Summer 2007

The A-B Signal Detection Theory Model

Ernesto A. Bustamante
Old Dominion University

THE A-B SIGNAL DETECTION THEORY MODEL

by

Ernesto A. Bustamante
M.S. May 2005, Old Dominion University
B.S. August 2001, Old Dominion University

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirement for the Degree of

DOCTOR OF PHILOSOPHY

PSYCHOLOGY: HUMAN FACTORS

OLD DOMINION UNIVERSITY
August 2007

Approved by:

James P. Bliss (Director)

Terry L. Dickinson (Co-Director)

James M. Henson (Member)

Rao Chaganty (Member)

ABSTRACT

THE A-B SIGNAL DETECTION THEORY MODEL

Ernesto A. Bustamante
Old Dominion University, 2007
Director: Dr. James P. Bliss

The purpose of this research was threefold: 1) Present the *a-b* SDT model as an alternative framework to overcome the limitations of the underlying SDT model and the traditional measures of sensitivity and criterion setting, 2) Provide empirical support to validate the adequacy of the *a-b* SDT model, and 3) Conduct a Monte Carlo Study to compare and contrast the strengths and weaknesses of both the traditional and the *a-b* SDT models across the full spectrum of response values with the goal of providing researchers and practitioners with recommendations regarding the adequacy of each model. The results from this research have both theoretical implications and practical applications. The findings from the empirical study suggest that Green and Swets (1966)'s contention that the detection and response processes are independent from each other is questionable. Furthermore, the findings from the Monte Carlo Study suggest that the *a-b* SDT model provides more accurate measures to capture the dependency between these two processes. This is particularly important for researchers and practitioners who are interested in studying human-automation interaction factors and how sensory and perceptual factors may affect humans' response biases while interacting with automated systems.

I would like to dedicate this work to my brother, Alfredo E. Bustamante.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my brother, Alfredo E. Bustamante, for helping me overcome several struggles throughout my life and my career. Without him, I would have never made it this far.

Second, I would like to thank my advisor, Dr. James P. Bliss, for providing me with the professional and personal guidance, help, and support that I needed to overcome my hurdles throughout my graduate career.

Third, I would like to thank Dr. Terry L. Dickinson, who has challenged me throughout my entire career and has helped me become a better professional.

Fourth, I would like to acknowledge the contributions of Dr. James M. Henson and Dr. Rao Chaganty to the development of this dissertation.

Fourth, I would Like to thank Amy R Thompson for helping me conduct part of this research.

Fifth, I would like to thank my parents, who despite our differences, have always been there for me and have helped me in their own way throughout my life.

Last, but not least, I would like to thank my fiancé, Ashley K. Smithberger, for providing me with the moral support that I needed to complete my graduate training and for giving me a new purpose in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Although Signal Detection Theory (SDT) was first introduced into the field of psychology as a means for studying humans' abilities to detect sensory stimuli (Green & Swets, 1966), researchers have postulated its use for analyzing the performance of humans and automated systems in a variety of different areas (Swets, 1996). Stanislaw and Todorov (1999) suggested that SDT could be applied to any situation in which an observer had to make a decision under some degree of uncertainty. Swets (1973) advocated that SDT could be applied to areas other than psychophysics, such as the study of vigilance, recognition memory, attention, imagery, learning, conceptual judgment, personality, reaction time, manual control, speech recognition, and information retrieval systems.

Within the context of SDT, Green and Swets (1966) emphasized the existence of two different and separate processes. According to Green and Swets (1966), detection or recognition is the process of identifying whether the psychological experience was caused by just noise or the signal. Conversely, the decision process depends on the amount or extent of the psychological experience required by the detector to make an affirmative response. Swets (1973) distinguished between a process of covert discrimination and a process of overt response. He argued that these two processes have a complex relationship, influenced by a number of factors, including, but not limited to, probability, expectations, and motivation.

This dissertation adheres to the format of the *Human Factors Journal*.

The main contribution of SDT to the study of the performance of humans and automated systems is the capability of SDT to separate the discrimination process from the response process by distinguishing between independent measures of sensitivity and measures of criterion setting. However, the usefulness of the traditional SDT model is questionable in most applied settings because observers do not make decisions based on an underlying psychophysical continuum. For example, research suggests that humans use cognitive heuristics (Kahneman, Slovic, & Tversky, 1982) and rely on naturalistic decision making (Klein, Orasanu, Calderwood, & Zsambok, 1993) to make decisions in a nonlinear fashion.

SDT and Psychophysics

SDT was first used in the field of psychology to study human performance in psychophysical tasks (Green & Swets, 1966). One goal of traditional psychophysical methods was to determine the absolute threshold, which is defined as the minimum strength of a sensory stimulus that was necessary for humans to detect it. Another goal of psychophysical methods was psychophysical scaling, which consisted of mapping changes in the physical characteristics of stimuli to changes in humans' psychophysical experience. There were three commonly used methods in psychophysics (Green & Swets, 1966).

One traditional method was the *method of adjustment*. In this case, participants adjusted the magnitude of a physical stimulus until they considered it to be just noticeable. One problem with this method was that participants were not very accurate at adjusting the magnitude of the stimulus with the controls available to them (Green & Swets, 1966).

Another psychophysical method was the *method of serial exploration*, which was also commonly referred to as the *method of limits*. In this method, experimenters presented participants with a given stimulus and varied its magnitude in either an ascending or a descending order. Participants had to tell the experimenter when they could no longer detect the stimulus or when they could just detect it, depending on the presentation order. The main problem with this technique was that it allowed participants to build an expectation about the magnitude of the next stimulus, which affected their willingness to respond (Green & Swets, 1966).

The third psychophysical method was the *method of constant stimuli*. This method mitigated some of the limitations of the method of limits. Experimenters presented participants with different magnitudes of the stimulus in a random order, and participants had to respond when they detected the given stimulus. The random nature of the presentation order limited participants' response bias associated with their expectation.

All of these methods, however, were based on the assumption of an absolute sensory threshold (Green & Swets, 1966). According to this assumption, participants would only emit an affirmative response if the magnitude of the stimulus exceeded the threshold. Swets (1961) argued that the problem with psychophysical methods could be analyzed within the context of what he referred to as the fundamental decision problem. According to Swets (1961), when people perform a detection or decision-making task, they are limited to responding to a given interval or trial predefined by the experimental or applied condition. Therefore, for each trial, people's responses are not indicative of whether or not they detected the stimulus. Instead, people's responses are indicative of which response option (i.e., yes or no) they considered to be more appropriate for the given trial. Taking this into consideration, Green and Swets (1966) proposed the yes-no task as an alternative method for assessing humans' ability to detect stimuli.

In the yes-no task, experimenters provided a cue (in a different sensory modality) to indicate to participants when the stimulus was going to be presented. For example, if the task consisted of detecting an auditory stimulus, experimenters would flash a light to indicate to participants that the stimulus was being presented. Participants were instructed to respond either "yes" or "no", depending on whether or not they perceived the stimulus. In contrast to psychophysical tasks, the yes-no task presented participants with either white noise or the target stimulus on each trial. Therefore, each trial had two possible true states: noise or the signal plus noise. Similarly, each trial led to two possible responses: yes or no. Consequently, all possible outcomes could be examined using a two-by-two contingency table as shown in Table 1.

TABLE 1: Two-by-Two Contingency Table

	True States of the World	
	Signal + Noise	Noise
Response		
Yes	Hit	False Alarm
No	Miss	Correct Rejection

Green and Swets (1966) stressed the fact that when analyzing this contingency table, it is important to refer to the cells as conditional probabilities based on the two possible states of the world to allow comparisons to be made regardless of the prior probabilities of each event. Moreover, Green and Swets (1966) called attention to the fact that only two of these probabilities are necessary to assess overall performance because the remaining two are merely their complements. One conditional probability is that an affirmative response (i.e., yes) will be emitted given that the signal is present (i.e., signal-plus-noise). This is commonly known as the *hit rate* or p(HI). The second conditional probability is that an affirmative response (i.e., yes) will be emitted given that the stimulus is not present (i.e., noise). This is commonly referred to as the *false alarm rate* or p(FA).

Under the traditional SDT model, the two possible states of the world (i.e., noise and signal-plus-noise) have a differential effect on participants' psychophysical experiences, which could be represented by two overlapping probability density functions as shown in Figure 1. Sensitivity (*d'*) is defined as the mean difference between the means of the two probability density functions. Criterion setting (*c*) is defined as the point along the psychophysical continuum above which a participant will make an affirmative response.

Traditional psychophysical theories based on an absolute sensory threshold predicted that participants would make affirmative responses in noise trials only by chance. However, as Swets (1961) indicated, even in the early studies it became apparent that noise trials led to a significant proportion of affirmative responses. Also, Swets (1961) emphasized the notion that participants' response criteria were affected by a number of non-sensory variables such as the prior probability of signal trials and different payoffs associated with different responses. Swets (1996) argued that varying participants' response criteria allows researchers to obtain a number of hit and false alarm rate combinations. These combinations can be plotted in what is frequently known as the receiver operating characteristic (ROC) curve (see Figure 2).

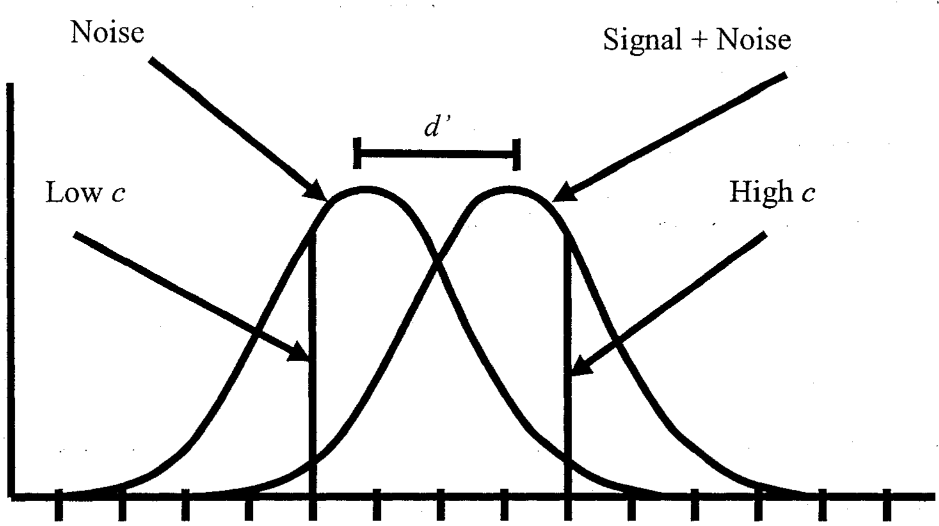


Figure 1. Noise and signal + noise distributions.

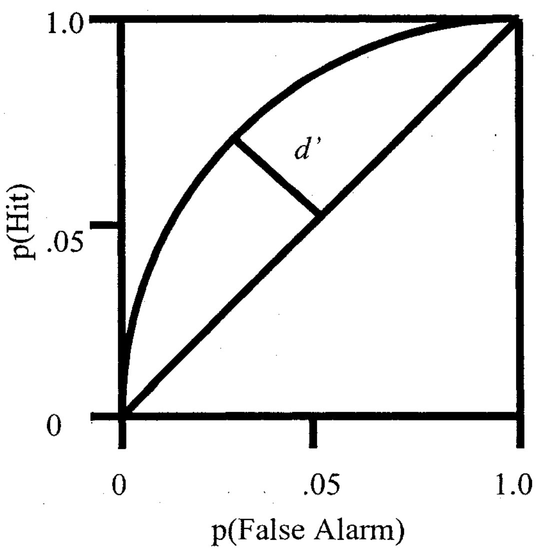


Figure 2. Receiver operating characteristic curve

Figure 2 shows a plot of hit rate values along the ordinate and false alarm rates along the abscissa. The straight line that cuts across the bottom left vertex and the top right vertex represents a sensitivity value at chance performance. The curved line represents a sensitivity value greater than chance performance. The line that connects the center of the straight line with the center of the curved line represents the exact sensitivity value.

Swets (1961) pointed out that traditional psychophysical methods, such as the method of adjustment, the method of limits, and the method of constant stimuli, were not able to differentiate between sensitivity and criterion setting. The main reason for this claim was that changes in performance could be attributed to sensitivity only if criterion setting was believed to be constant, whereas changes in performance could be attributed to changes in criterion setting only if sensitivity was assumed to remain constant. According to Green and Swets (1966), SDT provides the means for distinguishing sensitivity level from criterion setting.

Since Green and Swets published their seminal work in 1966, researchers have applied SDT to a variety of different domains that are well outside of the realm of psychophysics. Some of these areas include, but are not limited to, warning system performance (Bustamante, Bliss, & Anderson, in press; Lehto & Papastavrou, 1998; Parasuraman & Hancock, 1997), operator responses to alarms (Bustamante, 2005; Bustamante, Fallon, & Bliss, 2004; Meyer & Ballas, 1997; Sorkin & Woods, 1985), pilot weather judgments (Coyne, 2005), pilot terrain avoidance performance (Peterson, 1999), air combat training (Eubanks & Killeen, 1983), air traffic control (Bass, 2006), driver decision making performance (Wolf, Algom, & Lewin, 1988), decision making performance in supervisory control (Bisseret, 1981), group decision making (Sorkin, Hays, & West, 2001), automated speech recognition system performance (Deller, Desai,

& Yang, 2005), speech perception (Burlingame, Sussman, & Gillam, 2005), expert judgment performance (Harvey, 1992), luggage screening (Madhavan & Gonzales, 2006), and medical diagnosis (McFall & Treat, 1999; Mota & Schachar, 2000). The appropriateness of extending the traditional SDT model to areas that are outside of the realm of the psychophysical domain is questionable because the fundamental theoretical foundation of SDT may not serve as an adequate framework for conducting research in such areas.

Traditional SDT Model

Under the traditional SDT framework, researchers have postulated sensitivity measures as either the degree of separation between signal and signal-plus-noise probability density functions along a psychophysical continuum or the area underneath the ROC space (Donaldson, 1993). There are a number of SDT models and measures. Some examples of sensitivity measures include: d' and A_G (Green & Swets, 1966), A' (Pollack & Norman, 1964), E (McCornack, 1961), and Grier (1971)'s extension of the A' measure. Some examples of response bias or criterion setting measures include: c (Snodgrass & Corwin, 1988), β (Green & Swets, 1966), B'' (Hodos, 1970), and B_D'' (Donaldson, 1992). The most recently developed SDT model is Fuzzy SDT (Parasuraman, Masalonis, & Hancock, 2000). Fuzzy SDT provides measures of sensitivity (d') and response bias (β). Although these measures are conceptually similar to those proposed by Green and Swets (1966), they are fundamentally different because they are based on fuzzy logic and decision making. As such, fuzzy SDT is not appropriate for analyzing dichotomous decisions based on dichotomous states of the world (for applications of fuzzy SDT, see Hancock, Masalonis, & Parasuraman, 2000).

