Adaptation of Perceived Depth Related to Changes of the Effective Interpupillary Distance in Computer-Graphic Stereoscopic Displays

by

David Walter Schloerb

S.B. Physics
Massachusetts Institute of Technology, 1977

S.M. Mechanical Engineering
Massachusetts Institute of Technology, 1982

Submitted to the Department of Mechanical Engineering in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mechanical Engineering

at the

Massachusetts Institute of Technology

May 9, 1997

Signature of Author .....
Department of Mechanical Engineering
May 9, 1997

Certified by .....
Nathaniel I. Durlach
Senior Research Scientist
Department of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Ain A. Sonin
Chairman, Department Committee on Graduate Students

Eng.

Adaptation of Perceived Depth Related to Changes of the Effective Interpupillary Distance in Computer-Graphic Stereoscopic Displays

by

David Walter Schloerb

Submitted to the Department of Mechanical Engineering on May 9, 1997 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mechanical Engineering

ABSTRACT

Telepresence and virtual environment systems impose transformations that alter normal sensory-motor relationships. Engineers must know how operators will adapt to such transformations in order to make rational trade offs among design alternatives. Fundamental research is required because there are no general models to predict the extent of adaptation or how long it will take. This investigation focuses on how people adapt to a single visual transformation, a change of the effective interpupillary distance (IPD), while performing tasks with a computer-graphic stereoscopic display.

Experiments were performed with human subjects in which the effective IPD was either x1 or x2 the subject's actual IPD. Calibration experiments were performed in which the positions of virtual objects presented by the apparatus were judged relative to a physical probe. A forced-choice depth discrimination experiment was also performed using two virtual targets, where the targets were separated laterally so that discrimination was not limited by the size of the pixels on the display. Finally, an adaptation experiment was performed to observe how bias and resolution in an absolute depth identification task change over time when the effective IPD is changed. The exposure in the adaptation experiment involved explicit correct-answer feedback after each response.

An unexpected result was that subjects adapted in the absence of explicit feedback. Alternative feedback mechanisms are proposed. The fact that computer displays are composed of discrete pixels was found to be a subtle problem that confounded some of the experiments. When this was taken into account, virtual objects were perceived at the expected locations (for the specific stimuli of the test), depth discrimination improved as expected when the effective IPD was increased, and shifts in

bias and resolution were observed (although in one case there was no change in depth resolution when the effective IPD was changed). The results are roughly consistent with a preliminary adaptation model proposed by Shinn-Cunningham (1994). Limitations of the model are identified and design rules are proposed based on the results of the thesis research.

Thesis Supervisor: Nathaniel I. Durlach
Title: Senior Research Scientist

ACKNOWLEDGMENTS

I had two principal advisors in my doctoral program: Thomas B. Sheridan (my Faculty Advisor in the Department of Mechanical Engineering) and Nathaniel I. Durlach (my Thesis Supervisor). I am thankful to Tom for leading me back into school 10 years ago to work on the Ph.D., and I will be forever grateful to Nat for leading me back out again. Nat has been a source of funding, good advice, and emotional support, without which I might never have succeeded.

I am also grateful to the other members of my thesis committee, Richard M. Held and J. Kenneth Salisbury, Jr., who were both helpful and encouraging over the years. I would also like to thank Barbara Shinn-Cunningham, who provided valuable feedback in the preparations for my thesis defense. Several members of the lab also made contributions to the project. Specifically, I would like to thank Thomas E. von Wiegand, Ben Sachtler, Evan Wies, and Lukasz Weber for their help. My experimental subjects also deserve thanks for enduring hours of tedious tests and, finally, I am grateful to my family and friends for their endurance and support.

This work was supported by ONR-AASERT Grant #N00014-94-1-1079.

TABLE OF CONTENTS

ABSTRACT **2**
ACKNOWLEDGMENTS **4**

1. INTRODUCTION **10**
1.1 Thesis Problem 10
1.1.1 Focus of the Research 11
1.1.2 Objectives 12
1.2 Summary of the Experiments 13
1.2.1 Calibration (Chapter 4) 14
1.2.2 Discrimination (Chapter 5) 14
1.2.3 Identification (Chapter 6) 15
1.3 Overview of the Document 15

2. BACKGROUND **16**
2.1 Depth Perception 16
2.1.1 Accommodation and Convergence 16
2.1.2 Monocular Cues 17
2.1.3 Binocular Disparity 17
2.1.3.1 Stereopsis Regions 19
2.1.3.2 Stereoacuity and Range 20
2.1.3.3 Related Issues 20
2.2 Stereoscopic Displays 21
2.2.1 Principles of Operation 21
2.2.1.1 Virtual Space and the X_{ipd} Transformation 24
2.2.2 Limitations 24
2.2.2.1 Accommodation and Convergence 24
2.2.2.2 Crosstalk 25
2.2.2.3 Superposition of Real and Virtual Objects 25
2.2.2.4 Spatial Resolution and Display Pixelation 26
2.3 Adaptation 27
2.3.1 Experimental Paradigm 28
2.3.2 General Effects 29
2.3.3 Altered IPD 30
2.3.4 Shinn-Cunningham 32
2.4 Preliminary Theory of Intensity Resolution 33
2.4.1 Decision Model 33
2.4.2 Internal-Noise Model 36
2.5 Preliminary Adaptation Model 37

3. GENERAL METHODS 43
3.1 Subjects 43
3.1.1 Knowledge of Purpose 44
3.1.2 Preliminary Tests 44
3.2 Apparatus 45
3.2.1 Displays 45
3.2.1.1 CrystalEyes 52
3.2.1.2 Mirror, Experimental Space, LEDs, and Lighting 53
3.2.1.3 Image Color and Crosstalk 54
3.2.1.4 Eye Position 55
3.2.1.5 Reference LEDs 55
3.2.2 Controls 55
3.2.2.1 Keyboard 56
3.2.2.2 Trackball 56
3.2.2.3 PHANToM 56
3.2.2.4 Slider 56
3.2.3 Computers 57
3.2.4 Software 58
3.2.4.1 Versions, Types of Tests, and File Names 58
3.2.4.2 Calibrated Parameters 60
3.2.4.3 S_POINT Algorithm and the X_{ipd} Transformation 60
3.2.4.4 Jitter 67
3.2.4.5 Rendering 68
3.2.4.6 EYE_CAL Algorithm 69
3.2.5 Calibration 71
3.3 Stimulus Depth Cues 71
3.3.1 Physical Objects 72
3.3.2 Virtual Objects 72
3.4 Procedures 74
3.4.1 Subject Screening and Initial Training 74
3.4.2 Lighting 75
3.4.3 Feedback and Speed 75
3.4.4 Eye Position Measurements 76
3.4.5 Test Schedule and Breaks 76

4. CALIBRATION EXPERIMENTS 78
4.1 3D Experiment 78
4.1.1 Method 79
4.1.1.1 Subjects 79
4.1.1.2 Apparatus and Stimuli 79
4.1.1.3 Procedure 80
4.1.1.4 Data Analysis 80
4.1.2 Results 82

4.1.2.1 Presentation of the Data 82
4.1.2.2 Bias 86
4.1.2.3 Resolution 86
4.1.3 Discussion 89
4.1.3.1 Unexpected Resolution Results 89
4.1.3.2 Systematic Position Errors 90
4.2 Z-axis Experiment 90
4.2.1 Method 91
4.2.1.1 Subjects 91
4.2.1.2 Apparatus and Stimuli 91
4.2.1.3 Procedure 93
4.2.2 Results 93
4.2.3 Discussion 95
4.3 Virtual-Virtual Experiment 97
4.3.1 Method 97
4.3.1.1 Subjects 97
4.3.1.2 Apparatus and Stimuli 97
4.3.1.3 Procedure 98
4.3.2 Results 98
4.4 Display Pixelation Analysis 100
4.4.1 Depth Resolution 102

5. DISCRIMINATION EXPERIMENT 105
5.1 Methods 105
5.1.1 Subjects 105
5.1.2 Apparatus and Stimuli 106
5.1.2.1 Depth Number Units 108
5.1.3 Procedure 109
5.2 Results 110
5.3 Discussion 112
5.4 Psychophysical Scale 114

6. IDENTIFICATION EXPERIMENTS 117
6.1 Relevance of the Experiments 118
6.1.1 Applied Adaptation Studies 119
6.2 Experiment 1 120
6.2.1 Methods 120
6.2.1.1 Subjects 120
6.2.1.2 Apparatus and Stimuli 121
6.2.1.3 Procedure 124
6.2.2 Results 125
6.2.3 Discussion 128
6.2.3.1 Display Pixelation 128

6.2.3.2 Why the resolution did not change overall 130
6.2.3.3 Why the resolution did change near zero 134
6.2.3.4 Finite Response Set Bias 135
6.2.3.5 Indirect Feedback 135
6.2.3.6 Central Tendency 138
6.3 Experiment 2 138
6.3.1 Methods 139
6.3.1.1 Subjects 139
6.3.1.2 Apparatus and Stimuli 140
6.3.1.3 Procedure 141
6.3.2 Results 141
6.3.3 Discussion 145
6.3.3.1 Pre- to Postexposure Shift 146
6.3.3.2 Difference in the Shape of the x2 IPD Curves 147
6.3.3.3 Resolution shift between T40 and T42 148

7. QUANTITATIVE ANALYSIS 149
7.1 Application of the Preliminary Adaptation Model 149
7.1.1 Definitions 150
7.1.2 Evaluation of Terms 152
7.1.3 Overview of the Evaluation of k(t) and G 154
7.1.3.1 Representative Data 155
7.1.3.2 Dealing with Edge Effects 157
7.1.3.3 How σ_g is Estimated 158
7.1.3.4 How k_g is Estimated 159
7.1.4 Final Equations of the Model 160
7.2 Initial Investigation of A and k 162
7.2.1 Procedure 162
7.2.2 Results 162
7.2.3 Discussion 165
7.2.3.1 Indirect Feedback 166
7.2.3.2 Bias in the Estimate of k 166
7.2.3.3 Variance of k 166
7.2.3.4 Initial Training and Exposure 167
7.2.3.5 Hypothesis 168
7.3 Evaluation of σ_g and k_g 173
7.3.1 Procedure and Results 174
7.3.2 Observations and Discussion 177
7.3.2.1 σ analysis 179
7.3.2.2 k analysis 183
7.3.3 σ against N 186
7.4 Evaluation of G 188

7.5 Predictions vs. Results 192
7.5.1 Mean Response against N 193
7.5.2 Resolution against N 193
7.5.3 Mean Resolution 198
7.6 Limitations of the Model 200
7.6.1 Perscriptive Model 200
7.6.2 Assumptions about k 200
7.6.3 Discrepancies between Predictions and Results 201

8. CONCLUSION 203
8.1 Summary of the Results 203
8.1.1 Calibration Experiments (Chapter 4) 203
8.1.2 Discrimination Experiment (Chapter 5) 205
8.1.3 Identification Experiments (Chapters 6 and 7) 205
8.1.3.1 Preliminary Adaptation Model 206
8.1.3.2 Secondary Effects 206
8.2 Design Rules 207
8.3 Future Work 209
8.3.1 Roving Depth Experiment 209
8.3.2 True Adaptation in Absolute Identification 209
8.3.2.1 Definitions, Mechanisms, and Generalization 211
8.3.3 Development of the Model 211
8.3.4 Beyond Simple Identification 212
8.3.5 Antialiasing 212

APPENDIX A. VIRTUAL WORKBENCH 213
APPENDIX B. SUBJECT SCREENING: PROCEDURES AND DATA 224
APPENDIX C. APPARATUS CALIBRATION 227
APPENDIX D. DATA CORRECTIONS 255
APPENDIX E. K CORRECTION 257

REFERENCES 260

1. INTRODUCTION

This thesis investigates how people adapt to a fundamental visual transformation while performing various tasks with a computer-graphic stereoscopic display. Specifically, experiments were performed with human subjects in order to observe how perceived depth changes over time when the effective interpupillary distance (IPD) is changed. An important objective of the thesis is to test the results against the *preliminary model of adaptation* developed by Shinn-Cunningham in connection with auditory localization (1994). Another goal is to find out whether task performance can be improved by increasing the effective IPD of the display relative to the human operator's actual IPD.

The experimental results show that depth resolution is improved, under certain conditions, when the effective IPD is increased, and that human operators adapt more or less completely to the transformation in an absolute identification task. In other words, the bias in the operators' response, initially caused by the transformation, gradually disappears. The fact that computer-graphic displays are divided into discrete pixels is found to have subtle consequences, however, that may limit performance in some cases. The observed shifts in bias and resolution are roughly consistent with the preliminary adaptation model, although limitations of the model are also identified.

The remainder of this chapter provides a brief description of the thesis problem area, the specific objectives of the thesis, and the empirical research performed in order to achieve these objectives.

1.1 Thesis Problem

There is widespread interest in telepresence and virtual environments and on-going development of the associated technology (Barfield and Furness 1995; Durlach and Mavor 1995; Ellis 1991; Schloerb 1995; Sheridan 1992; Vertut and Coiffet 1986a; Vertut and Coiffet 1986b). Transformations that alter the normal sensory-motor relationships of the human operators are inevitable with such systems. Some transformations are unavoidable due to limitations of the technology, while others may be imposed by design in order to improve some aspect of task performance. In either case, human operators must *adapt* to an altered sensory-motor

interaction[1] while using the system and then re-adapt to normal when they stop using the system.[2]

The effect on performance will vary depending on the type of transformation and the task. In some cases, as with most conventional tools[3], adaptation is not an issue. In other cases, performance may be degraded temporarily. In such cases, even a brief transition period could have serious consequences. For example, surgeons may someday use enhanced visual displays that make it possible to look directly inside the patient during an operation (Bucholz, Robinson, and McDurmont 1995; Dautraix and Magnin 1995; Rolland, Holloway, and Fuchs 1994; Satava 1994; Zamorano et al. 1994a; Zamorano et al. 1994b). If the display were to change the surgeon's normal hand-eye coordination it could cause a problem if he or she was forced to suddenly start or stop using the display during an operation. In still other cases adaptation may be incomplete or simply not possible, as in the case of a large time delay.

Engineers must know how the operator's performance will change as a result of adaptation in order to make rational trade offs among design alternatives. Unfortunately, even though sensory-motor adaptation has been studied extensively (Rock 1966; Welch 1978; Welch 1986; Welch and Warren 1986), there are no general models that can be used to predict the extent of adaptation or how long it will take (Durlach and Mavor 1995, pp. 103-104).[4] Consequently, fundamental research is required in this area to guide the development of telepresence and virtual environment systems.

1.1.1 Focus of the Research

Stereopsis is the perception of depth that results when the different views from the two eyes are fused into a single mental image. Stereoscopic displays create this perception by presenting each eye with a separate two-dimensional visual image corresponding to the appropriate

---

[1]The interaction is altered in comparison to the operator's normal interaction with the world.

[2]One assumes that the exposure does not cause permanent damage. Visual stress associated computer-graphic stereoscopic displays, particularly helmet-mounted displays, has been investigated by Mon-Williams, et al. (1995; 1993).

[3]It is important to note that all tools transform the interaction between the human operator and the task environment. Synthetic environments make it possible to transform the interaction at the most fundamental level, however, potentially leading to sensory conflicts to which the operator must adapt.

[4]Note that the preliminary model of adaptation is only able to predict certain aspects of performance in an absolute identification task after other aspects of adaptation have been measured (see section 7.6.1).

view. This is the most common type of three-dimensional visual display and is considered to be an essential component of telepresence and virtual environment systems. The effective interpupillary distance (IPD), which is equal to the separation that the operator's eye-pupils would have if the displayed scene was a physical environment that he or she viewed directly, is the most fundamental parameter of a stereoscopic display.

The primary concern of the thesis is how people adapt to a change of the effective IPD while performing an absolute identification task with a computer-graphic stereoscopic display. In this task the subject tries to identify one stimulus presented at random out of a discrete set. The term "computer-graphic" identifies the means by which the visual images are created in the display, namely, via a common computer monitor. An essential feature of the system is that the images consist of an array of discrete dots, or pixels that may be illuminated independently.

Absolute identification is studied here because it is a fundamental component of many practical tasks. For example, people need to estimate depth when reaching for an object in space. Absolute identification is also simple enough to model and it is the task that is considered in the preliminary adaptation model of Shinn-Cunningham. The relevance of the task to traditional adaptation studies is addressed in section 6.1.

1.1.2 Objectives

The principal objectives of the investigation are as follows:

1) Measure the effect of changing the effective IPD on human performance in certain fundamental tasks so that the results may be used as design criteria in future applications.

2) Observe changes in resolution resulting from changes of the effective IPD. This is important because most prior adaptation studies, including all investigations of adaptation to altered IPD, have looked only at changes in bias. The study of resolution is important because one expects that resolution will be improved when the effective IPD is increased, possibly leading to improved performance in certain tasks.

3) Compare the results of the experiments to predictions of the preliminary adaptation model developed by Shinn-Cunningham (1994). Shinn-Cunningham studied supernormal auditory localization in a virtual environment and successfully modeled shifts in bias and resolution resulting from changes of the effective distance between the ears (see

chapter 2). One expects that the model will also work for vision because it is based on central mechanisms of perception.

1.2 Summary of the Experiments

Table 1.1 presents a list of all the thesis experiments, giving the title, the section in the thesis where the experiment is presented, and the associated test designations. Note that the test designations identify specific tests in the experiments corresponding to the test type (and program version) as discussed in section 3.2.4.1.

Table 1.1. List of Experiments.

Title	Section	Test Designations
Calibration Experiments		
3D Experiment	4.1	P1 and P2 (v2.201)
Z-axis Experiment	4.2	T1 and T2 (v2.101)
Virtual-Virtual Experiment	4.3	T4 and T5 (v2.101)
Discrimination Experiment	5.	T10 and T11 (v2.102)
Identification Experiments		
Experiment 1	6.2	T80 (v2.101)
Experiment 2	6.3	T40 and T42 (v2.102)

Three main groups of experiments were performed: 1) calibration, 2) discrimination, and 3) identification. The transformation in all of the experiments is specified by the value of X_{ipd} (the ratio of the effective IPD of the display to the actual IPD of the human operator) and referred to as the *X_{ipd} transformation*. Only two values of X_{ipd} were used in the experiments: $X_{ipd} = 1$ and $X_{ipd} = 2$. In other words, in every experiment the effective IPD was either x1 or x2 the subject's actual IPD.

While all three types of experiments involved changes of the effective IPD, only the depth identification experiments investigated adaptation. In other words, only the identification experiments looked at changes in performance over time when the subjects were exposed to feedback about the change in the effective IPD. In contrast, the first two groups of experiments (which provided data in support of the main experiment) simply presented the subjects with visual stimuli, without any source of information that might indicate that the stimuli were transformed in some way.

The experimental apparatus, which is described fully in chapter 3, consists of a CrystalEyes stereoscopic display and a PHANToM manipulandum. The PHANToM is used for haptic position input and the display is viewed by the subject in a partially silvered mirror so that the visual and haptic fields may be superimposed. Following are brief descriptions of the experiments.

1.2.1 Calibration (Chapter 4)

Preliminary measurements were made in order to calibrate the apparatus. Then the "calibration experiments," discussed in chapter 4, were performed in order to verify that subjects perceive virtual objects, presented on the display, at the same locations as real objects. In the latter tests, the subjects manually positioned a light-emitting diode (LED), attached to the endpoint of the PHANToM, so that it appeared to be at the same location as a virtual target on the display. A follow-up experiment was also performed in which the relative positions of two virtual objects were compared. It was observed that the calculated and perceived positions of virtual objects agreed within the accuracy of the test, allowing for small discrepancies from the nominal (intended) position due to *display pixelation* (see section 2.2.2.4). Positioning accuracy also improved in the virtual space when the effective IPD was increased.

1.2.2 Discrimination (Chapter 5)

The subjects' ability to discriminate the depth of two virtual targets on the display was measured in a forced-choice experiment. This experiment, discussed in chapter 5, was performed in conjunction with the second identification experiment (section 6.3), so that the results could be used in the quantitative analysis described in chapter 7. Both experiments involved the same set of subjects as well as nearly identical visual depth stimuli in which the images, constituting the two virtual targets, were separated laterally so that depth resolution was not limited by the display.[5] The results show that discrimination of the given

---

[5]When the targets are close together typical human stereo thresholds are an order of magnitude smaller than the minimum displayable depth on the display, determined by the minimum horizontal spacing of the pixels. As a result, subjects easily detect the minimum depth change and resolution is limited by the display. Increasing the separation of the targets makes it harder for the subjects to judge the relative distance so that it becomes possible to measure the subjects' ability to discriminate depth using stimuli presented on the display.

stimulus improved, as expected, when the effective IPD was increased. A psychophysical scale was also derived.

1.2.3 Identification (Chapter 6)

The main experiment of the thesis is the second identification experiment, discussed in chapter 6. Adaptation to changes of the effective IPD was observed in this experiment. Two depth identification experiments were performed because the results of the first experiment (section 6.2) were confounded and no change in resolution was observed when the effective IPD was doubled.[6] The design of the second experiment (section 6.3) solved the problems by using a stimulus like the one used in the discrimination experiment.

In both identification experiments the task is to identify a visual depth stimulus presented at random out of a standard set of depths. Each subject is asked to perform the task repeatedly and the average performance is observed. Then, over the course of the experiment, the effective IPD is changed in order to see how performance varies.

1.3 Overview of the Document

Following the general introduction in this chapter, chapter 2 presents further background for the thesis. Chapter 3 describes the general experimental methods and, then, chapters 4, 5, and 6 are each devoted to one of the major groups of experiments as described above. Chapter 7 presents a quantitative analysis of the second identification experiment, where the results of the experiment are compared to the predictions of the preliminary adaptation model. Finally, the thesis concludes with chapter 8 which summarizes the results of the thesis, presents design rules based on the results, and suggests future work. Additional material is included in appendices A-E.

[6]Actually the resolution did increase for the smallest depths. See sections 6.2.2 and 6.2.3.

2. BACKGROUND

This chapter presents five main areas of prior work that are relevant to the thesis. Depth perception and stereoscopic displays are discussed first (sections 2.1 and 2.2), leading to the definition of display pixelation. Adaptation is discussed next (section 2.3), beginning with the basic experimental paradigm. Prior altered-IPD adaptation studies and Shinn-Cunningham's Ph.D. thesis are also reviewed. Then a brief summary of the preliminary theory of intensity resolution is presented (section 2.4), followed by a discussion of the preliminary adaptation model (section 2.5). The latter model is an extension of the former, developed by Shinn-Cunningham to explain the results of her thesis experiments related to auditory localization (1994). A major objective of the present thesis is to provide further evidence, either for or against the model, in connection with visual depth perception.

2.1 Depth Perception

Vision is the most reliable source of depth information for humans and visual depth perception has been widely studied (Arditi 1986; Boff and Lincoln 1988b, section 5.9; Buser and Imbert 1992; Goldstein 1989; Graham 1965; Sekuler and Blake 1985). Roughly speaking, depth perception is the ability to judge distances from the observer.[1] One must be careful, however, to distinguish absolute distance judgments (distance from the observer to an object) and relative distance judgments (the depth of the object itself) because they differ markedly (Sekuler and Blake 1985, p. 217). Indeed, relative depth judgments are many times more precise than absolute distance judgments. Following is a brief summary of the various depth *cues* that have been identified.[2]

2.1.1 Accommodation and Convergence

When a person with normal vision fixates on (i.e., looks at) an object, both eyes automatically accommodate (i.e., focus) and converge on the object. Normally, accommodation and vergence operate together such that changes in one are accompanied by appropriate changes in the other (Edgar and Bex 1995). Consequently, one may define quantities such as the

---

[1]See Webster's Third New International Dictionary, 1965 ed., s.v. "depth perception."

[2]Cue theory seeks to identify the information that is provided by specific physical stimuli (cues). According to this theory, perception involves the combination of individual cues. Another theoretical approach, the ecological approach proposed by J. J. Gibson, focuses on information that is directly available in the environment without the need for further mental processing (Goldstein 1989, p. 227).

accommodative-vergence gain. By definition this quantity is equal to the change in accommodative-vergence (i.e., vergence that is associated with a given level of accommodation) per unit change in accommodation (Fisher and Ciuffreda 1990).

Kinesthetic sensations associated with the muscles used in accommodation and vergence provide oculomotor cues to the absolute distance of the object (Sekuler and Blake 1985, p. 218). Specifically, knowing the convergence angles, one could triangulate on the fixation point. Similarly, the closer an object, the more the lenses of the eyes must be deformed to bring the object into focus. Accommodation also provides a relative depth cue because points nearer or farther than the fixation point are blurred (Graham 1965, p. 504).

Both accommodation and convergence are near-field depth cues, probably only effective at distances of less than 2 or 3 meters (Goldstein 1989, p. 229). And, even within that range, neither is a particularly effective cue (Graham 1965, p. 520). Accommodation appears to be the weaker of the two cues, in that convergence is found to dominate accommodation in the case of a conflict (ibid.). Ultimately, however, stronger cues to depth are provided by the visual images that are presented to the eyes.

2.1.2 Monocular Cues

Normally there are several visual depth cues that are available to an observer using only one eye (Boff and Lincoln 1988b, pp. 1054-1055). Static sources of monocular depth information include: interposition, relative size, linear perspective, light and shade, and aerial perspective. Other monocular depth cues involve motion of the observer in the environment, for example motion parallax and perspective change (see ibid., or any introductory text on visual perception, for descriptions of these and other cues). Further discussion of monocular depth cues is presented in section 3.3.

2.1.3 Binocular Disparity

Of all the depth cues, binocular disparity is the most important in regards to relative distance judgments in the near field.[3] Binocular

---

[3]Clearly monocular cues can dominate in some cases (e.g., interposition). Also, Surdick, et al. (1994) performed a systematic study of depth cues using a stereoscopic display and found that the effectiveness of binocular disparity is roughly the same, or slightly worse than certain

disparity is the basis for the perception of depth (relative distance) known as *stereopsis* (Patterson and Martin 1992, p. 669). In normal viewing, disparity results from the lateral separation of the two eyes as illustrated in figure 2.1. In this figure, the points p and p' represent the centers of the pupils and the arcs below p and p' represent the two retinas.

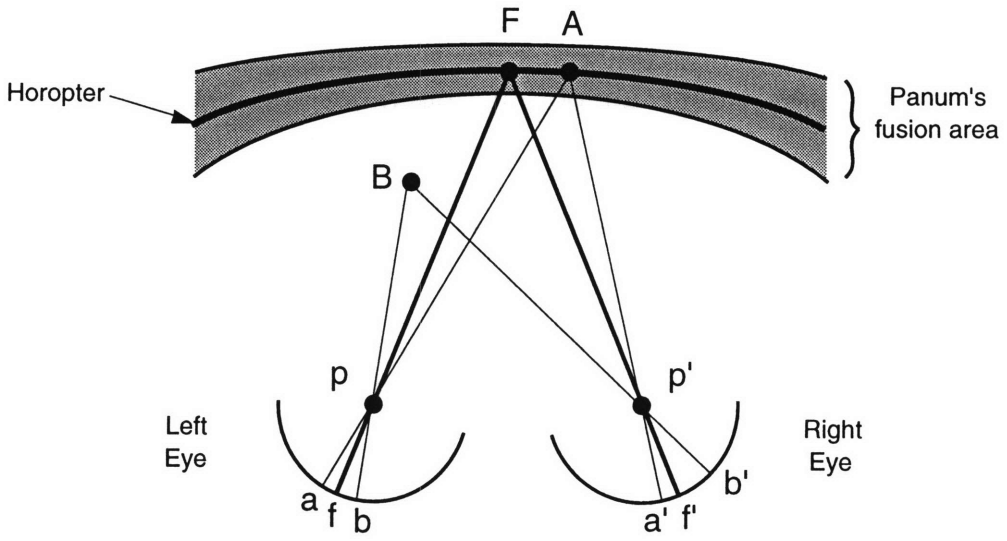


Figure 2.1. Top view depicting two eyes fixating point F (Patterson and Martin 1992, p. 670).

When a person fixates on a point F as shown in the figure, the images of the point fall on the retinas of the left and right eyes at f and f', respectively. The points f and f' are *corresponding points* on the two retinas (Goldstein 1989, p. 239), located at the centers of the foveas (Hodges and Davis 1993, p. 34). Indeed, the images of any fixated point would fall on f and f', because the eyes converge on the fixation point (i.e., they look directly at it), as shown in the figure for F. Similarly, every point on the retina of one eye has a corresponding point on the retina of the other eye. Specifically, the correspondence is such that if you placed one retina on top of the other, all of the corresponding points would line up.

---

monocular cues (foreshortening, linear perspective, and texture gradient) at a distance of 1 m, and significantly worse at 2m. The results are strongly dependent on the specific stimuli used to represent these cues in the test, however, and may not contradict the idea that binocular disparity is the most important cue *in general*. This is particularly true in the near field, as in the thesis experiments, at a distance of less than 1 m.

The *horopter* for a given fixation point (F) is defined as the surface in space consisting of all points whose images on the two retinas are corresponding points (Goldstein 1989, p. 272). In general, a given point X in space lies on the horopter for F if and only if the angle at the center of the pupil formed by the line segments to the retinal images of X and F is the same for both eyes. Thus, for example, in figure 2.1, point A is on the horopter for F ($\angle apf = \angle a'p'f'$), whereas point B is not ($\angle bpf \neq \angle b'p'f'$).

The lateral retinal image disparity (η) is defined as the difference between these angles (Kalawsky 1993, p. 55). Thus, disparity is zero for points on the horopter (e.g., $\angle apf - \angle a'p'f' = 0$), whereas disparity is non-zero for points not on the horopter (e.g., $\angle bpf - \angle b'p'f' \neq 0$). Further, the disparities of points located on opposite sides of the horopter have opposite signs. Typically this is described by saying that points in front of the horopter (e.g., B) have *crossed* disparity while points behind the horopter have *uncrossed* disparity. In general, the magnitude and the direction (crossed or uncrossed) of the disparity provides a cue about the distance of an object relative to the distance of the fixation point.

2.1.3.1 Stereopsis Regions: Objects that lie within a certain region of space around the horopter are perceived as three-dimensional. In other words, the brain fuses the two two-dimensional images from the eyes into a single three-dimensional percept. As the object moves further from the horopter (i.e., as the magnitude of disparity increases), however, it becomes impossible to fuse the images. And, as a result, the object appears double; a condition known as *diplopia*.

The region around the horopter in which fusion occurs is known as *Panum's fusion area* (see figure 2.1). In addition to Panum's area, two other stereopsis regions have been defined (Barfield and Furness 1995, pp. 153-155). The first region corresponds to quantitative or *patent* stereopsis. Depth judgments of objects within this region are monotonically related to the disparity. The region of patent stereopsis includes and extends beyond the region of fusion. In other words, quantitative estimates of depth are possible even in the case of double vision. When the disparity gets large enough, however, the perceived depth no longer increases with disparity.

Beyond the limit of patent stereopsis is the region of *qualitative* stereopsis. The perception of depth in this region is little better than the ability to discriminate whether an object is nearer or further than the fixation point. Also, vision is always diplopic and may result in

discomfort. Finally, as disparity increases further, the limit of qualitative stereopsis is reached where stereoscopic depth perception becomes impossible. Beyond this limit, objects are perceived at the fixation distance or may not appear to be at any definite distance.

The limits of the different regions vary widely across observers and between individual studies (Boff and Lincoln 1988b, pp. 1112-1113). The limits also depend on the lateral distance from the fixation point or, correspondingly, the size of the object that is observed. Patterson and Martin (1992, pp. 673-675) suggest the following representative limits for objects subtending a visual angle of 1.0 to 6.6 degrees. The same limits apply to both crossed and uncrossed disparity.

Table 2.1. Maximum Disparity for each Stereopsis Region.

Region	Disparity Limit
Panum's fusion area	20 arc-min
Patent Stereopsis	2 degrees
Qualitative Stereopsis	8 degrees

2.1.3.2 Stereoacuity and Range: In a well-known experiment, Howard measured the disparity threshold (η_t), the disparity for which observers were able to discriminate a depth difference 75% of the time.[4] "Of the 106 subjects tested, the 14 with the best stereoscopic acuity had values of η_t lying between 1.8 and 2.07 sec; the 24 with the worst acuity had values between 10.6 and 136.2 sec" (Graham 1965, p. 526).

A value of 10 arc-sec seems like a reasonably *typical* value of stereoacuity for use in calculations. Stereoacuity is affected by a large number of factors, however, as listed by Boff (1988b, pp. 1088-1089). Of particular interest in regards to the thesis experiments is the fact that acuity is reduced when the targets in the test are separated laterally (see chapter 5). Graham estimates a maximum range for stereopsis on the order of 500 meters (1965, p. 525).[5]

2.1.3.3 Related Issues: There are many other aspects of binocular disparity that are discussed in the references cited here. Following are three issues that are relevant to the thesis experiments: 1) vertical disparity, 2) color stereopsis, and 3) stereoblindness.

---

[4]This type of threshold is also called the just noticeable difference (JND).

[5]This calculation assumes a conservative value of η_t (30 arc-sec) and an IPD of 6.5 cm.

The type of retinal disparity discussed up to this point is, of course, lateral (horizontal) disparity. Vertical disparity is also possible, although it has different perceptual consequences. In particular, it does not improve stereopsis; however, if large enough, it does cause double vision. This is generally more disturbing than diplopia arising from horizontal disparity. In the case of a stereoscopic display, it is recommended that vertical disparity not exceed 3.4 arc-minutes (Boff and Lincoln 1988b, pp. 1064-1065).

Color stereopsis refers to the case where there is a difference in the apparent depth because the stimuli are different colors (Yeh 1993, pp. 58-59). The effect is caused by refraction in the eyes, leading to disparity between images of different colors.

Finally, it has been suggested that as many as 5 to 10 percent of the general population may be stereoblind (Sekuler and Blake 1985, p.230). Clearly any experiment, or other application involving a stereoscopic display, must take this factor into account.

2.2 Stereoscopic Displays

Stereoscopic displays date back to the first report, by Wheatstone in 1838, that disparity is the cue for stereopsis (Patterson and Martin 1992, p. 669). Although the technology has evolved from the stereoscopes of the 19th century, present-day systems still operate on the same basic principle. The essential idea in all these displays is that separate two-dimensional images are presented to each eye, with viewpoints that are offset in order to induce laterally disparity. A number of people have worked on the problem of modeling these displays in recent years (Davis and Hodges 1995; Grinberg, Podnar, and Siegel 1994; Grinberg, Podnar, and Siegel 1995; Hodges and Davis 1993; Hodges and McAllister 1993; Robinett and Holloway 1995; Robinett and Rolland 1992; Rolland, Ariely, and Gibson 1995). The following section summarizes the operating principles for the type of display used in the thesis experiments and identifies some relevant limitations.

2.2.1 Principles of Operation

A time-multiplexed or field-sequential stereoscopic display is used in the present experiments (Bos 1993; Lipton 1993; Lipton, Akka, and Meyer 1991). This type of display presents the images to the two eyes alternately, on the same video monitor, at a rate of about 60 Hz. The human operator views the display through special glasses with liquid

crystal shutters that actively turn on and off in sync with the images. The arrangement is such that each eye sees only the appropriate view and the shuttering happens too fast for the operator to detect. Further details of the system are discussed in section 3.2.1.1.

Figure 2.2 illustrates some important features of the display. Starting on the left, one imagines a *virtual space* filled with virtual objects that are to be displayed to the human operator. Equivalently, one might think of this as a real scene viewed by two cameras located at the effective eye positions. Projections of the objects onto the virtual display plane are calculated for the two projection centers, corresponding to the effective eye positions in the virtual space.

Note that there is some controversy about what part of the eye corresponds to the center of projection (Rolland, Ariely, and Gibson 1995, pp. 31-32). The present study uses the entrance pupil to define the *eye position* in accordance with a suggestion by Ogle (ibid.); hence the prominence of IPD throughout. The algorithm used to do the projection calculation in the experiments is discussed in section 3.2.4.3.

Moving to the right in the figure, the calculated images are presented to the human operator on the display screen. This, of course, happens in the real (i.e., physical) *experimental space*.[6] The screen view at the top of the figure shows the display screen as it would appear to an observer not wearing the special glasses. In other words, it shows both the left and the right-eye views at the same time. Note that the left and right images of some objects are displaced laterally (horizontally) relative to one another. The horizontal distance that the left and right images are displaced on the screen is called the *parallax.*

Parallax is the fundamental parameter of the display that is related to lateral disparity and, hence, stereopsis.[7] Note that virtual objects located near the virtual display plane have little or no parallax, while the magnitude of the parallax increases as objects move away from this plane. The sign of the parallax depends on whether the virtual object is in front or behind the display plane. Specifically, the right-eye view image is to the left of the left-eye view image on the screen when the virtual object is in front (i.e., between the effective eye positions and the virtual display plane) and vice versa when it is behind.

---

[6]Nominally one might call this the viewing space, but the term "experimental space" is used here to be consistent with discussion in later chapters.

[7]Note that the disparity also depends on the convergence angle (see section 2.2.2.1).

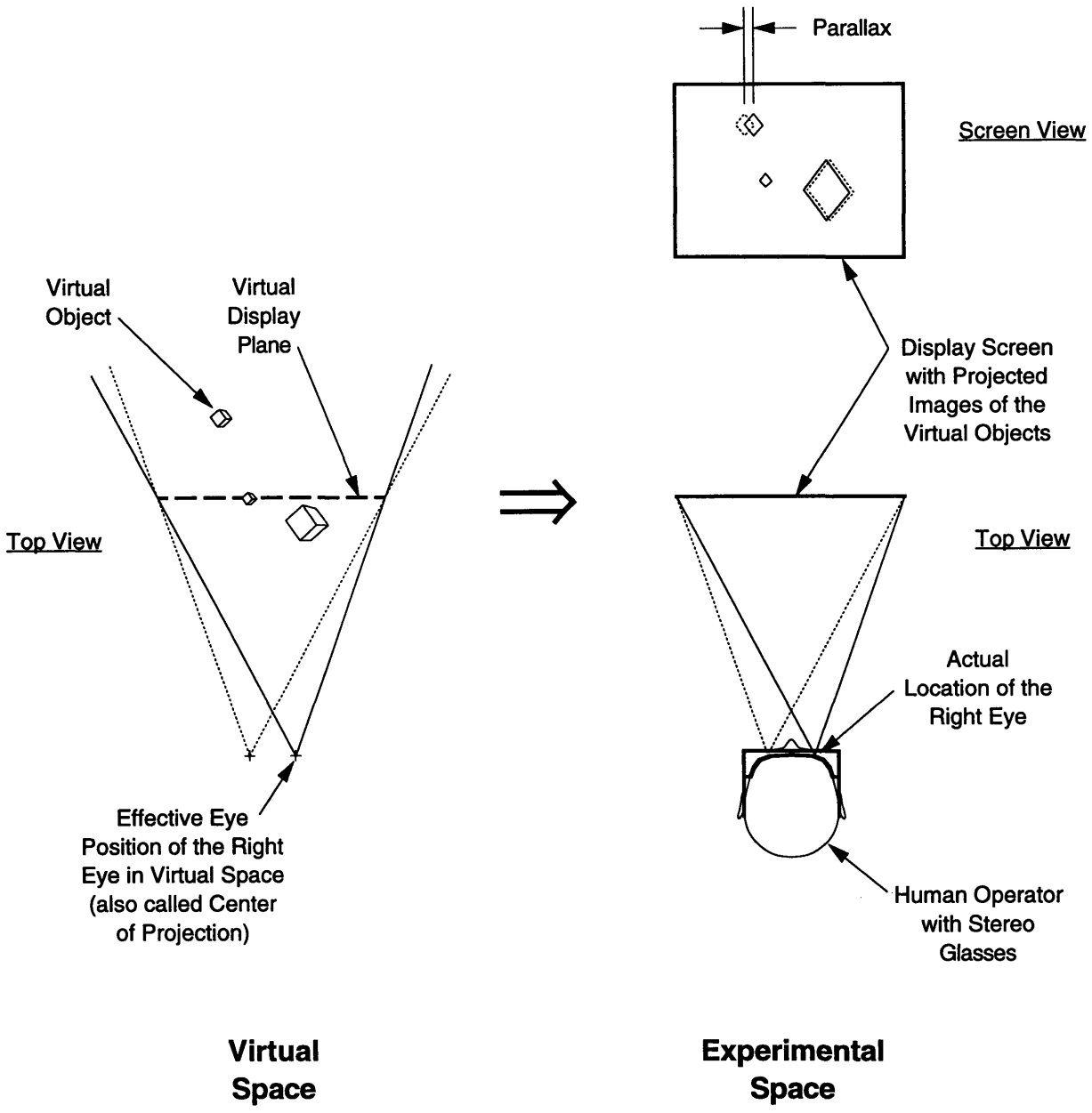


Figure 2.2. Important features of a stereoscopic display like the one used in the thesis experiments.

2.2.1.1 Virtual Space and the X_{ipd} Transformation: Assuming that the system is working properly, i.e., that the glasses are synchronized with the display and the projections of the virtual objects have been calculated correctly, the human operator should see three-dimensional objects corresponding to the objects in the virtual space. Further assuming that the virtual and experimental spaces are exactly similar, the size and distance of the objects perceived by the operator should be the same as in the virtual space. In other words, assuming that 1) the distances between the effective eye positions and the actual eye positions are equal, 2) the distances to the display planes are equal, and 3) all of the angles are similar, then one would expect the perceived space to correspond exactly to the virtual space.

Although this seems straightforward when one looks at figure 2.2, it can be confusing if you think about actually trying to determine the distances in virtual space. After all, what are the units in virtual space? In other words, what is the scale? This is a significant issue in the present study where the fundamental transformation, X_{ipd}, is defined as the ratio of the effective IPD to the actual IPD of the human operator (see section 1.2).

A simple answer to the question of scale in the virtual space is that it is determined by the size of the display plane. Specifically, the distance between any two points on the virtual display plane is identical to the distance between the corresponding points on the physical display screen. In other words, given two points on the physical display screen that are 1 cm apart, the corresponding points on the virtual display plane define the distance of 1 cm in the virtual space. Using this scale one can measure the effective IPD and the value of X_{ipd} is defined unambiguously.[8]

2.2.2 Limitations

A wide range of studies have looked at issues of perception and performance with stereoscopic displays (Hsu et al. 1996; Nemire and Ellis 1993; Patterson, Moe, and Hewitt 1992; Schor 1987; Surdick et al. 1994; Utsumi et al. 1994a; Utsumi et al. 1994b; Yeh 1993; Zhai and Milgram 1994). Following is a discussion of four limitations that are relevant to the present experiments.

2.2.2.1 Accommodation and Convergence: One of the major limitations of existing stereoscopic displays is the conflict between

---

[8]The value of X_{ipd} may also be defined unambiguously for a stereo television system.

accommodation and convergence associated with viewing images that have non-zero parallax (i.e., virtual objects that are not in the plane of the display). This happens because, when the human operator fixates on the images of a virtual point, the eyes converge until there is no disparity for that point just as they would when fixating on a physical object at the corresponding location. In order to see the point clearly, however, the operator must focus on the display screen. Thus, unlike normal vision, where accommodation and convergence operate together (see section 2.1.1), the level of accommodation must stay constant (corresponding to the distance of the screen) while convergence varies as the operator fixates on different points in the display space.

This conflict may cause discomfort, blurring, double vision, and possibly disturbances of binocular vision over time (Mon-Williams and Pascal 1995; Mon-Williams, Rushton, and Wann in press 1995; Mon-Williams, Wann, and Rushton 1993; Rushton, Mon-Williams, and Wann 1994; Rushton and Wann 1993; Wann, Rushton, and Mon-Williams 1995; Yeh 1993). At the very least it limits the useful depth range of the display. A practical rule of thumb is that the absolute magnitude of the parallax should not subtend a visual angle of more than 1.5° (Lipton, Akka, and Meyer 1991, p. 10).

2.2.2.2 Crosstalk: Crosstalk refers to the case where the image intended for one eye is also seen faintly by the other eye (Lipscomb and Wooten 1994; Lipton, Akka, and Meyer 1991). This effect, also called ghosting, sometimes looks like a faint shadow of the displayed object that becomes more noticeable with increasing parallax.

Crosstalk may occur because the liquid crystal shutters in the glasses do not close quickly enough or because they do not block all of the light when they are closed. Persistence of the image on the display is another culprit. In other words, the previous image (for the other eye) is still faintly visible when the view for the current eye is drawn. The latter effect depends on the color of the image, because the afterglow of the currently used phosphors tends to persist longer for blue and green than for red (Bos 1993, p. 115).

2.2.2.3 Superposition of Real and Virtual Objects: Ellis and Bucher (1995; 1994) report an effect where a virtual object appears to move closer to the observer when it is visually superimposed on a physical background. The effect was measured by having the subjects position a physical probe next to the virtual object and, then, comparing the probe position with and without a physical checkerboard object

superimposed. The effect appears to be related to a change in vergence when the physical object is introduced, although the underlying mechanism remains to be clarified.

2.2.2.4 Spatial Resolution and Display Pixelation: An obvious limitation of computer-graphic stereoscopic displays (or any discrete display like the one used in the present experiments) is that only discrete depth intervals may be presented. In particular, only integral values of parallax are possible. Further, the depth resolution of most existing stereoscopic displays, including the one used in this investigation, is well above the normal threshold of human stereoacuity. For example, McKenna and Zeltzer (1992, p. 423) calculate that a display with a 40° x 32° field of view would require the (currently unattainable) screen resolution of 4800 x 3840 pixels in order to match the visual acuity of most observers at a distance of 46 cm. Specifically the calculation assumes that foveal acuity is 30 arc-sec. The best stereoacuity thresholds are an order of magnitude smaller, however, so that the horizontal screen resolution would actually need to be ten times greater (i.e., 48000 x 3840 pixels) in order to match the depth resolution of most human observers.

More subtle effects are also observed with discrete stereoscopic displays. For example, in one experiment, displayed objects were observed to appear, then disappear, and then reappear again as they approached from a distance (Pfautz 1996, p. 50). Similarly, the size of objects will grow and shrink intermittently as they approach (or recede). Both of these effects result from rounding errors when the boundaries of virtual objects are displayed.

In general, spatial resolution on a discrete stereoscopic display is determined by the location of the observers eyes with respect to the display and the size of the pixels on the screen (Davis and Hodges 1995; Hodges and Davis 1993; McKenna and Zeltzer 1992). Figure 2.3 illustrates the geometry of the situation where the display space is partitioned into discrete indistinguishable diamond-shaped regions.[9] In other words, all points within these regions, called *stereoscopic voxels* by Davis and Hodges (1993, p. 38), are displayed by illuminating the same two pixels[10] on the screen.

The term *display pixelation* is defined here to describe this partitioning of the display space and the associated effects. Beyond the

---

[9]Pixels on the screen are assumed to be rectangular in this discussion.

[10]One pixel is the left-eye view and the other is the right-eye view of the points in the region.

obvious meaning, that the visual image is divided into pixels, the term also refers to the fact that the information is confused in the process.[11] These ideas are discussed further in section 4.4 in connection with a detailed analysis of the calibration experiments and a prior experiment (Rosenberg 1993). Indeed, display pixelation is an issue in all of the experiments (see sections 6.2.3.1 and 6.3). Further, it is related to the well-known problem of *aliasing* in computer graphics (Foley et al. 1990). The use of *antialiasing* techniques to reduce the effects of display pixelation is discussed in connection with future work (see section 8.3.5).

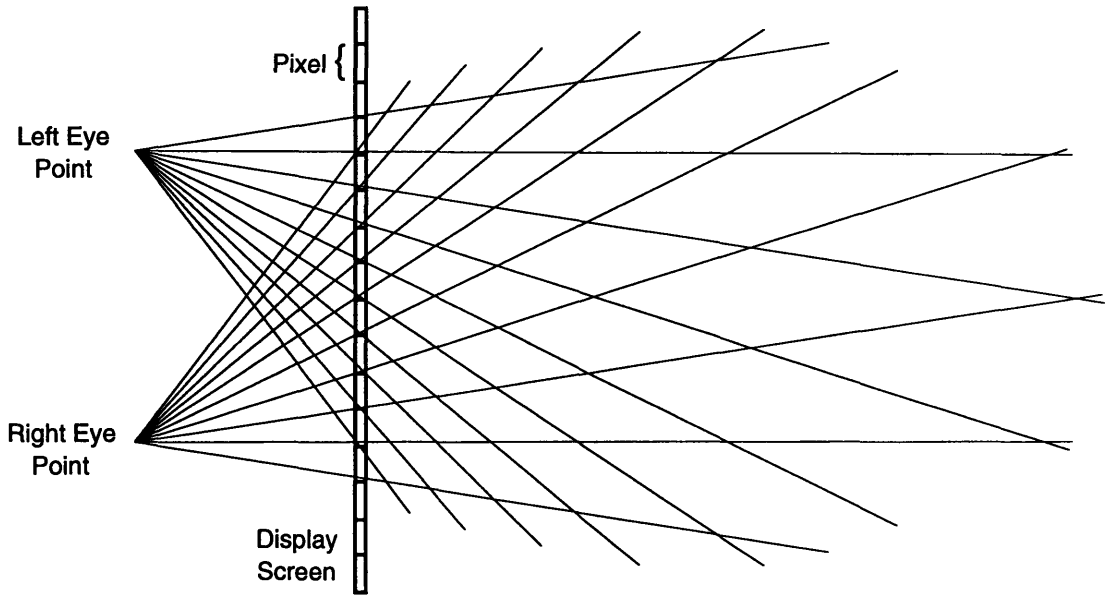


Figure 2.3. Cross section of stereoscopic voxels (Hodges and Davis 1995, p. 39).

2.3 Adaptation

Adaptation has been studied extensively in connection with perception and perceptual-motor performance (Rock 1966; Welch 1978; Welch 1986; Welch and Warren 1986). In these studies the subject's normal sensory-motor interaction with the world is transformed[12] and the affect on performance is observed for a given task. For example, in a test that is

---

[11]The latter meaning is derived from the word "pixilation" which is defined in Webster's Third New International Dictionary, 1965 ed., s.v. "pixilation." Pixilation refers to the quality or state of being pixilated; i.e., 1) mentally unbalanced: touched, daffy, 2) not altogether clear or coherent: bewildered, confused.

[12]The technique is also called "rearrangement" (Held 1972, p. 372) in order to suggest that the experiments focus on transformations where the same information is presented to the subject; just in a different manner.

representative of many early studies, the subject looks through prisms that displace the visual field to the right and the affect on mean pointing accuracy is observed. Figure 2.4 uses the example to illustrate the basic adaptation paradigm. Note that not all of the indicated steps are included or easily distinguishable in every experiment.

2.3.1 Experimental Paradigm

In the paradigm illustrated in figure 2.4, the subject's *initial performance* is first observed without the transformation. The figure shows that the subject in the example is initially able to point directly at the target with zero mean error. Then the *initial effect* of the transformation is observed. In the example, the subject points at the apparent position of the target to the right of the actual position.

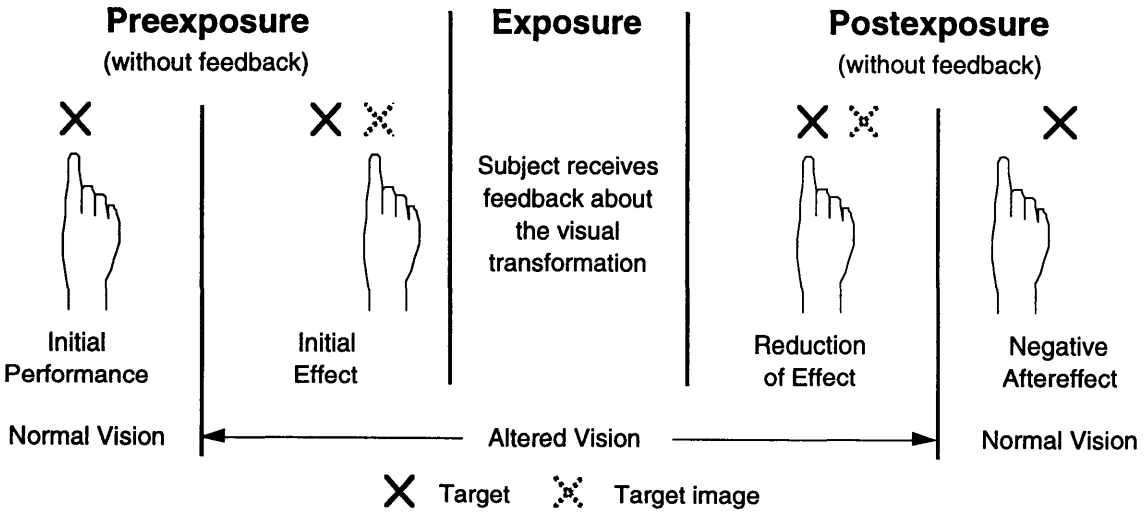


Figure 2.4. Adaptation paradigm illustrated by an experiment in which the subject attempts to point at a target that is visually shifted to the right (Welch 1978, p. 7).

Note that in both of these observations, one assumes that the subject does not receive feedback about the correct response. In the example, the subject might make a single, quick (open loop) movement so that his or her response is not significantly affected by visual feedback. Alternatively, repeated trials could be made where the subject's hand is not visible. For example, the target might be placed on a table and the subject would point at the target while reaching under the table.

After the initial measurements, the subject is given feedback about his or her performance in the presence of the transformation. In the example,

the subject might perform repeated pointing trials while looking at his or her hand through the prisms. This experimental step, called the *exposure*, is intended to affect the subject's performance. Note that the word "exposure" refers to the fact that the subject is given some kind of feedback about the transformation, not simply that the transformation is present. The transformation is also imposed on the subject during the *altered* pre- and postexposure measurements.

Nominally, during the exposure when adaptation takes place, the performance returns to the level that was observed before the transformation was imposed. In other words, it returns to the *initial performance* level.

The figure shows the subject in the example to be fully adapted after the exposure. It can be seen in the first postexposure measurement, that the subject now points at the actual target position in spite of the visual transformation imposed by the prisms. The shift in performance with the transformation, from preexposure to postexposure, is called the *reduction of effect*. A further measure of adaptation is the presence of a *negative aftereffect* when the transformation is finally removed. The figure shows the fully-adapted subject, at this stage in the example, pointing to the left of the target by an amount equal and opposite to the initial effect.

The adaptation rate may be observed either during the exposure period or through repeated pre/postexposure measurements (with the transformation) in a series of short exposures. The rate at which the subject re-adapts to normal, after the transformation is removed, is observed in a similar manner via periodic postexposure measurements (without the transformation).

2.3.2 General Effects

The degree and the rate of adaptation vary significantly depending on the type of transformation and the specific experimental task. Indeed, the situation is extremely complex and still poorly understood. In particular, there are no models that may be used to predict the results of adaptation experiments in general (Durlach and Mavor 1995, pp. 103-104).

For obvious practical reasons, adaptation experiments are typically designed to investigate transformations and tasks where the subject adapts relatively quickly (on the order of minutes to hours). In many cases special care is needed to observe aftereffects because subjects re-adapt

too rapidly when the transformation is removed. Re-adaptation is also spontaneous in some cases.

While complete or nearly complete adaptation is observed in some cases, in most experiments the subjects do not fully adapt (Welch 1978, p 278). Indeed in some cases no adaptation is observed. Three cases where little or no adaptation is observed are (Welch 1986, pp. 24-8 and 24-9): 1) inconsistent or unstable transformations, 2) experiments where the exposure does not provide the subject with salient information about the transformation, and 3) transformations that include a significant time delay (Held and Durlach 1987, p. 28-7).

2.3.3 Altered IPD

There have only been a handful of prior studies of adaptation to altered IPD.[13] In all of the prior investigations the subjects viewed physical objects through a telestereoscope.[14] The telestereoscope uses mirrors to increase or decrease the subject's effective IPD.

In the first such study (Wallach, Moore, and Davidson 1963), the subjects observed a rotating wire form through a telestereoscope at a fixed distance. Two mirror arrangements were used in the experiment, as shown in figure 2.5. In the first case, the effective eye separation was increased by 7.6 cm (corresponding to a mean X_{ipd} of about 2); in the second case, it was reduced by 2.8 cm (corresponding to a mean X_{ipd} of about 0.6).[15] The arrangement of the mirrors was such that vergence was approximately the same as for a direct view of the wire form.

Two main effects were observed. First, qualitatively, subjects reported that the wire form appeared to distort as it rotated during the exposure. Also, immediately after the exposure when the form was viewed directly (i.e., without the telestereoscope), the form appeared to distort in the

---

[13]Prior related work has been done to study performance effects of altered IPD in teleoperation and other applications (Kenyon 1981; Pepper, Cole, and Spain 1983; Pepper, Smith, and Cole 1981; Spain 1986), however, these were not adaptation studies. In other words, they did not investigate how performance changes over time as a result of exposure to altered IPD.

[14]Perhaps the first description of the telestereoscope is given by Helmholtz (Helmholtz 1962, pp. 310-312; Judge and Bradford 1988). Note Helmholtz's *Treatise on Physiological Optics* also contains one of the first descriptions of prism adaptation (1962, pp. 246-247) although he states in a footnote, "The experiment, practically as described here, was given by Czermak in Wiener Berichte. XVII, pp. 575-577."

[15]The authors use the value of 6.5 cm for the normal mean interocular distance although it seems unlikely that this is the actual mean IPD, specifically calculated for the subjects, since the same value is used for different groups of subjects.

opposite sense (i.e., the expected negative aftereffect occurred). Second, in a quantitative test, it was observed that subjects estimates of the depth of the wire form changed after watching it rotate for 10 minutes. The depth estimates were made immediately before and after the exposure while looking through the telestereoscope with the wire form standing still. Decreases in the mean estimated depth of 16% and 20% were observed when the effective IPD was increased ($X_{ipd} \approx 2$) in two tests, while mean increases of 15% and 18% were observed in two tests where the IPD was decreased ($X_{ipd} \approx 0.6$) (Wallach, Moore, and Davidson 1963, pp. 193-196).

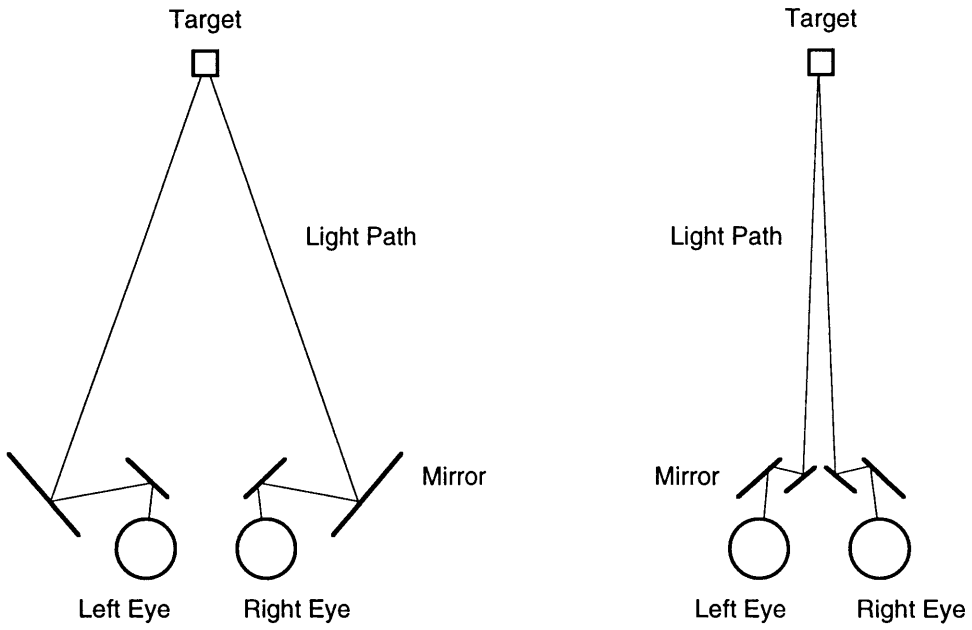


Figure 2.5. Telestereoscope arrangements for enhancing (left) and reducing (right) retinal disparity (Wallach, Moore, and Davidson 1963, p. 192).

These initial findings were investigated further in the same laboratory (Wallach and Karsh 1963a; Wallach and Karsh 1963b) and later reproduced by others (Epstein 1968). The proposed explanation for the effects is based on the fact that binocular disparity and the kinetic depth effect (KDE) provide conflicting depth cues during the exposure. The discrepancy is presumed to cause a recalibration of disparity by the KDE.

A later study (Fisher and Ebenholtz 1986) disputed the initial explanation for the observed adaptation. Specifically, the later experiments (using an arrangement similar to Wallach's apparatus) demonstrated that the KDE was neither necessary nor sufficient to

produce a significant reduction in binocular apparent depth. Instead, the later investigators suggest that the observed effects result from a shift in perceived distance to the wire form. See Ritter (1977; 1979) for an investigation of the relationship between disparity, viewing distance, and perceived depth.

Other studies investigated adaptation resulting from a more "natural" exposure to altered IPD. In these studies, the mirrors of the telestereoscope are aligned parallel to one another such that vergence is only correct when fixating on objects at infinity (see figure 2.6). Two groups observed changes in various oculomotor quantities such as the accommodative-vergence gain (Fisher and Ciuffreda 1990; Judge and Miles 1985) following a 30 minute exposure. The effect on performance in a one-handed ball-catching task was also investigated (Judge and Bradford 1988). In the latter case, the performance, which was measured repeatedly during the exposure, recovered fully to normal (i.e., 100% adaptation was observed) in tens of trials.

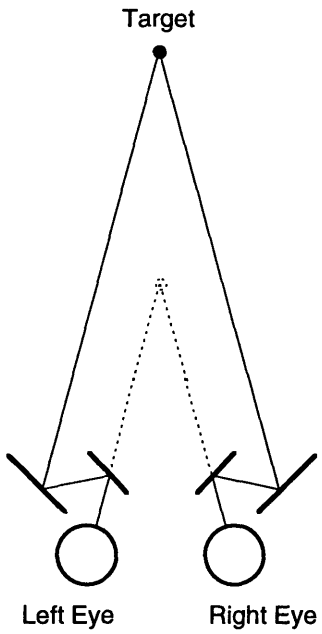


Figure 2.6.

2.3.4 Shinn-Cunningham

Most prior adaptation studies neglect resolution. Shinn-Cunningham was the first to investigate changes in both bias and resolution in an adaptation experiment (Shinn-Cunningham 1994). A virtual environment is used in these experiments to investigate adaptation to transformed auditory localization cues. Specifically, changes in resolution and bias in an absolute identification task are observed when the effective spacing between the listener's ears is increased. The exact transformation was non-linear (see equation 2.6) due to limitations of the apparatus, such that localization cues are enhanced in front of the subject and diminished to the side.

In the main version of the experiment, the subjects were randomly presented with a synthesized acoustic stimulus and asked to identify the direction of the sound. The subjects responded by selecting one of 13 positions indicated by visual targets. The targets' positions ranged from -60° to +60°, in 10° increments, with 0° directly in front of the subject.

Shinn-Cunningham observed the expected shifts in bias: initial effect, reduction of effect, and aftereffect. She concluded that, while the

exposure reduced the initial shift in mean error by less than 50%, the subjects adapted almost completely to a linear approximation of the transformation. Resolution in front of the subject was also improved initially, as expected, when the transformation was first imposed. Interestingly, the resolution also decreased slightly during the exposure although the postexposure resolution was still significantly improved over the untransformed case.

The preliminary adaptation model, developed by Shinn-Cunningham, successfully explains the observed shifts in resolution and bias. That model, in turn, is based on the preliminary theory of intensity resolution.

2.4 Preliminary Theory of Intensity Resolution[16]

The preliminary theory of intensity resolution is the first step in an effort to develop a unified theory of intensity perception by Durlach, Braida, and their associates (Durlach and Braida 1969). Subsequent experiments and theoretical developments are presented in a series of papers (Braida and Durlach 1972; Braida et al. 1984; Lippmann, Braida, and Durlach 1976; Pynn, Braida, and Durlach 1972). The preliminary theory is composed of two parts: a *decision* model and an *internal-noise* model.

2.4.1 Decision Model

The decision model is a special case of Thurstone's *law of categorical judgment* (Torgerson 1958, chapter 10). The model, which is also related to the theory of signal detection (Gescheider 1985; Green and Swets 1988), describes a probabilistic mapping between a set of physical stimuli presented to an observer in a perception experiment and the set of subsequent responses.[17] In the model it is assumed that each presentation of a physical stimulus gives rise to a different value of a random *decision variable*, where the decision variable is represented on a *decision axis* in the observer's mind.

---

[16]Although the theory was originally developed in connection with the perception of sound *intensity*, the underlying models are quite general and have been applied to a wide range of perceptual phenomena.

[17]Note that one may interpret the observer's response in such experiments as being directly correlated to the percept that is experienced by the observer. Specifically, if A is a given stimulus in the experiment, one assumes that the observer responds "A" when he or she perceives A. The relationship between decision theory and perception is also discussed by Swets, Tanner, and Birdsall (1964, pp. 54-55).

Figure 2.7 illustrates an example where three physical stimuli (S_1, S_2, and S_3) give rise to random distributions of sensations on a particular decision axis. In the example the observer is presented with one of the stimuli at random and his or her task is to identify it. Following a presentation of any one of the stimuli, a particular value of the decision variable is generated. The observer must then decide (make his or her best guess) about which stimulus caused the sensation.

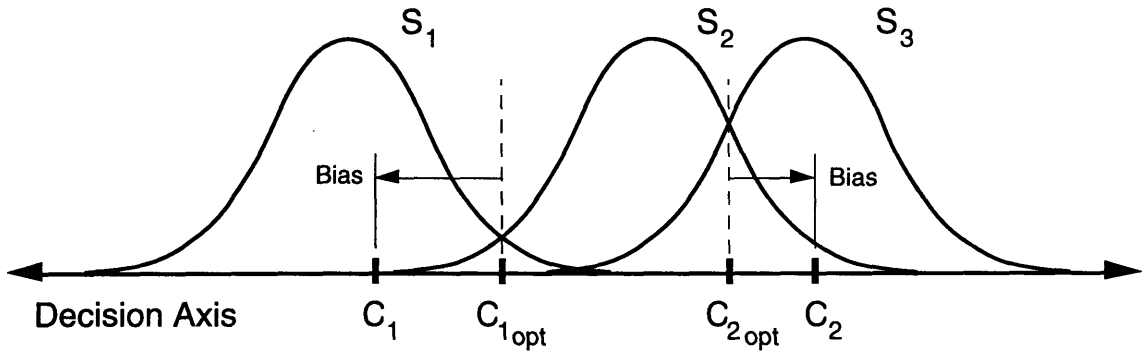


Figure 2.7. Decision Model.

The model assumes that the observer makes his or her response based on where the sensation falls on the *decision axis* with respect to certain *criteria*. In the example shown in the figure, there are two criteria: C_1 and C_2 (C_{1opt} and C_{2opt} identify optimum locations of the two criteria as will be discussed shortly). If the given value of the decision variable is less than C_1, the observer responds that the stimulus is S_1. The observer selects S_2 if the value is between C_1 and C_2 , and S_3 if the value is greater than C_2.

An important feature of the decision model is that it allows one to distinguish sensitivity and bias. Sensitivity is a measure of the observer's ability to distinguish (or resolve) different stimuli, while bias is a preference for giving one response over another independent of the physical stimuli. For example, the observer in the figure should be able to distinguish S_1 from S_2 more reliably than he or she can distinguish S_2 from S_3. Specifically, the resolution is better between S_1 and S_2 because there is less overlap of the two distributions.

Ultimately, resolution between a pair of stimuli is a function of the distance between the means of the two distributions on the decision axis relative to the variance of the distributions. For example, in the case of Gaussian distributions with equal variance, the normalized resolution is given by:

$$d_{ij}' = (\mu_j - \mu_i)/ \sigma \qquad (2.1)$$

where:

d_{ij}' = normalized resolution between any two stimuli $i < j$
μ_c = mean of the distribution of stimulus c ($c = i, j$)
σ = standard deviation of the distributions.

Bias, on the other hand, is related to the placement of the criteria. The optimal locations of the two criteria in the example are shown in the figure as C_{1opt} and C_{2opt}. *Optimum* performance is defined here, and for the remainder of the discussion, as the case where the observer is most likely to select the correct stimulus.

The observer's criteria may differ from optimum for many reasons. Indeed, a fundamental feature of the model is that the observer's criteria may change depending on the circumstances. In particular, the observer may use a different measure of performance than maximizing the number of correct responses. For example, the figure illustrates a case where the observer's criteria differ from optimum such that he or she has a preference or *bias* for responding "S_2." This might be the case if the reward for identifying S_2 every time it occurs is high, while there is little penalty for incorrectly responding S_2 when it does not occur. The normalized criterion bias is defined by the following equation in the case of stimuli with Gaussian distributions of equal variance:[18]

$$B_i = (1/2 (\mu_i + \mu_{i+1}) - C_i)/ \sigma \qquad (2.2)$$

where:

B_i = normalized bias of criterion C_i
μ_c = means of the distributions of consecutive stimuli, c ($c = i, i+1$)
σ = standard deviation of the distributions.

---

[18]Equation 2.2 is only well defined when the number of stimuli and the allowed number of responses are equal (also see the discussion at the end of section 2.5). The equation is identical to the one used by Shinn-Cunningham (1994, p.82), with some nomenclature changed in the present discussion for clarity.

2.4.2 Internal-Noise Model

The internal-noise model of the preliminary theory separates sensory limitations from memory limitations. Specifically, the model assumes that two types of noise contribute to the uncertainty in the decision model (i.e., to the non-zero value of the variance σ). First, *sensation noise* is introduced when the physical stimuli are transformed into sensations. Then, *memory noise* is added when the sensations are subsequently mapped onto the decision axis.

Sensory noise depends only on the stimuli and the observer, while memory noise also depends on the experimental paradigm. For example, one expects performance in a discrimination task, where the observer judges two stimuli at the same time, to be limited primarily by sensation noise. Performance in a more complex task may be limited by memory noise however. For example, in an identification task where the observer selects one out of a large number of stimuli presented at random, one expects much of the uncertainty in the response will be due to the difficulty of remembering all of the different stimuli.

Formally, the internal-noise model assumes that the mean (μ) and the variance (σ^2) of the distribution for a given stimulus are given, respectively, by the following equations:

$$\mu = \alpha(I) \tag{2.3}$$

where:

$\alpha(I)$ = psychophysical scale relating the physical stimulus intensity, I, to the mean sensation, μ, on the decision axis

and

$$\sigma^2 = \beta^2 + \gamma^2 \tag{2.4}$$

where:

β^2 = variance of the sensation noise
γ^2 = variance of the memory noise.

Further, the variance of the memory noise in an identification task is determined by the effective range of intensities in the stimulus set as follows:[19]

$$\gamma^2 = G^2R^2 = G^2[\ \alpha(I_{max}) - \alpha(I_{min})\ ]^2 \qquad (2.5)$$

where:

R = effective stimulus range on the decision axis = $\alpha(I_{max}) - \alpha(I_{min})$
I_{max} = maximum stimulus intensity
I_{min} = minimum stimulus intensity
G = constant.

Further discussion of the theory is presented in connection with the analysis of the first identification experiment in sections 6.2.3.2 and 6.2.3.3.

2.5 Preliminary Adaptation Model

A major drawback of the preliminary theory of intensity resolution, with regards to adaptation, is the fact that it does not include any mechanism to predict changes in resolution and bias over time. Indeed, the authors of the original paper state, "We ignore...the transient effects at the beginning of an experimental run associated with the establishment of context" (Durlach and Braida 1969, p.374). The preliminary adaptation model accounts for these changes in two ways (Shinn-Cunningham 1994, pp. 154-159). First, the model assumes that the adaptive state of the observer affects the position of the criteria on the decision axis. Second, it assumes that the effective stimulus range changes as the observer adapts.

The state of adaptation is fully described in the model by the slope (k) of a linear approximation to the observer's responses. Figure 2.8 illustrates the situation in a graph of response position, θ, against *normal-cue* position, $f_n(\Theta)$. The observer's response is assumed to be equal to the perceived position in the experiment, while the normal-cue position

---

[19]The effective range (R), defined here, differs from the quantity, R (call it R' here), defined in the initial presentation of the theory (Braida and Durlach 1972; Durlach and Braida 1969). Specifically, R = KR' where K is a constant. The value of the constant G is also changed in a corresponding manner (i.e., G = G'/K where G' is the earlier value). Elimination of the constant K simplifies the present analysis and is consistent with the values of R and G used by Shinn-Cunningham (1994).

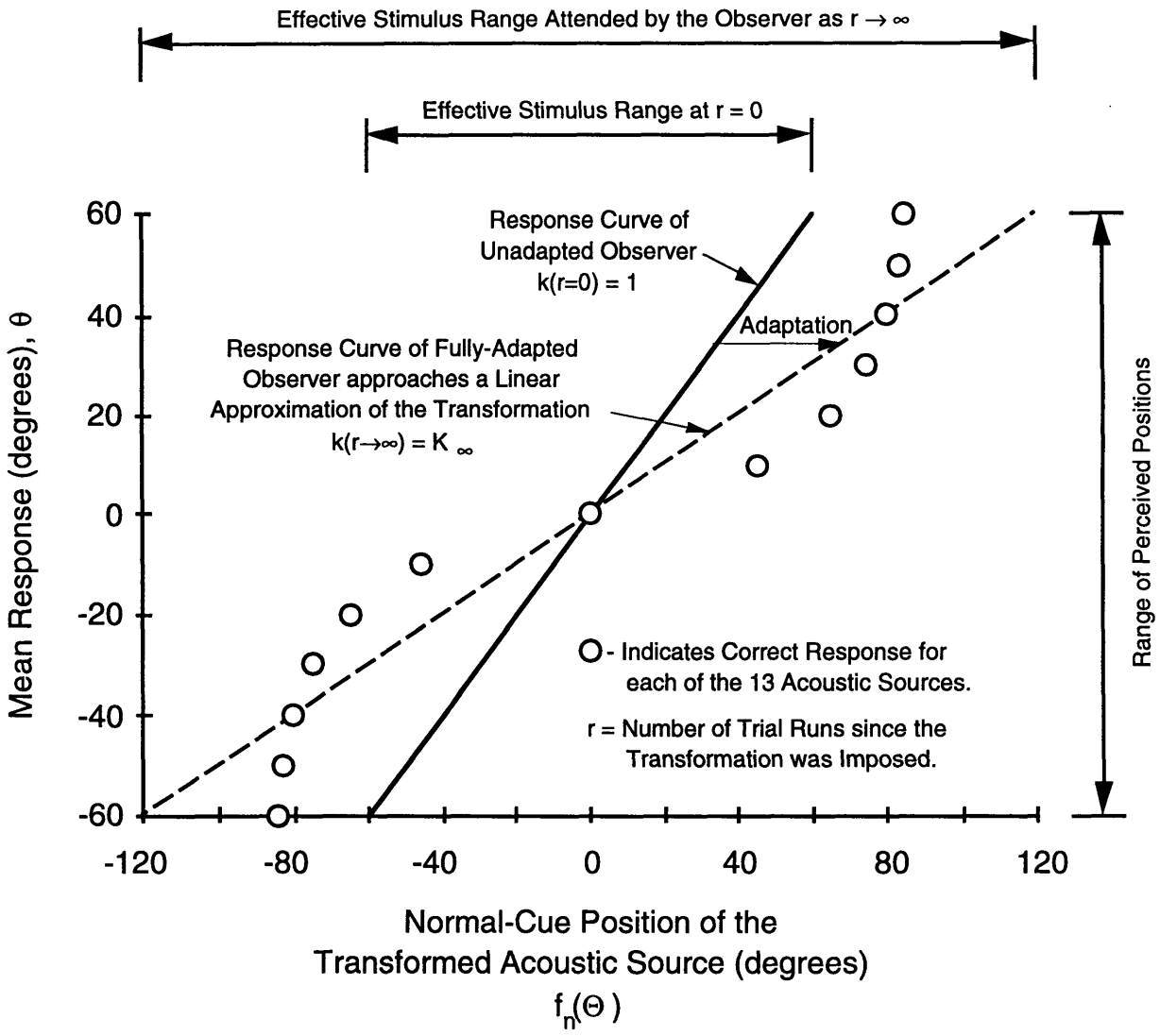


Figure 2.8. Schematic illustration of the adaptation model (Shinn-Cunningham 1994).

is a representation of the physical stimulus (ibid., p. 105). Specifically, the normal-cue position is the position that an unadapted observer would normally perceive when presented with the stimulus.

The normal-cue position differs from the "actual" source position when the acoustic cues are transformed (see section 2.3.4). The following equation defines the normal-cue position, as well as the acoustic transformation (ibid., p. 66), in Shinn-Cunningham's experiments:[20]

$$f_n(\Theta) = \frac{1}{2}\tan^{-1}\left[\frac{2n\sin(2\Theta)}{1-n^2+(1+n^2)\cos(2\Theta)}\right] \qquad (2.6)$$

where:

$f_n(\Theta)$ = normal-cue position (i.e., the transformed source position)
Θ = source position (i.e., the "actual" position that defines the correct response)
n = parameter that defines the transformation

Two lines are plotted in figure 2.8. The solid line corresponds to an unadapted observer ($k = 1$), while the dashed line corresponds to a fully-adapted observer. Initially, in the unadapted case, the observer responds by giving the maximum or minimum position (in the appropriate direction) when the stimulus is outside the effective range. Then, over time, the observer's responses adapt to a linear approximation of the transformation. The circles indicate the correct response for each of the positions.

In Shinn-Cunningham's experiments, the slope (k) is a function of the number of trial runs since the sensory transformation was first imposed on the observer, modeled by the following equation:[21]

$$k(r) = K_\infty - (K_\infty - K_0)\, e^{-br} \qquad (2.7)$$

where:

$k(r)$ = slope that describes the state of adaptation

---

[20]Shinn-Cunningham uses different nomenclature. The following substitution is made here to clarify the discussion: $\theta \rightarrow \Theta$.

[21]Shinn-Cunningham uses different nomenclature. The following substitution is made here to clarify the discussion: $T \rightarrow K_\infty$.

r = number of trial runs since the transformation was imposed[22]
K_∞ = value of k when the observer is no longer adapting[23]
K_0 = value of k in the trial prior to the transformation
b = constant related to the rate of adaptation.

Then, given that $\theta = k(r)f_n(\Theta)$ and assuming that the observer's criteria are midway between adjacent response positions, the criteria and the effective range are given by:

$$C_i(r) = \begin{cases} +\infty & \theta_i = \theta_{max} \\ \alpha\left(\frac{\frac{1}{2}(\theta_i + \theta_{i+1})}{k(r)}\right) & \theta_{min} \le \theta_i < \theta_{max} \\ -\infty & \theta_{i+1} = \theta_{min} \end{cases} \qquad (2.8)$$

and

$$R(r) = \alpha\left(\frac{\theta_{max}}{k(r)}\right) - \alpha\left(\frac{\theta_{min}}{k(r)}\right) \qquad (2.9)$$

where:

$C_i(r)$ = position of the i^{th} criterion after run r
R(r) = effective range after run r

---

[22]Note that r equals the number of trial "runs" where each run consists of 26 trials. Also *test* runs (no feedback) and training runs (feedback) alternate through the test such that the subjects only receive direct (i.e., explicit correct-answer) feedback about half of the time. A different variable, t, is used in place of r in later discussion as appropriate for the present experiments (see chapter 7).

[23]In Shinn-Cunningham's experiments the slope asymptote was found to be roughly equal to 1.05 times the slope of the least-squares linear approximation relating the mean perceived position and the correct position of the stimulus. The fact that the asymptote is greater than the best-fit line might be explained by the idea that the observer scales his or her responses to the minimum and maximum stimuli in a linear fashion. In other words, the line corresponding to the fully-adapted value of k in figure 2.8, would intersect (approximately) the correct-response points (circles) of the minimum and maximum stimuli. The idea that observers may linearly rescale all of their criteria at once, based on information about the location of only a few criteria, is proposed in section 6.2.3.5 in connection with the present experiments. Note that if the observer's state of adaptation was described by the dashed line in the figure, he or she would almost never give a maximum or minimum response (depending on the magnitude of σ). In contradiction to the explanation proposed here, however, Shinn-Cunningham states that the ratio of the asymptote to the best-fit line does not show any significant dependence on the strength of the transformation (ibid., p. 177).

$\alpha(f_n(\Theta))$ = psychophysical scale relating the physical stimulus (normal-cue position defined in equation 2.6) to the perceptual quantity, μ, on the decision axis

$k(r)$ = slope that describes the state of adaptation

r = number of trial runs since the transformation was imposed

θ_c = c^{th} response position (c = i, i+1, min, max).

Then, combining equations 2.1-2.9, one obtains the following functions for resolution (d_{ij}') and bias (B_i):[24]

$$d_{ij}'(r) = \frac{\alpha\left(f_n(\Theta_j)\right) - \alpha(f_n(\Theta_i))}{\left(\beta^2 + G^2R(r)^2\right)^{\frac{1}{2}}} \tag{2.10}$$

and

$$B_i(r) = \frac{\frac{1}{2}\left(\alpha(f_n(\Theta_i)) + \alpha(f_n(\Theta_{i+1}))\right) - \alpha\left(\frac{1}{2k(r)}(\theta_i + \theta_{i+1})\right)}{\left(\beta^2 + G^2R(r)^2\right)^{\frac{1}{2}}} \tag{2.11}$$

Note that equations 2.10 and 2.11 are written in terms of elements of both the stimulus and the response sets: Θ_i and θ_i, respectively. In the case of resolution (equation 2.10), where θ_c enters the equation via R(r), there is no difficulty. Specifically, if there is a different number of elements in the two sets, then k(r) is simply rescaled to match. In the case of bias, however, equation 2.11 is not well defined when the sets are unequal. This is not a problem for Shinn-Cunningham because her stimulus and response sets are identical. In other words, equation 2.11 is defined with the understanding that $\Theta_i = \theta_i$.

