar with the task. Later, you will be given a series of experimental trials. These trials will be presented to you on a standard television screen (similar to one you might have at home).

Experimental sessions. During the practice and experimental trials, a visual scene will be presented and you will see what a person might see if they were “walking down a path” while looking straight ahead. Your task will be to identify red squares, triangles or circles as you walk down the path. You may or may not see the color, but you should see the shape of the object. You should respond verbally as soon as you see the object. For example, if a circle is detected, you would respond by saying: “Circle.”

Please be sure you can see the object before you respond -- don't guess.

If you cannot identify an object, just say so. But don't give up too early. Even though we want you to be accurate, it is important to remember that you are not being “tested.” Instead you are helping us evaluate a system and we want you to report only what you see on the television monitor. So, relax and enjoy your visual journey.

At the end of each trial, you will be asked to rate how identifiable the objects were during the trial. It is important to stress that there is not a right or wrong answer for your rating. Be honest about what you feel. You will use the scale shown below.

Please rate the overall trial:

easy	-4	-3	-2	-1	0	1	2	3	4	hard
to	^	^	^	^	^	^	^	^	^	to
identify									identify	

Please feel free to ask the experimenter any questions you have now or at any time during the practice trials. Let the experimenter know when you are ready to begin.

Appendix C - Geisser-Greenhouse correction factor

The results were adjusted for the Geisser-Greenhouse correction factor as suggested by Keppel (1991). Within-subjects designs are subject to violation of the homogeneity of variance and sphericity assumptions. Keppel states that if the ratio of variances is less than or equal to 3, then the homogeneity of variance assumption may be violated without concern. However, sphericity is more sensitive to violations of assumptions. Sphericity is simply the assumption that the variances of difference scores is equal across the treatments. The Geisser-Greenhouse correction factor corrects for these potential problems by adjusting the F-critical value to a more stringent level.

For this research, the correction factor was applied by using the following formulas as suggested by Howell (1992).

main effects $df = 1, g(n-1)$

interactions $df = (g-1), g(n-1)$

where

g = number of groups

n = number of subjects per cell

The critical F values in the ANOVA summary tables shown in Appendices D and E were compared to the Geisser-Greenhouse F-values and all results remained significant after this comparison. Howell (1992) notes, that “very large values of F are significant regardless of the form of the sphericity matrix.”

Appendix D - Detailed statistical analysis - Experiment 1

Table D1.

Experiment 1 - Time to target

ANOVA summary table

Source	DF	SS	MS	F	P
Between subjects					
Decision Type (DT)	1	46742.0	46742.0	390.49	0
Ss w/in DT	28	3352.6	119.7		
Within subjects					
Luminance	3	3669.2	1223.1	102.4	0
Luminance*DT	3	863.7	287.9	24.1	0
Luminance*Ss w/in DT	1381	16496.6	11.9		
Total	1416	71124.1			

Geisser-Greenhouse critical F values (against which F ratios in Table were compared)

Decision Type - $F_{(.05)}(1,28) = 4.2$ Luminance - $F_{(.05)}(1,56) = 4.0$ Luminance*Decision Type - $F_{(.05)}(3,56) = 2.68$

Table D2.

Experiment 1 - Time to target

Means, standard deviations and standard errors

Decision Type	N	N*	Mean	StDev	SEMean
Detection	720	0	15.60	5.52	0.206
Identification	699	21	4.14	1.91	0.072
Luminance					
3.2 cd/m ²	345	15	7.85	6.87	0.370
4.5 cd/m ²	357	3	9.16	5.92	0.314
8.6 cd/m ²	357	3	10.93	7.22	0.382
16.5 cd/m ²	360	0	11.85	7.53	0.398

N* = number of missing values

Table D3.

Experiment 1 - Subjective rating

ANOVA summary table

Source	DF	SS	MS	F	P
Between subjects					
Decision Type (DT)	1	66.15	66.15	5.56	0.026
Ss w/in DT	28	333.25	11.90		
Within subjects					
Luminance	3	684.02	228.01	189.19	0
Luminance*DT	3	6.57	2.19	1.82	0.145
Error	204	245.85	1.21		
Total	239	1335.83			

Geisser-Greenhouse critical F values (against which F ratios in Table were compared)

Decision Type - $F_{(.05)}(1,28) = 4.2$

Luminance - $F_{(.05)}(1,56) = 4.0$

Luminance*Decision Type - $F_{(.05)}(3,56) = 2.68$

Table D4.