Despite the wide range of SDT measures, the most commonly used and widely accepted measures of sensitivity and criterion setting are *d'* and *c*, respectively. The sensitivity measure *d'* is defined as the difference between the mean of the signal-plus-noise distribution and the mean of the noise distribution (see Figure 1), or

$$d' = \Phi^{-1}[p(HI)] - \Phi^{-1}[p(FA)] \tag{1}$$

where,

d' = sensitivity

Φ^{-1}[p(HI)] = *z* score corresponding to the point below which the area under the standard normal distribution equals the proportion of hits

Φ^{-1}[p(FI)] = *z* score corresponding to the point below which the area under the standard normal distribution equals the proportion of false alarms

The criterion-setting measure *c* is defined as the point along the continuum above which an observer makes an affirmative response (see Figure 1), or

$$c = (-1) * \{.5 * \Phi^{-1}[p(HI)] + .5 * \Phi^{-1}[p(FA)]\} \tag{2}$$

where,

c = criterion setting

Φ^{-1}[p(HI)] = *z* score corresponding to the point below which the area under the standard normal distribution equals the proportion of hits

Φ^{-1}[p(FI)] = *z* score corresponding to the point below which the area under the standard normal distribution equals the proportion of false alarms

The main advantage of these measures is that they can be estimated from a single combination of proportions of hits and false alarms. However, research suggests that *d'* and *c* have questionable properties when based on extreme responding (Craig, 1979), which is often the case in real-world settings (Snodgrass & Corwin, 1988) such as traffic-collision warning systems (Parasuraman, Hancock, & Olofinboba, 1997), medical diagnosis (Li, Lin, & Chang, 2004), monitoring complex cockpit displays (Bailey & Scerbo, 2005), and luggage screening (Drury, Ghylin, & Holness, 2006). The reason for this is that the Φ^{-1} function is undefined for values of 0 and 1. Therefore, if an observer has a perfect hit rate of 1 or a false alarm rate of 0, the original hit and false alarm rates need to first be transformed. The problem is that transformations may lead to biased estimates (Hautus, 1995).

Limitations of the Traditional SDT Model

The applicability of the traditional SDT model is questionable for a variety of different domains. There are conceptual and practical reasons why extensions of traditional SDT are debatable (Long & Waag, 1981). First, the underlying SDT model rests on the assumption of the existence of the two probability density functions associated with signal and signal-plus-noise trials along a continuum (see Figure 1). Swets (1961) argued that the exact nature of the sensory excitation produced by either the noise or the stimulus was not an issue. According to Swets (1961), what matters is that sensory excitation varies from trial to trial even if the magnitude of the stimulus is held constant, and that excitation can be quantified in terms of a single continuous variable, which could be thought of as the decision variable. However, in most applied settings, this argument is questionable.

Second, one criterion for assessing the adequacy of measures is the capability to assign scores even when observers do not commit any errors (Craig, 1979). However, as previously mentioned, there are limitations related to traditional SDT measures in the presence of extreme responding (Long & Waag, 1981).

Goals of this Research

The purpose of this research was fourfold: 1) Present the *a-b* SDT model as an alternative framework to overcome the limitations of the underlying SDT model and the traditional measures of sensitivity and criterion setting, 2) Provide empirical support to validate the adequacy of the *a-b* SDT model, 3) Conduct a jackknifing study based on the data obtained from the empirical study to determine if differences between the *a-b* SDT model and the traditional SDT model were due to sampling error or some systematic variation, and 4) Conduct a Monte Carlo Study to compare and contrast the strengths and weaknesses of both the traditional and the *a-b* SDT models across the full spectrum of response values with the goal of providing researchers and practitioners with recommendations regarding the adequacy of each model.

CHAPTER II

THE A-B SDT MODEL

The *a-b* SDT model is based on the work of Bustamante, Fallon, and Bliss (2006), who offered alternative measures of sensitivity or accuracy (*a*) and response bias (*b*) that do not rely on the underlying assumptions of the traditional SDT model. Instead, *a* and *b* are based simply on the outcome matrix (see Table 1), defined by the proportion of hits and false alarms. It is important to note that the *a-b* SDT model is not atheoretical, and it shares many similarities about the detection and response processes as the traditional SDT model. Within the *a-b* SDT model, sensitivity or accuracy is conceptually defined as the tendency to make correct responses (i.e., hits and correct rejections). Response bias, on the other hand, is conceptually defined as the tendency to make affirmative responses (hits and false alarms).

In most applied settings, researchers are concerned with the ability of humans and automated systems to make accurate decisions. Therefore, Bustamante et al. (2006) first replaced the term "sensitivity" with "accuracy" (*a*) and defined it as the weighted sum of the proportion of correct affirmative and negative responses, or

$$a = .5 * p(HI) + .5 * p(CR) \qquad (3)$$

where,

a = accuracy

p(HI) = proportion of hits

p(CR) = proportion of correct rejections

Bustamante et al. (2006) defined response bias (*b*) as the weighted sum of the proportion of correct and incorrect affirmative responses, or

$$b = .5 * p(HI) + .5 * p(FA) \tag{4}$$

where,

b = response bias

p(HI)= proportion of hits

p(FA) = proportion of false alarms

Advantages of the a-b SDT Model over the Traditional SDT Model

The *a-b* SDT model has several advantages over the traditional SDT model. Comparing Formulas 1 and 2 with Formulas 3 and 4, it is evident that the *a-b* SDT model is more parsimonious than the traditional SDT model. There are three main reasons for this. First, Bustamante et al. (2006) made no reference to an underlying decision continuum. Swets (1961) argued that the exact nature of the sensory excitation produced by either the noise or the stimulus could be quantified in terms of a single continuous variable (the decision variable). However, this argument does not apply well to domains where individuals and automated systems make decisions based on multiple sources of information and different decision-making algorithms.

This lack of reliance on the assumption of an underlying decision continuum is one of the strengths of the *a-b* SDT model because in most applied settings, humans and automated systems do not make decisions based on a single underlying continuum. Researchers have suggested that many factors may influence human decision making in a nonlinear fashion. Examples include the perception of risk (Ayres, Wood, Schmidt, & McCarthy, 1998), the amount of effort involved in choosing a particular alternative

(Wogalter, Allison, & McKena, 1989), the use of cognitive heuristics (Kahneman et al., 1982), workload (Broadbent, 1978), fatigue (Krueger, 1989), and expertise (Klein et al., 1993). With regard to automated system decision making, designers typically use highly complex algorithms that are also nonlinear, such as decision trees, Monte Carlo Studys, and neural networks.

A second advantage of the *a-b* SDT model is that it does not require transformations of original hit and false alarm rates for extreme responses. Because the *a-b* SDT model is not based on an underlying continuum, there is no need to assume the existence of probability density functions associated with the different signal and signal-plus-noise trials. The *a-b* SDT model simply describes the decision outcome matrix shown in Table 1 using measures of accuracy and response bias that are uncorrelated with each other. In contrast, traditional measures of performance, such as hit rate, false alarm rate, overall percentage of correct decisions, and the ratio of hit rate to false alarm rate, are all inadequate measures of accuracy and response bias because they are correlated with factors that affect both the detection and the response processes (Stanislaw & Todorov, 1999).

A third advantage of the *a-b* SDT model is that the alternative *a* and *b* measures may be interpreted more intuitively. With regard to *a*, a score of 0 indicates the complete lack of ability to make accurate decisions. A score of .5 indicates performance at chance level, and a score of 1 indicates perfect decision-making accuracy. With regard to *b*, a score of 0 indicates a lack of affirmative responsiveness. A score of .5 indicates an unbiased level of responsiveness, and a score of 1 indicates a complete response bias toward affirmative responses. These metrics may be more appealing to human factors researchers as well as system designers and decision makers responsible for implementing human factors research findings because of their intuitive interpretative nature.

CHAPTER III

EMPIRICAL VALIDATION OF THE A-B SDT MODEL

Prior research has shown evidence to support the advantages of the *a-b* SDT model over the traditional SDT model (Bustamante, Anderson, Thompson, Bliss, & Scerbo, in press; Bustamante et al. 2006; Bustamante, Spain, Newlin, & Bliss, 2007). However, most of this research has relied on the use of Monte Carlo Studys. Consequently, there is a need to perform an empirical evaluation of the *a-b* SDT model to complement the previously conducted simulation-based research.

Consistent with the second goal of this research, this study aimed to provide an empirical validation of the *a-b* SDT model. The goal of this study was to gather empirical data from a traditional signal detection study to assess the adequacy of the *a-b* SDT model. The basic premise of SDT is that in a traditional SDT task, two distinct and independent processes take place: a covert discrimination process and an overt response process (Swets, 1973). Furthermore, according to the traditional SDT model, these two processes are independent of each other and are affected by different factors (Green & Swets, 1966). Therefore, to test the adequacy of the *a-b* SDT model, two factors that should affect each of these processes independently were manipulated within a traditional SDT task.

One of these factors was the probability of occurrence of the target stimulus. Prior research suggests that changes in the probability of occurrence of the target stimulus affects people's response bias (Bliss, Gilson, & Deaton, 1995). This effect is commonly known as probability matching. The second factor was based on a derivation of Weber's

Law, which is used to predict people's abilities to detect just noticeable differences between two stimuli. According to Weber's Law, the ratio of the difference between a target stimulus and a baseline stimulus equals a constant *K*, or

$$K = \frac{\Delta I}{I} \tag{5}$$

where,

K = Weber's fraction

ΔI = Difference between the baseline and target stimuli

I = Intensity of baseline stimulus

Based on Weber's Law, it follows that as the difference between the baseline and target stimuli increases, people's abilities to discriminate between the two stimuli should also increase. Within the context of SDT, this implies that increasing the difference between the baseline and target stimuli should increase people's accuracy levels.

Hypotheses

The empirical validation of the adequacy of the *a-b* SDT model rests on two hypotheses: 1) Increasing the probability of occurrence of the target stimulus should increase participants' response bias without affecting their accuracy; 2) Increasing the difference between the baseline and target stimuli should increase participants' accuracy without affecting their response bias.

METHOD

Experimental Design

A 3 x 3 repeated-measures design was used for this study. The probability of occurrence of the target stimulus was manipulated at three levels (.10, .50, and .90). The frequency difference between the baseline and target stimuli was also manipulated at three levels (5 Hz, 10 Hz, and 15 Hz).

Participants

A power analysis revealed that approximately 20 participants would be necessary to obtain statistically significant effects at a .01 alpha level, assuming a power of .80 and a medium effect size for each factor (Cohen, 1988). Therefore, 20 (10 females, 10 males) undergraduate and graduate students from Old Dominion University in Norfolk, VA participated in this study. Participants ranged from 18 to 34 years of age ($M = 23.05$, $SD = 3.69$). They all had normal or corrected-to-normal vision and hearing. Participants were compensated with one research credit as a form of incentive to participate in this study. In addition to this, a prize of $100 was awarded to the participant with the best performance to motivate participants to perform at their maximum level.

Materials and Apparatus

This study took place in a sound-attenuated room with an average ambient noise level of 45 dB(A). Participants performed a traditional yes-no SDT task, which consisted of discriminating between baseline and target auditory stimuli that varied in their fundamental frequency. All stimuli were generated using the NCH Tone Generator software and lasted 100 milliseconds. The baseline stimulus consisted of a simple sine wave of 500 Hz. Depending on the experimental condition, the target stimulus consisted of a simple sine wave of 505 Hz, 510 Hz, or 515 Hz. Stimuli were presented to participants through a set of sound-attenuated stereo headphones at 55 dB(A) using a fixed inter-stimulus ratio of 2.5 seconds. A Microsoft Visual Basic program was developed and loaded on a Dell Inspiron 600m laptop computer to 1) Collect participants' demographic and contact information (see Appendix A), 2) Present participants with the instructions of the study (see Appendix B), 3) Familiarize participants with the baseline and target stimuli prior to each session (see Appendix C), 4) Present participants with information regarding the type of session (i.e., practice or experimental), the probability of occurrence of the target stimulus (i.e., .10, .50, or .90), and the frequency difference between the baseline and target stimuli (i.e., 5 Hz, 10 Hz, or 15 Hz; see Appendix D), 5) Present participants with the baseline and target stimuli throughout each practice and experimental session, 6) Provide participants with feedback regarding their performance by updating their performance score (see Appendix E), and 7) Record their responses.

Procedure

Participants came to the laboratory individually. First, the experimenter greeted them and provided them with the informed consent form (see Appendix F). Second, the experimenter asked participants to silence or turn off their cellular phones if they had one. Third, the experimenter assigned each participant an identification number. Fourth, the experimenter asked participants if they had any questions regarding the nature of the study. If participants decided to participate, the experimenter asked them to sign and date the informed consent form. Fifth, the experimenter asked participants to complete the background and contact information form. Sixth, the experimenter showed participants the instructions for completing the study and asked them to read them carefully. Seventh, the experimenter instructed participants to place the set of stereo headphones on their heads and adjust them to fit comfortably. Eighth, the experimenter showed participants the familiarization screen and instructed them about how to use the graphical user interface of the program. Ninth, the experimenter showed participants how to navigate through the program to the first session and explained all the information displayed on the screen. Last, the experimenter answered any final questions participants had regarding the completion of the study.

As part of this experiment, participants performed nine one-minute practice sessions and nine five-minute experimental sessions, which varied according to the probability of occurrence of the target stimulus (i.e., .10, .50, .90) and the frequency difference between the baseline and target stimuli (i.e., 5 Hz, 10 Hz, 15 Hz). Each practice session preceded its corresponding experimental session. Practice sessions consisted of 20 trials, whereas experimental sessions consisted of 100 trials. All sessions were fully counterbalanced according to the ascending or descending nature of each factor (i.e., the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli). Furthermore, to avoid a potential vigilance decrement, the experimenter instructed participants to take a short break after each experimental session.

Participants' task consisted of pressing the number one key on top of the keyboard for trials in which they perceived that the target stimulus was presented and pressing the number zero key on top of the keyboard for trials in which they perceived that the target stimulus was not presented. To maintain experimental control, the experimenter asked participants to place and keep their left middle finger on top of the number one key and their right middle finger on top of the number zero key during each session.

Throughout each trial, participants received feedback about the accuracy of their responses through the changes in their performance score. Participants started each session with a score of zero. For each individual trial, if participants made an accurate response (i.e., a hit or a correct rejection), they received one point, which was added to their total performance score. Similarly, for each incorrect response (i.e., a false alarm or a miss), they lost one point, which was subtracted from their total performance score.

RESULTS

Given that the purpose of this study was to provide empirical support for the adequacy of the *a-b* SDT model, four 3 x 3 repeated-measures ANOVAs were conducted to assess the effects of the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli on the *a* and *b* measures as well as on the traditional *d'* and *c* measures. Due to the large number of statistical tests, statistical significance for all inferential tests was set *a priori* at $p < .01$. Similarly, only statistically significant results are reported.

The a-b SDT Model

Accuracy. A 3 x 3 repeated measures ANOVA showed a statistically significant effect of frequency difference on participants' accuracy levels (*a*), $F(2, 38) = 44.23$, $p < .01$, partial $\eta^2 = .70$. A follow-up trend analysis showed a statistically significant linear trend, $F(1, 19) = 58.47$, $p < .01$, partial $\eta^2 = .76$, and a statistically significant quadratic trend, $F(1, 19) = 16.14$, $p < .01$, partial $\eta^2 = .46$. Table 2 shows the means and standard deviations of participants' accuracy levels in each of the three frequency difference conditions.

Table 2: Means and Standard Deviations of Participants' Accuracy Levels across Different Frequency Difference Conditions

Frequency Difference	M	SD
5 Hz	.71	.17
10 Hz	.88	.16
15 Hz	.91	.13

These results are also graphically depicted in Figure 3.