It is important to realize that this is not the same as saying that the stimulus and the response are equal! In order to keep the matter straight one makes the following distinctions in the nomenclature used here:

---

[24]For clarity, the detailed forms of k(r) and R(r), given by equation 2.7 and 2.9, respectively, are not substituted into equations 2.10 (ibid., p. 159, eqn. 40) and 2.11 (ibid., p. 164, eqn. 42). Note that θ is poorly defined in ibid., equations 40, 41, 42, and 43.

Θ = continuous (real) source position variable used for calculation.

Θ_i = i^{th} source position, in other words one of the possible virtual positions used to calculate the physical stimulus in the experiment, contained in the set $\{\Theta_{min} \leq \Theta_i \leq \Theta_{max} \mid \Theta_i \text{ an integer}\}$.

θ = continuous response position variable, used for calculation and to represent the observer's mean response, contained in the set $\{\theta_{min} \leq \theta \leq \theta_{max} \mid \theta \text{ real}\}$.

θ_i = i^{th} response position, in other words one of the allowed responses, contained in the set $\{\theta_{min} \leq \theta_i \leq \theta_{max} \mid \theta_i \text{ an integer}\}$, where $\Theta_{min} = \theta_{min}$ and $\Theta_{max} = \theta_{max}$ for equation 2.11.

Further discussion of the preliminary adaptation model is presented in connection with the analysis of the second identification experiment. In that analysis, the model is first used to explain the qualitative results of the experiment in section 6.3.3. Then, a quantitative analysis is presented in chapter 7.

3. GENERAL METHODS

The Virtual Workbench[1] (Durlach et al. 1995, pp. 39-44) was used to present various visual stimuli to human subjects and to record their responses (see Figure 3.1). The thesis experiments were performed by the author over the course of approximately one year, beginning in the summer of 1995. Following are general descriptions of the subjects, apparatus, and procedures used in the tests. In addition, the principal stimulus depth cues are identified in section 3.3. Further details of the experimental methods are presented in subsequent chapters (4, 5, and 6) when the specific experiments are described.

3.1 Subjects

Subjects were selected from the general population of undergraduate and graduate students at MIT. Two non-student friends of the experimenter also took part in the tests. Each subject was assigned a unique identification number for the duration of the experiments carried-out in connection with the thesis. Note that out of the total group of 27 subjects who were assigned a number in the overall series of experiments, only a few took part in any given test.[2]

Subject 1, the experimenter, served as a subject in all formal tests as well as numerous developmental tests. Subjects 2 through 7, who were all fellow graduate students or friends, took part in the tests without pay. Typically, the participation of the unpaid subjects involved only a single developmental test, although Subject 5 took part in three tests. She also assisted the experimenter in two cases in order to make it easier for him to self-administer the test. The remainder of the subjects (8 to 27) were paid an hourly wage. In the latter group, all but two were undergraduates.

No special criteria were used in the initial selection of subjects except for availability and the subject's willingness to take part. The paid subjects were solicited, almost exclusively, via e-mail bulletins sent to students in the Department of Mechanical Engineering and the Department of Electrical Engineering and Computer Science. The rate of pay was $7/hr for a new subject, gradually increasing as the subject became more

---

[1]The Virtual Workbench was built by Thomas E. von Wiegand and others of the RLE Sensory Communications Laboratory under the direction of Nathaniel I. Durlach. See appendix A for further details.

[2]Also note that some experiments are not reported in this thesis document due to time constraints. The experiments that are reported here are considered to be the essential body of the thesis research.

experienced such that the most experienced subjects were paid $9/hr at the end. The pay was not linked to performance; the only incentive was verbal encouragement by the experimenter to "do your best."

3.1.1 Knowledge of Purpose

None of the subjects (except for subject 1 of course) gave any indication of understanding the nature or purpose of the experiments beyond the general explanation that was given to them at the start. They were told that the purpose was "to study how a person's perception of depth varies, or adapts, over time when his or her view of the world is distorted. Specifically in some tests the apparatus simulates the way objects would look if the distance between the subject's eyes was changed" (see "IPD Experiment - Subject Consent Form" at the end of appendix B). If a subject asked a specific question about the purpose of the test, or the nature of the stimulus, he or she was told that the experimenter could not answer because it might influence the subject's response. In some such cases, for example if the subject also spontaneously remarked that the visual stimulus had changed (e.g., "It looks different now!"), he or she might also be reminded about the general purpose of the experiments.

3.1.2 Preliminary Tests

Preliminary tests were performed to measure each subject's interpupillary distance (IPD), visual acuity, and stereoacuity. The IPD, which was measured with a ruler within ±1 mm, was a parameter of the algorithm used to present stimuli in the tests. The visual and stereo acuities were measured using standard tests (a Snellen eye chart and a Randot® test, respectively) in order to insure that the experimental population fell within reasonable norms. Subjects with corrected vision wore their glasses, or contacts, in all of the preliminary tests as well as in the main experiments. The only exception to this rule was the IPD measurement, where glasses were removed. See appendix B for a detailed description of the procedures and a table of the results for all of the preliminary tests.

Although no subject was initially prevented from taking part in the tests because of poor visual or stereo acuity, one subject (10) was not asked to return for additional tests because he had the worst visual acuity observed up to that point. Two later subjects (17 and 22) scored equally badly on the Snellen test (20/40), but at the time it seemed reasonable to

reject subject 10 (who had a noticeable squint) for poor visual acuity. It turned out later that his performance in the test was actually quite good.

Four other subjects were not asked to return for additional tests or quit before completing their first test. Subject 17 (who also had 20/40 vision) was rejected because she missed the appointment for her second test. Subject 18 was rejected after the first test because he was judged to be exceedingly slow at the experimental task. Subject 19 had a problem with the muscles in one eye that made it impossible for him to look down into the apparatus. And subject 24 quit halfway through the first test because she had a headache.

Subject 24 said that she had the headache before the test, but that the test was making it worse. Later she reported that the headache lasted for two days and she decided not to return for further testing. The experimenter (subject 1) had a similar experience on two occasions. All of the subjects were routinely questioned about whether they experienced any problems during the tests, but none of the others reported a problem with headaches. A few subjects reported having "sore eyes," however, and many of them reported being tired.

3.2 Apparatus[3]

Figures 3.1 to 3.9 present various views of the Virtual Workbench that document the experimental arrangement. A schematic showing how the major components were interconnected for the tests is presented in figure 3.10. Following are descriptions of the hardware (displays, controls, and computer), the software, and the initial calibration procedures used to set-up the apparatus. Note that all constant parameter values given in the discussion are for the system *as calibrated* for the experiments. Adjustments to the Virtual Workbench would change some or all of the values.

3.2.1 Displays

The apparatus is equipped with three computer monitors (see figures 3.2 and 3.10). The *main*, or *subject's* monitor is physically mounted on top of the Virtual Workbench at a 45° angle (also see figures 3.1 and 3.3). This monitor, a Hitachi CM2198MU 21-inch color monitor driven by an ATI Mach

[3]Note that other arrangements of the Virtual Workbench that permit more complex graphics are also possible (Durlach et al. 1995, pp. 39-44), but the simpler arrangement described here is adequate for the present tests.

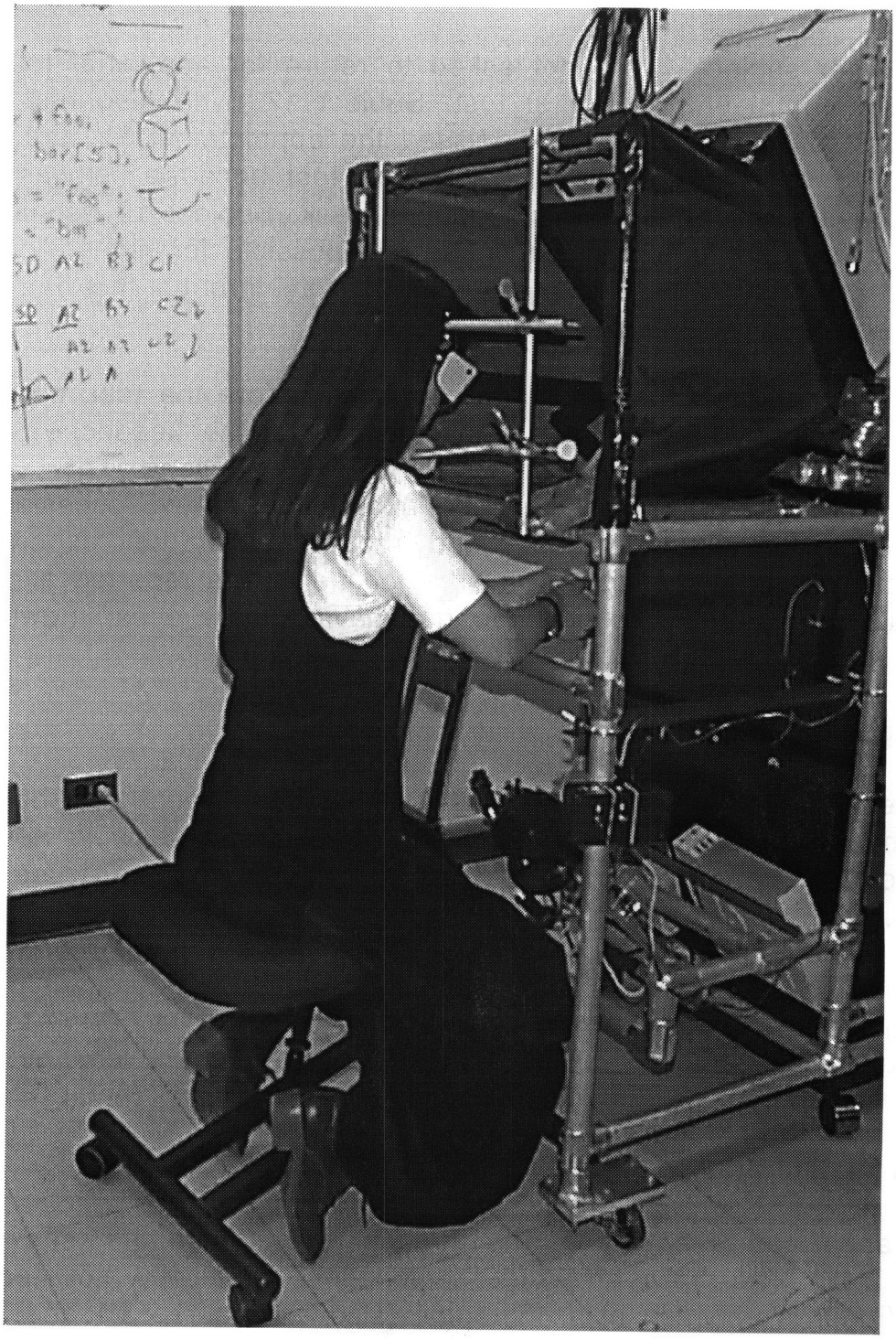


Figure 3.1. Virtual Workbench and Subject.

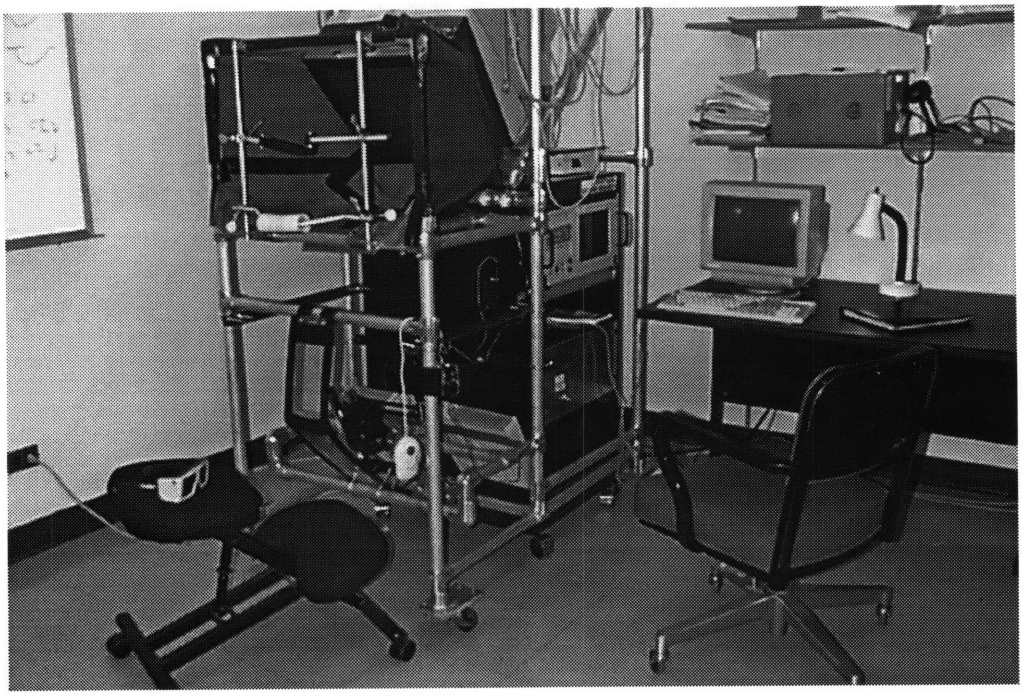


Figure 3.2. Virtual Workbench and Experimenter's Station.

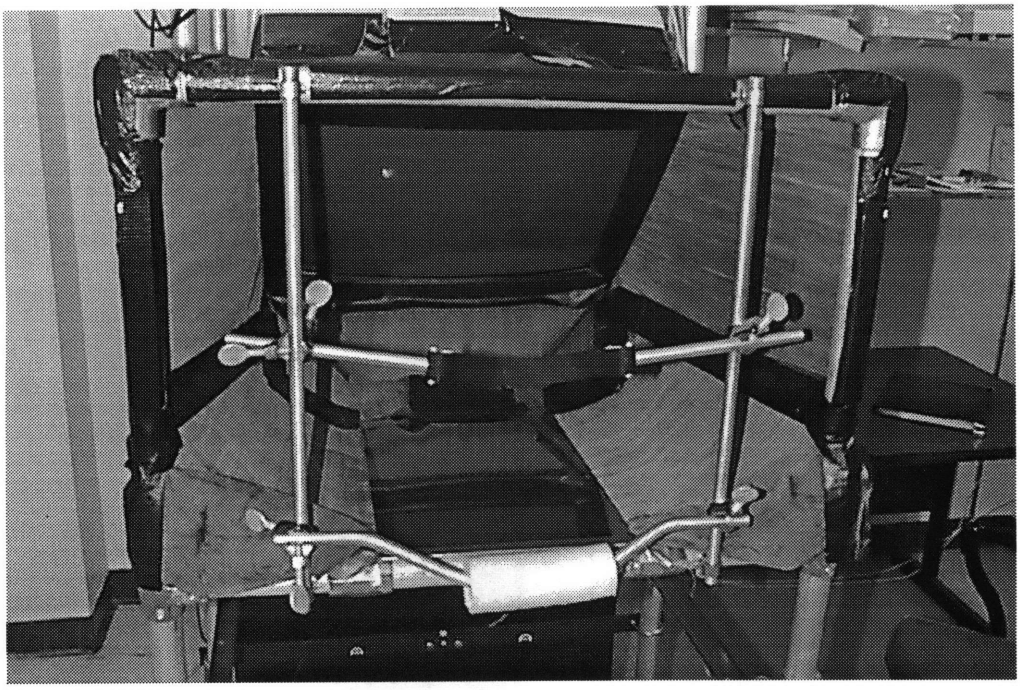


Figure 3.3. Close-up of Main Monitor, Mirror, and Headrest.

Figure 3.4. Trackball.

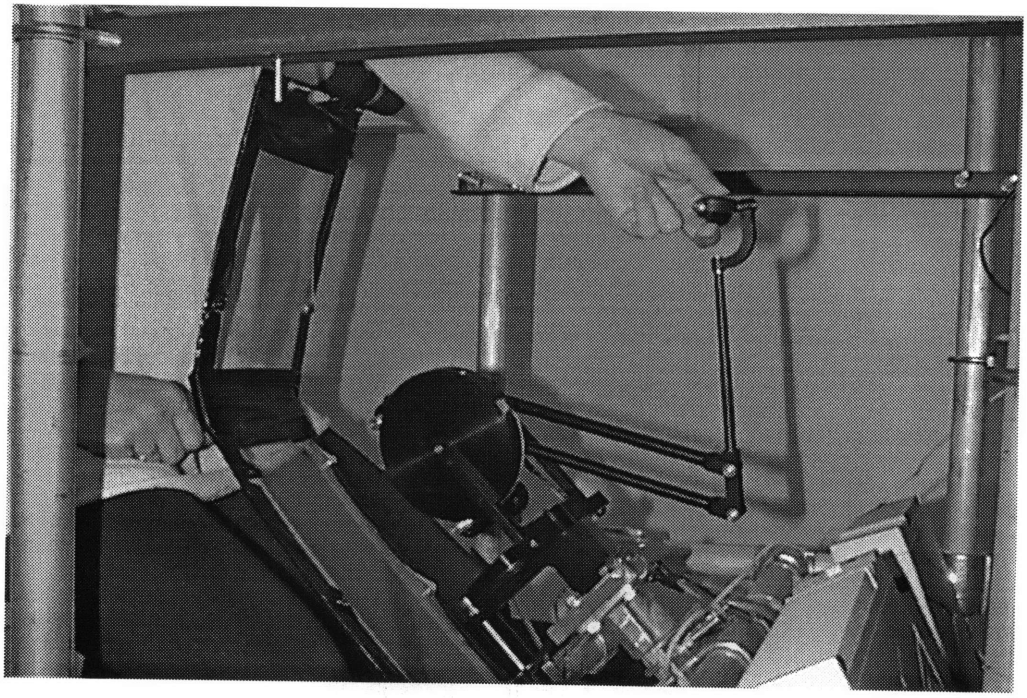


Figure 3.5. PHANToM.

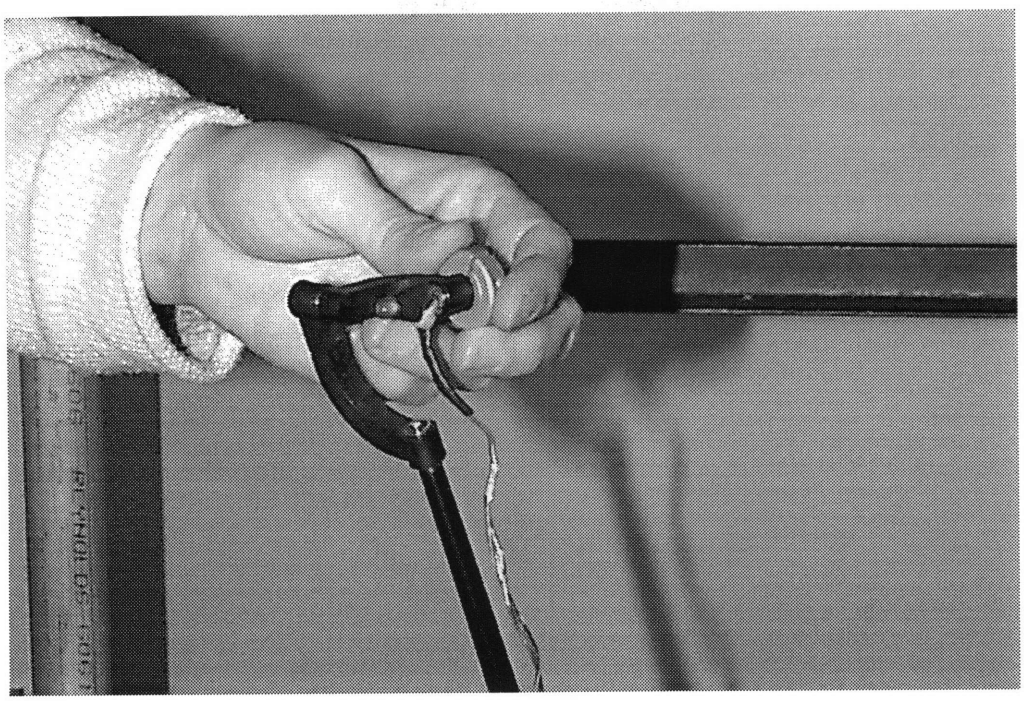


Figure 3.6. PHANToM LED Fixture.

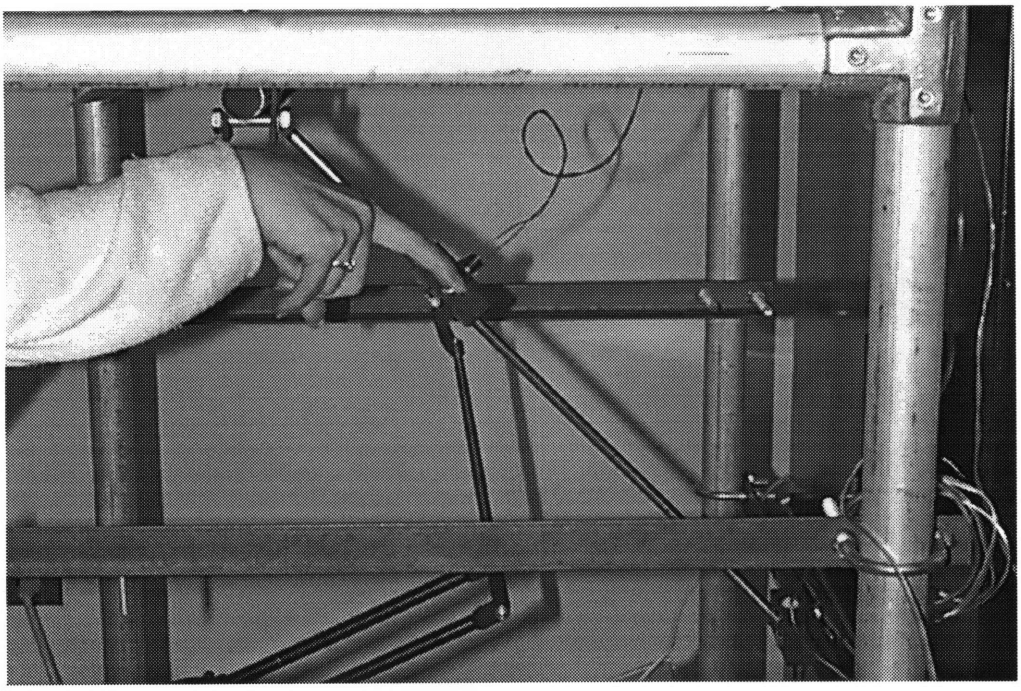


Figure 3.7. Slider.

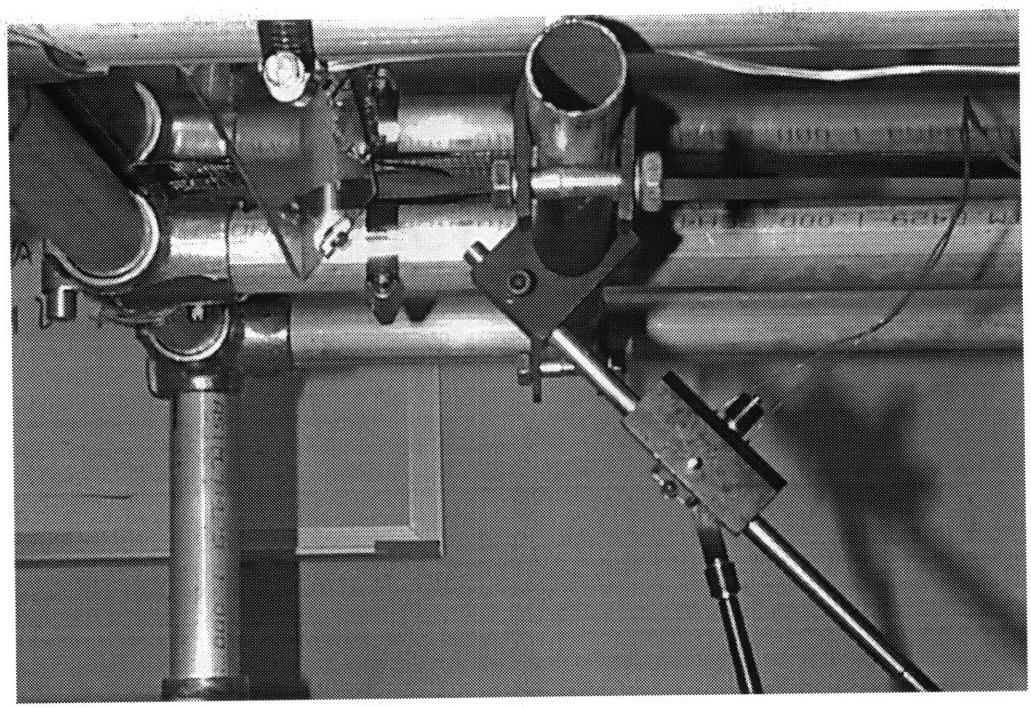


Figure 3.8. Slider Detail and Reference LED Assembly.

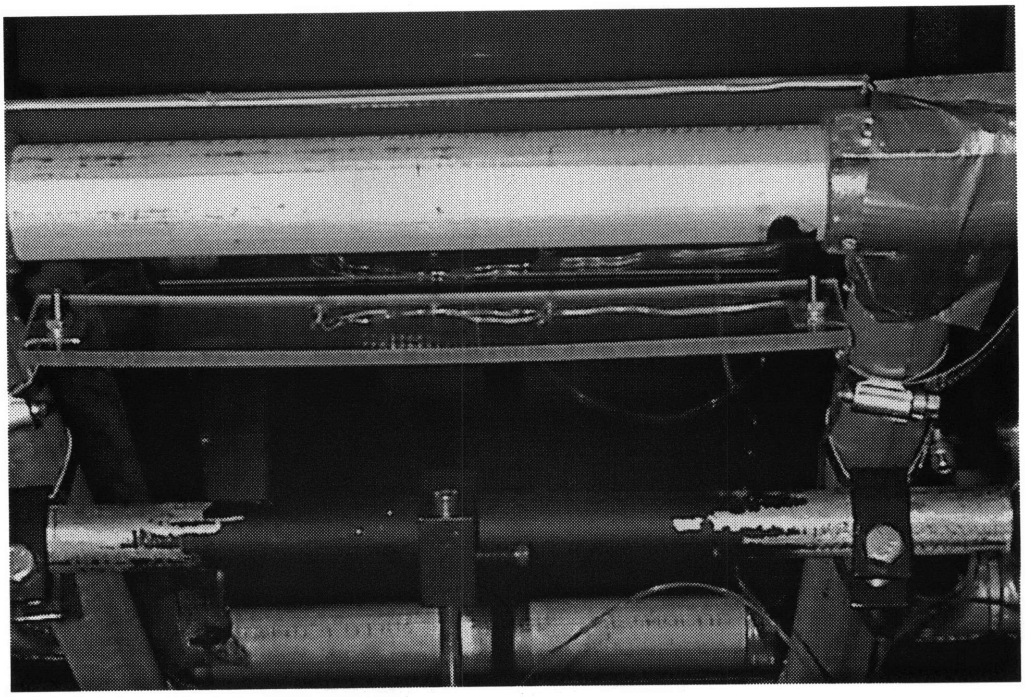


Figure 3.9. Reference LED Assembly.

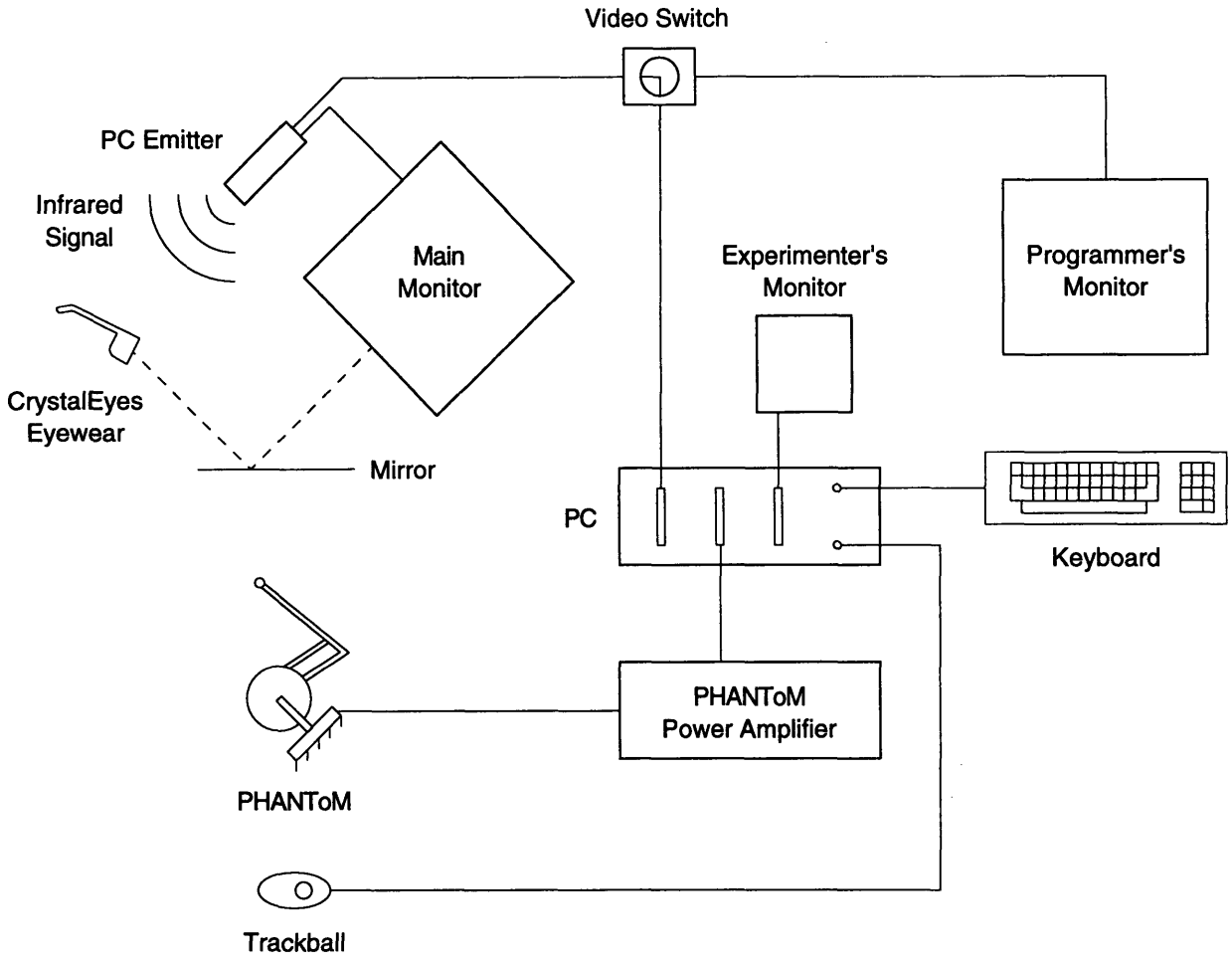


Figure 3.10. Schematic Diagram of the Apparatus.

64 Graphics Xpression video card, is used to present the visual stimuli to the subject. A secondary 9-inch black and white monitor with its own video card is physically mounted in the side of the Virtual Workbench. This *secondary* or *experimenter's* monitor is used to present text messages to the experimenter during the tests. The arrangement is such that both monitors can display information simultaneously, although re-directing the output of the computer from one monitor to the other unavoidably erases the latter monitor. A third monitor, used for programming and other operations not directly related to running tests, is located on the table beside the Virtual Workbench (see figure 3.2).

The *programmers* monitor shares the Graphics Xpression video card with the main monitor via a switch. The arrangement is such that the output from the card can be viewed on either monitor, but not on both at the same time. When the switch is set for the main monitor, the output signal from the video card passes through a StereoGraphics PC Emitter on the way to the main monitor input (see figure 3.10). The PC Emitter serves two purposes: 1) it adds an additional sync pulse each cycle causing the monitor to refresh at twice the normal rate, and 2) it transmits an infrared synchronization signal to the CrystalEyes glasses worn by the subject.

3.2.1.1 CrystalEyes: The StereoGraphics® CrystalEyes PC system, consisting of the PC Emitter and the CrystalEyes Eyewear in combination with any suitable monitor (illustrated in figure 3.11), is a field-sequential electrostereoscopic display (Bos 1993; Lipton 1993; Lipton, Akka, and Meyer 1991). Graphical images are drawn to the screen buffer of the video card so that the top half of the screen (when it is not viewed in stereo mode) presents the left-eye view of the image and the bottom half of the screen presents the right-eye view. Then in stereo mode, when the PC Emitter causes the screen to refresh at twice it's normal rate, the left and right views are presented alternately with each covering the entire screen. The lenses of the CrystalEyes glasses are liquid crystal shutters that actively turn on and off in sync with the display such that each eye sees only the appropriate view.

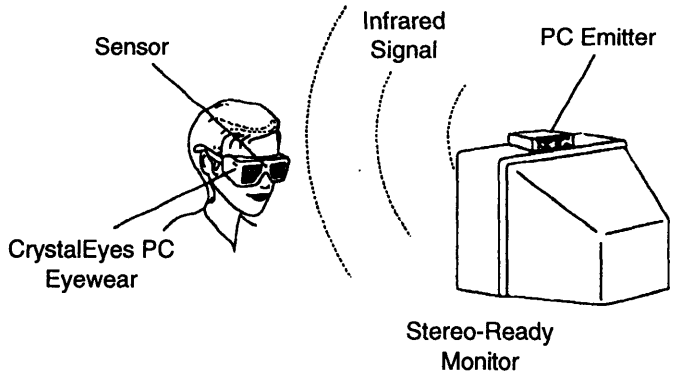


Figure 3.11. CrystalEyes PC System (StereoGraphics Corporation).

Both eye views are presented on the screen 60 times per second (i.e., the main monitor is refreshed at 120 hz) such that the subject is able to fuse the two images into a single three-dimensional percept with no apparent flicker. While the horizontal resolution is unchanged in stereo mode, the vertical resolution of each eye-view is slightly less than 1/2 of the normal display resolution. Specifically, in these experiments, a standard VGA graphics driver was used that is capable of generating pixels with 16 colors in an array 640 pixels horizontally by 480 pixels vertically. When viewed in stereo mode the resulting display has a resolution of 640 by 218 pixels (44 vertical pixels are lost because of the time it takes for the electron beam that draws the image to return to the top of the screen between the left and right-eye views, see section 3.2.4.3).

3.2.1.2 Mirror, Experimental Space, LEDs, and Lighting: The subject views the images on the main monitor in a front-surface mirror so that he or she can reach into the experimental space without blocking the screen (see figures 3.1 and 3.3).

Figure 3.12 defines the coordinate system for the space. Note that in subsequent discussion the z = 0 plane is sometimes referred to at the "xy plane" or the "plane of the display." The latter description is used because the origin of the space is located at the midpoint of the virtual image of the main monitor screen. Also note that the same coordinate system describes both the physical experimental space and the virtual space of the visual stimuli (see section 3.2.4.3).

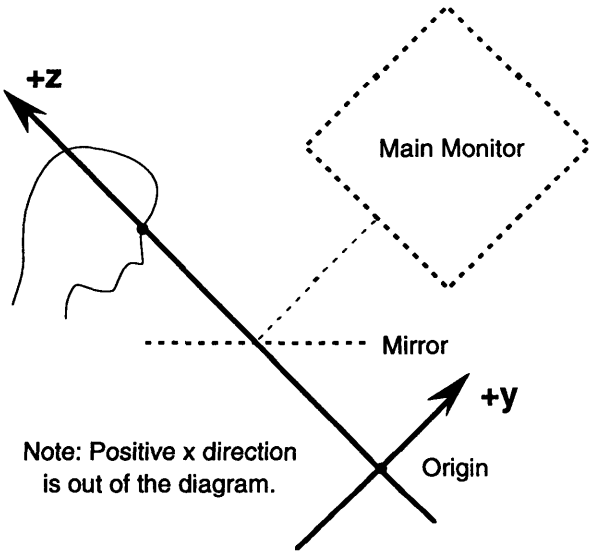


Figure 3.12. Experimental Space Coordinates.

A partially-silvered mirror is used so that the position of physical objects located in the experimental space under the mirror can be compared to the calculated position of virtual objects presented by the display. Red LEDs (Light-Emitting Diodes), that could be seen clearly though the mirror, served as the physical objects in those tests (see

figure 3.6 and section 3.2.2).[4] Note that all tests where a subject viewed the stereoscopic display were performed in relative darkness (see section 3.4.2).

The main lights in the room were turned off and flat-black sheet material was taped around the main monitor and the mirror in order to block stray light. A green piece of construction paper was also placed under the mirror to block stray light when it was not necessary to look through the mirror as part of the test.[5] Sources of stray light included: 1) a 25 W incandescent lamp located on the table next to the programmer's monitor and used by the experimenter for taking notes, 2) the subject's monitor itself, 3) the experimenter's monitor, 4) LED indicators on various pieces of equipment, and 5) cracks around the door to the room.[6]

Qualitatively the lighting was such that the subject could see the faint outline of the bezel around the main monitor when he or she looked into the mirror. It was also possible to see faint outlines of objects under the mirror when the green piece of construction paper was not in place. Both effects were hardly noticeable when the lights were first turned off, but became readily apparent as the subject adapted to the dark. None of the subjects reported that these artifacts were distracting.

3.2.1.3 Image Color and Crosstalk: Unless otherwise noted the graphical images are always drawn on the main monitor in red against a black background. The black background is designed to match the dark room and red images are used in order to minimize crosstalk on the display (see section 2.2.2.2). The effect of crosstalk in the present experiments is the appearance of a faint shadow of the displayed object on the plane of the display (z = 0 plane). Many subjects did not notice the effect until near the end of the test if at all. Using only one color also eliminates effects associated with color stereopsis.

---

[4]All LEDs used in connection with these experiments were Radio Shack, Red, Light-Emitting Diodes, size T 1-3/4 . The illuminated portion is approximately spherical, 0.5 cm diameter, and noticeably brighter viewed from the end than from the side. The rated optoelectrical characteristics of the LEDs (at 10 ma) are: Forward Voltage = 2.1 v, Luminous Intensity = 6.3 mcd. The LEDs were powered by a 6 v DC supply with a resistor in series such that the current was about 10 ma.

[5]Note that the color of the paper is largely irrelevant--any dark color would have worked just as well. The fact that the paper was green is mentioned only to help identify it in the discussion.

[6]The experiments were actually performed in two different rooms. Approximately half of the experiments (those prior to 1/29/96) were performed in a larger room with slightly more stray light and acoustic noise.

3.2.1.4 Eye Position: The left and right-eye views are drawn on the display assuming that the subject's eyes are at pre-defined locations in the experimental space. Specifically, the subject is positioned so that his or her pupils[7] are located as follows with tolerances as indicated:[8]

Table 3.1. Nominal Eye Coordinates.

Eye	X ±1cm	Y ±1cm	Z ±2cm
Left	$-x_{nom}$	0 cm	60 cm
Right	$+x_{nom}$	0 cm	60 cm

where x_{nom} is approximately 1/2 the subject's IPD, including a correction for the convergence of the eyes (see section 3.2.4.3). Note that (allowing for the small correction for convergence) the difference between the x coordinates of the left and right pupils is simply the subject's IPD; which is known from a preliminary test within ±0.1 cm (see section 3.1.2). Similarly the differences between the y and z coordinates for the two eyes are controlled within the same tolerance as the absolute positions for each eye (i.e., the absolute tolerances do not add). In other words, for example, the z coordinate of the left eye never differs by more than 2 cm from the z coordinate of the right eye.

A headrest is used to help the subject hold his or her head still at the standard position (see figures 3.1 to 3.3). The apparatus is also equipped with an adjustable chair to allow for subjects of different heights.

3.2.1.5 Reference LEDs: Two LEDs are mounted permanently to the frame of the Virtual workbench under the mirror during the experiments (see figures 3.8 and 3.9). The positions of the *reference* LEDs are carefully calibrated (see section 3.2.5) so that they may be used to control and measure the position the subject's head (see sections 3.2.4.6 and 3.4.4).

3.2.2 Controls

Three devices are used to input the subject's responses to the visual stimuli presented on the stereoscopic display: keyboard, trackball, and PHANToM. The experimenter also uses numerous power switches and other equipment controls to operate the apparatus at various times.

---

[7]Note that the algorithm used to draw the stimuli (see section 3.2.4.3) assumes that the centers of projection are coincident with the pupils as discussed in section 2.2.1.

[8]Note that the actual nominal z-coordinate used by the computer, 59.9 cm, is reported here with the appropriate number of significant digits.

3.2.2.1 Keyboard: A standard computer keyboard is used by the experimenter to input verbal responses. The keyboard is also the experimenter's primary means of controlling the apparatus during a test.

3.2.2.2 Trackball: A small hand-held trackball with two buttons[9] (see figure 3.4) is used by the subject to make various simple responses. For example, the subject might make a yes-no response by clicking the appropriate button. The trackball is also used by the subject to adjust the position of graphical objects on the display. The same trackball is also used by the experimenter for programming and other computer operations before and after the test.

3.2.2.3 PHANToM: The third input device is the PHANToM (Massie 1993; Massie 1996; Massie and Salisbury 1994), version 1.5, manufactured by Sensable Technologies, Inc. (see figure 3.5). The PHANToM, which was used only for position input in these experiments, is a small three degree-of-freedom (dof) force-feedback manipulandum that is equipped with a three dof gimbal at the end. The end gimbal is centered about the PHANToM's endpoint so that the user can hold onto the end of the manipulandum without applying any significant torque while moving through the workspace.

The end gimbal of the PHANToM is designed to attach to various fixtures. For example, figure 3.6 shows a fixture built for one of the experiments that holds an LED at the PHANToM's endpoint. Another fixture built for the experiments is called the slider.

3.2.2.4 Slider: The slider, shown in figures 3.7 and 3.8, is designed to constrain the motion of the PHANToM's endpoint to the z axis of the experimental space (see figure 3.12). Specifically, the PHANToM's endpoint is held within the following tolerances over the entire range of motion: $x = 0 \pm 5$ mm and $y = 0 \pm 5$ mm. This approach is simpler and more accurate than using the PHANToM's force capabilities for the same purpose. The slider shaft (marked with a pen) also provides a ready reference for testing the PHANToM's calibration (see section 3.2.5). The assembly is also designed to be easily removed so that it does not interfere with other experiments on the Virtual Workbench. Reference

[9]The Primax Electronics Ltd. serial trackball also has a third button that was not used in the experiments. The third (small round) button is located in the middle of the trackball and is designed to lock the left button in the pressed position.

marks make it possible to quickly re-install the assembly within the specified tolerances.

The slider consists of a 0.95 cm (3/8 inch) ground stainless steel shaft mounted on the Virtual Workbench frame and carefully aligned so that it's axis coincides with the z axis. A black Delrin block attached to the PHANToM's end gimbal slides freely on the shaft with detectable, but insignificant play. A finger-hold is provided for the subject by a 0.32 cm (1/8 inch) diameter stainless steel pin sticking 1.4 cm out of the side of the block. The axis of the pin is perpendicular to the z axis and aligned with the PHANToM endpoint.

The arrangement is such that when the subject holds the pin between thumb and forefinger, the two fingers straddle the z coordinate of the endpoint and the tips of the fingers are displaced approximately 1 cm in the positive x direction by the block.[10] The y coordinate of the finger-hold is nominally zero, but can vary depending on how the slider is held because the block rotates slightly about the shaft. The extent of rotation is limited by the connection to the PHANToM such that the y coordinate of the point where the axis of the pin intersects the side of the block is 0.0 ± 0.5 cm.

An LED is mounted underneath the slider block in the same z plane as the PHANToM endpoint and the finger-hold. The LED points toward the positive x direction, perpendicular to the subject's line of sight. The center of the illuminated portion of the LED is located as follows: x = 0.2 ± 0.5 cm (due to rotation), y = -1.0 ± 0.1 cm, z = PHANToM endpoint z ± 0.1 cm.

3.2.3 Computers

A 120 MHz Pentium PC,[11] physically incorporated in the Virtual Workbench, served all control functions for the tests including: 1) generating the graphics for the display according to a pre-determined list of conditions, 2) controlling the various input devices, and 3) recording

---

[10]The slider block could be mounted on the shaft for either right or left-hand operation, but it was always used in the right-hand configuration as shown in the photographs. In the left-hand configuration the subject's hand would have been displaced in the negative x direction from the z axis.

[11]The generic rack-mounted PC (IBM compatible Personal Computer based on the Intel Pentium microprocessor) sold by Net Computers International is actually a dual-Pentium PC. The software used to run the experiment (under MS DOS) made use of only one of the two Pentium chips however.

the subject's responses. This was possible because the visual stimuli involved only very simple (and in most cases static) graphics on the display and because the PHANToM was only used for position input. This arrangement greatly simplified the development of software and other aspects related to running the experiments.

A separate Apple Macintosh IIci computer (not connected to the Virtual Workbench) was used to generate the list of test conditions prior to the test and to analyze the data afterwards.[12]

3.2.4 Software

A program written in C++ was developed to run the experiments. Parts of the program are compiled as two separate applications: CreateTest and RunTest. The CreateTest application is compiled using Symantec C++ (version 7.0) and designed to run under Apple's Macintosh operating system (version 7.1). The RunTest application is compiled using Borland C++ (version 4.5) and designed to run under Microsoft DOS (version 6.22) on the PC. Note that the RunTest application used the "egavga.bgi" graphics driver from the standard Borland C++ graphics library and a Microsoft-compatible mouse driver from Mitsumi Electronics Corporation.

At the heart of the program are *test documents* that specify the test conditions and store the data for each particular test. Test documents are text files created on the Macintosh with the CreateTest application. The RunTest application then runs the test on the Virtual Workbench PC according to the test document and stores the data in the test document. Each test document is then saved as a record of one (it's) specific test.

Test documents are transferred between the Macintosh and the PC via 1.4 MB floppy disks using the Apple File Exchange application. The data from the tests was analyzed on the Macintosh using Matlab (version 4.1, The Math Works, Inc.). Figure 3.13 illustrates the flow of work in an experiment.

3.2.4.1 Versions, Types of Tests, and File Names: Various versions of the program were written corresponding to different groups of experiments (e.g., v2.101).[13] Each version is designed to run one or more

---

[12]Other Macintosh computers were also used at various times for these purposes, although the IIci was the primary off-line computer. Other Macs and PCs were also used for software development.

[13]All of the experiments reported here are based on the same underlying program library (version 2), hence all of the version numbers start out "v2." followed by three digits that

general types of tests where each general type corresponds to a different test document format. Each general type of test is, in turn, further differentiated into specific types of tests (e.g., T1) depending on how the controlled variables are defined in the test document.

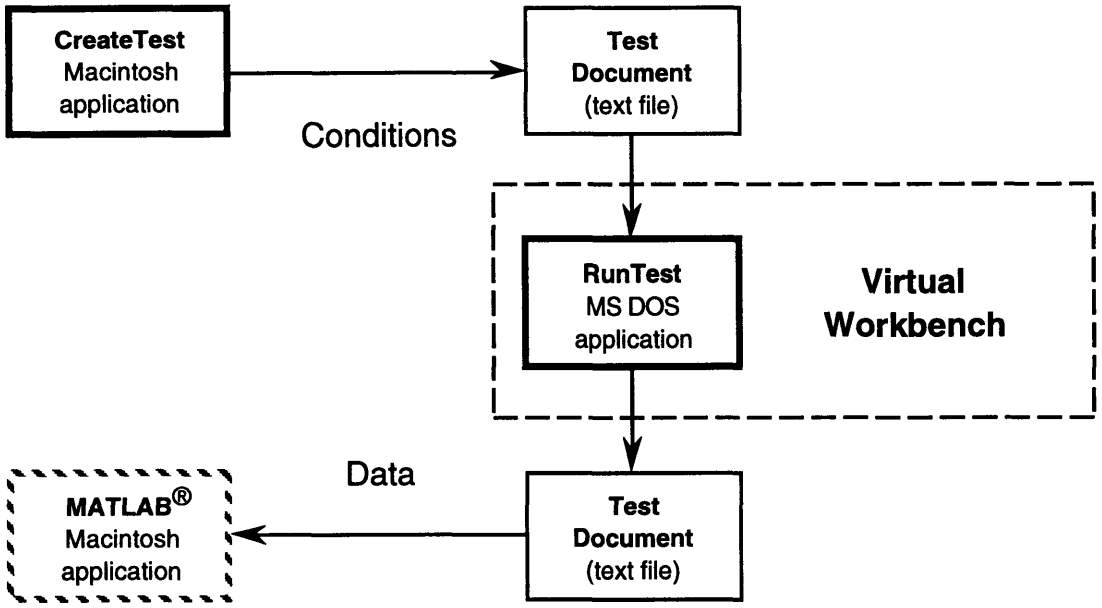


Figure 3.13. Experiment Work Flow.

A separate derived creator class[14] of the CreateTest application is used to create the test documents for each specific type of test. The application's user interface allows the operator to input the following information in order to create a specific test: file name, subject name, subject number, subject IPD, and creator type. The creator type identifies the specific type of test to be created and the application simply generates the desired test document when the operator clicks on the "Create" button.

---

identify the specific group of experiments. The first following digit identifies the general category of the experiment and the second two digits indicate progressive refinements. For example, v2.101 can be thought of as the first refinement of the 100's series of experiments. Note that while some "refinements" involved only minor changes like fixing bugs (observe that none of the tests in table 1.1 of chapter 1 has a version number that ends in 00), different version numbers are distinct experiments that may be performed in a radically different way. For example, v2.102 is very different from v2.101.

[14]For the purposes of the discussion in this paragraph, readers unfamiliar with object-oriented programming can think of the specific creator class as a subroutine. In later discussion, the reader needs only understand that a "class" in C++ involves a collection of variables and functions (i.e., subroutines) that work together in the program.

Although test documents can be created with any file name, the following code is used by convention in most cases: TxSyNz, where x is the test type number, y is the subject number, and z is a number used to distinguish cases where the same subject performs the same test more than once. Note the same test type number might be used by different versions of the experiment. Hence, test documents for different experiment versions are saved in separate directories and, in general, it is necessary to specify both the test document file name and the program version number in order to identify a particular test.[15]

3.2.4.2 Calibrated Parameters: Algorithm parameters that were determined during the calibration of the apparatus (see section 3.2.5) are compiled directly into the RunTest application as constants. Calibrated display constants are defined in the source file ENVR_VW.CPP and PHANToM constants are defined in the source file PHANTOM.CPP.

3.2.4.3 S_POINT Algorithm and the X_{ipd} Transformation: The s_point class of the program's underlying library is at the heart of all 3D visual stimuli presented to the subjects in the experiments. The algorithm transforms the coordinates of a point in virtual space into left and right views on the screen. Figure 3.14 illustrates how the left and right views of a point are projected onto the virtual image of the main monitor's screen. Note that the algorithm does not take screen curvature into account (a correction for this problem was used in some experiments however). Also note that the figure actually represents two cases that need to be distinguished (also see section 2.2.1).

1) In the first case the point is located in a virtual space that coincides identically with the physical experimental space except for the position of the subject's eyes. This space might be called the *experimenter's* virtual space because, in effect, it corresponds to a space in the experimenter's mind (for simplicity, however, it will just be called the *virtual space*). Specifically, the experimenter defines a point (x,y,z) in the virtual space along with effective positions for the subject's eyes. Then, given these inputs, the program's function s_point::UpdateDraw

[15]Unique file names that include the program version number are not possible because MS DOS limits the length of file names to 8 characters plus a three digit extension. Also note that Matlab requires that the extension be ".m" in order to read the file. The RunTest application automatically adds the ".m" extension to the test document file name when the test is completed.

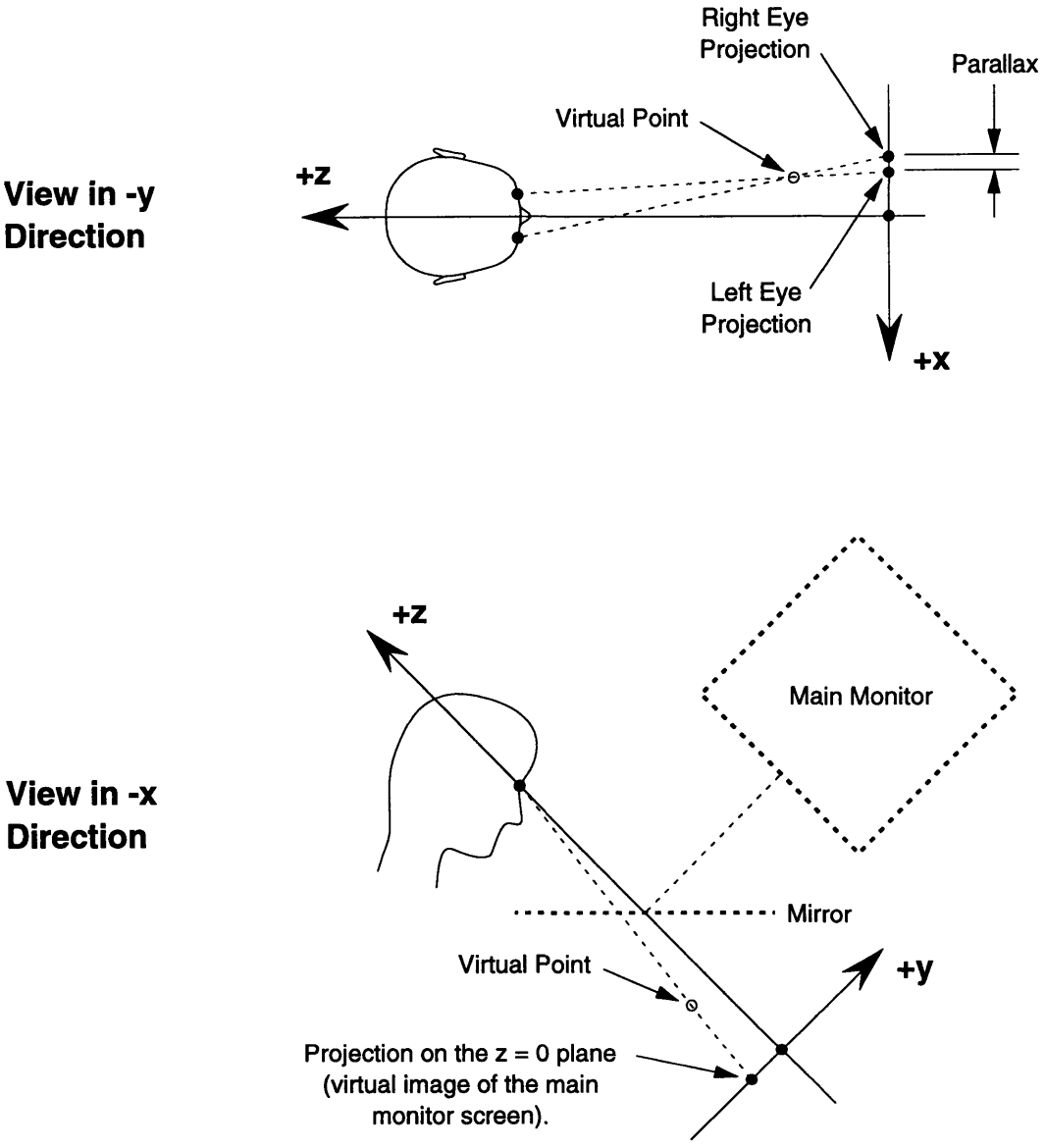


Figure 3.14. Illustration of the left and right view projections of a virtual point on the display.

calculates the projections of the point (x,y,z) onto the display screen from the effective positions of the eyes.[16] In other words, as shown in the figure, points on the screen are calculated corresponding to the left and right effective eye views of the point (x,y,z).

The effective eye positions are, of course, the transformed eye positions corresponding to the transformed locations of the subject's pupils (see section 2.2.1). Recall that the fundamental visual transformation studied here, called the X_{ipd} transformation, is simply the ratio of the effective IPD to the subject's actual IPD (see section 1.2). Specifically, the effective eye positions are defined as:

Left pupil coordinates = $(-x_{eye}, y_{eye}, z_{eye})$
Right pupil coordinates = $(+x_{eye}, y_{eye}, z_{eye})$

and the X_{ipd} transformation is given by the equations:

$$\begin{aligned} x_{eye} &= X_{ipd} * x_{nom} \\ y_{eye} &= Y_{nom} \\ z_{eye} &= Z_{nom} \end{aligned} \qquad (3.1)$$

where:

X_{ipd} = ratio of the effective IPD to the subject's actual IPD
x_{nom} = nominal distance between the z axis and the actual position of the subject's pupil in the x direction, corrected approximately for convergence as shown in figure 3.15.
= $(1 - R_{eye}/Z_{ref}) * d_{ip}/2$ (cm)
Y_{nom} = nominal y coordinate of the pupil = 0 cm
Z_{nom} = nominal z coordinate of the pupil = 59.9 cm

d_{ip} = subject's IPD measured with a ruler while subject is looking straight ahead (see appendix B)
R_{eye} = approximate distance from the pupil to the center of rotation of the eye = 1 cm (Boff and Lincoln 1988a, p 46)
Z_{ref} = reference distance determined during initial calibration of the apparatus corresponding to the distance from the plane of the virtual image of the main monitor in the mirror to the center of rotation of the subject's eye = 60.9 cm.

---

[16]The s_point::UpdateDraw function is based on a standard algorithm (Lipton, Akka, and Meyer 1991, pp. 31-35) with some minor sign changes and a major modification designed to correct for non-linear horizontal pixel spacing on the main monitor.

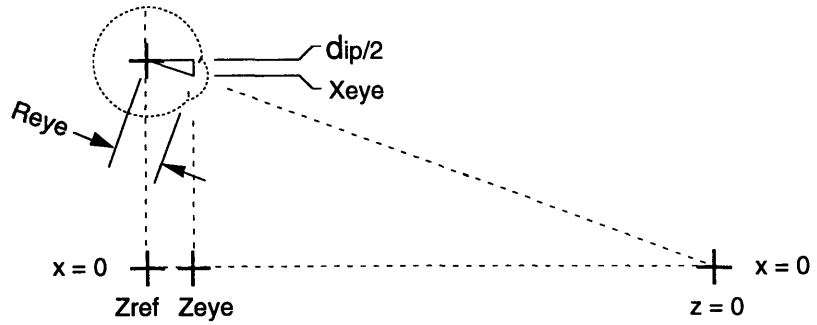


Figure 3.15. Convergence Correction Diagram.

Given a virtual point (x,y,z) and the effective (transformed) locations of the subject's eyes, the s_point::UpdateDraw function calculates the screen coordinates to be displayed as follows. First, as illustrated in figure 3.14, the screen coordinates of the left and right views (L_x,L_y and R_x,R_y respectively) are calculated in the nominal units of the experimental space (cm):

$$\begin{aligned} L_x &= (((x + x_{eye})/(z_{eye} - z)) * z_{eye}) - x_{eye} \\ L_y &= (((y - y_{eye})/(z_{eye} - z)) * z_{eye}) + y_{eye} \\ R_x &= (((x - x_{eye})/(z_{eye} - z)) * z_{eye}) + x_{eye} \\ R_y &= (((y - y_{eye})/(z_{eye} - z)) * z_{eye}) + y_{eye}. \end{aligned} \qquad (3.2)$$

Note that equations (3.2) for the left and right views are identical except for the sign of the x_{eye} terms. Also note that the y equations are simplified in practice because y_{eye} equals zero and that the z coordinate of any point on the display screen is zero by definition (see figure 3.14).

Next, the screen coordinates are normalized to the size of the display (i.e., they are transformed to dimensionless units) so that the values range from -1 to +1. Calculated screen coordinates that fall outside the boundary of the screen (i.e., when the normalized value is less than -1 or greater than +1) are simply not displayed. The normalized screen coordinates are then used to calculate the two pixels on the display corresponding to the left and right views of the point (x,y,z) as follows:[17]

[17]The (int) operator, defined as a "cast" in the C programming language (Kernighan and Ritchie 1988), transforms the floating point number on the right into an integer on the left by removing the fractional part of the number. In other words, the floating point number is rounded to the integer closer to zero.

$$L_h = (int)\ (\ (H_{stereo} - H_{mid})(L_x')^2 + H_{stereo}L_x' + H_{mid}\ )$$
$$L_v = (int)\ (\ (V_{stereo} - V_{mid})(L_y')^2 + V_{stereo}L_y' + V_{mid}\ )$$
$$R_h = (int)\ (\ (H_{stereo} - H_{mid})(R_x')^2 + H_{stereo}R_x' + H_{mid}\ ) \qquad (3.3)$$
$$R_v = (int)\ (\ (V_{stereo} - V_{mid})(R_y')^2 + V_{stereo}R_y' + V_{mid} + V_{offset}\ )$$

where:

L_x' = x coordinate of left-eye projection of (x,y,z) on the display (normalized screen units)

L_y' = y coordinate of left-eye projection of (x,y,z) on the display (normalized screen units)

R_x' = x coordinate of right-eye projection of (x,y,z) on the display (normalized screen units)

R_y' = y coordinate of right-eye projection of (x,y,z) on the display (normalized screen units)

L_h = horizontal coordinate of left-eye view pixel in the video card screen buffer (pixels)

L_v = vertical coordinate of left-eye view pixel in the video card screen buffer (pixels)

R_h = horizontal coordinate of right-eye view pixel in the video card screen buffer (pixels)

R_v = vertical coordinate of right-eye view pixel in the video card screen buffer (pixels)

$H_{stereo} = H_{max} / 2$

H_{max} = maximum horizontal screen coordinate (note that this is one less than the width of the pixel array in the screen buffer because the first pixel coordinate is zero) = 639 pixels

H_{mid} = horizontal coordinate at the center of the screen = 315 pixels

$V_{stereo} = (V_{max} - V_{blank}) / 4$

V_{max} = maximum vertical screen coordinate (note that this is one less than the height of the pixel array in the screen buffer because the first pixel coordinate is zero) = 479 pixels

V_{blank} = vertical blanking interval = 45 pixels[18]

V_{mid} = vertical coordinate at the center of the screen = 107 pixels

$V_{offset} = V_{blank} + 2V_{stereo}$.

---

[18]The vertical blanking interval is required in order to allow for the time it take the main monitor's electron gun to return to the top of the screen between the left and right-eye views. Note that actually only 44 rows of pixels in the screen buffer are nominally blank because the 45th row of the blanking interval is the first row of the right-eye view.

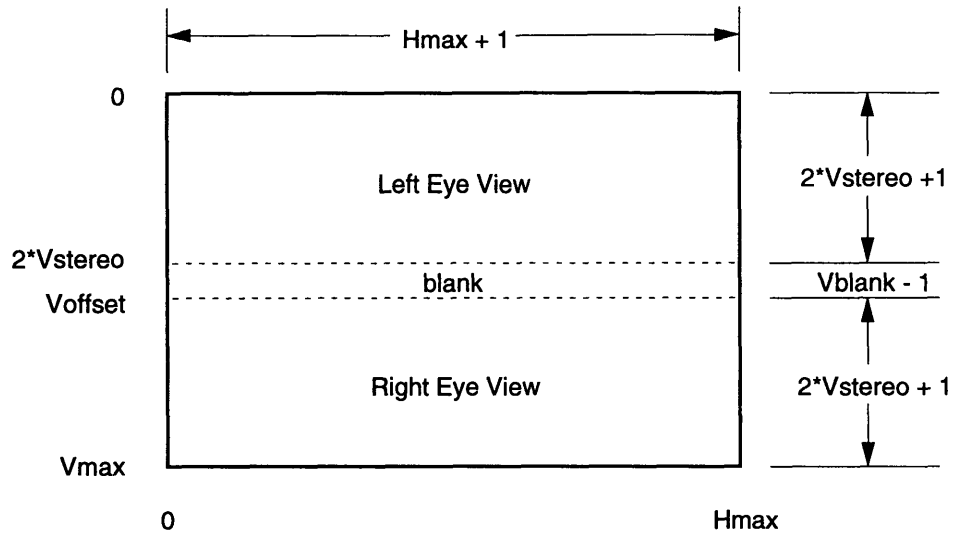
Note that the equations (3.3) are identical for the left and right views except for the V_{offset} term. Recall that the image in the screen buffer of the video card, which is identical to the image on the screen when the monitor is not in stereo mode, is divided into separate left and right-eye views (see section 3.2.1.1). Specifically, the left-eye view is presented at the top of the image and the right-eye view is presented at the bottom so that when the monitor is in stereo mode, running at twice the normal refresh rate, the two views are superimposed on the monitor's screen.

Figure 3.16 shows the layout of the screen buffer image and presents the coordinate system used to locate pixels: pixel (0,0) is in the upper left-hand corner and pixel (H_{max},V_{max}) is in the lower left. Also shown is the pixel coordinate system for the virtual image of the screen in stereo mode superimposed with the experimental space coordinate system. Note that the pixel coordinate system of the virtual image is inverted vertically relative to the screen image because the main monitor is viewed in the mirror.

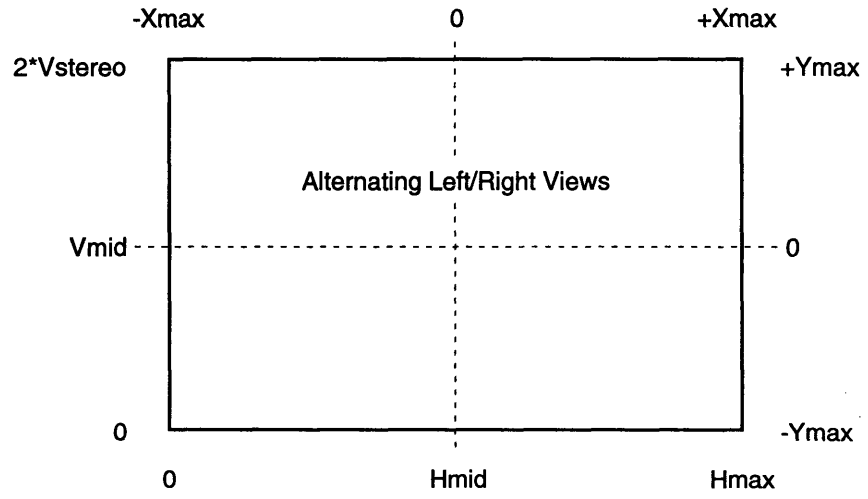
Most of the parameters used in the equations (3.1-3.3) were determined as part of the apparatus calibration discussed in section 3.2.5. The non-linear equations (3.3) were developed because it was observed during the calibration that the sizes of the pixels vary across the screen. Specifically, the width and (to a lesser extent) the height of the pixels get smaller, respectively, as a pixel's horizontal and vertical coordinates increase. Consequently H_{mid} and V_{mid} are slightly less than half of H_{max} and $2V_{stereo}$ respectively.

The value of the parameter V_{blank} is adjusted so that the left and the right-eye views are nearly aligned vertically when the display is in stereo mode. It was not possible to align the two views exactly. The right view is displaced downward relative to the left view by approximately 0.05 cm (1/2 the vertical pixel spacing in stereo mode). This is less than 1/2 the recommended maximum limit for vertical disparity (see section 2.1.3).

2) The second case to be considered is when the subject actually views the displayed point. In this case the space illustrated in figure 3.14 is identical with the physical experimental space. This might also be called the *stimulus* space because it corresponds the physical stimulus, but the name *experimental space* will continue to suffice. In this case the eye positions are the subject's actual eye positions and the screen coordinates are the actual locations of the two pixels on the display, corresponding to the left and right views of the virtual point (x,y,z).



Screen Buffer Image
(Note: identical with direct view of
the screen when not in stereo mode.)



Screen Image in Stereo Mode
(Note: the experimental space coordinate system is
superimposed on the mirror image of the screen.)

Figure 3.16. Screen coordinates of the screen buffer image and the virtual image of the screen in stereo mode.

In general the location of the point, call it p' in the experimental space, is different from the location of the point in virtual space, p. There are three causes of this shift in position. The main effect is due to the X_{ipd} transformation when $X_{ipd} \neq 1$. Another effect is due to the fact that the discrete location of the pixels on the screen in the experimental space are, in general, different from the calculated projections of the virtual point on the screen (e.g., L_x,L_y and R_x,R_y). The third effect is due to the curvature of the screen. Figure 3.17 illustrates the situation. Also see section 4.1.1.4 where a transformation from the experimental space to the virtual space is discussed.

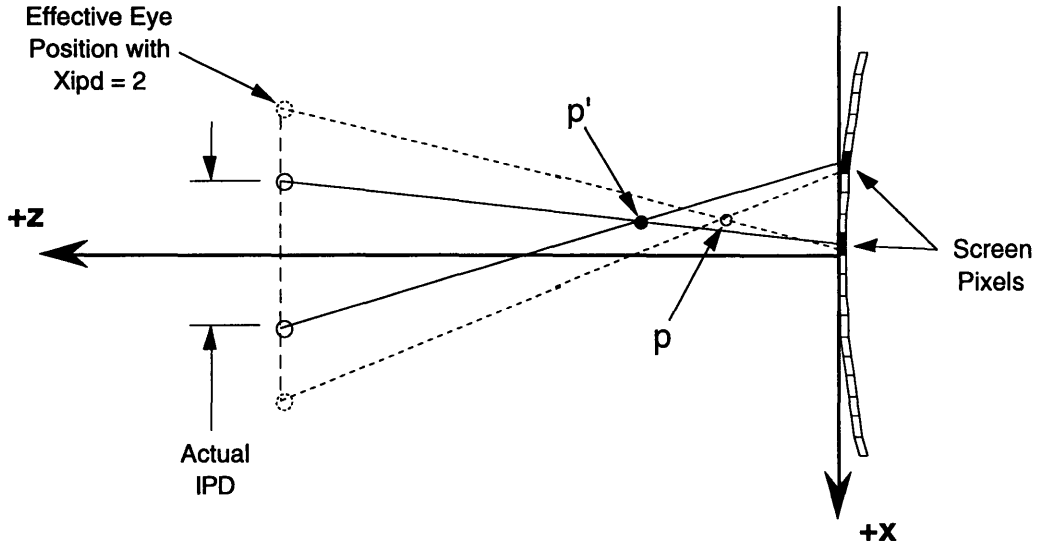


Figure 3.17. Transformation from a virtual point (p) to the physical stimulus point (p').

3.2.4.4 Jitter: Another function of the s_point class adds uncertainty to the screen coordinates of the displayed point (calculated by equation 3.3). This is done in some cases because it was observed that sudden pixel changes (particularly in the vertical direction) can give the subject a significant cue about the depth of a displayed object. When the point's "uncertainty" flag is set to TRUE, the function s_point::AddJitter randomly changes the calculated screen coordinates of the point by ± 1 pixel. Both the left and right views are changed by the same amount so that the left and right vertical coordinates agree and the parallax is not changed. The horizontal and vertical coordinates are effected independently with a 1/3 chance of being increased by 1 pixel, a 1/3 chance of being decreased by 1 pixel, and a 1/3 chance of being unchanged.

3.2.4.5 Rendering: The s_point class is used to render visual stimuli in the experiments in various ways. The general approach is described here along with one specific technique that is used to draw rectangles in many of the experiments. Other rendering techniques are discussed with the specific experimental methods as necessary.

In general, virtual objects are defined in the experiments by a set of s_points.[19] An object's position is given by the virtual space coordinates of the s_points and the left and right-eye views are given by the corresponding projected screen coordinates. For example, a pyramid may be defined by five vertices where each vertex is an s_point.[20] A wireframe rendering (Foley et al. 1990) of the pyramid is then accomplished by drawing lines between the projections of the vertices on the screen. Specifically, for each view of the pyramid, lines are drawn between each of the 4 corners of the base (around the perimeter) and then from each corner to the apex. Note that there are no *hidden lines* with this simple rendering technique. Also, the order in which the lines are drawn does not matter provided that they are all the same color.

Rectangular Stimuli: A slightly modified technique is used to draw two-dimensional rectangles (in space) when the virtual rectangle is always parallel to the xy plane. The position of the stimulus object in virtual space (x,y,z) is defined to be the center of the rectangle. The object itself is defined by two s_points, (x-w/2, y-h/2, z) and (x+w/2, y+h/2, z), corresponding, respectively, to the lower-left and upper-right corners of the rectangle. Note that w is the width of the rectangle in virtual space and h is the height.

The screen coordinates of the s_points, in turn, define two rectangular regions on the main monitor's screen corresponding to the left and right-eye perspective views of the object. An empty rectangle may be rendered by drawing lines between the corner points on the screen, around the perimeter of each region. Solid rectangles are drawn by filling-in the regions with a solid color. Note that the screen parallax of the two views is determined solely by the lower-left corner; with the width and height

---

[19]The word "s_point" is used here to refer to an *instance* (Eckel 1995; Winston 1994) of the s_point class. Each instance of the class (i.e., each *s_point*) corresponds to a unique set of coordinates in virtual space (x,y,z) as well as a unique set of screen coordinates (Lh,Lv,Rh,Rv). The screen coordinates are the pixel locations of the left and right-eye views of the virtual point in the video cards screen buffer as calculated using the s_point algorithm.

[20]Note that, while some preliminary experiments did involve three-dimensional shapes like pyramids, the ultimate experiments described in the thesis involve only two-dimensional shapes like the rectangles discussed in the next paragraph.

of the renderings based on the left view. Specifically, the two views of the virtual object are determined as follows.

The left views of the virtual points (x-w/2, y-h/2, z) and (x+w/2, y+h/2, z) are calculated, using the s_point algorithm, to determine the screen coordinates (Lh_{min}, Lv_{min}) and (Lh_{max}, Lv_{max}), respectively. Thus the left view of the object is the rectangle on the screen with corners at the points, moving clockwise: (Lh_{min}, Lv_{min}), (Lh_{min}, Lv_{max}), (Lh_{max}, Lv_{max}), and (Lh_{max}, Lv_{min}). The right-eye view of the object is determined somewhat differently. First, the lower-left corner is calculated in the same way as the left view in order to determine the screen parallax. In other words, the screen coordinates of the lower-left corner of the right view (Rh_{min}, Rv_{min}) are calculated from the virtual point (x-w/2, y-h/2, z) using the s_point algorithm. Then the screen coordinates of the upper-right corner are calculated by adding the width (w_{pix}) and the height (h_{pix}) of the left-view rectangle, measured in pixels. Specifically, the upper-right corner of the right view is ($Rh_{min} + w_{pix}$, $Rv_{min} + h_{pix}$), where $w_{pix} = Lh_{max} - Lh_{min}$ and $h_{pix} = Lv_{max} - Lv_{min}$.

This is done so that a small difference in the parallax of the upper-right and lower-left corners does not produce a one pixel difference in the parallax on the screen. One pixel differences of this type are easily detected by the subject, making the rectangle appear tipped in the z direction.

Jitter may also be added to the screen coordinates of the two s_points, as described in the previous section, in order to add uncertainty to the rectangular stimulus. When this is done the boundaries of the projection on the screen shift randomly by ±1 pixel. This can cause a ±1 pixel shift in the center of the rectangle and a change of ±2 pixels in the width and height. The parallax does not change however.

3.2.4.6 EYE_CAL Algorithm: The eye_cal class of the program's underlying library is used to control the position of the subject's eyes during a test. The algorithm requires the subject to line-up the reference LEDs (see section 3.2.1.5) with marks on the display. Figure 3.18 illustrates the geometry. The algorithm works in two ways.

In the first case a display calibrated for the subject's measured IPD is presented by the function eye_cal::HeadPosSetUp. The subject then adjusts the position of his or her head until marks on the display line-up with the

reference LEDs. The goal is to position the subject's eyes at the standard eye positions (see section 3.2.1.4).[21]

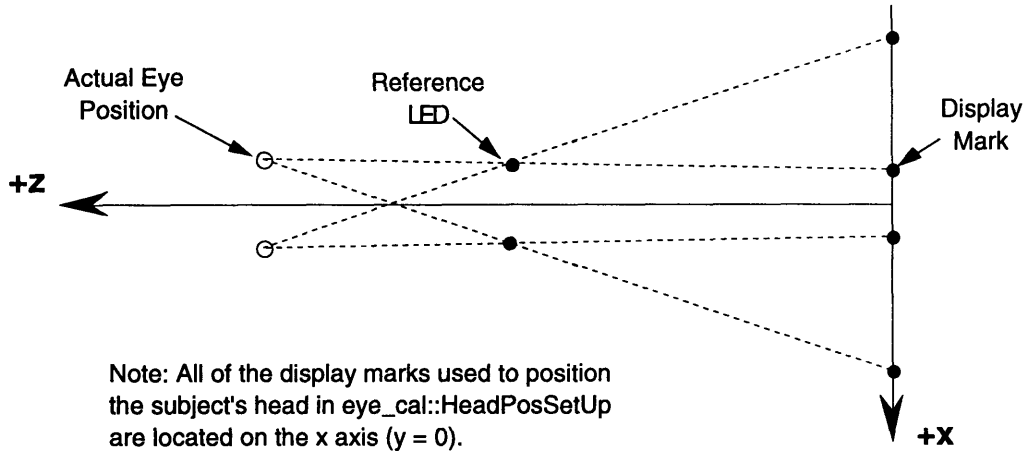


Figure 3.18. Illustration of EYE_CAL Geometry.

In the second case the subject adjusts the positions of marks on the display, using the trackball, until they line-up with the LEDs. The latter procedure, controlled by the function eye_cal::EyeCalib, is used to measure the position of the subject's eyes. This is done in order to verify that the subject is at the nominal position within the desired tolerance. Note that the algorithm is based on the center of rotation of the eyes rather than the location of the pupils. In other words the calculated eye positions should correspond to ($\pm$ $d_{ip}/2$, 0, Z_{ref}) rather than the nominal pupil positions ($\pm$ x_{nom}, Y_{nom}, Z_{nom}) as defined in equation 3.1 (see section 3.2.4.3).

The result of the EyeCalib measurement is presented on the experimenter's display so that the experimenter can monitor the subject's eye positions periodically during the test. A warning is presented when the eye positions are found to be out-of-tolerance. See section C.1.4.3 in appendix C for an analysis of the accuracy of the eye position measurements.

---

[21]A bug in the function eye_cal::CalcXmXe_x was discovered while writing appendix C after all of the experiments were finished. This function calculates the x coordinates of the marks on the display used to position the subject's head. Specifically, the pupil location (corrected for convergence) is used rather than the center of eye rotation. In theory this makes it impossible for the subject to line-up the marks and the LEDs exactly. In practice: 1) the discrepancy between the LEDs and the marks is less than 1/2 mm (i.e., 1/10th of the LED diameter) so that the subject probably can't tell and just averages the difference; 2) the display calibration is based on the subsequent eye position measurements--if those are OK then the initial set-up is irrelevant.

3.2.5 Calibration

The display was adjusted and parameters to be used in the experiments were measured so that a point in virtual space would be accurately displayed in the physical experimental space. The display was subsequently tested in the "calibration experiments" described in chapter 4. Appendix C presents a detailed discussion of the initial calibration and the related tests.

The PHANToM, which was used to measure position in the thesis experiments, was calibrated by Lukasz Weber. Weber measured the encoder values for a large number of endpoint positions and calculated the transformation to Cartesian coordinates. The PHANToM calibration was subsequently tested at known points on the z axis of the experimental space. These tests were performed in conjunction with all tests where the slider was used in order to verify Weber's calibration. The tests also served to insure that the PHANToM was properly calibrated in each specific test. In general, it was observed that the absolute accuracy of the PHANToM position (with respect to the experimental space) was better than $x \pm 4$ mm, $y \pm 4$ mm, $z \pm 2$ mm on the z axis (in the range $-10\text{cm} \leq z \leq 10$ cm) with repeatability better than $\pm$ 1/2 mm (see appendix C).

Following the initial calibration the only further calibration of the apparatus was to reset the PHANToM encoders periodically. This was done, as required prior to the start of each test, with the PHANToM at pre-defined reference positions in the experimental space. Calibration of the display also requires that the subject's eyes be at the proper positions as discussed in sections 3.2.1.4 and 3.4.4.

3.3 Stimulus Depth Cues

This section attempts to identify the principal depth cues that are available to the subjects in the experiments (see sections 2.1, 2.2, and 3.2 for background). The main cue is binocular disparity. Indeed, performance is shown to be significantly worse when this cue is removed in one of the calibration experiments (see section 4.2). Other cues are available, however, such that the subjects' depth judgments involve a complicated interaction among different cues. The cues are also different in the case of physical and virtual objects as discussed in sections 3.3.1 and 3.3.2, respectively.

For example, oculomotor cues associated with accommodation and convergence are present; and different for physical and virtual objects.

Also, almost the full spectrum of monocular cues is available to judge the depth of physical objects, while only a small subset of monocular cues is available in the case of virtual objects. Crosstalk also provides an additional cue for virtual objects as discussed in section 3.2.1.3.

The fact that the subjects do not move their heads during the test (see section 3.2.1.4) means that most of the monocular cues in the experiment are static. One exception, where a dynamic monocular cue is provided, is in the calibration experiments (see chapter 4). Specifically, the subjects move a probe in the z direction by hand. In this case, one expects that kinesthetic sensations of hand position and velocity will give the subjects more information about the experimental space. The subjects can also observe the rate of change of the visual size of the probe and the rate of perspective change.

3.3.1 Physical Objects

There are two key physical objects viewed by the subjects in the experiments: 1) the main monitor, and 2) the LED probe used in the calibration experiments (see sections 3.2, 4.1, and 4.2). In both cases one expects that the subjects will have a good idea of the true size and distance of the objects. In particular it seems likely that the image, of the monitor screen and bezel in the mirror, provides a good absolute distance reference in all of the experiments. This is true because the subjects see the monitor in normal room light before the test. At that time, of course, the monitor is right in front of them and they have all of the normal spatial cues available to judge its size and distance (including motion parallax). Further, once the test starts, the monitor is still partially visible in stray light (see section 3.2.1.2) and it doesn't move.