Experiment 1 - Subjective rating

Means, standard deviations, and standard errors

Decision Type	N	Mean	StDev	SEMean
Detection	120	6.44	2.03	0.185
Identification	120	5.39	2.56	0.234
Luminance				
	N	Mean	StDev	SEMean
3.2 cd/m ²	60	3.43	1.97	0.254
4.5 cd/m ²	60	5.43	1.91	0.246
8.6 cd/m ²	60	6.91	1.50	0.193
16.5 cd/m ²	60	7.91	1.14	0.147

Appendix E - Detailed statistical analysis - Experiment 2

Table E1.

Experiment 2 - Time-to-target

ANOVA summary table

Source	DF	SS	MS	F	P
Between subjects					
Decision Type (DT)	1	42099.4	46742.0	437.56	0
Ss w/in DT	28	2694.0	96.2		
Within subjects					
Sharpness	2	169.8	84.9	9.2	0
Sharpness*DT	2	27.8	13.9	1.5	0.224
Error	1046	9705.2	9.3		
Total	1079	54696.2			

Geisser-Greenhouse critical F values (against which F ratios in Table were compared)

Decision Type - $F_{(.05)}(1,28) = 4.2$ Sharpness - $F_{(.05)}(1,42) = 4.0$ Sharpness*Decision Type - $F_{(.05)}(2,42) = 3.15$

Table E2.

Experiment 2 - Time-to-target

Means, standard deviations, and standard errors

Decision Type	N	Mean	StDev	SEMean
Detection	540	17.89	4.45	0.192
Identification	540	5.41	1.88	0.081
Sharpness	N	Mean	StDev	SEMean
0.864 mm	360	11.93	7.27	0.383
0.711 mm	360	11.95	7.26	0.383
0.508 mm	360	11.09	6.81	0.359

Table E3.

Experiment 2 - Subjective rating

ANOVA summary table

Source	DF	SS	MS	F	P
Between subjects					
Decision Type	1	3.61	3.61	0.19	0.67
Ss w/in Decision Type	28	546.11	19.50		
Within subjects					
Sharp	2	5.21	2.61	1.94	0.147
Sharpness*Decision Type	2	0.13	0.07	0.05	0.952
Error	146	195.95	1.34		
Total	179	751.02			

Geisser-Greenhouse critical F values (against which F ratios in Table were compared)

Decision Type - $F_{(.05)}(1,28) = 4.2$

Sharpness - $F_{(.05)}(1,42) = 4.0$

Sharpness*Decision Type - $F_{(.05)}(2,42) = 3.15$

Table E4.

Experiment 2 - Subjective rating

Means, standard deviations, and standard errors

Decision Type	N	Mean	StDev	SEMean
Detection	90	6.46	1.79	0.189
Identification	90	6.17	2.28	0.240
Sharpness				
0.864 mm	60	6.31	2.28	0.295
0.711 mm	60	6.53	1.79	0.231
0.508 mm	60	6.11	2.06	0.266

Appendix F - MacCurveFit program

Raner (1996) provides the following overview of how the MacCurveFit program uses SSE and R^2 to find the best fit for a given curve. The software uses the Davidon-Fletcher-Powell algorithm for non-linear optimization to fit the data.

The object of least squares curve fitting is to minimize the sum of squares error. The program uses two quantitative indicators to determine the best fit for a curve.

The first indicator is the SSE which is defined as

$$SSE = \sum (y_i - f(x_i, a, b, c, d))^2 \quad \text{Eq. F1}$$

where x_i and y_i are the i_{th} data pair, f is the function and a , b , c , and d are the coefficients. For a perfect least squares fit, SSE would equal zero, the larger the value of SSE the poorer the fit.

The value R^2 is the correlation coefficient and is calculated as

$$R^2 = 1 - \frac{nSSE}{\sum y_i^2 - (\sum y_i)^2} \quad \text{Eq. F2}$$

where n is the number of data points. A perfect fit has a correlation coefficient of 1 and the lower the value the poorer the fit.

Appendix G - Hyperbolic tangent function

A hyperbolic tangent function was used for fitting the visual efficiency curves in this research. Beaton (1997) describes the hyperbolic tangent function in the following paragraphs. Note that for his example figures, both the x and y coordinates have been standardized on a scale of 0-1.