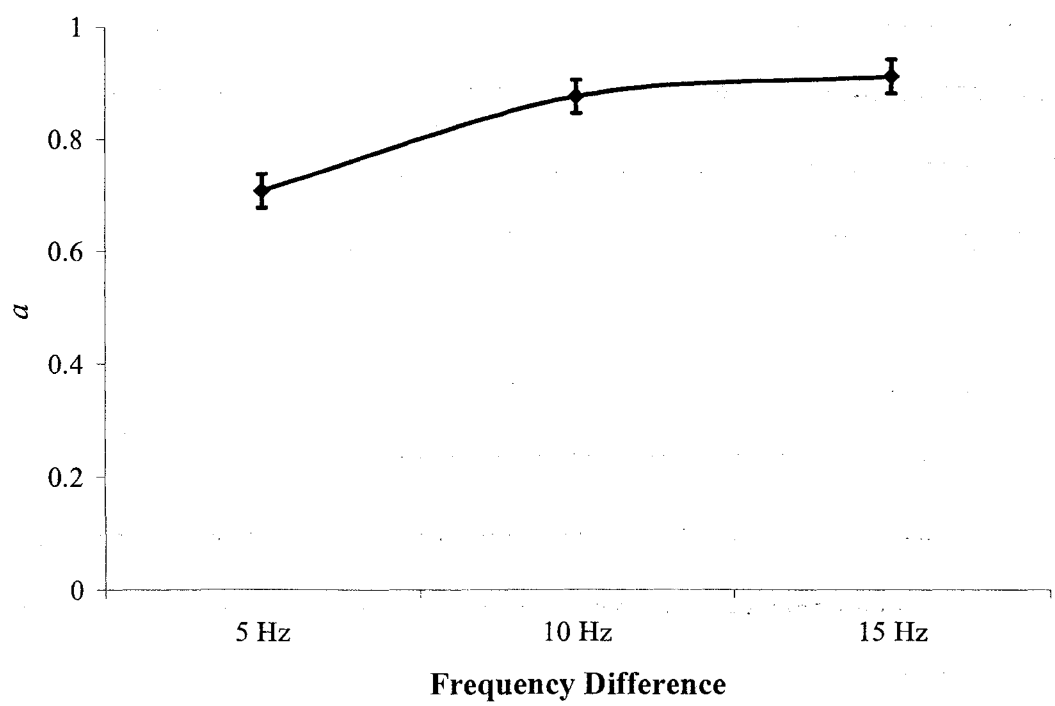


Figure 3. Participants' accuracy levels as a function of the frequency difference between the baseline and target stimuli.

Response Bias. A 3 x 3 repeated measures ANOVA showed a statistically significant effect of the probability of occurrence of the target stimulus on participants' response biases (*b*). However, because Mauchly's test indicated that the sphericity assumption was violated, the Greenhouse-Geisser correction was used to adjust the degrees of freedom, $F(1.07, 20.29) = 24.45, p < .01$, partial $\eta^2 = .56$. Results also showed a statistically significant interaction effect between the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli on participants' response biases (*b*), $F(2.01, 38.20) = 20.38, p < .01$, partial $\eta^2 = .52$.

To examine the nature of this interaction effect, three simple-effects follow-up analyses and Scheffe contrasts were conducted. The purpose of these analyses was to assess the strength of the effect of the probability of occurrence of the target stimulus at each level of the frequency difference factor. Table 3 shows the results of the simple-effects follow-up analyses.

TABLE 3: Simple-Effects Follow-Up Analyses for Participants' Response Biases

Frequency Difference	df	SS	MS	F
5 Hz	1.18	2.33	1.98	39.23*
10 Hz	1.10	0.40	0.36	9.94*
15 Hz	1.14	0.22	.19	8.29*

$^{*}p < .01$

Table 4 shows the means and standard deviations of participants' response biases in each of the three probability of occurrence of the target stimulus conditions across all three frequency difference conditions.

TABLE 4: Means and Standard Deviations of Participants' Response Biases

	Probability of Occurrence of Target Stimulus					
	.10		.50		.90	
Frequency Difference	M	SD	M	SD	M	SD
5 Hz	$.27_a$	.19	$.46_b$	.08	$.75_c$	.17
10 Hz	$.41_b$	.15	$.50_b$	.05	$.61_d$	.16
15 Hz	$.41_b$	.14	$.50_b$	.04	$.55_d$	.11

NOTE: Means with different subscripts are significantly different from each other at $p <$ .01 based on Scheffe contrasts.

These results are also graphically depicted in Figure 4.

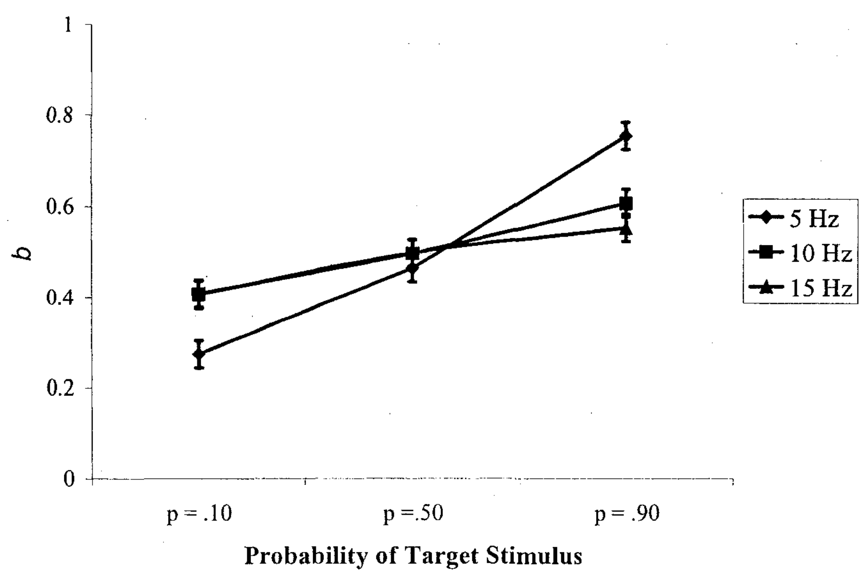


Figure 4. Participants' response biases as a function of the interaction between the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli.

As Figure 4 shows, the effect of the probability of occurrence of the target stimulus on participants' response biases decreased as the frequency difference between the baseline and target stimuli increased.

Traditional SDT Model

Because of the previously discussed limitation of the traditional SDT model for estimating sensitivity and criterion setting measures in the presence of extreme responses, the observed hit and false alarm rates were first transformed using the log-linear transformation, as recommended by Snodgrass and Corwin (1988). Subsequently, d' and c were calculated based on the transformed hit and false alarm rates.

Sensitivity. A 3 x 3 repeated measures ANOVA showed a statistically significant effect of frequency difference on participants' sensitivity levels (d'), $F(2, 38) = 57.01$, $p < .01$, partial $\eta^2 = .75$. A follow-up trend analysis showed a statistically significant linear trend, $F(1, 19) = 85.36$, $p < .01$, partial $\eta^2 = .81$, and a statistically significant quadratic trend, $F(1, 19) = 16.33$, $p < .01$, partial $\eta^2 = .46$. Table 5 shows the means and standard deviations of participants' sensitivity levels in each of the three frequency difference conditions.

TABLE 5: Means and Standard Deviations of Participants' Sensitivity Levels across Different Frequency Difference Conditions

Frequency Difference	M	SD
5 Hz	1.56	1.24
10 Hz	2.92	1.33
15 Hz	3.25	1.14

These results are also graphically depicted in Figure 5.

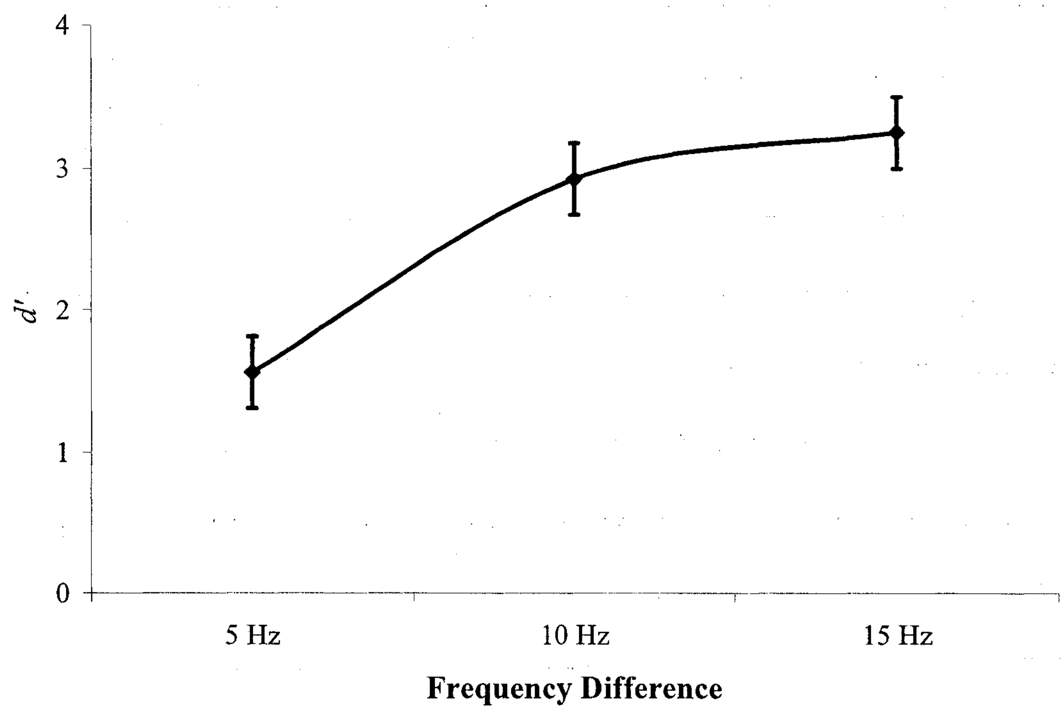


Figure 5. Participants' sensitivity levels as a function of the frequency difference between the baseline and target stimuli.

Criterion Setting. A 3 x 3 repeated measures ANOVA showed a statistically significant effect of the probability of occurrence of the target stimulus on participants' criterion settings (*c*). However, because Mauchly's test indicated that the sphericity assumption was violated, the Greenhouse-Geisser correction was used to adjust the degrees of freedom, $F(1.11, 21.10) = 63.79, p < .01$, partial $\eta^2 = .77$. Results also showed a statistically significant interaction effect between the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli on participants' criterion settings (*c*), $F(2.24, 42.56) = 9.73, p < .01$, partial $\eta^2 = .34$.

To examine the nature of this interaction effect, three simple-effects follow-up analyses and Scheffe contrasts were conducted. The purpose of these analyses was to assess the strength of the effect of the probability of occurrence of the target stimulus at each level of the frequency difference factor. Table 6 shows the results of the simple-effects follow-up analyses.

TABLE 6: Simple-Effects Follow-Up Analyses for Participants' Criterion Settings

Frequency Difference	df	SS	MS	F
5 Hz	1.22	35.10	28.72	49.07*
10 Hz	1.19	13.93	11.71	36.90*
15 Hz	1.28	11.29	8.83	53.24*

*$p < .01$

Table 7 shows the means and standard deviations of participants' criterion setting in each of the three probability of occurrence of the target stimulus conditions across all three frequency difference conditions.

TABLE 7: Means and Standard Deviations of Participants' Criterion Setting

	Probability of Occurrence of Target Stimulus					
	.10		.50		.90	
Frequency Difference	M	SD	M	SD	M	SD
5 Hz	$.87_a$	.64	$.11_b$	.29	$-.99_c$	.64
10 Hz	$.62_d$	.45	$.01_b$	.17	$-.57_e$	.47
15 Hz	$.62_d$	.38	$.05_b$	.18	$-.44_e$	.37

NOTE: Means with different subscripts are significantly different from each other at $p < .01$ based on the Scheffe test.

These results are also graphically depicted in Figure 6.

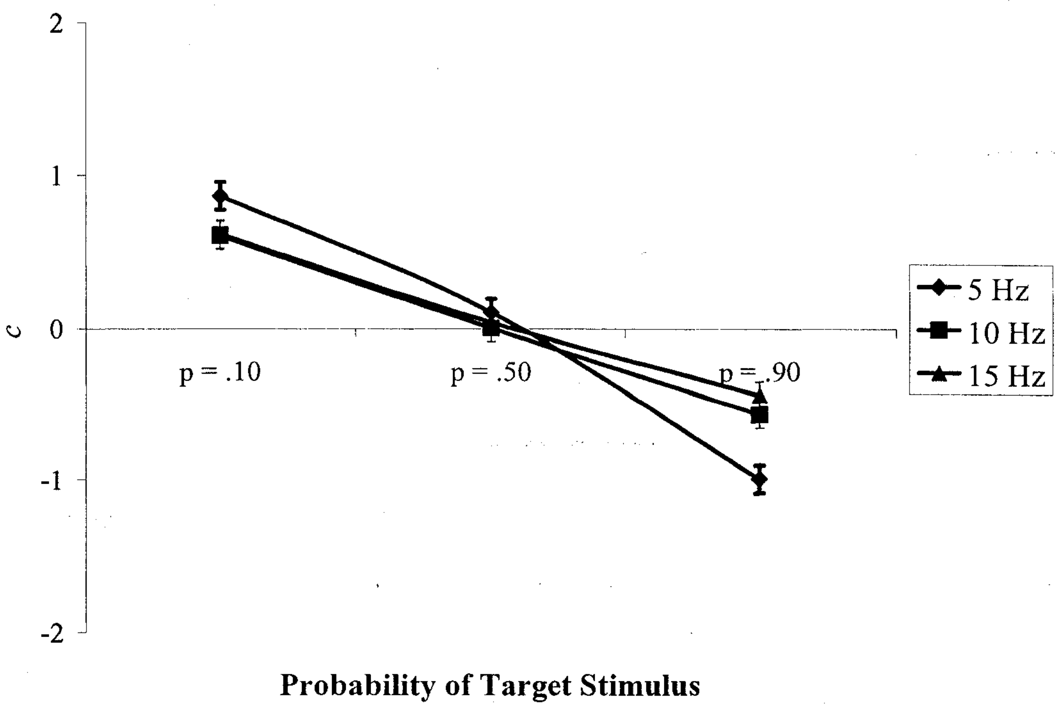


Figure 6. Participants' criterion settings as a function of the interaction between the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli.

As Figure 6 shows, the effect of the probability of occurrence of the target stimulus on participants' criterion settings decreased as the frequency difference between the baseline and target stimuli increased.

Correlative Comparisons Between and Within Models

A correlation analysis of the *a*, *b*, *d'*, and *c* measures revealed that both models provided statistically similar measures of accuracy or sensitivity and response bias or criterion setting respectively. Also, results from this analysis showed that both models provided uncorrelated measures of accuracy and response bias or sensitivity and criterion setting respectively. Table 8 shows the pattern of correlations of the measures between models as well as within each model.

TABLE 8: Correlations among Measures

Variable	1	2	3	4
1. Accuracy (*a*)	--			
2. Response Bias (*b*)	.04	--		
3. Sensitivity (*d'*)	.95*	.05	--	
4. Criterion Setting (*c*)	.01	-.94*	-.01	--

* $p < .01$

DISCUSSION

Results provided partial empirical support for the validity of the *a-b* SDT model. As expected, participants' accuracy levels increased as the frequency difference between the baseline and target stimuli increased. Furthermore, the probability of occurrence of the target stimulus did not affect participants' accuracy levels. Additionally, results showed a similar pattern for the traditional SDT measure of sensitivity (*d'*). Also, as predicted, results showed that participants' response biases increased as the probability of the target stimulus increased. Moreover, results showed a similar, yet reversed, pattern for the traditional SDT measure of criterion setting (*c*). It is important to note that the reason why the pattern was reversed was simply due to the fundamental conceptual and mathematical definitions of response bias and criterion setting. Greater values of response bias are indicative of people's tendency to make affirmative responses, whereas lower values of criterion setting are indicative of people's tendency to make affirmative responses.