Thus, even if the reduced level of light limits some depth cues, one expects veridical perceptions of the physical objects. Further, all of the depth cues are consistent with one another (i.e., they are related in the usual way). This is not true in the case of the virtual objects in the experiments.

3.3.2 Virtual Objects

As discussed previously in section 2.2.2.1, there is a conflict between accommodation and convergence in the case of virtual objects. Recall that the normal relationship exists when the subject fixates on a virtual point at $z = 0$ (i.e., in the plane of the display), while the magnitude of the conflict increases as the point moves further in front or behind the screen

(i.e., as the absolute value of the parallax grows). Note that the fact that the oculomotor cues agree for both physical and virtual objects in the plane of the display, but not elsewhere, further supports the suggestion that the monitor provides a distance reference in the experiments. Specifically, it provides a reference at $z = 0$.

The relationships between the other depth cues are nominally consistent with normal vision. Specifically, disparity, convergence, and the monocular cues should all agree when $X_{ipd} = 1$, allowing for display pixelation. In addition to the unavoidable *noise* introduced by pixelation, random noise is added to the monocular cues when the objects are rendered with Jitter (see section 3.2.4.4), making judgments based on these cues somewhat less certain. There are two principal monocular cues for virtual objects in the experiments: perspective change and relative size.

Perspective change refers to a relative distance cue associated with a changing point of view (Boff and Lincoln 1988b, p. 1054). While the subjects heads do not move significantly during the test, the virtual objects on the display do move relative to the subject. This is true to some extent even in the case of a static display, because the subject observes the test objects at different distances over the course of multiple trials.

Relative size is a cue based on the fact that the visual angle subtended by an object on the retina gets smaller as the object moves away from the observer. Relative size provides a cue to absolute distance when the observer knows the true size of the object (familiar size). Alternatively, it provides a relative distance cue when two objects of the same size are compared. Similarly, one might compare the visual angle of the same object at different times to estimate the change in distance, assuming that the true size of the object does not change.

Other monocular cues are eliminated for virtual objects in the experiment by design. For example, the virtual objects in most of the experiments consist of static rectangles parallel to the observer's line of sight. In all cases, the virtual rectangles are positioned so that they do not overlap each other or the border of the display, thus eliminating *interposition*, a powerful relative distance cue. Another purpose of this was to avoid any problems associated with the superposition effect that is discussed in section 2.2.2.3. Virtual objects are also rendered in only one color and with uniform intensity in order eliminate a range of minor

static cues, including color stereopsis.[22] The objects also lack fine detail that might be used as a cue.

3.4 Procedures

Subjects performed various experimental tasks depending on the specific test. Most of the data for the test was automatically gathered by the Virtual Workbench. The experimenter also recorded some observations in a notebook. The following general procedures were used in all of the experiments.

3.4.1 Subject Screening and Initial Training

All first-time subjects were interviewed prior to taking part in the experiments. In most cases, the preliminary tests were performed at the same time. In all cases, the first step was to measure the subject's IPD with a ruler so that the measured value could be used to create the test document for the first test. Typically, the subject read and filled-out various forms while the test document was created. In particular, all subjects signed a consent form in accordance with the rules of MIT's Committee on the Use of Humans as Experimental Subjects. A copy of the consent form is presented at the end of appendix B.

First-time subjects were then trained in the procedures used to control eye position (see section 3.2.4.6). Experienced subjects usually did not require additional training in the eye_cal procedures. The procedures are simple and they are repeated many times in each test. In any event, the first step in every test is to position the subject's head. So subject's were always forced to review the procedures before beginning the test.

Training in the specific experimental tasks varied from test to test. In most cases the task was simply described by the experimenter using hand gestures to represent the stimuli. Then the subject was allowed to practice with a demo version of the test or, in the case of very simple tasks, the subject just began the test while the experimenter watched over his or her shoulder. In all cases the subject was asked if he or she had any questions about the procedure and the experimenter watched to

---

[22]One exception is the calibration experiment described in section 4.2. In this case a *green* reference frame is presented on the display in some of the tests. Color stereopsis does not appear to be a factor, however, because the result with and without the reference frame are similar. Also, the actual depth judgment is between a red virtual rectangle and a red LED.

make sure, as far as possible (see section 3.4.3), that the procedures were in fact being followed.

3.4.2 Lighting

The headrest was adjusted so that the subject's head was in the proper position prior to the start of each test. This was done with the main room lights on in most (if not all) cases. Then the main lights were turned off for the duration of the test. Note that there was sufficient light in the room (without turning the main lights back on) to re-adjust the headrest if needed (see section 3.2.1.2). No special effort was made to allow the subject to adapt to the darkness, although it took at least 1 or 2 minutes to actually start the test once the lights were off. Subjects were also allowed to leave the room during breaks (to use the bathroom, get a drink of water, etc.) so that they were occasionally exposed to normal room light during a test. Then, like at the beginning, it took at least 1 or 2 minutes to restart the test after the subject returned to the dark room.

3.4.3 Feedback and Speed

In general, with only two exceptions, the experimenter did not have any means of judging the subject's performance until the test was complete. This limited possible influence on the subject, making it simple and honest for the experimenter to respond, "I don't know," if a subject asked how he or she was doing at the given task. One exception to the general rule is, of course, when feedback was intentionally given to the subject. All feedback was controlled by the Virtual Workbench, however, and in most cases without the experimenter having any specific knowledge about the subject's performance. The other exception is related to how quickly the subject performed the experimental task.[23]

The experimenter always told the subject how long the test was expected to take and usually indicated the pace required for each trial. Subjects who seemed to be taking too long were asked to try to speed-up, usually with some further explanation about the expected accuracy of the responses. For example, the experimenter might say, "the experiment is designed so that you will make some mistakes. That's how I measure your performance. So don't worry about doing it perfectly. Just look at the target, make your judgment, and click the button." Then (as a means of indicating the appropriate pace) the experimenter would add, "look and

---

[23]Note that the experimenter also knew the results of the eye position measurements, but that is really external to the experimental task.

click, look and click, look and click." Ultimately, however, the subject was always allowed to take as long as he or she needed to finish the test.

3.4.4 Eye Position Measurements

The subject's eye positions were measured before and after each group of trials using the eye_cal procedures (see section 3.2.4.6). When the eyes were found to be out of position, the subject was presented with the eye_cal::HeadPosSetUp display and asked to re-position his or her head. One exception is the case where it was the eye position measurement following the final group of trials. In the latter case the test simply ended. Otherwise, after the subject was re-positioned, the position was re-measured and the whole procedure was repeated, as necessary, until the eye positions were within the desired tolerance (x±1cm, y±1cm, z±2cm).[24]

3.4.5 Test Schedule and Breaks

Unless otherwise noted, subjects only performed one test per day. This was done in an effort to insure a uniform state of adaptation at the start of each test and so that the subjects would not become too tired. The tests were scheduled simply according to the availability of the subject and the experimenter. Subsequent tests might be less than one day apart or more than two weeks apart.

The tests were designed to last at most about two hours because this seemed like a reasonable endurance limit. The first test was usually designed to be somewhat shorter than the others in order to allow for the additional time required by first-time subjects (for preliminary tests, training, creating and transferring the test document to the Virtual Workbench, and so forth). After the first test, however, it was possible to have the apparatus set-up and waiting when the subject arrived.

The duration of the tests varied significantly from subject to subject (and sometimes from day to day for the same subject when a test was repeated) because subjects were allowed to proceed at their own pace. For example, a test designed to take about two hours took more than 2-3/4 hours for one subject. He repeated the test on another day in under 2 hours while another subject did the same test in less than 1-1/2 hours. Some of

---

[24]Note that the error allowed in the measured position (i.e., before the out-of-tolerance warning is displayed to the experimenter) is only half of the total tolerance (see sections 3.2.4.6 and C.1.4.3).

the variability resulted because subjects sometimes had a hard time getting comfortable in the headrest. This led to them spending more time re-adjusting head position. Some subjects also required more rest breaks than others.

Tests were divided into groups of trials primarily in order to give the subjects a chance to rest during the test. Rest breaks were allowed (at the subject's discretion and for as long as needed, typically about 5 minutes) at the end of any group of trials and after the position of the eyes was measured. In this way the eye positions were known before and after every group because the subject's head could be re-positioned after the break, before starting the next group. Of course, the subjects were free to stop at any time and some did (i.e., they stopped in the middle of a group), although this only happened in a few cases.

Whenever a subject moved his or her head prior to the eye position measurement at the end of the group (assuming that the subject also wanted to continue with the test), the subject was asked to reposition his or her head so that it *felt* about right. Then the subject continued with the test and the eye positions were measured at the end of the group as usual. The experimenter always made a note when this occurred.

During the breaks the subjects were allowed to do pretty much whatever they wanted to do (see section 3.4.2). The only exception to this general rule was during the first adaptation experiment (see section 6.2), where the subjects were asked to keep their eyes closed during the break.[25] This was done in order to avoid any affect on the state of adaptation, associated with normal viewing. The experimenter decided that the procedure was unnecessary, however, and it was discontinued prior to the second adaptation experiment (see section 6.3).

[25]The subjects were also asked to keep their eyes closed during the breaks in some other experiments that are not reported here.

4. CALIBRATION EXPERIMENTS

Two calibration experiments and one follow-up experiment were performed with the Virtual Workbench to verify that subjects perceived virtual objects presented on the display at the same locations as real objects. In the first two experiments, the subjects manually adjusted the position of a physical probe (an LED attached to the PHANToM) so that it appeared to be at the same location as a virtual target on the display. The first calibration experiment,[1] called the 3D experiment, spanned a rectangular volume of the virtual space and relied entirely on the PHANToM to measure the position of the LED. Consequently, some discrepancies in the 3D experiment between the predicted location of the displayed object and the measured position of the LED may be due in part to errors in the PHANToM calibration. The second experiment, which was restricted to the z axis of the space, was performed in conjunction with independent measurements of the PHANToM position in order to overcome this difficulty. The third experiment was identical to the second experiment except that the relative positions of two virtual objects were judged by the subject, rather than comparing a virtual object to the LED. This was done in order to investigate an artifact of the second experiment related to accommodation.

4.1 3D Experiment

In the 3D calibration experiment, virtual targets were presented randomly at 22 standard positions in virtual space, spanning the region $-12 \leq x \leq 12$, $-4 \leq y \leq 4$, $0 \leq z \leq 8$ (cm). Subjects manually adjusted a physical probe so that it appeared to coincide with each virtual target position in repeated trials. The targets were presented at both x1 and x2 IPD (i.e., with X_{ipd} equal to 1 and 2 respectively, see section 3.2.4.3, equation 3.1) in order to see how this would affect the subjects' responses.[2] The results generally show good correspondence between the position of the probe and the calculated position of the target, although some discrepancies were observed, and depth resolution improved in the virtual space when the effective IPD was increased.

---

[1]The 3D calibration experiment was actually performed after the z-axis and virtual-virtual experiments. It is presented first here in order to simplify the overall discussion.

[2]Note that one can assume here that there was no adaptation in the experiment because the subjects did not receive any feedback about the transformation.

4.1.1 Method

The experiment involved two tests, P1 and P2 (v2.201),[3] that were identical except for the trial group order. Two tests were used so that the group order could be counter-balanced among the subjects. The subjects were alternately assigned to one test or the other when they were scheduled to take the test.

4.1.1.1 Subjects: Five MIT students, 4 males and 1 female, served as subjects in the experiment. Subjects S1, S8, and S15 performed test P1 and subjects S11 and S12 performed P2. Each subject only performed one test, although both S12 and S15 had to repeat their respective tests a few days later because the initial data was unusable. This happened because the PHANToM was not properly reset on the first attempt. The results of the test with S8 are also compromised because of a problem with the PHANToM calibration (see section 4.1.2).

The subjects ranged in age from 20 to 40 years; IPDs ranged from 6.17 to 6.77 cm; and visual acuity (corrected in 3 cases) ranged from 20/13 to 20/25 as measured with a standard Snellen eye chart. Stereoacuity was measured using a Randot test and found to be 70 arc-sec or better in all cases. Two of the subjects (S8 and S12) did not consider themselves to be strictly right or left-handed (see appendix B), although they typically performed the primary task in these experiments with their right hands. The other three subjects were right-handed.

4.1.1.2 Apparatus and Stimuli: The LED fixture (shown in figure 3.6) served as the physical probe in the experiment. This fixture is designed so that the illuminated region of the red LED, roughly 0.5 cm in diameter, is centered at the endpoint of the PHANToM within ±0.2 cm. Note that the LED is seen by the subject through the partially silvered mirror (see section 3.2.1.2).

Figure 4.1 presents a full-scale plane view of the experimental stimulus from the direction of the subject. The virtual target presented on the display was an empty red square, 1.0 cm on a side in the virtual space, parallel to the xy plane. The target was rendered without jitter as described in section 3.2.4.5.

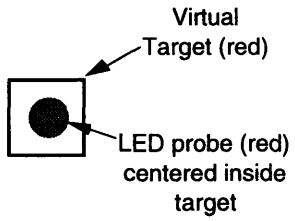


Figure 4.1.

[3]The test performed by subject 1 was actually run under an earlier version of the RunTest application (v2.200), however, the two versions are identical for the given test (P1).

4.1.1.3 Procedure: The test consisted of 4 groups of trials where the value of X_{ipd} alternated by group. In the P1 version of the test the group order was: x1, x2, x1, x2; and in P2 the order was: x2, x1, x2, x1. The groups were otherwise identical except for the random order of the trials. Five trials were performed at each of 22 target positions for a total of 110 trials per group in random order.

In each trial, the subject moved the LED probe so that the center of the illuminated part of the LED appeared to be centered inside the virtual target. More specifically, the LED was centered at the mid-point of the square (in x and y, as shown in figure 4.1) and so that it appeared to be at the same depth as the plane of the target (in z). The subject was free to use either hand to position the probe. When the position of the LED appeared to coincide with the position of the target, the subject used his or her other hand to click a button on the trackball to input the data and begin the next trial.

Training for the task was minimal. First, the experimenter explained the task briefly, answering any questions that the subject might have had about the procedure. Then the experimenter observed the subject during the first few trials of the test and asked questions to insure that the procedure was understood. The task procedure, while informal, was nominally as follows.

Subjects were asked to hold the probe so that the bright end of the LED pointed toward them and so that stray light from the probe did not illuminate their hands excessively. One reason for this was to avoid the superposition effect discussed in section 2.2.2.3. The goal was for the subject to see a stimulus consisting only of the virtual target and the LED, such that the two did not overlap. This was the essential appearance of the stimulus in all cases, even when other objects under the mirror were faintly visible to the subject.

Subjects were also asked to move the probe on each trial even if the target appeared to be at the same place. In the latter case the subject was supposed to move the probe away from the old location by at least a few centimeters before moving back to the target. This was done in order to insure independent responses on each trial.

4.1.1.4 Data Analysis: When $X_{ipd} = 1$, the experimental space and the virtual space are identical and it is straightforward to compare the response position (experimental space) to the target position (virtual space). Otherwise, as in the present experiment when $X_{ipd} = 2$, the

positions must be transformed into a common space before they may be compared. The approach taken here is to transform the subjects' responses (LED probe positions) into virtual space coordinates.

The following transformation is based on the same geometry that is used to derive the drawing algorithm in chapter 3 (see figure 3.17 in section 3.2.4.3). The basic idea is that one first calculates the projections of the point in the experimental space (x_e,y_e,z_e) onto the display plane according to equation 4.1. Note that equation 4.1 is identical to equation 3.2 except that all of the quantities, and in particular x_{eye}, now refer to the experimental space. The inverse transformation is then used to calculate the position of the corresponding point in virtual space (x_v,y_v,z_v) according to equation 4.2.

Transformation from (x_e,y_e,z_e) to (L_x,L_y,R_x,R_y):

x_{eye} = IPD/2 = actual eye position in experimental space

$$\begin{aligned} L_x &= (((x_e + x_{eye})/(z_{eye} - z_e)) * z_{eye}) - x_{eye} \\ L_y &= (((y_e - y_{eye})/(z_{eye} - z_e)) * z_{eye}) + y_{eye} \\ R_x &= (((x_e - x_{eye})/(z_{eye} - z_e)) * z_{eye}) + x_{eye} \\ R_y &= (((y_e - y_{eye})/(z_{eye} - z_e)) * z_{eye}) + y_{eye}. \end{aligned} \qquad (4.1)$$

Inverse transformation from (L_x,L_y,R_x,R_y) to (x_v,y_v,z_v):

x_{eye} = IPD*X_{ipd}/2 = effective eye position in virtual space

$$\begin{aligned} z_v &= ((R_x - L_x) * z_{eye}) / (R_x - L_x - (2.0*x_{eye})) \\ y_v &= L_y * ((z_{eye} - z_v)/z_{eye}) \\ x_v &= ((L_x + x_{eye}) * ((z_{eye} - z_v)/z_{eye})) - x_{eye}. \end{aligned} \qquad (4.2)$$

The key difference between equations 4.1 and 4.2 is the value of x_{eye}. Specifically, in equation 4.1 the actual eye position is used while in 4.2 one uses the effective eye position. Note however that, while this is the simplest way to conceptualize the transformation, any non-zero value of IPD gives the same result because IPD cancels when the two equations are combined. Also note that the overall transformation uses the ideal screen coordinates. In other words, display pixelation and screen curvature are not included.

4.1.2 Results

In general, the correspondence between the subjects' responses (i.e., the transformed position of the LED probe, see section 4.1.1.4) and the centers of the targets was roughly within ±0.5 cm in the virtual space. The error was sometimes more than 1 cm, however, particularly in the z direction. Also note that the standard deviation of the responses was smaller than the absolute error in many cases, suggesting that the subjects' errors were systematic (i.e., that the responses were biased). The variance of the responses was also smaller in the z direction when the effective IPD was doubled, corresponding to improved resolution.

The bias and the resolution results are described in more detail in the following sections (4.1.2.2 and 4.1.2.3, respectively) after the main figures that summarize all of the data are described in section 4.1.2.1.

4.1.2.1 Presentation of the Data: Figures 4.2 and 4.3 present the mean combined responses, at x1 and x2 IPD, respectively, of subjects S1, S11, S12, and S15. The x1 IPD results for S8 (calculated in a similar manner) are presented separately in figure 4.4, because of a problem with the calibration of the PHANToM that occurred after the other tests and before her test (see section C.2.3). S8's responses are consistent with the other subjects, just shifted to the right as discussed in section 4.1.2.2 and appendix C. Her x2 IPD results (not presented here) are similar.

Each figure presents two views of the virtual space (like a mechanical drawing), showing the targets as a reference. The xy plot (top) shows the subject's view and the xz plot (bottom) shows the view looking up from bottom of the space (see figure 3.12). In the xy plot the 22 targets appear to be squares, as they would appear to a subject, while from the bottom they appear to be line segments.

The other symbols in the figure mark the responses of the four subjects. Specifically, the center of each symbol corresponds to the mean response position (transformed per equations 4.1 and 4.2), while the height and the width of the symbol show the standard deviation of the responses in the given direction. Note that the distance from the center of each symbol to the edge (i.e., top, bottom, left, or right) is three times the standard deviation of the responses in that direction.

The mean responses in the figures were calculated as follows. First, the transformed results for all trial groups with the same value of X_{ipd} were combined. Then, the mean values of the 10 responses per target for

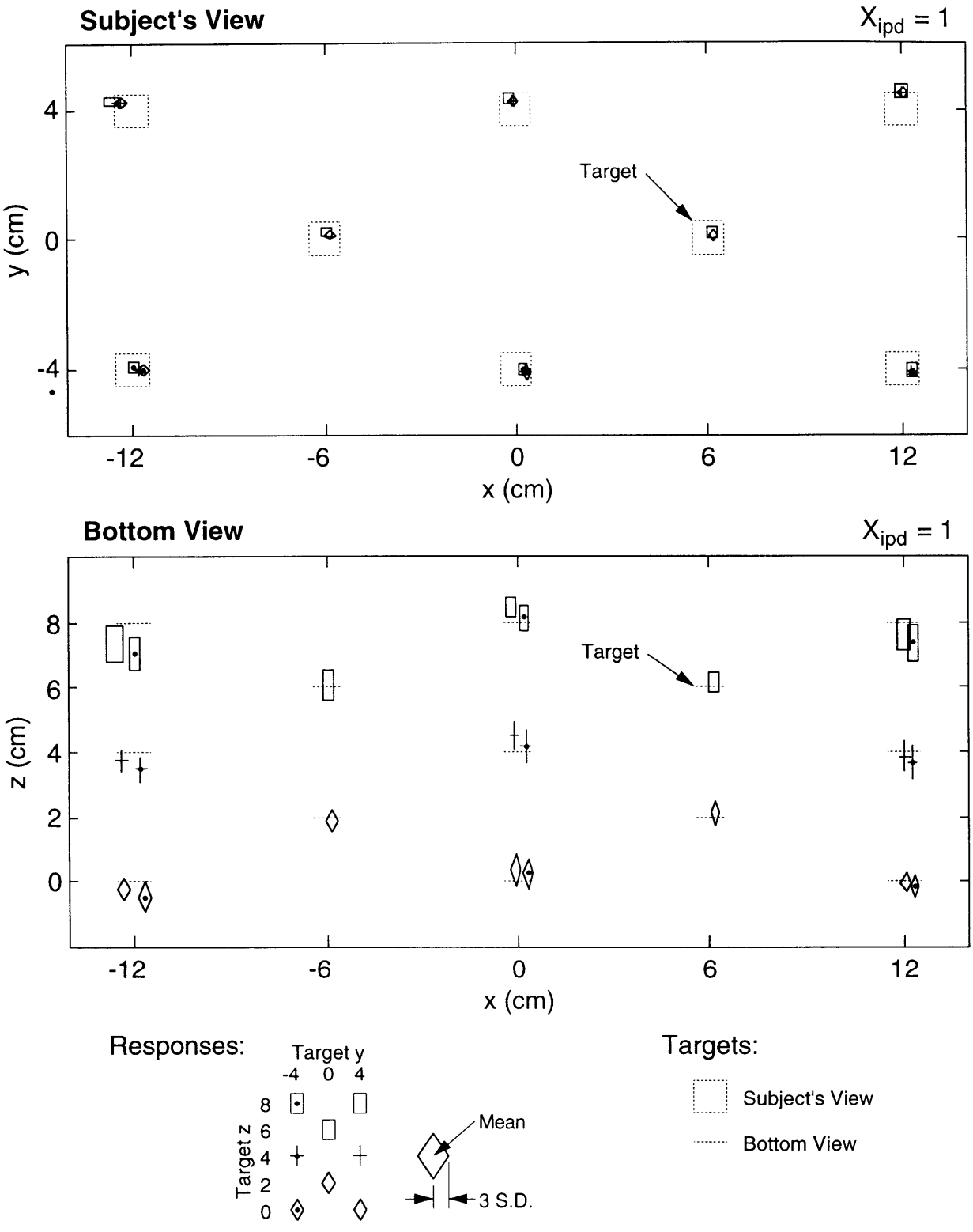


Figure 4.2. Mean combined positioning responses of subjects S1, S11, S12, and S15, at 22 target locations in virtual space at x1 IPD (tests P1 and P2, v2.201).

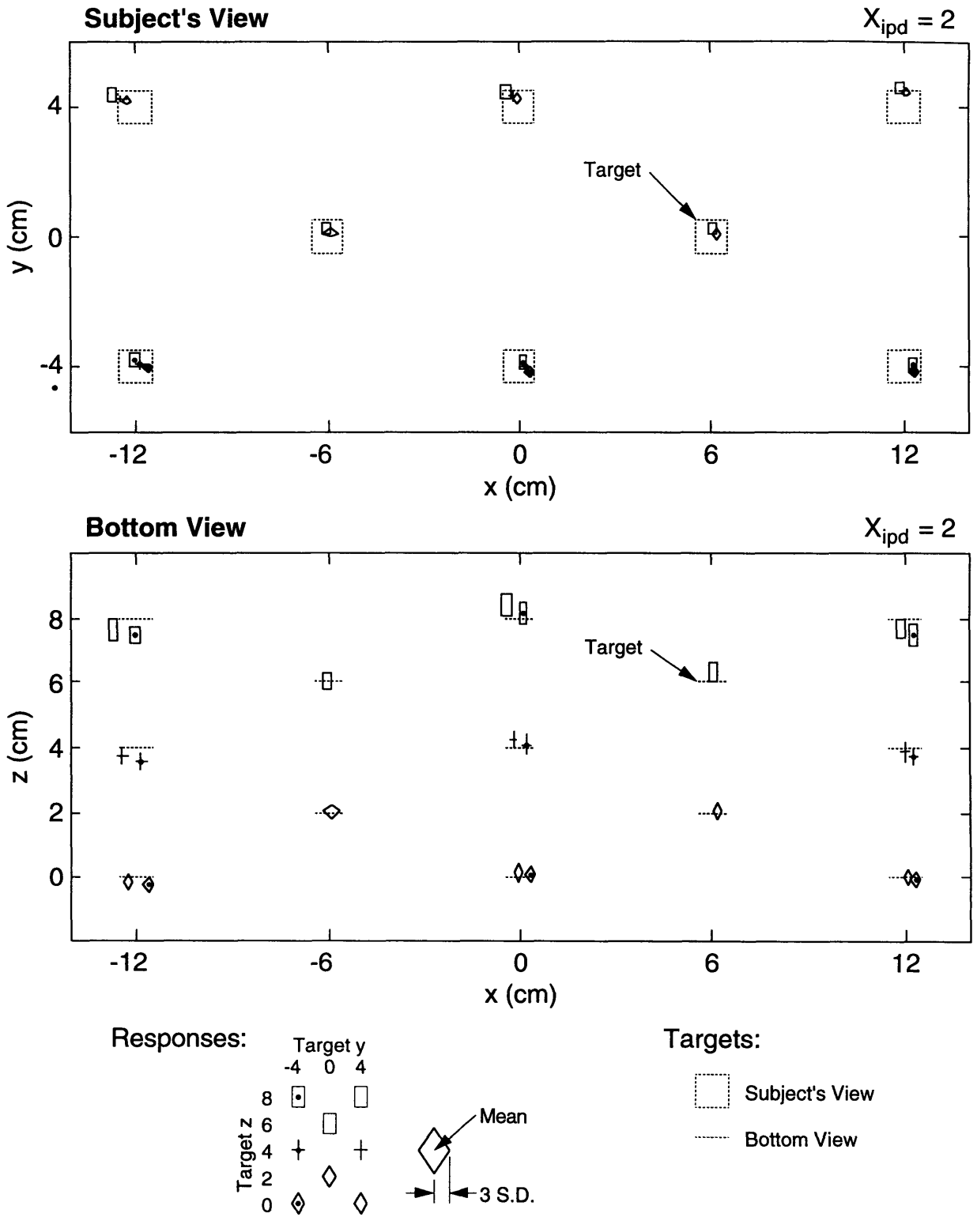


Figure 4.3. Mean combined positioning responses of subjects S1, S11, S12, and S15, at 22 target locations in virtual space at x2 IPD (tests P1 and P2, v2.201).

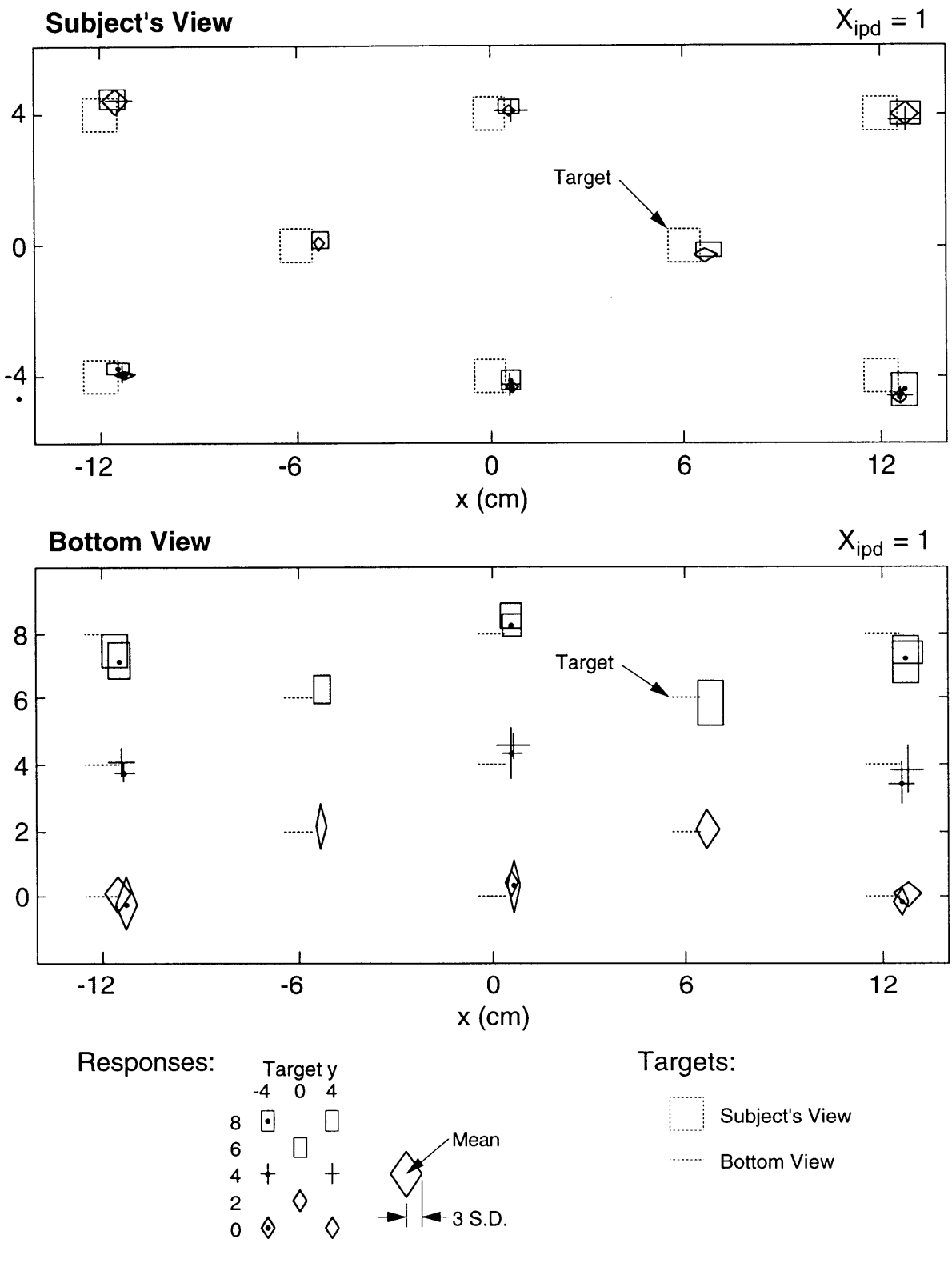


Figure 4.4. Mean positioning responses of subject S8, at 22 target locations in virtual space at x1 IPD (test P1, v2.201). Rightward shift due to a PHANToM calibration problem.

each subject were calculated. And, finally, the mean values of the subjects were averaged. Note that this calculation gives a separate mean value for each component direction (x, y, and z).

Similarly, the standard deviation of the responses was calculated as the mean (across subjects) of the standard deviation of each subject's (transformed) responses for the given target in the given direction. For example, to calculate the standard deviation of the responses in the x direction for a given target, one first calculates the standard deviation of the x coordinates of the 10 transformed responses (for that target) for each of the subjects. Then the standard deviations of each subject are averaged across all of the subjects whose data is included in the figure.

The same symbols are used for the corresponding points in both plots. Specifically, the symbols are ordered (□, +, ◊) from front to back in the xy plot. This corresponds to decreasing values of z as seen in the xz plot. Note that the (+) is not required for $y = 0$ where there are only two response points in each column. Also note that a small dot is used to distinguish the responses at $y = -4$ cm (with •) from the responses at $y = 4$ cm (without •) so that each position may be identified.

4.1.2.2 Bias: Referring to figures 4.2 to 4.4, the most noticeable result is the fact that the mean responses do not fall at the centers of the targets (i.e., the mean error is not zero). Two trends in the mean response are apparent in the figures. First, the mean responses at $y = 4$ cm are shifted to the left and slightly up from the corresponding responses at $y = -4$ cm. Interestingly, this trend in not apparent in the data for S8 (see figure 4.4). Second, the responses at $x = 0$ tend to miss the target in the positive z direction while responses at $x = \pm 12$ cm tend to err in the negative z direction. In other words, the response space viewed in the xz plot seems to bow-up in the middle. Note that this trend is also evident for S8.

4.1.2.3 Resolution: Figures 4.5 and 4.6, respectively, present the mean absolute error and the mean standard deviation of the responses (described in more detail shortly). In each figure, the given resolution measure (i.e., the mean absolute error or the mean standard deviation of the responses) is plotted against spatial direction (x, y, or z). The ordinate gives the resolution measure and the abscissa identifies the spatial direction, with the x1 and x2 IPD cases shown side by side. Specifically, the x1 IPD case is shown on the left (o) and the x2 IPD case is shown on the right (x).

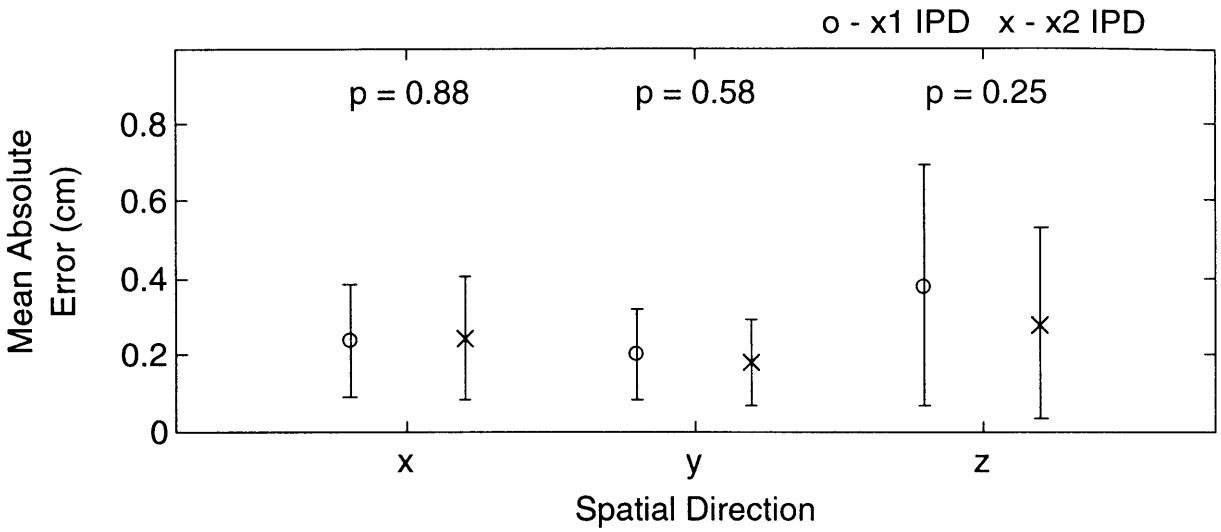


Figure 4.5. Mean absolute positioning error averaged over all 22 targets and subjects S1, S11, S12, and S15 (tests P1 and P2, v2.201).

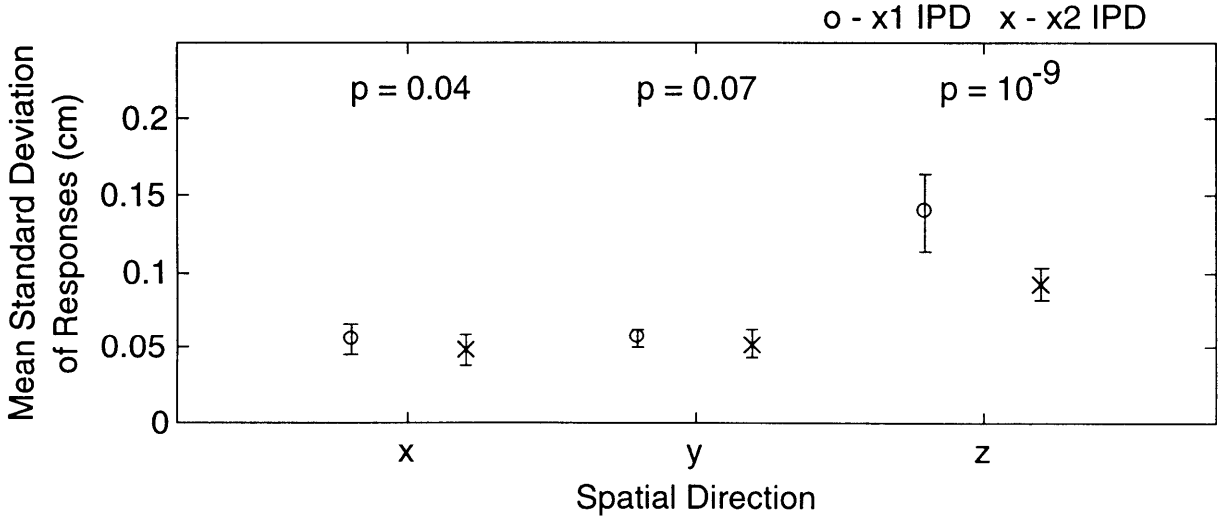


Figure 4.6. Mean standard deviation of the responses of subjects S1, S11, S12, and S15, averaged over all 22 targets (tests P1 and P2, v2.201).

The result of a hypothesis test, comparing the x1 and x2 IPD cases, is also presented above each pair of points. Specifically, "p" is the probability of the corresponding observation, given that the two sets of measurements are actually based on the same underlying distribution (i.e., that the true means are actually the same). The significance level is calculated in a T-test using the Matlab function "ttest2."

Figure 4.5 presents the mean absolute error in the task. Specifically, the error is equal to the mean (averaged across all 22 targets) of the absolute difference between the target position and the mean response (for that target), where the mean response is averaged across all of the subjects except S8. In other words, the figure gives the mean absolute error averaged across the 22 mean responses (one for each target) shown in either figure 4.2 or 4.3 (recall that figure 4.2 gives the 22 mean responses for x1 IPD, while figure 4.3 gives the 22 mean responses for x2 IPD). The error bars in each case show plus or minus one standard deviation, calculated for the corresponding set of 22 errors used to calculate the mean absolute error. In other words the error bars are related to the variance of the subjects' absolute error across the set of targets.

Note that there is no significant change in the mean absolute error when the effective IPD is increased. This is true in all directions at the 25% level, a result that is surprising in the case of the z direction. The observed error does get slightly smaller for the z direction, going from x1 to x2 IPD, but, as indicated above, the change is not very significant.

Figure 4.6 presents the mean standard deviation of the responses in each direction, corresponding to the average height and width of the symbols in figures 4.2 and 4.3 (see section 4.1.2.1). In other words, the standard deviation of the responses is averaged across the set of 22 targets for each direction (x, y, and z). Also, as in figure 4.5, the error bars in figure 4.6 show plus or minus one standard deviation of the given resolution measure across the set of 22 targets, where in this case the measure is the standard deviation of the responses.

There is a decrease in the mean standard deviation of the responses (ordinate) between x1 and x2 IPD in every direction. The change is clearly significant in the z direction, well beyond the 1% level. Surprisingly the mean standard deviation of the responses also changes slightly for x and y, although the change is small and only significant at a level of 5-10%. Note that the scale of the plot makes it hard to see the change, but when

the ordinate is magnified so that the error bars are about the same size as in the absolute error plot (figure 4.5), the difference is apparent.

Finally, note that the results for x and y are similar in both plots. Specifically, compare the magnitudes of the given resolution measure in each case, the size of the error bars, and the direction of change between x1 and x2 IPD. In contrast, the results for z are noticeably different from x and y.

4.1.3 Discussion

Some of the resolution results are approximately what one would expect. Specifically, similar results for x and y were observed, with little or no change when the effective IPD was changed, and there was improved resolution in the z direction when the effective IPD was increased. The latter conclusion is based on the fact that the standard deviations of the responses get smaller, while the absolute error stays the same or gets slightly smaller (see figures 4.5 and 4.6 and/or compare figures 4.2 and 4.3).

Other resolution results were not what one would expect however. These unexpected results are discussed next, followed by an analysis of the apparent systematic errors in mean response.

4.1.3.1 Unexpected Resolution Results: Two aspects of the resolution results were surprising. First, why is there so much evidence of change in the mean standard deviation of the responses for x and y (i.e., a 5-10% significance level)? And, second, why is there so little evidence of change in mean absolute error in the z direction between x1 and x2 IPD (i.e., only a 25% significance level)?

An answer to the first point (i.e., why there is a change in the mean standard deviations for x and y), might simply be that the increased effective IPD makes the targets appear closer in the experimental space. The nearer targets are visually bigger, making it easier to see the relative position of the probe. In any event it is a small effect.

One explanation for the second point (i.e., why there is relatively little change in absolute error in the z direction), might be that much of the error is due to systematic bias. In other words the subjects are being more precise, but at the wrong places. As a result, the improvement in precision that is apparent in the reduction of the standard deviation of the

responses, might not have a significant affect on the mean absolute error in this particular test.

4.1.3.2 Systematic Position Errors: If one compares the magnitude of the mean errors in figures 4.2 and 4.3 to the standard deviations of the responses (height and width of the symbols), it appears likely that the positioning errors for many of the targets are systematic rather than purely random. There are several possible reasons why this might be so.

A likely explanation for the first trend discussed in section 4.1.2.2, that the mean responses at $y = 4$ cm are shifted to the left and slightly up from the corresponding response at $y = -4$ cm, is that the relative calibrations of the PHANToM and the display are slightly off. Consistent with this explanation is the fact that the results for S8 (figure 4.4) do not show the same shift to the left, after the calibration of the PHANToM was disturbed.

The second trend discussed in section 4.1.2.2, namely that the response space seems to bow-up in the positive z direction, is almost certainly the result of screen curvature. In particular, the center of the display screen is about 0.5 cm closer to the subject than it is at $x = \pm 12$ cm. Thus, particularly at $z = 0$, one would expect to see the bowing-up that is observed in the experiment.

The fact that the responses tend to be slightly in front of the targets at $x = 0$ may be related to the effect described by Ellis et al. (see section 2.2.2.3), although given the magnitude of the other mean errors (see figures 4.2 and 4.3) it is more likely to be due to the inaccuracy of the calibration. Similarly, display pixelation may account for some of the observed bias. Ultimately, however, one cannot decide these issues without some measure of the accuracy of the PHANToM.

4.2 Z-axis Experiment

In the z-axis calibration experiment, virtual targets were presented randomly at 5 standard z coordinates along the z axis: -10, -5, 0, 5, 10 (cm). Subjects manually adjusted a physical probe so that it appeared to coincide with each target position in repeated random trials. In all cases the effective IPD was equal to the subject's nominal IPD (i.e., $X_{ipd} = 1$). The PHANToM calibration was also tested independently in connection with the tests and it was observed that the mean responses (probe position) agreed with the calculated target positions within the accuracy of the calibration measurements.

4.2.1 Method

The experiment involved two main tests, T1 and T2 (v2.101), that are identical except that T2 includes an additional virtual object at z = 0. The additional object was intended to serve as a visual reference for the subject, but it did not appear to have a significant effect on the results. Two preliminary tests (T2A and T2B) were also performed. These are identical to T2 except in one case (T2A) the subject's left eye is covered and in the other case (T2B) the effective IPD is set equal to zero.[4]

4.2.1.1 Subjects: Ten MIT students, 8 males and 2 females, served as subjects in the experiment. Four of the subjects (S1, S14, S15, and S16) performed test T1 and seven subjects (S1, S6, S8, S9, S10, S11, and S12) performed T2. Note that S1 took part in both tests while each of the other subjects only performed one test. S1 also performed the preliminary tests T2A and T2B.

The subjects ranged in age from 18 to 40 years; IPDs ranged from 6.17 to 6.77 cm; and visual acuity (corrected in 4 cases) ranged from 20/13 to 20/40 as measured with a standard Snellen eye chart. Stereoacuity was measured using a Randot test and found to be 70 arc-sec or better in all cases. All of the subjects were comfortable positioning the probe with the right hand, although two of the subjects (S8 and S12) did not consider themselves to be strictly right-handed (see appendix B).

4.2.1.2 Apparatus and Stimuli: The slider was used to constrain the motion of the PHANToM in the experiment and the red LED mounted on the slider served as the physical probe. The slider was designed so that the illuminated region of the LED, as seen by the subject through the partially silvered mirror, was approximately centered below the z axis at the same z coordinate as the endpoint of the PHANToM, within ±0.1 cm (see section 3.2.2.4).

---

[4]The latter was done by setting the value representing the subject's actual IPD in the test document (d_{ip} in eqn. 3.1) equal to zero (see section 3.2.4). A slightly different eye-positioning procedure (see section 3.4.4) was required in the case of T2B because version 2 of the underlying library is not designed to do eye position measurements for a subject with IPD = 0. The alternate procedure used here was to measure eye position before and after the test from the set-up menu. In doing this, the subject (who was also the experimenter) had to reposition his head in the headrest by feel. The measured eye positions before and after the test are as follows:

Before: L = (-3.28,0.07,61.16), R = (3.34,0.08,61.66)
After: L = (-3.34,0.15,61.35), R = (3.33,0.15,61.41).

Figure 4.7 presents a full-scale plane view of the stimulus from the direction of the subject. The virtual target presented on the display was a solid red rectangle parallel to the xy plane, located slightly above the z axis. The target was 2.0 cm wide (in the x direction) by 1.0 cm high (in the y direction), centered at x = 0 cm and y = 0.8 cm in the virtual space. A reference frame was also presented on the display in the case of test T2, but not in test T1. The reference frame consists of 3 concentric, empty, *green* squares parallel with the xy plane and centered at the origin.

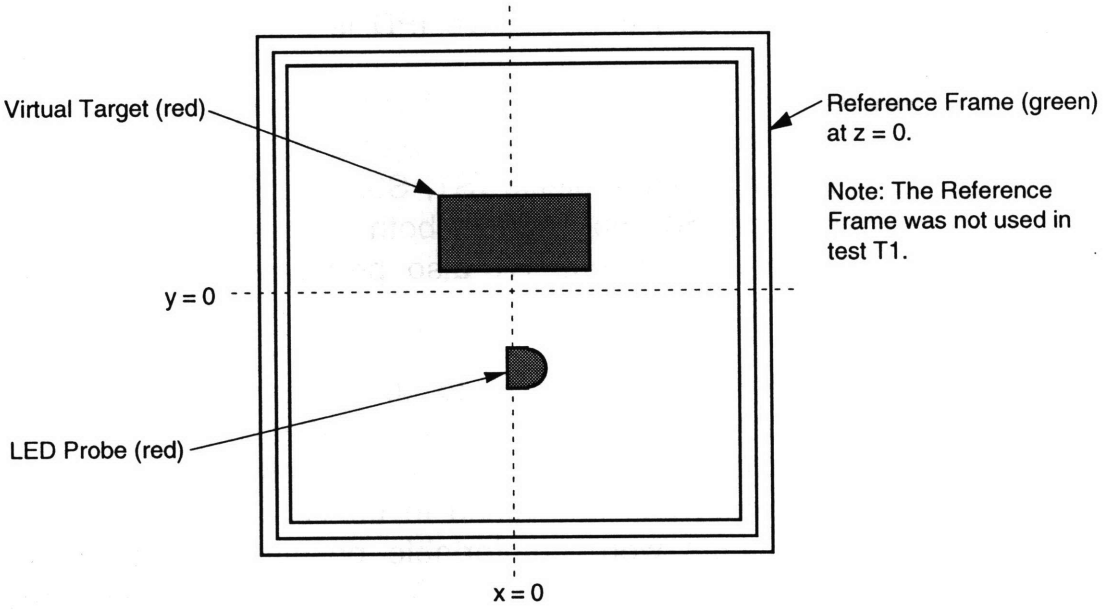


Figure 4.7. LED Probe and Virtual Target (tests T1, T2, T2A, and T2B, v2.101).

The virtual target was rendered with jitter in a manner similar to the one described in section 3.2.4.5, except that the screen coordinates of both corners were used directly. In other words, both the left and right views correspond to the projected regions where one s_point defines the lower-left corner and the other s_point defines the upper-right corner. This rendering algorithm sometimes caused the rectangle to appear tipped in the z direction when a small difference in the parallax of the two points was rounded to a one-pixel difference on the screen.

The reference frame (when it was used) was rendered in the same way as the virtual target, except without jitter. Note that the parallax of the reference frame was always identically zero. One consequence of this was that the frame never appears tipped in the z direction as may occur for the target. Another consequence was the fact that ghosting did not occur because both views were identical except for a 1/2 pixel vertical shift.

4.2.1.3 Procedure: The test consisted of 5 groups of 100 trials that were identical except for the random order of the trials. In each group, 20 trials were performed at each of the 5 target positions in random order. Note that the test was broken into groups simply to give the subject a chance to rest.

In each trial, the subject moved the slider using the right hand so that the probe appeared to be at the same distance (z coordinate) as the target. When the probe and the target appeared to be at the same distance, the subject used the left hand to click a button on the trackball. This action served to input the current position of the slider as the subject's response. It also served to begin the next trial.

Training for the task was minimal. First, the experimenter explained the task briefly, answering any questions that the subject might have had about the procedure. Then the experimenter observed the subject during the first few trials of the test and asked questions to insure that the procedure was understood. Subjects were asked to move the probe a few centimeters on each trial, even if the target appeared to be at the same place, in order to insure independent responses.

At some time before or after the test, while the same PHANToM calibration was in effect (i.e., before the encoders were reset) the experimenter tested the PHANToM calibration according to the procedure described in appendix C, section C.2.2.

4.2.2 Results

In general, there was good correspondence between the measured z coordinate of the LED probe and the calculated z coordinate of the virtual target. In particular, there was no evidence of the effect described by Ellis et al. (1995; 1994) in which a virtual object appears closer when it is visually superimposed on a physical background (see section 2.2.2.3). This suggests that the physical objects that were faintly visible under the mirror in the thesis calibration experiments (see section 3.2.1.2) did not significantly affect the results.

Figure 4.8A shows the combined results for 7 subjects in the two main versions of the experiment (T1 and T2). The plot presents the mean of all responses with error bars showing plus or minus three standard deviations. Specifically, for each target distance (z coordinate), all of the subjects' responses were combined into a single set and then the mean and

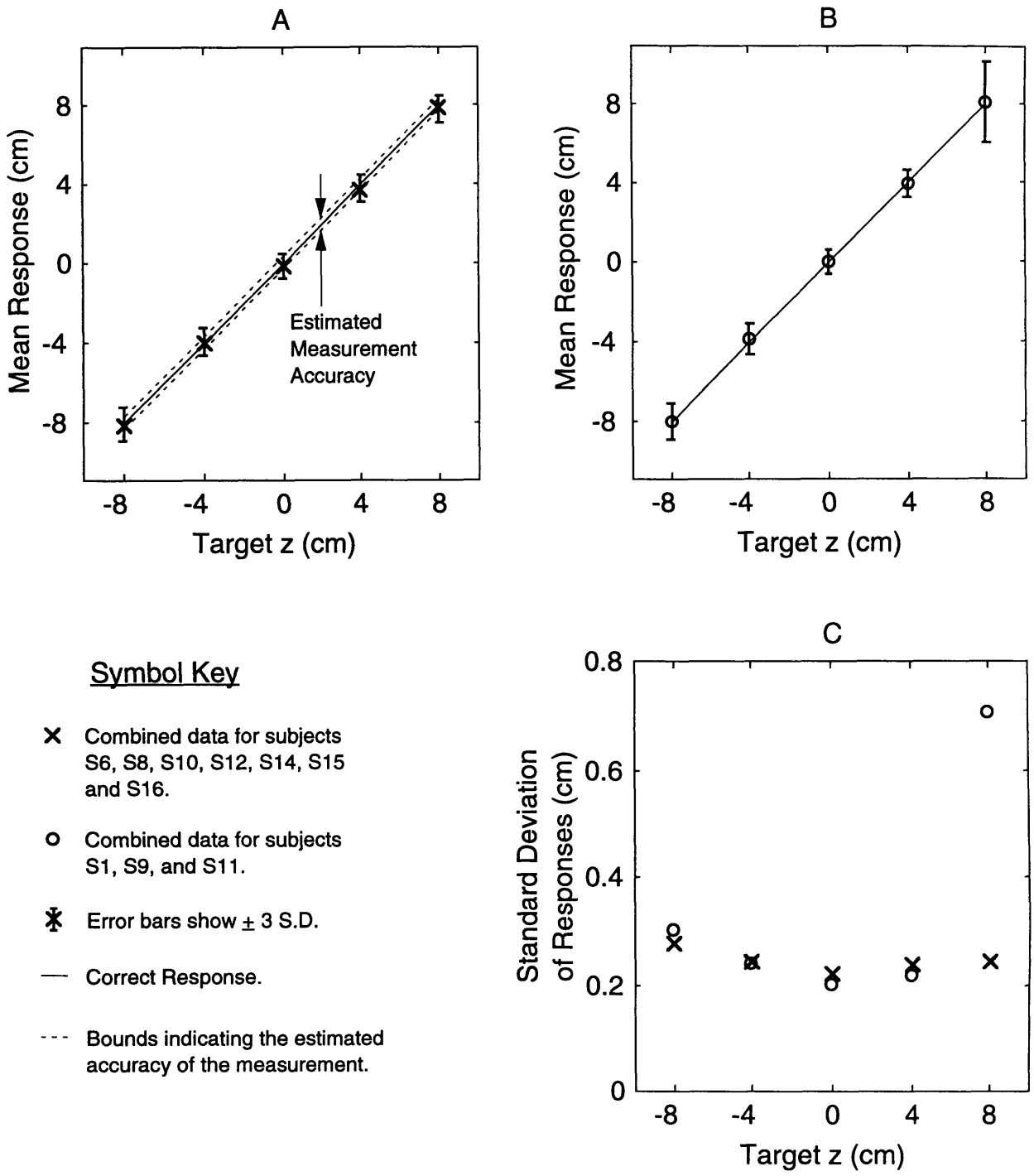


Figure 4.8. Mean and standard deviation of responses in manual positioning of a physical probe with respect to a virtual target on the z axis in virtual space at x1 IPD (tests T1 and T2, v2.101).

the standard deviation of that set were calculated to determine the mean response and the error bars for the given z coordinate. The solid line shows the correct response and the dashed lines indicate the estimated tolerance of the measurement.[5]

Figure 4.8B presents a similar plot for the three other subjects in the experiment (S1, S9, and S11) who are not included in figure 4.8A. These subjects were not included in the first figure because their performance was noticeably different than the others. Specifically, they had much larger variances in their responses for targets located at the maximum z value (8 cm). Subject 1 also had a higher standard deviation (0.43) at -8 cm.

Figure 4.8C shows the means of the standard deviations for the two groups of subjects for further comparison. Specifically, in this plot the standard deviations were first calculated for the responses of each subject (for each target distance), then the individual standard deviations were averaged across the given group of subjects (for each target distance). Note that figures 4.8B and 4.8C do not include the results for S1 in T1 in order to give equal weight to all three subjects; however, Subject 1 had similar results in both T1 and T2.

Finally, figure 4.9 presents the results of the preliminary tests (T2A and T2B) that were performed by S1. The individual results for S1 in test T2 are also shown for comparison. These results clearly show that performance in the task is severely degraded when stereoscopic cues have been eliminated by either closing one eye or setting the effective IPD to zero.

4.2.3 Discussion

The experimenter (S1) observed during his second test (T2) that it was subjectively more difficult to judge the targets at the maximum values of z (positive or negative) corresponding to the targets with the largest parallax. His dominant eye also became quite sore. One explanation for these effects is the conflict between accommodation and convergence that exists on the display for stimuli with large parallax. The problem is particularly acute in the present case where the subject attempts to

---

[5]The total tolerance in the z direction is estimated to be ± 0.3 cm as follows: 1) the LED on the slider block is aligned with the PHANToM endpoint within ± 0.1 cm (see section 4.2.1.2); 2) the result of the PHANToM calibration test suggests that the endpoint position is measured within ± 0.2 cm with respect to the experimental space (see section 3.2.5).

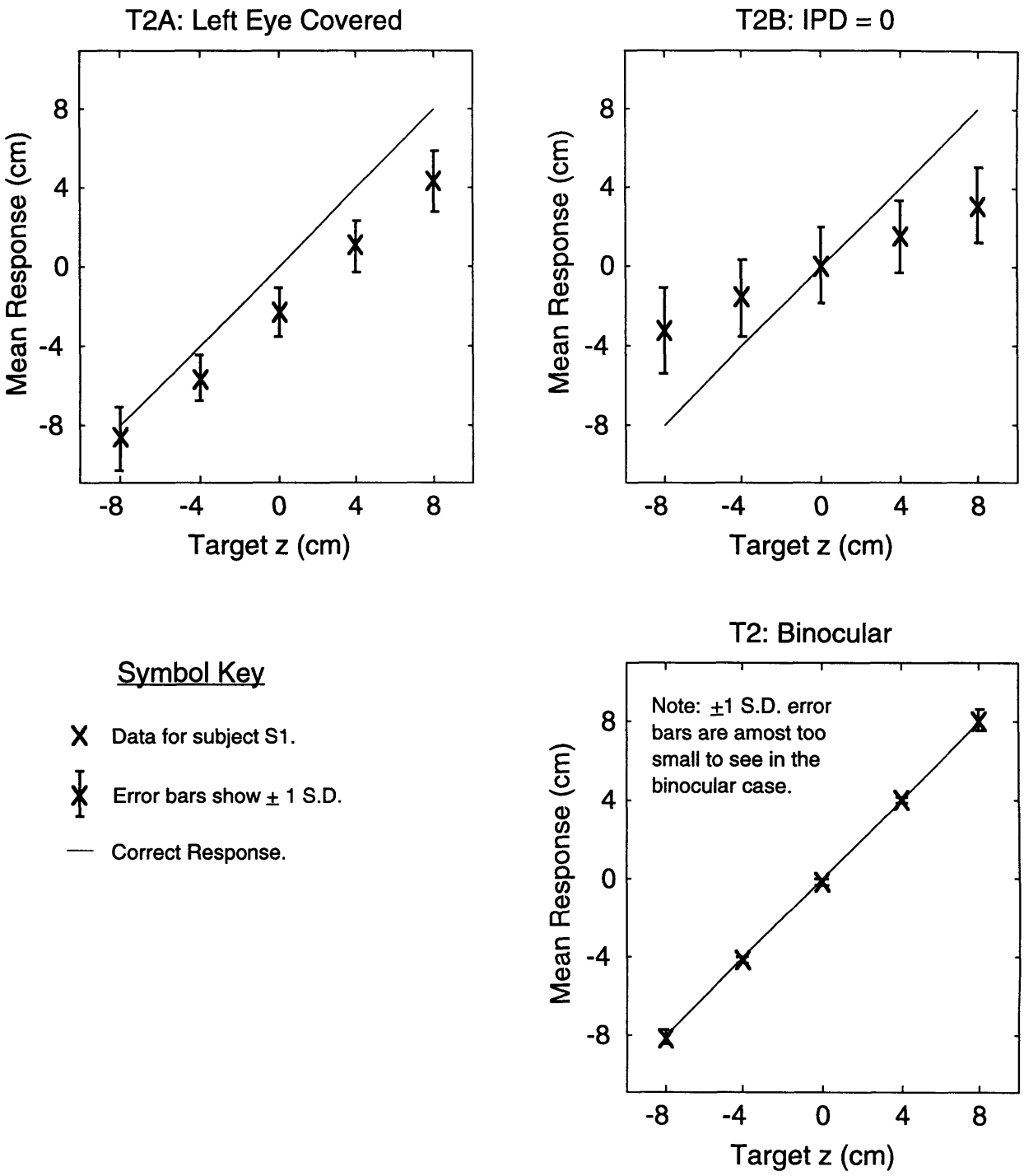


Figure 4.9. Comparison of monocular and binocular manual positioning of a physical probe with respect to a virtual target on the z axis in virtual space at x1 IPD (subject S1; tests T2, T2A, and T2B, v2.101).

fixate on both a physical object (the probe) and a virtual object (the target) at the same time. Perhaps the inability to focus on both objects makes some subjects more uncertain about the relative position.

4.3 Virtual-Virtual Experiment

The virtual-virtual calibration experiment was identical to the z-axis experiment, described in section 4.2, except that a virtual probe was used in place of the physical LED probe. The experiment was designed to investigate the fact that three of the subjects in the z-axis experiment had much larger variances in their responses for stimuli located at large positive or negative z values than did the other subjects. It was hypothesized that the cause of this artifact might be related to the different amount of accommodation required for the two stimuli (see section 4.2.3). Specifically, the given subjects might be more uncertain in their judgments of the relative distance when they are unable to focus on the virtual target and the LED in the same way. Consequently, it was predicted that the larger variance would disappear if the position of two virtual objects were to be compared because, then, the level of accommodation would be the same. The predicted result was observed.

4.3.1 Method

The experiment involved two tests, T4 and T5 (v2.101), that were identical except for the size of the virtual probe. The probe in T4 was the same size as the virtual target (2.0 cm x 1.0 cm), while the probe in T5 was approximately the same size as the LED (0.5 cm x 0.5 cm). This difference had no significant effect on the mean results.

4.3.1.1 Subjects: The experiment was performed by the three subjects (S1, S9, and S11) who had a larger variance in their responses than the other subjects in the z-axis experiment (see 4.2.2). Subject S1 performed test T4 and the other two subjects performed test T5.

4.3.1.2 Apparatus and Stimuli: Figure 4.10 presents a plane view of the experimental stimulus from the direction of the subject. The apparatus was identical to the apparatus in the z-axis experiment except for the probe. The virtual probe was an empty red rectangle rendered in the same manner as the virtual target except without jitter. Also note that the green piece of paper, that would normally be placed under the partially silvered mirror in tests where the subject does not need to look through the mirror, was not used in this test. In other words the subjects

could see the same background objects through the mirror, faintly illuminated by stray light, as in the z-axis experiment.

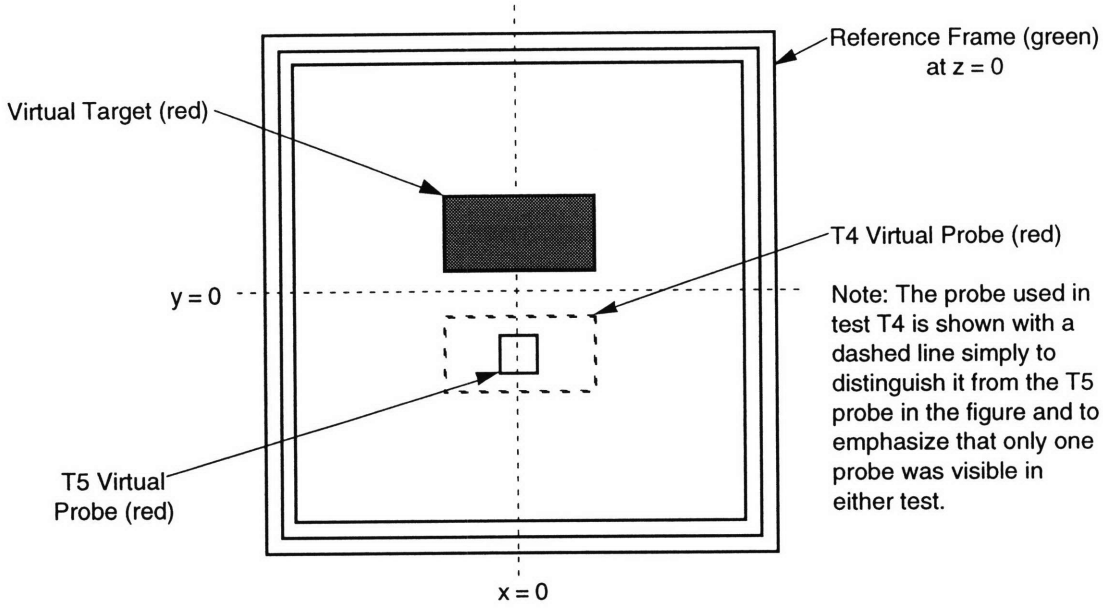


Figure 4.10. Virtual Probe and Target (tests T4 and T5, v2.101).

4.3.1.3 Procedure: The procedure was identical to the procedure used in the z-axis experiment (see section 4.2.1.2).

4.3.2 Results

Figure 4.11 presents the results of the experiment. The plots in the figure are similar to the ones that are used to present the results of the z-axis experiment (see section 4.2.2). Figures 4.11A-4.11C present the mean responses of the individual subjects (S1, S9, and S11, respectively) in the virtual-virtual experiment, with error bars showing plus or minus 3 standard deviations. The quantities in these plots were calculated in exactly the same way as in figures 4.8A and 4.8B except that the present plots are for single subjects rather than for groups of subjects.

Figure 4.11D shows the mean standard deviations for the three subjects in the virtual-virtual experiment. The quantities in this plot were calculated in an identical manner to the quantities in figure 4.8C. Specifically, the standard deviations were first calculated for the responses of each subject and then the individual standard deviations were averaged across the three subjects.

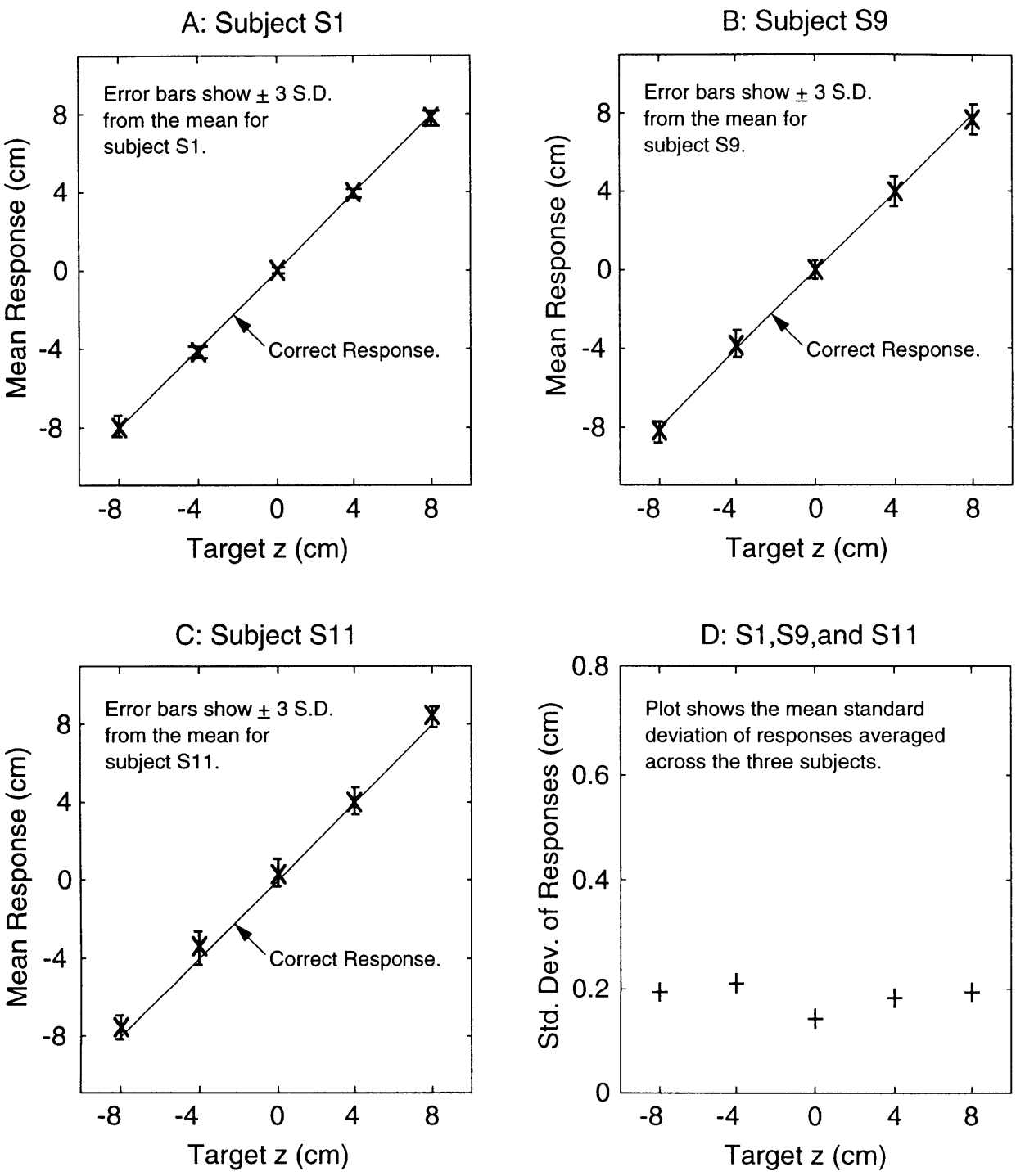


Figure 4.11. Mean and standard deviation of responses in manual positioning of a virtual probe with respect to a virtual target on the z axis in virtual space at x1 IPD (tests T4 and T5, v2.101).

It can be seen in the figure that the mean standard deviation of the three subjects is approximately independent of the target z coordinate. In other words, as expected, the variance is not larger at the extreme values (as it was in the z-axis experiment). Thus it would appear that a conflict between convergence and accommodation, while it doesn't alter the mean response, can reduce resolution in some cases.

4.4 Display Pixelation Analysis

The mean results of the calibration experiments tell only part of the story. A more detailed analysis reveals subtle affects of display pixelation (defined in section 2.2.2.4). Specifically, there are systematic trends in the subjects' positioning errors.

Figures 4.12A-4.12C present some representative plots from the virtual-virtual experiment (see section 4.3) that help to explain the results. In each figure, a response histogram is vertically juxtaposed above the parallax values for the virtual probe. The abscissa in all of the plots is the z coordinate in virtual space, the ordinate for the histograms (top plot in each figure) is the number of responses per bin (where "bins" are uniform segments of the z axis), and the ordinate of the parallax plots (bottom plot in each figure) is the number of pixels of parallax. The actual number of pixels is plotted for the lower-left corner of the probe, while the parallax of the upper-right corner is displaced-up by 4 pixels so that both values may be shown on the same plot.

Three quantities are shown in the parallax plots. First, the solid lines that look like uneven stair steps, give the actual parallax on the display screen as a function of z. Second, the dotted lines show the ideal parallax of the corners corresponding to real (i.e., non-integral) values of parallax. And, third, X's mark the intersections of these two curves. The latter corresponds to the nominal perceived location of the virtual probe for all values of z with the same actual parallax. For example, in the parallax plot of figure 4.12B, it can be seen that the parallax is zero for both corners of the probe in the range roughly from $z = -0.25$ cm to $z = +0.25$ cm. Note that the corresponding perceived locations of the virtual target are identified on the histograms by dashed lines.

Although all of the details about how the subjects respond are not yet clear, there are some obvious affects of display pixelation. In particular the subject's resolution depends on the width of the parallax region (step) at the target's position. For example, S1 had much better resolution at $z = 0$ than S9 (compare the size of the error bars at $z = 0$ in figures

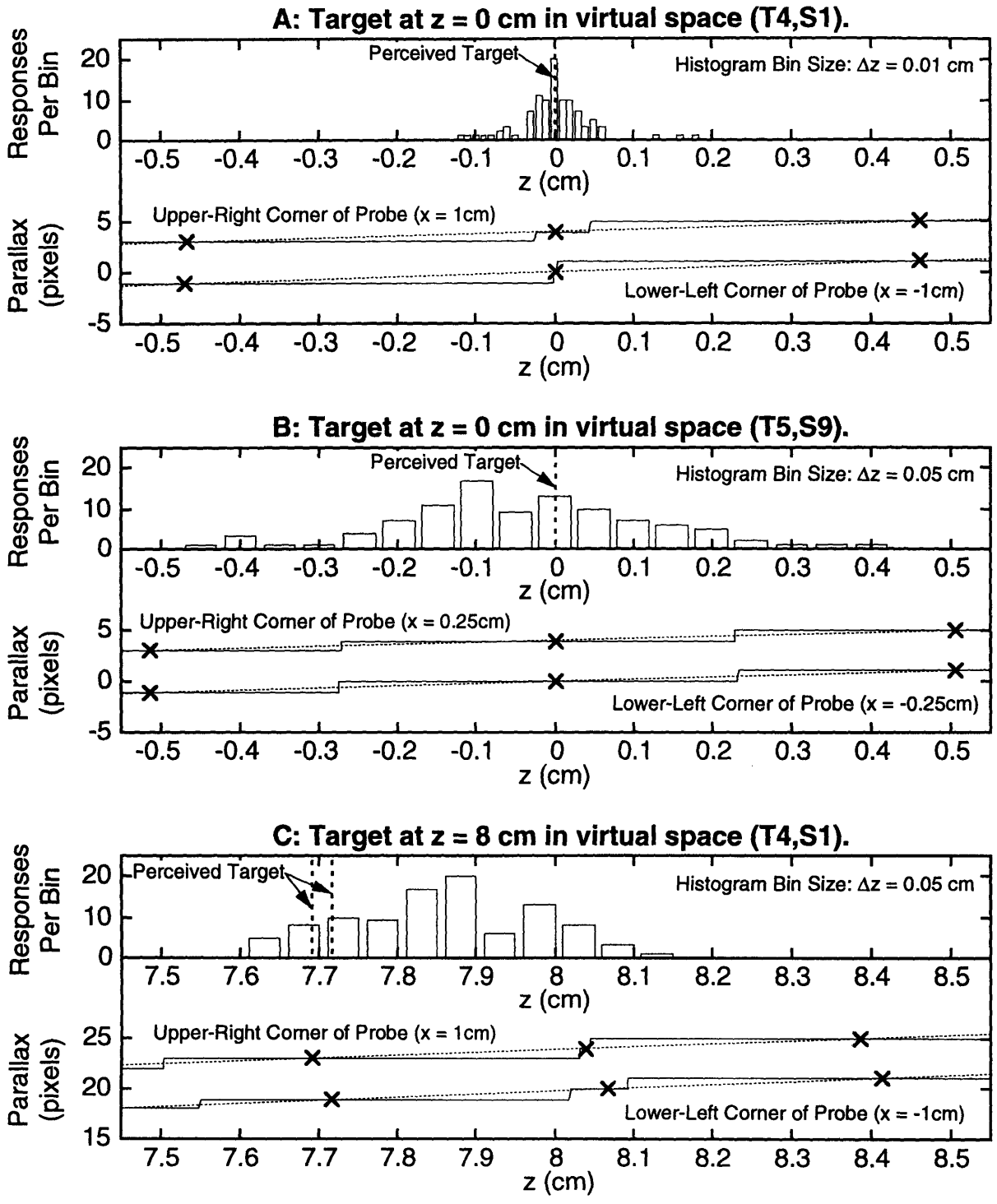


Figure 4.12. Selected response histograms from the virtual-virtual experiment (tests T4 and T5, v2.101) showing the screen parallax of the probe as a function of z (subtract 4 pixels from the indicated parallax of the Lower-Left Corner). Also shown are the discrete "perceived" locations of the probe (x) and the target (---)

4.11A and 4.11B in section 4.3.2), because the parallax region of the target and probe are narrower for S1 (compare figures 4.12A and 4.12B). Indeed, look at the spike in the histogram for S1 (figure 4.12A) corresponding to the very narrow region of parallax for the lower-left corner. Differences occur, of course, because parallax varies in a complex way with the position in virtual space and the effective IPD (see figure 2.3 in section 2.2.2.4).

One may also assume that the shape of the response distribution is related to: 1) manual positioning accuracy, and 2) a response strategy that allows for misalignment of the probe and target by one (or more) pixel on some percentage of the trials. Further, one would expect the mean response to fall at the perceived position of the target. Note that the relative calibration of the PHANToM and the display does not matter here because the subject is comparing two virtual objects.

Interestingly, while these rules do apply in a number of cases, as shown in figures 4.12A and 4.12B for a target at $z = 0$, there are other cases where the subject seems to be using a different response strategy. For example, as shown in figure 4.12C, the mean response of S1 is in front of the "perceived" location of the target that is nominally located at $z = 8$ cm. Specifically, the "perceived" location of the target (indicated by the dashed lines on the histogram) is about $z = 7.7$ cm, while the mean response is about 7.8-7.9 cm.

The cause of this last result is not yet understood. Overall, however, display pixelation does appear to be an important factor in the analysis of computer-graphic stereoscopic displays. In particular, it affects depth resolution.

4.4.1 Depth Resolution

One expects depth resolution to improve when the effective IPD is increased and this is demonstrated for the positioning task of the 3D calibration experiment (see section 4.1.2). In contrast, an earlier study (Rosenberg 1993) concluded that "No statistically significant increase in performance could be correlated to increasing interocular distances greater than 3 cm."[6] The task of that study was to align two virtual "pegs" on a stereoscopic display using a mouse. The pegs were rendered

[6]The interocular distance described in the paper is assumed to be the effective IPD.

and shaded to provide monocular depth cues and the performance measure was the mean position error.[7]

Figure 4.13 presents the reported results of the experiment, showing the observed mean error in the z direction against the effective IPD. Although the exact shape of the curve can only be understood by taking into account the detailed response strategies of the subjects (e.g., how they manipulate the mouse, how they trade off monocular and binocular depth cues, and so forth), the general asymptotic behavior can be explained in terms of display pixelation.

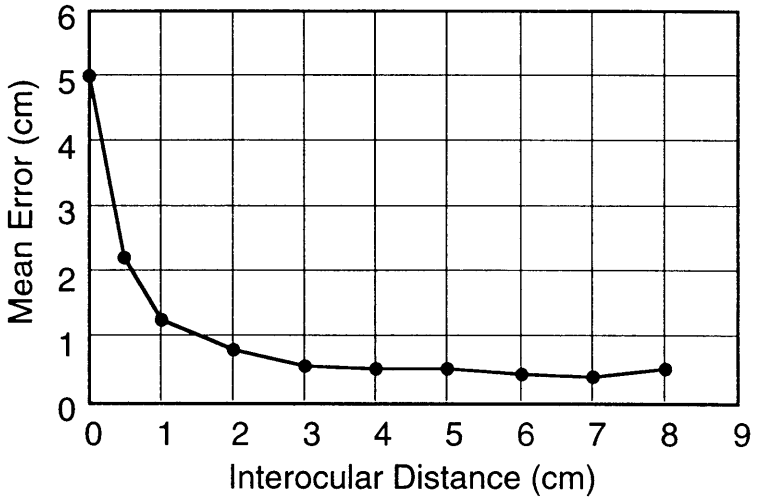


Figure 4.13. Mean Alignment Error (along the z axis) vs. Interocular Distance (Rosenberg 1993).

Specifically, assume that stereopsis will dominate as the effective IPD increases and that the subject's accuracy in controlling the position of the adjustable peg remains constant. Further assume that the mean error will approach some multiple of the minimum depth interval that can be displayed. For example, a subject might choose the exact parallax most of the time, but allow a response error of plus or minus one pixel on some fraction of the trials.

These "error" trials, in turn, correspond to the tails of the response distribution in the virtual space. Assuming that the fraction of responses

---

[7]Unfortunately it is not clear from the paper how the error was calculated, however, error is reported in the same units as the interocular distance (cm) so that the most straight-forward interpretation is that it represents a distance in the virtual space. In other words, the error is defined in the same space where the effective IPD is defined. One also assumes that it is the absolute error, such that this is a measure of resolution rather than a measure of bias.

in the tails remains constant, the shape of the response distribution is scaled by the minimum displayable depth interval.

Note that the "errors" discussed here are associated with the subject's response strategy rather than his or her ability to detect the change in parallax. In particular, it is assumed that the subject can detect a one pixel shift in parallax essentially 100% of the time.

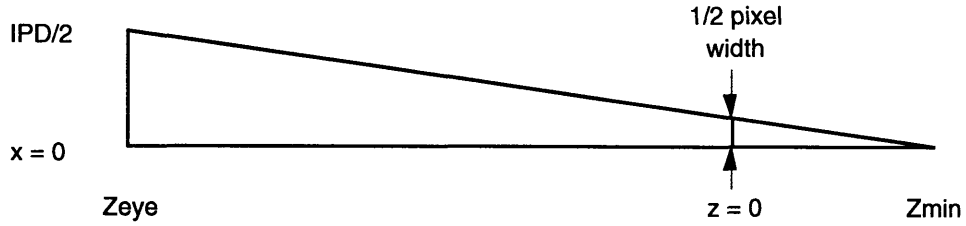


Figure 4.14. Minimum Displayable Depth Interval (Zmin).

Figure 4.14 presents the geometry for a simple case where the minimum depth interval (Z_{min}) is calculated on the z axis with respect to the z = 0 plane. Using this case as a model to understand the general behavior, it can be seen that the minimum depth approaches zero asymptotically as the effective IPD increases. Following is a table of values, calculated with the model, corresponding to the conditions of the Rosenberg experiment assuming a pixel width of 0.04 cm (40cm/1000 pixels):[8]

Table 4.1. Minimum Depth Interval for Various IPDs.

IPD (cm):	0	0.5	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
Z_{min} (cm):	∞	6.40	3.20	1.60	1.07	0.80	0.64	0.53	0.46	0.40

Ultimately, of course, depth resolution (like any type of task performance) depends upon the specific task. For example, depth resolution improves in the 3D calibration experiment, when the effective IPD is doubled, because the task is to position a physical probe rather than a virtual probe. Another depth resolution task of fundamental interest is discrimination.

---

[8]One explanation for how the mean error could be less than the minimum depth in the case of small IPDs (compare the table to figure 4.13), is that the subject might identify the endpoints of the response interval and then position the mouse at the midpoint.