Properties of the sigmoid curve

A hyperbolic tangent function (also known as a “sigmoid” curve) describes this general trend, as given by

$$f(x) = a \frac{e^{b(x+c)} - 1}{e^{b(x+c)} + 1} + d \quad (\text{Eq. G1})$$

Examining Eq. G1, we can deduce some properties of the sigmoid curve. First, we notice that when $x = -c$, $e^{b(x+c)}$ becomes one (since the exponent becomes zero), $f(x)$ becomes $a * (0/2) + d$ which is simply d . As x increases (i.e. approaches positive infinity), there are two cases: b is positive, or b is negative. However, as we shall see, b will be positive for many display parameters.

As x approaches positive infinity, $e^{b(x+c)}$ approaches positive infinity (since the exponent approaches positive infinity), the expression in parentheses in Eq. G1 approaches one, and $f(x)$ approaches $d+a$. As x decreases (i.e. approaches negative infinity), $e^{b(x+c)}$ approaches zero (since the exponent approaches negative infinity), the expression in parentheses in Eq. G1 approaches -1, and $f(x)$ approaches $d-a$.

To review, when $x = -c$, $f(x) = d$, as x increases from $-c$, $f(x)$ approaches $d+a$, and as x decreases from $-c$, $f(x)$ approaches $d-a$. These trends are illustrated in Figure G1:

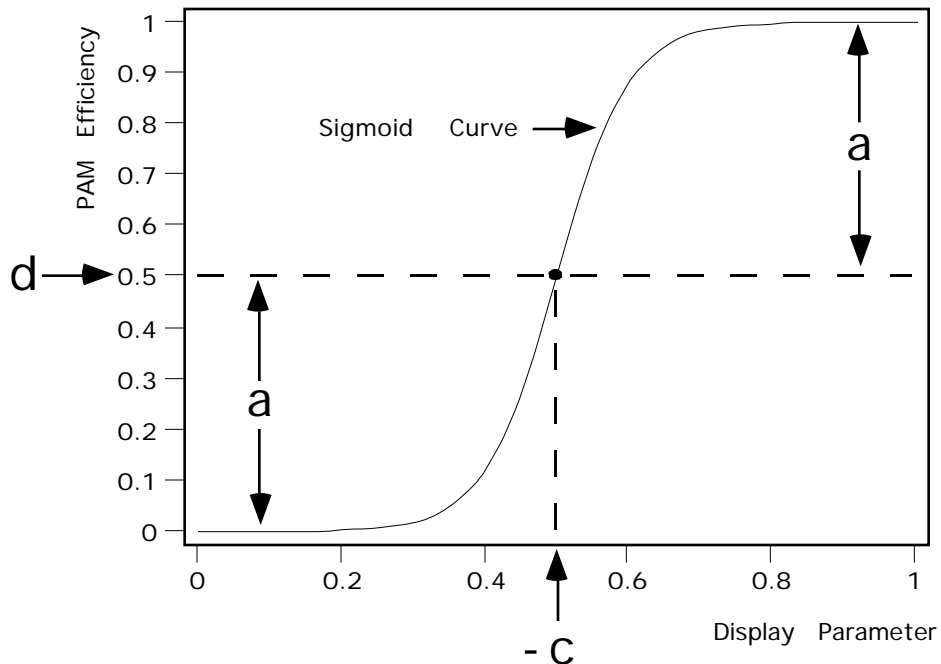


Figure G1. Hyperbolic tangent function and its parameters

What, then, is the interpretation of the b parameter? The larger the value of b , the greater the magnitude of $e^{b(x+c)}$, and therefore, the more rapidly $f(x)$ approaches $d+a$ as x increases from $-c$. Also, the larger b is, the more rapidly $f(x)$ approaches $d-a$ as x decreases from $-c$. The parameter b , therefore, affects the slope or steepness of the curve, particularly in the “middle” section of the curve, i.e. near $x = -c$. A larger value of b results in a steeper curve, as illustrated below:

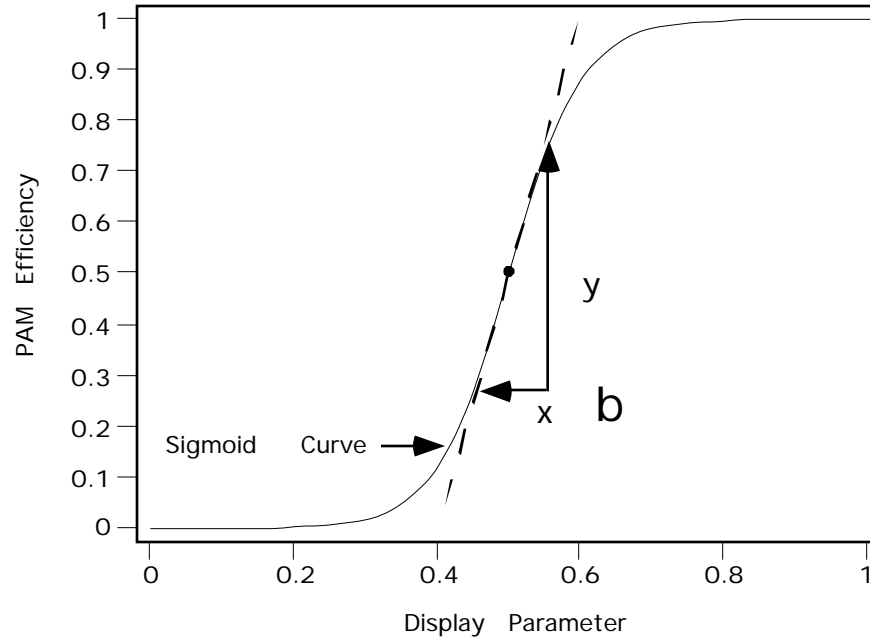


Figure G2. Slope of the hyperbolic tangent function

To review, the a , b , c , and d parameters have the following interpretations:

- a - The difference between the high and low proficiency values is $2a$. For display parameters, a will usually be 0.5, i.e., proficiency values range from 0% to 100%.
- b - proportional to the steepness of the middle of the sigmoid curve, i.e. how rapidly proficiency increases with increases in the display parameter value.
- c - determines the horizontal placement of the sigmoid curve ($x = c$ is the display parameter value corresponding to 50% proficiency)
- d - determines the vertical placement of the sigmoid curve. For display parameters, d will usually be 0.5, i.e. 50% proficiency.

Appendix H - Simple effects - Luminance-Detection conditions

Table H1.

Luminance-Detection condition - Time to target

Univariate F-test

Source	DF	SS	MS	F	P
Within subjects					
Subjects	14	2858.37	204.17		
Luminance	3	4040.15	1346.72	96.08	0
Subjects*Luminance	42	588.67	14.02		
Error	660	14434.17	21.87		
Total	719	21921.36			

Table H2.

Luminance-Detection condition - Time to target

Means, standard deviations, and standard errors

Luminance	N	Mean	StDev	SEMean
3.2 cd/m ²	180	12.52	6.52	0.486
4.5 cd/m ²	180	14.14	4.04	0.301
8.6 cd/m ²	180	17.40	3.91	0.291
16.5 cd/m ²	180	18.35	5.09	0.379

Table H3.

Luminance-Detection condition - Subjective rating

Univariate F-test

Source	DF	SS	MS	F	P
Within subjects					
Subjects	14	100.09	7.15		
Luminance	3	284.01	94.66	83.28	0
Subjects*Luminance	42	47.74	1.13		
Error	60	58.25	0.97		
Total	119	490.09			

Table H4.

Luminance-Detection condition - Subjective rating

Means, standard deviations, and standard errors

Luminance	N	Mean	StDev	SEMean
3.2 cd/m ²	30	4.18	1.66	0.304
4.5 cd/m ²	30	5.93	1.63	0.297
8.6 cd/m ²	30	7.45	1.06	0.194
16.5 cd/m ²	30	8.20	0.75	0.137

Appendix I - Simple effects - Sharpness-Detection conditions

Table I1.

Sharpness-Detection condition - Time to target

Univariate F-test

Source	DF	SS	MS	F	P
Within subjects					
Subjects	14	2235.77	159.7		
Sharpness	2	161.69	80.85	4.35	0.023
Subjects*Sharpness	28	520.26	18.58		
Error	495	7758.59	15.67		
Total	539	10676.31			

Table I2.

Sharpness-Detection condition - Time to target

Means, standard deviations, and standard errors

Sharpness	N	Mean	StDev	SEMean
0.864 mm	180	18.35	4.47	0.333
0.711 mm	180	18.22	4.61	0.343
0.508 mm	180	17.13	4.18	0.312

Table I3.