Additionally, consistent with prior simulation research (Bustamante et al. 2006), the pattern of correlations among the *a*, *b*, *d'*, and *c* measures showed two important properties of the *a-b* SDT model. First, the high relationship between *a* and *d'* and *b* and *c* respectively showed empirical support for the construct validity of the *a-b* SDT model. Second, the absence of significant correlations between *a* and *b* showed partial support for the notion that the *a-b* SDT model provides independent measures of accuracy and response bias.

Nevertheless, the data from this study showed unexpected results, which bring into question the adequacy of both models. The fact that there was an interaction effect between the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli on both response bias and criterion setting suggests that neither model provides independent measures of accuracy or sensitivity and response bias or criterion setting. As Figures 4 and 6 show, as the frequency difference between the baseline and target stimuli increased, and, consequently, participants' accuracy or sensitivity levels increased, the effect of the probability of occurrence of the target stimulus on participants' response biases or criterion settings decreased. Furthermore, results showed a stronger interaction effect for response bias than criterion setting, suggesting that the nature of the dependency between the measures within each model may vary across models.

However, the differences in the effect size of the interaction effect could have been due to sampling error or some systematic difference between the two models.

Therefore, a jackknifing study was conducted based on the data obtained from the empirical study to determine if differences between the *a-b* SDT model and the traditional SDT model were due to sampling error or some systematic variation that will require future research.

CHAPTER IV

JACKKNIFING STUDY

The difference in the effect size of the interaction between the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli on response bias and criterion setting between the *a-b* SDT model and the traditional SDT model raised an important issue of concern that the empirical study alone could not address. The purpose of this jackknifing study was to determine if the difference in the effect size of the interaction on response bias and criterion setting was due to sampling error or some systematic variation.

METHOD

Given that the empirical study had a sample size of 20 participants, this jackknifing study consisted of 20 iterations based on a sample size of 19 participants for each iteration. A 3 x 3 repeated-measures ANOVA was conducted for each iteration, and the effect size of the interaction was recorded. The independent variables were the frequency difference between the baseline and target stimuli (5 Hz, 10 Hz, 15 Hz) and the probability of occurrence of the target stimulus (.10, .50, .90).

RESULTS

Statistical analyses consisted of calculating the mean and standard error of the effect size of the interaction. A dependent-samples t-test was used to determine if the difference between models was statistically significant. To maintain consistency, statistical significance was set *a priori* at an alpha level of $p < .01$.

Results showed a statistically significant difference in the interaction effect between the two models, $t(19) = 66.09, p < .01$. The interaction effect between the probability of occurrence of the target stimulus and the frequency difference between the baseline and target stimuli was significantly greater for the *a-b* SDT model ($M = .52$, $SD = .02$) than the traditional SDT model ($M = .34$, $SD = .02$). These results are graphically depicted in Figure 7.

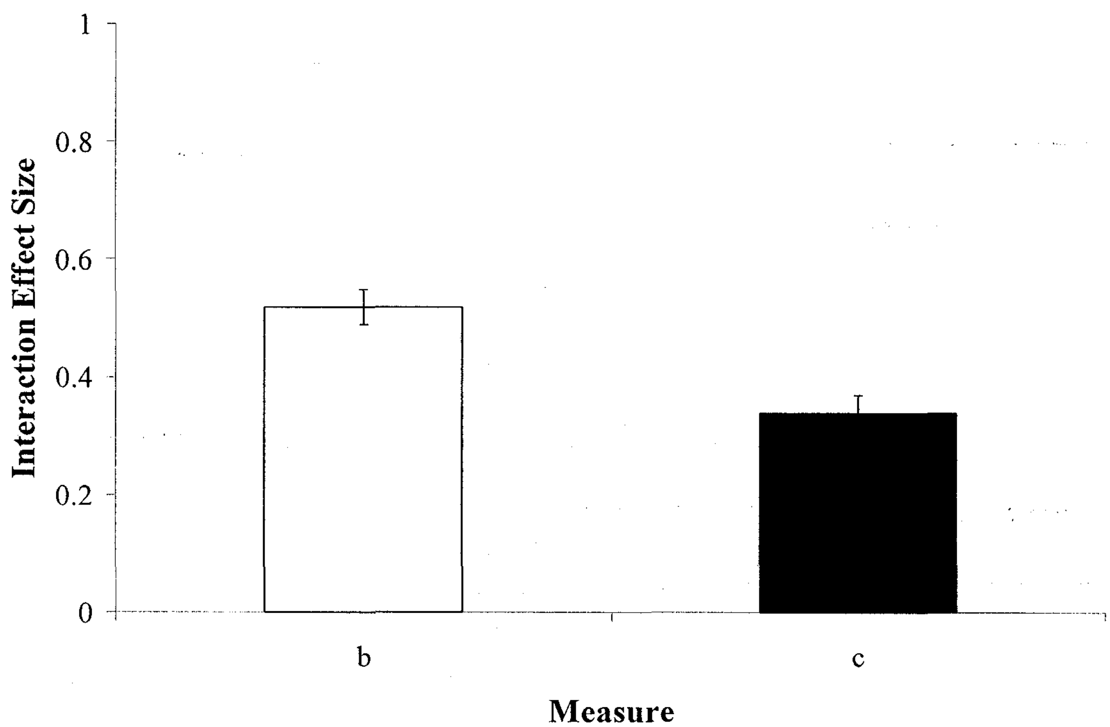


Figure 7. Differences in the interaction effect size between the *a-b* and traditional SDT models.

DISCUSSION

Results from the jackknifing study provided additional insight regarding the differences between the *a-b* SDT model and the traditional SDT model. The results from the jackknifing study indicated that the differences in the interaction effect size between the *a-b* SDT model and the traditional SDT model were not due to sampling error. These findings suggest that there are systematic differences in the degree of dependency between the *a* and *b* measures and the *d'* and *c* measures respectively. Furthermore, these findings served as an additional foundation for the Monte Carlo Study.

CHAPTER V

MONTE CARLO STUDY

As previously mentioned, the main contribution of SDT is that it allows researchers to examine the covert detection process independently of the overt response process. A fundamental requirement to satisfy this claim is to have a model from which to derive independent measures of accuracy or sensitivity and response bias or criterion setting. The lack of a correlative relationship between measures of accuracy or sensitivity and measures of response bias or criterion setting is necessary but not sufficient to conclude that such measures are independent. The only condition in which the lack of a correlative relationship would be sufficient to establish statistical independence between measures of accuracy or sensitivity and measures of response bias or criterion setting would be if both measures within a given model were bivariate normally distributed (Papoulis & Pillai, 2002).

In general though, two random variables, X and Y, are statistically independent if and only if, the conditional probability of a value of X given a value of Y equals the marginal probability of the value of X and vice versa (Papoulis & Pillai, 2002), or

$$p(X = x | Y = y) = p(X = x) \tag{6}$$

and,

$$p(Y = y | X = x) = p(Y = y) \tag{7}$$

Neither the *a-b* SDT model nor the traditional SDT model satisfies this property. Given the intuitive nature of the calculation and interpretation of *a* and *b*, this can be clearly shown using the *a-b* SDT model as an example. Based on Formulas 3 and 4, it is clear that in cases of an extreme accuracy score (either 0 or 1), the probability of *b* being equal to .5 is 1 and 0 for any other value of *b*. Likewise, given an extreme value of *b* (either 0 or 1), the probability of *a* being equal to 0 is 1 and 0 for any other value of *a*. However, given the complex nature of the computations of the traditional SDT measures, demonstrating lack of independence for them is not as simple. Furthermore, results from the empirical study and the jackknifing study suggest that this lack of independence may vary depending on the value of each measure. Therefore, the purpose of the Monte Carlo Study was to compare both models to examine the severity of their lack of independence across their full spectrum of potential values. Based on the results from the empirical study and the jackknifing study, the degree of dependency between *a* and *b* and *c* and *d* respectively was expected to vary according to the range of values of each measure.

METHOD

To maximize estimation accuracy, the Monte Carlo Study was based on a population of 100,000 cases. Based on prior research (Bustamante et al. 2006), hit and false alarm rates were randomly drawn from a uniform distribution. Both measures of each model were then calculated based on the simulated hit and false alarm rates. The Statistical Package for the Social Sciences (SPSS 14.0) was used to conduct the simulation and analyze the data (see Appendix G).

RESULTS

Given the lack of effective statistical tests of independence, the relationship between the measures within each model was analyzed graphically. The purpose of the graphical analysis was to examine the range of possible values of response bias or criterion setting given different ranges of values of accuracy or sensitivity.

Preliminary analyses consisted of frequency distributions and normal Q-Q plots of each measure to assess the nature of their univariate distributions (Tabachnick & Fidell, 2001)

Figures 8 to 11 show the frequency distributions of *a, b, d',* and *c* values.

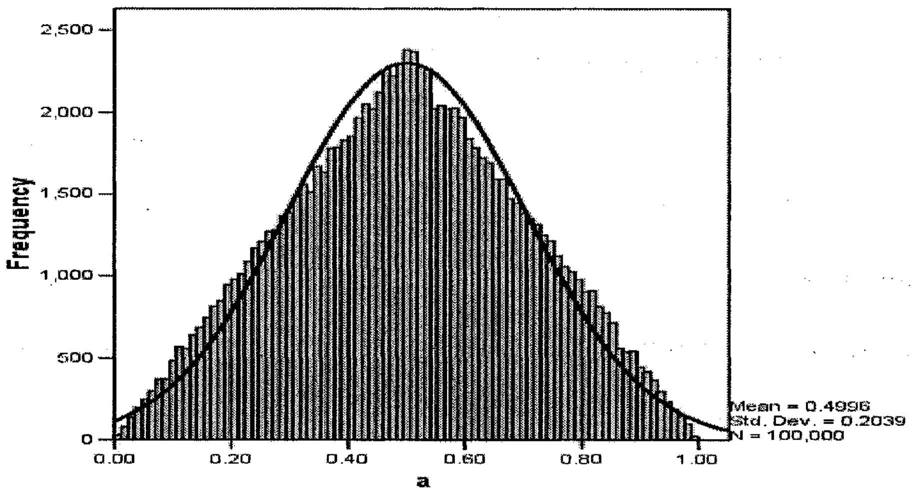


Figure 8. Frequency distribution of *a* values.

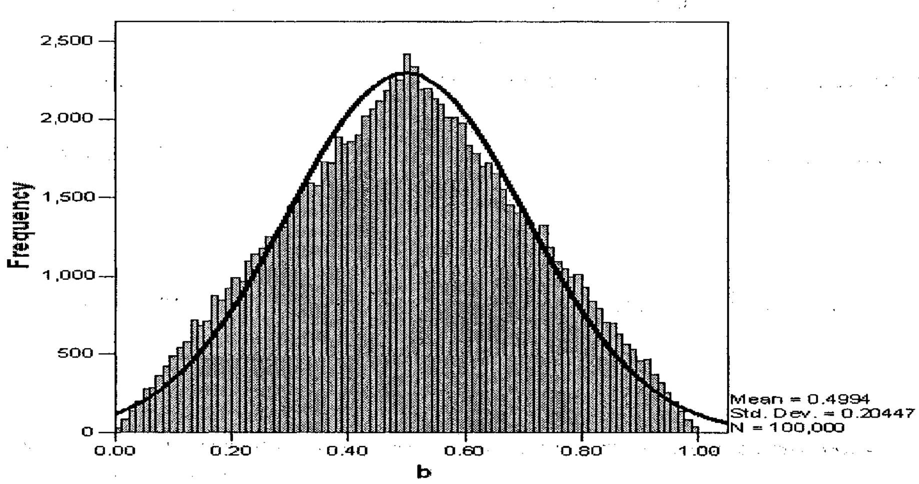


Figure 9. Frequency distribution of *b* values.

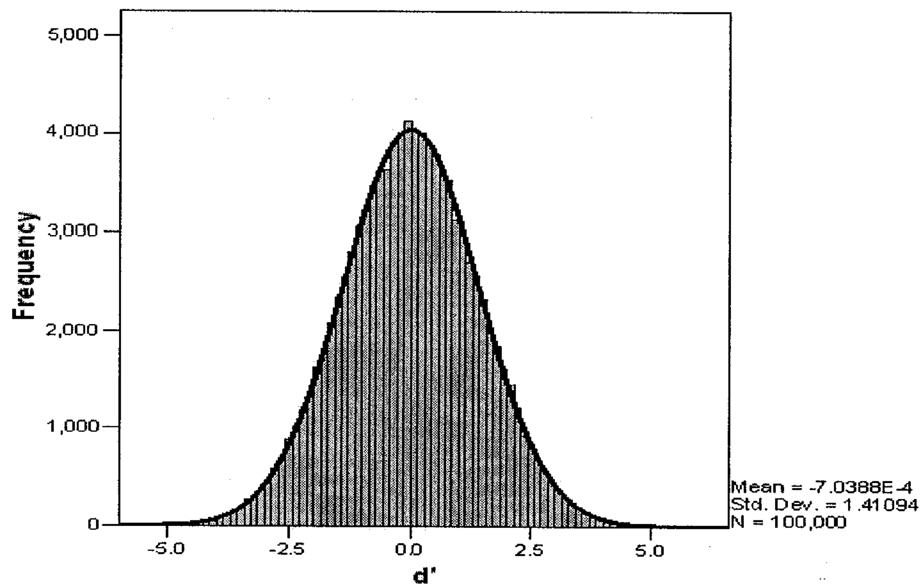


Figure 10. Frequency distribution of *d'* values.

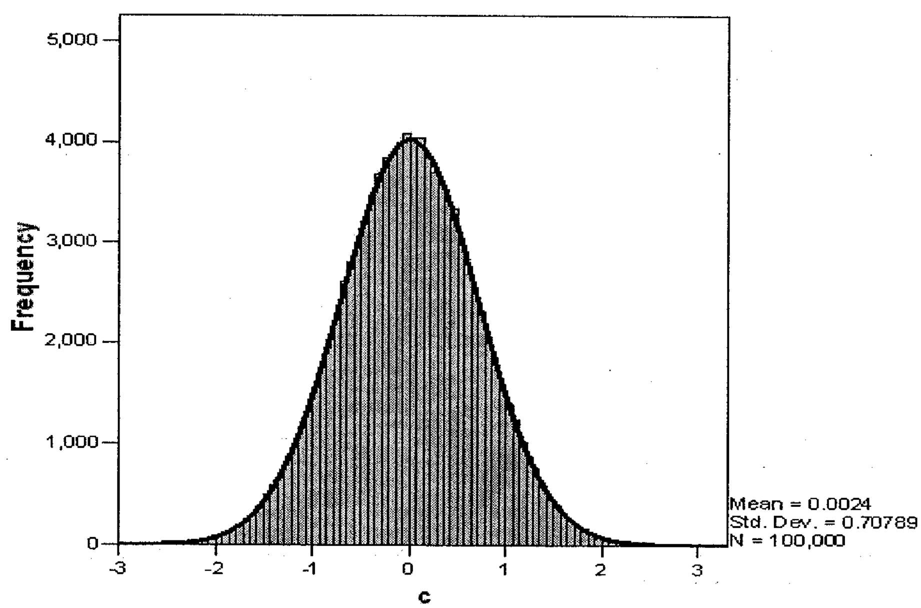


Figure 11. Frequency distribution of *c* values.