5. DISCRIMINATION EXPERIMENT

This chapter presents a forced-choice experiment in which the subjects' ability to discriminate the depth of two virtual targets was measured. At first look, it does not appear that this would be possible with the apparatus used in the experiments. This is so because the minimum displayable depth interval is many times larger than the just noticeable difference (JND) for most subjects. Specifically, a one pixel parallax on the Virtual Workbench (as configured for the current experiments) corresponds to a disparity of about 180 arc-sec. In contrast, the poorest stereoacuity measured for any of the subjects (using the Randot test) was 70 arc-sec and typical thresholds are on the order of 10 arc-sec (see section 2.1.3.2).

Prior research (Boff and Lincoln 1988b, pp. 1092-1093) has shown that stereoacuity is significantly degraded, however, when the two targets are separated laterally. This fact was used to design a stimulus where depth resolution is limited by the subject's sensory ability rather than the performance of the display. Note that the stimulus was designed in conjunction with the second depth identification experiment (see section 6.3), so that the results of both experiments could be used in the analysis of the preliminary adaptation model in chapter 7.

5.1 Methods

The depth discrimination experiment involved two forced-choice tests, T10 and T11 (v2.102), that were identical except for the trial group order. Two tests were used so that the group order could be counter-balanced among the subjects. The subjects were alternately assigned to one test or the other when they were scheduled to take the test.

5.1.1 Subjects

Nine MIT students, 5 males and 4 females, served as subjects in the experiment. Five of the subjects (S1, S8, S22, S25, and S27) performed test T10 and the other four subjects (S12, S21, S23, and S26) performed T11. All of the subjects, except S8, performed only one test.

The subjects ranged in age from 19 to 41 years; IPDs ranged from 6.10 to 6.90 cm; and visual acuity (corrected in 3 cases) ranged from 20/13 to 20/40 as measured with a standard Snellen eye chart. Stereoacuity was measured using a Randot test and found to be 40 arc-sec or better in all cases. Six of the subjects were right-handed, one was left-handed, and

two did not consider themselves to be strictly either right or left-handed (see appendix B).

S8 was asked to repeat her test about 5 weeks later because the results of her first test are somewhat anomalous. In particular, of all the subjects in the experiment, S8 had the worst resolution in the discrimination test, but the best resolution in the second identification test (see section 6.3). Given the fact that both sets of results are used in the quantitative analysis of chapter 7, and the fact that the relative performance of the other subjects was pretty much the same in both cases, the experimenter decided that it would be worthwhile for S8 to repeat the test.

Both the experimenter and S8 also remembered that she was particularly tired on one of the prior tests (although they were not certain it was T10 rather than one of the identification tests). In any event, her performance improved dramatically in the second test, with the overall frequency of correct responses going from 65% to over 85%. The latter is the third highest score among the subjects.

5.1.2 Apparatus and Stimuli

The experiment was performed on the Virtual Workbench, configured as described in section 3.2. The stimulus consisted of two virtual rectangles, or targets, 1.05 cm wide x 4.17 cm high. The two empty red rectangles, centered approximately at $y = 0$ and (respectively) at $x = \pm 10$ cm, were rendered with jitter as described in section 3.2.4.5.[1] In contrast to earlier tests, however, the screen coordinates were calculated in advance by the CreateTest application. The RunTest application then drew the rectangles at the pre-calculated coordinates. This was done so that only physically meaningful stimuli were presented to the subjects.

The essential idea is that the corners of the targets are determined according to the usual algorithm (see sections 3.2.4.3 to 3.2.4.5), except that the z coordinates in virtual space are specially selected to correspond to integral (pixel) values of parallax. For example, figure 5.1 presents a plot of screen parallax against z for points in virtual space at

[1]Additional uncertainty is also added to the position of the new stimulus, when the initial projection of the rectangle has an odd number of pixels in either the horizontal or the vertical direction. Specifically, when this occurs, and before the AddJitter function of section 3.2.4.5 is called, the center point of the target image is shifted either plus or minus 1/2 pixel on the screen, in the given direction, with a 50% probability. Nothing happens when there are an even number of pixels.

x = 10 cm and y = 0 cm. Note that the figure, which is based on an effective IPD of 6.34 cm, is similar to the parallax plots in figure 4.12. Specifically, three quantities are shown: 1) the actual parallax on the display screen (solid line), 2) the ideal parallax (i.e., the non-integral parallax shown by the dotted line), and 3) the intersections of the two curves (marked by X's). The intersections correspond to the displayable depths (i.e., values of z) an observer would nominally perceive, assuming no adaptation.[2] And it is these values that are selected for the stimuli in this experiment.

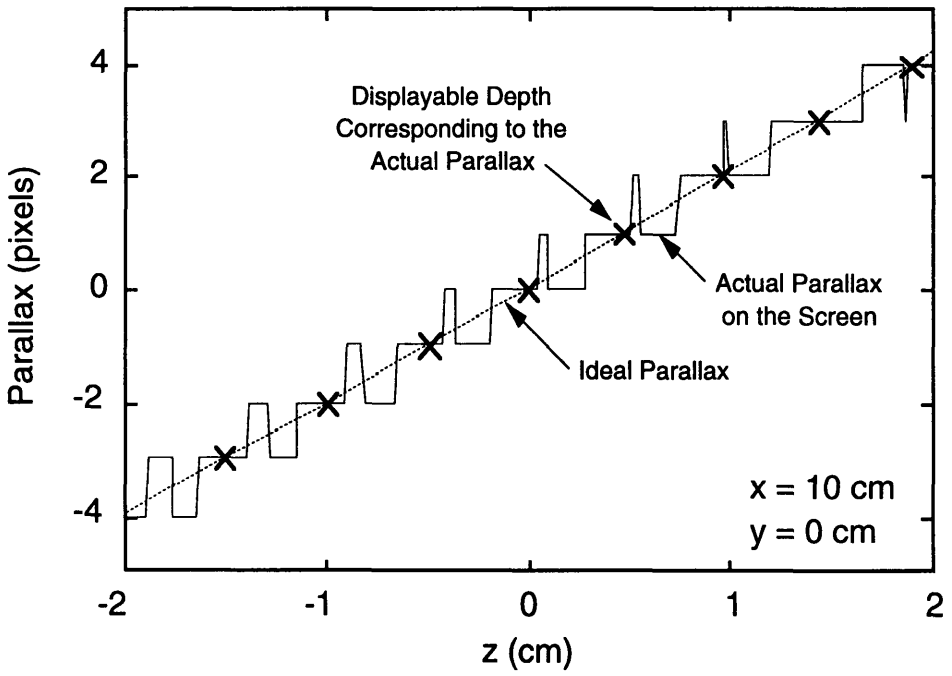


Figure 5.1. Plot of parallax against z for a point in virtual space showing the discrete set of depths that can be displayed.

The result of the calculations is a discrete set of displayed rectangles that appear to be more or less uniformly spaced in the z direction, with the spacing increasing slightly as the distance from the subject increases. Further, the targets appear to be approximately centered at the desired x and y coordinates with the desired width and height.

There are, of course, small discrepancies between the ideal positions and the nominal perceived positions of the rectangles. Specifically, display pixelation and screen curvature cause the targets to be seen at slightly different locations than if they were real (i.e., physical) objects

[2]Note that one assumes that there is no adaptation in the discrimination experiment because the subjects do not receive any feedback about the transformation.

viewed directly; or if viewed through a telestereoscope in the case of altered IPD. Any small discrepancies in the x and y directions, however, including width and height, are effectively masked by jitter. Further, although in theory there is a slight difference in perceived depth resulting from the horizontal separation of the two sides of the rectangle (both displayed with the same parallax), the experimenter was not able to detect this difference. Indeed, the difference is at least an order of magnitude smaller than the depth associated with a one pixel change in the parallax (e.g., see the parallax plot for the target at 8 cm in figure 4.12). The effect would also be masked somewhat by jitter in the xy plane.

5.1.2.1 Depth Number Units: A significant difference between the stimuli in this experiment and stimuli displayed in the more direct manner, as in the calibration experiments, is that different sets of virtual space coordinates are used for subjects with different IPDs. Correspondingly, the subjects will perceive the targets at different z coordinates in the experimental space. This does not appear to be a problem, however, because the stimuli are consistent from test to test for the same subject. Thus, any difference in performance between subjects that is caused by differences in the stimuli may simply be considered part of the normal variation between individuals. More importantly, however, the principal physical stimulus in the experiment (i.e., the screen parallax) is the same for all subjects.

In light of the preceding discussion, it seems clear that the depth should be specified in the experiment in terms of the parallax rather than the virtual space coordinates. But, then, what does it mean to change the effective IPD? Fortunately, there is a very simple relationship between the parallax and the value of X_{ipd}. Specifically, they are proportional (as illustrated in figure 5.2). Thus, a virtual rectangle with a parallax of one pixel at x1 IPD has a parallax of two pixels at x2 IPD.

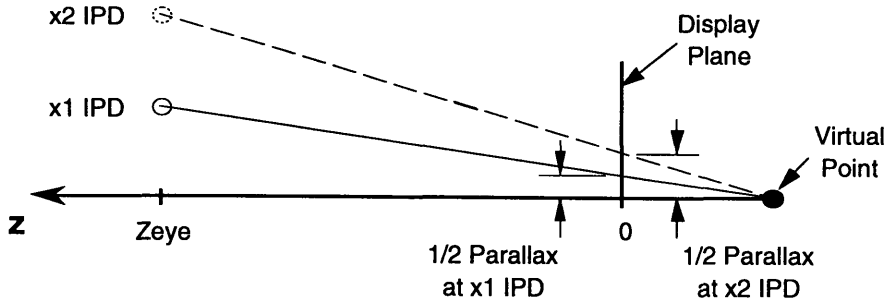


Figure 5.2. Relationship between X_{ipd} and Parallax.

This leads to the definition of depth number units (dnu) that is used here and in chapters 6 and 7 (see section 7.1.1):

N dnu = N pixels of parallax at x1 IPD. (5.1)

Thus a target at a depth of N dnu in virtual space has a parallax of N pixels when $X_{ipd} = 1$ (by definition) and a parallax of 2N pixels when $X_{ipd} = 2$. Also note that z = 0 dnu corresponds to z = 0 cm in the virtual space and that the sign of the depth is the same for both sets of units. In other words, positive virtual depths measured in dnu are in front of the plane of the display (z = 0), like positive virtual depths measured in cm (see figure 3.12 in section 3.2.1.2), and negative virtual depths measured in dnu are behind the plane of the display.

5.1.3 Procedure

The two virtual rectangles of the stimulus are referred to as "Left" (x = -10 cm) and "Right" (x = +10 cm), with one of the two being defined as a reference. Note that the fact that one target is considered to be a reference by the experimenter is unknown to the subjects. The test consists of 8 groups of trials where the value of X_{ipd} and the reference (L = left rectangle, R = right rectangle) alternate as shown in the following table (5.1).

Table 5.1. T10 and T11 (v2.102) Test Groups.

	T10		T11	
Group	X_{ipd}	Ref	X_{ipd}	Ref
1	1	L	2	R
2	2	L	1	R
3	1	R	2	L
4	2	R	1	L
5	1	L	2	R
6	2	L	1	R
7	1	R	2	L
8	2	R	1	L

The groups are otherwise identical except for the random order of the trials. In each trial the reference target is presented at one of three standard depths (N_{ref} = -6, 0, +6 dnu) and the other target is located relative to the reference at one of 4 possible depths (N_{ref} + n, where n = -2, -1, +1, +2 dnu). Ten trials are performed at each condition for a total of 120 trials per group in random order.

In each trial the two targets are presented simultaneously and the subject is supposed to judge which of the two appears closer. If the right rectangle appears to be closer than the left rectangle, the subject clicks the right trackball button. Alternatively, if the left rectangle appears closer he or she clicks the left button. The selected rectangle disappears first, when the button is initially pressed. This is done in order to give the subject feedback about whether he or she selected the intended rectangle. Then, when the button is released, the other rectangle disappears, the subject's response is recorded, and the next trial begins automatically. No feedback about the correct response is provided to the subjects in any of the trials.

Training for the task was minimal. First the experimenter explained the task briefly, answering any questions that the subject might have about the procedure. A demo was available so that the subjects could practice the task prior to the start of the test, but most subjects simply began the test without practice (the experimenter's notes only identify two tests where the demo was used). In any event, the experimenter always observed the subject during the first few trials of the test and asked questions to insure that the procedure was understood.

5.2 Results

Figure 5.3 presents the combined results for all of the subjects on four standard ROC (receiver operating characteristic) plots (Gescheider 1985; Green and Swets 1988). The plots analyze the results in terms of two probabilities, estimated by the frequency of the subjects' responses. The abscissa is the frequency with which subjects respond "Left" given that the right rectangle is closer, p(Llr). The ordinate is the frequency with which subjects respond "Left" given that the left rectangle is closer, p(Lll). In general, the closer a point is to the upper left-hand corner, the better the performance.

Each data point corresponds to the combined results for all of the subjects at a single test condition. In other words, the raw data for all of the subjects is pooled before the probabilities are calculated. One reason for doing this is that some of the subjects (under some conditions) did not make any mistakes. Thus, it would have been impossible to calculate sensitivity (discussed shortly) on an individual basis.

The same data is presented on all four plots using different symbols to identify the different conditions in the experiment. Figures 5.3A and 5.3B

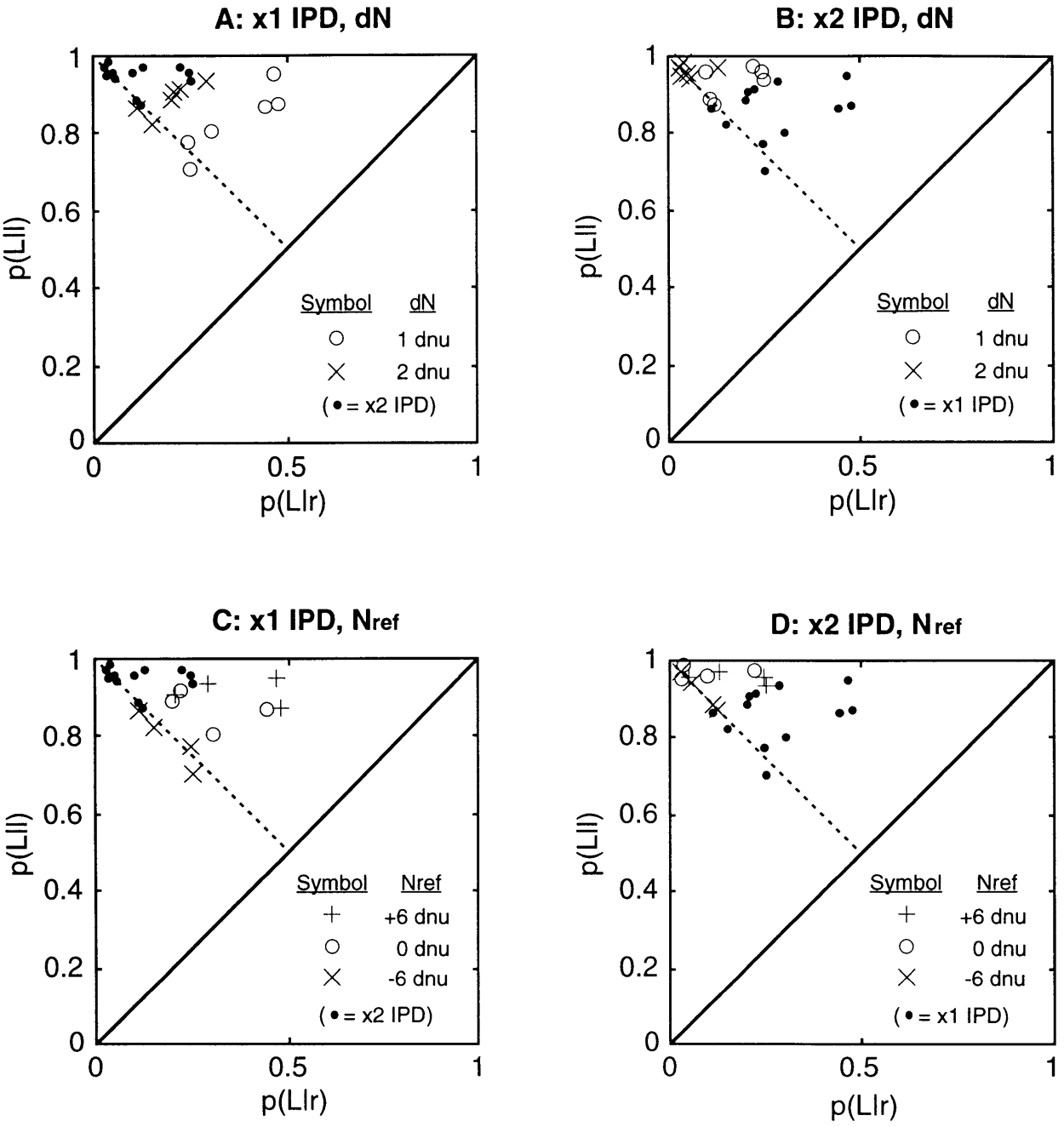


Figure 5.3. ROC plots of the combined data for all of the subjects in the discrimination experiment (tests T10 and T11, v2.102). The four plots show the same data using different symbols to identify the different conditions.

break-down the results by the absolute depth difference between the two rectangles for x1 and x2 IPD trials, respectively. It can be seen in the figures that the subjects are more sensitive to a 2 dnu difference (x) than a 1 dnu difference (o). Discrimination is also clearly better for $X_{ipd} = 2$ than for $X_{ipd} = 1$.

Figures 5.3C and 5.3D break-down the results by the depth of the reference rectangle. These figures show that the subjects are relatively unbiased in their responses when the depth number is -6 (the x symbols), but they become more biased as the targets get closer (the o and + symbols). Specifically, in the latter case the subjects have a bias toward responding "Left." This was unexpected and is not understood.

The dots (•) in the figures show the *other* X_{ipd} data. Specifically, in the x1 IPD plots (figures 5.3A and C), the dots show the x2 IPD data, while in the x2 IPD plots (figures 5.3B and D), the dots show the x1 IPD data.

Figures 5.4 and 5.5 presents the normalized sensitivities (d') calculated according to case V of Thurstone's *law of comparative judgment* (Gescheider 1985, p. 152, equation 6.8). The latter is based on the correct-response frequencies in the experiment, assuming that the distributions on the decision axis are Gaussian with equal variance. Figure 5.4 plots d' against the depth difference of the two targets (in virtual space) and figure 5.5 plots d' against the parallax difference.

5.3 Discussion

It can be seen in figure 5.4 that discrimination in the virtual space was improved when the effective IPD was increased. The reason for this is apparent in figure 5.5. Namely, discrimination improved because sensitivity is directly related to the screen parallax (i.e., the physical stimulus) and the parallax doubles when the effective IPD is doubled.

The solid lines in figures 5.4 and 5.5 show the best-fit line to the *parallax* data excluding the results for ±4 pixels. The dotted line in figure 5.5 has 2 times the best-fit slope as required to fit the x2 IPD data in that figure. Note that it is reasonable to exclude the ±4 pixel data because it is likely that the corresponding values of d' are underestimated. This is true because estimates of large values of d' are unreliable.

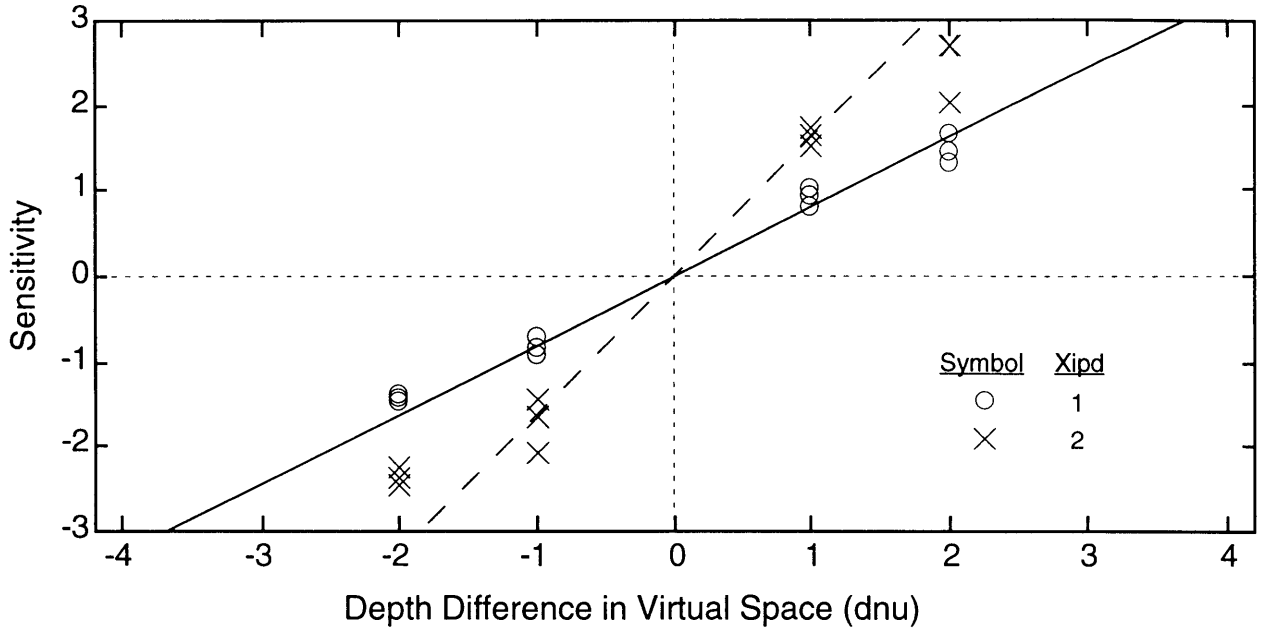


Figure 5.4. Sensitivity (d') plotted against the depth difference of the targets in virtual space (tests T10 and T11, v2.102).

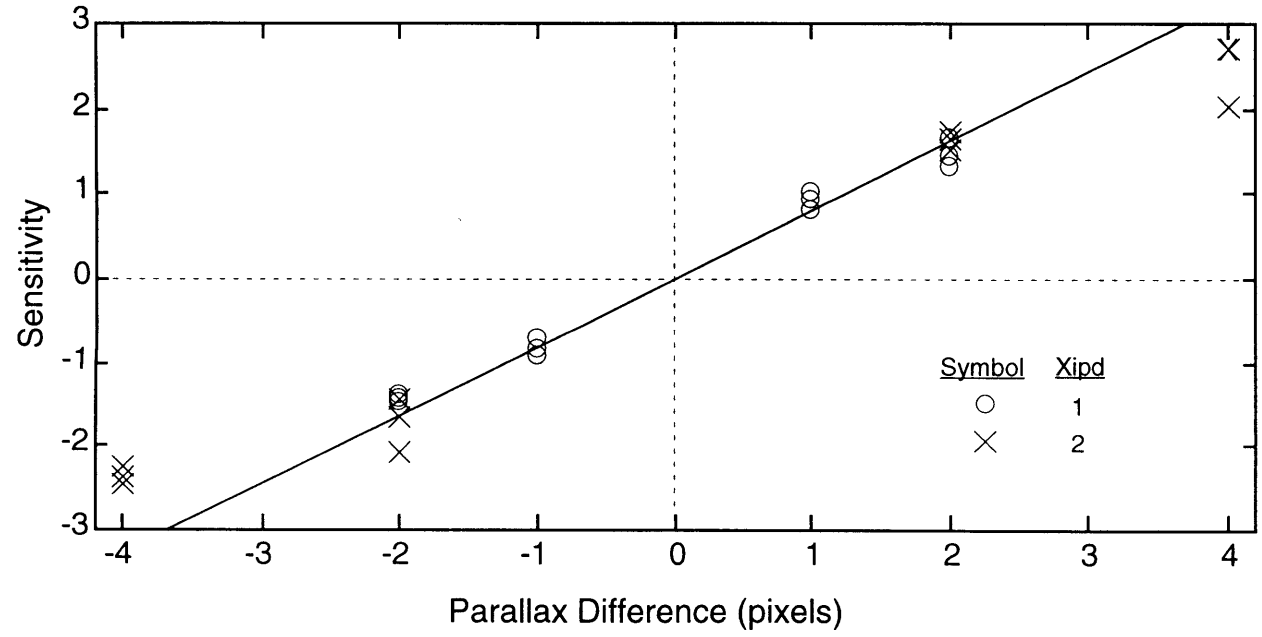


Figure 5.5. Sensitivity (d') plotted against the difference in screen parallax of the targets (tests T10 and T11, v2.102).

5.4 Psychophysical Scale

A psychophysical scale was constructed based on the observed values of d'. Specifically, a function (α) is constructed that relates the physical stimulus in the experiment (P = pixels of parallax) to the mean of the corresponding distribution on the decision axis (see section 2.4, equation 2.3). Note that d' is the *normalized* sensitivity such that the corresponding unit on the decision axis is the standard deviation of the distribution. The latter quantity is further assumed to be solely related to sensation noise (i.e., it is equal to β, as defined in section 2.4.2).

The first step in the construction of the scale is to arbitrarily equate zero on the decision axis with zero parallax. This is clearly the simplest choice and, given that a linear model is ultimately derived here, it seems justified. Specifically, a different choice for zero would only involve an additive constant that would cancel out in the final analysis (see section 7.1).

The next step in the analysis is to note that the observed values of d' are related to parallax *differences*. Thus, in the case of $N_{ref} = 0$, one obtains the values of $\alpha(P)$ directly from the observed d', while in the case of $N_{ref} = \pm 6$ dnu, one can only estimate the slope of the function. Also note that 6 dnu corresponds to P = 6 pixels in the x1 IPD groups and P = 12 pixels in the x2 IPD groups (by definition, see equation 5.1).

When the ±4 pixel data is excluded as discussed in section 5.3, the estimated slope is roughly constant. This, in turn suggests a linear model such that, using the best-fit slope shown in figure 5.5, one obtains the following function:

$$\alpha(P) = 0.8\ P. \qquad (5.2)$$

Figure 5.6 plots the psychophysical scale (equation 5.2) together with the data from the experiment (excluding ±4 pixel data as previously noted). The difference data from the experiment (i.e., the observed values of d') are plotted on the absolute scale for $N_{ref} \neq 0$ by using the model (i.e., equation 5.2) to determine the locations of the reference points. For example, in the case of $N_{ref} = +6$ dnu and $X_{ipd} = 2$ (i.e., $P_{ref} = +12$ pixels), one plots the data point for n = +1 dnu (i.e., $\Delta P = +2$ pixels) at $P_n = +14$ pixels and $\alpha(P_n) = 0.8\ P_{ref} + d'_n$.

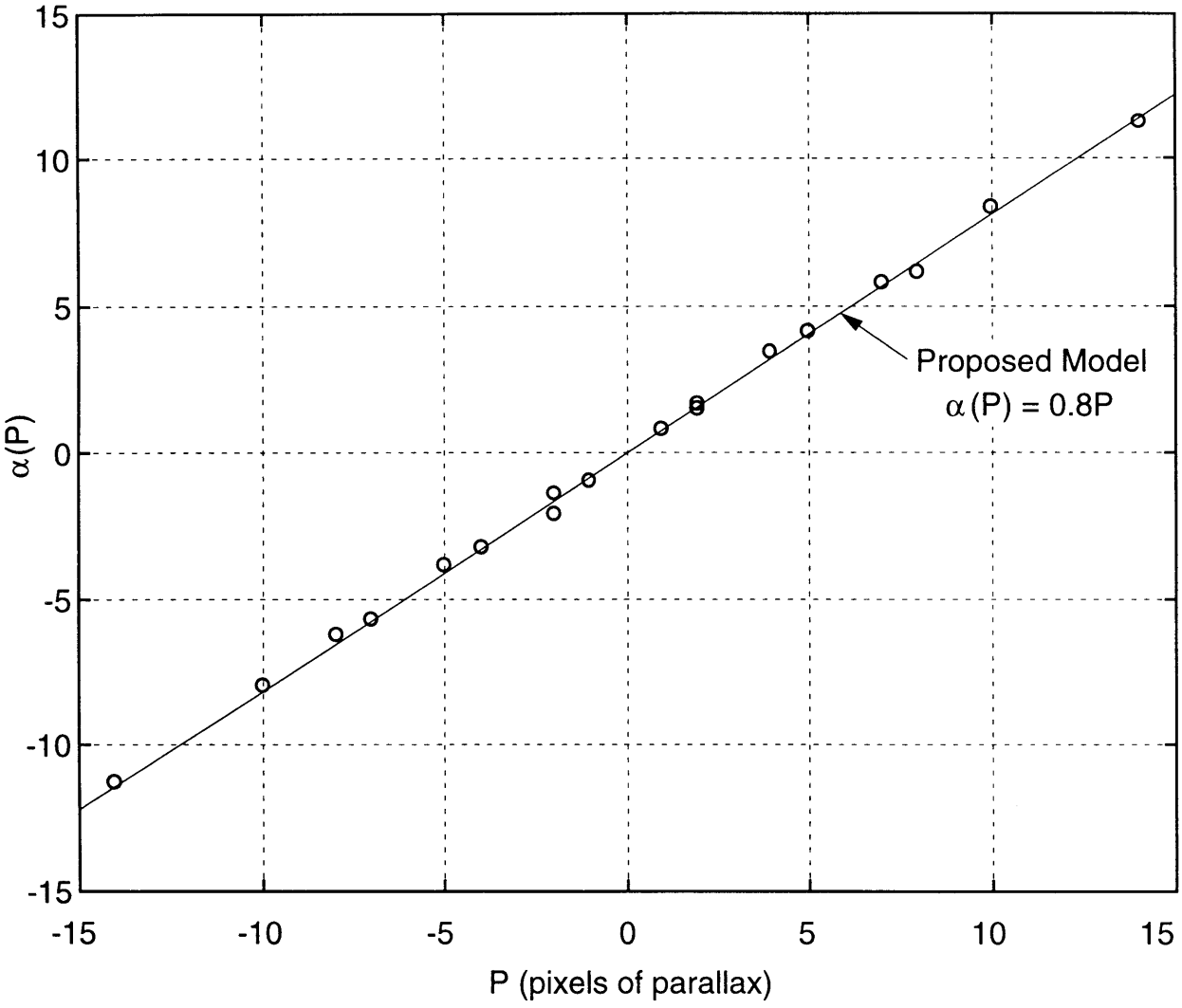


Figure 5.6. Plot of the proposed psychophysical scale, α (P), against screen parallax, P, showing the relative sensitivity data that was determined in the discrimination experiment (tests T10 and T11, v2.102).

In conclusion, the reader is cautioned that the function is not determined quite as well as it might appear in figure 5.6. For example, if the data points were all relative to $P = 0$ the standard error of the slope would be ± 0.006. Unfortunately, the data only gives estimates of the slope with an accuracy that is better illustrated in figure 5.5. Specifically, the standard deviation of the slope estimates is 0.1, suggesting that the single significant digit in equation 5.2 is appropriate. Nevertheless, unless one believes that the true function exhibits some unexpected high-order behavior,[3] piecing the segments together as in figure 5.6 cannot be far off.

[3]Nominally, one would expect a first-order power law after Stevens (Gescheider 1985, p. 217).

6. IDENTIFICATION EXPERIMENTS

This chapter presents two adaptation experiments in which the subjects performed an absolute identification task. In both experiments, the task was to identify a visual depth stimulus presented at random out of a standard set of depths. Each subject was asked to perform the task repeatedly and the average performance was observed. Then, over the course of the experiment, the effective IPD was changed in order to see how the performance varied.

One expects that increasing the effective IPD will initially cause the subject to overestimate the depth of the stimulus; assuming that he or she was able to identify the depth correctly on average before the change. In other words, one expects a shift in the subject's mean response error, or *bias*, due to the sensory transformation. Given feedback about the correct response, however, the subject should adapt to the transformation over time.[1] Specifically, the bias should tend to zero as the subject learns to identify the stimulus depth in the new context. In addition to shifts in bias, one also expects the subject's *resolution* (i.e., his or her ability to discriminate one stimulus from another) to improve when the effective IPD is increased. Further discussion about the expected shifts in bias and resolution is presented with other background material in chapter 2 (see sections 2.3-2.5).

Experiment 1: The results of the first experiment were initially somewhat surprising in that, while the expected shifts in bias were observed, there was no change in resolution[2] when the effective IPD was doubled. This result is apparently related to the fact that performance in an absolute identification task depends on both sensory and cognitive abilities. Part of the explanation may be given by the preliminary theory of intensity resolution, introduced in section 2.4 and discussed further in connection with the first experiment in section 6.2.3. Display pixelation is also a significant confounding factor in the experiment.

Experiment 2: The second experiment corrected problems with the first experiment by using stimuli similar to the ones used in the

---

[1]Note that the experiments in chapters 4 and 5 are not adaptation experiments, even when the subjects are exposed to an altered IPD, because the subjects do not receive feedback about the transformation. Specifically, the subjects are simply presented with a stimulus such that they do not perceive any conflict with other sensory input and there is nothing to which they must adapt.

[2]Actually the resolution did increase for the smallest stimulus depths. See sections 6.2.2 and 6.2.3.

discrimination experiment (see section 5.1.2). In particular, the stimuli involved significant lateral separation. The expected results were observed in the second experiment as discussed in sections 6.3.2 and 6.3.3. A quantitative analysis of the results that provides evidence for and against the preliminary adaptation model is presented in chapter 7.

6.1 Relevance of the Experiments

The identification experiments are clearly relevant to the objective of providing further evidence either for or against the preliminary adaptation model. The model, originally developed in connection with auditory localization, is applied to a new sensory modality. Further discussion is required, however, because the experimental paradigm is not entirely consistent with existing definitions of adaptation.

For example, Welch's definition of adaptation requires a "registered discrepancy between or within sensory modalities" (Welch 1978, p. 8; Welch 1986, p. 24-3). No such discrepancy exists in the present case.[3] Instead, the subjects modify their behavior on the basis of feedback about the correct response. As a result one might argue that true perceptual adaptation does not take place because the subjects are consciously relearning the stimulus depths. Indeed, Welch states that "conscious correction of one's errors during exposure to the distortion does not qualify as true adaptation" (1986, p. 24-12); a view that is shared by Rock (1966, p. 11).

Similarly, Held states that adaptation has "often been interpreted as resulting from recognition of error and subsequent correction. We doubted this interpretation because of our conviction that a single mechanism underlies both adaptation to rearrangement in the adult and the development of the young. An error-correcting process could hardly explain the original acquisition of coordination." (Held 1972, pp. 373-374)

While all of this may be true (the discrepancy requirement of Welch's definition and Held's doubts not withstanding), it does not rule out the possibility that conscious error-correction may alter perception in some cases. Indeed, in the present experiments, as well as those performed by Shinn-Cunningham (see section 2.3.4), a negative aftereffect is observed, a result that is frequently cited as evidence of true adaptation (Rock

---

[3]This is not entirely true because the relationship between accommodation and convergence is modified (see section 2.2.2.1), however, this is only considered to be a minor confounding factor in the experiment.

1966, p. 11; Welch 1986, p. 24-15). Further, the observed changes in bias are accounted for, in the preliminary adaptation model, by shifts in the subject's criteria.

Assuming that the underlying theory is correct, such shifts represent a fundamental change in perception. Specifically, for a particular criterion in the decision model (see section 2.4), a given sensory magnitude leads to one percept if it falls on one side of the criterion and it leads to another percept if it falls on the other side. If the criterion moves on the decision axis, then the observer's perception is altered.

Perceptual modification of this kind would appear to fit the requirements of "true" adaptation. First of all, the criteria in the decision model represent an inherently modifiable aspect of perception. It is still not clear how this is possible. In other words, we are not any closer to answering the following question posed by Held (1972, p. 372): "How can the visual control of spatially coordinated action be stable under normal circumstances and yet sufficiently modifiable to allow recovery from transformation?" Nevertheless it is so, and this mechanism is presumably present at all stages of development.

Finally, even though the shifts are the result of conscious error-correction, the subjects do not appear to control the positions of the criteria at will. In other words, the objection that the subjects are simply giving the response that they know to be correct, while perceiving something different, may not apply. This point could be put to the test by challenging subjects to try to retain two sets of criteria in an experiment similar to the present ones (see section 8.3.2).

6.1.1 Applied Adaptation Studies

Although, as discussed immediately above, it is believed that the effects observed in the present experiments represent "true" adaptation, other approaches are no less important. For example, an engineer working on a specific application, where a particular transformation is imposed, might choose a practical approach where the task itself is the primary concern.[4] In this case the initial effect, the reduction of effect, and the subsequent aftereffect may be viewed as independent measures of adaptation.[5] The questions are simply: 1) what are the initial and final

---

[4]The work of Rolland, et al. (1995) is a good example.

[5]Note that the "reduction of effect" may simply be viewed as the final level of performance which need not depend on the initial effect.

levels of performance when the transformation is imposed, 2) how long does it take to adapt, 3) what is the magnitude of the aftereffect, and 4) how long does it take to re-adapt to normal when the transformation is removed? Any type of transformation is fair game and the engineer need not care about the underlying mechanisms.

This is a perfectly reasonable and expedient approach to the problem, particularly since the underlying mechanisms are so poorly understood. The traditional approach is also important (even for engineers) because the results may be applied generally. This *study seeks to meet the objectives of both approaches.*

6.2 Experiment 1

In the first adaptation experiment a virtual target was presented randomly at any one of ten standard positions on the z axis in virtual space, equally spaced from z = 0.4 cm to z = 4.0 cm in 0.4 cm increments. Subjects judged the depth of the target relative to a fixed reference frame and responded by adjusting a graphical scale using a trackball. The depth stimuli were numbered from 1 to 10, in order of increasing z, and the computer only allowed the subject to make one of 10 corresponding responses. Both the reference frame and the scale were presented on the display at z = 0 (i.e., with zero parallax).

The test involved pre- and postexposure measures without feedback, at both x1 and x2 IPD, according to the standard adaptation paradigm (see section 2.3.1). The exposure (x2 IPD) involved the same task, except with feedback. Specifically, in each exposure trial a reference mark was presented on the scale to indicate the correct response after the subject made his or her response. Similar training trials were also performed, except at x1 IPD, before and after the pre- and postexposure measures, respectively.

6.2.1 Methods

The experiment involved a single test: T80 (v2.101). The test was repeated by some subjects two or three times (see section 6.2.1.1), although no subject did more than one test per day.

6.2.1.1 Subjects: Six MIT students, 3 males and 3 females, served as subjects in the experiment. Subject 1 performed the test 3 times; subjects 8, 12, 13, and 15 each performed the test twice; and subject 16 performed the test once for a total of 12 experimental runs. The subjects

ranged in age from 19 to 40 years; IPDs ranged from 5.50 to 6.77 cm; and visual acuity (corrected in 3 cases) ranged from 20/13 to 20/25 as measured with a standard Snellen eye chart. Stereoacuity was measured using a Randot test and found to be 70 arc-sec or better in all cases. Four of the subjects were right-handed while the other two subjects did not consider themselves to be strictly either right or left-handed (see appendix B).

6.2.1.2 Apparatus and Stimuli: The experiment was performed on the Virtual Workbench, configured as described in section 3.2, except that the green piece of paper was not placed under the mirror during the first 2 or 3 tests.[6] Figure 6.1 presents a full-scale plane view of the experimental stimulus from the direction of the subject. The virtual target was a solid red square, 2.0 cm on a side, parallel to the xy plane, and centered on the z axis as shown in the figure.

The target was rendered with jitter as described in section 3.2.4.5, except during the first three tests. In those first tests the target was rendered in a similar manner, except using the screen coordinates of the upper-right and lower-left corners directly as described in section 4.2.1.2. The rendering technique was modified for the subsequent tests because the first algorithm sometimes makes the target appear tipped in the z direction.

The reference frame consisted of 3 concentric, empty, red squares parallel with the xy plane and centered at the origin. The reference frame was identical to the frame used in the calibration experiments except for the color (see section 4.2.1.2). A red frame was used in the present experiments in order to eliminate any possible effects related to color stereopsis (Boff and Lincoln 1988b, p. 1120).

Figure 6.1 also shows the graphical scale used to input the subject's response and to provide feedback. The scale was located under the depth stimulus near the bottom of the screen. A white pointer, controlled by the subject via the trackball, indicated the current scale setting. Rolling the ball forward (nominally away from the subject) caused the pointer to

---

[6]Specifically, the first three tests were S1's first two tests and S12's first test. The green piece of paper was not used in S1 and S12's first tests, and (perhaps) S1's second test. Unfortunately, the experimenter's notes in the lab book are not clear in regards to S1's second test. All subsequent tests did use the green piece of paper, however, assuming there was no mistake. In any event it is unlikely that this had a serious affect on the results. Recall that the purpose of the paper is to prevent the subjects from seeing objects under the mirror that are made faintly visible by stray light (see section 3.2.1.2).

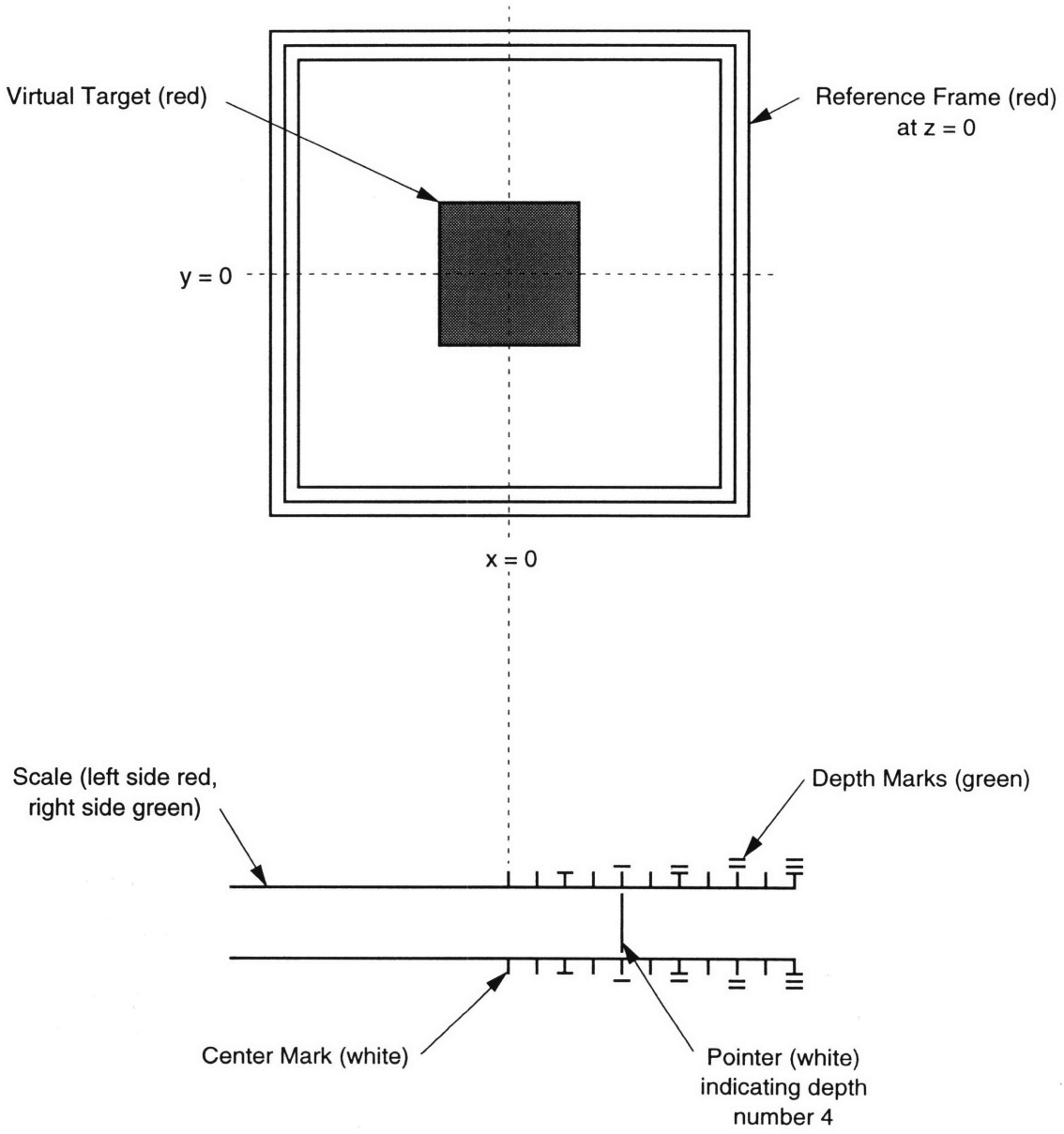


Figure 6.1. Full-scale plane view of the virtual target, reference frame, and scale used in the first identification experiment (T80, v2.101).

move left, while rolling the ball backward (nominally toward the subject) moved the pointer to the right.

The idea is that the horizontal scale represents positions on the z axis. Specifically, $z = 0$ is in the middle of the scale (indicated by two white marks). Recall that $z = 0$ corresponds to the location of the virtual image of the display screen (see figure 3.12). The right side of the scale (rendered in green) corresponds to $z > 0$ (i.e., toward the subject). Similarly, the left side of the scale (rendered in red) corresponds to $z < 0$ (i.e., behind the plane of the display).[7]

Green marks on the right side of the scale indicate the 10 stimulus depths. The marks are located to scale in the virtual space. In other words, the first mark is at $x = 0.4$ cm (corresponding to the z coordinate of the first stimulus depth), the second mark is at $x = 0.8$ cm (corresponding to the z coordinate of the second depth), and so forth out to the tenth mark at 4.0 cm. A system of horizontal hash marks was used to help the subjects recognize depths in the middle of the range.[8]

The scale was rendered so that it always appeared exactly the same in all of the trials. In other words, the scale always had the same size, the spacing of the marks did not change, and it was rendered with zero parallax so that the appearance was not affected by changes of X_{ipd}. Also, the initial setting of the pointer was always on the first stimulus depth.

The subject input his or her response by positioning the pointer at the desired depth mark. He or she then pressed both trackball buttons at the same time to input the current setting. The latter action ended the trial and the next trial began automatically when the buttons were released. Note that the pointer jumped between the discrete depth settings, when the scale was adjusted, in such a way that it was only possible to input one of the 10 allowed responses.

In feedback trials a green reference mark appeared on the scale when the buttons were pressed and it remained until they were released. The reference mark, which was identical to the pointer except for the color,

---

[7]The fact that the scale has a left side is an artifact of the software development. It was anticipated that both positive and negative z values might need to be indicated on the scale. The present experiment only involves positive depths however.

[8]One would think that a simpler approach would have been to use numbers. Unfortunately, it proved to be a difficult matter to generate text upside-down with the PC (as needed on the Virtual Workbench where the computer screen is viewed in a mirror). Fortunately, the given marking system proved to be satisfactory.

indicated the correct response. If the subject's response was correct, the effect was that the pointer appeared to change its color from white to green.

6.2.1.3 Procedure: The test consisted of 7 groups of 100 trials, where each of the stimulus depths was presented 10 times in random order. Aside from the random order of the trials, the groups differed only in the value of X_{ipd}, and whether or not feedback was given, as shown in table 6.1.

Table 6.1. T80 (v2.101) Test Groups.

Group	X_{ipd}	Feedback
1. Initial training	1.0	yes
2. Preexposure, nominal	1.0	no
3. Preexposure, altered	2.0	no
4. Exposure (adaptation training)	2.0	yes
5. Postexposure, altered	2.0	no
6. Postexposure, nominal	1.0	no
7. Final training	1.0	yes

The subjects were given little training prior to group 1 of their first test. A demonstration of the scale was given in at least one case to help the subject understand how to do the task. It is possible that this was done in some other cases as well (the experimenter's notes are unclear on the matter), however, the fact that the demonstration is not mentioned suggests that in most cases the task was simply described and then the test was started.

Although the experimenter's exact instructions varied from test to test, the notes in the lab book (and the experimenter's memory) suggest that the principal message was as follows:

1) The subjects were instructed to learn the task based on the feedback provided by the reference mark. The experimenter indicated that the procedure was first to train the subjects (with feedback), *then* test them (without feedback), and then repeat the training *and testing in* subsequent groups of trials.

2) Subjects were asked to give their "best estimate" of the stimulus depth in the trials without feedback. At least one subject (S16) interpreted this instruction to mean that she should respond "the way it

looked," as she indicated at the end of the test, rather than on what she believed the "correct response" to be, based on the training. It is not known whether this was simply her interpretation or whether it is based on some additional instruction given to her (and possibly other subjects) by the experimenter. It is assumed that most subjects made no such distinction. They simply tried to respond as they were trained.

3) In some cases the experimenter suggested to the subject that the purpose of the test was to see how well they could learn the task under different conditions (without specifying the conditions). Otherwise the experimenter said nothing about the purpose of the experiment beyond the general introduction (see section 3.1.1) given on a previous day.

4) During the test, the experimenter made no reference to any change in the experimental conditions from group to group unless specifically asked about a change by the subject. In the latter case the subject was told that the apparatus distorts the space, but his or her task was still the same, i.e., to learn the stimulus depths in accordance with the feedback. Typically the experimenter introduced each new group of trials by saying something like, "This group is just like the last one." The subject was also told whether he or she would receive feedback in the group.

6.2.2 Results

Figure 6.2 presents the combined results for all of the tests. The top graph in the figure plots the mean estimated depth number against the stimulus depth number in order to show the shifts in bias. In other words, the measure of bias is mean error.[9] The bottom graph plots the observed resolution against the stimulus depth. Specifically, for any two adjacent depths the resolution is calculated as follows:[10]

---

[9]To the extent that the standard deviations of the responses (σ_i, see equation 6.1) are constant for all of the stimulus depths (i = 1,..,10), this measure of bias is equivalent to Shinn-Cunningham's "raw" normalized bias estimate (1994, p.83) only multiplied by the constant factor σ (the common value of σ_i). Note that the raw normalized bias is equal to the difference between the mean response and the correct response divided by the standard deviation (σ_i) for the given stimulus (i). Indeed, when mean error (mean estimated depth - stimulus depth) is plotted against stimulus depth in the present experiment, there is little difference whether the error is normalized or not (i.e., σ_i is more or less constant and approximately equal to one). This is also true of the bias results in the second identification experiment discussed in section 6.3.

[10]The quantity d_{ij}^* (equation 6.1) is one estimate of the quantity d_{ij}' (equation 2.1). The other quantities in equation 6.1 (μ and σ) are calculated from the experimental measurements and, hence, are estimates of the corresponding quantities in equation 2.1. Also note that d_{ij}^* is equivalent to the "raw" resolution estimate used by Shinn-Cunningham (1994, p.83).

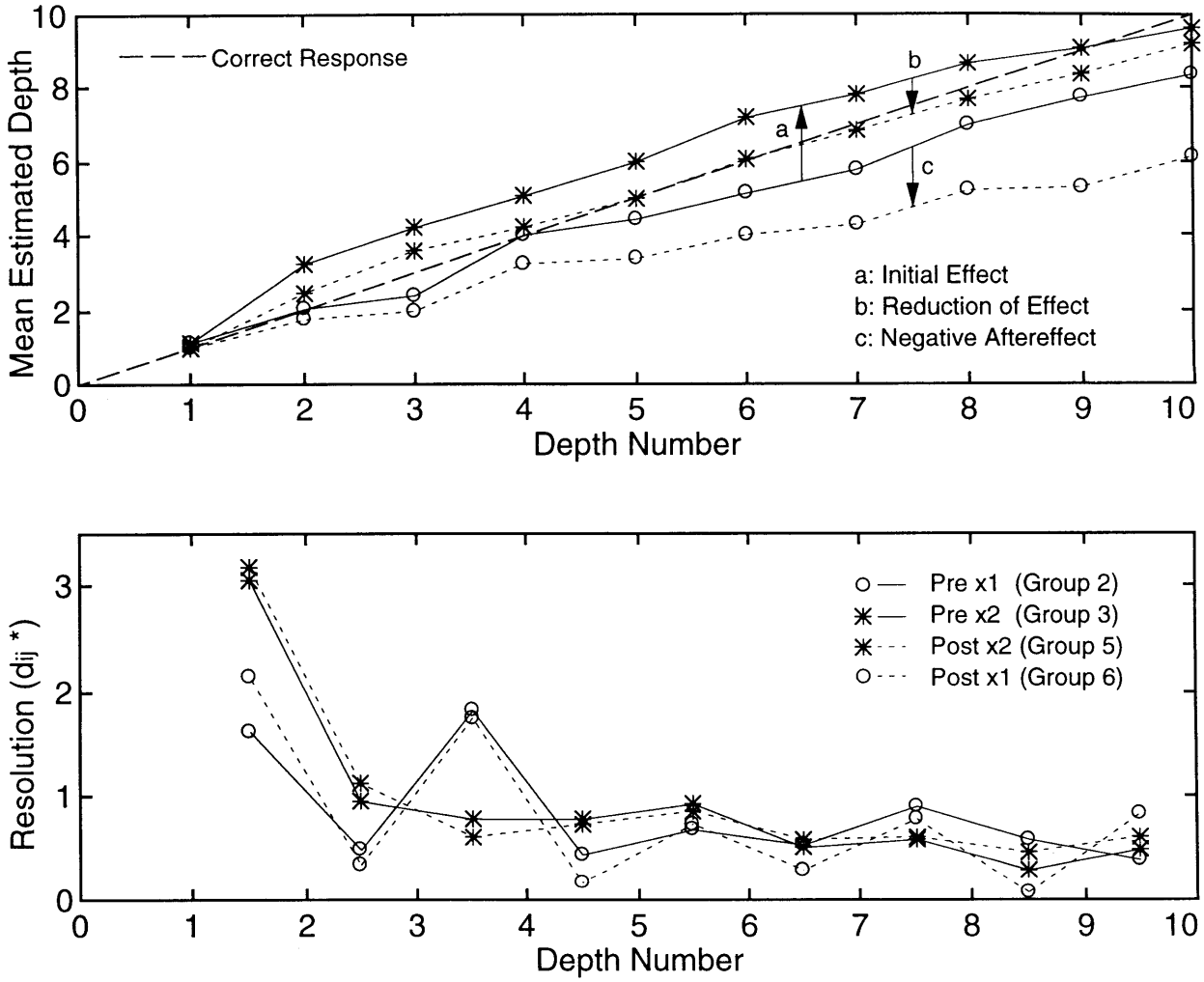


Figure 6.2. Mean estimated depth (showing shifts in bias) and resolution against depth number for all subjects in the first identification experiment (test T80, v2.101).

$$d_{ij}^* = (\mu_j - \mu_i) / (1/2\ (\sigma_i + \sigma_j)) \qquad (6.1)$$

where:

d_{ij}^* = observed resolution between adjacent depth stimuli $i < j$
μ_c = mean estimated depth when the stimulus depth number is $c = i, j$, combining the data for all subjects.
σ_c = standard deviation of subjects' estimated depth responses when the stimulus depth number is $c = i, j$.

Note that d_{ij}^* is plotted midway between the i^{th} and j^{th} stimulus depths in the figure. Also note that both the bias and resolution measures combine the results for all of the subjects, treating them like a single observer. In other words, all of the data was combined together and then the means and standard deviations for the whole group were calculated for each stimulus depth. This was done, rather than calculating the results for each individual and then averaging across the subjects, because it was not possible to calculate d_{ij}^* individually in some cases due to the small number of trials in the experiment. Specifically, some subjects gave the same response every time they were presented with a given stimulus. This led to infinite values of d_{ij}^* in some cases (when both σ_i and σ_j were found to be zero).

Overall, the subjects' mean depth estimates show the expected shifts in bias (see section 2.3.1): a) initial effect, b) reduction of effect, and c) negative aftereffect. Interestingly, however, there is almost no shift for the first stimulus depth. Also the initial effect and the aftereffect are not symmetrical and the magnitude of the shifts are less than one might expect, based purely on the magnitude of the change in the physical stimulus. Subjects also show an inherent bias toward underestimating the largest depths, as shown by the results of the preexposure x1 IPD group.

The other main result is that, in contrast to the experimenter's initial expectation, there is little difference in resolution between the various conditions (pre- and postexposure, x1 and x2 IPD). The expected increase in resolution is observed for the depths nearest zero however. There is also an unexpected peak in the resolution at x1 IPD between stimulus depths 3 and 4. Possible explanations of these results are presented in the following discussion.

6.2.3 Discussion

The main question that needs to be answered is: why is there no general change in resolution when the effective IPD is changed by a factor of two? Furthermore, the answer to this question must also take account of the change in resolution that does occur in the case of the smallest depths. Other questions are: 1) why is there a peak in the resolution at x1 IPD, 2) why is there no shift in bias for the first stimulus depth, 3) why are the initial effect and the aftereffect asymmetrical and smaller than expected, and 4) why do the subjects underestimate the larger depths in the preexposure measurement at x1 IPD?

The discussion begins with a look at the confounding role of display pixelation in the experiment. The analysis of display pixelation answers questions 1 and 2 above, and also provides insights relevant to the main question about resolution. The other questions are discussed in turn.

6.2.3.1 Display Pixelation: Display pixelation is a significant issue in the present experiment because the virtual stimuli are spaced very near the depth resolution of the display. Indeed, the spacing is too close (in virtual space) to distinguish some pairs of adjacent depths on the display. Figure 6.3 shows the screen parallax of the target as a function of the stimulus depth for each of the subjects.

It can be seen in the figure that, while on average the parallax changes slightly less than 1 pixel between each of the stimulus depths at x1 IPD (or twice that rate at x2 IPD), the parallax may change by 0, 1, 2, or 3 pixels between any given pair of adjacent stimulus depths. Also, the step changes in the parallax occur at different values of z, depending on the subject's IPD, so that different subjects are presented with different physical stimuli in the experiment.

When the physical stimulus (i.e., the parallax) is taken into account, the causes of two experimental results become immediately apparent, while a third result is made more puzzling.

First, the fact, that there is no shift in bias for the first stimulus depth, is immediately explained by the fact that the screen parallax is always the same in that case. Specifically, the target presented to every subject for the first stimulus depth has a parallax of 1 pixel for both x1 and x2 IPD.

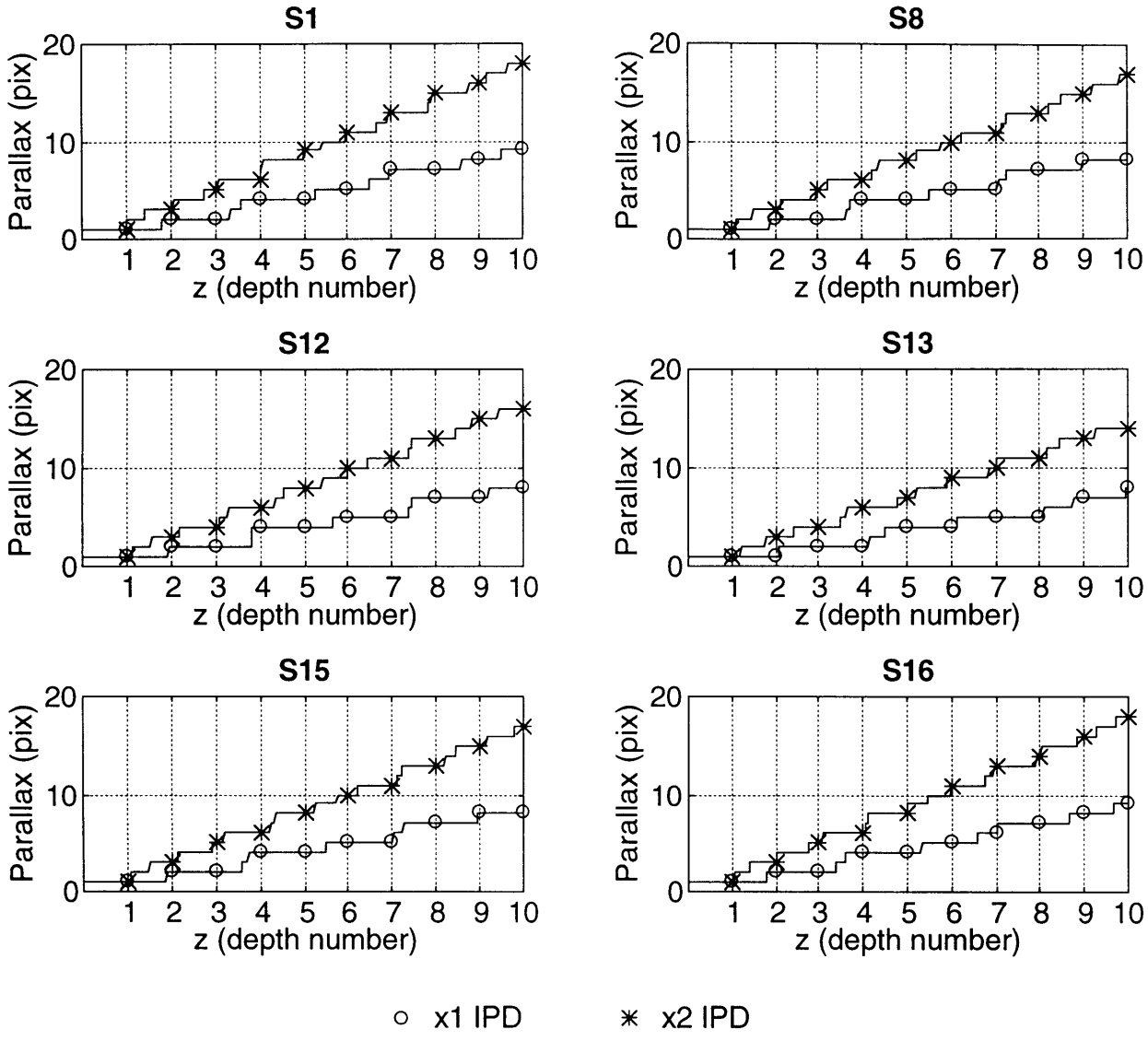


Figure 6.3. Screen parallax of the virtual target against z in the virtual space for each subject in the first identification experiment (test T80, v2.101). The symbols "o" and "*" correspond to the 10 depths in the experiment viewed at x1 and x2 IPD, respectively. Note that the stimuli are uniformly spaced on the z axis, 0.4 cm apart.

Second, it can be seen in the figure that every subject, except S13, was presented with the same pattern of physical stimuli at x1 IPD in the depth range between 2 and 5. In particular, stimulus depth pairs (2,3) and (4,5) are indistinguishable while there is a 2 pixel change between depths 3 and 4. This artifact caused the peak in resolution that is observed between the third and forth stimulus depths. Note that different patterns in the way the parallax changes for the different subjects tends to smooth out the results for the other stimulus depths.

Independent of display pixelation, however, one would expect the effect of pixelation on resolution to be reduced by averaging over several stimulus depths. Figure 6.4 presents a plot where the *smoothed resolution* at stimulus depth j is given by:

$$d_j^* = (d_{ij}^* + d_{jk}^*)/2, \qquad (6.2)$$

where i, j, and k are consecutive stimulus depths.

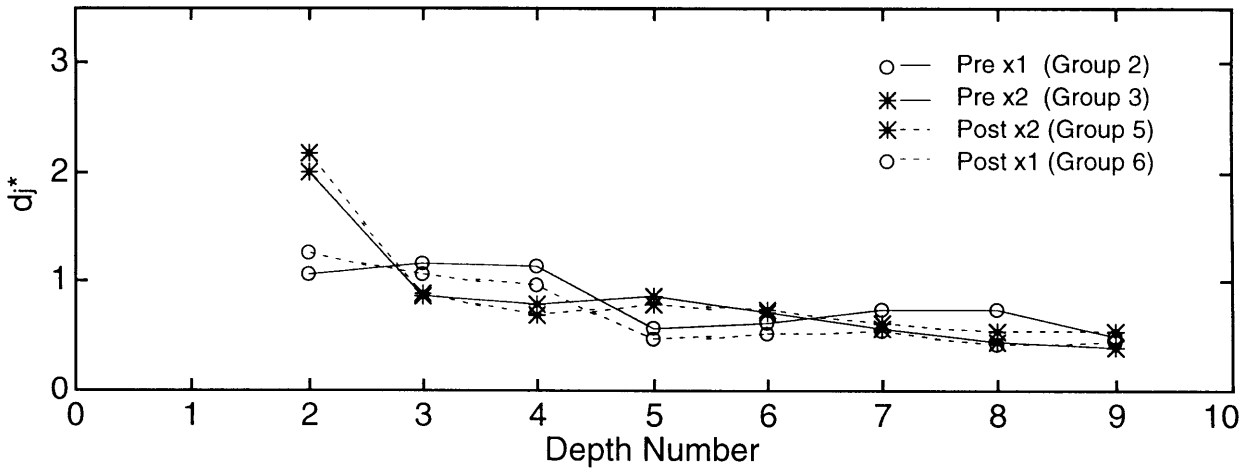


Figure 6.4. Smoothed Resolution (test T80, v2.101)

It can be seen in this figure that, except for smoothing-out the peak between stimulus depths 3 and 4, the results are essentially unchanged. Unfortunately, this only makes the main question about resolution more apparent. How can the resolution be the same (on average) for most of the stimulus depths at x1 and x2 IPD when the parallax (on average) is different by a factor of two?

6.2.3.2 Why the resolution did not change overall: The reason that the expected overall change in resolution was not observed in the experiment, may be partially explained by the preliminary theory of

intensity resolution (see section 2.4). Other factors must be taken into account as well, however, and the final explanation is not entirely satisfying. In any event, the theory predicts the general character of the performance illustrated in figure 6.5. In this figure, the predicted total sensitivity for an identification experiment is plotted against the effective range according to the following equation:

$$\Delta' = \frac{R}{\left(\beta^2 + G^2R^2\right)^{1/2}} \qquad (6.3)$$

where:

Δ' = total sensitivity = d_{ij}' (where $I_j = I_{max}$ and $I_i = I_{min}$)
I_{max} = maximum stimulus intensity
I_{min} = minimum stimulus intensity

R = effective range = $\alpha(I_{max}) - \alpha(I_{min})$
$\alpha(I)$ = psychophysical scale function equal to the mean of the distribution on the decision axis corresponding to the physical stimulus intensity, I

β^2 = variance of the sensation noise
G^2R^2 = variance of the memory noise
G = constant.

The equation may be derived from equations 2.1-2.5 (chapter 2) and the definition of total sensitivity. Total sensitivity is the sum of the sensitivities d_{ij}' (i.e., the normalized resolutions given by equation 2.1) between every pair of adjacent stimuli. Equivalently, because d' is additive, the total sensitivity is the normalized resolution between the minimum and maximum intensities of the stimulus set.

In addition to equation 6.3, the figure is also based on the results of an identification experiment where the total sensitivity was measured in a series of tests with different ranges (Braida and Durlach 1972, experiment 4, see figure 4d). Specifically, the parameters β and G are based on the values given by Braida and Durlach, showing a good fit to their data.[11] The abscissa is also transformed from R into units of just

[11]Note that some of the quantities are defined differently in equation 6.3 than by Braida and Durlach (1972). See the footnote in section 2.4.2 in connection with equation 2.5. Also note

noticeable difference (JND) by assuming a constant JND of 1/2 dB at all intensity levels in the experiment.[12]

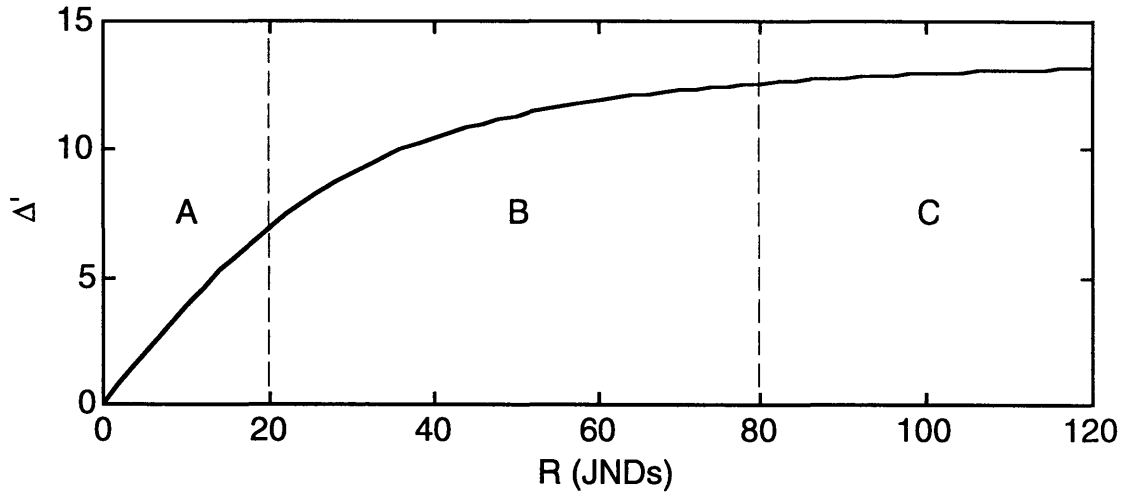


Figure 6.5. Relationship between total sensitivity (Δ') and effective range (R) predicted by equation 6.3.

Referring to figure 6.5, it can be seen that the predicted curve may be divided into three regions: A, B, and C. In region A, where the stimuli span less than 20 JND, the total sensitivity is roughly proportional to the range. Then, as the range increases in region B, the total sensitivity begins to approach an asymptote, such that it is approximately constant in region C (R > 80 JND). Another way to think of this is that resolution is limited by sensation noise in region A (i.e., β dominates performance) whereas memory noise dominates in region C.

The total sensitivity is, of course, proportional to the average resolution (and the number of stimuli is the same in all cases) so that the behavior shown in the figure is directly related to the results of the experiment (i.e., d*). Also note that even though the curve in the figure is based on a hearing experiment, one expects similar results for vision because the model is concerned with central mental processes. A simple explanation for the main resolution result, then, is that the subjects are operating in region C. In other words, resolution is limited by memory rather than sensation noise. As a result, doubling the disparity between the stimulus depths does not improve resolution because memory noise doubles as well and memory noise is the limiting factor.

---

that, in the case of hearing, $R = \log(I_{max}/I_{min})$ and decibels are the appropriate unit for the JND.

[12]The number of JNDs in the total range (R) is equal to 10R(dB)/JND(dB).

Testing the Hypothesis: At best, the above reasoning is valid only to the extent that the range of disparities in the experiment contains a large number of disparity JNDs. Using the stereoacuity results from the preliminary tests for the estimates of the JNDs for individual subjects (see the right-most column of the table in appendix B) and averaging over all six subjects in the experiment, the mean number of JNDs in the range is 50 for the x1 IPD case.[13] This estimate falls in the middle of region B, a result that is not entirely satisfying; however, even at this level, doubling the effective IPD (and, hence, the stimulus range) only increases the total sensitivity by about 15%.

It is also possible that the value of the JND used in the calculation is overestimated. This may be true because the stereoacuity was not measured very accurately in the preliminary tests, which were only intended to check that the subjects had roughly normal vision (see section 3.1.2). Indeed, both of the subjects who were retested at a later time showed improved scores.[14] For example, the observed stereoacuity for S8 went from 70 to 30 arc-sec. Also the best score on the Randot test that was used to measure the stereoacuity is 20 arc-sec. Hence, the three subjects (half of the subjects in the experiment) who scored this well, might have done better in a more accurate test. In any event the hypothesis that resolution is limited primarily by memory noise in this experiment does appear to account for the essential feature of the results.

Other Factors: It must also be pointed-out, however, that this is not the whole story. In particular, the preliminary theory of intensity resolution does not predict any change in the effective range (R) during a test (see section 2.5 where the preliminary adaptation model is introduced). While the preliminary adaptation model does predict such changes, one does not expect the effective range to change in the absence of feedback. Also, the "R" in the numerator of equation 6.3 becomes a different term from the "R" in the denominator. Specifically, in the preliminary adaptation model, the numerator of equation 6.3 is purely a function of the physical stimulus (see equation 2.10), whereas the

---

[13]Note that the disparity range is different for different subjects. The range for each subject was calculated by multiplying the stimulus range (in pixels of parallax) shown in figure 6.3 times 180 arc-sec/pixel, where the latter value corresponds to a pixel width of 0.0525 cm at a distance of 60 cm. The number of JNDs in the range was then calculated for each subject individually (using the Randot results in appendix B) and, finally, the individual results were averaged across the six subjects.

[14]The only subjects to be retested were S8 and S12, who were retested 10 months later prior to the second identification experiment.

effective range ("R" in the denominator) is a function of the state of adaptation and the size of the response set (see equation 2.9).[15]

Thus, independent of the initial hypothesis, the preexposure x2 IPD test group (3) should show better resolution than the preexposure x1 IPD group (2). This is expected, according to the preliminary adaptation model, because the stimulus range (numerator of equation 6.3) increases in the shift from x1 to x2 IPD, while the effective range ("R" in the denominator) remains constant in the absence of feedback. Recall that there is no explicit feedback in either preexposure group.

Ultimately, however, the mechanism of the initial hypothesis should apply after the exposure when adaptation is complete. The model thus predicts that the resolution in the postexposure x2 IPD group (5) should be equal to the resolution in the preexposure x1 IPD group (2), as observed.

The fact that no change in resolution was observed in the preexposure groups (2 and 3) when the effective IPD was doubled, may be related to the fact (in addition to the initial hypothesis) that the magnitude of the shifts in bias are less than expected. The latter result is explained by a hypothesis that adaptation may occur in the absence of direct (i.e., explicit correct-answer) feedback as discussed in section 6.2.3.5. Nevertheless, one would still expect some resolution change because the adaptation in this case is incomplete.

Another factor may be that additional cognitive noise is introduced by display pixelation. For example, in some cases the subjects received different feedback for the same physical stimulus (see section 6.2.3.1). If such noise were to dominate sensory and memory noise, it might cause the observed result (i.e., that there was no overall change in resolution for any of the groups). Unfortunately this is pure speculation. Further research is obviously required to settle the issue.

6.2.3.3 Why the resolution did change near zero: The reason that the resolution is greater at x2 IPD for the smallest depths, might be explained by the fact that the reference frame provides a perceptual anchor. This result is consistent with the perceptual anchor model of context-coding (Braida et al. 1984). The model predicts that resolution should be better near perceptual anchors like the reference frame, because memory noise is relatively less at these locations. The effect may also be

---

[15]Also see the related discussion of pre- to postexposure resolution shifts in section 6.3.3.1.

enhanced because the physical stimulus corresponding to the first depth is always the same (see figure 6.3).

6.2.3.4 Finite Response Set Bias: The reason that the initial effect is less for the largest stimulus depths is simply due to the fact that it is not possible for the subjects to give a response larger than 10. This also explains the asymmetry between the initial effect and the aftereffect (see figure 6.2). Specifically, while the subject cannot give a response larger than the largest stimulus depth when the effective IPD is increased (initial effect), there is no such limitation when the IPD is reduced in the postexposure groups (negative aftereffect). In the latter case, at least for the largest depths, the subjects are free to estimate that the depth is half the actual value.

6.2.3.5 Indirect Feedback:[16] Analysis of the response limitations imposed by the finite size of the response set also gives some insight into why the overall magnitudes of the bias shifts are less than expected. Specifically, when the subject is presented with a stimulus that is larger than the largest magnitude that he or she expects, the subject will tend to scale-up his or her expectations. Indeed, the resulting rescaling of criterion may also change the subject's perception of the smaller depths as well.

One hypothesis is that people are able to linearly rescale all of their criteria at once, based on information about the location of only a few criteria. For example, a subject might *stretch* the entire set of criteria when a sensation is too large, uniformly increasing the spacing between each adjacent pair, rather than simply moving the maximum criterion. This would be consistent with Shinn-Cunningham's results, in which the subjects adapted almost completely to a linear approximation of the transformation (see section 2.3.4). Also note that this assumption would not necessarily rule-out the possibility that an observer could adapt fully to a non-linear transformation, although adapting to the linear approximation could be much faster and might dominate any higher-order process.

In any event, it seems clear that the subject can receive *feedback* about the sensory transformation in the experiment, even when feedback is not directly (i.e., explicitly) given. Hence, adaptation may occur in the preexposure x2 IPD group, causing the *initial effect* to be less than

---

[16]Indirect feedback was unexpected and not understood until after the second identification experiment, even though it is discussed here in connection with the first experiment.

expected. A similar process does not work at the small end of the range in this experiment, because the first stimulus depth provides a minimum bound: nothing ever appears smaller than stimulus number 1. Other indirect feedback mechanism are possible however.

Mechanisms: In general, feedback mechanisms that operate in the absence of direct feedback about the correct response, depend on the subject's memory of prior stimuli and responses. Some possible mechanisms are as follows. First, as already suggested, the observer might judge a sensation to be outside the normal range. This mechanism would require the existence of two additional criteria at the extreme ends of the range. Note that the perceptual anchors, proposed in the perceptual anchor model of context coding (Braida et al. 1984, p.723), could serve this purpose.

Alternatively, the observer might judge relative distances on the decision axis, independent of the criteria. Specifically, a sensation might be judged to be more than one stimulus depth number larger than a sensation previously judged as the maximum (or similarly at the minimum end of the range). Another possibility is that the observer might be able to judge the variance of the sensations. Thus, a new cluster of sensations might be perceived in between two clusters, previously judged to be adjacent stimuli. The latter would require more trials than the former mechanisms, suggesting that subjects who receive only indirect feedback would adapt more quickly to an increase in the stimulus range.

Still another mechanism for adaptation is that the observer might simply guess differently based on the *a priori* probability. For example, as commented-on by some subjects in the experiment, toward the end of the postexposure x1 IPD group (6) one becomes aware that no large depths were perceived in the group. As a result the observer might start to guess that the largest perceived sensations actually correspond to the maximum depth (or similarly at the minimum end of the range). One would also expect observers to fill-in any gaps in the range in a similar manner. Any judgments based on *a priori* probabilities are inherently slow compared to the mechanisms discussed earlier however.

Apparently, one hundred trials is not enough for a subject to fully adapt in the experiment via even the fastest indirect feedback mechanisms. Over a very large number of trials, however, one would expect the subject to learn to identify the stimulus depths correctly in this way (or at least as well as is possible via direct feedback). As discussed below, this effect is

illustrated in figure 6.6, where the subjects' depth estimates are presented for each of the four pre- and postexposure groups.

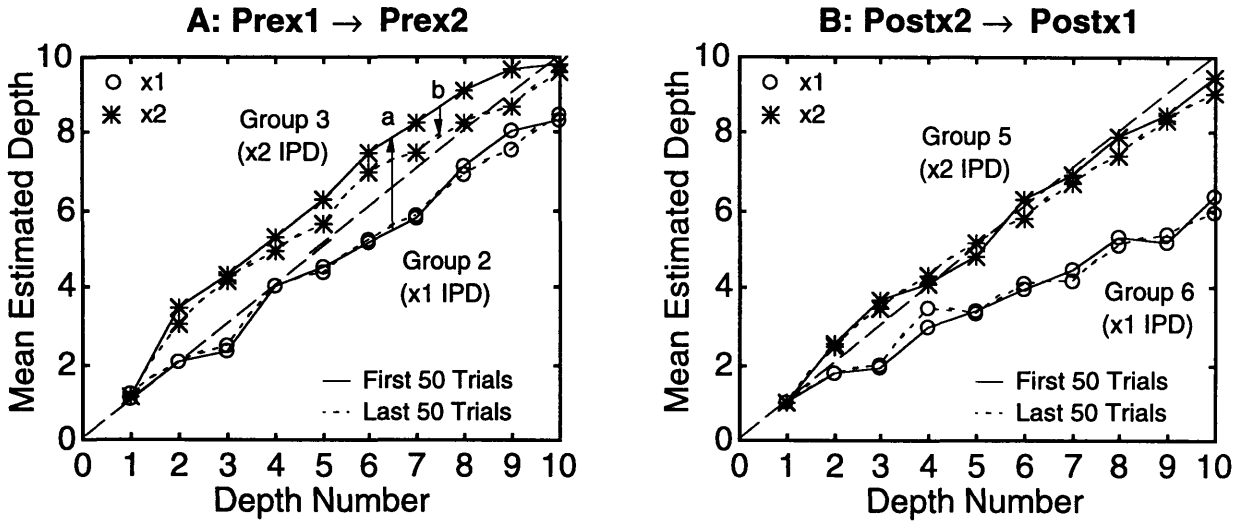


Figure 6.6. Evidence of adaptation in the absence of explicit feedback in the first identification experiment (T80, v2.101).

Direct Evidence: Figure 6.6A shows the preexposure transition from x1 to x2 IPD in the experiment (groups 2 and 3). Similarly, figure 6.6B shows the postexposure transition from x2 to x1 IPD (groups 5 and 6). There is, of course, no direct (i.e., explicit correct-answer) feedback for any of the groups shown in the figure.

The two plots are similar to the previous plot (top of figure 6.2) that shows the bias results for the experiment. In all of the plots, the symbols "o" and "*" indicate x1 and x2 IPD data, respectively. However, the meanings of the solid and the dashed lines are different in the present plots (figure 6.6) from their respective meanings in the previous plot (figure 6.2). In figure 6.6, the solid lines indicate the first 50 trials of the given group and the dashed lines indicate the last 50 trials.

The two estimates (solid and dashed lines) are the same for the initial groups (2 or 5) in both plots as one would expect. In other words there is no apparent shift in bias during the initial groups. This makes sense because the subjects were just trained with the same value of effective IPD in the prior group (1 or 4), such that they are reasonably well adapted at the start of the "initial" group shown in the plots. Thus, even if there was explicit feedback, one would not expect much difference in performance between the beginning and the end of the initial groups.

In contrast, a difference is apparent for the preexposure x2 IPD group (i.e., the final group in figure 6.6A). Specifically, the mean depth estimates are larger during the first fifty trials of the group. In other words, the initial effect (a) is larger. This, in turn, corresponds to a reduction of effect (b) between the beginning and the end of the group, showing direct evidence of adaptation in the absence of explicit feedback.