Sharpness-Detection condition - Subjective rating

Univariate F-test

Source	DF	SS	MS	F	P
Within subjects					
Subjects	14	221.74	15.84		
Sharpness	2	2.24	1.12	0.82	0.450
Subjects*Sharpness	28	38.09	1.36		
Error	45	24.75	0.55		
Total	89	286.82			

Table I4.

Sharpness-Detection condition - Subjective rating

Means, standard deviations, and standard errors

Sharpness	N	Mean	StDev	SEMean
0.864 mm	30	6.48	2.06	0.376
0.711 mm	30	6.63	1.56	0.285
0.508 mm	30	6.25	1.77	0.323

Appendix J - Simple effects - Luminance-Identification conditions

Table J1.

Luminance-Identification condition - Time to target

Univariate F-test

Source	DF	SS	MS	F	P
Within subjects					
Subjects	14	494.55	35.32		
Luminance	3	567.35	189.12	77.82	0
Subjects*Luminance	42	101.96	2.43		
Error	637	1371.88	2.15		
Total	696	2535.74			

Table J2.

Luminance-Identification condition - Time to target

Means, standard deviations, and standard errors

Luminance	N	N*	Mean	StDev	SEMean
3.2 cd/m ²	165	15	2.76	1.64	0.127
4.5 cd/m ²	177	3	4.05	1.71	0.128
8.6 cd/m ²	177	3	4.34	1.87	0.141
16.5 cd/m ²	180	0	5.28	1.52	0.113

N* = number of missing values

Table J3.

Luminance-Identification condition - Subjective rating

Univariate F-test

Source	DF	SS	MS	F	P
Within subjects					
Subjects	14	233.15	16.65		
Luminance	3	406.57	135.53	64.60	0
Subjects*Luminance	42	88.11	2.01		
Error	60	51.75	0.86		
Total	119	779.59			

Table J4.

Luminance-Identification condition - Subjective rating

Means, standard deviations, and standard errors

Luminance	N	Mean	StDev	SEMean
3.2 cd/m ²	30	2.67	1.98	0.361
4.5 cd/m ²	30	4.92	2.05	0.375
8.6 cd/m ²	30	6.37	1.68	0.307
16.5 cd/m ²	30	7.62	1.38	0.252

Appendix K - Simple effects - Sharpness-Identification conditions

Table K1.

Sharpness-Identification condition - Time to target

Univariate F-test

Source	DF	SS	MS	F	P
Within subjects					
Subjects	14	474.01	33.86		
Sharpness	2	29.29	14.65	3.64	0.039
Subjects*Sharpness	28	112.74	4.03		
Error	495	1098.46	2.22		
Total	539	1714.59			

Table K2.

Sharpness-Identification condition - Time to target

Means, standard deviations, and standard errors

Sharpness	N	Mean	StDev	SEMean
0.864 mm	180	5.50	1.69	0.126
0.711 mm	180	5.59	2.06	0.153
0.508 mm	180	5.06	1.52	0.113

Table K3.

Sharpness-Identification condition - Subjective rating

Univariate F-test

Source	DF	SS	MS	F	P
Within subjects					
Subjects	14	324.37	23.16		
Sharpness	2	3.11	1.55	1.44	0.254
Subjects*Sharpness	28	30.23	1.08		
Error	45	102.88	2.29		
Total	89	460.58			

Table K4.

Sharpness-Identification condition - Subjective rating

Means, standard deviations, and standard errors

Sharpness	N	Mean	StDev	SEMean
0.864 mm	30	6.13	2.50	0.457
0.711 mm	30	6.41	2.01	0.368
0.508 mm	30	5.96	2.33	0.425

VITA

Dean H. Orrell received his undergraduate degree in Human Factors Engineering at Wright State University and his Masters degree in Experimental-Human Factors Psychology from the University of Dayton. Prior to coming to Virginia Tech he served as an Assistant Professor at the United States Air Force Academy in Colorado Springs as part of the Behavioral Sciences and Leadership Department teaching human factors engineering, psychology, and statistics.

As a human factors engineer, he designs and evaluates systems from the user's perspective. His special interests include how people interact with and learn to use computer systems and in particular: software usability testing, multimedia development, graphical user interface design, and user education and training.

Dean H. Orrell