Figures 12 to 15 show the normal Q-Q plots of *a, b, d'*, and *c* values.

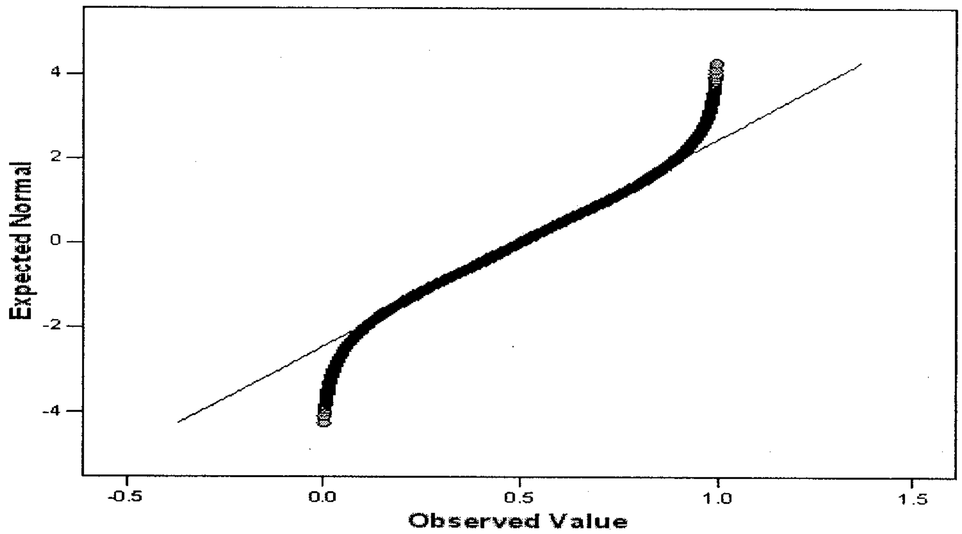


Figure 12. Normal Q-Q plot of *a* values.

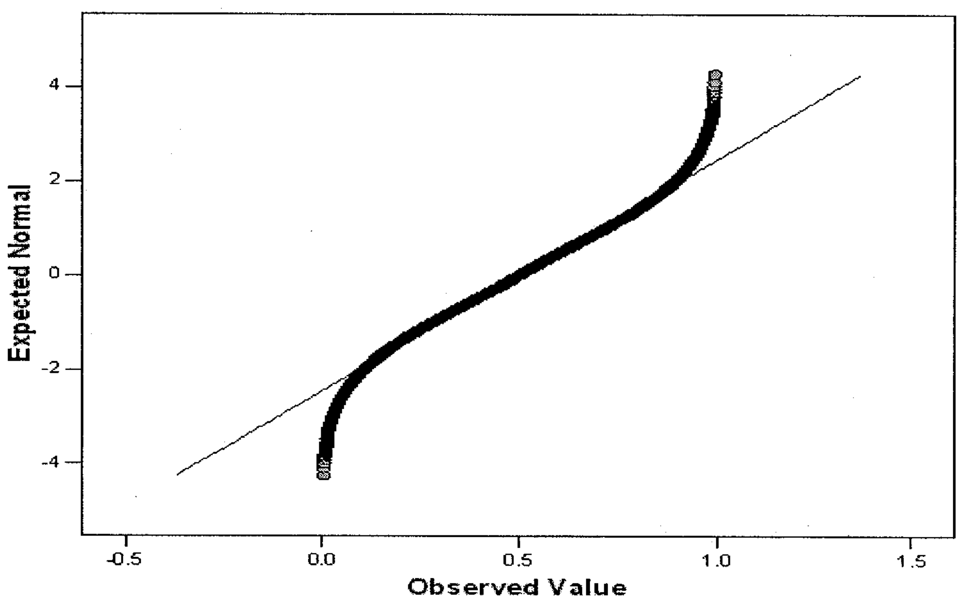


Figure 13. Normal Q-Q plot of *b* values.

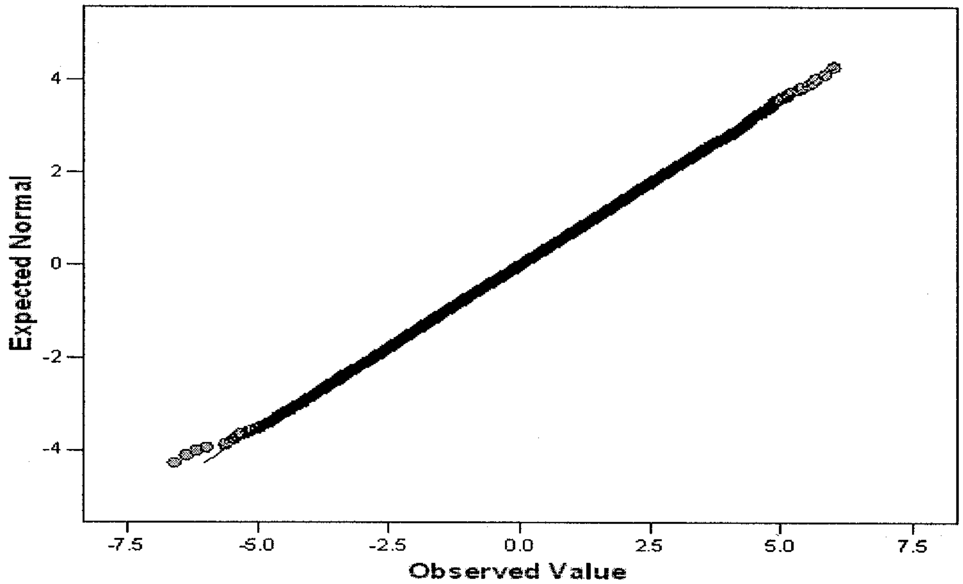


Figure 14. Normal Q-Q plot of *d'* values.

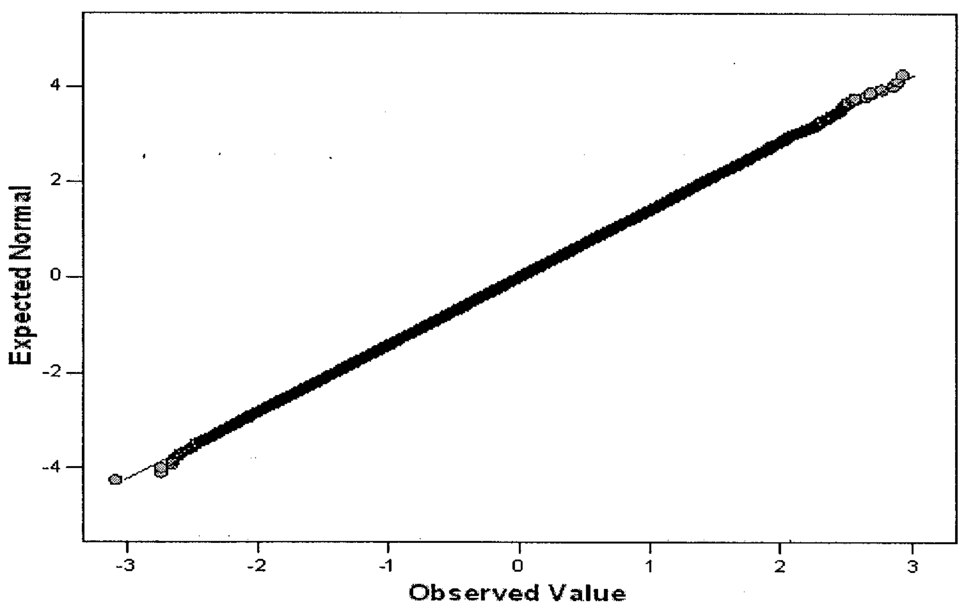


Figure 15. Normal Q-Q plot of *c* values.

Table 9 shows the correlations among the simulated measures.

TABLE 9: Correlations among Simulated Measures

Variable	1	2	3	4
1. Accuracy (*a*)	--			
2. Response Bias (*b*)	.00	--		
3. Sensitivity (*d'*)	.99*	.00	--	
4. Criterion Setting (*c*)	.00	-.98*	.00	--

* $p < .01$

Table 10 shows the results of the Kolmogorov-Smirnov test of univariate normality for each measure.

TABLE 10: Kolmogorov-Smirnov Test of Univariate Normality

Measure	K-S	p
a	.02	.00
b	.02	.00
d'	.00	.20
c	.00	.20

NOTE: *p* values were adjusted with the Lilliefors Significance Correction

As table 10 shows, *a* and *b* values were not normally distributed, whereas *d'* and *c* values were. Furthermore, as Figures 12 and 13 show, *a* and *b* values were not normally distributed as they approached their extreme scores.

Each set of measures (i.e., *a* and *b*, *d'* and *c*) was then plotted jointly to examine the nature of their relationship across the full spectrum of their values.

Figures 16 and 17 show the scatter plots of the *a-b* SDT measures and the traditional SDT measures across their full range of possible values.

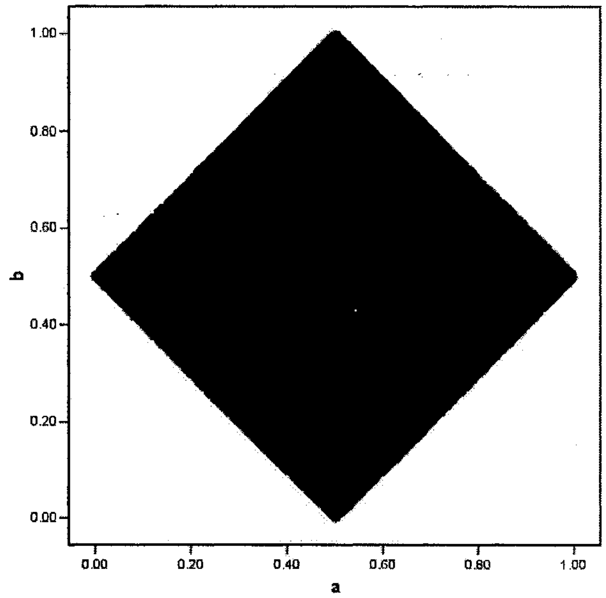


Figure 16. Scatter plot of the *a-b* SDT measures across their full range of possible values.

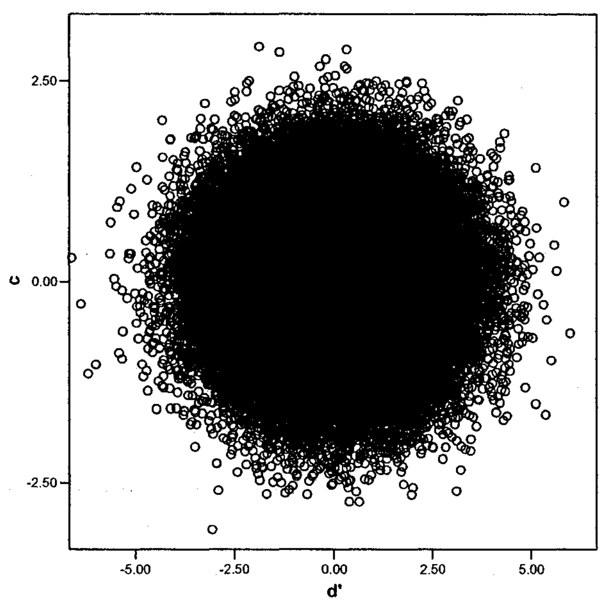


Figure 17. Scatter plot of the traditional SDT measures across their full range of possible values.

Figures 18 and 19 show the 3-D scatter plots of the *a-b* SDT measures and the traditional SDT measures across their full range of possible values.

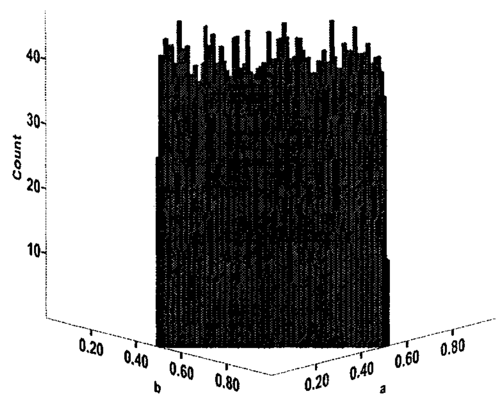


Figure 18. 3-D scatter plot of the *a-b* SDT measures across their full range of possible values.

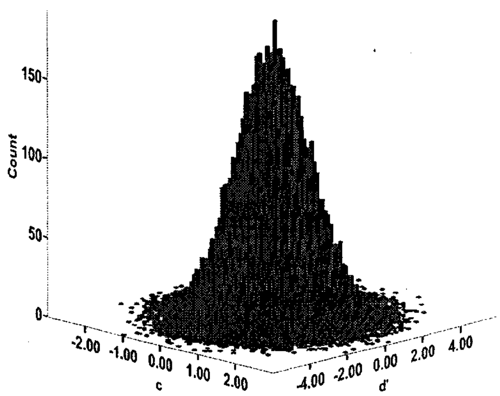


Figure 19. 3-D scatter plot of the traditional SDT measures across their full range of possible values.

Results from the 3-D scatter plot of the traditional SDT measures suggest that *d'* and *c* are bivariate normally distributed. Last, each set of measures was plotted jointly while restricting the range of one of the measures (i.e., *a* and *d'*) to explore the nature of the frequency distribution of the other measure (i.e., *b* and *c*) given the restricted set of values for the former measure. Given the intuitive interpretative nature of *a* values, scatter plots were restricted based on intervals of .10 across possible *a* values (see Appendix H). Only the *a* and *d'* values above chance performance were restricted for the subsequent graphical analyses. There were two reasons for doing this. First, given the symmetric nature of the scatter plots of *a* and *b* values and *d'* and *c* values respectively (see Figures 16 and 17), the graphical analyses above chance performance would be identical to those below chance performance. The second and most important reason for doing this was that in most settings, researchers and practitioners are interested in assessing human and automated system performance above chance level. Figure 20 shows the scatter plots of *a* and *b* values and *d'* and *c* values when .50 >= *a* < .60.

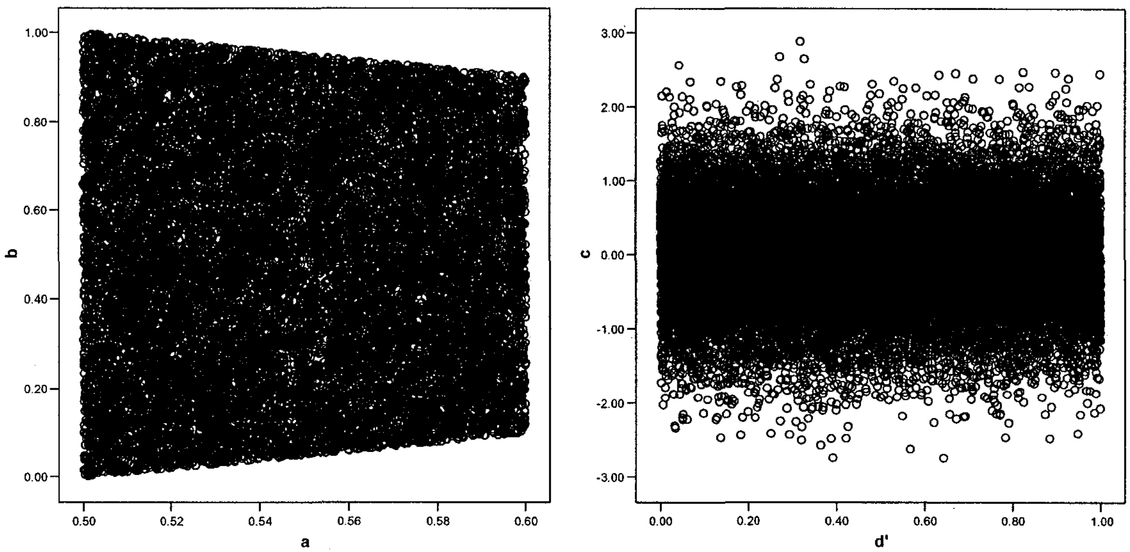


Figure 20. Scatter plots of *a* and *b* values and *d'* and *c* values when .50 >= *a* < .60.