Interestingly, a similar result is not observed in the case of the negative aftereffect as shown in figure 6.6B. The latter is consistent with the earlier suggestion that indirect feedback is more effective when the stimulus range is increased. Sections 7.2.3.1, 7.2.3.5, and 7.3.2.2 discuss indirect feedback further in connection with the quantitative results of the second identification experiment.

6.2.3.6 Central Tendency: One reason subjects might underestimate the larger depths in the preexposure measurement at x1 IPD is the general tendency of human observers to give intermediate responses (Helson 1964, pp. 94-102). One possible source of the central tendency in the present experiment is related to the finite size of the response set. Specifically, given that subjects can never overestimate the largest stimulus depth, any error results in a mean response that underestimates the correct response. The same argument applies to responses at the other end of the stimulus range. As a result, at best, the subject can only approach the correct-response value at the ends. These ideas are discussed further in connection with the second identification experiment. See the analysis presented in chapter 7 (sections 7.1.3.2, 7.2.3.3, 7.3.2.2) and appendix E.

6.3 Experiment 2

The second identification experiment was designed to eliminate most of the confounding factors of the first experiment. This was accomplished with stimuli that are essentially identical to the ones used in the depth discrimination experiment discussed in chapter 5. Indeed, the two experiments were designed as a pair so that the results of both might be used in the analysis of the preliminary adaptation model in chapter 7.

As in the depth discrimination experiment, the present stimuli consisted of two virtual targets separated horizontally on the display in order to make it more difficult for the subjects to discriminate the depth. The only difference from the stimuli used in the discrimination experiment was that the left target included marks inside the rectangle so that it could be used as a scale to input responses. The left target also served as the reference (always with zero parallax), while the right

target was presented randomly at one of 13 standard depths: from -6 to +6 dnu in 1 dnu increments. Recall that depth number units (dnu) are equal to the number of pixels of parallax when the target is viewed at x1 IPD (see section 5.1.2.1).

The principal advantage of the new design is that sensation noise becomes a significant factor in resolution. Thus, one expects that resolution will improve when the effective IPD is increased, unlike the results observed in the first experiment where memory noise appears to dominate such that there is no change in resolution. Another advantage is that display pixelation effects are nearly eliminated because the targets have integral parallax.

The present experiment also followed a slightly different adaptation paradigm from the first. In the new paradigm, both $X_{ipd} = 1$ and $X_{ipd} = 2$ were used as the nominal transformation (with the other value serving as the altered transformation in each case).[17] After the first experiment, it was realized that this is possible because the identification task is arbitrary and independent of other spatial judgments. Also, no adverse effects were observed in any of the prior experiments when subjects had to stop for a break in the middle of a test. A related change in the paradigm is that the final training (group 7 in the first experiment, see section 6.2.1.3) was eliminated. The advantages of these changes are, respectively, to provide additional information and to shorten the tests.

6.3.1 Methods

The experiment involved two tests, T40 and T42 (v2.102), that were identical except for the value of X_{ipd}. Two tests were used so that the condition (nominal = x1 IPD, altered = x2 IPD) could be counter-balanced with the condition (nominal = x2 IPD, altered = x1 IPD). The tests were performed on different days with the subjects being alternately assigned to take one test or the other on the first day. Each subject then performed the remaining test on a subsequent day.

6.3.1.1 Subjects: Nine MIT students, 5 males and 4 females, served as subjects in the experiment. All of the subjects performed both tests once.

---

[17]Typically, one would think of the "nominal" transformation as the null transformation corresponding to the subject's normal interaction with the world (thus "nominal" would typically correspond to $X_{ipd} = 1$) and the "altered" transformation would be some type of change from the nominal (e.g., $X_{ipd} = 2$). In the second identification experiment, however, "nominal" simply refers to the value of X_{ipd} during the initial training and the "altered" value of X_{ipd} is simply the other value used in the test.

Five of the subjects (S1, S12, S21, S22, and S27) performed test T40 first and the other four subjects (S8, S23, S25, and S26) performed T42 first. Note that this is the same group of subjects who took part in the depth discrimination experiment discussed in chapter 5.

6.3.1.2 Apparatus and Stimuli: The apparatus was identical to the first identification experiment (see section 6.2.1.2), except for the graphical scale used to input the subjects' responses. Similarly, the stimuli were identical to the discrimination experiment (see section 5.1.2), except for the scale and the exact parallax values (although the range of parallax was approximately the same in the two cases). In the present experiment, the scale was incorporated inside the left target such that it also served as the depth reference in the identification task. Both the right and left targets, including all of the scale markings, were rendered in red.

Figure 6.7 presents a full-scale drawing of the left target/scale. Aside from the internal makings, the scale differed from the left target in the discrimination experiment in that it was always presented with zero parallax and without jitter.[18] Thus the appearance of the scale was the same in every trial, except for the position of the pointer. The latter consisted of two horizontal lines that straddled the indicated depth mark.

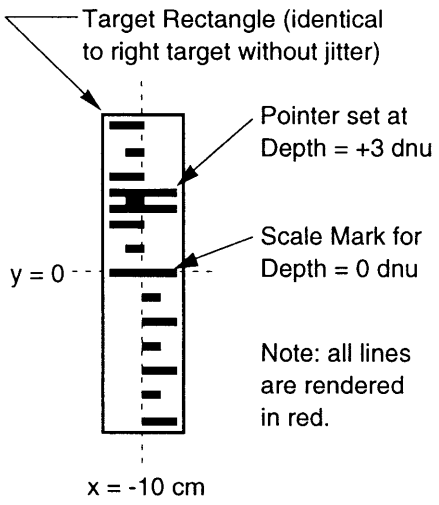


Figure 6.7. Scale (T40 and T42, v2.102).

There were 13 marks on the scale, corresponding to the 13 stimulus depths in the test. The mark in the middle (at $y = 0$) corresponded to the zeroth depth stimulus (i.e., zero parallax or $z = 0$). The marks in the upper half of the scale corresponded to positive depths ($z > 0$) and the marks in the lower half corresponded to negative depths ($z < 0$). The initial reading on the scale at the start of every trial was zero (dnu), and the subject input his or her response in the same manner as in the first

[18]Specifically, the AddJitter function of section 3.2.4.5 was not called. And, while the additional function that jitters the center of the right target image by ±1/2 pixel (see footnote in section 5.1.2) was called, the latter had no effect because the width and the height of the scale had an even numbers of pixels.

experiment. Feedback was also given to the subject as in the first experiment (i.e., a green reference pointer appeared on the scale, see section 6.2.1.2).

6.3.1.3 Procedure: The test consisted of 8 groups of 130 trials where each of the stimulus depths was presented 10 times in random order. Aside from the random order of the trials, the groups differed only in the value of X_{ipd}, and whether or not feedback was given, as shown in table 6.2.

Table 6.2. T40 and T42 (v2.102) Test Groups.

Group	T40 X_{ipd}	T42 X_{ipd}	Feedback
1. Initial training	1.0	2.0	yes
2. Initial training	1.0	2.0	yes
3. Preexposure, nominal	1.0	2.0	no
4. Preexposure, altered	2.0	1.0	no
5. Exposure (adaptation training)	2.0	1.0	yes
6. Exposure (adaptation training)	2.0	1.0	yes
7. Postexposure, altered	2.0	1.0	no
8. Postexposure, nominal	1.0	2.0	no

Double initial training and exposure groups were used in an attempt to insure that each subject's performance approached a reasonable asymptote prior to the pre- and postexposure measurements. Otherwise the procedure was similar to that used in the first identification experiment (see section 6.2.1.3). Some corrections were made to the data by the experimenter prior to the analysis (see appendix D).

6.3.2 Results

Figures 6.8 and 6.9 present the results for T40 and T42, respectively. The figures are similar to figure 6.2 in section 6.2.2, which presents the results of the first experiment (T80, v2.101). Like figure 6.2, the top graph in the figures presents the subjects' mean estimates and the bottom graph plots the resolution, d_{ij}^* (see equation 6.1).

In general, as in the first experiment, the overall shifts in bias are as one would expect. Also, in particular, the shifts show the same asymmetries (and the other detailed bias effects not related to display pixelation) observed in the first experiment. For example, the initial effect (a) in T40 (preexposure, x1 $\rightarrow$ x2 IPD) is limited by the finite size of the response set, whereas in T42 (preexposure, x2 $\rightarrow$ x1 IPD) it is not.

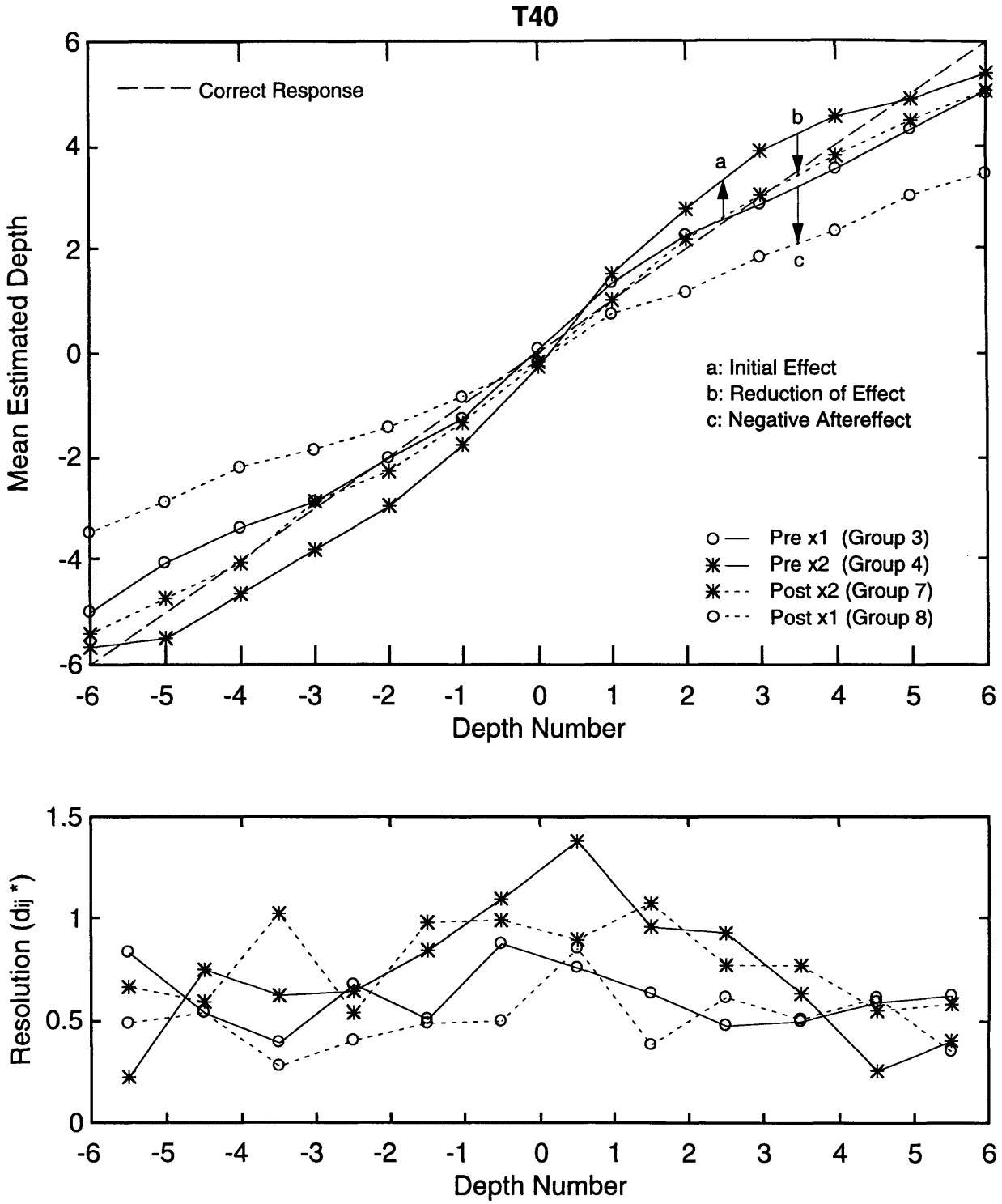


Figure 6.8. Mean estimated depth (showing shifts in bias) and resolution against depth number for all subjects in test T40 of the second identification experiment (v2.102).

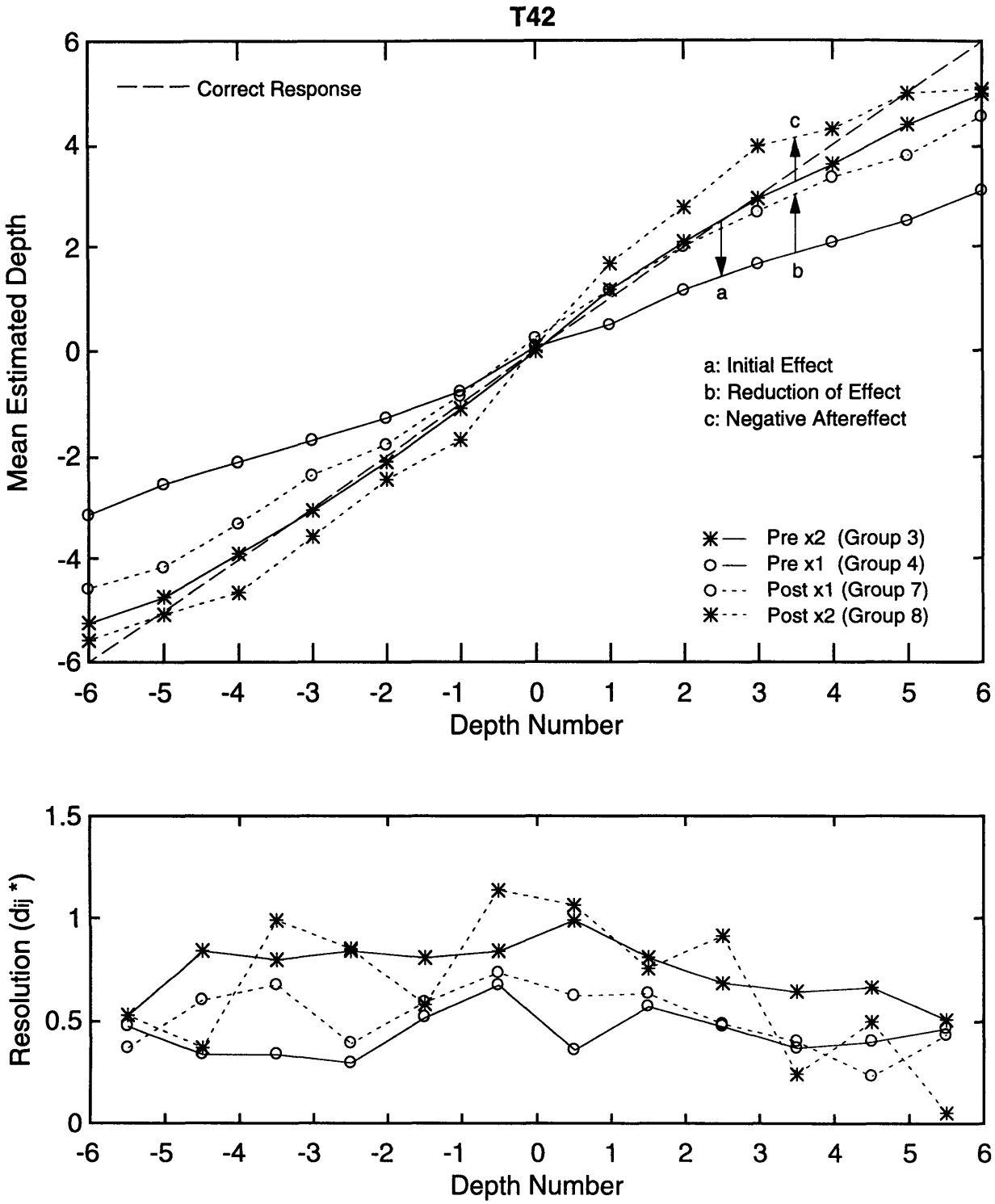


Figure 6.9. Mean estimated depth (showing shifts in bias) and resolution against depth number for all subjects in test T42 of the second identification experiment (v2.102).

Also, as expected, resolution is larger overall at x2 than at x1 IPD. Interestingly, the shapes of the resolution curves are slightly different in the two tests. Specifically, in T40 (figure 6.8) the preexposure x2 IPD curve is more peaked in the middle than the postexposure x2 IPD curve. In the latter case the resolution is roughly constant across the stimulus depths. In contrast, in T42 (figure 6.9), the preexposure x2 IPD curve is nearly constant while the postexposure curve is more peaked (at least the resolution seems to drop off more near the ends). Also note that all of the x1 IPD resolution curves are roughly constant with depth, and that the x2 IPD resolution falls to near the x1 IPD level (or less) at the ends in all cases.

Unfortunately, there is considerable noise in the raw resolution curves, even though each data point is based on 900 pairs of trials! Figure 6.10 presents the mean resolution across stimulus depth in an effort to make the results more clear. Specifically, the figure plots the mean resolution for various intervals around the center of the stimulus range. The averaging intervals are defined by the radius (r) from depth number zero to either end of the interval (inclusive). For example, r defines the region around depth zero such that r = 1 includes the middle three depth stimuli [-1,0,1] and r = 6 includes all of the stimuli [-6,...,6]. This analysis assumes that the "true" resolution curves are symmetric about zero.

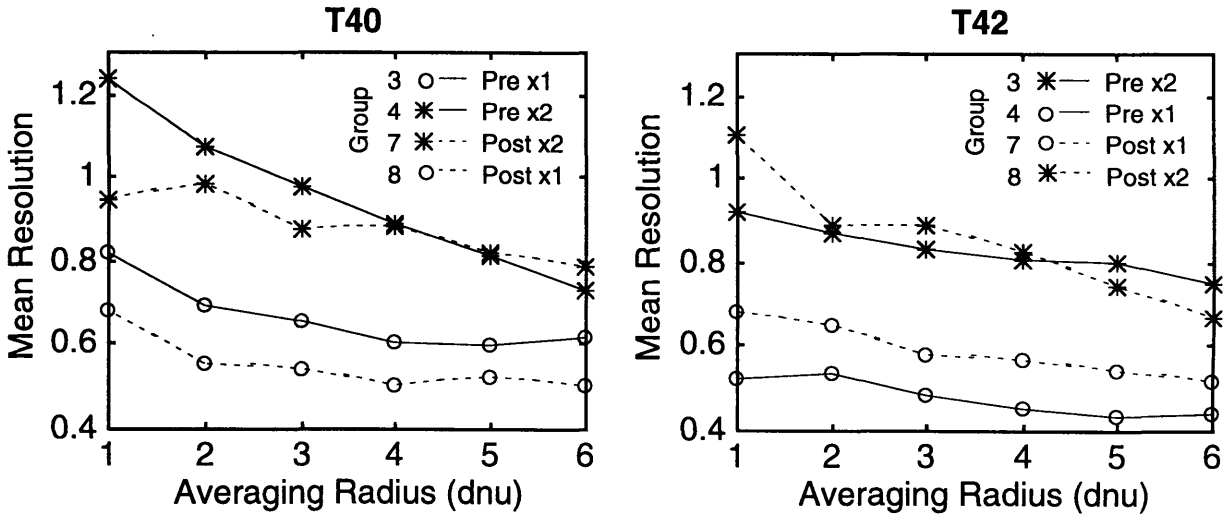


Figure 6.10. Mean resolution averaged across depth intervals symmetric about zero (tests T40 and T42, v2.102).

Figure 6.10 highlights most of the resolution results observed in the previous two figures. First, the x2 IPD resolution is clearly greater than the x1 resolution in all cases. Second, the curves that have a noticeable peak at depth number zero in the raw resolution plots (T40 preexposure x2 IPD and T42 postexposure x2 IPD) have an obvious downward slope with increasing radius. This indicates greater resolution in the center of the stimulus range (i.e., near zero) in agreement with the raw plots. The other three curves in each graph are roughly horizontal, indicating, as observed earlier, that the resolution is roughly constant with depth in these cases.

Three new features of the results are made apparent in the mean resolution plots. First, even the curves that are roughly constant with depth still have a slight downward slope, where the mean resolution decreases as the radius of the averaging interval increases. This indicates that (even in these cases) the resolution is slightly better in the center of the stimulus range (i.e., near zero) as one would expect given that the reference target is located at zero. Second, the resolution changes between pre- and postexposure measurements for both x1 and x2 IPD. This is a very interesting result: why does the resolution change in these cases? Third, there is a clear symmetry between the two tests. Specifically, if the T40 curves in figure 6.10 were to be shifted down about 0.07 (dimensionless units), the entire set would be roughly superimposed on the T42 curves (assuming the data for both tests were on the same plot), except that the pre- and postexposure groups would be switched.

6.3.3 Discussion

The observed shifts in bias have already been explained in connection with the first experiment (see sections 6.2.3.4 to 6.2.3.6). Further discussion of both the bias and the resolution results is also presented in chapter 7 in connection with the quantitative analysis. Before moving on to chapter 7, however, this section seeks to answer the following qualitative questions about the resolution results.

First, what causes the shift in resolution between pre- and postexposure measurements? Second, while the symmetry of the two tests (i.e., the fact that the two sets of curves in figure 6.10 may be roughly superimposed on one another) is presumably due to the symmetric ordering of the groups, how does one explain the asymmetries? Specifically: a) why are the shapes of the x2 IPD resolution curves different between pre- and postexposure, and b) why are the T40 curves in figure 6.10 higher than the corresponding T42 curves?

6.3.3.1 Pre- to Postexposure Shift: The shifts in resolution apparent in figure 6.10 may be explained by a change of effective range, as predicted by the preliminary adaptation model. Recall that the effective range is the range on the decision axis attended to by the subject between the means of the minimum and maximum stimulus distributions (see equation 2.9), and that the range is a function of the state of adaptation (see section 2.5). If the effective range increases, the resolution decreases (and vice versa). This may be seen in the following equation that presents the relevant features of the model (see equation 2.10):

$$d_{ij}' = \frac{\mu_j - \mu_i}{\left(\beta^2 + G^2R^2\right)^{\frac{1}{2}}} \qquad (6.4)$$

where:

d_{ij}' = normalized resolution between any two stimuli i < j
μ_c = mean of the distribution of stimulus c (c = i, j)
β^2 = variance of the sensation noise
G^2R^2 = variance of the memory noise
G = constant
R = effective stimulus range.

Shinn-Cunningham suggested this mechanism to explain similar results in her experiment (Shinn-Cunningham 1994, p.169). Specifically, in her case, a slight decrease in resolution was observed as the subjects adapted to the auditory transformation (see section 2.3.4). The same mechanism serves to explain all of the pre- to postexposure shifts in resolution that are observed here (except as discussed in the sections 6.3.3.2 and 6.3.3.3 below).

For example, the case in the present experiments that corresponds most closely to Shinn-Cunningham's results is the shift in T40 between groups 4 and 7 shown in figure 6.10 (T40). In this case, the subjects are initially trained and tested at x1 IPD (groups 1-3). Then, resolution improves when the subjects are tested in the next group (4) at x2 IPD. This happens because the distributions are further apart on the decision axis (the quantity μ_j-μ_i is greater), while the denominator remains relatively constant. Next, as the subjects adapt between groups 4 and 7,

the effective range increases[19] causing the resolution to decrease (at least in the middle of the stimulus range, see section 6.3.3.2). The quantity μ_j-μ_i remains constant, of course, because the physical stimuli are the same in the two cases.

The same explanation also works for the *decrease* in resolution that is observed in T42, between the pre- and postexposure groups at x2 IPD (groups 3 and 8, respectively). In the later case, the effective range is larger in the preexposure x2 IPD group (3), following the initial training at x2 IPD, than in the postexposure x2 IPD group (8), following the exposure at x1 IPD. The observed pre- to postexposure resolution shifts in the x1 IPD groups are also consistent with the proposed explanation.

6.3.3.2 Difference in the Shape of the x2 IPD Curves: The relatively higher resolution observed in the middle of the range (compared to the ends, see figure 6.10) for the T40, preexposure x2 IPD group (4) and the T42, postexposure x2 IPD group (8), is related to the finite size of the response set and the effective range. In both cases, the effective range is smaller than the optimum correct-response range. Thus, stimuli near the ends of the stimulus range tend to be identified closer to the minimum or maximum depths, as appropriate. The latter effect is the same as the response bias already discussed in section 6.2.3.4.

As a result of the response bias, the mean responses are closer together near the ends than in the middle of the range (see top of figures 6.8 and 6.9). This, in turn, would (and apparently does) cause the observed resolution (d*) to be relatively higher in the middle of the range. The latter is expected, assuming the response variance is roughly constant across the range (see equation 6.1).[20]

In contrast, the other x2 IPD resolution curves (T40, postexposure, group 7, and the T42, preexposure, group 3) are more constant across the stimulus range. The difference in shape is consistent with the present

---

[19]Note that the effective range is larger in group 4 than in group 3, as discussed in section 6.2.3.5, but it does not reach the level of complete adaptation for x2 IPD (see chapter 7, figure 7.11). Hence, the effective range still increases between groups 4 and 7.

[20]Section 7.3.3 presents evidence that the variance of the distributions on the decision axis is roughly constant across the stimulus range. However, that analysis does not include responses at the extreme ends of the range (intentionally, to reduce bias, see section 7.1.3.2), so that the assumption here is not entirely verified. Indeed, the bias at the ends of range, for the measured standard deviations of the responses that are used to determine d*, would tend to make d* larger. Fortunately, consideration of this additional factor only strengthens the present explanation about why the shape of the curves is different. Note that one expects less σ-bias in the cases discussed in the following paragraph.

explanation, because the effective range is closer to the correct-response range in the latter cases. Consequently, the mean responses are more uniform across the stimulus range in these cases and, hence, so is the observed resolution.

6.3.3.3 Resolution shift between T40 and T42: One hypothesis about why the two sets of curves are shifted is that the subjects do not adapt to the altered condition as completely during the exposure as they do to the nominal condition during the initial training. In other words, after learning the task one way (during the initial training), it is harder for them to relearn it a new way only a short time later (during the exposure). Thus in T40 (where the initial transformation is x1 IPD), the subjects always attend to a slightly smaller effective range on the decision axis than they do in T42 (where the initial transformation is x2 IPD). A smaller effective range, in turn, has less memory noise and, hence, better resolution (at least in the middle of the range) for the same pair of physical stimuli.

This effect is, of course, precisely what is observed. The hypothesis also predicts that both x2 IPD resolution curves should be more peaked near zero in T40 than in T42. The peaks occur for the reasons already discussed in section 6.3.3.2 for one pair of the curves (T40 preexposure x2 IPD and T42 postexposure x2 IPD). Indeed, the peak does seem to be larger in T40 for this pair of curves. The difference is apparently too small to see, however, in the case of the other pair of curves (T40 postexposure x2 IPD and T42 preexposure x2 IPD). Sections 7.3.2.1,3 and 7.3.2.2 discuss this hypothesis further.

7. QUANTITATIVE ANALYSIS

While the preliminary adaptation model is useful for explaining the qualitative results of the second identification experiment (just discussed), the real potential power of the model is its ability to make quantitative predictions. This chapter focuses on the second identification experiment (T40 and T42, v2.102), looking more closely at the results and testing predictions of the model against them. The analysis provides further insight into some of the qualitative results as well as evidence, both for and against the model.

The chapter is divided into six sections. In the first section (7.1), specific equations that may be applied to the experiment are derived from the equations presented in chapter 2 (see section 2.5). The derived equations are ultimately used to predict the results of the second identification experiment in section 7.5. In that section, the predictions of the model are compared to the experimental results and the two are found to be in rough agreement. Some differences are observed, however, and the chapter concludes with a discussion of the model's limitations in section 7.6.

On a first reading, the reader may choose to skip the intervening sections (7.2-7.4) without loss of understanding at the end. The analysis presented in the middle of the chapter has two purposes. First, unknown parameters of the model are evaluated from the results of the second identification experiment. This is done so that the parameters may be used in the predictions of section 7.5 (an overview of the evaluation is presented in section 7.1.3). Second, the behavior of the parameters is observed and discussed within the context of the preliminary adaptation model. These observations are the source of many theoretical insights.

7.1 Application of the Preliminary Adaptation Model

The first step of the analysis is to define the relations, variables, and other parameters of the model. All definitions presented here are consistent with earlier definitions (e.g., chapter 2). They are simply restated in terms of the second identification experiment, with more precision, or simply for convenience. Then, each of the terms is evaluated, leaving only k(t) and G unknown (see definitions below). The evaluation of these two remaining terms is motivated, next, through a look at some representative data. The discussion includes an overview of the evaluation of the terms in subsequent sections of the chapter (7.2-7.4). Section 7.1 then concludes with the derivation of the final equations of the model.

7.1.1 Definitions

This subsection defines all of the quantities that are used in the chapter except for some definitions that are presented in sections 7.1.3.3 and 7.1.3.4. The definitions are listed, beginning with a mapping (7.1) that summarizes the preliminary adaptation model and ending with the definitions of some special units (*dau* and *dnu*).[1] These definitions will not be repeated in subsequent discussion. The model is summarized as follows:[2]

$$\mathbf{N} \xrightarrow{f_{xipd}} \mathbf{P} \xrightarrow{r_s} \mathbf{Y} \xrightarrow{r_m} \mathbf{Q} \xrightarrow{f_r} \mathbf{M} \tag{7.1}$$

with functions and relations given by:

$$f_{xipd}:\mathbf{N} \rightarrow \mathbf{P} \; : \; P = X_{ipd}N \;\; (P \in \mathbf{P}, N \in \mathbf{N}) \tag{7.1a}$$

$$r_s:\mathbf{P} \rightarrow \mathbf{Y} \quad : \mathbf{Y} \text{ is a random variable with Gaussian distribution, mean } = \alpha(P), \text{ variance } = \beta^2 \tag{7.1b}$$

$$r_m:\mathbf{Y} \rightarrow \mathbf{Q} \quad : \mathbf{Q} \text{ is a random variable with Gaussian distribution, mean } = \alpha(P), \text{ variance } = \beta^2 + \gamma^2 \tag{7.1c}$$

$$f_r:\mathbf{Q} \rightarrow \mathbf{M} \quad : M = M_i \text{ iff } C_{i-1} < Q \leq C_i \;\; (Q \in \mathbf{Q}, M, M_i \in \mathbf{M}, C_{i-1}, C_i \in \mathbf{C}) \tag{7.1d}$$

with fundamental sets and variables:[3,4,5]

---

[1]Depth number units (dnu) are also defined previously in section 5.1.2.1, equation 5.1.

[2]Refer to sections 2.4 and 2.5 for further description of the model.

[3]The same variable may be used to represent, both, the "true" quantity of the model and the observed quantity in the experiment. This is done because using different symbols for the two cases becomes unwieldy. Indeed, experimental values are used as parameters of the model. The distinction should be clear in the context of the discussion.

[4]Convention: a quantity that is defined with a subscript (e.g., N_i) may be used without the subscript (e.g., N). In the latter case the quantity is assumed to have the same essential meaning, but without a specific value assigned to it (e.g., N might represent any stimulus depth number). Further, it is sometimes convenient to use such a variable to represent quantities that are not included in the original set. For example, M might be used to represent the mean response. The meaning in each case should be clear from the context.

[5]Convention: in the case of quantities that are defined as functions of an argument (e.g., $C_i(t)$), the argument may not be explicitly stated in all cases (e.g., C_i) for convenience. In such cases, a

$\mathbf{N}$ = set of virtual stimulus depth used in the experiment
= $\{ N_{min} \leq N \leq N_{max} \mid N \text{ an integer}\}$
N_i = i^{th} stimulus depth (dnu)

$\mathbf{P}$ = set of screen parallaxes used in the experiment
= $\{ X_{ipd}N_{min} \leq P \leq X_{ipd}N_{max} \mid P \text{ an integer} \}$
P = screen parallax (pixels)

$\mathbf{Y}$ = unidimensional continuum of sensations (modeled as identical with the real numbers)
Y = particular value of $\mathbf{Y}$ resulting from the presentation of stimulus P in the experiment (dau)

$\mathbf{Q}$ = unidimensional continuum that serves as the *decision axis* (modeled as identical with the real numbers)
Q = particular value of $\mathbf{Q}$ resulting from the presentation of stimulus P in the experiment (dau)

$\mathbf{M}$ = set of allowable subject responses (identical to $\mathbf{N}$)
= $\{ M_{min} \leq M \leq M_{max} \mid M \text{ an integer}, M_{min} = N_{min}, M_{max} = N_{max}\}$
M_i = i^{th} response depth (dnu)
M_i^* = mean response following the presentation of N_i (dnu)[6]

$\mathbf{C}$ = set of criteria (one element larger than $\mathbf{M}$)
= $\{ -\infty = C_{min-1} \leq C_{min} \leq \ldots \leq C_i \leq \ldots \leq C_{max-1} \leq C_{max} = \infty \mid C_i \in \mathbf{C} \subset \mathbf{Q} \}$
$C_i(t)$ = position of the i^{th} criterion on the decision axis (dau)
C_{Mi} = position of the i^{th} criterion on the decision axis rescaled to depth number units (dnu)

and other quantities and parameters:

$\alpha(P)$ = psychophysical scale relating the physical stimulus, P, to the mean values of the perceived quantities Y and Q (dau)
β^2 = variance of sensation noise (dau^2), assumed constant in the experiment

---

subscript may serve to identify specific values of the parameter (e.g., k_g). This is particularly convenient in the case of functions of t, because the exact dependence on t is not established.

[6]An asterisk is used to differentiate M_i^* from M_i in this case, even though the convention is to distinguish different forms of variables via context (e.g., estimates from "true" values), because both terms are used in equation 7.16. Also note that M_i^* is a real number, not contained in $\mathbf{M}$.

Φ = normal cumulative distribution function (see equation 7.15)

γ^2 = variance of memory noise (dau^2), assumed equal for all depths
= $G^2R(t)^2$ (by assumption, see equations 2.5 and 2.9)

σ^2 = variance of the distributions on the decision axis (dau^2), assumed equal for all depths
= $\beta^2 + \gamma^2$ (by assumption, see equations 2.4 and 7.1c)

σ_g = mean value of σ for some group of trials identified by g (dau), see equation 7.11 for other definitions of σ

$d_{ij}'(t)$ = normalized resolution between any two stimuli $i < j$

d_{ij}^* = observed resolution given by equation 6.1

G = constant that specifies the relative affect of sensation and memory noise (dau)

k(t) = adaptation coefficient (dnu/pixels)[7]

k_g = mean value of k(t) for some group of trials identified by g (dnu/pixels)

N_{min} = minimum stimulus depth (dnu)

N_{max} = maximum stimulus depth (dnu)

R(t) = effective range of stimuli on the decision axis (dau)

t = number of trials completed since the current test condition was imposed on the subject[8]

X_{ipd} = ratio of effective IPD/actual IPD (the sensory transformation that is imposed on the subject with units of pixels/dnu in equation 7.1a)

where the special units are:

dau = decision axis unit, defined by $\alpha(P)$ such that $\beta = 1$.

dnu = depth number unit, defined such that $N = P$ when $X_{ipd} = 1$.

7.1.2 Evaluation of Terms

The appropriate psychophysical scale, $\alpha(P)$, was determined in the discrimination experiment as follows (see section 5.4, equation 5.2):

---

[7]Consistent with the discussion in section 2.6, k is defined as the mean value of M/P for any given state of adaptation. Section 7.1.3.4 describes how k is estimated in the present experiment and defines the related coefficients k' and A.

[8]The parameter, t, replaces the parameter, r, that is used in chapter 2 (see equation 2.7). Note that different nomenclature is used because the two parameters are defined differently. Also, the exact forms of the functions are not established in chapter 7. Instead, t serves only to identify a presumed functional dependence.

$$\alpha(P) = 0.8\ P\ (dau) \tag{7.2}$$

where the coefficient in the equation is defined such that when P is given in pixels, $\alpha(P)$ is defined in units equal to the standard deviation of the sensation noise. In other words, given equation 7.2,

$$\beta \equiv 1\ (dau) \tag{7.3}$$

on the decision axis.

This scale is based on the mean sensitivity of the same nine subjects who took part in both the discrimination experiment and the second identification experiment (see sections 5.1.1 and 6.3.1.1). While the exact form of $\alpha(P)$ may differ for other subjects or experiments, it should apply reasonably well in the present case.

The minimum and maximum stimulus depths are given in section 6.3 as follows:

$$\begin{aligned} N_{min} &= -6\ (dnu) \\ N_{max} &= +6\ (dnu). \end{aligned} \tag{7.4}$$

Thus, given that f_{xipd}, N, and M correspond to f_n, Θ, and θ, respectively (the latter are defined in chapter 2), equations 2.8-2.10 may be restated and simplified as follows in the present case:[9]

$$C_i(t) = \left\{ \alpha\left(\frac{\frac{1}{2}(M_i + M_{i+1})}{k(t)} \right) \right\}_{-\infty}^{+\infty} = \begin{cases} +\infty & M_i = M_{max} \\ 0.4\,\dfrac{M_i + M_{i+1}}{k(t)} & M_{min} \le M_i < M_{max}\ (dau) \\ -\infty & M_{i+1} = M_{min} \end{cases} \tag{7.5}$$

$$R(t) = \alpha\left(\frac{M_{max}}{k(t)} \right) - \alpha\left(\frac{M_{min}}{k(t)} \right) = \frac{10}{k(t)}\ (dau) \tag{7.6}$$

---

[9]The coefficients in the equations, 7.5-7.7, are defined with the appropriate units so that the units of each of the parameters is as defined in section 7.1.1. The resolution, d_{ij}' is, of course, dimensionless.

$$d_{ij}'(t) = \frac{\alpha\left(f_{xipd}(N_j)\right) - \alpha\left(f_{xipd}(N_i)\right)}{\left(\beta^2 + G^2R(t)^2\right)^{\frac{1}{2}}} = \frac{0.8X_{ipd}\left(N_j - N_i\right)}{\left(1+100\frac{G^2}{k(t)^2}\right)^{\frac{1}{2}}}. \qquad (7.7)$$

Note that the normalized bias, B_i (equation 2.11), is not used in the following analysis. Instead, the model is used to predict the subjects' mean responses in accordance with the presentation of the results in chapter 6 (see sections 6.2.2 and 6.3.2). In other words, the measure of bias is mean error. This measure is approximately the same as B_i[10] and it makes little difference in the present experiment whether the error is normalized or not because the observed σ is more or less constant for all of the stimulus depths (see section 7.3.3). As a result, normalizing the error simply scales it by a constant factor. Indeed, because σ also happens to be approximately equal to one in the present case (see section 7.3.1), normalizing has little affect on the results.

7.1.3 Overview of the Evaluation of k(t) and G

The two remaining unknown terms, k(t) and G, involve key assumptions of the model. The constant G is assumed in the preliminary theory of intensity resolution for the particular type of memory noise (context mode) that is appropriate here (Durlach and Braida 1969, p. 377). Similarly, a key assumption of the preliminary adaptation model is that the subject's state of adaptation may be fully described by the slope (k) of a linear approximation to the observer's responses (Shinn-Cunningham 1994, p.107). The latter assumption is particularly reasonable in the present case where the sensory transformation, f_{xipd}, is linear.

Assuming that the model is correct, the two terms are related by the following equation,

$$\sigma^2 = 1+100\frac{G^2}{k(t)^2}\ (dau^2). \qquad (7.8)$$

Thus, once k is known, G can be estimated from the variance of the responses.

---

[10]Compare $B_i = (1/2(\mu_i + \mu_{i+1}) - C_i)/\sigma_i$ to Mean Error = μ_i - $1/2(C_{i-1} + C_i)$.

The analysis presented in sections 7.2-7.4 partitions the test into sets of trials where σ and k are assumed constant and, then, the mean values are used to estimate G. Specifically, the mean values σ_g and k_g are estimated in section 7.3 for each of the 8 main groups of (130) trials in the tests (see section 6.3.1.3). A nearly identical calculation is also performed for 24 subgroups where the main groups are divided approximately into thirds. The results of these calculations (the pairs of σ_g and k_g) are, then, used to estimate G in section 7.4.

A finer partitioning of the test, into 10-trial intervals, is also performed in section 7.2 in order to gain further insight into how k varies with t. Note that the analysis does not lead to an equation for k(t) corresponding to equation 2.7 in chapter 2. Instead, the ultimate predictions of the model in section 7.5 are based on the observed mean values of k (i.e., k_g) determined in section 7.3.

Accepting for the moment that the approach will work, the question is: how does one estimate σ and k for a given group of trials? The answer is motivated by a look at some representative data.

7.1.3.1 Representative Data: Figure 7.1 presents the response histograms for one subject (S1) in one test (T40), as an example of the type of data to be analyzed. The figure consists of four separate plots, corresponding to the four indicated test groups (preexposure x1 and x2 IPD, postexposure x2 and x1 IPD). The ordinate in the plots is the response depth, M, while the abscissa represents both the stimulus depth, N (as indicated), *and* the number of responses. A separate histogram is presented for each stimulus depth with the zero-response level plotted at the corresponding value of N. The horizontal bars of the histograms, corresponding to the number of responses at each depth, are scaled so that one response equals 1/10 dnu on the abscissa. All of the responses for the group are presented in the given plot.

Assume for the purposes of the immediate discussion that, while σ and k might change from group to group, they are constant in each of the plots. This is consistent with the initial assumptions of the experiment, because there is no direct (i.e., explicit correct-answer) feedback in any of the groups considered in the figure. It is also a reasonable approximation that will be used in the subsequent analysis, even though k and, consequently, σ (according to equation 7.8) change in the no-feedback groups as discussed in section 6.2.3.5. Sections 7.2 and 7.3 investigate how k changes as a function of t. For now, however, it helps to keep things simple.

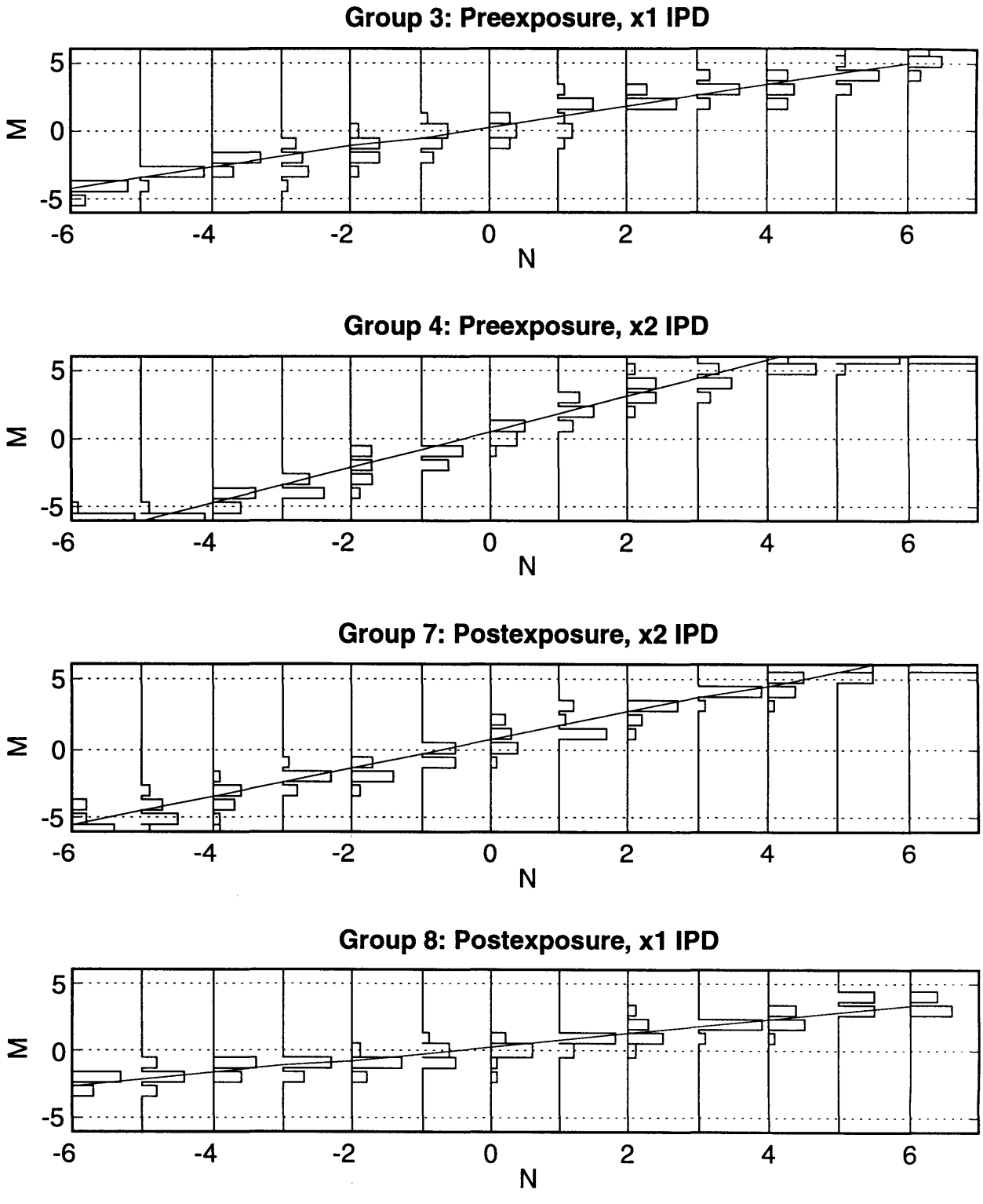


Figure 7.1. Response (M) histograms for one subject (S1) in one test (T40, v2.102) plotted against stimulus depth (N).

There are several important features of the results that are illustrated in figure 7.1. First, the effect of the finite size of the response set is apparent, particularly in the preexposure x2 IPD case (group 4) where the depths appear larger to the subject than in the previous group. Allowing for the effects at the ends, however, the mean responses appear to be approximately linear with stimulus depth number. Note that the best-fit line calculated for the intermediate responses is drawn on the plots (to be discussed shortly, see equation 7.12). Further, the slope changes from group to group in accordance with the subject's state of adaptation. The slope in the figure (call it $k' = M/N$, in the intermediate range) is linearly related to k as follows:

$$k = k'/X_{ipd} \text{ (dnu/pixels)}. \tag{7.9}$$

The figure is also consistent with the assumption, allowing for the effects at the ends and the small number of trials,[11] that the response distributions are roughly Gaussian with equal variance across the stimulus range.[12] Note that the histograms in the figure approximate the probability density functions on the decision axis, simply rescaled by 0.8/k. This is true because the response depths, M, map onto the decision axis in a linear fashion according to the following equation:

$$Q = \alpha(M/k) = 0.8\ M/k \text{ (dau)}. \tag{7.10}$$

As a result, σ, and hence, G can be estimated from the experimental results given a sufficient number of trials where σ and k are constant. The additional assumption, that σ is the same for all stimulus depths, can also be tested at the same time. Further, since k clearly changes in the experiment (e.g., compare groups 4 and 7 in the figure), if G is approximately the same for different pairs of σ and k, there would be significant evidence that G is a constant. The only remaining issue, before the detailed calculations are presented, is: how does one deal with the effects at the ends of the response range?

7.1.3.2 Dealing with Edge Effects: The approach taken here is to eliminate all data related to either a minimum or a maximum response (i.e., $M = M_{min}$ or $M = M_{max}$). The basic idea behind the approach is that a response of M_{min} or M_{max} gives no information about the value of Q on the

---

[11]There are 10 trials per stimulus depth per group (see section 6.3.1.3).

[12]It may help to turn the figure on its side in order to see the distributions more clearly.

decision axis, other than that it is outside the corresponding (finite) minimum or maximum criterion. For example, a response of M_{min} only tells us that the sensation (Q) was less than C_{min}. The unwanted data is removed from the analysis as follows.

Any stimulus that resulted in at least one response of M_{min} or M_{max} is eliminated from the analysis. In other words, all trials where that stimulus was presented are removed from the group of trials in the analysis. Beyond this, all stimulus depths less than or equal to the one that resulted in a response of M_{min} (or greater than or equal to a stimulus that resulted in a response of M_{max}), are eliminated from the analysis. For example, in figure 7.1 the data for N = 5 and N = 6 would be eliminated in the analysis of group 3. Similarly, only N in the range from -4 to 3 would be included for group 4.

Unfortunately this approach throws-away a great deal of data, making the result less certain. There is also a bias in the determination of k because data points that cause k' to be smaller (and, hence, k to be smaller) are more likely to be included than data that makes k' larger.[13] The estimate of σ is also biased, particularly near the ends of the range, because responses far from the mean are not included in the measurement. As a result, the observed value of σ is likely to be smaller than the true value. Nevertheless, biases in the determinations of k and σ are less than if M_{min} and M_{max} were included (see group 4 in figure 7.1). These issues will be discussed further throughout the chapter.

7.1.3.3 How σ_g is Estimated: An estimate of the standard deviation of the distribution on the decision axis is derived from the standard deviation of the responses, M, for a given stimulus, N (e.g., the standard deviation for one of the histograms in figure 7.1). The standard deviation is calculated with the Matlab function, "std," that "computes the *sample* standard deviation" normalized by n-1, where n is the number of responses (MathWorks 1992, p.482). This is an unbiased estimate of the standard deviation (Wonnacott and Wonnacott 1987, pp. 61-62), except as noted in the preceding paragraph.

Equation 7.10 is then used to transform σ_M to σ, i.e.,[14]

---

[13]This mechanism was already proposed in section 6.2.3.6 as a possible source of the subjects' central tendency of judgment related to the finite size of the response set.

[14]Note that equation 7.11 may be used directly with any of the various versions of σ, for example, the values corresponding to individual stimuli (σ_{mi} to σ_{qi}) or the "true" modeled

$$\sigma_{qi} = 0.8\ \sigma_{mi}/k \quad \text{(dau)} \tag{7.11}$$

where:

σ_{qi} = estimated standard deviation of the distribution on the decision axis (i.e, the estimated value of σ) for a given stimulus, N_i, and for one subject (dau)

σ_{mi} = observed standard deviation of responses, std(M), for a given stimulus, N_i, and for one subject (dnu)

and

σ_q = mean value of σ_{qi} across all available stimulus depths (dau)[15]

σ_m = mean value of σ_{mi} across all available stimulus depths (dnu).

The quantity, σ_g, is then estimated as the mean value of σ_q for all of the subjects in the experiment for the observed group of trials, g. Note that the experimental results are sometimes reported as σ_{Mg} in order to avoid the involvement of k in the calculation of σ_g as follows:

σ_{Mg} = mean value of σ_M for some group of trials identified by g (dnu)

where:

σ_M = value of σ on the decision axis rescaled to depth number units (dnu)

and σ_{Mg} is estimated as the mean value of σ_m, across subjects, for the observed group of trials, g.

7.1.3.4 How k_g is Estimated: Regression (Wonnacott and Wonnacott 1987, pp. 18-22) is used to estimate k, as illustrated in figure 7.2, according to the model:

$$M = A + k'N \quad \text{(dnu)} \tag{7.12}$$

---

values (σ_M to σ). The distinction between these various quantities should be clear in the context of the discussion.

[15]As discussed in section 7.1.3.2, some values of N are not included in the calculation.

where:

A = zero intercept of the best (least squares) fit line (dnu)
k' = $X_{ipd}k$ = slope of the line (as in figure 7.1).

Note that A represents a uniform (distance) bias in the response and that k may be calculated directly from k' as shown (also see equation 7.9).

The value of k_g (i.e., the mean value of k) may be estimated in this manner for any set of (more than 1) trials where k is approximately constant. Note that the value of k_g, that is presented in the following discussion, is the mean value of the estimates of k, in the given interval, for each subject. Also, in the following analysis the estimate of k is arbitrarily assigned to the middle of the interval. For example, k might be estimated at t = 25 for the interval of trials, 21-30.

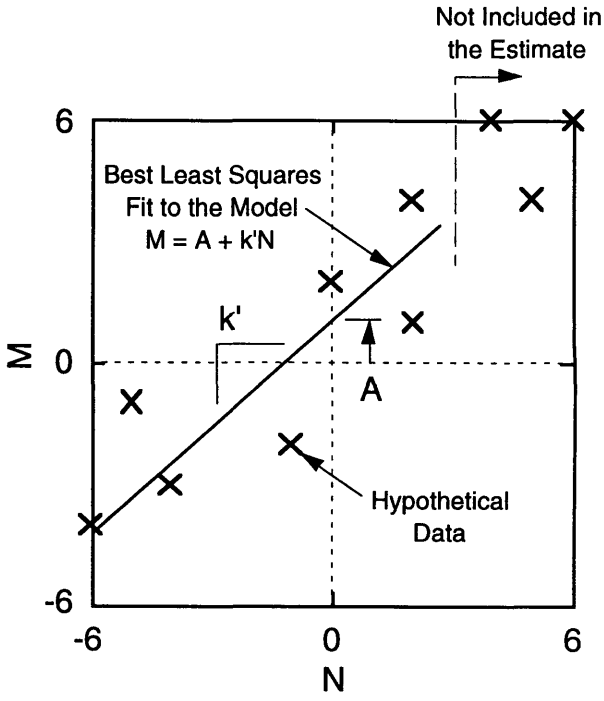


Figure 7.2. Estimation of k.

7.1.4 Final Equations of the Model

The derivation of the equations, used to predict the experimental results in section 7.5, is simplified by rescaling the decision axis into depth number units according to equation 7.10. Thus equation 7.5 is transformed as follows:

$$C_{Mi} = \begin{cases} +\infty & M_i = M_{max} \\ \frac{1}{2}(M_i + M_{i+1}) & M_{min} \le M_i < M_{max} \text{ (dnu)} \\ -\infty & M_{i+1} = M_{min} \end{cases} \qquad (7.13)$$

and σ_M is derived from equation 7.8:

$$\sigma_M = \frac{k_g}{0.8}\left(1+100\frac{G^2}{k_g{}^2}\right)^{\frac{1}{2}} \text{ (dnu)} \qquad (7.14)$$

where k_g is substituted for k(t) in accordance with the rest of the analysis.

The mean value, k_g, is used to approximate k(t) in the model primarily because this approach makes the analysis simpler. It is a reasonable approximation because 1) the observed quantities in the test that one hopes to predict are themselves mean values for groups of trials and 2) k(t) is approximately constant over most of the trials within each of these groups (see section 7.2). Note that the predictions in section 7.5 are for the test groups 3, 4, 7, and 8 (see table 6.2 in section 6.3.1.3) and that these are the same groups for which the results are presented in section 6.3.2.

In the analysis, it is also convenient to use the *normal cumulative distribution function*,

$$\Phi(c,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{c} x e^{-(x-\mu)^2/2\sigma^2} dx, \qquad (7.15)$$

where Φ gives the probability that a Gaussian random variable, with mean, μ, and standard deviation, σ, will be less than or equal to c. Note that the parameters μ and σ refer to any Gaussian distribution in the context of equation 7.15. Also note that x is a dummy variable under the integral and that Φ is calculated, here, using the Matlab function "normcdf."

The mean response (M_i*) is then estimated (using equations 7.13 to 7.15) as the weighted sum of all possible values of M_j as follows:

$$M_i{}^* = \sum_{j=min}^{j=max} M_j\left(\Phi\left(C_{Mj}, k_g X_{ipd} N_i, \sigma_M\right) - \Phi\left(C_{M(j-1)}, k_g X_{ipd} N_i, \sigma_M\right)\right) \text{ (dnu)} \qquad (7.16)$$

and, in turn, the observed resolution (see equation 6.1) is estimated as:

$$d_{ij}{}^* = \frac{M_j{}^* - M_i{}^*}{\sigma_M}. \qquad (7.17)$$

7.2 Initial Investigation of A and k

Figures 7.3 and 7.4 present the results of an analysis where the tests, T40 and T42, respectively, are partitioned into 10-trial intervals. This is done primarily to investigate how the zero intercept, A, and the adaptation coefficient, k, vary over the course of the tests. In the following section, the analysis procedure is discussed first. Then, important features of the results are identified and discussed, leading to a hypothesis about certain aspects of how people change their criteria on the decision axis.

7.2.1 Procedure

In the analysis, the coefficients (A and k) are calculated separately for each subject on each 10-trial interval and assigned to the 5th trial of the interval (see section 7.1.3.4). Then, the mean values of the coefficients are calculated (across subjects) for each interval after some processing to remove outliers.[16] The ordinate of the plots gives the mean (·) plus or minus one standard deviation (points on the jagged lines). The abscissa is the trial number in the test; beginning with the (non-existent) 0th trial, before the start of the test, and ending on trial 1040 (equal to the sum of all of the trials in the test). The vertical lines in the plots indicate group boundaries. The horizontal lines show the values corresponding to the case where the subject is always correct. The latter will be referred to as the *correct-response* value in subsequent discussion.

7.2.2 Results

The first thing to observe in the figures is that, while there is some variation among the subjects, the mean value of A is approximately zero. There does not appear to be any systematic variation of the coefficient across all the trials of both tests. Also the variance across subjects (i.e., the distance between the upper and lower jagged lines) is about the same throughout both tests, independent of the experimental conditions. This is

---

[16]The following processing is performed on the raw data after it is corrected as discussed in appendix D. Note that the raw data consists of nine pairs of values, A and k for each subject, in each of the 104 intervals. First, all coefficient pairs (A,k) are removed from the data set, where k equals inf, -inf, or NaN. Such values result when the calculation fails, for example, because it is necessary to divide by zero. Next, all coefficient pairs (A,k) are removed from the data set, where k is either the high or low value in the interval. The latter does not bias the mean value and it eliminates a great many extreme values that are clearly inconsistent with the rest of the data in the interval and in neighboring intervals.

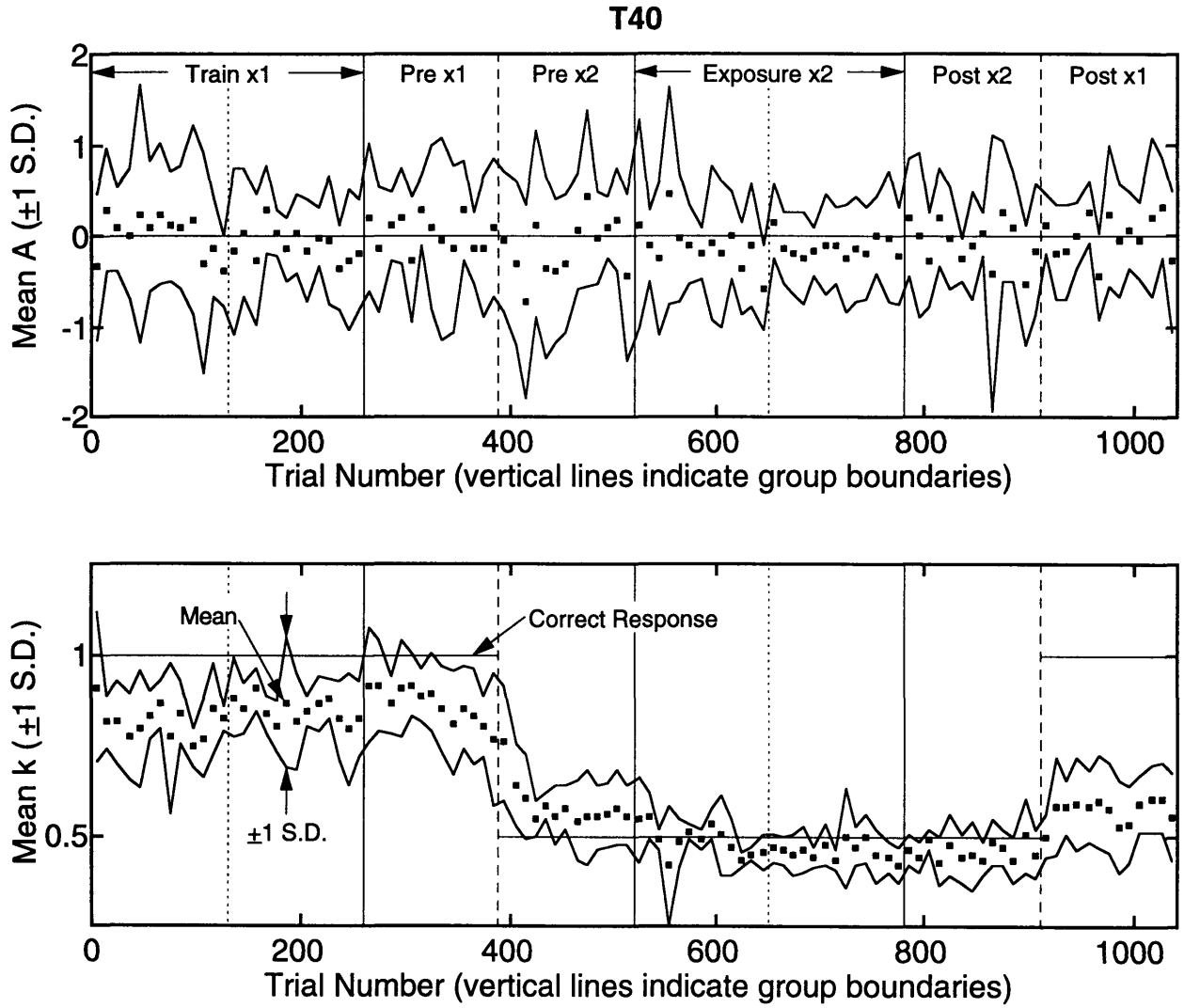


Figure 7.3. Means and standard deviations (across subjects) of 10-trial estimates of A and k in test T40 (v2.102).

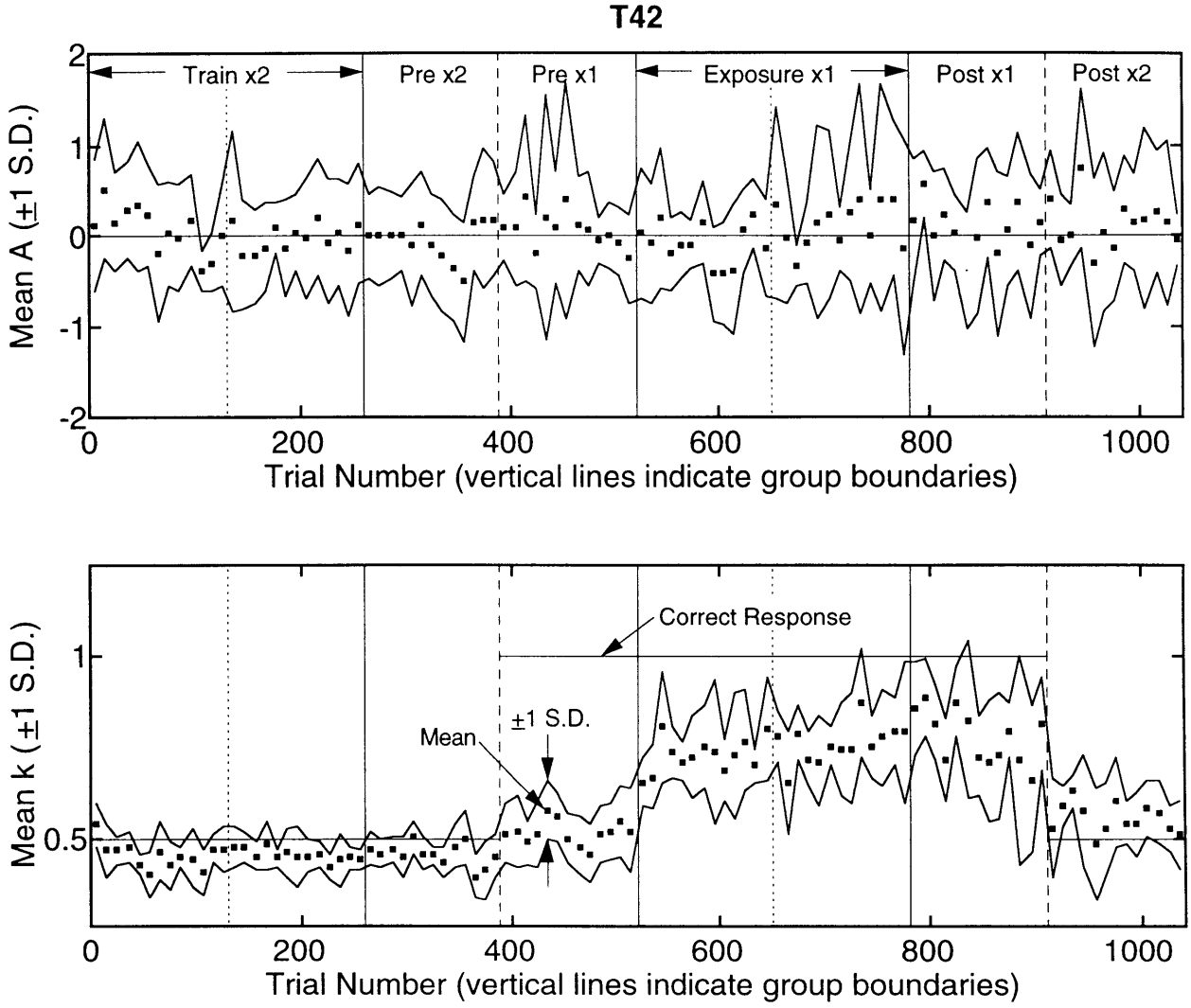


Figure 7.4. Means and standard deviations (across subjects) of 10-trial estimates of A and k in test T42 (v2.102).

a convenient result because the preliminary adaptation model does not take this coefficient into account. In other words, the equations of the model would need to be revised if A was non-zero and, particularly, if systematic changes were observed. It is also not surprising, given the perceptual anchor provided by the reference target, and it will not be discussed further. In contrast, there are systematic changes in k that are similar for both tests.

Beginning at the start of the tests, it can be seen in both figures that the mean value of k is approximately constant throughout the initial training period (G1-2). The observed value of k is slightly less than the correct-response value in both tests (indicated by horizontal lines). Note that the slightly higher value of k that is observed during the first interval of T42 (t = 5, figure 7.4) may represent a real trend, although it only differs from the adjacent point by about one standard deviation (across subjects).

The coefficient remains constant, as one might expect, during the nominal preexposure test group (G3), but then changes in the altered group (G4). A slight downward trend is apparent during group 3 in the case of T40 (figure 7.3), although it is hard to have much confidence in this result given the large variance in the measurements.[17] During the exposure (G5-6), k gradually approaches an asymptote that is slightly below the new correct-response value. Then, after the exposure, just as observed following the initial training, k remains at the same level in the first (altered) postexposure test group (G7) and it changes in the second (nominal) postexposure group (G8).

Finally, it may be observed that the variance of k across subjects (related to the distance between the upper and lower jagged lines) is greater for x1 IPD groups than for x2 IPD groups. There does not appear to be any systematic changes in the variance within these regions however. Specifically, it was anticipated that there would be more variance at the start of the initial training, but this is not observed. Further analysis of k is presented in section 7.3.2.2.

7.2.3 Discussion

Several features of the results are surprising, or otherwise deserve comment. The following section discusses: 1) indirect feedback, 2) the

[17] If this is a real trend, one explanation is that the subjects become less certain about the placement of their criteria over time in the absence of feedback (see section 7.3.2).

fact that k seems to approach a value that is less than the correct-response value during initial training and exposure, 3) changes in the variance of k, and 4) the fact that the mean value of k is approximately constant during the initial training. Initial training is also compared to the exposure and a hypothesis is presented at the end.

7.2.3.1 Indirect Feedback: Although the effect was already discussed in section 6.2.3.5, it is striking how quickly the value of k changes in the absence of direct (i.e., explicit correct-answer) feedback; particularly when one compares group 4 (no feedback) to group 5 (feedback) in the case of T40 (figure 7.3). This behavior was not anticipated prior to the experiments. For further discussion, see section 7.3.2.2.

7.2.3.2 Bias in the Estimate of k: At least part of the explanation of why the observed k tends to approach a value that is less than the correct-response value during training is the fact that the estimate of k is biased (see third paragraph of section 7.1.3.2). Further discussion of the issue is left until section 7.3.2.2, after the results are corrected for this source of bias.

7.2.3.3 Variance of k: Another interesting result is the difference in variance of k across subjects between the x1 and x2 IPD groups (specifically, the distance between the upper and lower jagged lines in the mean k plots is larger for x2 IPD groups). One explanation for the effect is related to the fact that the subjects have poorer resolution in the x1 IPD case (see section 6.3.2). Poorer resolution corresponds to greater variability in the responses and, hence, more variability in the experimental determination of k. In other words, the increased variance of k across the subjects is not due to increased inter-subject variability, but rather, it is due to a general increase in the experimental variability of k itself. And hence, the variability of the mean k increases because the variability of the observed value of k for each individual subject increases.

In order to investigate this hypothesis, the mean and the standard deviation of k were analyzed for individual subjects. The analysis was performed across regions of trials where the mean value of k (across the subjects) is approximately constant. Specifically, groups 2 and 3 (combined) and groups 6 and 7 (combined) were used (see figures 7.3 and 7.4). Note that the value of X_{ipd} is different in different regions and in different tests.

Two quantities were calculated. The first quantity is the mean (across the subjects) of the standard deviation of the calculated k values for each individual subject (across the region). The second quantity is the standard deviation (across the subjects) of the mean k values for each subject (across the region). The results of the analysis are presented in table 7.1.

Table 7.1. Variance in Regions of Constant k.

Experimental Condition	mean(std(k))	std(mean(k))
T40, G2-3, $X_{ipd} = 1$	0.14	0.09
T42, G2-3, $X_{ipd} = 2$	0.06	0.02
T42, G6-7, $X_{ipd} = 1$	0.20	0.13
T40, G6-7, $X_{ipd} = 2$	0.07	0.02

As expected, the results show that the variance of k is greater in the x1 IPD regions than in the x2 IPD regions. Also, in agreement with the hypothesis, the effect is evident in the variance of the subjects' individual k values, mean(std(k)). Indeed, this variance is consistently greater than the variance of k across the subjects, std(mean(k)). Thus, it seems likely that the proposed explanation is correct, although there is also clearly variability among the subjects. Indeed, the existence of variability among the subjects helps to explain how the mean value of k can be approximately constant in the first group.

7.2.3.4 Initial Training and Exposure: It seems very unlikely that the subjects would begin both tests approximately calibrated to the task by chance. Nevertheless, in both T40 and T42 (figures 7.3 and 7.4, respectively) the mean value of k is essentially constant and equal to the apparent asymptotic value during the initial training (G1-2). Part of the explanation, of course, is that all of the subjects are not calibrated at the start.

Instead, some subject may be too high in their initial depth estimates while others are too low. On average, this hypothesis suggests that the value of k should be equal to the correct-response value, unless the subjects begin the test with a common bias. Such is the case during the exposure (G5-6), where the subjects are already adapted to another value of k. And, in agreement with the hypothesis, there appears to be a learning curve in that case (see the figures).

A common bias is also possible at the start of the test, because each test is the second test taken by the subject about half of the time. Hence an initial value of k between 0.5 and 1.0 might be observed. There is little evidence of this effect in the data however.

Over time, one also expects that the subjects will converge on the correct-response value during the training (perhaps with some bias). Hence, the variance across the subjects should be greater at the start of training than at the end. Surprisingly, this effect is not observed either during the initial training or the exposure. In order to investigate this issue, the raw k values[18] for each of the subjects are presented in figures 7.5 to 7.8, for the initial training and the exposure groups in both tests.

Four things are apparent in the individual data. One, the random changes in k are at least as big as the mean change one hopes to observe. An explanation, then, about why the inter-subject variance does not appear to get smaller during the training, is that the change may be small compared to the overall level of noise. Two, the subjects do begin the training with different initial values of k, some high and some low. Thus the initial hypothesis is at least partially confirmed, although a third observation about the data is that, in many cases, the subject does happen to start with k near the final asymptotic value. In those cases, as one would expect, k tends to stay at about the same level throughout the training. Finally, one may observe that there is very little difference between the initial training and the exposure data, beyond the fact that there is a more consistent initial bias in the exposure (as already observed for the mean data, see figures 7.3 and 7.4).

7.2.3.5 Hypothesis: The third observation in the previous paragraph raises an interesting question. Do subjects calibrate themselves very quickly at the start of the test, perhaps in the first few trials, so that only the first one or two intervals of trials are affected? One hypothesis is that the subjects may initially establish criteria on the decision axis in a more or less conscious way. The accuracy of the initial calibration would be limited by factors related to the overall level of noise in k. Specifically, after only a few trials, one would expect k to fall within a range about the correct-response value roughly equal to plus or minus the mean change in k (call it Δk). Via this mechanism, subjects could go from being completely untrained to a level of performance where k is roughly equal to the correct-response value (allowing for possible bias) in a very short time.

---

[18]In other words without the processing to remove outliers, used for figures 7.3 and 7.4.

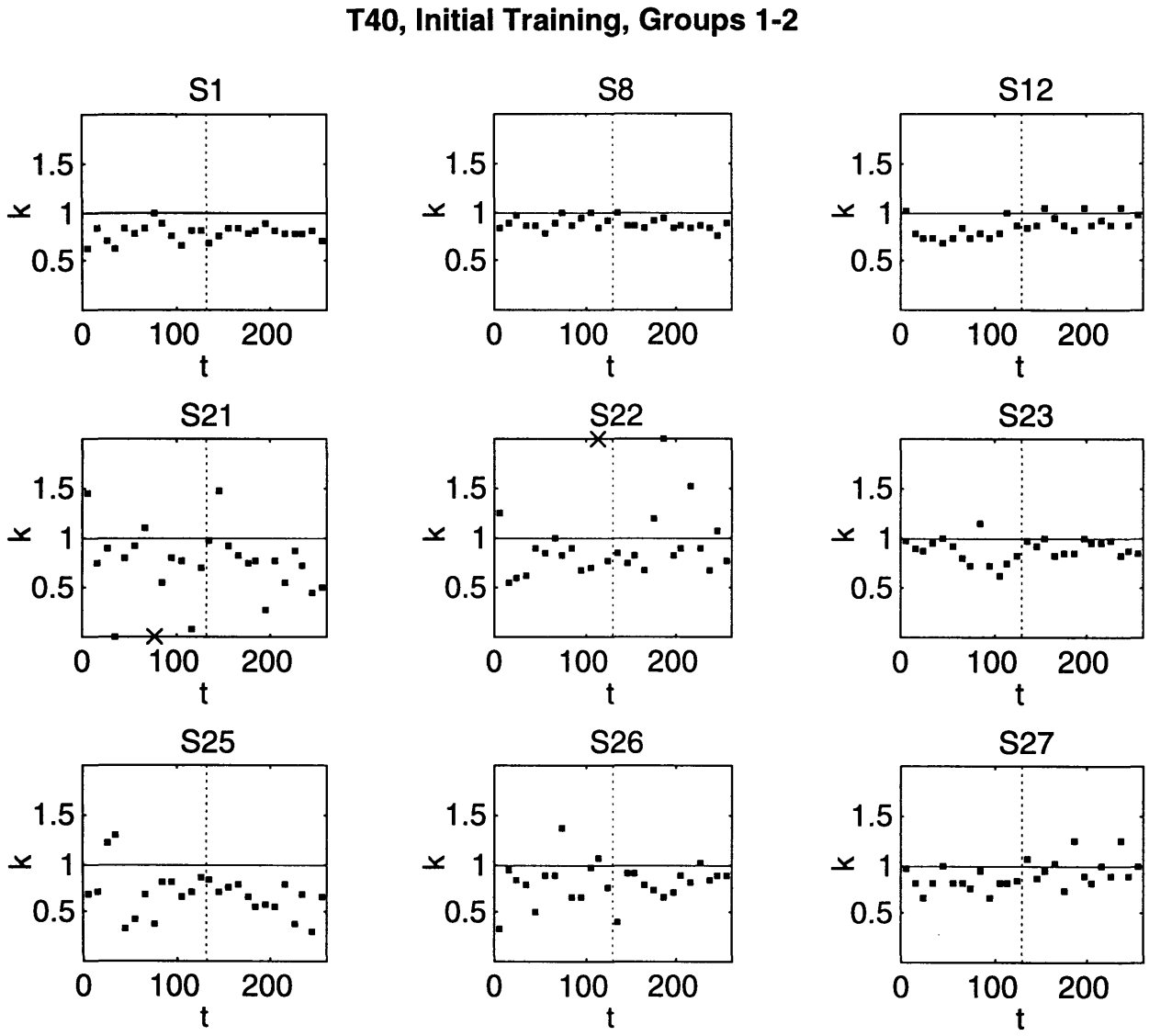


Figure 7.5. Individual 10-trial estimates of k for each subject in test T40 (v2.102) during the initial training, groups 1 and 2. Note that "x" indicates an estimate that is outside the bounds of the plot. No value is shown when the calculation failed.

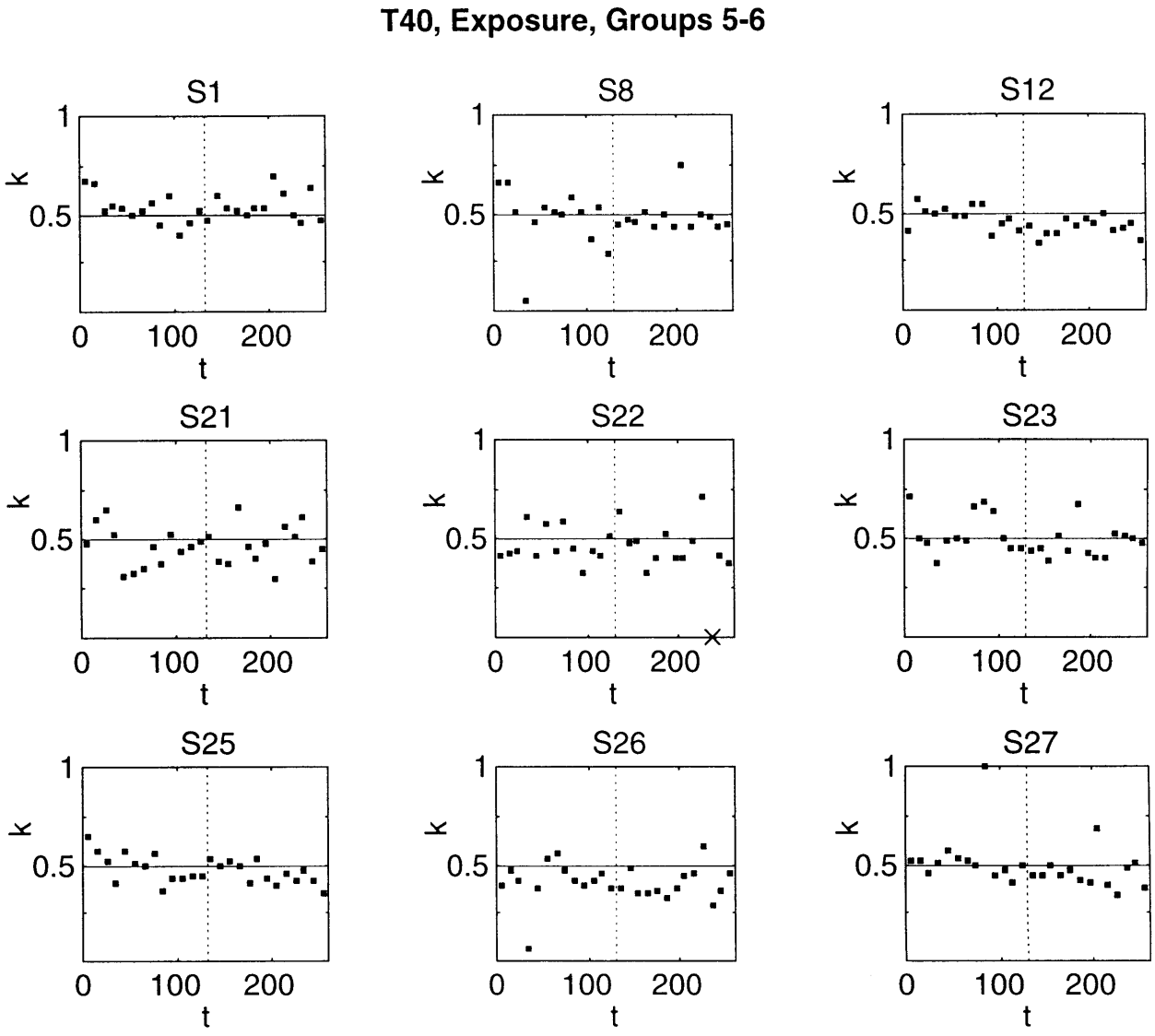


Figure 7.6. Individual 10-trial estimates of k for each subject in test T40 (v2.102) during the exposure, groups 5 and 6. Note that "x" indicates an estimate that is outside the bounds of the plot. No value is shown when the calculation failed.

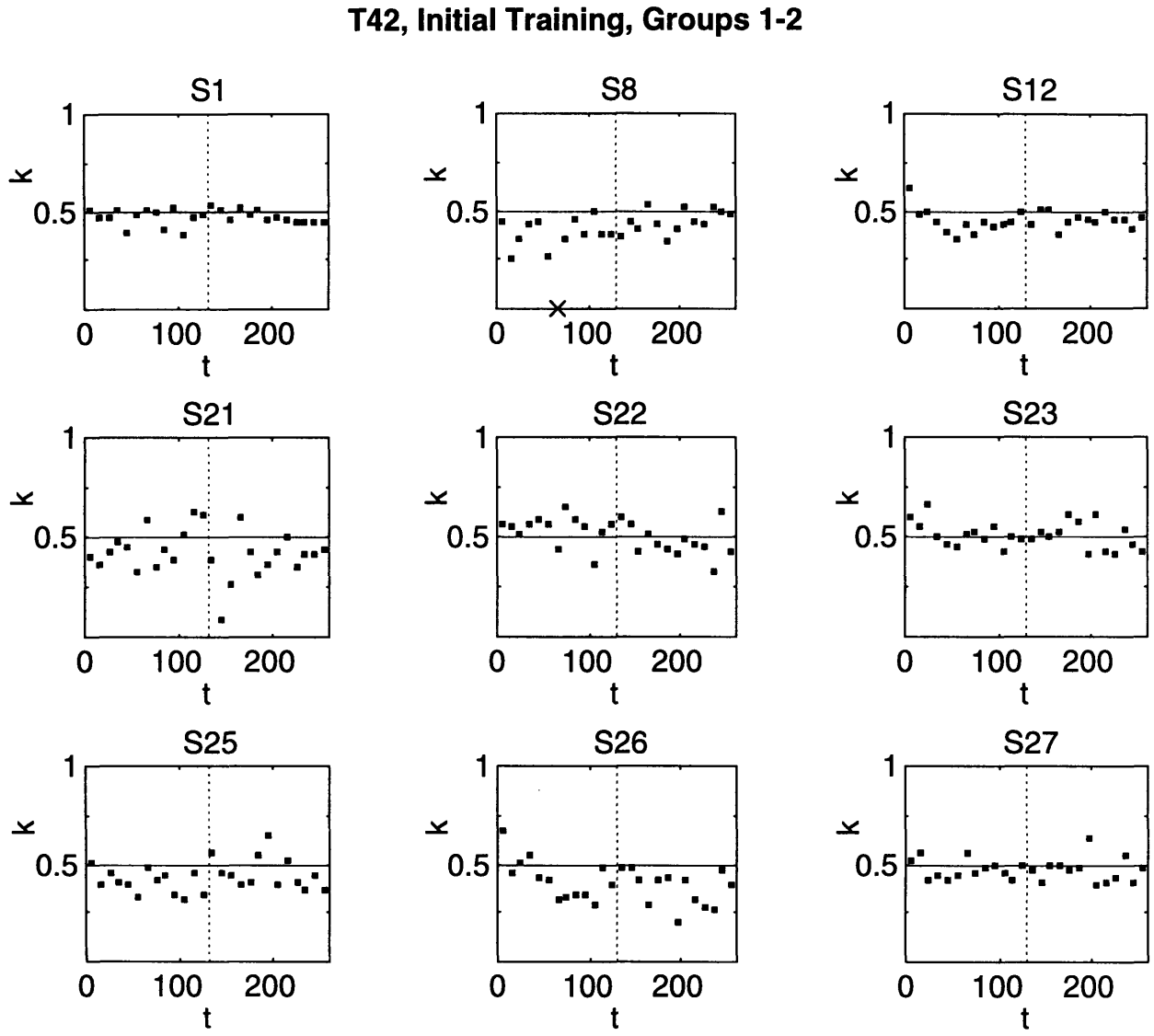


Figure 7.7. Individual 10-trial estimates of k for each subject in test T42 (v2.102) during the initial training, groups 1 and 2. Note that "x" indicates an estimate that is outside the bounds of the plot. No value is shown when the calculation failed.

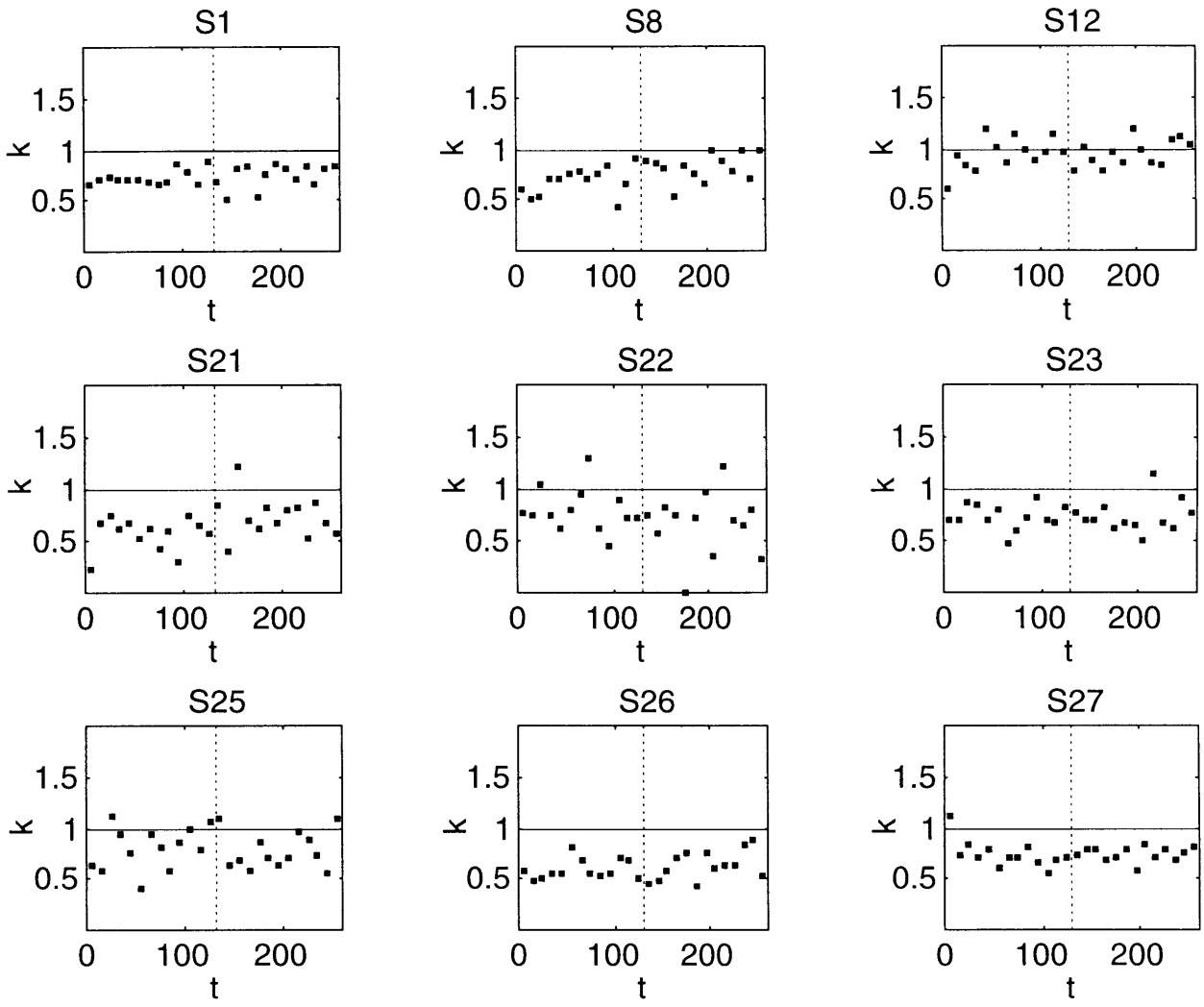


Figure 7.8. Individual 10-trial estimates of k for each subject in test T42 (v2.102) during the exposure, groups 5 and 6. Note that "x" indicates an estimate that is outside the bounds of the plot. No value is shown when the calculation failed.

After the initial level of performance is reached, however, further improvement would require a different mechanism. Specifically, if the subject continues to readjust his criteria based on each new value of Q (see equation 7.1), k would simply fluctuate about the asymptotic value within plus or minus Δk. To improve, instead of following the noise, the subject would need to follow his or her nose (by this I mean, "follow his or her own self-directed course"). In other words, after a brief period of conscious learning where the subject becomes roughly calibrated, he or she must settle on a particular set of criteria.

Once the criteria are established, further improvement in the task would require a mechanism that could integrate over a large number of trials in order to filter-out the noise. In other words, the subject would only change the position of the criteria based on the results of many trials. Such a mechanism might be unconscious and would, necessarily, have a slower time constant than during the initial period of learning.

Presumably this is what is happening during the exposure where, despite direct feedback, it takes time for the subjects to adapt to the new correct-response value of k (see figures 7.3 and 7.4). Of course, the more rapid form of learning would come back into play anytime the subject is presented with clear evidence that the criteria are in the wrong place. For example, the perception that a given stimulus is well outside the normal bounds, could cause the subject to rescale the decision axis very quickly. Thus, one might explain how there can be a rapid change in the mean value of k in group 4 of test T40 (where there is no feedback), followed by a much slower change in group 5 (where there is feedback).

7.3 Evaluation of σ_g and k_g

In this section, as discussed in section 7.1.3, each test (T40 and T42) is partitioned into sets of trials where σ and k are assumed to be approximately constant and the mean values of σ and k are estimated for each interval. The main result of the analysis is presented and discussed first, leading to a number of observations about how the two parameters vary over the course of the test. Of particular interest in this discussion is the suggestion that σ might vary independently of k (see section 7.3.2.1, 2). A secondary result, derived from the same calculations, is presented next in section 7.3.3. The latter investigates the assumption of the model that σ is the same for all stimulus depths.

7.3.1 Procedure and Results

Table 7.2 presents the estimated values of σ_g and k_g for the 8 main groups of trials in each test (see section 6.3.1.3). The reported values are the mean of the individual subject values that are observed (see section 7.1.3). The table also includes the corresponding standard deviations of each quantity (in parentheses) to show the inter-subject variability.

Table 7.2. σ_g (dau) and k_g (dnu/pixel) with (standard deviation across subjects) for T40 & T42.

Group	T40 σ_g	T40 k_g	T42 σ_g	T42 k_g
1	0.96 (0.33)	1.00 (0.21)	1.5 (0.44)	0.50 (0.05)
2	1.0 (0.32)	0.89 (0.11)	1.4 (0.48)	0.49 (0.04)
3	0.94 (0.43)	0.95 (0.18)	1.4 (0.53)	0.48 (0.07)
4	1.1 (0.43)	0.71 (0.20)	1.3 (0.60)	0.54 (0.07)
5	1.4 (0.22)	0.54 (0.06)	1.3 (0.49)	0.79 (0.13)
6	1.5 (0.52)	0.51 (0.05)	1.2 (0.44)	0.87 (0.11)
7	1.5 (0.52)	0.50 (0.06)	1.0 (0.35)	0.95 (0.27)
8	1.3 (0.41)	0.60 (0.14)	1.5 (0.50)	0.63 (0.13)

Figures 7.9 and 7.10 present the results of table 7.2 in graphical form, with the given quantities plotted against trial number. Figure 7.10 differs from 7.9 in that it presents σ_{Mg} (i.e., the directly observed standard deviations of the responses, see section 7.1.3.3) rather than the calculated values of σ_g. The values corresponding to the 8 main groups of the tests (table 7.2) are represented by solid horizontal line segments. Each line segment extends over the range of trials corresponding to the given group (note that the vertical lines indicate group boundaries). The dotted horizontal lines show the correct-response values in the k-plots and provide arbitrary references in the σ-plots.

Also presented in the figures are the results of a similar analysis where each group is divided approximately into thirds.[19] The resulting 24 estimates of each quantity are represented by square dots (▪) in the figures. The resulting subgroup estimates are inherently less reliable than

[19]Specifically each group is partitioned into the following sets of trials: [1,43], [44,87], [88,130].

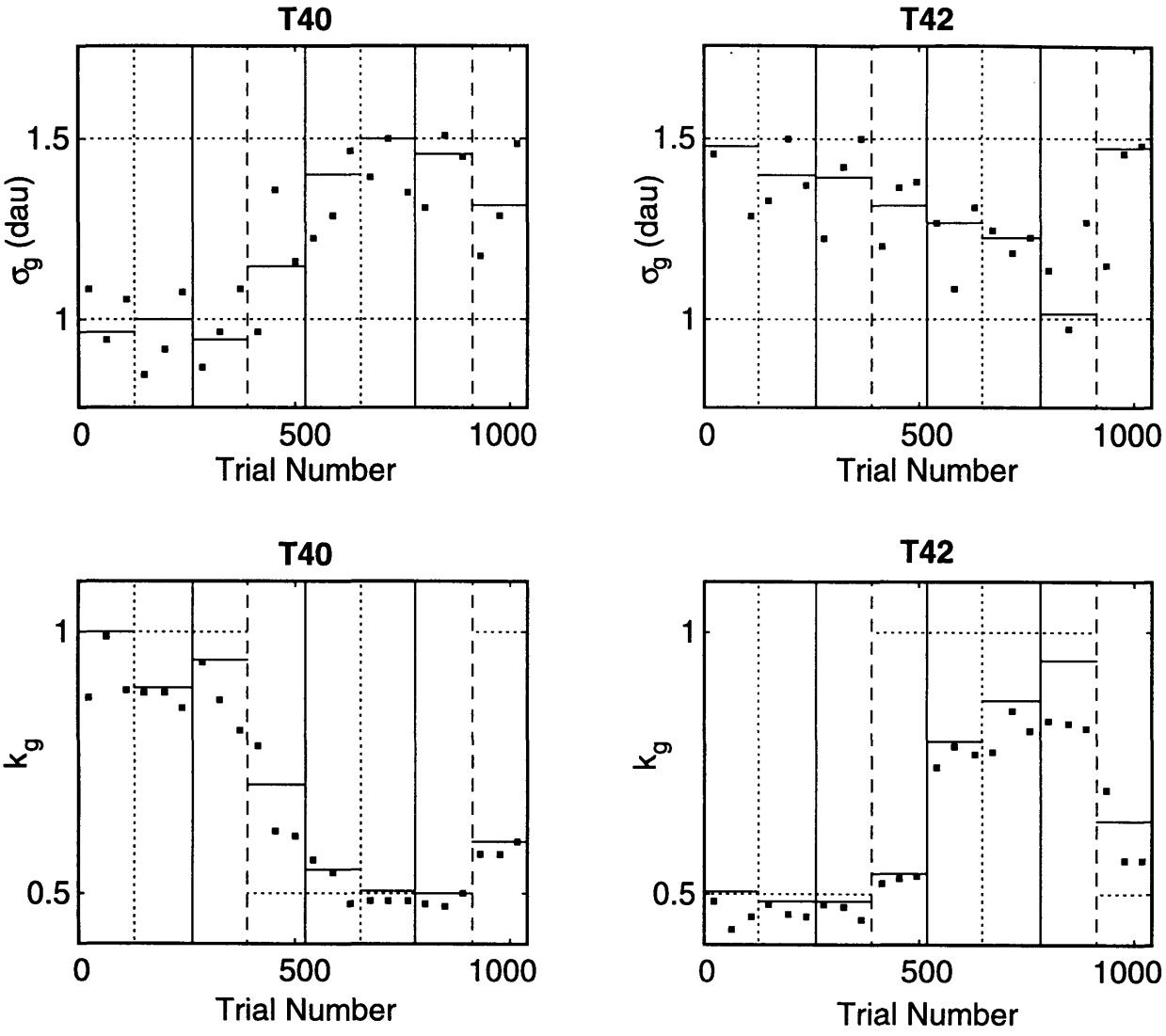


Figure 7.9. Mean group and subgroup estimates of σ_g (dau) and k_g (dnu/pixels) across all subjects in the second identification experiment (tests T40 and T42, v2.102).

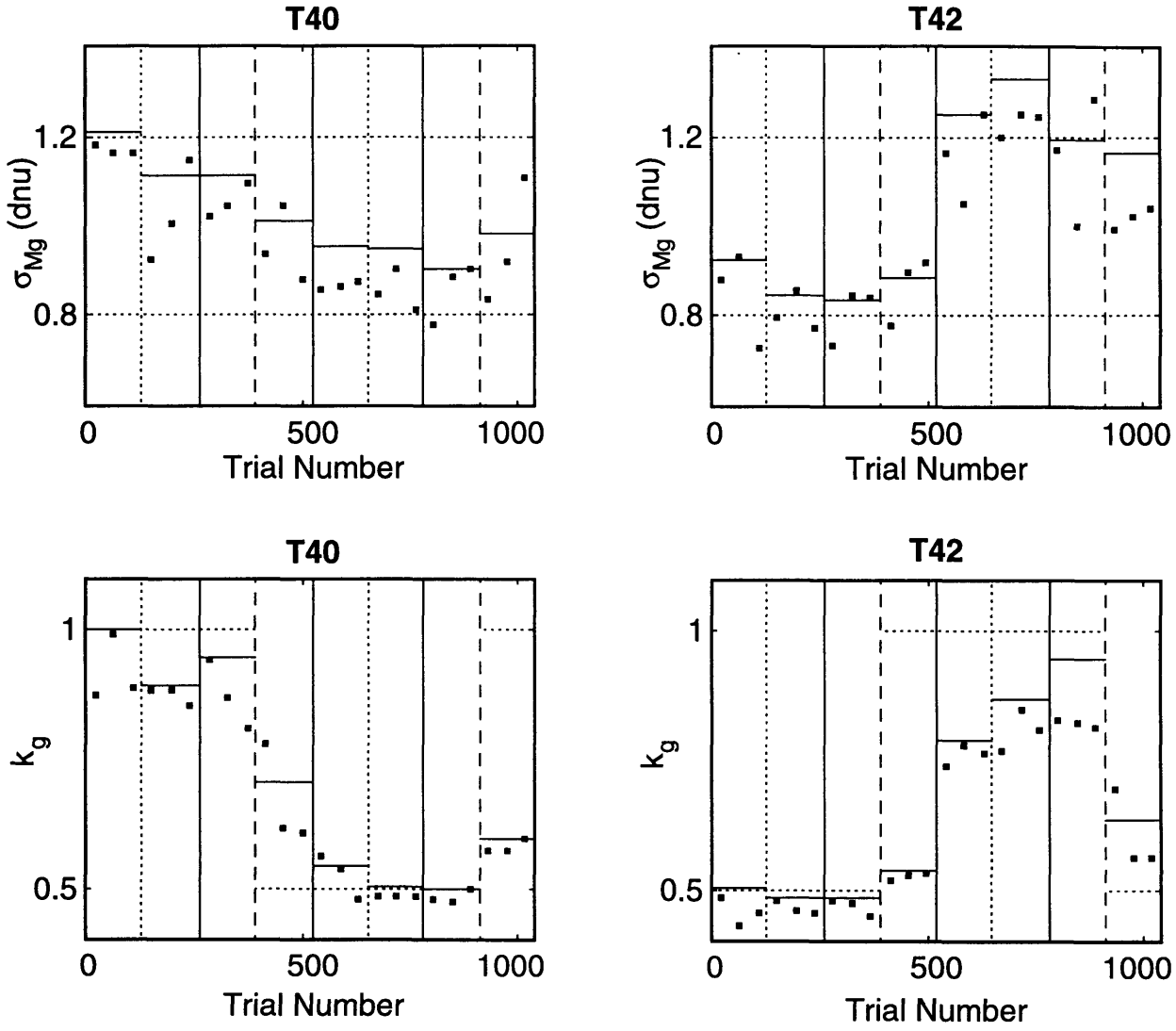


Figure 7.10. Mean group and subgroup estimates of σ_{Mg}(dnu) and k_g (dnu/pixels) across all subjects in the second identification experiment (tests T40 and T42, v2.102).

the full-group estimates because they involve fewer trials. Indeed, the number of trials is further reduced in the subgroup estimates of σ, because stimuli presented three or fewer times in an interval are not included in the calculation.

Another difference between the two σ-estimates is that a lower bound is used in the subgroup analysis. Specifically, in cases where all of the responses are the same for a given stimulus value (resulting in $\sigma_{mi} = 0$), a minimum value is substituted by assuming that all but one of the responses are at one depth, while the other differs by one dnu. This is considered a lower bound on the resolution of the measurement. The affect of the correction is to increase the estimated values of σ slightly, bringing the group and subgroup σ-estimates closer together, but otherwise leaving the general character of the data unchanged.

Finally, figure 7.11 presents the results of an analysis where the estimated values of k are corrected for bias related to the finite size of the response set. The figure, which is otherwise similar to the previous k-plots, superimposes the corrected data for all three analyses: group (—), subgroup (o), and the 10-trial estimates (•) discussed in section 7.2. The correction is based on the fact that sensations (Q) outside of the nominal bounds on the decision axis (C_{min} and C_{max}) are not weighted as heavily, in the estimate of k, as sensations in the middle of the response range. Details of the calculation are presented in appendix E.

Note that using the corrected values of k in the calculation of section 7.4, does not significantly alter the value of G. Also, while the correction does bring the different estimates of k closer together, the overall effect is quite small (compare figures 7.3, 7.4, and 7.9 with 7.11). The correction also has little affect on the predictions of section 7.5. Thus, while the corrected data (figure 7.11) is useful as a best-estimate for the qualitative analysis in section 7.3.2, the correction is not used in any subsequent calculations.

7.3.2 Observations and Discussion

One can make several observations about the results that are presented in figures 7.9, 7.10, and 7.11. Specific comments related to the behavior of σ and k are presented in the following subsections (7.3.2.1 and 7.3.2.2, respectively). First, however, the discussion begins with two general observations.

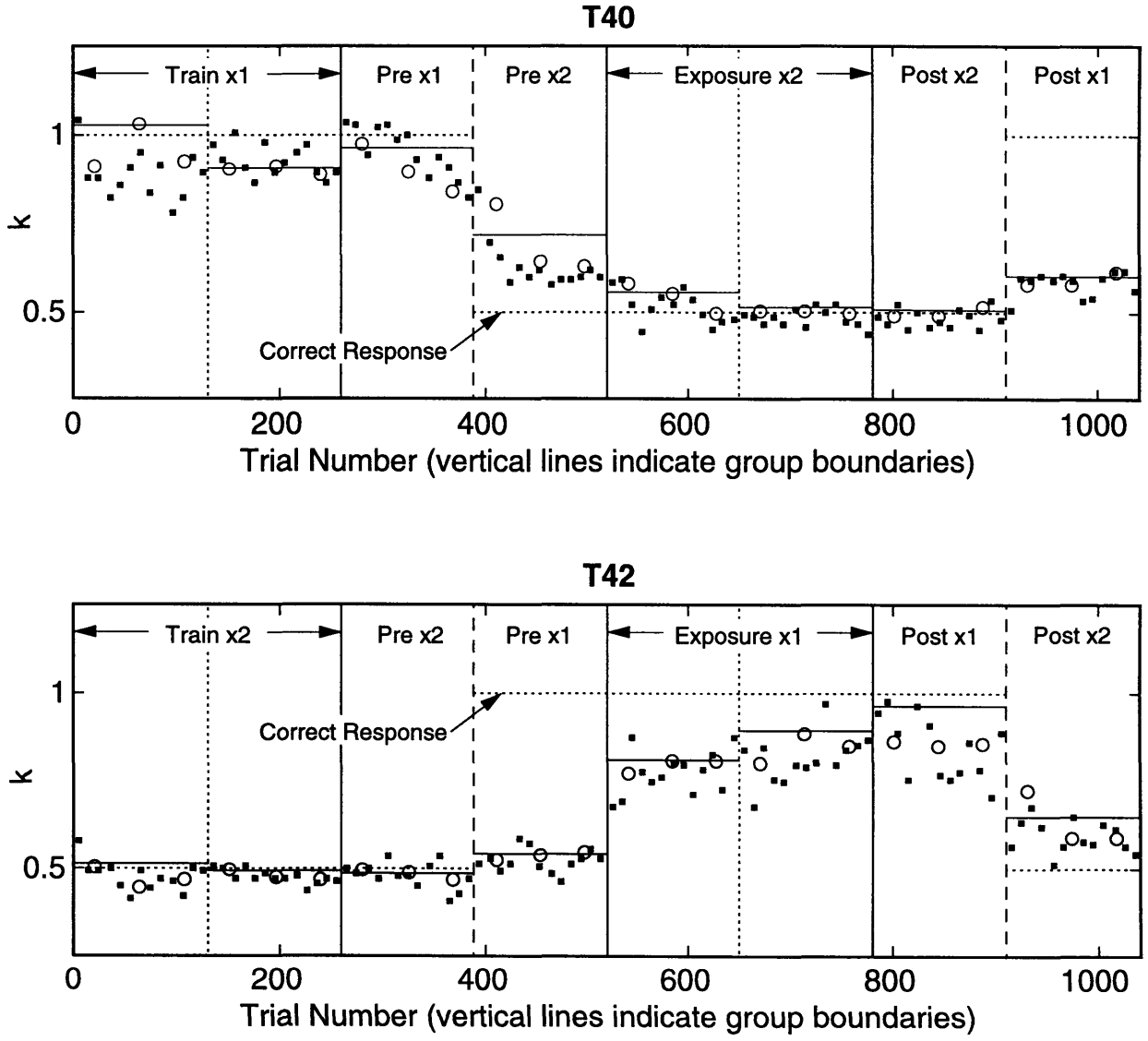


Figure 7.11. Corrected k estimates averaged across all subjects in the second identification experiment (tests T40 and T42, v2.102).

The first general observation is that the two sets of estimates (group and subgroup) appear to be consistent with one another, allowing for noise, although the subgroup estimates tend to be smaller than the group estimates. One exception to the latter tendency is the fact that the two σ_g-estimates in figure 7.9 appear to be approximately the same.[20] The k-estimates are also consistent with the general trends observed in figures 7.3 and 7.4 for the 10-trial estimates (as shown in figure 7.11).

The second general observation is that, while the data suggests many interesting trends, it is hard to have confidence in all of the fine details. For example, as discussed previously in section 7.2.2, there appears to be a downward trend in k_g in group 3 of test T40 (apparent in the subgroup analysis). One explanation for this (if it is a real trend) is that the subjects become more uncertain about their responses in the absence of feedback. And, as a result, they tend to give more responses in the middle of the range. In agreement with the hypothesis, the value of σ_{Mg} increases in the group (see figure 7.10). On the other hand, the change observed in σ_{Mg} in group 3 is small compared to the change that is observed in the previous group, where one has every reason to expect the subject's performance is stable.

With this general warning in mind, the following discussion identifies and seeks to explain the gross trends observed in the data. In order to help evaluate the reliability of the results, counterexamples (that might indicate a particular observation is purely a matter of chance) are also presented when they are known to the author.

7.3.2.1 σ analysis: There are three general σ-results to be discussed. The first is related to the value of σ on the decision axis and, hence, figure 7.9 (where σ_g is presented in dau). The other two results are independent of the units, however, and are best interpreted by looking at σ_{Mg} (figure 7.10). This is true because σ_{Mg} does not include the additional noise associated with k_g that is present in σ_g (see equation 7.11).

1) Referring to figure 7.9 (and table 7.2), one may observe that σ_g is approximately equal to 1 dau in the first 3 groups of T40. In other words, in these groups, the variance of the distributions on the decision axis

---

[20]This happens because the ratios of the two estimates of σ_{Mg} and k_g are approximately equal (see equation 7.11). This might be purely a matter of chance or it might indicate a common mechanism in the bias of the estimates of σ_{Mg} and k_g.

appears to be entirely due to sensation noise ($\sigma \approx \beta$) and memory noise is essentially zero ($\gamma \approx 0$). See sections 7.1.1, 7.1.2 (equation 7.3), and 2.4.2 for the necessary background.

Similar results are also observed in group 7 of T42. Indeed, σ_g is less than 1 dau in two group estimates and seven subgroup estimates. The latter result, which corresponds to $\sigma < \beta$, contradicts the present model because one assumes that the total variance is the sum of the variance due to sensation noise plus the variance due to memory noise (i.e., $\sigma^2 = \beta^2 + \gamma^2$). The most reasonable interpretation, however, is that the model is basically sound and that σ_g is less than β as a result of experimental error. Specifically, β might be underestimated in the determination of α in chapter 5 and the present analysis probably does underestimate σ_g as discussed in section 7.1.3.2. Also, the observed values of σ_g are only less than β by a fraction of one standard deviation, as shown in table 7.2. Also note that the smallest values of σ are observed when the effective stimulus range is smallest and k is largest (see equation 7.6), in accordance with the model.

Although the preliminary theory of intensity resolution and, hence, the preliminary adaptation model predict that σ will be greater than β in the experiment, it is not too surprising that the two would be approximately equal in the middle of the stimulus range in some cases. This is true because the subjects are judging the position of the variable target relative to a reference at 0 dnu. Consequently, the subject's task for depths near zero is similar to the task in the discrimination experiment of chapter 5 (where β is effectively determined). And one might expect the resolution to be approximately the same in both cases.

This situation is explained by the perceptual anchor model of context-coding (Braida et al. 1984), which might be used in a future refinement of the present adaptation model. According to the perceptual anchor model, however, one also expects σ to increase as the variable target gets further from the reference. Hence, the mean value of σ across all of the stimulus depths should still be larger than β.

The fact that this is not observed may be partly explained by an additional source of bias in the estimate of σ, not previously discussed. Specifically, the approach used to deal with edge effects in the analysis preferentially selects the middle of the stimulus range (see section

7.1.3.2). Thus, if σ is smaller in the middle of the range, the estimated mean value is low.

2) The next σ-effect is apparent in figure 7.10. Specifically, beginning at the start of the tests, one may observe that the variance of the subjects' responses decreases from group 1 to group 2 in both T40 and T42. This observation is based primarily on the group estimates although the subgroup estimates are consistent on average. The variance also stays roughly the same from group 2 to group 3.

This is a perfectly reasonable result. As the subjects are trained they become more consistent in their responses.[21] Then they stay at the same level in group 3, where the only change in the experimental conditions from the previous group is that feedback is no longer given. On the other hand, there is very little change in σ_{Mg} during the exposure of T40 and the *opposite* trend is observed during the exposure of T42. Also, σ_{Mg} changes in the group immediately after the exposure in both tests (group 7).

Certainly, one explanation for the difference is that the changes in σ_{Mg}, during and immediately after the training and exposure groups, are simply the result of chance. On the other hand, given that the observed trend in σ_{Mg} during the first three groups is very reasonable, perhaps the differences in the later groups (5-7) can be explained in terms of the state of adaptation. One must be careful here, however, because the apparent effect is not explained by the preliminary adaptation model. Specifically, k appears to be approximately constant during the first three groups, so that, if σ is changing, then the state of adaptation is not fully described by k (as assumed in the model).

One possibility might be to modify the model by dividing memory noise into *task* and *learning* components. Task noise would be the asymptotic value of memory noise (presently represented by G^2R^2), while learning noise would represent additional uncertainty that is present when the task is unfamiliar (or, to be more specific, that is present until the subject settles on a fixed set of criteria). Learning noise is, thus, an additional free parameter (related to adaptation) that goes to zero as the subject learns.

Assuming that the model can be modified in some such way, the differences in the observed trends of σ_M in groups 1-3 and 5-7 might be

---

[21]Note that this is individual consistency, not across subjects!

explained by changes in k. The two quantities are, of course, interrelated (see equation 7.14). Specifically, in contrast to the first three groups, it may be observed in the figures that k decreases during groups 5-7 of T40 and it increases during groups 5-7 of T42. Referring to equation 7.14, it may be seen that the model predicts that an increase in k could counteract the proposed trend, where σ_M decreases independently of k as the subject learns (assuming that the observed trend in σ_M would otherwise be the same in the two sets of groups). This is encouraging, given that the performance in T42 differs most from the proposed trend. On the other hand, the decreasing value of k in T40 should cause σ_M to decrease even more, in groups 5-7, than during groups 1-3. The fact that σ_{Mg} changes less in groups 5-7 of T40, than it does in groups 1-3, makes the "experimental noise" explanation (i.e., that this is all just the result of chance) more appealing. Also, it is not immediately clear why σ_{Mg} would change between groups 6 and 7, for any reason other than noise.

One possible explanation is that the subjects have a harder time adapting during the exposure than during the initial training (as proposed previously in section 6.3.3.3). As a result, given that the mean locations of the criteria on the decision axis are still in flux (as indicated by the changing value of k), it might not be possible for the subjects to make the finer level of adjustment observed in the first three groups. For example, the learning noise (proposed above) might not start to decay until k becomes relatively constant.

Indeed, some response inconsistency may be advantageous during the initial stage of adaptation (where k is changing). Specifically, k will only change if the subject's criteria are free to move on the decision axis, so adaptation might be improved when the positions of the criteria are more uncertain. Further, the drop in σ_{Mg} that is observed between groups 6 and 7 might be explained by saying that, once feedback is no longer given in the test, the subject is forced to adopt a more consistent response strategy. In other words, the learning noise simply goes to zero. While this is all pure speculation (and it is proposed only to suggest possible directions for future research), the general explanation (i.e., that the subjects do not adapt as fully during the exposure as they do during the initial training) is consistent with the next (and last) σ-result that is apparent in the figures.

3) The last apparent σ-effect in the experiment is an asymmetry in the results of the two tests. Specifically, the range of σ_{Mg} is less in T40 than in T42. This is significant because the only difference between the two

tests is the order of the groups. The effect appears to be due to the fact that σ_{Mg} is larger, for the same test conditions, during the second half of the test. In other words, subjects are relatively more consistent in their responses when they are only trained one way, during and immediately after the initial training. While retraining (i.e., the exposure) immediately after the initial training does not achieve the same level of performance, at least in the same number of trials.

This idea is consistent with the general hypothesis (first proposed in section 6.3.3.3) that the subjects do not adapt as fully during the exposure as they do during the initial training. It is also consistent with the idea, proposed in the second part of this section (2), that there may be an additional parameter of adaptation associated with *learning noise*. The latter idea is left to future research while the former hypothesis is explored further in the following section (see "Resolution shift between T40 and T42").

7.3.2.2 k analysis: Figure 7.11 presents the best estimates of k as a function of t. One expects all of the different types of estimates (i.e., group, subgroup, and 10-trial) to give the same result, with the common value being an unbiased estimate of the "true" value of k. This is expected because the estimates are corrected as discussed in appendix E. To a large extent the different estimates do correspond, although there are a few notable exceptions.

Learning Curves: The first apparent differences, between the different estimates of k, occur in T40, group 4, and T42, group 8. These results are consistent with the hypothesis that adaptation takes place in the absence of direct feedback, particularly (as is the case here) in a transition from x1 to x2 IPD (see section 6.2.3.5). Specifically, the subgroup estimates suggest a learning curve with the group estimates taking-on an intermediate value. The 10-trial estimates are also consistent with a learning curve, although the results for T42 are less clear. In any event, there is a change in k associated with these groups (i.e., adaptation is taking place), so the learning-curve interpretation seems like the most reasonable explanation. Similar learning trends are also apparent during the exposure (groups 5 and 6) of both tests, although the consistency of the estimates in T42 is not very good.

Noise: Indeed, there is no explanation for the differences between the estimates of k in a number of cases. Specifically, there is no clear explanation for the differences that are apparent in T40, groups 1 and 3, or T42, groups 6 and 7. Note that these are all x1 IPD groups and that the

experimental variability of k is greatest in these groups (as discussed in section 7.2.3.2). Similar results are also observed in one case when X_{ipd} is equal to 2 (i.e., T42, group 1), although the differences are scaled-down by the factor of 2. Beyond that, however, the x2 IPD estimates do appear to be more consistent than the x1 IPD estimates overall.

The latter suggests that the unexplained differences are the result of stochastic noise. While this is a very likely *explanation* , it also severely limits the interpretation of other trends in the data. Following are summaries of the trends that may be extracted from the data for the two tests.

Summary of Trends: The general behavior of k during the tests has already been described in section 7.2.2, in connection with the uncorrected 10-trial estimates. The same trends are observed here, although the magnitude of k is different. In particular, one notable difference with the previous observations is that the asymptotic values of k are closer to the correct-response value (indicated by the horizontal dotted lines) during the first three groups of the tests. Indeed, the observed value of k may not be less than the correct-response value in T42, although the mean still appears to be less than the correct-response value in T40. Similar observations may be made about the value of k after the exposure.

Indirect Feedback and Adaptation Rate: The changes in the observed value of k from group to group, give strong evidence of adaptation in the absence of direct (i.e., explicit correct-answer) feedback. Also, larger changes occur in transitions from x1 to x2 IPD (T40, group 4, and T42, group 8) than in the transitions from x2 to x1 IPD (T40, group 8, and T42, group 4), consistent with the previous discussion in section 6.2.3.5. Indeed, as discussed above, learning curves are evident in the x1 to x2 IPD transitions.

Similar learning curves are evident in the exposure groups as well, although it is difficult to asses the relative rates given the assumed level of noise in the data. Interestingly, however, the changes observed in the x2 to x1 IPD transition groups seem to happen abruptly. This is particularly apparent in T40, group 8. The latter observation does not appear to be consistent with the idea that adaptation may be modeled as a first-order exponential function. The possibility that adaptation might occur at different rates (under the same conditions) is proposed in section 7.2.3.5, although the mechanism discussed there does not seem to apply in the present case. Of course, comparing the magnitude of the change in T40, group 8, to the change in T40, group 3, on might argue that the apparent

abruptness may be nothing more than noise. Otherwise, frankly, little more can be said about it except that the results might suggest an interesting direction for future research.

Resolution shift between T40 and T42: Continuing with the discussion at the end of section 7.3.2.1, one can see that k does not quite reach the level, following the exposure (see groups 6 and 7), that it does following the initial training (see groups 2 and 3). This is, of course, consistent with the hypothesis that the subjects do not adapt as fully during the exposure as they do during the initial training. The observed differences are small, however, and may not be significant.

One can also compare the observed values of k for the symmetric group pairs of figure 6.10. Table 7.3 presents the group estimates of k_g for this purpose. Note that the group pairs in the table correspond to cases where the only difference between the experimental conditions is whether the group is presented during the first or the second half of the test. In other words, whether the group is after the initial training or after the exposure.

Table 7.3. Ratios of k_g for Symmetric Group Pairs in T40 & T42.

Group Pairs	T40 k_g/T42 k_g
T40G3/T42G7	0.95/0.95
T40G7/T42G3	0.50/0.48
T40G4/T42G8	0.71/0.63
T40G8/T42G4	0.60/0.54

It can be seen in the table that the ratios of k_g for the two tests are generally greater than one (in other words, T40 k_g > T42 k_g). The only exception being the first group pair (T40G3/T42G7) and in that case one might argue that the *group* k-estimate in the denominator is large by chance and, hence, the true ratio is greater than one. Compare the group, subgroup, and 10-trial estimates for T42, group 7 in figure 7.11. This result, in turn, corresponds to a downward shift in d' between T40 and T42, as discussed in section 6.3.3.3 (also see equation 7.6).

Central Tendency: Finally, despite the correction for bias related to the finite size of the response set (see appendix E), it appears that the asymptotic value of k may still be less than the correct-response value. Specifically, look at the estimated values of k in figure 7.11 for groups 2

and 3. Almost all of the estimates of k in those groups are less than the correct-response value, the only exceptions being a few 10-trial estimates. This suggests that there may be some other source of central tendency in the subjects' responses, in addition to the bias related to the finite size of the response set (see section 6.2.3.6).

Alternatively, it may be that the correction proposed in appendix E is incomplete. For example, the estimates of σ are probably small (see section 7.1.3.2) which would make the corrected value of k small as well. If that is the case, then the proposed mechanism of the correction might account for all of the subjects' central tendency in the experiment. Further research is required to see if this is indeed the case or, if not, what other mechanisms might be involved. The correction clearly accounts for some of the central tendency in any event.

7.3.3 σ against N

Figure 7.12 presents a plot of the mean standard deviation of the subjects' responses against stimulus depth. The plot combines the data for both tests and all of the subjects, with the ordinate calculated as follows. First, σ_{mi} is calculated for each subject, at each stimulus depth (within the allowable range, see section 7.1.3), and in each of the 8 main groups of the test (see section 6.3.1.3). Then, all of the standard deviations (σ_{mi}) are normalized (divided) by the mean standard deviation (σ_m) of the corresponding subject and group. Finally, all of the normalized values are averaged for each stimulus depth.

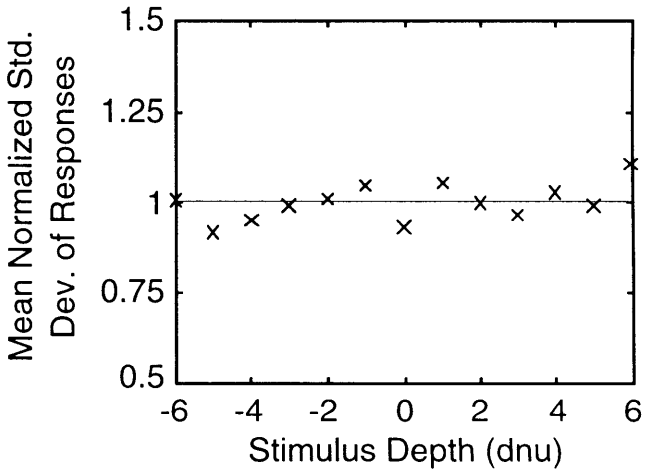


Figure 7.12. Normalized σ against N (tests T40 and T42, v2.102).

In agreement with the model, the figure shows that the normalized response standard deviation (and hence σ)[22] is roughly the same for all of the stimulus depths. Figure 7.13 presents further support, showing the

[22]Converting σ_M to σ involves multiplying by a constant that is canceled when the quantity is normalized (see equation 7.11).

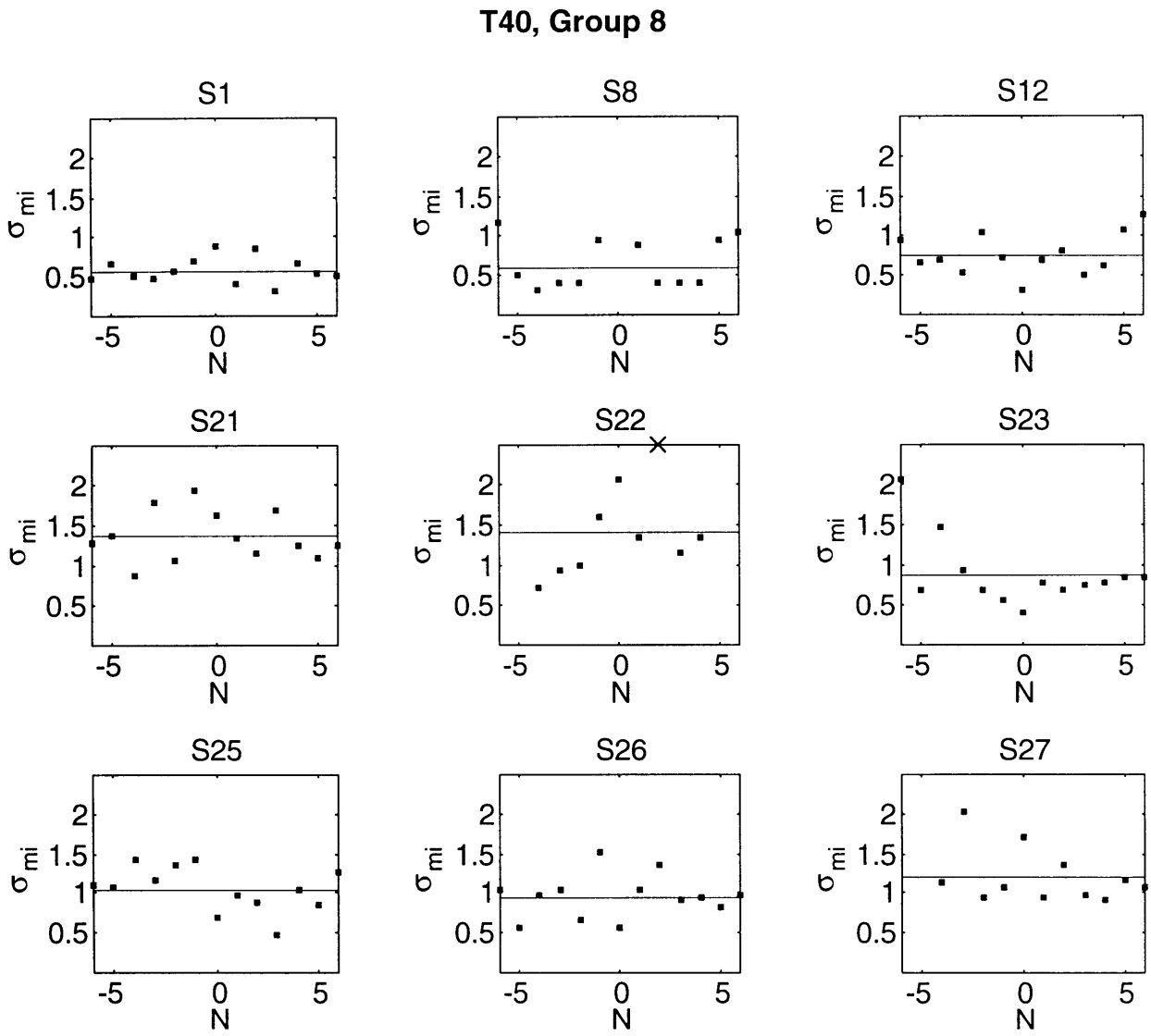


Figure 7.13. σ_M(dnu) against N (dnu) for each subject in one group (8) of test T40 (v2.102). Note that "x" indicates an estimate that is outside the bounds of the plot. No value is shown when the calculation failed.

data for individual subjects in one group (8) of one test (T40). Note that this particular trial group is selected primarily because it includes most of the stimulus depths for most subjects. One can see in the figure that, although there is more noise in the individual data, σ_{mi} is roughly constant with stimulus depth (N) for all of the subjects.

These (nominally encouraging) results may be somewhat misleading, however, because the estimates of σ are biased. In particular, one expects that the estimated value of σ will be smaller than the true value and that the bias will be largest near the ends of the range (see section 7.1.3.2). In other words, the fact that σ appears constant may mean that it actually increases near the minimum and maximum values of N. There are also theoretical reasons for expecting σ to be smaller in the middle of the stimulus range as discussed in section 7.3.2.1,1. Unfortunately, it is not possible to evaluate the effect in the present measurements without some quantitative estimate of the bias. This is left to future analysis.

7.4 Evaluation of G

The values of σ_g and k_g, presented in section 7.3, are used with equation 7.8 to determine G for each group and subgroup of trials. In the cases where the observed value of σ_g is less than 1 dau (see section 7.3.2.1,1), the analysis assumes $\sigma = 1$ in the calculation to avoid imaginary results. Consequently, in the modified cases the estimated value of G is zero.

Figures 7.14 to 7.16 present the results of the analysis for both tests in plots of G against group number, σ_g, and k_g, respectively. The cases where G is zero stand out in the plots, reinforcing the idea that these estimates are the result of experimental error as discussed in section 7.3.2.1,1. For example, compare the zero and non-zero estimates of G for the same groups, in the subgroup analysis shown in figure 7.14. Note that the magnitude of the difference is also an artifact of equation 7.8 when σ is near one (for example, look at the trend of the data in figure 7.15). Based on these observations, it seems reasonable to exclude the cases where G is zero from the remainder of the analysis. Also note that, even if those values are included in the estimate for G, the mean value is only 20% smaller and the ultimate predictions of the model in section 7.5 are essentially the same.

The horizontal line in each figure corresponds to the mean value of G, excluding the cases where G is zero. The calculated mean value is as

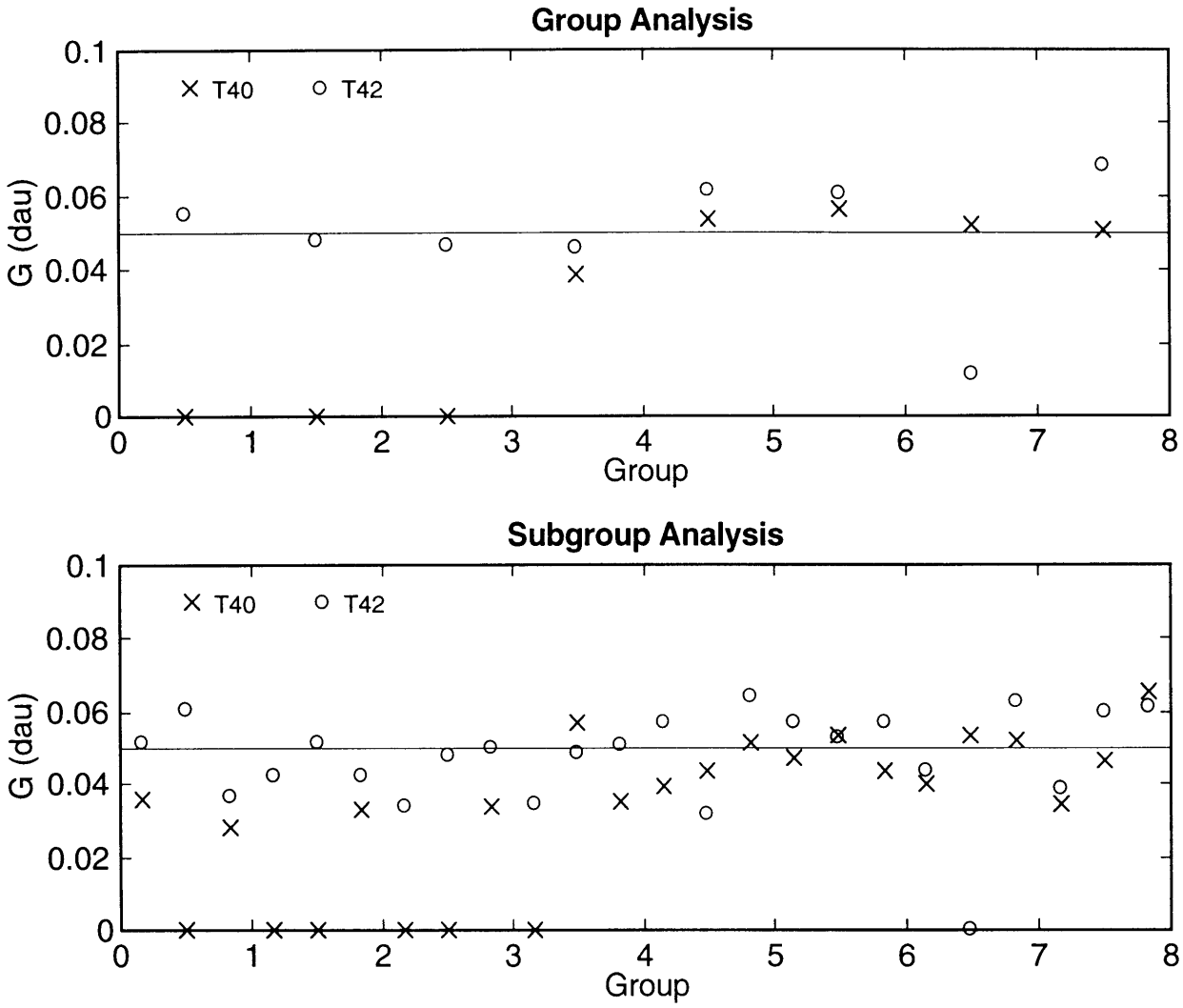


Figure 7.14. Values of G estimated from the results of the second identification experiment (tests T40 and T42, v2.102) plotted against group number.

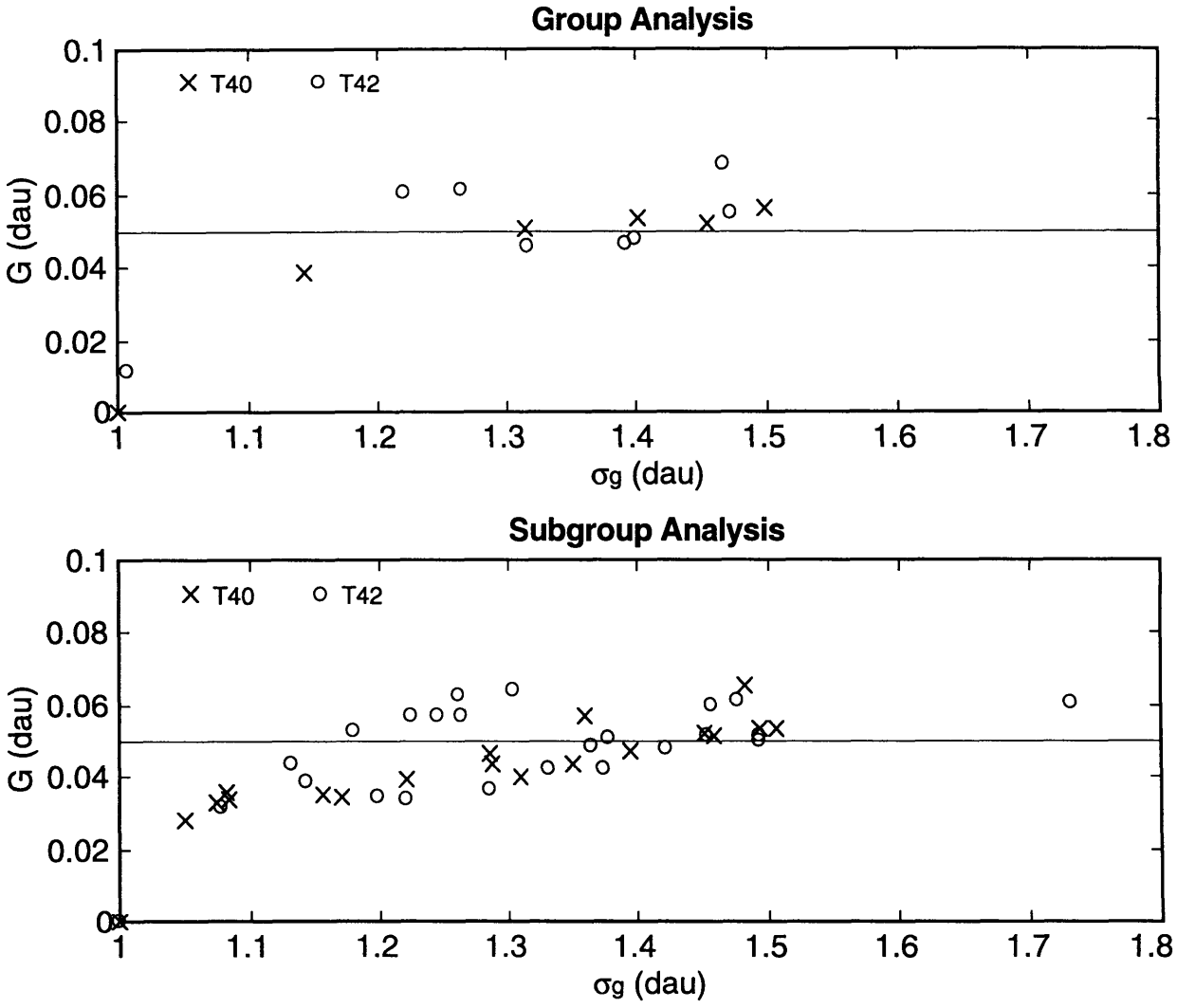


Figure 7.15. Values of G estimated from the results of the second identification experiment (tests T40 and T42, v2.102) plotted against σ_g.

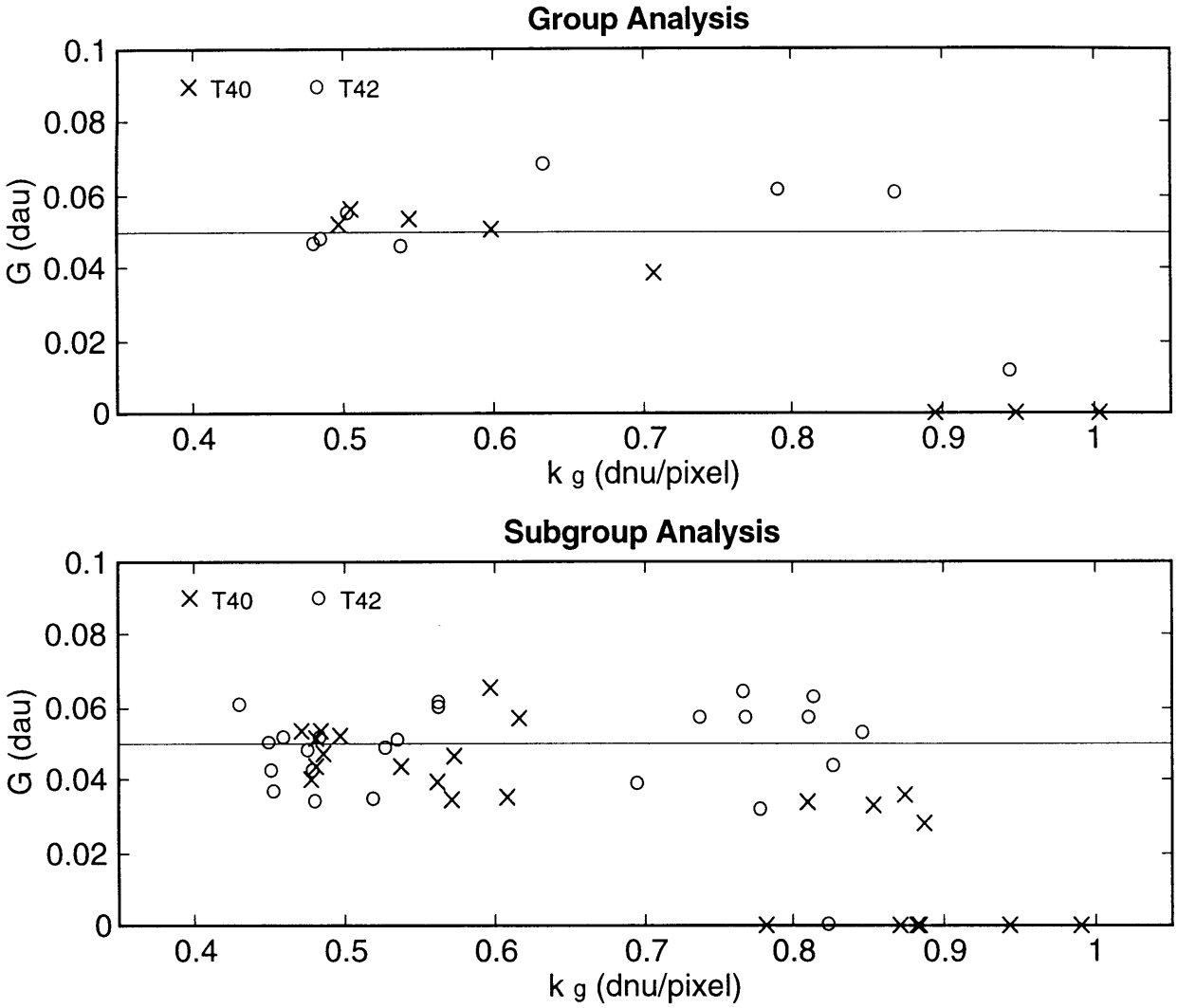


Figure 7.16. Values of G estimated from the results of the second identification experiment (tests T40 and T42, v2.102) plotted against k_g.

follows (with the standard deviation across the estimates in parentheses):[23]

$$G = 0.05\ (0.01)\ (\text{dau}). \tag{7.18}$$

In agreement with the model, the figures are reasonably convincing that G is constant in the experiment. Although, in the case where G is plotted against σ_g (figure 7.15), the slope of the best (least-squares) fit line is found to be significantly different from zero at the 1% level (excluding the G = 0 estimates). In all of the plots the slope is small, however, such that it might be modeled as zero. Also, in the other two cases (figures 7.14 and 7.16), the slope is not significantly different from zero at the 1% level (excluding the G = 0 estimates). A summary of the analysis is presented in table 7.4.

Table 7.4. Summary of Null-Hypothesis Test of G Estimates.

regressor of G	analysis	slope	p-value
group number	group	0.0002	0.9
"	subgroup	0.002	> 0.01
σ_g	group	0.07	0.007
"	subgroup	0.04	10^{-6}
k_g	group	-0.03	0.2
"	subgroup	-0.008	0.5

7.5 Predictions vs. Results

This section presents the predictions of the model together with the experimental results, previously presented in figures 6.8 to 6.10 (see section 6.3.2). The predictions are based on equations 7.7, 7.16, and 7.17, using the values of k_g that are presented in table 7.2 and the estimated value of G given by equation 7.18. Note that k_g is substituted for k(t) in equation 7.7 and that the equations are presented in sections 7.1.2, 7.1.4, and 7.4, respectively. Table 7.2 is presented in section 7.3.1.

[23]Both the group and the subgroup estimates gave the same result with one significant digit.

7.5.1 Mean Response against N

Figures 7.17 and 7.18 present the predicted mean responses against stimulus depth for tests T40 and T42, respectively. Each figure is divided into 4 plots, corresponding to the 4 *test* groups: nominal preexposure (3), altered preexposure (4), altered postexposure (7), and nominal postexposure (8). The groups are presented in order, beginning in the upper-left corner of each figure, and moving left to right and down the page as you would read text.

In all cases the model predicts the observed mean response reasonably well in the middle of the stimulus range. Near the ends of the range, however, the model overestimates the response for positive stimuli and it underestimates the response for negative stimuli. Also, the deviation at the ends, between prediction and experimental result, is larger for larger values of k. Interestingly, in most cases there is also a hint of some higher-order behavior in the experimental results. Specifically, if one connects the data points, the resulting experimental curve oscillates about the predicted curve in a pattern that is symmetric about zero. Of course, this effect may be nothing more than a slight underestimation of the slope (k') and/or experimental error. There is no apparent difference in any of these observations between the two tests.

7.5.2 Resolution against N

Figures 7.19 and 7.20 present the predicted values of d' and d* against stimulus depth for tests T40 and T42, respectively. The figures are divided into 4 plots, organized like the two previous figures. The two predicted quantities, d' and d* agree (with one another) in the middle of the stimulus range, but diverge near the ends of the range. Specifically, d* falls-off at the ends, while d' is constant across stimulus depth by assumption. Note that the amount that d* diverges from d' at the ends is larger for larger values of k. This result is expected because d* depends on the difference between predicted mean responses, which is less at the ends of the range for larger k (see equation 7.17).

When the predicted d* curve is compared to the observed mean values of d* in the figures, the predicted curves do correspond to the general trends of the results. For example, when the observed resolution is relatively constant with stimulus depth, so is the predicted curve (see figure 7.19, group 8 and figure 7.20, group 4). Also, in groups where the observed resolution falls-off near the ends of the stimulus range, the

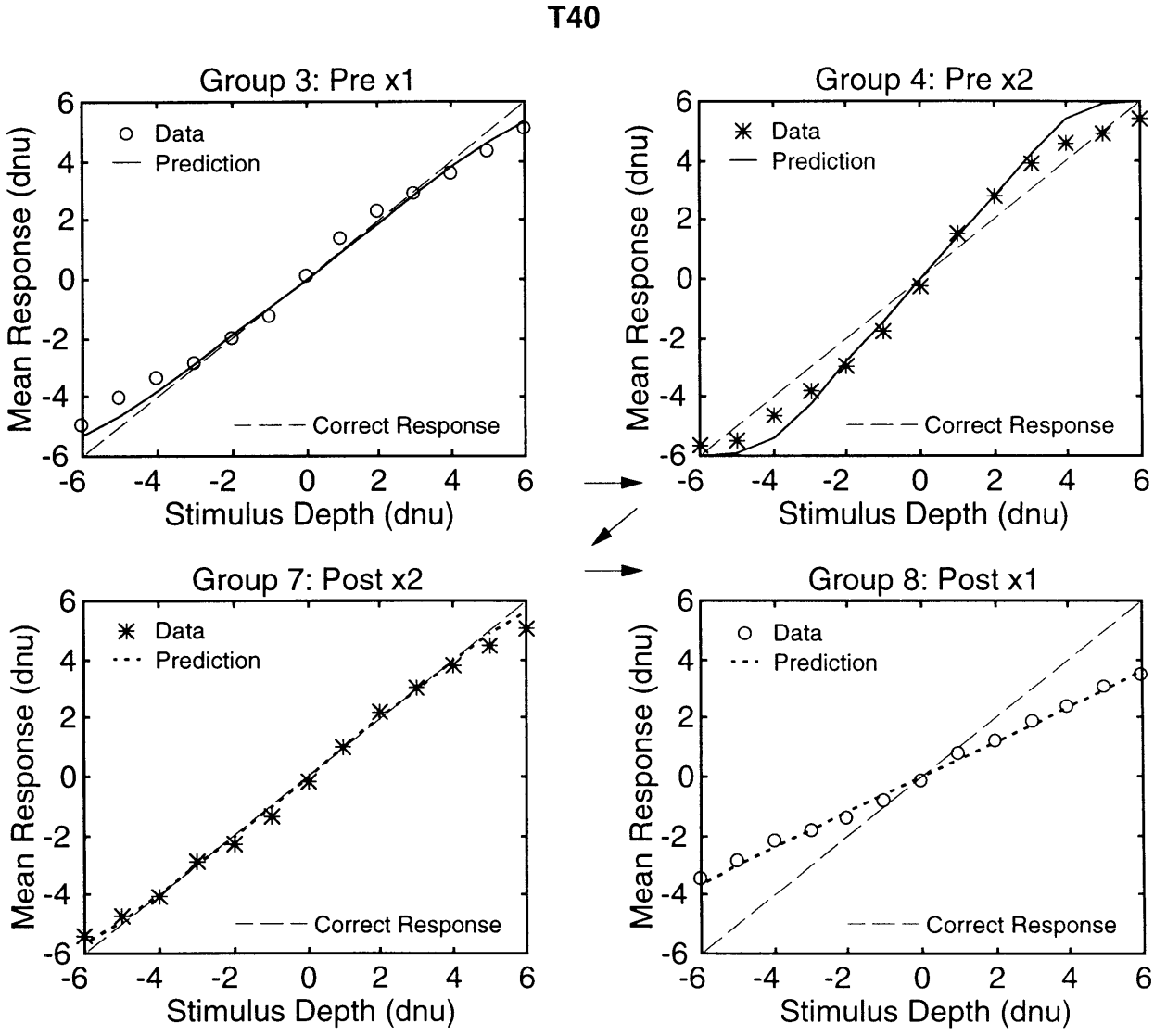


Figure 7.17. Predicted and observed mean response for the no-feedback groups of test T40 (v2.102).

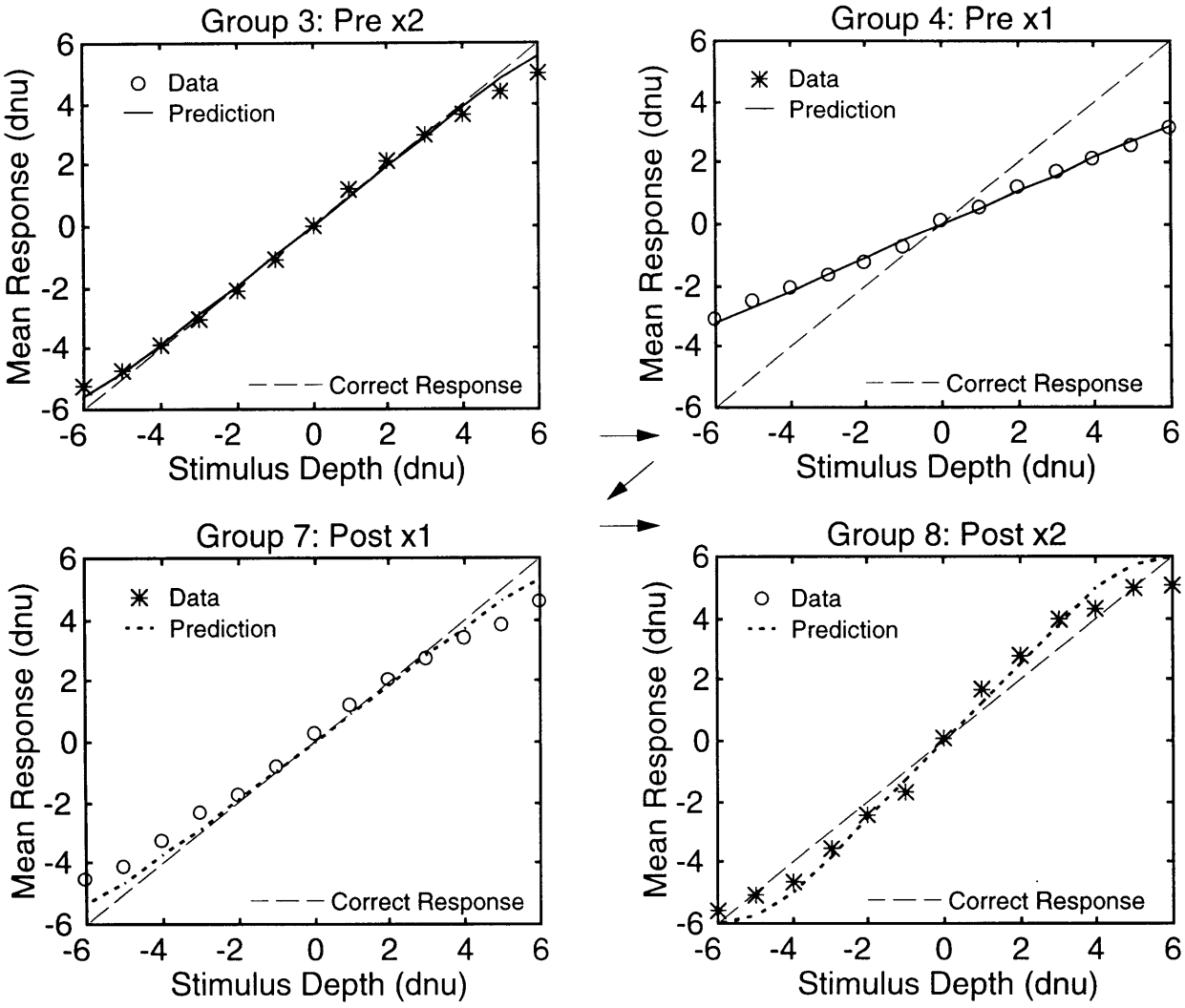


Figure 7.18. Predicted and observed mean response for the no-feedback groups of test T42 (v2.102).

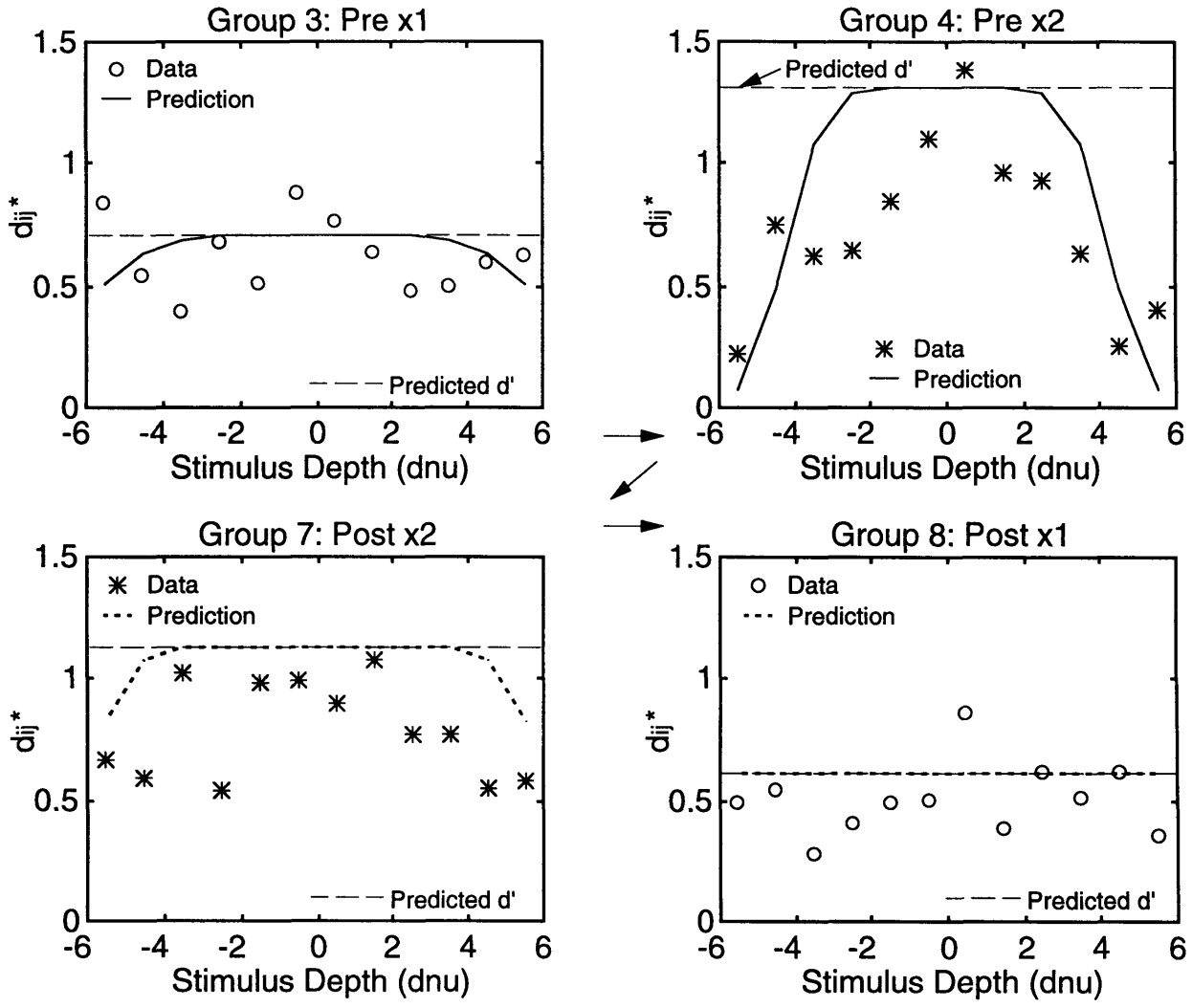


Figure 7.19. Predicted and observed resolution (d*) for the no-feedback groups of test T40 (v2.102). The predicted value of d' is also shown.

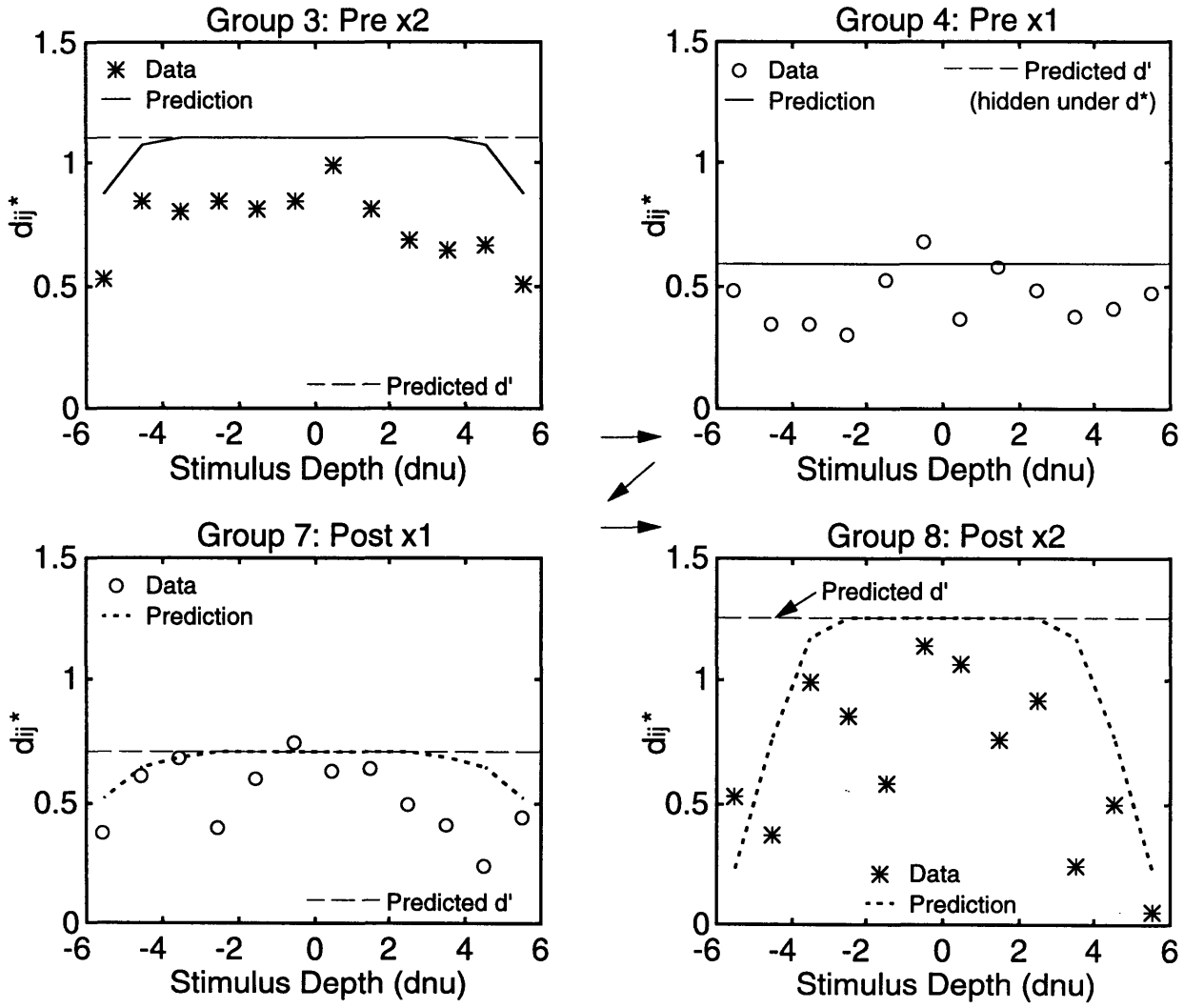


Figure 7.20. Predicted and observed resolution (d*) for the no-feedback groups of test T42 (v2.102). The predicted value of d' is also shown.

same trend is predicted by the model (see figure 7.19, group 4 and figure 7.20, group 8).

On the other hand, the predictions tend to overestimate the observed resolution. Also, noise in the observed results makes it difficult to assess the true shape of the resolution curves and, hence, the correctness of the predicted curves. For example, one could imagine that the resolution does not fall-off as predicted in figure 7.19, group 3. Also, the data seems to suggest a narrower peak in the resolution curve than predicted in figure 7.19, group 4, although the difference may simply be a matter of noise and a mismatch in the overall level.

7.5.3 Mean Resolution

Shortcomings of the resolution predictions are somewhat more clear in figure 7.21, where the mean values of d* are presented for both tests. The figure is divided into 4 plots such that the top 2 plots correspond to T40 and the bottom 2 plots correspond to T42, with the predictions on the left and the data on the right. In each graph, as in chapter 6, the mean value of d* (across stimulus depth) is plotted against the radius of the averaging interval (see section 6.3.2).

The most striking feature of the figure is that the shapes of the top curves in the plots are different; see T40, group 4 (*—) and T42, group 8 (*--). Specifically, the predicted curves are flat in the middle of the stimulus range (radius < 4), whereas the observed curves accelerate upward as the radius goes to zero. Similar trends are apparent in most of the other curves, although the effect is not as obvious because there is less change in d* with radius.

Other differences, between the predictions and the experimental results, are also apparent in the figures. First, as already observed in figures 7.19 and 7.20, the overall level of the predictions tend to be higher than the observed resolution. There is also a difference in the relative magnitude of the resolution in different groups. For example, in T42, the predicted change in mean resolution between groups 3 (*—) and 7 (o--) is much larger than the observed change, whereas the difference between groups 4 (o—) and 7 (o--) is about the same.

On the other hand, the model does successfully predict the direction of the shift between the pre- and postexposure groups, as well as the overall order of the groups in terms of mean resolution. Also in agreement with

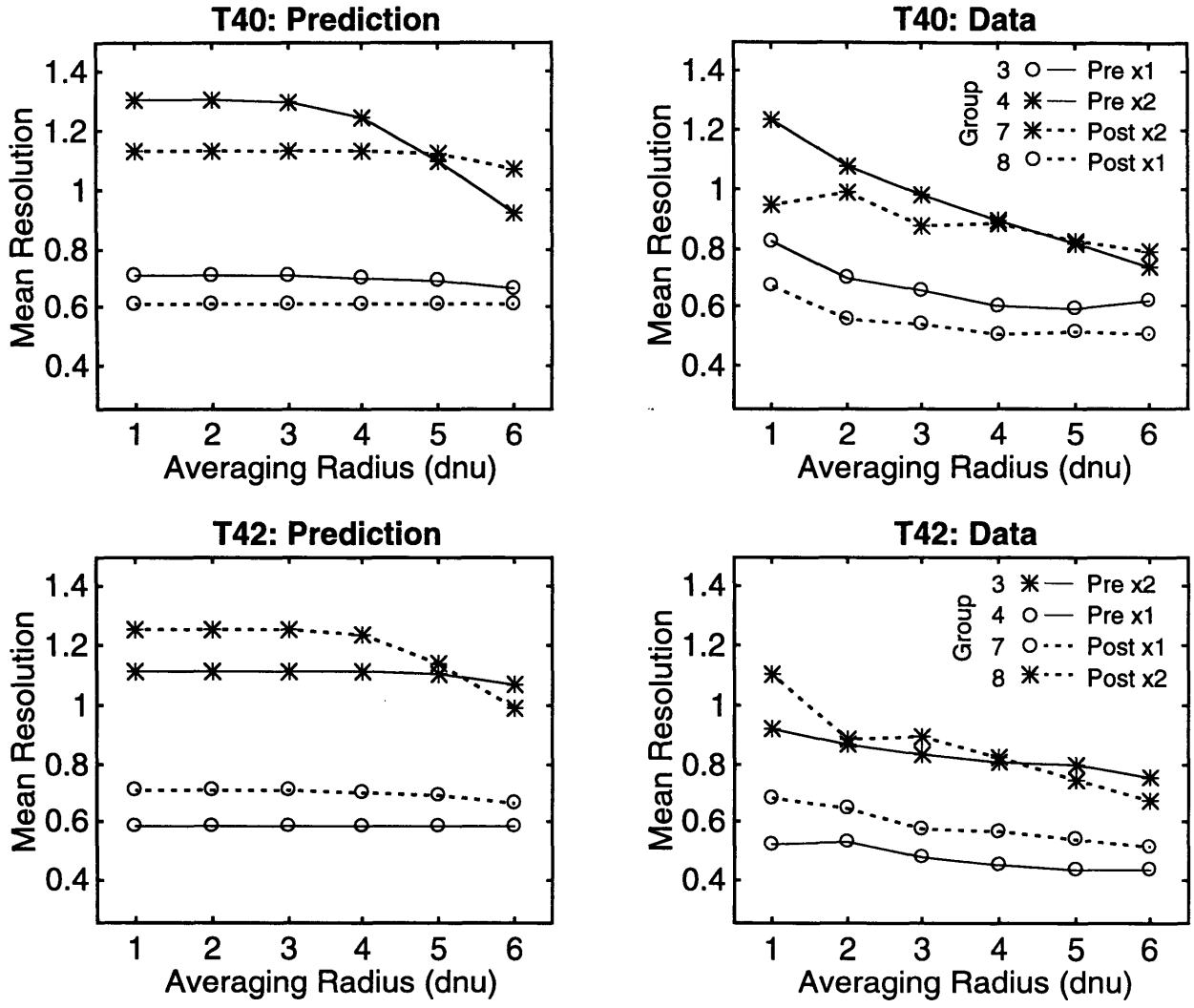


Figure 7.21. Predicted and observed resolution (d*) averaged across depth intervals symmetric about zero for the no-feedback groups of tests T40 and T42 (v2.102).

the observed results, while there is a discrepancy in the shape of the curves as discussed above, the top curves in the prediction plots do drop more with increasing radius than the other predicted curves.