Figure 21 shows the scatter plots of a and b values and d' and c values when .60 >= a < .70.

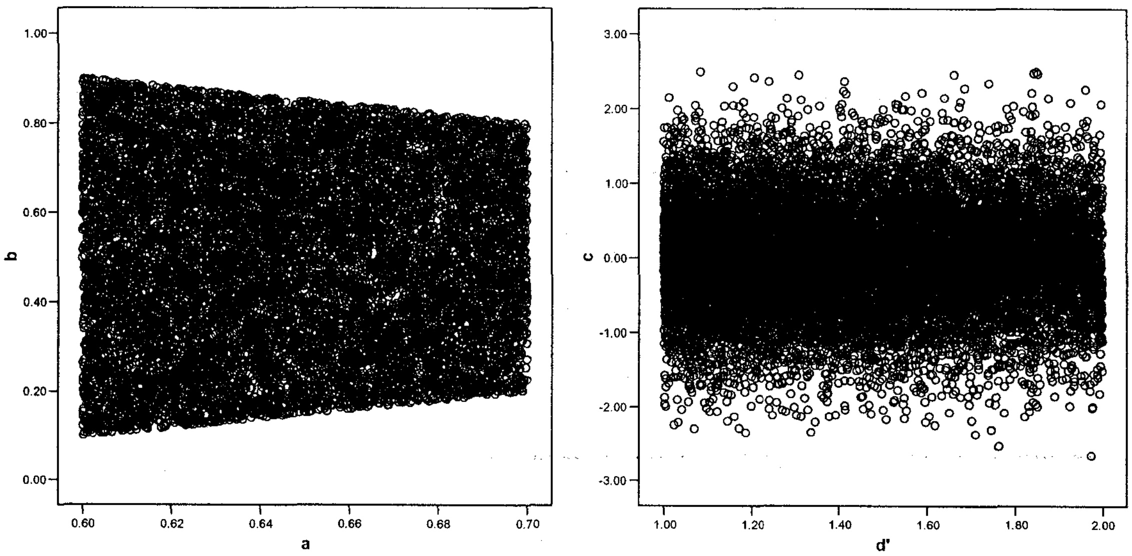


Figure 21. Scatter plots of a and b values and d' and c values when .60 >= a < .70.

Figure 22 shows the scatter plots of a and b values and d' and c values when .70 >= a < .80.

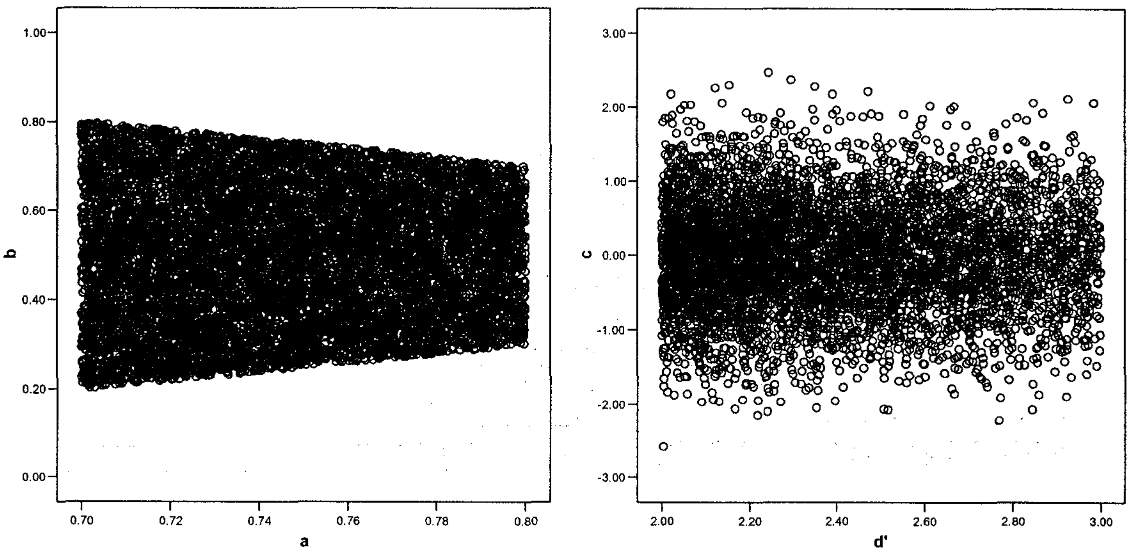


Figure 22. Scatter plots of a and b values and d' and c values when .70 >= a < .80.

Figure 23 shows the scatter plots of *a* and *b* values and *d'* and *c* values when .80 >= *a* < .90.

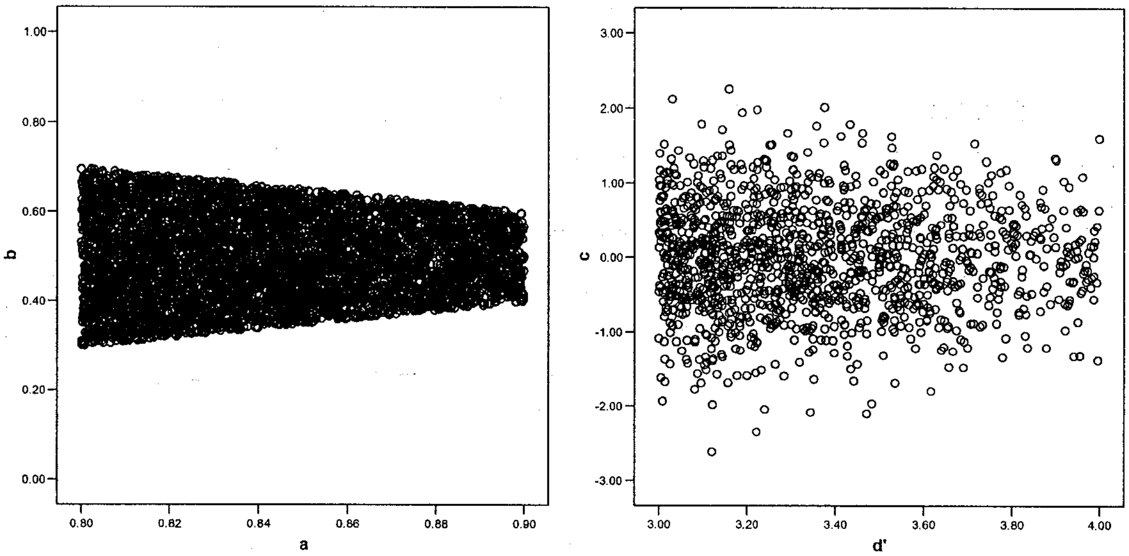


Figure 23. Scatter plots of *a* and *b* values and *d'* and *c* values when .80 >= *a* < .90.

Figure 24 shows the scatter plots of *a* and *b* values and *d'* and *c* values when .90 >= *a* < 1 .00.

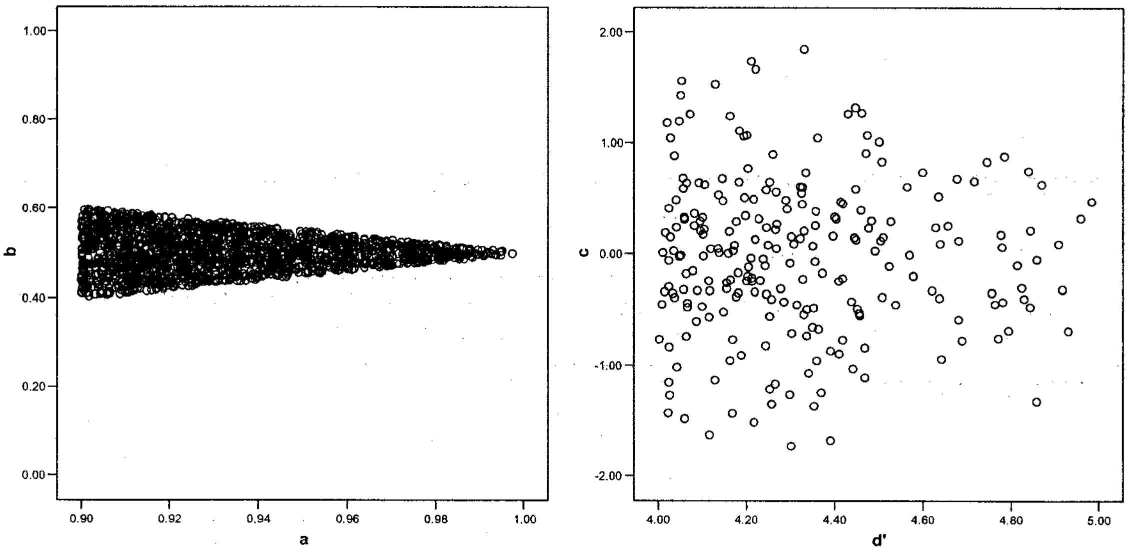


Figure 24. Scatter plots of *a* and *b* values and *d'* and *c* values when .90 >= *a* < 1.00.

DISCUSSION

Results from the preliminary analyses showed that unlike the traditional SDT measures, the *a-b* SDT measures were not normally distributed (see Table 10 and Figures 12 to 15). Results from the correlational analyses were consistent with prior research (Bustamante et al., 2006), indicating that both models provide uncorrelated measures of accuracy or sensitivity and response bias or criterion setting. However, results from the 3-D scatter plots of the *a* and *b* and *d'* and *c* measures across their full range of possible values showed that the relationship between the *a* and *b* measures is different than that of the *d'* and *c* measures (see Figures 18 to 19). More specifically, unlike the *a-b* SDT measures, the traditional SDT measures seem to be bivariate normally distributed. These findings suggest that the traditional SDT measures are statistically independent from each other, whereas the *a-b* SDT measures are statistically dependent on each other. Furthermore, as expected, as the value of *a* increased, the range of possible *b* values decreased (see Figures 20 to 24). Consistent with the empirical study and the jackknifing study, the results from the Monte Carlo Study suggest that the nature of the dependency between the *a* and *b* measures is stronger than that of the *d'* and *c* measures.

Based on Green and Swets (1966)'s argument that the detection and response processes are independent, the results from the Monte Carlo Study would seem to suggest that the traditional SDT model provides more adequate measures than the *a-b* SDT model. However, the results from the empirical study and the jackknifing study bring into question the viability of Green and Swets (1966)'s argument.

According to Green and Swets (1966), non-sensory factors, such as the probability of the target stimulus, should have no effect on sensitivity. Likewise, sensory factors, such as the frequency difference between the baseline and target stimuli, should have no effect on criterion setting. However, the results from the empirical study and the jackknifing suggested that the effect of sensory factors can impact the effect size of non-sensory factors on criterion setting or response bias. These are crucial findings because they suggest that factors that affect the covert detection process can also affect the overt response process, thereby establishing a dependency between these two processes.

Consequently, the results from this research suggest that contrary to Green and Swets (1966)'s argument, the detection and response processes are dependent, and the nature of their dependency increases as the difference between the baseline and target stimuli increases. Furthermore, the findings from this research suggest that the *a-b* SDT model is a more adequate framework for detecting the dependency of the detection and response processes.

An applied example of a domain in which the *a-b* SDT model may be a more adequate framework to analyze human performance is pilots' decision-making during potential weather threats. Research shows that in general, commercial aviation pilots have a tendency to deviate from their predetermined flight paths due to potential weather threats (Bliss, Fallon, Bustamante, Bailey, & Anderson, 2005). Two of the main reasons for this tendency to deviate from potential weather threats are passengers' safety and comfort. The problem is that making unnecessary flight path deviations can have negative effects, such as increased fuel consumption and flight delays.

Within the context of SDT, passengers' safety and comfort constitute non-sensory factors that increase pilots' biases toward deviating from their predetermined flight paths. Researchers and designers can use the *a-b* SDT model to examine how sensory factors, such as the characteristics of weather displays, can mitigate the effect of non-sensory factors. The purpose of this would be to increasing pilots' decision-making accuracy to the point where they would not be biased by such non-sensory factors and would avoid making unnecessary flight path deviations.

CHAPTER VI

CONCLUSION

The combined results from this research provide important theoretical and practical contributions to researchers and practitioners. Given its lack of reliance on the assumption of a single underlying continuum, the *a-b* SDT model is more generalizable and applicable than the traditional model. This greater generalizability and applicability is particularly important for researchers and practitioners who are interested in examining the performance of humans and automated systems in complex domains, such as aviation, driving, luggage screening, and medical diagnosis, in which neither humans nor automated systems make decisions based on a single underlying continuum.

More importantly though, the findings from this research suggest that Green and Swets (1966)'s contention that the detection and response processes are independent from each other does not hold true for either the *a-b* SDT model or the traditional SDT model. An important point to note is that although the traditional SDT model provides independent measures of sensitivity and criterion setting, the empirical data suggests that the covert detection and overt response processes are not independent. Therefore, the *a-b* SDT model provides more accurate measures to capture the dependency between these two processes.

This is particularly important for researchers and practitioners who are interested in examining not only the detection or decision-making accuracy of humans and automated systems, but also their response biases. More specifically, researchers interested in studying human-automation interaction factors, such as compliance, reliance, and trust, may benefit from using the *a-b* SDT model to assess how sensory and perceptual factors affect humans' response biases while interacting with automated systems.

Technological advances have made the use of automated systems a common practice in a variety of task domains, including aviation (Bliss, 2003), air traffic control (Masalonis & Parasuraman, 2003), ground transportation (Shinar, 2000), medicine (Weinger, 2000), mining (Mallett, Vaught, & Brnich, 1993), ship handling (Kerstholt, Passenier, Houttuin, & Schuffel, 1996), and nuclear power control (Bransby, 2001). The increased used of automated systems has changed the role of humans from operators to system monitors (Parasuraman & Riley, 1997). Human monitors are notoriously ineffective in complex situations characterized by high levels of workload (Woods, 1995). Engineers and designers have developed automated alarm systems to assist human monitors (Papadopoulos & McDermid, 2001).

Advanced sensor technologies and fault-diagnosis algorithms have allowed alarm systems to detect the presence of dangerous conditions effectively (Tumer & Bajwa, 1999). The primary purposes of alarm systems are to detect dangerous conditions and attract operators' attention so that they can either avoid or escape problems (Xiao & Seagull, 1999). Ideally, systems should issue alarms only when there is an actual underlying problem present. However, because of legal implications, system designers tend to follow the engineering fail safe approach, setting the threshold of alarm systems low enough to alert operators of even the slightest possibility of a problem (Swets, 1992). Moreover, the rare occurrence of dangerous conditions makes it difficult for designers to develop alarm systems that emit a low number of false alarms (Parasuraman & Hancock, 1999). Consequently, most alarm systems generate many false alarms (Getty, Swets, Pickett, & Gonthier, 1995). Frequent false alarms cause a cry-wolf effect, which leads to a loss of trust in the system and a decrease in operator compliance with alarm signals (Breznitz, 1983). As a result of this cry-wolf effect, operators often ignore or cancel alarms without searching for additional information that could help them detect the presence of dangerous conditions (Sorkin, 1988).