7.6 Limitations of the Model

The preliminary adaptation model does a reasonably good job of predicting the results of the second identification experiment, as discussed in section 7.5. The assumption that G may be modeled as a constant is also verified in section 7.4. Similarly, section 7.3.3 presents evidence that σ is approximately the same for all stimulus depths (another assumption of the model), although the analysis does not include the affect of bias in the estimate of σ. If the expected bias is taken into account, the analysis would suggest that σ actually increases at the ends of the stimulus range. Other limitations of the model are also apparent and provide a guide to future work.

7.6.1 Prescriptive Model

The main limitation of the preliminary adaptation model is that it does not provide any mechanism for describing the process of adaptation. In other words, the model is merely descriptive about how the observer adapts. This means that in order to use the model one must first observe how k changes over time and, then, plug the observed values of k into the model. The latter clearly limits the model's predictive power. For example, adaptation in the absence of direct feedback was an entirely unexpected result in the present experiments. Adaptation mechanisms, that might serve as a guide to developing a more fully prescriptive model, are proposed in section 6.2.3.5 in an attempt to explain the unexpected result. Also see section 7.2.3.5.

7.6.2 Assumptions about k

The analysis in chapter 7 also raises questions about three of the model's assumptions. The possibility that σ may not be constant across the stimulus range was already mentioned in the first paragraph of the section. Two other observations are related to the adaptation coefficient (k).

A fundamental assumption of the preliminary adaptation model is that the observer's state of adaptation is fully described by k. The discussion in section 7.3.2.1 suggests, however, that σ changes independently of k

during the experiment. While this may only be a minor effect (indeed, the present results are not conclusive that it occurs), it must be considered in any attempt to develop a prescriptive model.

Another assumption that is challenged by the present results is that the behavior of k may be described by a simple first-order exponential model. Specifically, section 7.2.3.5 suggests that the subjects may adapt very quickly, by one mechanism, during the first few trials of the experiment and, then, shift to a slower form of adaptation during the subsequent training trials. Abrupt changes in k are also observed in the absence of direct feedback, following a transition from x2 to x1 IPD (see section 7.3.2.2, Indirect Feedback and Adaptation Rate), although this result may be nothing more than noise. While one would hope that an exponential model may still be used to describe individual adaptation mechanisms, the point is that there appear to be multiple mechanisms and that the subject can shift abruptly from one to the other. If this is true then a more complicated model is required.

7.6.3 Discrepancies between Predictions and Results

Three discrepancies between the predictions of the model and the results stand out in section 7.5: 1) the mean response is underestimated and overestimated at the low and high ends of range, respectively, 2) the resolution estimates tend to be too high, and 3) the shapes of the mean resolution curves are different. Each discrepancy is discussed in turn.

1) The simplest explanation for the discrepancies of the mean response predictions at the ends of the range (see section 7.5.1) is that σ is underestimated. Specifically, if σ_M is made larger, then the predicted value of M_i^* will be moved more toward the center of the response range (see equation 7.16). This is particularly true in the case of large values of k where the observed discrepancy is the greatest. Increasing σ would also have little effect in the middle of the range where the present predictions are relatively good.

Note that the fact that the mean response is well predicted in the middle of the range is not surprising given the way k is defined and estimated. Specifically, recall that k is estimated as the slope of the best-fit line to the subjects' responses in the middle of the range (see sections 7.1.3.2 and 7.1.3.4)!

Aside from the fact that the proposed explanation is simple, it is also promising because one already expects that the estimated value of σ is low (see sections 7.1.3.2, 7.3.2.1,1, and 7.3.3). Indeed, the hypothesis is further supported by the fact that the expected bias is largest near the ends of the range (i.e., such that the true value of σ would be larger near the ends of the range than the estimated value), since that is where the discrepancy is greatest. Of course, changing the estimated value of σ would also affect the resolution predictions.

2) Clearly increasing the estimated value of σ in the model will reduce the predicted resolution overall, as needed to reduce the observed discrepancy in resolution magnitude. Simply increasing the uniform value of σ in the model does not appear to be the answer, however. Specifically, when the value of G is artificially increased in order to increase the estimate of σ (a radical change of about three times the estimated value), the discrepancy in the mean response is reduced, but not eliminated. While, at the same time, the discrepancy in the magnitude of the resolution is reversed (i.e., the predicted resolution now tends to be too low). This fact is further evidence in support of the idea that the model may need to be modified so that σ varies along the decision axis, such that it is larger near the ends of the stimulus range.

3) The discrepancy in the shape of the predicted mean resolution curves also suggests that there are aspects of the experiment that are not explained by the model. Changing the model so that σ is larger near the ends of the stimulus range would reduce the observed discrepancy. Unfortunately, this does not appear to be the whole answer. When the estimate of σ is artificially changed in this manner, the shape of the top predicted curves (T40 preexposure x2 and T42 postexposure x2) can be made to coincide with the results. Unfortunately, when this is done, the slopes of the other x2 IPD predicted curves (T40 postexposure x2 and T42 preexposure x2) become much steeper than the slopes of the actual curves (see section 7.5.3). Further research is required to develop a suitable model, possibly along the lines of the perceptual anchor model of context-coding (Braida et al. 1984).

8. CONCLUSION

The thesis makes several contributions. First, the preliminary adaptation model is investigated for a different sense modality than the one for which it was originally developed. In addition to providing evidence both for and against the model, the thesis addresses the relevance of the experiments to traditional adaptation research (see section 6.1). The investigation is also the first to observe changes in resolution associated with adaptation to altered IPD. Further, the discovery of indirect feedback in the present experiments led to the proposal of various possible mechanisms of adaptation that, in turn, suggest directions for future research. The thesis also adds to the general body of knowledge related to computer-graphic stereoscopic displays. In particular, display pixelation is used to explain various experimental results, present and prior.

The remainder of the chapter summarizes the results that are the basis of these contributions. Other contributions are also identified in the course of the discussion. Finally, design rules are proposed, based on the results, and future work is suggested.

8.1 Summary of the Results

The experiments are logically divided into three main groups, presented in separate chapters. Chapter 4 describes the calibration experiments in which a physical probe is adjusted so that it appears to be at the same location as a virtual target. Chapter 5 presents the discrimination experiment that uses the same depth stimulus as the second identification experiment. Chapter 6 describes the two absolute identification experiments in which adaptation is observed. A quantitative analysis of the second identification experiment is presented in chapter 7. The following discussion presents the results of each group of experiments in turn.

8.1.1 Calibration Experiments (Chapter 4)

It was observed that the subjects perceived virtual objects near the predicted locations, based on the algorithm that is presented in section 3.2.4.3. Resolution in the virtual space also improved when the effective IPD was increased. Specifically, the variances of the subjects' responses in positioning the physical probe in the 3D Experiment with x2 IPD were only slightly larger in the physical experimental space compared to the x1

IPD responses. This, in turn, corresponds to a smaller variance in the virtual space and, hence, better resolution (see section 4.1.2).

While there are some discrepancies between the predicted position of the virtual object and the observed mean response, this discrepancy is accounted for by an analysis of display pixelation. Indeed, display pixelation was found to be a significant factor that must be taken into account in the analysis of computer-graphic stereoscopic displays. Although display pixelation has been investigated previously by others (see section 2.2.2.4), the present analysis applies the concept to explain various experimental results.

An understanding of display pixelation was essential for the design and interpretation of the subsequent experiments. The concept is also applied to explain the results of a prior experiment by Rosenberg (see section 4.4.1). Limitations of computer-graphic stereoscopic displays in psychophysical research and other applications are also made apparent. Specifically, displays like the one used in the present experiments can only generate a small (and relatively non uniform) subset of visual stimuli. A possible solution to this problem is discussed in section 8.3.5.

The fact that virtual objects are perceived at the predicted positions suggests that the available depth cues (see section 3.3) are sufficient to locate near-field objects. The results also confirm that accommodation is not a significant distance cue (see section 2.1.1).

On the other hand, while the mean response was as predicted, some of the subjects were less consistent in their responses to stimuli with large (positive or negative) parallax. It is hypothesized that the additional uncertainty is related to the conflict between accommodation and convergence in the display. The result of a follow-up experiment, in which the positions of two virtual objects were compared, is consistent with this hypothesis (see sections 4.2 and 4.3).

Another interesting observation made in connection with these experiments (not previously mentioned) is that when a virtual probe is moved by hand, the human operator does not perceive discrete changes in the visual depth of the probe. This is true even when the operator, knowing that the parallax changes in discrete steps, tries to observe the corresponding discrete changes in depth. In contrast, discrete changes in the size of the probe are plainly visible.

8.1.2 Discrimination Experiment (Chapter 5)

The depth discrimination experiment was performed in conjunction with the second identification experiment, so that the results of the former experiment could be used in the quantitative analysis of the latter. Both experiments involved the same set of subjects as well as nearly identical visual depth stimuli. The stimulus consisted of two virtual targets, separated horizontally on the display so that depth resolution was not limited by display pixelation. The results of the discrimination experiment show that discrimination of the given stimulus improves, as expected, when the effective IPD is increased.

The discrimination results were analyzed using signal detection theory and Thurstone's law of comparative judgment. Then, the calculated sensitivity was used to determine a psychophysical scale that relates the physical stimulus (parallax) to the mean of the corresponding distribution on the decision axis. The scale was found to be linear and is used in the quantitative analysis of chapter 7.

Surprisingly, the subjects demonstrated a bias in the experiment. Specifically, they were more likely to respond "Left" than "Right," when the targets were in front of the plane of the display. This was unexpected and is not understood.

8.1.3 Identification Experiments (Chapters 6 and 7)

Two depth identification experiments were performed because the results of the first experiment were confounded by display pixelation and because it was impossible to assess the preliminary adaptation model using the results of the first experiment. The latter is true because there was no change in resolution when the effective IPD was changed.

Although the exact cause of this result is not fully understood, a partial explanation is provided by the preliminary theory of intensity resolution. Specifically, resolution may not have changed in the first experiment because it was limited by memory noise rather than sensation noise (see section 6.2.3.2). Additional cognitive noise related to pixelation may also have been a factor. Various artifacts in both the resolution and the bias results are also directly attributable to display pixelation in the first experiment (see section 6.2.3.1).

The design of the second experiment eliminates these problems and, apparently as a result, the expected gross shifts in resolution and bias are

observed (see sections 2.3.1 and 2.3.4). In particular, resolution improves when the effective IPD is increased, while the bias returns to the preexposure level as the subject is exposed to the transformation. Overall, the results are consistent with the preliminary adaptation model, although some limitations are identified. Various secondary effects are also identified and (sometimes) explained.

8.1.3.1 Preliminary Adaptation Model: Aside from the general agreement between the predictions of the model and the experimental results, discussed in section 7.5, the results also support the model's assumption that G is a constant (see section 7.4). Questions are raised about other assumptions of the model in section 7.6.2, however, and various discrepancies between the predictions and the results are discussed in section 7.6.3. In particular, the possibility is raised in section 7.3.2.1,2 that a new adaptation parameter (in addition to k) may be required. Also, changes in k are not fully described by a first-order exponential function. And, the difference in the shape of the mean resolution curves (see section 7.5.3) suggests that there are aspects of the experiment that are not explained by the model.

Most of these limitations are evident in the secondary effects that are discussed in the following section. The main limitation, however, is that the model does not provide any mechanism for describing the process of adaptation (see section 7.6.1). This is discussed further in section 8.3.3.

8.1.3.2 Secondary Effects: Several resolution effects are identified in connection with the second identification experiment. In particular, the magnitude and the shape of the resolution curves vary from group to group, depending on the value of X_{ipd} and the group order. Hypotheses are proposed to explain the effects in section 6.3.3 based on the preliminary adaptation model. One hypothesis suggests that the subjects do not adapt as fully during the exposure as they do during the initial training (see section 6.3.3.3). Evidence for this is also presented in the quantitative analysis (see section 7.3.2), although this analysis also suggests that the variance of the subject's responses may change independently of the adaptation coefficient, k (in contradiction to a key assumption of the preliminary adaptation model, as noted in section 8.1.3.1). It is proposed that the latter effect might be explained by an independent parameter associated with learning noise (see section 7.3.2.1,2). Unfortunately, there is insufficient evidence to confirm this hypothesis.

Secondary bias effects are also observed in both the first and second identifications experiments. Specifically, the initial effect and the

aftereffect in the subjects responses are asymmetrical and smaller than expected. Also, the subjects appear to underestimate the larger depths in the nominal preexposure measurements. These effects appear to be related to the finite size of the response set, indirect feedback, and the central tendency discussed in sections 6.2.3.4 to 6.2.3.6. Further evidence of the effects, and discussion of the related hypotheses, are presented in connection with the quantitative analysis in chapter 7, where the behavior of k is observed.

Indirect Feedback: An important, unexpected result of the experiments is that the subjects adapt in the absence of direct (i.e., explicit correct-answer) feedback. This is not to say that the subjects adapt without *any* feedback, but rather that other *indirect* feedback mechanisms exist in the experiment. A number of adaptation mechanisms are proposed in an attempt to explain the results. The primary importance of the work is that it may lead to a more prescriptive adaptation model.

Multiple Adaptation Mechanisms: The detailed analysis of the behavior of k also suggests the existence of multiple adaptation mechanisms in the experiments. Specifically, section 7.2.3.5 suggests that the subjects may adapt very quickly, by one mechanism, during the first few trials of the experiment and, then, shift to a slower form of adaptation during the subsequent training trials. Abrupt changes in k may also occur in the absence of direct feedback, following a transition from x2 to x1 IPD (see section 7.3.2.2, Indirect Feedback and Adaptation Rate). No explanation is proposed for the latter effect.

8.2 Design Rules

A few simple rules are derived from the experimental results to aid engineers in selecting the best effective IPD for a given application, and to guide the design of systems that use computer-graphic stereoscopic displays in general. See section 2.2 for references to additional information in this area. The rules are as follows:

1) Changing the effective IPD may be acceptable in many cases. There is always some performance cost, however, so one should minimize the number of changes in a given application and be sure that there is some benefit to outweigh the cost. To find out how depth resolution will be affected in a given task, calculate the ratio:

$$R_d = \frac{\text{total depth range of the task}}{\text{binocular disparity threshold}}.$$

In other words, R_d is the number of binocular disparity JNDs in the range of the task. R_d equals one in a discrimination task and it equals the magnitude of the difference between the maximum and minimum stimuli in an absolute identification task.

If R_d is small (less than about 20)[1], then resolution should change proportionally with the effective IPD. Hence, one might improve resolution in a task by increasing the effective IPD. On the other hand, if R_d is large (more than about 80), then resolution does not change. Hence, one might reduce conflicts between accommodation and convergence by reducing the effective IPD without affecting resolution. Some change in resolution may be observed for R_d in the intermediate range. Of course, these predictions assume that performance in not limited by display pixelation. This problem leads to the second rule.

2) Design virtual objects with visually meaningful dimensions when spatial judgments are required. In other words, the objects should be designed so that their intended projections on the screen correspond to the integral number of pixels with which they are actually drawn. Similarly, depth intervals presented on the display should have an integral parallax.

The reason for this rule is obvious: the visual stimulus corresponds to the image on the display, not the model in the computer. The rule must be emphasized, however, because binocular depth discrimination is one or two orders of magnitude better than the resolution of existing displays. As a result, small shifts related to display pixelation can significantly alter the information presented by the display. Further, the image of a given virtual object changes depending on the effective IPD and where the object is located on the display.

3) A corollary to the second rule is that one should design systems with a single value of effective IPD whenever possible. This is simpler, of course, because one does not need then to calibrate the system for each operator. The point here, however, is that changing the effective IPD also changes the visual stimulus such that different operators are presented with different information.

One exception to this rule is the case where virtual images are superimposed on a direct (untransformed) view of the real world. The

---

[1]The numerical estimates for small and large R_d are based on hearing data (Braida and Durlach 1972). See figure 6.5 in section 6.2.3.2.

reason for this exception is that the effective IPD must be the same in both cases for proper registration. The next design rule applies to this type of application.

4) Keep virtual objects near the plane of the screen (z = 0) for accurate depth judgments relative to real objects in augmented reality applications. This final rule is based on the fact that some operators become less consistent in their responses when large parallax values are used.

Note that this problem does not exist if the real world view is presented on a stereoscopic television display, because then accommodation is the same for both the "real" and the virtual images. Also, the exception to the third rule need not apply with a stereoscopic television system because both the virtual images and the real world view may be transformed in the same way (i.e., they can have the same effective IPD). Hence, the effective IPD can be different from the operator's actual IPD without affecting the registration of the real and virtual images.

8.3 Future Work

There are, of course, many unanswered questions. Indeed, the thesis is but one step in a long journey. Following are some suggested directions for future research.

8.3.1 Roving Depth Experiment

A direct extension of the identification experiments in chapter 6 would be one in which the reference target is randomly presented at different distances on successive trials. This should be done in order to distinguish relative depth judgments from absolute, or egocentric distance judgments in the experiment (see section 2.1). A roving reference target was not used in the thesis experiments because of the limited depth range of the display.

8.3.2 True Adaptation in Absolute Identification

Section 6.1 makes the case that the present experiments involve "true" adaptation. Perhaps the main argument against that position is that the subjects are consciously relearning the depth stimuli in an absolute identification experiment. As a result, *the argument* goes, the subjects do not perceive the stimuli differently when their vision is transformed.

Instead, they simply learn to respond differently, consciously correcting for the transformation. For example, the subjects might respond "2" when stimulus depth number 2 is presented, even though the effective IPD is double and it looks like depth 4.[2] The subjects do this consciously (according to *the argument* that this is not true adaptation) because they have learned that when it looks like depth 4, "2" is the correct response.

One could put *the argument* to the test by telling the subjects about the transformation and then challenging them to give the correct response. In other words, one would perform an experiment like the identification experiments in chapter 6, except that the subjects would be explicitly told when the effective IPD is changed and asked to compensate.

Further work is needed to determine the proper group order and the exact nature of the instructions to the subjects in the new experiment, but the basic idea is to see if the subjects can consciously correct for the transformation. More specifically, one asks the question: can they maintain two (or more) sets of criteria corresponding to different values of effective IPD? The extent to which the subjects cannot do this when explicitly challenged to do so, would seem to provide evidence as to the extent to which *the argument* is false.

For example, it seems likely that some effect of the subject's attempts to make conscious corrections would be apparent in the results, but would some remnant of the usual shifts in bias (initial effect, reduction of effect, and negative aftereffect) also be present? If so, then these remaining shifts would seem to contradict the idea that adaptation in the experiment is purely conscious.

Indeed, the resolution shifts that were observed in the second identification experiment (section 6.3) and Shinn-Cunningham's resolution results (1994) already provide such evidence, since the observed pre- to postexposure changes in resolution can hardly be the result of conscious correction.

The proposed experiment should provide additional evidence, either for or against *the argument*. It would also be interesting to see if the degree to which bias may be controlled by conscious correction is balanced by a

---

[2]One may wonder whether statements like, "it looks like depth 4," are meaningful, given that the stimulus designations in an absolute identification experiment are arbitrary. A simple response is that they are meaningful to the extent that *the argument* is correct. In other words, if the subjects can identify the stimulus as "depth 4" (even if only to themselves) then it is meaningful.

loss in resolution. The proposed experiment is also directly related to the issue of dual adaptation.

8.3.2.1 Definitions, Mechanisms, and Generalization: If adaptation in an absolute identification experiment is not purely conscious (i.e., if *the argument* against true adaptation discussed above is at least partly false), then the implication is that true adaptation is observed in the experiment. But what does this really mean? Further work is required to refine the definition of adaptation. More work is also needed, generally, to identify the underlying mechanisms and the extent to which the present results may be generalized.

8.3.3 Development of the Model

As stated in section 7.6.1, the main limitation of the preliminary adaptation model is that it does not provide any mechanism for describing the process of adaptation. This limits the model's predictive power because, in order to use the model, one must first observe how people adapt. A model that describes the detailed process by which observers change their criteria is required. Although this may prove to be a difficult task, it would seem to be a useful and logical extension of the theory.

Some adaptation mechanisms are proposed in connection with the analysis of indirect feedback in the present experiments (see section 6.2.3.5). Related issues are also discussed in section 7.2.3.5. Such mechanisms could be developed into a model of adaptation by 1) establishing a set of rules that would predict how observers change their criteria given a particular sensation (Q) on the decision axis (under a given experimental condition), and 2) estimating the probability that the sensation will occur given that a particular stimulus is presented. One could then estimate the mean change in the criteria over a series of trials in an experiment like those of the thesis.

A number of issues might be investigated in the process of developing such a model, for example: display pixelation noise (re section 6.2.3.2), learning noise (re section 7.3.2.1,2), abrupt changes in k (re section 7.3.2.2), how σ varies with N (re section 7.3.3), and why adaptation might be slower after initial training (re section 6.3.3.3). One might also perform new versions of the thesis experiments with a non-linear set of depths to test the idea that people rescale their criteria in a linear fashion (re section 6.2.3.5). The reason that the predicted mean resolution curves are a different shape than the actual curves should also be investigated (see section 7.6.3,3).

8.3.4 Beyond Simple Identification

Although, the simple case of unidimensional absolute identification is still not fully understood, ultimately we must move beyond it. One extension of the present work might be to consider a multidimensional identification task. One could then observe the affect on one dimension when a transformation is imposed on another dimension. Training might be interrelated among the dimensions in some way in order to induce (or enhance) the effect. Such an experiment might more closely correspond to traditional adaptation studies where a discrepancy between sense modalities is involved (see section 6.1).

The present adaptation model might also be extended to cross-modal estimates of spatial quantities, since the latter appear to be nothing more than absolute identification.[3] Of course, the extent to which this is *not* true could teach us something about how we perceive space. Another direction of research might be to study adaptation in an absolute identification task using exposures other than direct feedback about the correct response. For example, the exposure might involve wearing a telestereoscope while walking around the lab. Hand-eye tasks should also be investigated because of their practical importance.

8.3.5 Antialiasing

Display pixelation was a significant problem in the experiments. These effects, in turn, are related to the well-known problem of aliasing in computer graphics, where pixelation causes images to appear jagged. Antialiasing techniques exist for smoothing images by gradually varying the intensities of adjacent pixels (Foley et al. 1990). While these techniques were not used in the thesis experiments because of time constraints, they would appear to be a straightforward (at least partial) solution to the problem of display pixelation.

Indeed, Rolland et al. (1995, p. 45) report that the depth threshold observed in their experiment for discriminating two virtual objects was less than one would predict because antialiasing was used. In spite of antialiasing, however, the depth threshold observed for two virtual objects was still roughly seven times larger than the threshold they observed for two real objects. Further experiments should be performed to investigate the affect of antialiasing on depth perception in computer-graphic stereoscopic displays.

---

[3]Some experiments were performed in this area by the author that are not reported here because of time constraints.

APPENDIX A. VIRTUAL WORKBENCH

The following appendix presents the contents of the World Wide Web page devoted to the Virtual Workbench and selected links as of March 29, 1997. This computer site was maintained when this thesis was written by the Massachusetts Institute of Technology, Research Laboratory of Electronics, Sensory Communications Laboratory. The internet addresses of the Virtual Workbench page and the selected links were as follows:

VWB Page

http://mimsy.mit.edu/people/tew/VWBpage.html

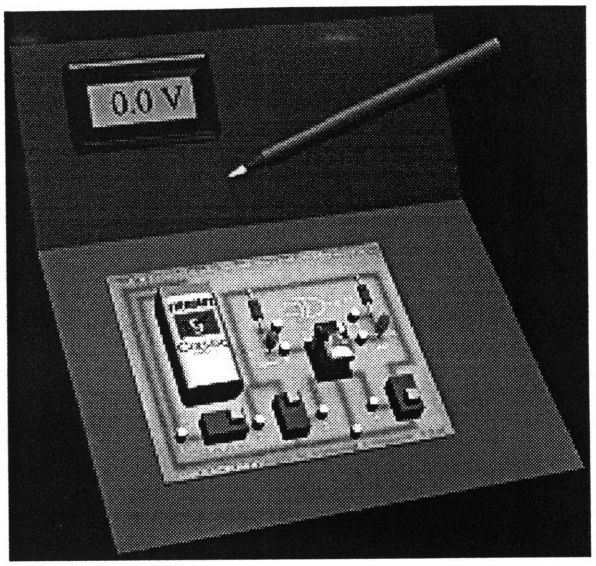
Links

1. http://mimsy.mit.edu/people/tew/VWBschematic.gif
2. http://mimsy.mit.edu/people/tew/VWBtomevanjim.gif
3. http://mimsy.mit.edu/people/tew/VWBtomben.gif

4. http://mimsy.mit.edu/people/tew/VWBwSubj.gif

5. http://mimsy.mit.edu/people/tew/VWBphantom1.gif
6. http://www.ai.mit.edu/projects/handarm-haptics /haptics.html

7. http://mimsy.mit.edu/people/tew/VWBcktBdReal.gif
8. http://mimsy.mit.edu/people/tew/VWBcktNonAlign.gif
9. http://mimsy.mit.edu/people/tew/VWBcktAlign.gif



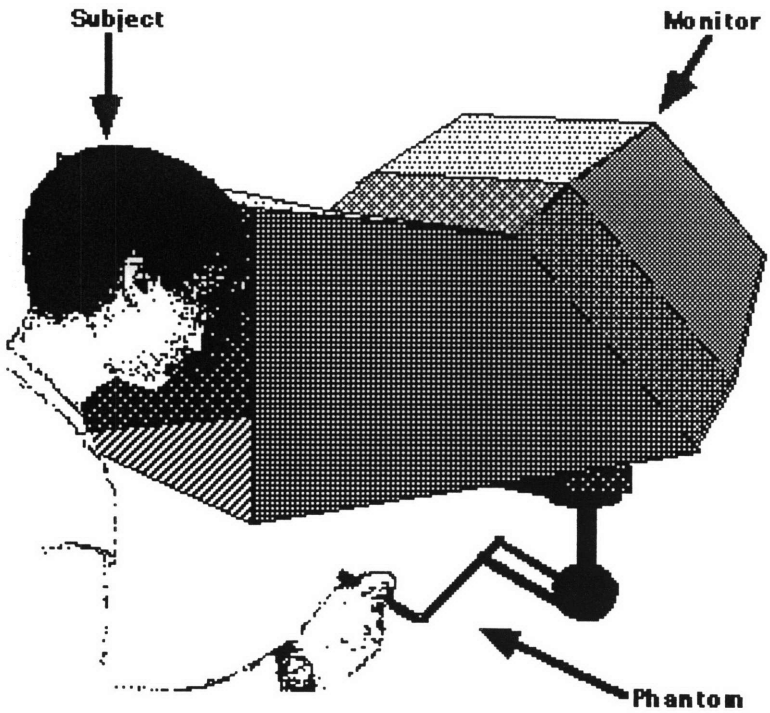
The MIT/RLE/ Virtual Environment Technology *for* Training Virtual Workbench

Project Staff:
Nathaniel I. Durlach (P.I.)
Dr. Thomas E. von Wiegand , Dr. Ben Sachtler, David Schloerb
Evan F. Wies, David C. Lossos, Jim Bandy , Lukasz Weber, Robert Powers, Erika Carmel, Stephen Baird
Nora Luongo , Secretary.

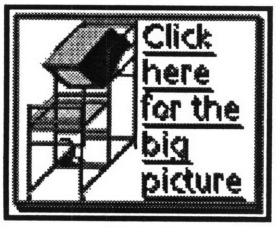
Introduction

The multimodal Virtual Workbench (VWB) was developed to both study human sensorimotor capabilities in near-field space and to explore cognitive performance and training in complex tasks involving spatialization of information in Virtual Environments (VEs). The system combines state-of-the-art haptic feedback and visual display technologies, with audio capabilities, to create an alternative «front-end» interface for the core Testbed.

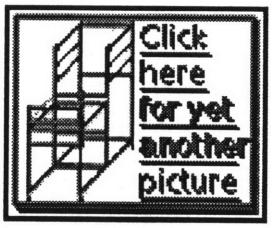
Currently available head-mounted displays (HMDs), and their associated tracking and graphics-rendering systems, operate at the limits of current technology. Despite their great cost, these systems are still insufficient to present a virtual world free of performance-reducing and nauseogenic effects.



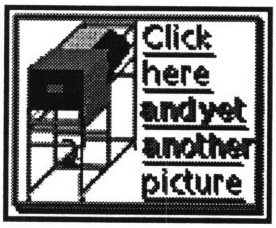
In the Virtual Workbench, we combine visual display and haptic feedback technologies into a form in which the user looks and reaches into a virtual space (as opposed to trying to immerse the entire body). In this way we are able to sidestep many of the technical shortcomings of HMDs, enabling the Virtual Workbench to serve as a precision instrument for measuring human sensory and cognitive performance.

¤ 

Link 1

¤¤ 

Link 2

¤¤¤ 

Link 3

The Virtual Workbench was specifically designed for precise overlay of visual and haptic spaces, providing a cohesive world percept. The effects (on human performance) of deviations from precise registration of the spaces, as well as distortions, delays, and other manipulations (e.g. modification of forces in haptic space), will be studied using this system. In addition to these sensorimotor effects, we are preparing experiments designed to examine the influences of the above manipulations on cognitive performance (using memory and reasoning tasks).

Electronic Troubleshooting Task

As an initial demonstration, part of our development efforts are being directed toward producing the imagery for a highly veridical training task involving manual interaction with virtual electronic circuit boards and components. This «Electronic Troubleshooting Task» will be a synthesis of the efforts of our sponsors at the Naval Air Warfare Center / Training Systems Division, with those of the haptics and core Testbed teams at MIT, and the training team at Bolt Beranek and Newman.

More visuals:

¤ **View of subject seated at the VWB** (Link 4)

The VWB is constructed about a framework of 1" aluminum pipe. The piping is assembled using adjustable fittings to accommodate both changes in equipment configuration and variations in user anatomy. This mounting technique also provides a solid mechanical ground for haptic «force-feedback» devices, and includes a self-contained rack mount area for control equipment. Lockable wheels allow the VWB to be easily moved to the various laboratory areas as needed.

¤ **Close up of the PHANToM®** (Links 5 & 6)

The PHANToM® is capable of imparting forces along three translational degrees of freedom. These translational forces are decoupled from the rotational degrees of freedom through the use of a gimbal. A scaled-up version of the PHANToM® is installed in the VWB. We are planning to add a second such device in order to enable two-finger grip, or alternatively, to impart rotational forces as defined by two points on a line. (The PHANToM® is a computer-controlled force-reflecting device that is vastly improved over earlier teleoperator force-reflecting linkages.)

¤ **Circuit board for demo of training task** (Link 7)

The simple test case circuit board includes a relay operated by a «two-way switch» (XOR). The position of the relay's SPDT switch is indicated by two LEDs. The circuit allows both «proper» manual interaction (by actuating the slide switches) as well as «atypical» interaction (by forcibly moving the actuator arm of the relay). The *virtual* relay responds to the atypical interaction just as the *real-world* component does: as the actuator arm moves through the position between N.C. and N.O. both LEDs will momentarily extinguish. Incorporating the non-ideal behavior of real-world devices will be an important aspect of training in VETT projects.

¤ **View of PHANToM® and virtual circuit board** (Link 8)

The PHANToM® work-volume is located coincidently with the visual space below the mirror. In this way, the user can manipulate three-dimensional virtual objects presented in the workspace through natural bidirectional kinesthetic interaction, i.e. the user feels (and sees) the virtual objects that he is manipulating in the virtual space.

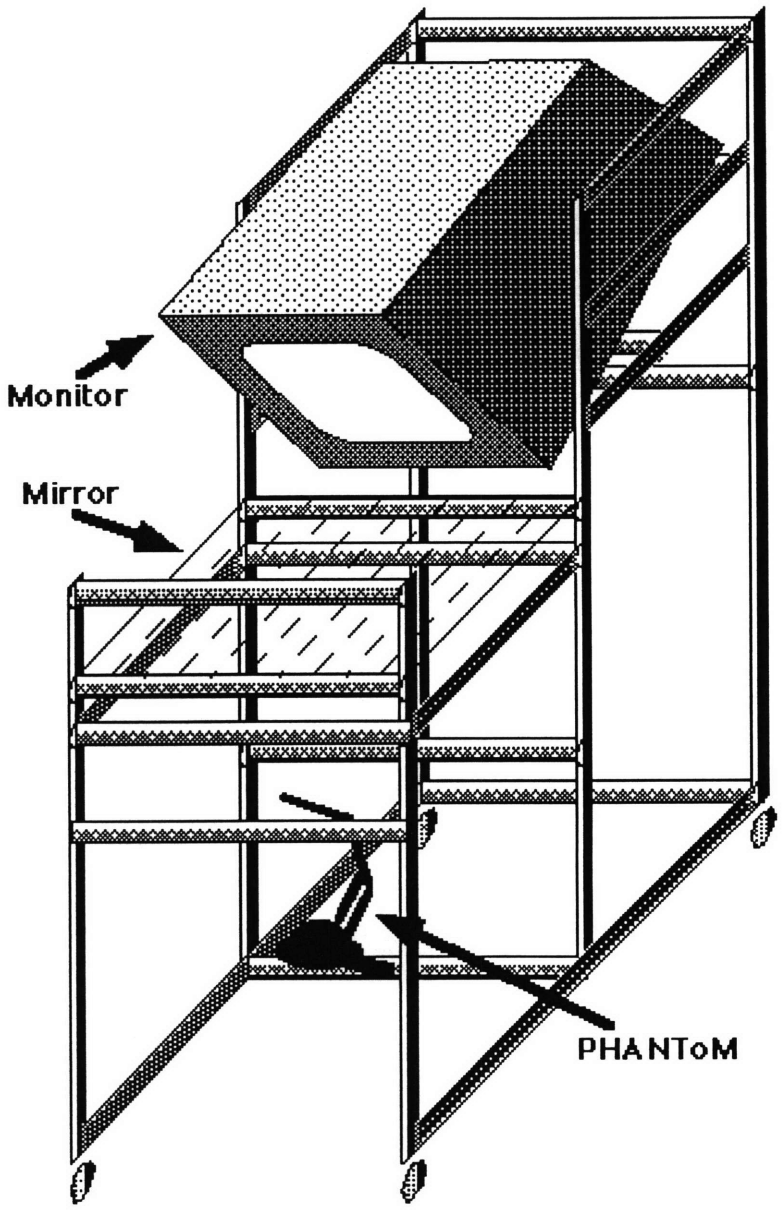
¤ **Subject's overlaid view of virtual circuit board** (Link 9)

The photograph shows the registration of the three-dimensional graphic representation of the circuit board with the stylus attached to the PHANToM®. In actual experiments the image below the semi-silvered mirror is blocked: the subject sees only the virtual objects, and the virtual probe that moves in synchrony with the stylus. The subject feels the virtual objects in their proper places, and can interact with them (i.e. measure voltages, actuate switches, etc.).

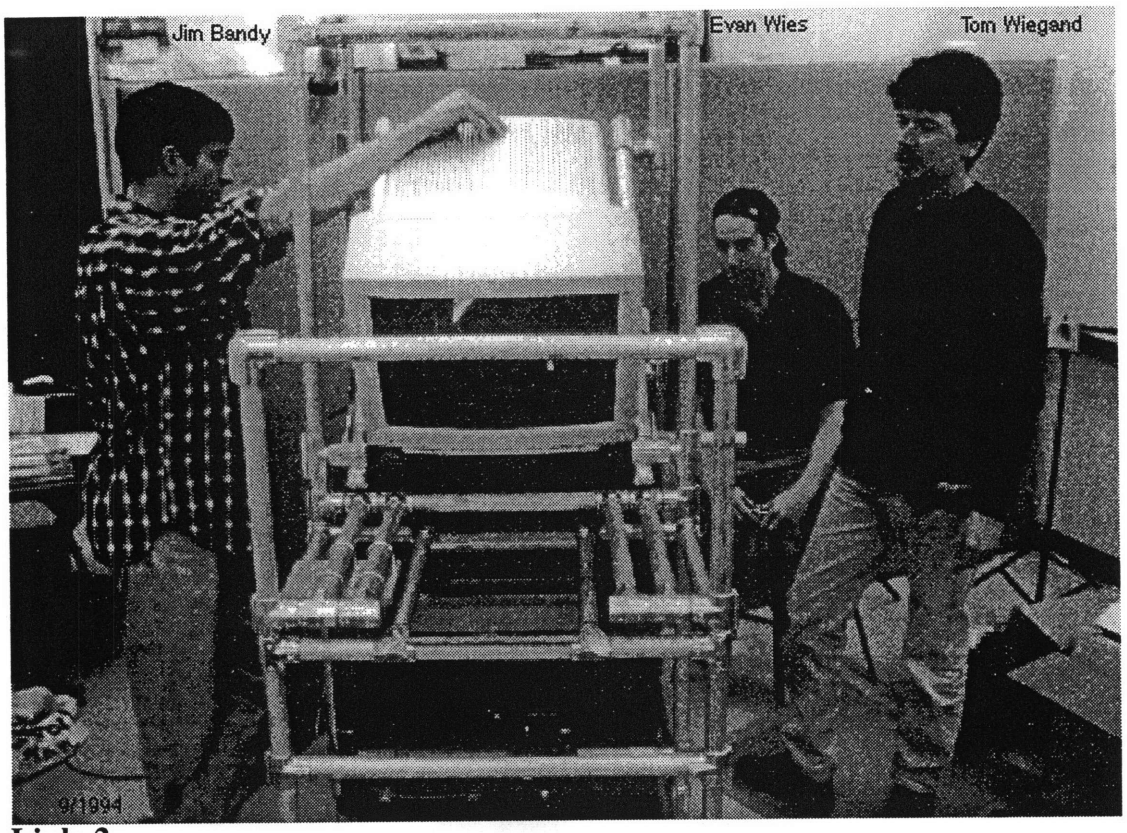
E-mail comments to:

tew@cbgrle.mit.edu
Dr. Thomas E. von Wiegand

Massachusetts Institute of Technology
/ Research Laboratory of Electronics
77 Massachusetts Ave., Room 36-755
Cambridge MA 02139

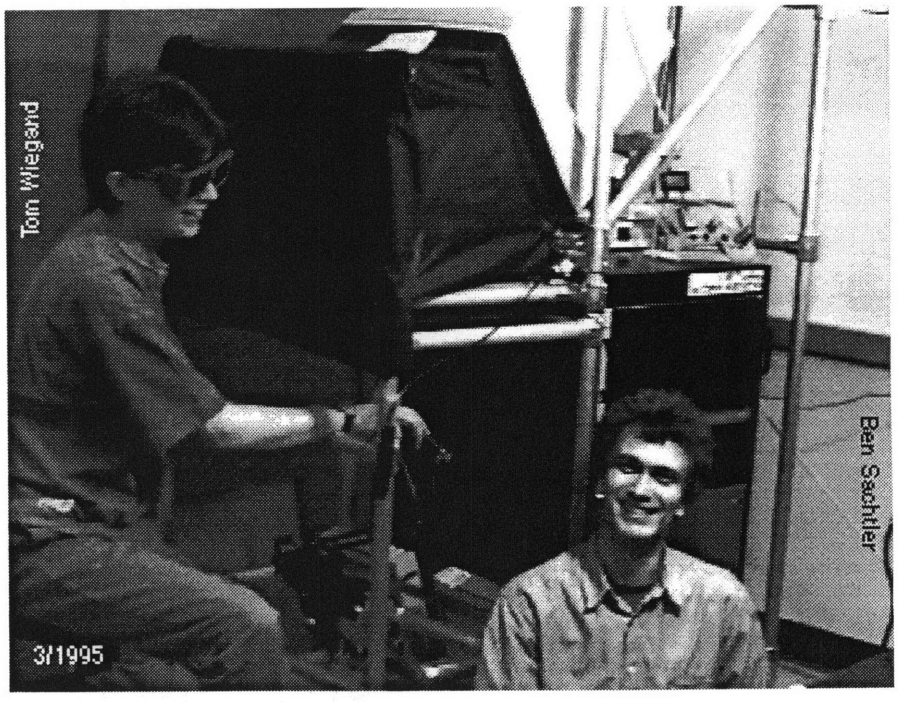


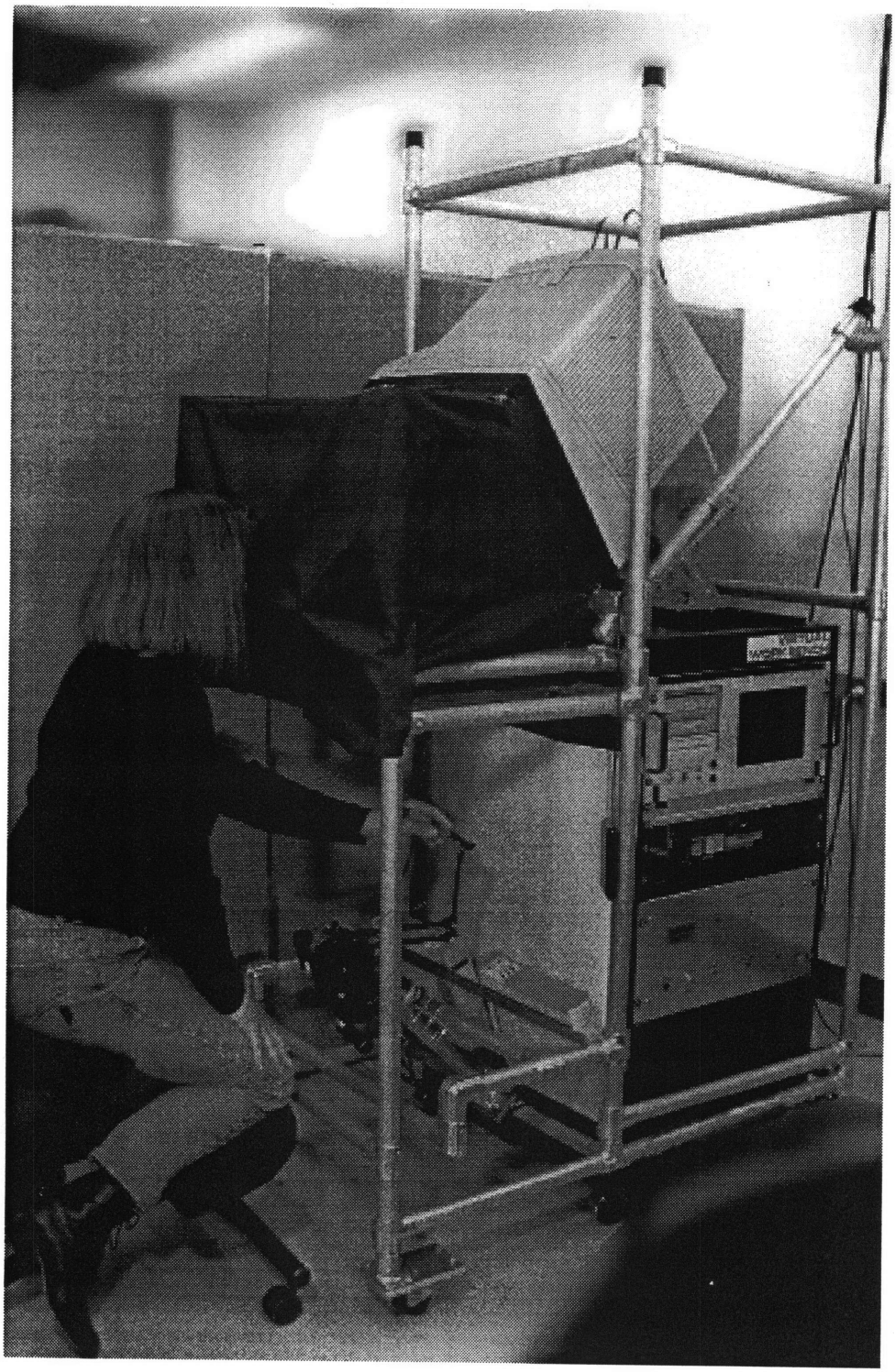
Link 1


Jim Bandy
Evan Wies
Tom Wiegand
9/1994

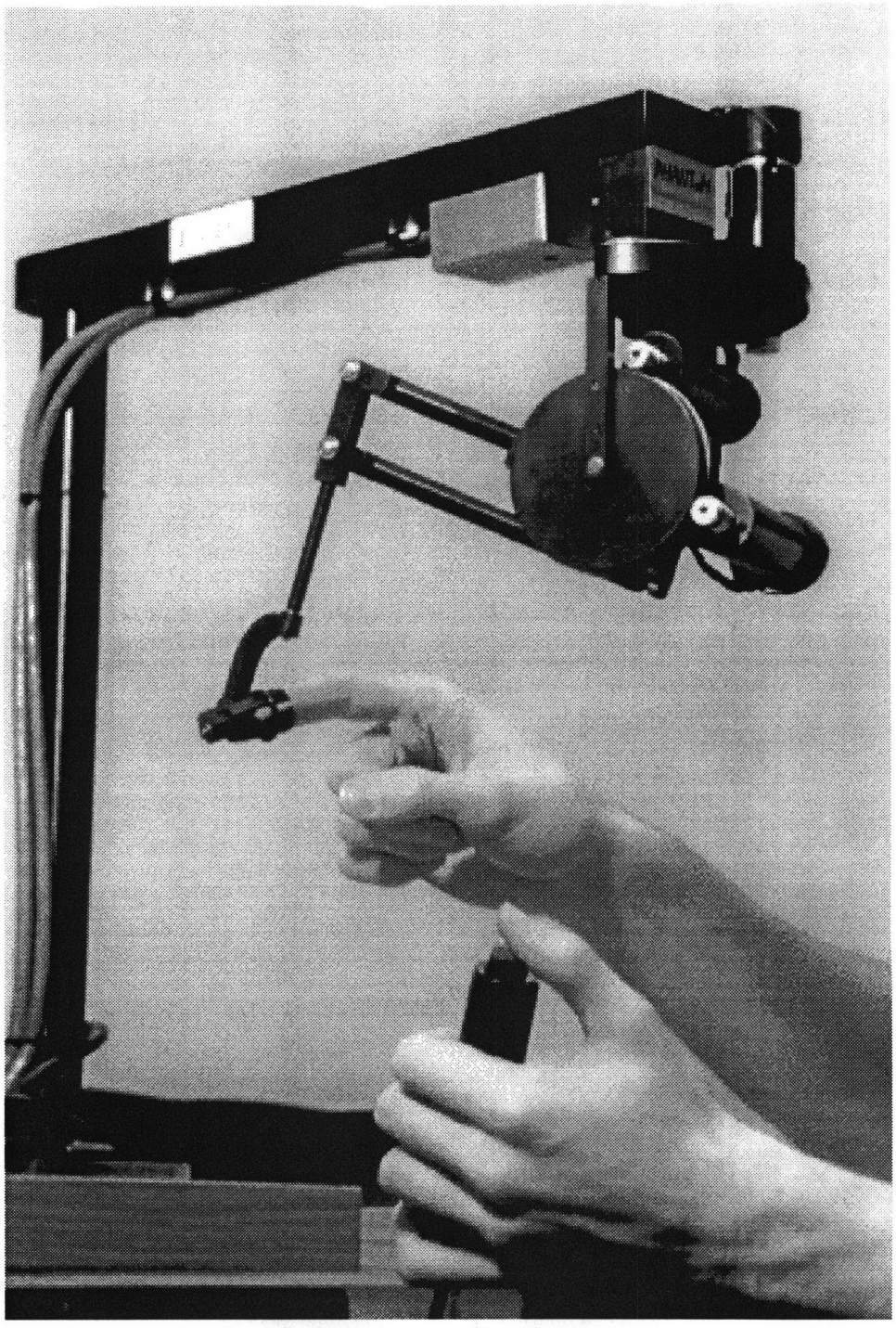
Link 2

Link 3


Tom Wiegand
Ben Sachtler
3/1995



Link 4



Link 5

Haptics

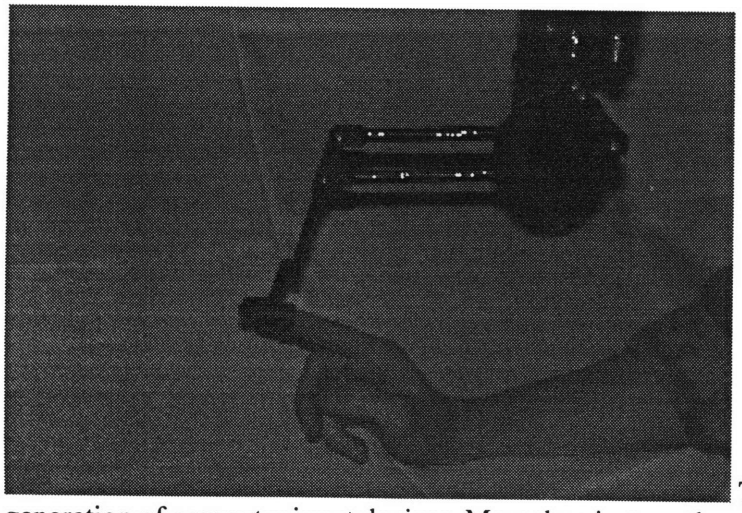
The haptics group at the MIT Artificial Intelligence Lab conducts research in a wide variety of topics related to touch perception and feedback including the development of devices to effect haptic stimulation and software to render the sensations of touch. The group is headed by Dr. Kenneth Salisbury. Another group headed up by Ken is the Vision and Touch Guided Manipulation group.

The people in the Haptics group are:

- Brian Anthony
- Donald Green
- Craig Latimer
- Akhil Madhani
- Thomas Massie
- Derek Schulte
- Nick Swarup (graduated)
- Chris Tarr
- Criag Zilles (graduated)

One of our first developments was the Phantom Haptic Interface (distributed by SensAble Devices, Inc.):

Touching is Believing



The PHANToM represents a new generation of computer input devices. More than just another passive, 3D mouse, the PHANTom

allows users to actually feel virtual objects. Unlike buzzing tactile stimulators, the PHANToM actively exerts an external force on the user's finger tip - creating the illusion of interactions with solid virtual objects. Smoothe spheres, flat walls, sharp corners, and even texture can be convincingly conveyed to the human haptic sytem using the device. This is not a bulky exoskelatal device - one simply inserts his or her finger tip into the PHANToM and interacts with the virtual environment.

High Fidelity

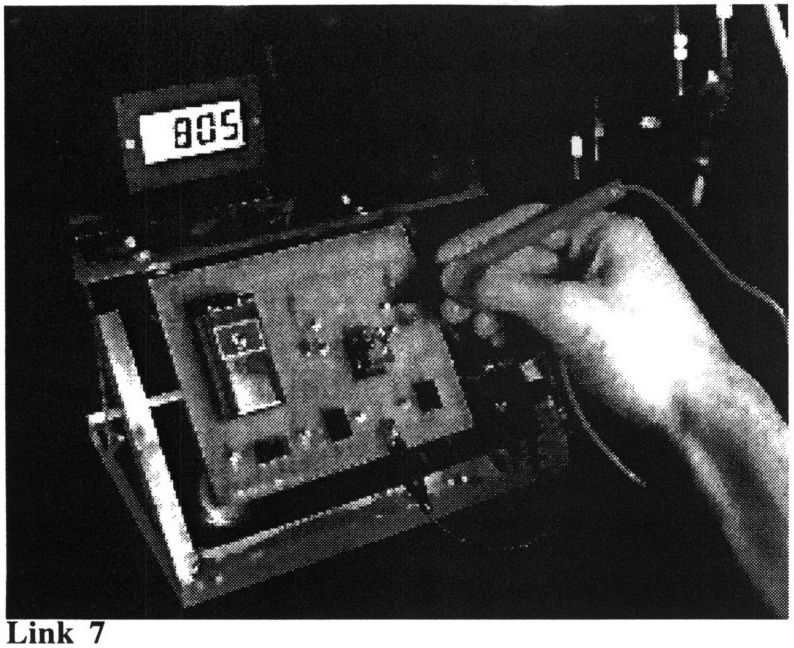
The PHANToM contains 3 motors which control the x, y, and z forces exerted on the user's finger tip. Mounted on each motor is an optical encoder to determine the x, y, and z position of the user's finger tip. The torque from the motors is transmitted through a proprietary cable transmission to a stiff, light-weight linkage. At the end of this linkage is a passive, 3 degree of freedom gimbal attached to a thimble. The passive gimbal allows the thimble to rotate so that a user's finger tip may assume any comfortable orientation. The user's finger tip can then be modelled as a point or frictionless sphere within the virtual world. The device has low friction, low inertia, and no unbalanced weight so movements through free virtual space are unimpeded.

Versatility

The PHANToM is best described as a universal force-reflecting interface. In fact, a stylus can be substituted for the thimble. A user can then manipulate the stylus to control the tip of a virtual pencil or paint brush. Users can actually feel the tip of the stylus touch virtual surfaces. The PHANToM can be used as a high precision force-reflecting master for teleoperation or as a 3 DOF input device for CAD. Artists can mold clay within the computer and surgeons can practice procedures on virtual patients.

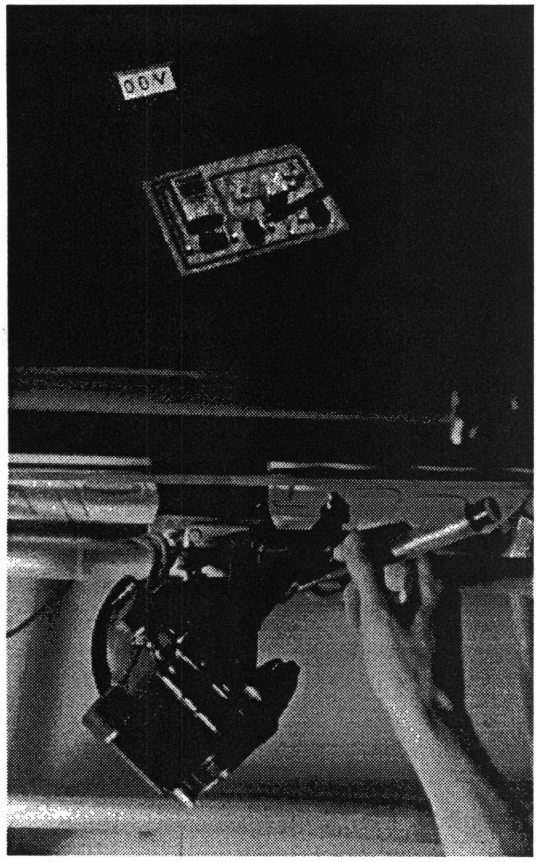
---

special-effects@ai.mit.edu

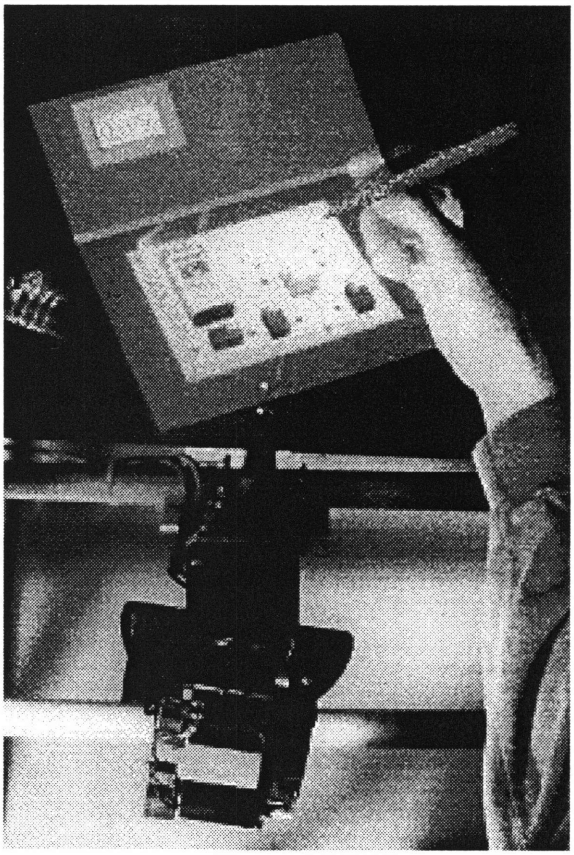


Link 7

Link 8



Link 9



APPENDIX B. SUBJECT SCREENING: PROCEDURES AND DATA

The following appendix describes the procedures that were used in the three preliminary screening tests. Then, results of the tests are presented in a table along with other information about the subjects. A blank copy of the consent form that was signed by each subject is also presented at the end. See sections 3.1 and 3.4 for related discussion.

IPD Measurement Procedure: The subject's IPD is measured as follows. The subject and the experimenter sit facing each other at somewhat less than arms length (about 30 cm). A 15 cm ruler is placed as close as possible to the subject's eyes (about 1 cm), horizontally, directly under the pupils. The ruler is held by the subject on a horizontal support to help steady the ruler and the subject. The experimenter closes one eye and instructs the subject to look at his open eye. Then the experimenter reads the scale marking under the corresponding eye of the subject. Specifically, if the experimenter's right eye is open then he reads the scale marking under the center of the subject's left pupil. Next the procedure is repeated for the other eye. Specifically, if the experimenter's right eye was open, he now opens the left eye and reads the scale marking under the subject's right pupil. The experimenter records the two readings and then the entire procedure is repeated at least three times. Next, the IPD is calculated for each pair of readings and the mean of all the readings is used as the estimate of the IPD.

Visual Acuity Procedure: Visual acuity is measured by having the subject read a Snellen eye chart, with both eyes, at a distance of 20 $\pm$ 0.5 ft (6.1 m). The smallest line that can be read without mistakes is taken to represent the acuity. The results are reported in the standard way as the ratio of the test distance (20 ft) to the distance (ft) where a person with a visual acuity of 1 minute of arc would be able to read the same line on the chart.

Stereo Acuity Procedure: Stereoacuity is measured using standard *Randot Stereotests*, SO-002, 071294 (Stereo Optical Co., Inc. 3539 N. Kenton Ave., Chicago, IL 60641, 1-800-344-9500) where the subject views random dot stereograms. Both the "Forms" and "Circles" tests were used. The subjects all had perfect scores on the "Forms" test (where the minimum acuity is 250 seconds of arc), however, so only the results of the "Circles" test are reported in the following table.

Preliminary Test Results and Subject Data

No.	Sex	Age	Height	Hand	IPD (cm)	Vis.Cor.	Snellen	Randot (sec)
1	M	40	5'9"	R	6.77	none		
		41					20/13	20
2	M				6.50			
3	M	≅35			6.53			20
4	M	<30			6.27		20/13	40
5	F	43	5'2"	R	5.93	none	20/25	40
6	M	≅32			6.73	none	20/25	20
7	M	<30			6.47	glasses	20/20	20
8	F	20	5'2"	(n1)	6.37	glasses	20/20	70
		20			6.43 (n2)	contacts	20/20	30
9	M	18	5'11"	R	6.17	none	20/20	25
10	M	20	5'11"	R	6.40	none	20/40	30
11	M	21	6'2"	R	6.40	glasses	20/20	50
12	M	29	5'11"	(n1)	6.17	none	20/25	30
					6.20 (n2)	none	20/25	20
13	F	19	5'6"	R	5.50	contacts	20/20	30
14	M	18	5'11"	R	6.20	glasses	20/25	25
15	M	20	5'9"	R	6.40	glasses	20/25	20
16	F	20	5'4"	R	6.70	none	20/15	20
17	F	18	5'7"	R	6.43	glasses	20/40	30
18	M	19	5'6"	L	6.00	contacts	20/15	70
19	M				6.77	contacts	20/30	25 (n3)
20	M	22	6'3"	R	6.93	glasses	20/20	20
21	F	20	5'7"	R	6.73	contacts	20/30	20
22	F	19	5'4"	R	6.20	none	20/40	40
23	F	21	5'5"	L	6.10	none	20/13	40
24	F	21	5'6"	R	5.90	contacts	20/25	70 (n4)
25	M	19	6'2"	R	6.90	glasses	20/25	20
26	M	19	6'1"	R	6.67	none	20/13	25
27	M	23	5'8"	R	5.93	none	20/20	20

Notes:

n1) Subject 8 writes, eats, brushes teeth with left; plays sports, open doors with right. Subject 12 writes with Left (considers himself left-handed), but throws with Right. Both did tests with Right.

n2) The preliminary tests were repeated prior to the v2.102 tests and the new IPD's were used.

n3) Subject 19 couldn't tell it was a plus (at 250 sec of arc), just that something was there (scored as correct). In any event, he did not do the experiment because of a problem with his vision.

n4) Subject 24 couldn't judge 50 arcsec (so score is 70), but got 40 and 30 arcsec correct.

IPD Experiment - Subject Consent Form

In this experiment the subject is asked to make repeated judgments about the depth of visual stimuli that are presented on a stereographic computer display. Two general types of tests are performed. In one test the subject is asked to adjust a visual scale on the display using a trackball so that the indicated scale distance corresponds to the stimulus depth. In the other test the subject adjusts a physical pointer with his or her hand until it appears to be at the same point in space as a given part of the stimulus.

The purpose of the experiment is to study how the subject's perception of depth varies, or adapts, over time when his or her view of the world is distorted. Specifically in some tests the apparatus simulates the way objects would look if the distance between the subject's eyes was changed. People quickly re-adapt to normal when visual distortions of this type are removed and each test session (where there is a distortion) includes a period of re-adaptation at the end that is intended to insure that the subject leaves the session pretty much the way he or she came in. Indeed, because the effect is quite subtle, it is sometimes important for the subject to complete the entire test without being exposed to normal viewing conditions, for example, by stopping and looking around the room. When this is necessary brief rest periods will be allowed during the test with eyes closed. It is anticipated that each test will take about one hour.

Participation in this experiment is entirely voluntary and you may choose to stop at any time. All information that you provide us is strictly confidential, and your data will be processed anonymously.

Please read the following and sign below:

In the unlikely event of physical injury resulting from participation in this research, I understand that medical treatment will be available from the MIT medical department, including first aid emergency treatment and follow-up care as needed, and that my insurance carrier may be billed for the cost of such treatment. However, no compensation can be provided for medical care apart from the foregoing. I further understand that making such medical treatment available, or providing it, does not imply that such injury is the investigator's fault. I also understand that by my participation in this study I am not waiving any legal rights (further information may be obtained by calling MIT's Office of Insurance and Legal Affairs at 253-2822) and that I may contact the Chairman of MIT's Committee on the Use of Humans as Experimental Subjects at 253-1772 if I feel that I have been treated unfairly as a subject.

I have read and understood the above statements and agree to participate as a subject in these experiments.

Signature:_____ Date:_____

APPENDIX C. APPARATUS CALIBRATION

This appendix describes the procedures that were followed during the initial calibration of the Virtual Workbench display for the experiments of the thesis. The procedures used to calibrate the PHANToM are not discussed here because that work was done independently by Lukasz Weber under the direction of Ben Sachtler. Rather than attempt to document someone else's work, the approach taken in this investigation was simply to test the PHANToM calibration. These tests were performed simultaneously with other tests in order to insure that the PHANToM was properly calibrated each time it was used. The following discussion assumes that the reader is familiar with the apparatus as described in section 3.2.

Important Note: While it should go without saying, the calibration depends on the fact that the positions of the display, mirror, and PHANToM are all fixed relative to each other. Any changes that distort the Virtual Workbench frame or otherwise move the components relative to one another will alter the calibration.

C.1 Display Calibration

There are two key assumptions inherent in the algorithm that is used to draw images on the display (see figure C.1). First, the physical location of pixels illuminated on the display screen must correspond to the calculated projection points in the experimental space. Second, the subject's pupils must be located at the specified nominal positions. Both factors depend on the experimental space being carefully defined with known physical reference points.

Figure C.2 shows the display reference points established on the Virtual Workbench for these experiments. The V_Space[1] Reference Display is a special grid pattern presented on the main monitor by the RunTest application (including grid lines spaced 1 cm apart, not shown in the figure). This display defines various important points in the experimental space; specifically: the origin, the boundary of the pixel array, and two marks used to line-up the reference LEDs. Note that the actual reference positions coincide with the virtual image of the display screen, as viewed in the mirror, and that the display is viewed in stereo mode.

---

[1]Note that "v_space" is used in program documentation to refer to the virtual space of the experiments.

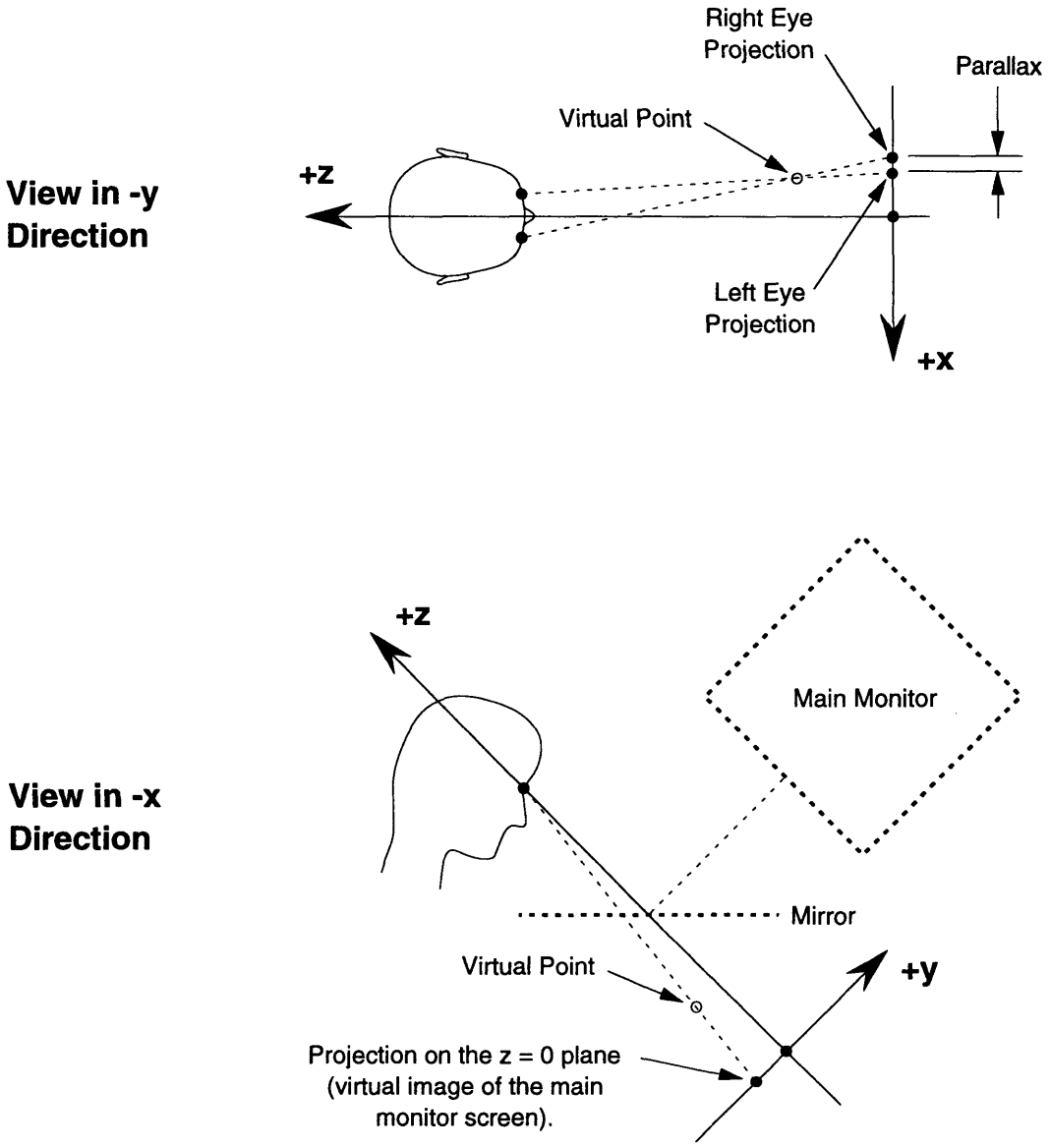


Figure C.1. Illustration of the left and right view projections of a virtual point on the display.

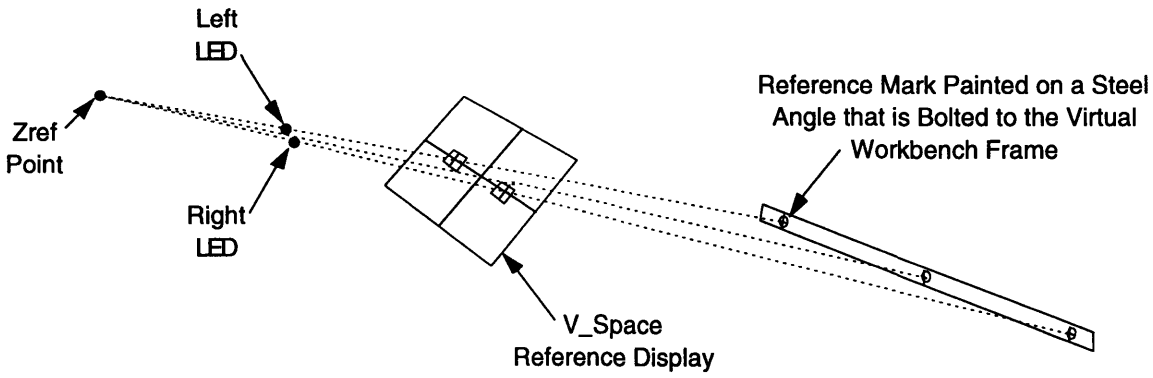


Figure C.2. Alignment of Zref Point, Reference LEDs, V_Space Reference Display, and Physical Reference Marks.

The Z_{ref} point is a special reference point that lies on the z axis at a known distance from the origin. The z coordinate of the reference point is the same as the nominal z coordinates for the centers of rotation of the subject's eyes. The exact point (Z_{ref} = 60.9 cm) was determined during the initial calibration when a bracket was physically mounted on the Virtual Workbench at that location. The Z_{ref} point is used to help locate the reference LEDs (which fall on lines between the Z_{ref} point and the two marks on the V_Space Reference Display as shown in the figure). Permanent physical reference marks are also located on the Virtual Workbench frame to make it easier to re-locate the Z_{ref} point in the future.

The calibration procedure involves four major steps; each building on the proceeding step. First the V_Space Reference Display is calibrated. Then the Z_{ref} point is located and this, in turn, is used to help locate the reference LEDs. Then the eye position algorithm is tested at the end.

C.1.1 Calibration the V_Space Reference Display

When any RunTest application based on version 2 of the underlying library is started on the Virtual Workbench, the first thing to come up on the monitor is the Main Menu. Item 3 of the Main Menu transfers control of the program to the Envr Set-up menu that is used to calibrate the display (among other things). Item 2 of the latter menu is the V_Space Reference Display. Items 21 (Vblank Set-up) and 22 (PixScreen Set-up) of the Set-up menu present still other calibration displays.

C.1.1.1 Display Measurements: In all of the following discussion the displays are described as if the experimenter is looking at the main monitor in the mirror. In other words "up" refers to the positive y direction. If the screen is viewed directly then the vertical orientation is reversed (i.e., the positive y direction is down). The horizontal orientation of the display is the same in either case (i.e., the positive x direction is to the right).

One reason that the mirror-view convention is used here is that the display is measured while looking in the mirror. Specifically, the display is measured by holding a ruler next to the virtual image of the display. Recall that the mirror is partially-silvered so that it is possible to see the ruler directly through the mirror while looking at the reflection of the display. A lamp carefully placed under the mirror can be used to illuminate the ruler without obscuring the display.

A flat panel, mounted under the mirror so that it is aligned with the virtual image of the display, serves as a hand rest while using the ruler. Note that a special procedure is needed to do the measurements because the monitor screen is curved. Specifically, the ruler should be held so that the measured points on the display do not move relative to the scale markings on the ruler when the experimenter moves his or her head from side to side.

C.1.1.2 Vblank Set-up: The first step in the calibration is to run the Vblank Set-up routine in order to determine the desired value of the parameter V_{blank}. This parameter controls the vertical alignment of the left and right-eye views on the display. The menu item draws two horizontal lines just to the left of the origin in the left-eye view; one at y = 0 cm and the other at y = 0 cm - 1 pixel. Note that the y = 0 line is on top.

The routine also draws a single horizontal line in the right-eye view just to the right of the origin at y = 0 cm. Ideally, when the value of V_{blank} is set properly and the display is viewed in stereo mode (without the glasses), the left and right lines at y = 0 should be at the same height. As it turns out, however, the left and right views are offset by half a pixel in the present implementation. Thus the two views are considered to be aligned when the right line is centered between the two left lines.

V_{blank} must be properly set before proceeding to the next step of the calibration. In order to adjust V_{blank} it is necessary to change the value of the extern constant "Vblank" in the source file ENVR_VW.CPP (or other

envr init file as discussed in ENVR.H) and then recompile the program. Note that increasing V_{blank} causes the right line to move up. The value of V_{blank} used in the thesis experiments is 45 pixels.

C.1.1.3 PixScreen Set-up: The next two steps in calibrating the V_Space Reference Display use the PixScreen Set-up display. This menu item actually allows the experimenter to select from 3 related displays using the keyboard: B) both eye views, L) left eye view only, and R) right eye view only. Figure C.3 presents a representative drawing that illustrates all three displays in stereo mode. Note that the displays should be viewed in stereo mode, but without the CrystalEyes glasses when doing the calibration.

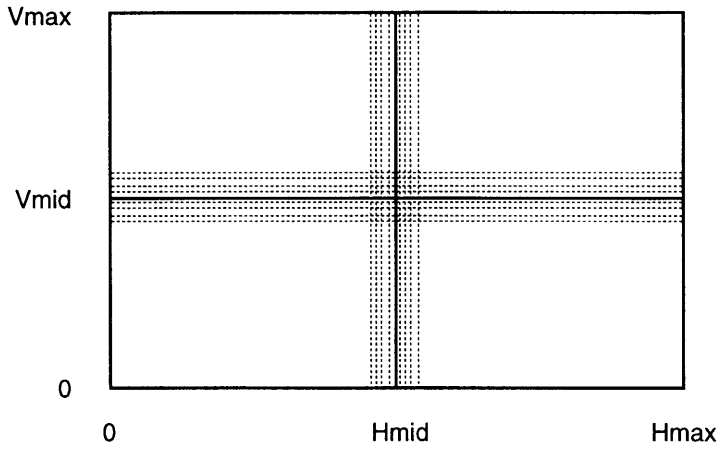


Figure C.3. PixScreen Set-up Display.

The key feature of the 3 displays is that they are all based directly on the pixel screen coordinates. All of the displays include the same white border showing the minimum and maximum pixel values. White lines are drawn at H_{mid} (vertical line) and V_{mid} (horizontal line). Several red lines are also drawn on the L and R displays parallel and adjacent to the H_{mid} and V_{mid} lines; spaced 2 pixels apart.

All of the lines are drawn across the entire screen buffer image. In other words, in terms of the screen coordinates (h,v), vertical lines (i.e., where h = constant) are drawn from v = 0 to v = V_{max} and horizontal lines (i.e., where v = constant) are drawn from h = 0 to h = H_{max}. Part of each vertical line is thus drawn in both the left and right views. This has no effect on the display in stereo mode, however, where the two parts of the line are exactly superimposed. Thus the vertical lines are identical on all three displays.

Horizontal lines that are not part of the border are distinguished, however, by the fact that V_{offset} is added to the vertical coordinate of the right view. For example, in the case of the V_{mid} line, $v = V_{mid}$ for the left view while $v = V_{mid} + V_{offset}$ for the right view. Assuming that V_{blank} is calibrated as described in section C.1.1.1 (recall that V_{offset} depends on V_{blank}), the white V_{mid} line and its associated red lines are all displaced down 1/2 pixel on the R display relative to the same lines on the L display. Similarly, the V_{mid} line on the B display is actually two lines spaced 1/2 pixel apart.

Main Monitor Adjustment: It is essential that the main monitor be properly adjusted and the geometry settings saved before proceeding with the rest of the calibration. These settings determine the locations of the pixels on the display that are the basis for all of the remaining calibration. While it should be easier to re-calibrate the apparatus after the initial calibration (particularly if the physical reference points are unchanged), changing the geometry settings on the monitor would, at least, make it necessary to review each of the steps in the procedure. The monitor's Contrast, Brightness, and Color Balance can be set at any time, but should be held constant for the duration of the experiments.[2] The operation of the monitor controls is described in the Hitachi CM2198MU User's Manual; in particular see pages 19-20 for the geometry adjustments.

The monitor is adjusted as follows:

1) Start the RunTest application and select the PixScreen, B display with the monitor in stereo mode.
2) Adjust the Horizontal and Vertical Size to the smallest possible settings. Note that the LED by the button blinks when the limit is reached.
3) Adjust the Horizontal Position, Vertical Position, and Side Pincushion approximately by eye.
4) Repeat the Side Pincushion adjustment while comparing the sides of the display image to the edge of a ruler. Note that this step and subsequent steps use the measurement procedure described in section C.1.1.1. In the thesis experiments the sides of the display appeared to be straight within ± 1 mm at all points.

---

[2]To the best of the author's knowledge the main monitor settings were not changed during the experiments. The Color Balance was observed to be standard setting No. 2 at the end. This setting corresponds to a color temperature of approximately 6500K according the Hitachi manual.

5) Adjust the Trapezoid so that the top and bottom of the display are the same length (horizontally) and verify that the width is the same in the middle of the display (i.e., measure the length of the V_{mid} line). Repeat the Side Pincushion and Trapezoid adjustments as necessary until the display image is the same width from top to bottom. In the current experiment the width was observed to be 33.5 cm.
6) Measure the distance between the boundary of the display and the edge of the screen (dark edge on the face of the tube) at various points around the perimeter. Adjust the Horizontal Position, Vertical Position, and Rotation as necessary until the display image is centered and square with the face of the tube.
7) Push the store button to save the geometry settings. Note that the monitor is able to save user geometry settings for up to 24 different video modes. In other words the geometry can be set differently for different modes such that the monitor automatically restores the saved setting when a given mode is used. The manual warns that two video modes with similar horizontal and vertical frequencies might not be distinguished however. Also note that the monitor only saves one Rotation setting (i.e., different Rotation settings are not saved with the other geometry settings for different video modes). Similarly the monitor saves only one setting each for Contrast, Brightness, and Color Balance.

Pixel Position Measurements: The next step in the calibration is to determine the values of X_{max}, Y_{max}, H_{mid}, and V_{mid}, using the PixSreen L and R displays. This is done as follows:

1) Select either the L or R display with the monitor in stereo mode.
2) Locate the center of the display using the measurement procedure described in section C.1.1.1. Note that (X_{max} = display width (cm)/2) and (Y_{max} = display height (cm)/2).
3) If H_{mid} and V_{mid} are set correctly then the horizontal and vertical center points on the display will fall exactly on the white H_{mid} and V_{mid} lines. Otherwise the center points will not fall on the white lines. In the latter case use the red lines, that are spaced 2 pixels apart, to estimate the number of pixels from H_{mid} or V_{mid} to the middle of the display. Repeat the measurements at various locations using both the L and R displays.
4) Adjust the parameters H_{mid}, V_{mid}, X_{max}, and Y_{max} as required. The parameters are adjusted by changing the value of the corresponding extern constants in the source file ENVR_VW.CPP (or other envr init file as discussed in ENVR.H). Note that the program must be recompiled to effect the change.

5) Repeat steps 1 to 4 until the parameters are set correctly (as well as possible).

C.1.1.4 Verification: At this point the parameters of the V_Space Reference Display should be calibrated (or very nearly calibrated). Put the V_Space Reference Display on the screen and verify its dimensions using the procedure described in section C.1.1.1. Revise the values of the parameters determined in section C.1.1.3 as necessary.

In the thesis experiments small changes (1 pixel in magnitude) were made to the parameters at this stage of the calibration after repeated measurements. Two types of discrepancies were observed. First, small relative distances might be in error by up to ± 1 pixel (about ± 1 mm vertically or ± 0.5 mm horizontally). This problem is, of course, due to pixelation. Second, several short distances add up to a larger number than the total measured distance. This problem is due to screen curvature which is neglected in the s_point algorithm.

The final values used in the program are: H_{mid} = 315 pixels, V_{mid} = 107 pixels, X_{max} = 16.8 cm, and Y_{max} = 10.8 cm. Allowing for measurement errors and pixelation, it is estimated that the absolute position of a point on the calibrated display is within ± 2 mm in the y direction (vertically) or ± 1 mm in the x direction (horizontally) of the intended projection point on the z = 0 plane.

Screen Curvature: The z coordinates of points on the screen can differ significantly from z = 0 (the value assumed by the s_point algorithm). Specifically, with z defined to be zero at the midpoint on the screen, the screen curves back (toward negative z) as one moves away from the origin. The difference is relatively small near the center of the screen, but the edge of the V_Space Reference Display (as calibrated for the thesis experiments) is nearly 1 cm behind the z = 0 plane. The following table presents the results of measuring the curvature in the x direction at y = 0:

x (± cm):	0.0	2.0	4.0	6.0	8.0	10.0	12.0	14.0	16.0	16.8
Δz (-cm):	0.00	0.00	0.05	0.10	0.20	0.30	0.47	0.60	0.82	0.95

These results were obtained by forming a draftsmen's flexible curve-drawing tool to the screen and then transferring the curve to a piece of paper. Except in the case of the origin, the observed Δz is the mean of the two measurements at plus and minus the indicated x coordinate.