Researchers have tried to mitigate the cry-wolf effect by focusing on increasing operators' biases toward responding to alarm signals. To that end, researchers have manipulated hearsay information, the perceived urgency of such signals, and reaction modalities (Bliss, 1997; Bliss, Dunn, & Fuller, 1995). Although such solutions may result in higher response rates to true signals, they also increase responses to false alarms. This, in turn, may increase their level of workload and hinder primary-task performance (Gilson & Phillips, 1996).

Researchers have also tried to adapt the traditional SDT model to better characterize operators' interactions with automated alarm systems (Bustamante et al., 2004; Lehto & Papastavrou, 1998; Meyer, 2004; Sorkin & Woods, 1985). The underlying purpose for adapting SDT to analyze human responses to alarms is to examine system characteristics that may increase operators' abilities to distinguish between true and false alarms, thereby increasing operators' responses to only true alarms and decreasing their responses to false alarms. However, one of the main problems with adopting the traditional SDT model to this domain is that when operators interact with automated alarm systems in complex environments, they do not make decisions based solely on a single underlying psychophysical continuum. Instead, when responding to alarms, operators need to take into account several factors that vary according to the task at hand. Some of these factors include: the reliability of the system (Bliss, Gilson, et al., 1995; Getty et al., 1995), the type of response modality (Bliss, 1997), the perceived urgency of alarm signals (Bliss, Dunn, et al., 1995; Edworthy & Loxley, 1991), the presence of likelihood and task-critical information (Bustamante, 2005), workload (Bustamante & Bliss, 2005), and the acoustic parameters of alarm signals (Edworthy, Hellier, & Hards, 1995). Therefore, given its lack of reliance on the assumption of a single underlying psychophysical continuum, the *a-b* SDT model may serve as a more adequate framework for characterizing operators' interactions with automated alarm systems.

More importantly though, the findings from this research showed evidence to support the superiority of the *a-b* SDT model over the traditional SDT model. Aside from being more parsimonious and generalizable, the *a-b* SDT provides measures of accuracy and response bias that more adequately capture the dependency between the covert detection and overt response processes. As such, researchers and practitioners can use the *a-b* SDT model to more adequately examine how sensory system characteristics that can improve operators' accuracy interact with non-sensory factors to affect operators' reliance on and compliance with automated systems.

REFERENCES

Ayres, T. J., Wood, C. T., Schmidt, R. A., & McCarthy R. L. (1998). Risk perception and behavioral choice. *International Journal of Cognitive Ergonomics, 2*(1-2), 35-52.

Bass, E. J. (2006). Comparison of four quantitative methods of measuring human judgment performance in a simulated horizontal air traffic conflict prediction task. *International Journal of Applied Aviation Studies, 6*(1), 25-46.

Bailey, N. R., & Scerbo, M. W. (2005). Concurrent monitoring for multiple critical signals in a complex display: A vigilance perspective. In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1518-1522). Santa Monica, CA: Human Factors and Ergonomics Society.

Bisseret, A. (1981). Application of signal detection theory to decision making in supervisory control: The effect of the operator's experience. *Ergonomics, 24*(2), 81-94.

Bliss, J. P. (1997). Alarm reaction patterns by pilots as a function of reaction modality. *The International Journal of Aviation Psychology, 7*(1), 1-14.

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *The International Journal of Aviation Psychology, 13*(3), 249-268.

Bliss, J. P., Dunn, M. C., & Fuller, B. S. (1995). Reversal of the cry-wolf effect: An investigation of two methods to increase alarm response rates. *Perceptual and Motor Skills, 80*, 1231-1242.

Bliss, J. P., Fallon, C. K., Bustamante, E. A., Bailey, W. R., & Anderson, B. L. (2005). *Reactions of air transport flight crews to displays of weather during simulated flight.* (Final Report, ODU Contract 130971). Langley, VA: NASA Langley Research Center.

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics, 38*(11), 2300-2312.

Bransby, M. (2001). Design of alarm systems. In J. Noyes & M. Bransby (Eds.), *People in control: Human factors in control room design* (pp. 207-222). Edison, NJ: Institution of Electrical Engineers.

Breznitz, S. (1983). *Cry-wolf: The psychology of false alarms.* Hillsdale, NJ: Lawrence Erlbaum.

Broadbent, D. (1978). The current state of noise research: Reply to Poulton. *Psychological Bulletin, 85*, 1052-1067.

Burlingame, E., Sussman, H. M., & Gillam, R. B. (2005). An investigation of speech perception in children with specific language impairment on a continuum of formant transition duration. *Journal of Speech, Language, and Hearing Research, 48*(4), 805-816.

Bustamante, E. A. (2005). A signal detection analysis of the effects of workload, task-critical and likelihood information on human alarm response. In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1513-1517). Santa Monica, CA: Human Factors and Ergonomics Society.

Bustamante, E. A., Anderson, B. L., Thompson, A. R., Bliss, J. P., Scerbo, M. W. (in press). Robustness of the *a b* signal detection theory model of decision making. In *Proceedings of the Human Factors and Ergonomics Society 51th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Bustamante, E. A., & Bliss, J. P. (2005). Effects of workload and likelihood information on human response to alarm signals. In *Proceedings of the 13th International Symposium on Aviation Psychology* (pp. 81-85). Oklahoma City, OK: Wright State University.

Bustamante, E. A., Bliss, J. P., & Anderson, B. L. (in press). Effects of varying the threshold of alarm systems and workload on human performance. *Ergonomics*.

Bustamante, E. A., Fallon, C. K., & Bliss, J. P. (2004). Adapting signal detection theory for analyzing human response to alarm signals: A theoretical approach [Abstract]. *American Psychological Association Divisions 19 & 21 Midyear Symposium*. Alexandria, VA, March 4-5.

Bustamante, E. A., Fallon, C. K., & Bliss, J. P. (2006). A more parsimonious approach to estimating signal detection theory measures. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 1711-1715). Santa Monica, CA: Human Factors and Ergonomics Society.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Coyne, J. T. (2005). Crisp and fuzzy signal detection theory and pilot weather judgment: Implications for VFR flights into IMC. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*(1-B), 597.

Craig, A. (1979). Nonparametric measures of sensory efficiency for sustained monitoring tasks. *Human Factors, 21*(1), 69-78.

Deller, J. R., Desai, K. H., & Yang, Y. P. (2005). A decision-theoretic framework for the evaluation of language models used in speech recognizers. *Natural Language Engineering, 11*(4), 363-393.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121*, 275-277.

Donaldson, W. (1993). Accuracy of *d'* and *A'* as estimates of sensitivity. *Bulletin of the Psychonomic Society, 31*, 271-274.

Drury, C. G., Ghylin, K. M., & Holness, K. (2006). Error analysis and threat magnitude for carry-on bag inspection. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 1189-1193). Santa Monica, CA: Human Factors and Ergonomics Society.

Edworthy, J., Hellier, E., & Hards, R. (1995). The semantic association of acoustic parameters commonly used in the design of auditory information and warning signals. *Ergonomics, 38*(11), 2341-2361.

Edworthy, J., & Loxley, S. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors, 33*(2), 205-231.

Eubanks, J. L., & Killeen, P. R. (1983). An application of signal detection theory to air combat training. *Human Factors, 25*(4), 449-456.

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1*, 19-33.

Gilson, R.D., & Phillips, M.J. (1996). Alert or sound alarm? *Aerospace Engineering*, 21-23.

Green. D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424-429.

Hancock, P. A., Masalonis, A. J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical applications. *Theoretical Issues in Ergonomics Science, 1*(3), 207-230.

Harvey, L. O. (1992). The critical operating characteristic and the evaluation of expert judgment. *Organizational Behavior and Human Decision Processes, 53*(2), 229-251.

Hautus, M. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d*'. *Behavior Research Methods, Instruments, & Computers, 27*, 46-51.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74*, 351-354.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.

Kerstholt, J.H, Passenier, P.O., Houttuin, K., & Schuffel, H. (1996). The effect of a priori probability and complexity on decision making in a supervisory control task. *Human Factors,* 38(1), 65-78.

Klein, G. A., Orasanu, J., Calderwood, R. & Zsambok, C. E. (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex Publishing.

Krueger, G.P. (1989). Sustained work, fatigue, sleep loss and performance: A review of the issues. *Work and Stress, 3*(2) 129-141.

Lehto, M. R., & Papastavrou, J. D. (1998). A signal-detection theory based perspective on design of warning. *Perceptual and Motor Skills, 86*, 720-722.

Li, C. R., Lin, W. Chang, H. (2004) A psychophysical measure of attention deficit in children with attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology, 113*(2), 228-236.

Long, G. M., & Waag, W. L. (1981). Limitations on the practical applicability of *d*' and β measures. *Human Factors, 23*(3), 285-290.

Madhavan, P. & Gonzales, P. (2006). Effects of sensitivity, criterion shifts, and subjective confidence on the development of automaticity in airline luggage screening. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 334-338). Santa Monica, CA: Human Factors and Ergonomics Society.

Mallett, L.G., Vaught, C., & Brnich, M.J., Jr. (1993). Sociotechnical communication in an underground mine fire: A study of warning messages during an emergency evacuation. *Safety Science, 16*, 709-728.

Masalonis, A. J., & Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics, 46*(11), 1045-1074.

McCornack, R. L. (1961). *Inspector accuracy: A study of the literature*. Albuquerque: Sandia Corporation.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50*, 215-241.

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors, 46*(2), 196-204.

Meyer, J. & Ballas, E. (1997). A two-detector signal detection analysis of learning to use alarms. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 186-189). Santa Monica, CA: Human Factors and Ergonomics Society.

Mota, V. L., & Schachar, R. J. (2000). Reformulating attention-deficit/hyperactivity disorder according to signal detection theory. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(9), 1144-1151.

Papadopoulos, Y., & McDermid, J. (2001). Automated safety monitoring: A review and classification of methods. *International Journal of Condition Monitoring and Diagnostic Engineering Management, 4*(4), 1-32.

Papoulis, A., & Pillai, S. U. (2002). *Probability, Random Variables, and Stochastic Processes*. New York, NY: McGraw-Hill.

Parasuraman, R., & Hancock, P. A. (1999). Using signal detection theory and Bayesian analysis to design parameters for automated alarm systems. In M. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 63-67). Mahwah, NJ: Erlbaum.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230-253.

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics, 40*(3), 390-399.

Parasuraman, R., Masalonis, A. J., & Hancock, P. A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*(4), 636-659.

Peterson, D. A. (1999). The effects of experience and training in the use of a terrain enhanced primary flight display for terrain avoidance (pilot training). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 60*(6-B), 2998.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1*, 125-126.

Shinar, D. (2000). Fleet study evaluation of an advanced brake warning system. *Human Factors, 42*(3), 482-489.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34-50.

Sorkin, R. D. (1988). Why are people turning off our alarms? *Journal of the Acoustical Society of America, 84,* 1107-1108.

Sorkin, R. D., Hays, C. J., & West, R. (2001). Signal-detection analysis of group decision making. *Psychological Review, 108*(1), 183-203.

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human Computer Interaction, 1*, 49-75.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Swets, J. A. (1961). Is there a sensory threshold? *Science, 134*, 168-177.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science, 182*(4116), 990-1000.

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist, 47*(4), 522-532.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers.* Mahwah, NJ: Lawrence Erlbaum Associates.

Tabachnick, B. G., & Fidell, L. S. (2001). *Computer-assisted research design and analysis*. Needham Heights, MA: Allyn & Bacon.

Tumer, I. Y., & Bajwa, A. (1999). A survey of aircraft engine health monitoring systems. In *Proceedings of the 35th Joint Propulsion Conference, Aerospace Systems Condition Monitoring Session*, (pp. 1-7). Los Angeles, CA: American Institute of Aeronautics and Astronautics.

Weinger, M. B. (2000). Auditory warnings in the medical work domain: An overview of critical issues. In *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 211-214). Santa Monica, CA: Human Factors and Ergonomics Society.

Wogalter, M. S., Allison, S. T., & McKenna, N. A. (1989). Effects of cost and social influence on warning compliance. *Human Factors, 31*, 133-140.

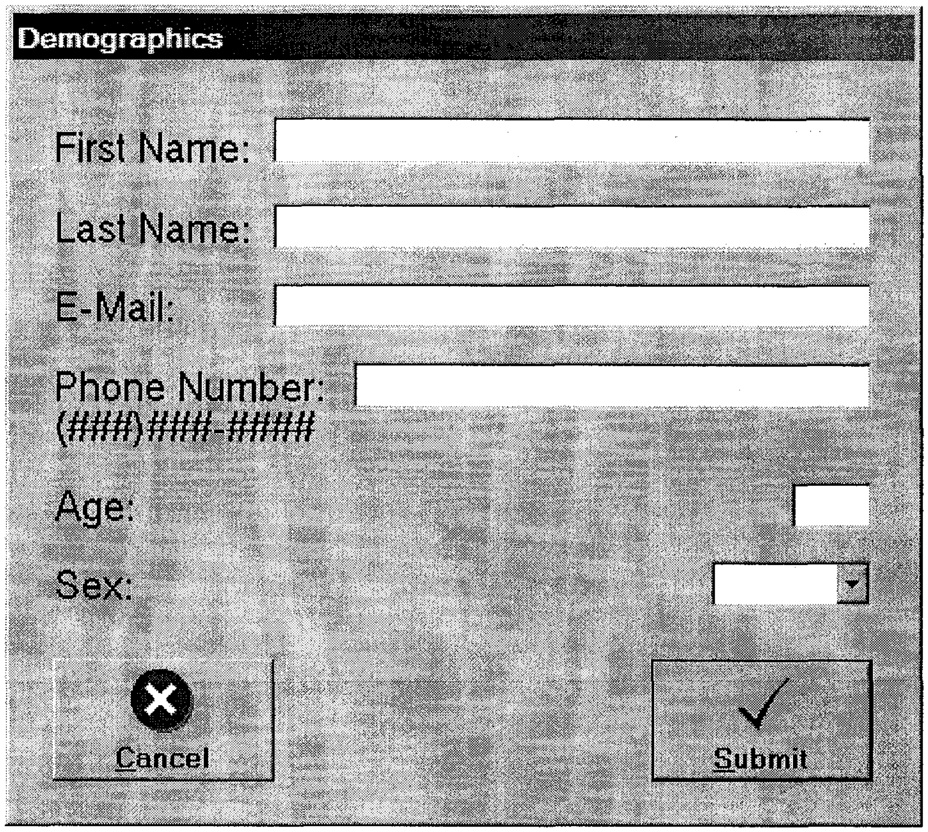
Wolf, Y., Algom, D. Lewin, I. (1988). A signal detection theory analysis of a driving decision task: Spatial gap acceptance. *Perceptual and Motor Skills, 66*(3), 683-702.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics, 38*(11), 2371-2393.

Xiao, Y., & Seagull, F.J. (1999). An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks. *Proceedings of the Human Factors and Ergonomic Society 43rd Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 256-260.