C.1.2 Locating the Z_{ref} Point

Once the V_Space Reference Display is calibrated, the next step is to locate the Z_{ref} point. This was initially done in the thesis experiments by mounting a bracket (the "Z_{ref} support") at approximately the right location. Then, through a series of measurements and small adjustments, the support was finally located at an acceptable Z_{ref} point as follows: $x = 0 \pm 0.5$ cm, $y = 0 \pm 0.5$ cm, $z = 60.9 \pm 0.1$ cm. Note that the indicated tolerances are estimated from measurements made from the Z_{ref} point to the corners of the V_Space Reference Display. A special laser assembly was used to do the measurements.

C.1.2.1 Laser Measurement Procedure: Figure C.4 presents a photograph of the laser assembly that was used in the thesis experiments. The assembly consists of two small lasers (removed from pen lecture-pointers), mounted 16-17 cm apart on an adjustable frame. A center support is mounted approximately in the middle of the bar ("cross bar") connecting the two lasers. The center support, which extends approximately 2.5 cm perpendicular to the axis of the cross bar, has a spherical end ("ball") that measures 7 mm in diameter.[3]

This apparatus was given to the experimenter by Ben Sachtler who developed it, with Tom Wiegand, for doing related work on the Virtual Workbench. It proved to be useful for measuring distances in the experimental space in cases where the interposition of the mirror prevents a more direct method. The measurement procedure is as follows:

1) Place the end of the center support at a point that defines one end of the distance to be measured. For example, the ball might be placed on the Z_{ref} support so that its tip (i.e., the part furthest from the cross bar) is at the Z_{ref} point.
2) Align the lasers so that they converge to the point that defines the other end of the distance to be measured. For example, the lasers might be directed into the mirror to converge on the center of the V_Space Reference Display (i.e., the origin of the experimental space). Note that the arrangement is such that the distance from the Z_{ref}

[3]Although the laser assembly proved adequate for the task, following are three suggestions for improving it in the present application: 1) the ball at the end of the center support should be replaced by a sharp point, 2) the mechanical linkages of the assembly should be redesigned for fine adjustment, and 3) lasers with a smaller diameter beam should be used.

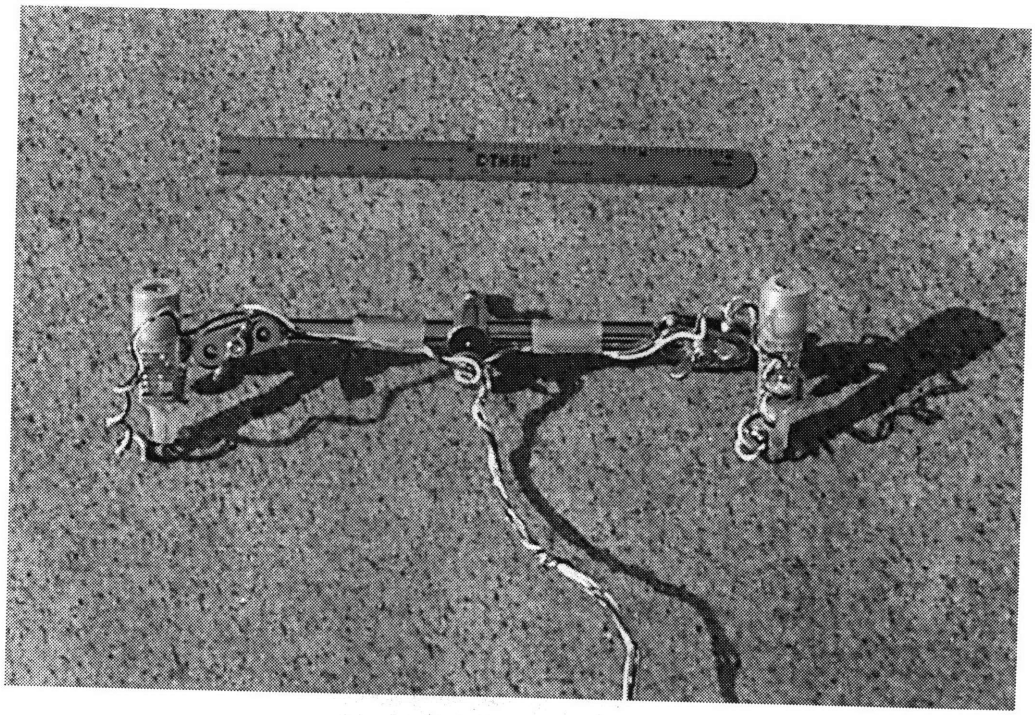


Figure C.4. Laser Assembly.

point to the physical monitor screen (via the mirror) is the same as the distance to the virtual image of the screen.

3) Now, being careful not to readjust the laser assembly, take the assembly aside and measure the distance from the ball to the point where the two lasers converge. This, of course, is the desired distance. A simple way to do this measurement is to clamp a meter stick perpendicular to a wall so that one end of the stick butts-up against the wall. Then position the assembly so that the ball is on the scale and the two lasers converge to a point on the wall next to the end of the meter stick. The distance is then read off the scale.
4) Although there is some uncertainty in the procedure, repeated measurements produce mean values within reasonable tolerances. Uncertainties involve: 1) judging the point on the ball used in the measurement, 2) the fact that the mechanical linkages sometimes appear to stick in the same place, and 3) judging when the two spots are converged.

C.1.2.2 Initial Alignment of the Z_{ref} Support: The Z_{ref} support was constructed from 1/2" (1.3 cm) diameter aluminum rod and lab-stand clamps like those used to make the headrest. The support consists of a short length of rod (approximately 15 cm long) clamped vertically to the middle of a horizontal rod. The horizontal rod, in turn, is mounted between the two vertical headrest rails. The arrangement is such that the top end of the vertical rod provides a small platform that may be used to physically mark the location of the Z_{ref} point. A pen mark on the end of the rod locates the exact Z_{ref} point at the edge nearest to the Virtual Workbench (see figure C.5).

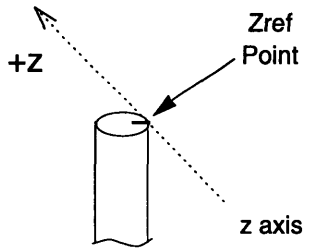


Figure C.5. Z_{ref} Support

Initial alignment of the Z_{ref} support is done by eye so that the marked point is approximately on the z axis and 60 cm (or slightly further) from the origin of the experimental space. Next the ball of the laser assembly is placed at the marked point on the support and the lasers are adjusted so that they converge on one corner of the V_Space Reference Display. The

lasers are then pointed at the other corners of the display (without changing the adjustment) to see if the distance to each of the corners is the same.

Z_{ref} support is adjusted as required until the point marked on the support is equidistant from all of the corners of the display. The goal is to locate the marked point precisely on the z axis at approximately the right distance from the origin. The exact value of Z_{ref} is then measured with the laser assembly.

C.1.2.3 Laser Measurements: Once the Z_{ref} support is approximately in position (using the procedure described in section C.1.2.2), repeated measurements are made (according to the procedure of section C.1.2.1) to verify the exact position. Specifically, one takes the mean of several measurements of the distance to each of the corners of the display. Fine adjustments of the Z_{ref} support are made as required and the measurements are repeated until the marked point is on the z axis within the desired tolerance.

Following are the results of the final set of laser measurement at the Z_{ref} point. The mean of 12 measurements for each distance is reported (± one standard deviation):

Display Point		Distance to Z_{ref} point (cm)
origin (Z_{ref})	=	60.92 ± 0.17
top left	=	65.37 ± 0.17
bottom left	=	65.39 ± 0.20
top right	=	65.33 ± 0.22
bottom right	=	65.44 ± 0.16

C.1.2.4 Physical Reference Marks: Physical reference marks are painted on a steel angle bolted to the Virtual Workbench frame as shown in figure C.2. These reference marks were added to make it easier to relocate the Z_{ref} point in the future (if necessary) after the Z_{ref} support was removed following the calibration. The procedure is as follows:

1) Set-up the V_Space Reference Display on the main monitor.
2) Mount a single laser so that its beam just grazes the edge of the Z_{ref} support and also intersects the origin of the experimental space (i.e., the laser spot is centered on the middle of the V_Space Reference Display).

3) Mark the location of the laser spot on the steel angle attached to the Virtual Workbench frame. Note that a white piece of paper can be taped to the angle to make the spot more visible. The paper is first marked with a pencil and the location of the mark is confirmed by sighting along the z axis. Then, when the mark is in the correct place, transfer it through the paper with a scribe and permanently mark the angle with white paint.
4) Repeat steps 2 and 3 for the Reference LED marks on the V_Space Reference Display. Note that the location of the marks is determined by the parameter, M, that may vary depending on the exact value of Z_{ref} (see section C.1.3).

C.1.3 Mounting the Reference LEDs

The exact procedure used to mount the Reference LEDs during the initial calibration of the thesis experiments was not recorded. The author's notes and memory suggest, however, that the procedure is approximately as follows:

1) Install a new 1/4" (0.6 cm) thick Lexan bar for mounting the Reference LEDs (i.e., a bar without holes drilled for the LEDs) under the mirror. Note that the bar should have mounting holes with generous clearance. It should also be mounted square with the mirror and with its face parallel to the xy plane. Use the laser assembly to measure the distance from Z_{ref} to the bar according to the procedure described in section C.1.2.1 and make an initial estimate of the parameter Z_m (the z coordinate of the Reference LEDs).
2) Set-up the V_Space Reference Display on the main monitor. Note that M (the absolute value of the x coordinate of the Reference LED marks on the display) depends on the values of Z_{ref} and Z_m. Specifically M is selected as large as possible, but so that subjects with large IPDs will still be able to line-up the LEDs with points inside the boundary of the display. In the thesis experiments the round value of M = 10.0 cm was selected.
3) Put scotch tape on the Lexan bar at the approximate locations of the two LEDs and the z axis.
4) During the procedure described in section C.1.2.4 (or using a similar procedure) make and verify pencil marks on the scotch tape that locate the two Reference LEDs and the z axis. Also verify that the distances between the marks agree with calculations based on the values of Z_{ref} and Z_m.
5) Remove the Lexan bar and drill holes at the three indicated locations. The LED mounting holes are drilled with a 0.116" drill and

counter-sunk for the LED flange. Note that the operation should be done with a milling machine so that the holes are all on the same line and precisely spaced. In other words the pencil marks are only used as a guide.

6) Press the LEDs into the mounting holes in the Lexan bar and solder on the wire leads so that they are as unobtrusive as possible.
7) Re-install the Lexan bar so that the center hole and the Reference LEDs are properly aligned with Z_{ref} and the V_Space Reference Display.
8) Mount a horizontal bar on the Virtual Workbench side (and flush with the top) of the Z_{ref} support.[4] The bar should be aligned with the x axis (V_{mid} line on the display) and parallel with xy plane. Note that this bar is also used in the eye_cal tests discussed in section C.1.4.
9) Mark the top of the horizontal bar at the center adjacent to the Z_{ref} point. Also mark two points on either side, equidistant from the center, with the same spacing as the Reference LEDs.
10) Measure the distance from the two outer points on the horizontal bar to approximately corresponding grid points on the V_Space Reference Display (using the procedure of section C.1.2.1) and verify that the bar is parallel to the xy plane.
11) Measure the distance from the two outer points on the horizontal bar to the Reference LEDs. Note that the distance is measured from the side of the bar away from the Virtual Workbench, i.e., at the same z coordinate as Z_{ref}. Also note that paper targets can be taped over the face of the holes in order to make it easier to see the laser spots on the assembly. Adjust the position of the Reference LED assembly as required so that both LEDs are at the same z coordinate.
12) Use the value of z determined in step 11 for the final value of Z_m and update the value of that parameter in the source file ENVR_VW.CPP (or other envr init file as discussed in ENVR.H). Note that the program must be recompiled to effect the change. The final value used in the thesis experiments is Z_m = 38.0 cm.
13) As a final note it should be pointed out that small errors in the physical alignment of the Reference LEDs can be compensated for by modifying the algorithms of the eye_cal class.

---

[4]In the current experiments a 1/8" (0.3mm) thick extruded aluminum bar was clamped to the vertical headrest rails so that its -z face was 4 mm towards the Virtual Workbench horizontally from the Zref point and the top of the bar was less than about 1/2 mm below the top of the Zref support.

C.1.4 EYE_CAL Tests

The following section describes tests that were performed in order to verify the eye_cal algorithm prior to the start of the experiments. These initial tests led to a modification of the program's underlying library before the start of the experiments; the main change was the addition of correction factors to the eye_cal::CalcEyeCoords function. It is assumed that these corrections are needed primarily because the Reference LEDs are not exactly in the right places and to allow for curvature of the screen. The exact kinematics of the eyes is not fully understood however.

A summary of the results of actual eye position measurements that were made routinely during the experiments is also presented. It was observed that the measured eye positions cluster around the nominal positions within the estimated tolerance (x±1cm, y±1cm, z±2cm).

C.1.4.1 Verification Method: The eye_cal algorithm was tested using the horizontal bar described in step 8 of C.1.3. The bar was marked at several points corresponding to simulated IPDs of 3.5, 4.5, 5.5, 6.5, and 7.4 cm.[5] A series of tests was then performed where the experimenter made repeated measurements of the simulated eye positions by sighting over the bar with one eye. Figure C.6 shows the experimental set-up.

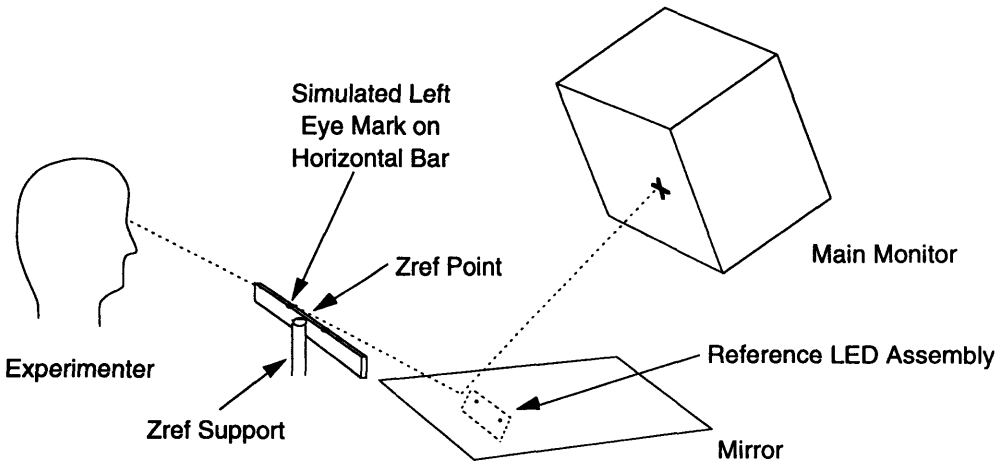


Figure C.6. EYE_CAL Test Set-up.

[5]The IPD of 7.4 cm was originally 7.5 cm, but with that distance it was not possible to adjust the right-most mark on the display far enough to the right. The problem was solved by moving the simulated left-eye position 1 mm to the right. This resulted in the eye positions not being symmetric about x = 0 (specifically, Lx = -3.65 cm and Rx = +3.75 cm), but that does not affect the results of the test.

The tests were performed using an early version of the RunTest application (v2.100) with test documents that were manually created for each test. The test documents defined a test with one group consisting of one (dummy) trial. The actual data was collected during the (normally routine) eye position measurements before and after the group. Note that the experimenter had considerable flexibility in performing the tests because the application allows the experimenter to repeat eye position measurements as many times as needed; saving all the observed positions. The experimenter only had to keep track of which measurements corresponded to which simulated IPDs in his lab notebook.

In some of the later tests a separate test document was used for each simulated IPD. This was done because the IPD (one of the standard parameters of the test document) is used to calculate some of the correction factors that were added to eye_cal algorithm. Even in these cases, however, it was possible (and sometimes necessary) to subtract out (and/or add in) corrections after the test was complete. The latter was done with Matlab when the data was analyzed.

C.1.4.2 Results and Correction Factors: Figure C.7 presents the uncorrected results of the final eye_cal tests that involved 10 measurements at each simulated eye position.[6] The figure presents the mean value (x) with error bars showing plus or minus one standard deviation.[7] Note that the simulated eye positions differed from the nominal centers of rotation of the eyes because it was necessary to look over the bar. Specifically, the simulated eye positions were measured to be $x = \pm$ IPD/2 (except as noted for IPD = 7.4), $y = 0.44$ cm, and $z = 60.64$ cm; within an estimated tolerance of $\pm$ 1 mm relative to the Z_{ref} point.

It can be seen in the figure that there are systematic errors with respect to the simulated eye positions. Specifically, the calculated x coordinates are displaced uniformly from the simulated positions (indicated by horizontal lines in the figure) by about 1/2 mm, independent of the IPD. The y-coordinates show no systematic error and the z coordinates show a significant change in error with IPD.

---

[6]The data presented in figures C.7 and C.8 is from tests EYECAL81.M-EYECAL85.M (10/24/95) and EYECAL91.M-EYECAL95.M (10/25/95). The number 9 tests included all corrections in the RunTest application while the number 8 tests included only the correction for z. Corrections are later added or subtracted as needed for the figures using Matlab. Also note that there were some later tests at other simulated eye positions that are not reported here.

[7]The error bars are very hard to see in the x and y plots because they are so small.

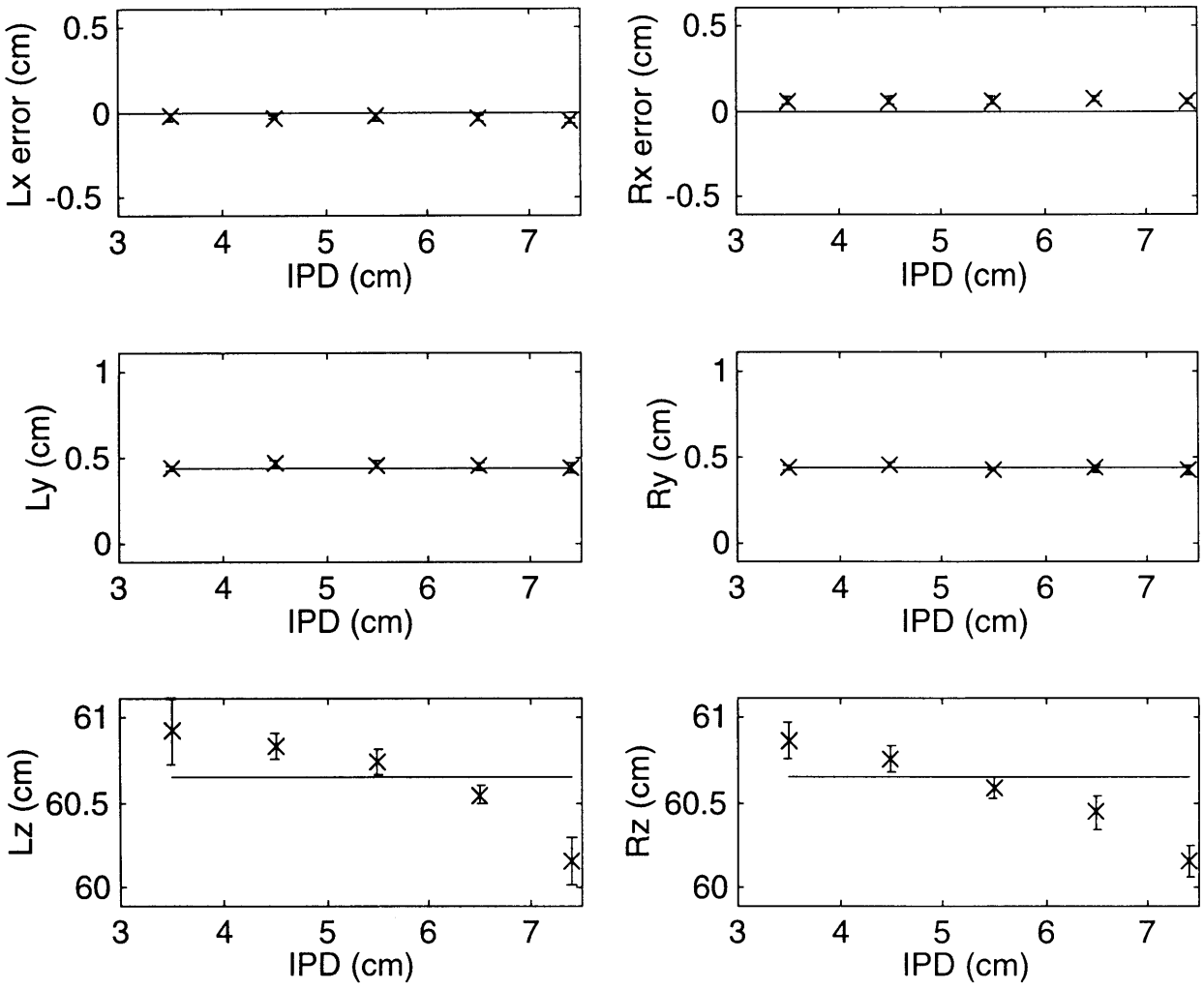


Figure C.7. Uncorrected EYE_CAL Results.

Figure C.8 presents the corrected results for the same tests in a similar plot. The correction factors, calculated based on a least-squares fit to the observed error,[8] are given by the following equations:

$$\begin{aligned} Lx_{corrected} &= Lx_{uncorrected} + 0.0308 \\ Rx_{corrected} &= Rx_{uncorrected} - 0.0680 \\ \\ Ly_{corrected} &= Ly_{uncorrected} \\ Ry_{corrected} &= Ry_{uncorrected} \\ \\ Lz_{corrected} &= Lz_{uncorrected} + 0.2051\ d_{ip} - 1.1314 \\ Rz_{corrected} &= Rz_{uncorrected} + 0.1638\ d_{ip} - 0.8430 \end{aligned} \qquad \text{(C.1)}$$

where:

L_i = calculated left-eye coordinate (i = x,y,z; corrected, uncorrected)
R_i = calculated right-eye coordinate (i = x,y,z; corrected, uncorrected)
d_{ip} = subject's IPD measured with a ruler while subject is looking straight ahead (see appendix B).

C.1.4.3 Summary of Eye Position Measurements: In an effort to assess the accuracy of the eye_cal algorithm and the subjects' eye positions, the eye_cal data from all v2.102 tests (T10, T11, T40, and T42) was investigated. Figure C.9 plots the measured eye positions for each of the 9 subjects. It can be seen in the figure that the measurements cluster around the nominal eye positions (centers of eye rotation) as desired; with somewhat more variability for z than for x and y. The variability observed in the figure, along with various other calibration measurements, may be summed to estimate the eye position tolerances that are presented in section 3.2.1.4 (x±1cm, y±1cm, z±2cm).

Specifically, the laser measurements of section C.1.2.3 are used to estimate that the absolute position tolerance of Z_{ref} is about x±0.5cm, y±0.5cm, z±0.1cm. The simulated eye positions in the eye_cal test are estimated to be correct ±0.1cm relative to the Z_{ref} point. The results presented in figure C.8 suggest that the eye position tolerances of the eye_cal procedure are about x±0.1cm, y±0.1cm, z±0.2cm. Finally, using the

---

[8]The z-corrections are actually based on the errors observed in tests EYE_CAL4.M, EYE_CAL5.M, EYE_CAL6.M, EYECAL71.M-EYECAL75.M (10/21/95 to 10/23/95) which had a combined total of 20 measurements at each eye position. The tests were repeated in EYE_CALs 8 and 9 because the part of the algorithm used to calculate x was changed. The x-corrections are based on tests EYECAL81.M-EYECAL85.M (10/24/95) that involved 5 measurements at each position.

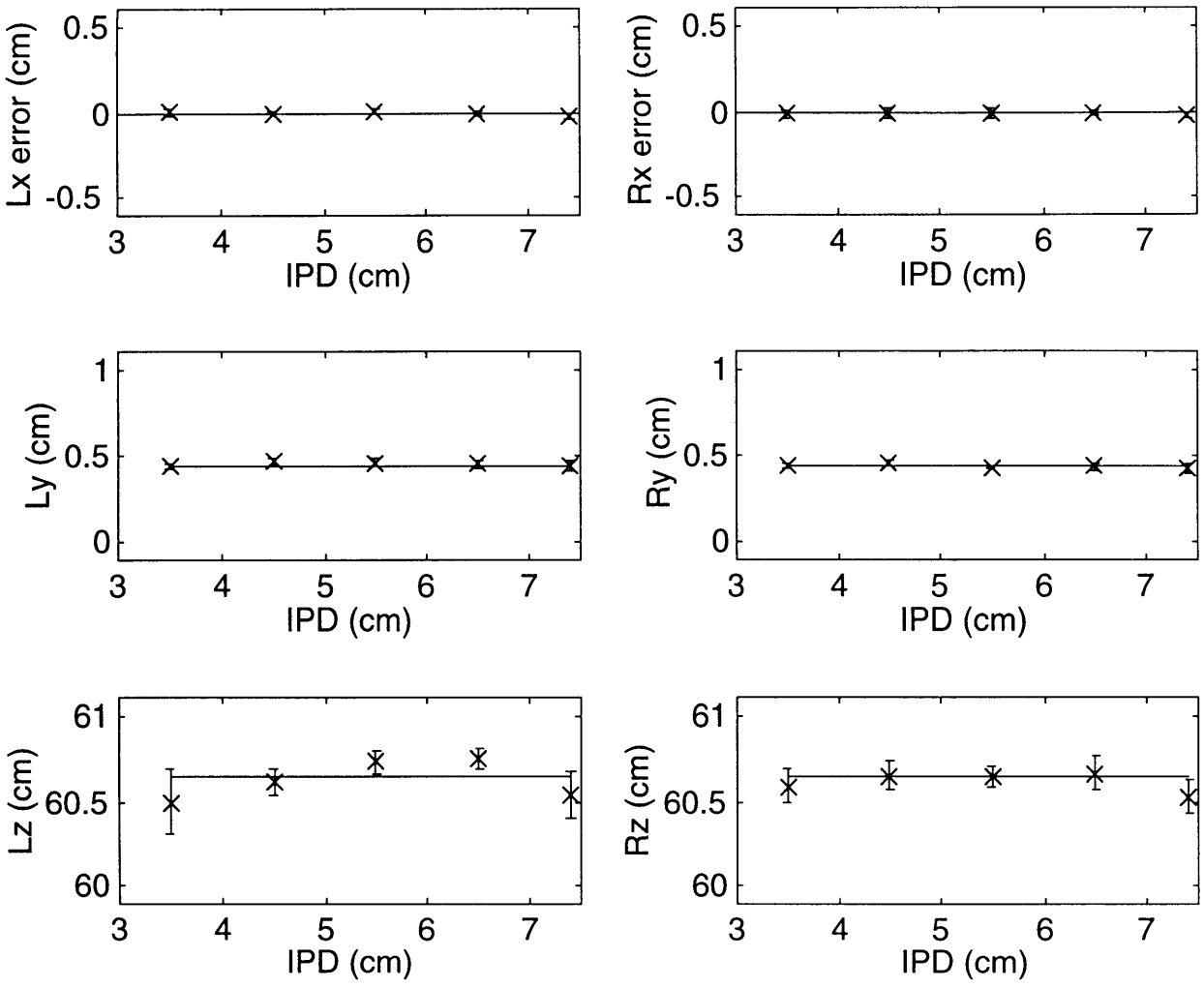


Figure C.8. Corrected EYE_CAL Results.

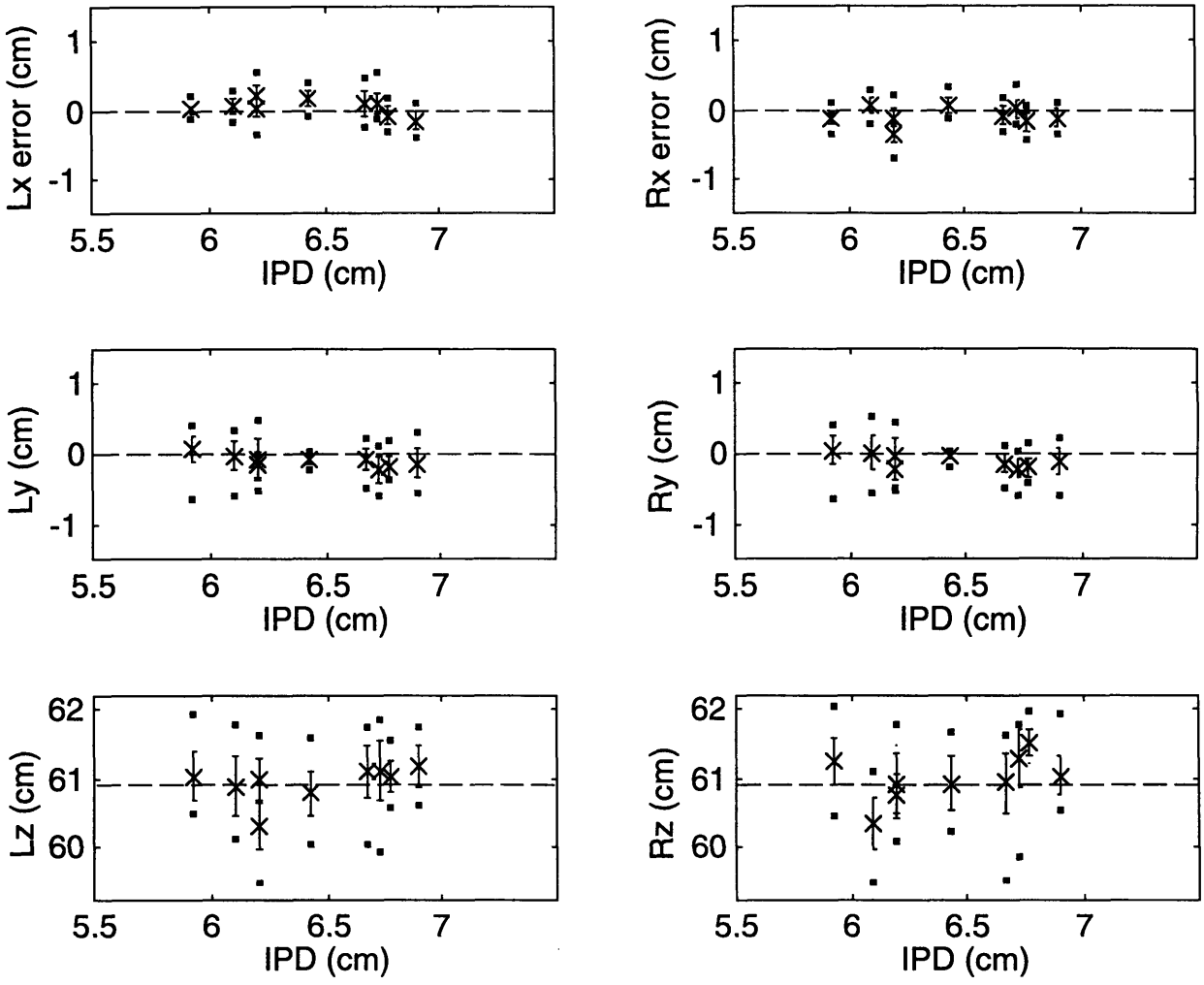


Figure C.9. Summary of all eye_cal measurements in v2.102. Plots show means (±1 std. dev.) for each subject. Dots give maximum and minimum values. Dashed line is the nominal coordinate (note that x plots show error = mean - nominal).

max and min points in figure C.9, subject variability is observed to be generally within x±0.5cm, y±0.5cm, z+1.0/-1.5cm. Uncertainty in R_{eye} (assuming it is only known within ±50%) increases the uncertainty of the pupil coordinates as follows: x±0.1cm, y±0.0cm, z±0.5cm. Summing the individual tolerances one gets: x±1.3cm, y±1.2cm, z±1.9/-2.4cm. Allowing for the crude nature of the analysis, one rounds off to the stated eye position tolerances.

Figure C.10 presents a second analysis of the eye_cal data, comparing the values of IPDs measured by the eye_cal algorithm with the IPDs measured using a ruler. It can be seen in the figure that the eye_cal procedure tends to underestimate the IPD by about 0.1 or 0.2 cm on average. The reason for this difference is not known. It may represent an actual shift in the centers of rotation of the eyes when they converge on the display. Subjects might also do the calibration without looking directly at any of the marks on the display so that the measured position corresponds more to the pupil position than the center of rotation. The discrepancy might also be due to a bug in the function eye_cal::CalcXmXe_x that is used to position the subject's head. In the latter case the difference would result because the subject learns to line-up the LEDs and the reference marks in the wrong way.

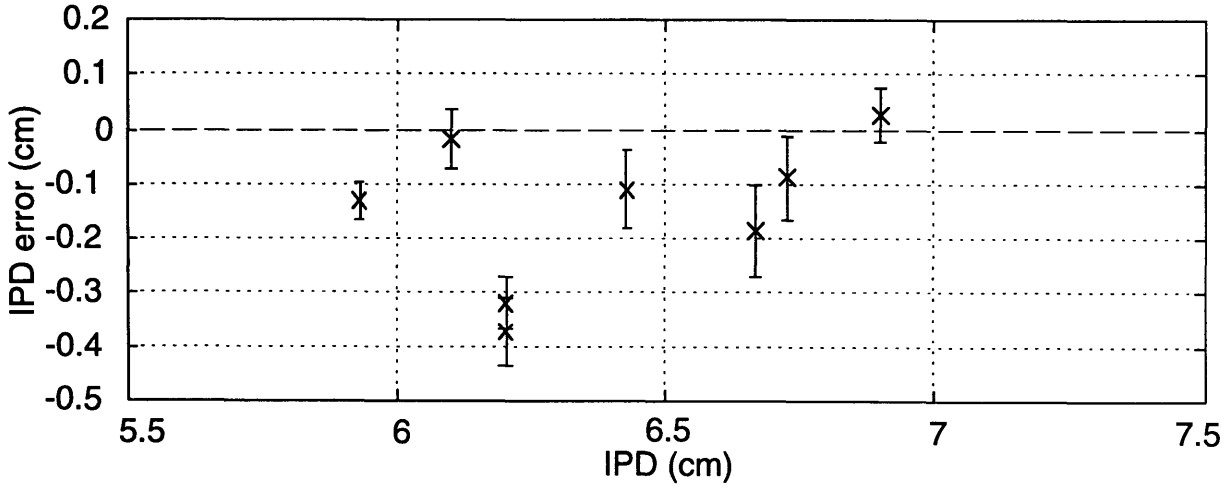


Figure C.10. Difference between the IPD measured with a ruler (see appendix B) and the IPD measured with the eye_cal algorithm in the last eye_cal measurement prior to the start of each group in every v2.102 test (except for subject S1 tests). IPD error = eye_cal measurement - ruler measurement.

C.2 PHANToM Calibration Tests

Two tests of the PHANToM calibration were performed in conjunction with the thesis experiments. The absolute position accuracy on the z axis was measured in every test where the slider was used. An alternative relative calibration test, like the ones described in section 4.1 except with many fewer trials, was also used to check the calibration of the PHANToM in experiments where the slider was not used.

In the absolute tests it was observed that the absolute accuracy of the PHANToM position was better than x ± 0.4 cm, y ± 0.4 cm, z ± 0.2 cm on the z axis (in the range -10cm $\leq$ z $\leq$ 10 cm) with repeatability better than ± 0.5 cm. This calibration test is possible because the slider end-fixture provides a convenient position reference. Hence a necessary part of measuring the absolute position, that will be discussed first, is the initial alignment of the slider itself. Also note that uncertainty in the alignment of the slider shaft is responsible for half of the reported PHANToM tolerance.

C.2.1 Slider Alignment

Various reference marks were established during the initial alignment of the slider assembly so that it could be removed and re-installed easily (see following discussion of the initial alignment procedure). First, the lower shaft mounting bracket (the "flag pole" bracket) was permanently attached to the Virtual Workbench so that the -z end of the slider shaft could always be positioned exactly[9] at the same point relative to the frame. The upper mounting brackets of the assembly were also permanently attached to the frame and the assembly's horizontal bar was marked so that it could be re-installed each time at approximately the same location relative to the upper brackets. The arrangement is such that the +z end of the shaft was always located within ±0.1cm of the nominal z axis defined by the Z_{ref} point.

More generally, assuming the -z end of the shaft is located within the same tolerance, any point along the axis of the slider shaft is located relative to the nominal z axis at the coordinates: x = 0.0 ± 0.1 cm and y = 0.0 ± 0.1 cm. Allowing for the uncertainty in the absolute position of the Z_{ref} point relative the experimental space (i.e., x±0.5cm, y±0.5cm,

[9]The end of the stainless steel shaft bottomed against the steel angle and was clamped lightly inside the aluminum flag pole bracket by a nylon screw. The positioning tolerance of metal clamped against metal is at least within ±0.005cm.

z±0.1cm), one can estimate that the axis of the shaft is aligned to the absolute z axis of the experimental space within ±0.1cm at the origin and ±0.2cm in the range: -10cm $\leq z \leq$ 10 cm.

In addition, the distance from the origin of the experimental space to the -z end of the shaft was measured (as discussed in the following procedure to locate z = 0) so that z coordinates could be marked along the shaft for the absolute position tests described in section C.2.2. The location of these marks is estimated to be within ±0.2cm of the absolute z coordinate and ±0.1cm relative to one another.

Initial alignment procedure:

1) Mount the lower (flag pole) bracket so that the shaft mounting hole is centered on the middle reference mark (see section C.1.2.4) when viewed from the Z_{ref} point.
2) Mount a single laser on the headrest rails so that its beam is aligned with the z axis of the experimental space. Specifically the beam should intersect the Z_{ref} point, the center hole in the reference LED assembly, the origin on the V_Space Reference Display, and the physical reference mark centered inside the flag pole bracket hole.
3) Install the slider assembly so that the laser spot is centered on the +z end of the slider shaft.
4) Tighten the hose clamps holding the upper brackets to the Virtual Workbench frame. Also tighten the bolts that clamp the slider assembly's horizontal bar into the upper brackets. Note that the assembly is properly installed when the upper brackets are tight, the laser spot is centered on the end of the slider shaft, and the shaft slides freely through the block on the horizontal bar.
5) Tighten the set screws at the top and the bottom of the shaft. Make them sung as in normal use.
6) Observe the slider from a point about 1 cm to the side of the Z_{ref} point. The block should appear to move less than ±0.1cm in x and y, relative to the V_Space Reference Display, as the slider is moved over its entire length of travel. This was observed during the initial alignment of the slider assembly in the thesis experiments.
7) Repeat steps 4 to 6 as needed to meet the desired position tolerance and, then, mark the slider assembly's horizontal bar at the inside edges of the two upper brackets.

Procedure to locate z = 0:

1) Clamp the laser assembly near the Z_{ref} point and adjust the two lasers so that they converge on the V_Space Reference Display about one cm to the right of the origin.
2) Move the slider until the laser spots converge on the front face of the finger-hold pin sticking out of the side of the slider block. Then lock the slider in place by putting a clamp on the shaft under the block. Note that a piece of paper held against the front of the pin will serve to make the spots more visible.
3) Estimate the region where the laser spots converge by moving the piece of paper back and forth in the z direction (note that the pin can be rotated out of the way). Also read the PHANToM position using the "Read Phantom Position" option (item 4) on the Envr Set-up menu.
4) Repeat steps 2 and 3 a few times to verify that the z = 0 point is properly located; the lasers should be readjusted each time as in step 1. Note that in the thesis experiments the convergence region was observed to be ±0.1cm and the PHANToM position was -0.1cm. Allowing for the radius of the pin, which displaces the endpoint of the PHANToM -0.15 cm in the z direction, the calibration of the PHANToM was found to agree with the z = 0 point in the experimental space within the accuracy of the laser measurement (i.e., ±0.1cm).
5) Make a mark on the slider shaft with a permanent marking pen such that the -z edge of the mark coincides with the +z face of the slider block when the slider is at the z = 0 point (i.e., allow for the radius of the finger-hold pin by making the mark about 1 mm in front of the face of the block when it is located using this procedure).
6) Measure the distance from the -z edge of the z = 0 mark to the -z end of the slider shaft. This distance was observed to be 22.0 cm in the thesis experiments.

C.2.2 Absolute Position Tests

Figure C.11 presents a summary of the PHANToM calibration tests performed on the z axis. The *pcal* data was taken in connection with the v2.101 z-axis calibration experiment discussed in section 4.2 (note: the name "pcal" stands for PHANToM calibration). In these tests the slider shaft was marked with a permanent marker pen. The marks were located at 5 nominal z coordinates (-10, -5, 0, 5, 10 cm) on the z axis. They were located by placing the slider shaft on a table and measuring the distance from each mark to the end of the shaft within ±0.1cm. The shaft needed to be remarked almost every time the slider was used because the marks

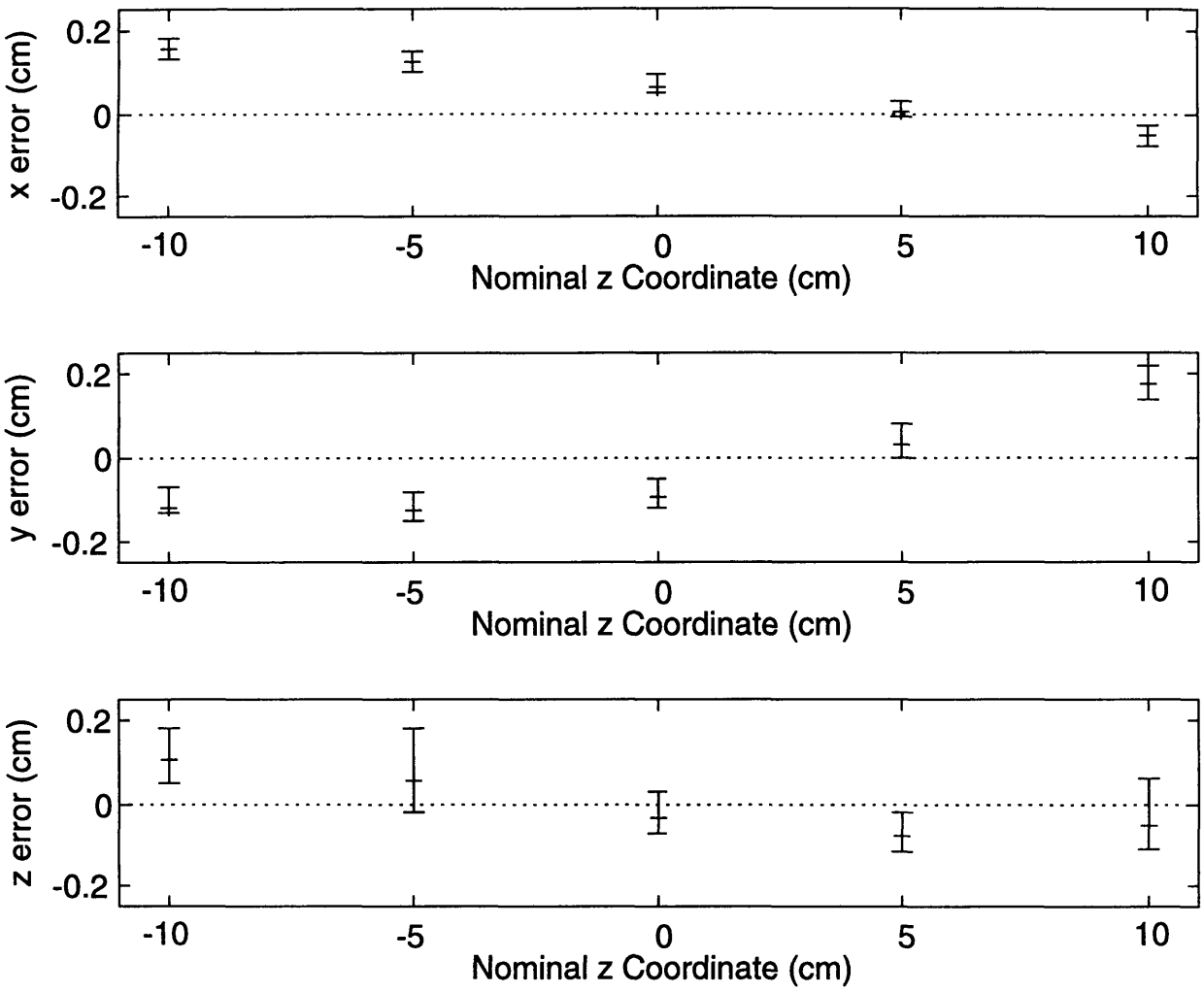


Figure C.11. Summary of all v2.101 absolute position tests of the PHANToM connected with the z-axis calibration experiments. Center mark is the mean. Error bars show the maximum and minimum values.

started to wear off after a few hundred trials. Sometimes the shaft was simply remarked in place by marking over the old marks.

In these tests the experimenter reset the PHANToM encoders at the reference point located at the bottom of the slider shaft. The original PHANToM reference point (where the PHANToM is moved to the extreme lower-left corner of its workspace, determined by a stop on the device) was also used in some cases. The pcal test was then performed either before or after the main test while the encoder calibration was unchanged.

In the pcal test the experimenter positioned the slider at the appropriate mark on the shaft and then pressed return on the keyboard to input the position. Two trials were performed at each nominal position in a standard order such that the PHANToM was moved between each measurement. Test documents from the tests were typically saved with a file name according to the following code: Pmmddyyn.m, where "mm" is the month (01 to 12), "dd" is the day of the month (01 to 31), "yy" is the year (e.g., 95), and "n" is the number of the test on that day (1 to 9).

The results show that the position measured by the PHANToM did not deviate from the nominal position by more than ± 0.2 cm in x, y, or z on the z axis (in the range -10cm $\leq z \leq$ 10 cm). Indeed, since most of the variation in the measurements is due to errors in positioning during the test, it is estimated that the repeatability of the PHANToM measurements is better than ± 0.05 cm. The latter is consistent with other observations.

The absolute calibration of the PHANToM is estimated to be better than x ± 0.4 cm, y ± 0.4 cm, z ± 0.2 cm on the z axis in the observed range. This estimate is based on the estimated alignment of the slider shaft (see section C.2.1). The absolute z-tolerance is somewhat better than for x and y for 3 reasons: 1) the mean z-errors in figure C.11 are within ± 0.1 cm; 2) the absolute calibration was observed to be within ± 0.1 cm at the origin (see C.2.1, Procedure to locate z = 0, step 4); and 3) the relative tolerance of the marks on the shaft is ± 0.1 cm.

C.2.3 Relative Position Tests

Figure C.12 presents a summary of the relative position PHANToM calibration tests. The *scal* data was taken in connection with the v2.201 probe-gate tests (note: the name "scal" stands for spatial calibration). In these tests the experimenter reset the PHANToM encoders at a reference point marked on the Virtual Workbench frame prior to each test. The scal test was performed either before or after the probe-gate test while the

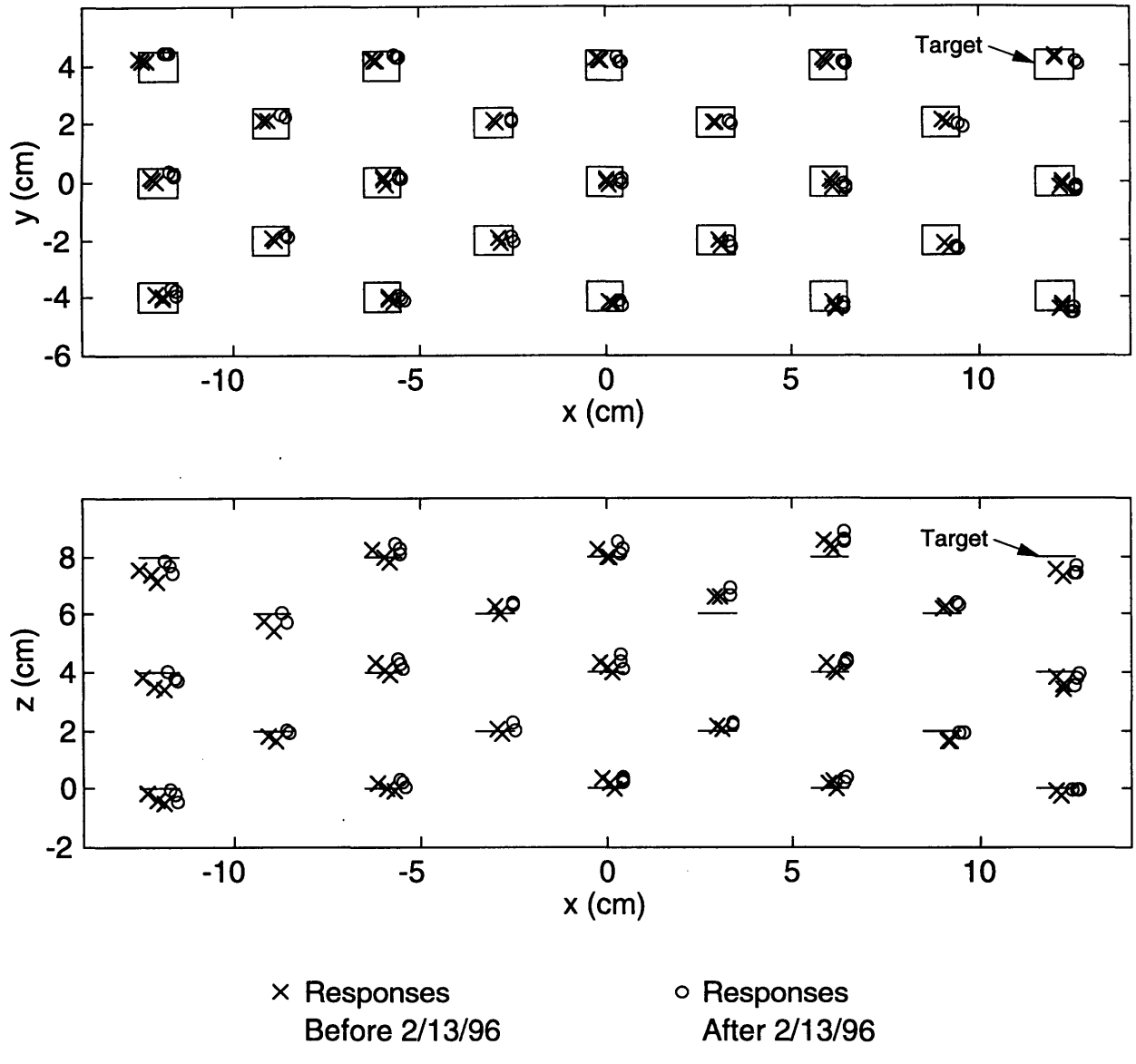


Figure C.12. Summary of v2.201 relative position tests of the PHANToM showing an apparent change in the calibration. The figure shows individual responses from the 3 scal tests immediately before and the 3 scal tests immediately after 2/13/96.

encoder calibration was unchanged. In the scal test the experimenter positioned an LED mounted on the endpoint of the PHANToM so that it appeared to be inside a virtual target presented on the display. The test involved one trial at each of 61 target positions (as shown in the figure) in a standard order. The procedure was similar to the 3D Calibration Experiment discussed in section 4.1.

The results are comparable to the results of the 3D Calibration Experiment except that it was observed that a significant shift occurred on or about February 13, 1996. Specifically, it can be seen in the figure that the x coordinates of the responses shift about 1/2 cm to the right. The cause appears to be that some work was done on the Virtual Workbench frame at that time that moved the PHANToM reference point. It is unlikely that this calibration shift had any effect on the results of the probe-gate tests.

APPENDIX D. DATA CORRECTIONS

Ten responses in the second identification experiment (see section 6.3) were reversed by the experimenter out of a total of 18,720 trials. Specifically, the original response that was recorded in the test document was changed from minus to plus (or vice versa) prior to the quantitative analysis. These modifications are shown in table D.1.

Table D.1. Modifications of data prior to analysis.

Test	Group	Trial	Change
T40S12N1	1	129	+4 → -4
"	2	59	+3 → -3
"	2	85	+6 → -6
"	5	79	+3 → -3
"	6	91	+4 → -4
T40S22N1	2	15	-6 → +6
T40S23N1	4	117	+6 → -6
T42S8N1	2	115	+6 → -6
T42S21N1	6	28	-6 → +6
T42S27N1	2	24	+6 → -6

These changes were made in cases where there is a strong probability that the subject moved the scale pointer up rather than down (or vice versa) by mistake. For example, in group 2 of test T40, subject S22 responded "-6" on trial 15 when the stimulus was +6 dnu. The response is found to be clearly out of place in comparison to the subject's other responses. This may be seen in table D.2 where all of the subject's responses for group 2 are presented, sorted into columns by stimulus. The original response in trial 15 is indicated by an asterisk.

Table D.2. T40S22N1 group 2 responses.

	Stimulus												
	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
	-6	-5	-4	-4	-4	-3	-1	-2	-1	0	3	1	4
R	-5	-5	-5	-5	-4	-2	1	-1	3	-3	4	4	4
e	-5	-5	-5	-5	-4	-2	0	-1	3	0	0	6	-6*
s	-5	-6	-5	-3	-4	-2	2	2	3	3	6	1	5
p	-5	-5	-4	-4	-4	-2	1	-2	2	1	2	4	3
o	-6	-5	-4	-5	-2	-3	-3	-1	2	4	1	5	5
n	-6	-5	-5	-3	-2	-1	0	3	0	1	4	2	6
s	-6	-5	-4	-4	-4	-2	0	0	0	3	4	4	5
e	-5	-5	-6	-4	-4	-2	0	0	6	0	2	3	5
	-6	-4	-5	-2	-1	-4	-4	0	3	3	4	6	4

Initially, there was no plan to edit any of the data, even though S12 reported making the 5 errors shown in table D.1 during his test. He was the only one who reported reversing any of the responses in the two tests, and he only did so during the groups where there is feedback, and only on one day. There are, of course, many reasons why he and the other subjects might not report reversing responses. For example, the subjects might not have been aware of the mistakes (particularly likely during no feedback trials). Also the experimenter did not emphasize the need to report errors. Thus it seems likely, particularly in light of the number of reports by S12 when he was aware of the problem, that there are many unreported reversal errors. Changing the few reported errors seemed unnecessary.

The changes were ultimately made, however, in connection with the analysis discussed in section 7.3. In that analysis, all of the data in a given group is eliminated when a response of -6 is given for a stimulus that is greater than, or equal to, a stimulus resulting in a response of +6 (see section 7.1.3.2). The analysis program flagged groups where this occurred and, when the data was subsequently inspected, clear response reversals, like the one shown in table D.2, were observed. In the course of correcting the obvious maximum/minimum reversals, the reported reversals (S12) were also inspected, found at the reported places, and changed.

In the case of S21 the reversal is not as clear-cut as in the other cases. In general, there is considerable variance in her responses and trial 28 of group 6 may simply be an extreme case. It was changed, however, because not to do so would have eliminated the entire group from the analysis, which would bias the combined results toward less variance.

APPENDIX E. K CORRECTION

This appendix describes the correction of the estimated values of the adaptation coefficient, k, that is used in figure 7.11 of section 7.3.1. The correction is based on the fact that sensations (Q) outside of the nominal bounds on the decision axis (C_{min} and C_{max}) are not weighted as heavily, in the estimate of k, as sensations in the middle of the response range. This idea may provide some insight into one mechanism of the central tendency of human observers in general, although the following analysis focuses exclusively on the problem of calculating the bias in the second identification experiment (see section 6.3 and chapter 7).

At the heart of the calculation is the determination of the expected value of k that will be observed (k_{obs}) based on the true values of k (which one hopes to discover) and σ. The correction for the initial estimate of k (k_g, per section 7.1.3.4) is then calculated by an iterative procedure, using the Matlab function "fminu."

In the procedure the square of the difference between k_{obs} and k_g is minimized in order to estimate the true value of k. The true value of σ_M is estimated by σ_{Mg} for the group and subgroup estimates (see sections 7.1.3.3 and 7.3), while equation 7.14 provides the estimate of σ_M in the 10-trial case. The latter is based on the value of G determined in section 7.4. Note that using the corrected values of k in the calculation of section 7.4, does not significantly alter the value of G.

Figure E.1 illustrates the key features of the calculation of k_{obs}. The calculation is simplified by noting that the response space, M, is symmetric about zero. It is also assumed that the probability that a sensation will fall outside the opposite boundary ($-C_{Mb}$) is negligible. The quantity k_{obs} is then calculated as the weighted sum of expected values of k, for estimates that include different numbers of N_i, using the following equations (E.1-E.6). Recall that the estimated value of k in the experiment (k_g) is calculated using only N_i that are less than all stimuli that resulted in a response of M = 6 (see sections 7.1.3.2 and 7.1.3.4).

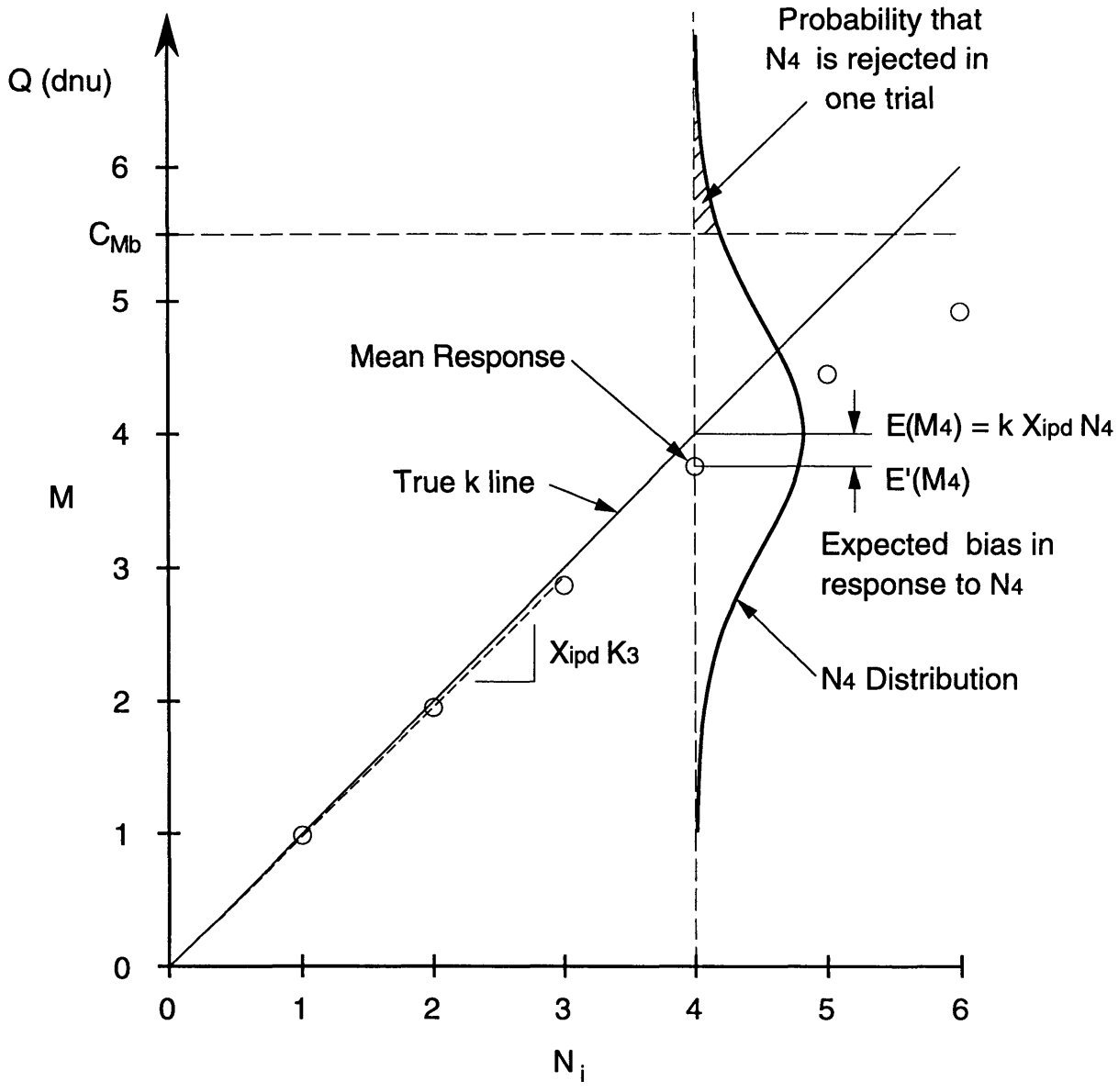


Figure E.1. Illustration of the k correction showing one distribution (N_4) and one estimated slope (K_3) of the calculation. Note M is presented on a continuous axis corresponding to Q (dnu).

$$k_{obs} = \sum_{i=1}^{N_{max}} p(max = i) K_i \quad \text{(E.1)}$$

$$p(max = j) = \left(1 - p_{j+1}\right) \prod_{i=1}^{j} p_i \quad \text{(E.2)}$$

$$p_i = \Phi\left(C_{Mb}, E(M_i), \sigma_M\right)^{T \frac{N_{max}}{2N_{max}+1}\left(1 - \frac{N_{max}-1}{N_{max}}\right)} \quad \text{(E.3)}$$

$$K_j = \frac{\sum_{i=1}^{j} N_i E'(M_i)}{X_{ipd} \sum_{i=1}^{j} N_i^2} \quad \text{(E.4)}$$

$$E'(M_i) = E(M_i) - \frac{\frac{\sigma_M}{\sqrt{2\pi}} e^{-(C_{Mb} - E(M_i))^2 / 2\sigma_M^2}}{\Phi\left(C_{Mb}, E(M_i), \sigma_M\right)} \quad \text{(E.5)}$$

$$E(M_i) = k X_{ipd} N_i \quad \text{(E.6)}$$

where the quantities in E.1-E.6 are:

- C_{Mb} = upper bound criterion on the decision axis (scaled to dnu) = 5.5 dnu
- dnu = depth number unit, defined such that N = P when $X_{ipd} = 1$
- $E(M_i)$ = expected value of M_i when $C_{Mb} = \infty$
- $E'(M_i)$ = expected value of M_i when the tail of the distribution above C_{Mb} is excluded.
- Φ = normal cumulative distribution function (see equation 7.15)
- k = adaptation coefficient (dnu/pixels)
- K_j = least-squares fit slope to $E'(M_i)$ for all N_i with $i \leq j$
- k_{obs} = expected value of the observed value of k
- M_i = i^{th} response depth (dnu)
- N_i = i^{th} stimulus depth (dnu)
- N_{min} = minimum stimulus depth (dnu)
- N_{max} = maximum stimulus depth (dnu)
- $p(max = i)$ = probability that the largest stimulus depth included in the estimate of k is N_i
- p_i = probability that N_i is not rejected (when it is presented) over the course of T trials
- σ_M = standard deviation of the distributions on the decision axis rescaled to depth number units (dnu)
- T = number of trials in the interval where k_g is estimated.
- X_{ipd} = ratio of effective IPD/actual IPD (pixels/dnu).

REFERENCES

Arditi, Aries. 1986. Binocular Vision. In *Handbook of Perception and Human Performance*, edited by K. R. Boff, L. Kaufman and J. P. Thomas. New York: Wiley-Interscience, John Wiley and Sons.

Barfield, Woodrow, and Thomas A. III Furness, eds. 1995. *Virtual Environments and Advanced Interface Design.* New York: Oxford Unversity Press, Inc.

Boff, Kenneth R., and Janet E. Lincoln, eds. 1988a. *Engineering Data Compendium: Human Perception and Performance.* 3 vols. Vol. 1. Dayton, OH: Harry G. Armstrong Aerospace Medical Research Laboratory Wright-Patterson Air Force Base.

Boff, Kenneth R., and Janet E. Lincoln, eds. 1988b. *Engineering Data Compendium: Human Perception and Performance.* 3 vols. Vol. 2. Dayton, OH: Harry G. Armstrong Aerospace Medical Research Laboratory Wright-Patterson Air Force Base.

Bos, Philip J. 1993. Liquid-Crystal Shutter Systems for Time-Multiplexed Stereoscopic Displays. In *Stereo Computer Grapics and Other True 3D Technologies*, edited by D. F. McAllister. New Jersey: Princeton University Press.

Braida, Louis D., and Nathaniel I. Durlach. 1972. Intensity Perception: II. Resolution in One-Interval Paradigms. *The Journal of the Acosutical Society of America* 51 (2 (Part 2)):483-502.

Braida, L. D., J. S. Lim, J. E. Berliner, N. I. Durlach, W. M. Rabinowitz, and S. R. Purks. 1984. Intensity Perception. XIII. Perceptual anchor model of context-coding. *The Journal of the Acosutical Society of America* 76 (3):722-731.

Bucholz, Richard D., Eric Robinson, and Lee McDurmont. 1995. Head mounted display for use during image guided surgery.

Buser, Pierre, and Michel Imbert. 1992. *Vision.* Translated by Kay, R. H.: The MIT Press.

Dautraix, Isabelle, and Isabelle Magnin. 1995. Stereoscopic computer graphics for ultrasonic medical data. Paper read at Stereoscopic Displays and Virtual Reality Systems II, 7-9 February 1995, at San Jose, California.

Davis, Elizabeth Thorpe, and Larry F. Hodges. 1995. Human Stereopsis, Fusion, and Sterescopic Virtual Environments. In *Virtual Environments and Advanced Interface Design*, edited by W. Barfield and T. A. Furness, III. New York: Oxford University Press.

Durlach, Nathaniel I., and Louis D. Braida. 1969. Intensity Perception. I. Preliminary Theory of Intensity Resolution. *The Journal of the Acoustical Society of America* 46 (2 (part 2)):372-383.

Durlach, Nathaniel I., and Anne S. Mavor, eds. 1995. *Virtual Reality: Scientific and Technological Challanges*. Washington, D.C.: National Academy Press.

Durlach, Nathaniel I., Thomas E. v. Wiegand, David Zeltzer, M. Srinivasan, K. Salisbury, Lee Beauregard, David Brock, Ben Sachtler, Jonathan Pfautz, David Schloerb, and Corrie Lathan. 1995. Virtual Environment Technology for Training (VETT): Annual Report for MIT Work Performed During Year 2. Naval Air Warfare Center, Training Systems Division, 12350 Research Parkway, Orlando, FL 32826: Masschusetts Institute of Technology, Research Laboratory of Electronics.

Eckel, Bruce. 1995. *Thinking in C++*. Englewood Cliffs, New Jersey: Prentice Hall.

Edgar, Graham K., and Peter J. Bex. 1995. Vision and displays. In *Simulated and Virtual Realities*, edited by K. Carr and R. England. London, England; Bristol, PA, USA: Taylor & Francis.

Ellis, Stephen R. 1991. Nature and Origins of Virtual Environments: A Bibliographical Essay. *Computing Systems in Engineering* 2 (4):321-347.

Ellis, Stephen R. 1995. The Relationship of Binocular Convergence and Errors in Judged Distance to Virtual Objects. Paper read at 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems, June 27-29, 1995, at Massachusetts Institute of Technology, Cambridge, MA, USA.

Ellis, Stephen R., and Urs J. Bucher. 1994. Distance Perception of Stereoscopically Presented Virtual Objects Optically Superimposed on Physical Objects by a Head-Mounted See-Through Display. Paper read at Human Factors and Ergonomics Society 38th Annual Meeting.

Epstein, William. 1968. Modification of the Disparity-Depth Relationship as a Result of Exposure to Conflicting Cues. *American Journal of Psychology* 81:189-197.

Fisher, S. Kay, and Kenneth J. Ciuffreda. 1990. Adaptation to optically-increased interocular separation under naturalistic viewing conditions. *Perception* 19:171-180.

Fisher, S. Kay, and Sheldon M. Ebenholtz. 1986. Does perceptual adaptation to telestereoscopically enhanced depth depend on the recalibration of binocular disparity? *Perception & Psychophysics* 40 (2):101-109.

Foley, James D., Andries van Dam, Steven K. Feiner, and John F. Hughes. 1990. *Computer Graphics: Principles and Practice*. 2nd ed. Reading, Massachusetts: Addison-Wesley Publishing Company.

Gescheider, George A. 1985. *Psychophysics: Method, Theory, and Application*. second ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Goldstein, E.Bruce. 1989. *Sensation and Perception*. 3rd ed. Belmont, CA: Wadsworth.

Graham, Clarence H. 1965. Visual Space Perception. In *Vision and Visual Perception*, edited by C. H. Graham. New York, NY: John Wiley & Sons, Inc.

Green, David M., and John A. Swets. 1988. *Signal Detection Theory and Psychophysics*. reprint ed. Los Altos, California: Peninsula Publishing.

Grinberg, Victor S., Gregg Podnar, and M. W. Siegel. 1994. Geometry of Binocular Imaging. Paper read at Stereoscopic Displays and Virtual Reality Systems, 8-10 February 1994, at San Jose, California.

Grinberg, Victor S., Gregg Podnar, and M. W. Siegel. 1995. Geometry of binocular imaging II: The augmented eye. Paper read at Stereoscopic Displays and Virtual Reality Systems II, 7-9 February 1995, at San Jose, California.

Held, Richard. 1972. Plasticity in Sensory-Motor Systems. In *Perception: Mechanisms and Models*, edited by R. Held and W. Richards. San Francisco: W. H. Freeman and Company.

Held, R., and N. Durlach. 1987. Telepresence, Time Delay, and Adaptation. Paper read at Spatial Displays and Spatial Instruments, August 31 - September 3, 1987, at Asilomar, CA.

Helmholtz, Hermann von. 1962. *Treatise on Physiological Optics*. Translated by Southall, James P. C. 3 vols. Vol. 3. New York: Dover Publications, Inc.

Helson, Harry. 1964. *Adaptation-Level Theory: An Experimental and System Approach to Behavior*. New York: Harper & Row.

Hodges, Larry F., and Elizabeth Thorpe Davis. 1993. Geometric Considerations for Stereoscopic Virtual Environments. *Presence: Teleoperators and Virtual Environments* 2 (1):34-43.

Hodges, Larry F., and David F. McAllister. 1993. Computing Stereoscopic Views. In *Stereo Computer Grapics and Other True 3D Technologies*, edited by D. F. McAllister. New Jersey: Princeton University Press.

Hsu, Jean, Zygmunt Pizlo, David M. Chelberg, Charles F. Babbs, and Edward J. Delp. 1996. Issues in the Design of Studies to Test the Effectiveness of Stereo Imaging. *IEEE Transactions on Systems, Man, and Cybernectics--Part A: Systems and Humans* 26 (6):810-819.

Judge, Stuart J., and C. Mary Bradford. 1988. Adaptation to telestereoscopic viewing measured by one-handed ball-catching performance. *Perception* 17:783-802.

Judge, Stuart J., and Frederick A. Miles. 1985. Changes in the coupling between accommodation and vergence eye movements induced in human subjects by altering the effective interocular seperation. *Perception* 14:617-629.

Kalawsky, Roy S. 1993. *The Science of Virtual Reality and Virtual Environments*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Kenyon, Robert V. 1981. The Effects of Enhanced Disparity on Manual Control Stereopsis and Tracking Performance: Man Vehicle Laboratory, Massachusetts Institute of Technology.

Kernighan, Brian W., and Dennis M. Ritchie. 1988. *The C Programming Language*. 2nd ed. Englewood Cliffs, New Jersey: Prentice Hall.

Lippmann, R. P., Louis D. Braida, and Nathaniel I. Durlach. 1976. Intensity Perception. V. Effect of payoff matrix on absolute identification. *The Journal of the Acosutical Society of America* 59 (1):129-134.

Lipscomb, James S., and Wayne L. Wooten. 1994. Reducing crosstalk between stereoscopic views, 8-10 February 1994.

Lipton, Lenny. 1993. Composition for Electrostereoscopic Displays. In *Stereo Computer Grapics and Other True 3D Technologies*, edited by D. F. McAllister. New Jersey: Princeton University Press.

Lipton, Lenny, Robert Akka, and Lhary Meyer. 1991. *The CrystalEyes Handbook*. San Rafael, CA: StereoGraphics Corporation, 2171 East Francisco Blvd., San Rafael, CA 94901.

Massie, Thomas H. 1993. Design of a Three Degree of Freedom Force-Reflecting Haptic Interface. S.B., Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Massie, Thomas H. 1996. Initial Haptic Explorations with the Phantom: Virtual Touch Through Point Interactions. S.M., Department of Mechanical Engineering, Massachusetts Institute of Technology.

Massie, Thomas H., and J. Kenneth Salisbury. 1994. The PHANToM Haptic Interface: A Device for Probing Virtual Objects. Paper read at ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, November, 1994, at Chicago.

MathWorks. 1992. MATLAB® Reference Guide. The MathWorks, Inc., 24 Prime Park Way, Natick MA 01760.

McKenna, Michael, and David Zeltzer. 1992. Three Dimensional Visual Display Systems for Virtual Environments. *Presence: Teleoperators and Virtual Environments* 1 (4):421-458.

Mon-Williams, Mark, and Eve Pascal. 1995. Virtual Reality Systems: Implications for Optometrists. *Optometry Today*, January 30, 1995, 30-33.

Mon-Williams, Mark, Simon Rushton, and John P. Wann. in press 1995. Investigation the reciprocal cross-links between accommodation and vergence: Implications for virtual reality displays. *Ophthalmic and Physiological Optics*.

Mon-Williams, Mark, John P. Wann, and Simon Rushton. 1993. Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics* 13 (October):387-391.

Nemire, Kenneth, and Stephen R. Ellis. 1993. Calibration and evaluation of virtual environment displays. Paper read at IEEE 1993 Symposium on

Research Frontiers in Virtual Reality, October 25-26, 1993, at San Jose, California.

Patterson, Robert, and Wayne L. Martin. 1992. Human Stereopsis. *Human Factors* 34 (6):669-692.

Patterson, Robert, Linda Moe, and Tiger Hewitt. 1992. Factors that Affect Depth Perception in Stereoscopic Displays. *Human Factors* 34 (6):655-667.

Pepper, Ross L., Robert E. Cole, and E. H. Spain. 1983. The influence of camera separation and head movement on perceptual performance under direct and TV-displayed conditions. *Proceedings of the society for information display* 24 (1):73-80.

Pepper, Ross L., David C. Smith, and Robert E. Cole. 1981. Stereo TV improves operator performance under degraded visability conditions. *Optical Engineering* 20 (4):579-585.

Pfautz, Jonathan D. 1996. Distortion of Depth Perception in a Virtual Environment Application. B.S., M.E., Massachusetts Institute of Technology.

Pynn, C.T., Louis D. Braida, and Nathaniel I. Durlach. 1972. Intensity Perception. III. Resolution in Small-Range Identification. *The Journal of the Acoustical Society of America* 51 (2 (part 2)):559-566.

Ritter, Manfred. 1977. Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics* 22 (4):400-407.

Ritter, Manfred. 1979. Perception of depth: Processing of simple position disparity as a function of viewing distance. *Perception & Psychophysics* 25 (3):209-214.

Robinett, Warren, and Richard Holloway. 1995. The Visual Display Transformation for Virtual Reality. *Presence: Teleoperators and Virtual Environments* 4 (1):1-23.

Robinett, Warren, and Jannick P. Rolland. 1992. A Computational Model for the Stereoscopic Optics of a Head-Mounted Display. *Presence: Teleoperators and Virtual Environments* 1 (1):45-62.

Rock, Irvin. 1966. *The Nature of Perceptual Adaptation.* New York: Basic Books, Inc.

Rolland, Jannick P., Dan Ariely, and William Gibson. 1995. Towards Quantifying Depth and Size Perception in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 4 (1):24-49.

Rolland, Jannick P., Todd Barlow, Frank A. Biocca, and Anantha Kancherla. 1995. Quantification of Adaptation to Virtual-Eye Location In See-Thru Head-Mounted Displays.

Rolland, J. P., R. L. Holloway, and H. Fuchs. 1994. A comparison of optical and video see-through head-mounted displays. Paper read at Telemanipulator and Telepresence Technologies, 31 October-1 November 1994, at Boston, Massachusetts.

Rosenberg, Louis B. 1993. The Effect of Interocular Distance upon Operator Performance using Stereoscopic Displays to Perform Virtual Depth Tasks. Paper read at IEEE Virtual Reality Annual International Symposium, September 18-22, 1993, at Seattle, Washington.

Rushton, Simon, Mark Mon-Williams, and John P. Wann. 1994. Binocular vision in a bi-ocular world: new generation head-mounted displays avoid causing visual deficit. *Displays* 15 (4):255-260.

Rushton, Simon, and John P. Wann. 1993. Problems in Perception and Action in Virtual Worlds. Paper read at Virtual Reality International 93: Proceedings of the third annual conference on Virtual Reality, April 1993, at London.

Satava, Richard M. 1994. Telemanipulation, telepresence and virtual reality for surgery in the year 2000. Paper read at Telemanipulator and Telepresence Technologies, 31 October-1 November 1994, at Boston, Massachusetts.

Schloerb, David W. 1995. A Quantitative Measure of Telepresence. *Presence: Teleoperators and Virtual Environments* 4 (1):64-80.

Schor, Clifton. 1987. Spatial Constraints of Stereopsis in Video Displays. Paper read at Spatial Displays and Spatial Instruments, August 31 - September 3, 1987, at Asilomar, CA.

Sekuler, Robert, and Randolph Blake. 1985. *Perception*. first ed, *Alfred A. Knopf Series in Psychology*. New York, NY: Alfred A. Knopf, Inc.

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

Shinn-Cunningham, B.G. 1994. Adaptation to Supernormal Auditory Localization Cues in an Auditory Virtual Environment. Ph.D., M.I.T.

Spain, E.H. 1986. Effects of Extended Camera Baseline and Image Magnification on Target Detection Time and Target Recognition with a Stereoscopic TY System: Naval Ocean System Center.

Surdick, R. Troy, Elizabeth T. Davis, Robert A. King, Gregory M. Corso, Alexander Shapiro, Larry Hodges, and Kelly Elliot. 1994. Relevant cues for the visual perception of depth: is where you see it where it is? Paper read at Human Factors and Ergonomics Society 38th Annual Meeting.

Swets, John A., Wilson P. Tanner, Jr., and Theodore G. Birdsall. 1964. Decision Processes in Percption. In *Signal Detection and Recognition by Human Observers*, edited by J. A. Swets. Los Altos, California: Peninsula Publishing.

Torgerson, Warren S. 1958. *Theory and Methods of Scaling*. New York: John Wiley and Sons, Inc.

Utsumi, Akira, Paul Milgram, Haruo Takemura, and Fumio Kishino. 1994a. Effects of fuzziness in perception of stereoscopically presented virtual object location. Paper read at Telemanipulator and Telepresence Technologies, 31 October-1 November 1994, at Boston, Massachusetts.

Utsumi, Akira, Paul Milgram, Haruo Takemura, and Fumio Kishino. 1994b. Investigation of Errors in Perception of Stereoscopically Presented Virtual Objects Locations in Real Display Space. Paper read at Human Factors and Ergonomics Society 38th Annual Meeting.

Vertut, J., and P Coiffet. 1986a. *Robot Technology, Volume 3A: Teleoperation and Robotics: Evolution and Development*: Prentice-Hall.

Vertut, J., and P Coiffet. 1986b. *Robot Technology, Volume 3B: Teleoperation and Robotics: Applications and Technology*: Prentice-Hall.

Wallach, Hans, and Eileen B. Karsh. 1963a. The modification of stereoscopic depth perception and the kinetic depth-effect. *American Journal of Psychology* 76:429-435.

Wallach, Hans, and Eileen B. Karsh. 1963b. Why the modification of stereoscopic depth-perception is so rapid. *American Journal of Psychology* 76:413-420.

Wallach, Hans, Mary E. Moore, and Linda Davidson. 1963. Modification of stereoscopic depth-perception. *American Journal of Psychology* 76:191-204.

Wann, John P., Simon Rushton, and Mark Mon-Williams. 1995. Natural Problems for Stereoscopic Depth Perception in Virtual Environments. *Vision Research* 35 (19):2731-2736.

Welch, Robert B. 1978. *Perceptual Modification: Adapting to Altered Sensory Environments*. Edited by E. C. Carterette and M. P. Friedman, *Series in Cognition and Perception*. New York, NY: Academic Press, Inc.

Welch, Robert B. 1986. Adaptation of Space Perception. In *Handbook of Perception and Human Performance*, edited by K. R. Boff, L. Kaufman and J. P. Thomas. New York: Wiley-Interscience, John Wiley and Sons.

Welch, Robert B., and David H. Warren. 1986. Intersensory Interactions. In *Handbook of Perception and Human Performance*, edited by K. R. Boff, L. Kaufman and J. P. Thomas. New York: Wiley-Interscience, John Wiley and Sons.

Winston, Patrick Henry. 1994. *On to C++*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Wonnacott, Thomas H., and Ronald J. Wonnacott. 1987. *Regression: A Second Course in Statistics*. Reprint Edition, original ©1981, by John Wiley & Sons, Inc. ed. Malabar, Florida: Krieger Publishing Company.

Yeh, Yei-Yu. 1993. Visual and Perceptual Issues in Stereoscopic Color Displays. In *Stereo Computer Grapics and Other True 3D Technologies*, edited by D. F. McAllister. New Jersey: Princeton University Press.

Zamorano, Lucia J., A. Majeed Kadi, Zhaowei Jaing, and Lutz Nolte. 1994a. Use of an infrared system for intraoperative digitization of laser and surgical instruments, at SPIE.

Zamorano, Lucia J., Jaing Zhaowei, Majeed Kadi, and Fernando Diaz. 1994b. Computer-assisted laser volumetric resection of intracranial lesions.

Zhai, Shumin, and Paul Milgram. 1994. Asymmetrical Spatial Accuracy in 3D Tracking. Paper read at Human Factors and Ergonomics Society 38th Annual Meeting.

3233-46