APPENDIX A

Demographic and Contact Information



APPENDIX B

Participant Instructions

Instructions

INSTRUCTIONS

The purpose of this study is to assess your ability to differentiate a target auditory signal from a baseline signal. We are specifically interested in examining how the probability of occurrence of the target signal and the difference in frequency between the target and baseline signals affects your performance.

As part of this experiment, you will perform nine one-minute practice sessions and nine five-minute experimental sessions, which will vary according to the probability of occurrence of the target signal (i.e., .10, .50, .90) and the difference in the frequency between the target and baseline signals (i.e., 5 Hz, 10 Hz, 15 Hz).

Your job will consist of pressing the #1 Key on top of the keyboard for those trials in which you perceive that the target signal was presented and pressing the #0 Key on top of the keyboard for those trials in which you perceive that the target signal was not presented. To maintain experimental control, please place and keep your left middle finger on top of the #1 Key and your right middle finger on top of the #0 Key during each session.

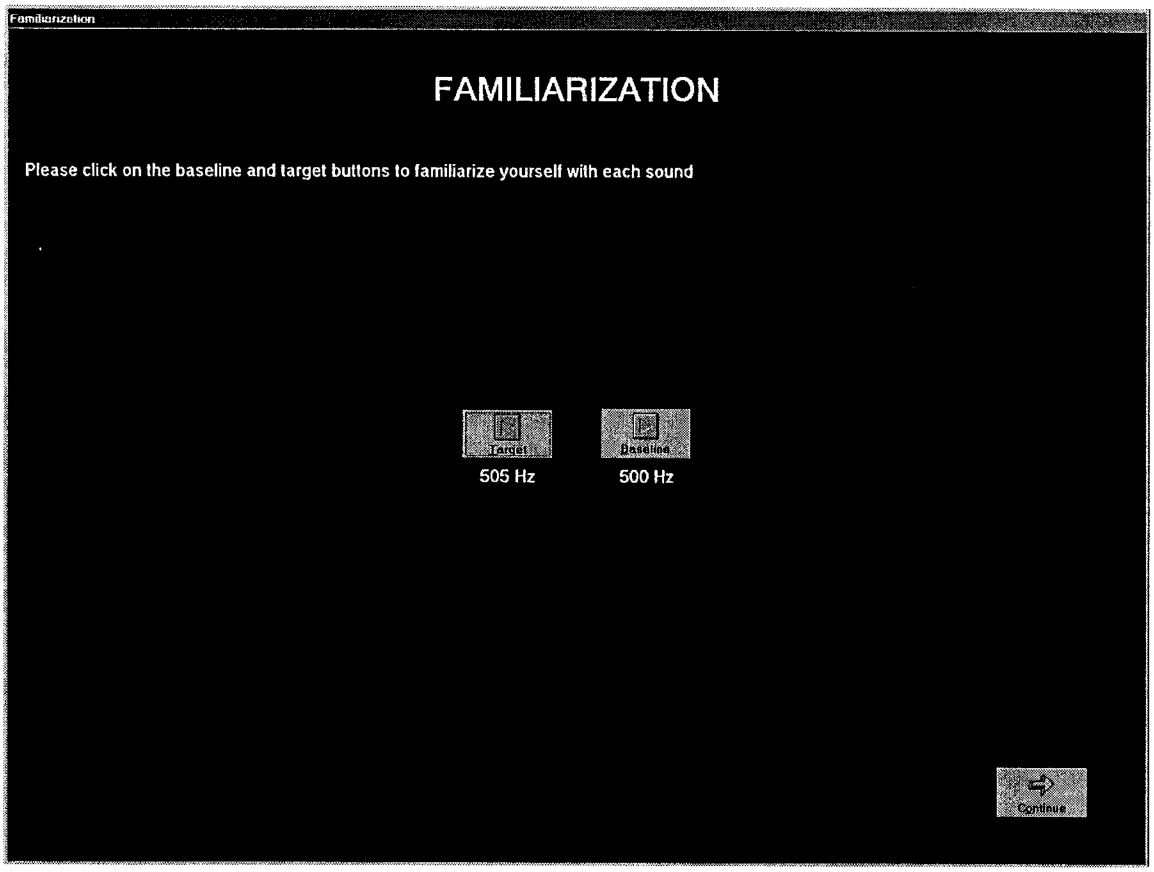
Throughout each trial, you will receive feedback on the accuracy of your decision through the changes in your performance score. You will start each experimental session with a score of zero. For each individual trial, if you make an accurate decision, you will receive one point, which will be added to your total score. Similarly, for each incorrect decision, you will lose one point, which will be subtracted from your total score. The maximum score you can receive in a given session is 100 points and the minimum is -100 points. The participant with the highest average score across sessions in the experiment will receive $100.

If you have any questions, please notify the experimenter at this time.

Continue

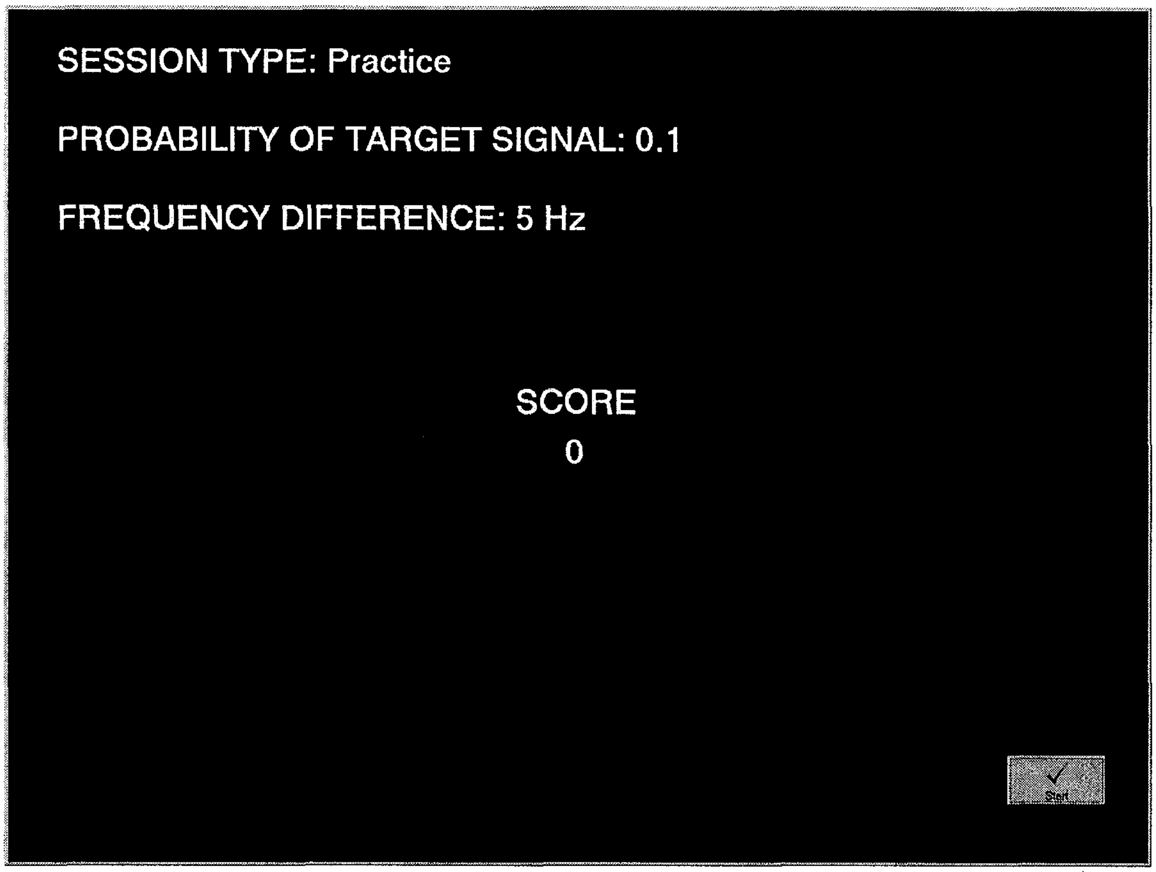
APPENDIX C

Familiarization



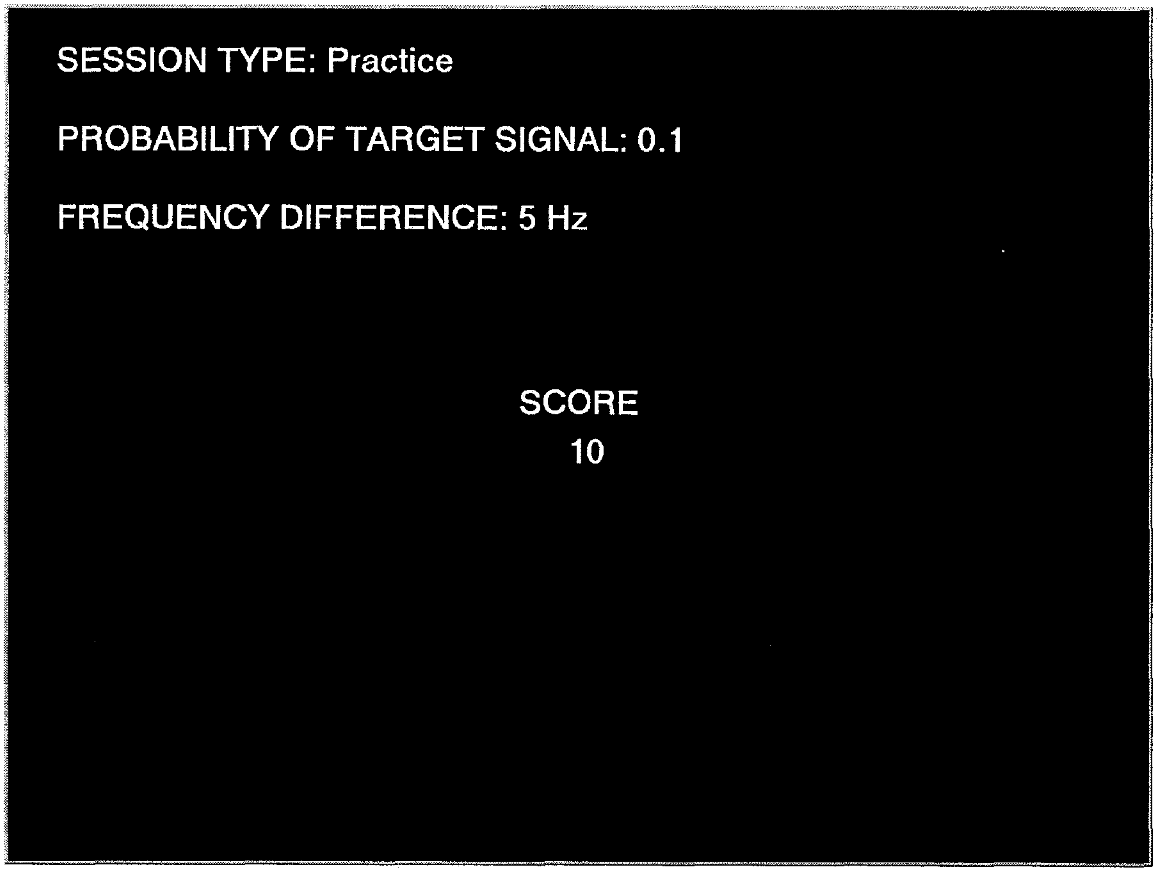
APPENDIX D

Session Information



APPENDIX E

Performance Feedback



APPENDIX F

Old Dominion University Informed Consent Form

PROJECT TITLE: Signal Detection

INTRODUCTION
The purposes of this form are to give you information that may affect your decision whether to say YES or NO to participation in this research, and to record the consent of those who say YES. It is your right and responsibility to inform the researcher if you wish to cease participation at any time.

RESEARCHERS
James P. Bliss, Ph.D., Associate Professor, College of Sciences, Psychology Department
Ernesto A. Bustamante, M.S., ABD, Graduate Student, College of Science, Psychology Department

DESCRIPTION OF RESEARCH STUDY
The purpose of this study is to assess your ability to differentiate a target auditory signal from a baseline signal. We are specifically interested in examining how the probability of occurrence of the target signal and the difference in frequency between the target and baseline signals affects your performance. As part of this experiment, you will perform nine one-minute practice sessions and nine five-minute experimental sessions, which will vary according to the probability of occurrence of the target signal (i.e., .10, .50, .90) and the difference in the frequency between the target and baseline signals (i.e., 5 Hz, 10 Hz, 15 Hz).

Your job will consist of pressing the #1 Key on top of the keyboard for those trials in which you perceive that the target signal was presented and pressing the #0 Key on top of the keyboard for those trials in which you perceive that the target signal was not presented.

Throughout each trial, you will receive feedback on the accuracy of your decision through the changes in your performance score. You will start each experimental session with a score of zero. For each individual trial, if you make an accurate decision, you will receive one point, which will be added to your total score. Similarly, for each incorrect decision, you will lose one point, which will be subtracted from your total score. The maximum score you can receive in a given session is 100 points and the minimum is -100 points. The participant with the highest average score across sessions in the experiment will receive $100.

If you decide to participate, you will join a study involving research on the development and refinement of the *a b* Signal Detection Theory Model of Decision Making. If you say YES, your participation may last up to one hour at the laboratory in Mills Godwin Building room 234. Approximately 20 of Old Dominion University students will be participating in this study.

EXCLUSIONARY CRITERIA
To the best of your knowledge, you should not have any diagnosed hearing or vision deficits that would keep you from participating in this study. If you do have any of these deficits, you must wear the required corrective lenses or hearing aid. You must be at least 18 years of age to participate.

RISKS AND BENEFITS
RISKS: If you decide to participate in this study, then you may face a risk of the common problems associated with computer usage. The researcher tried to reduce these risks by minimizing the amount of time in front of the computer. Also, as with any research, there is some possibility that you may be subject to risks that have not yet been identified.

APPENDIX F (continued)

BENEFITS: The main benefit to you for participating in this study is that you will receive 1 Psychology research credit that may be used for extra credit or to fulfill a class requirement. It is possible to acquire this credit in other ways without participating in this experiment. You will also earn $100 if you are the best performer of the entire participant pool.

COSTS AND PAYMENTS
The researchers want your decision about participating in this study to be absolutely voluntary. There is not cost to participate and the researchers will pay you $100 if you are the best performer of the entire participant pool.

NEW INFORMATION
If the researchers find new information during this study that would reasonably change your decision about participating, then they will give it to you.

CONFIDENTIALITY
All information obtained about you in this study is strictly confidential unless disclosure is required by law. The results of this study may be used in reports, presentations and publications, but the researcher will not identify you.

WITHDRAWAL PRIVILEGE
It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and walk away or withdraw from the study -- at any time. Your decision will not affect your relationship with Old Dominion University, or otherwise cause a loss of benefits to which you might otherwise be entitled. The researchers reserve the right to withdraw your participation in this study, at any time, if they observe potential problems with your continued participation.

COMPENSATION FOR ILLNESS AND INJURY
If you say YES, then your consent in this document does not waive any of your legal rights. However, in the event of harm, injury, or illness arising from this study, neither Old Dominion University nor the researchers are able to give you any money, insurance coverage, free medical care, or any other compensation for such injury. In the event that you suffer injury as a result of participation in this research project, you may contact Dr. James P. Bliss at 757-683-4051 or Dr. David Swain the current IRB chair at 757-683-6028 at Old Dominion University, who will be glad to review the matter with you.

VOLUNTARY CONSENT
By signing this form, you are saying several things. You are saying that you have read this form or have had it read to you, that you are satisfied that you understand this form, the research study, and its risks and benefits. The researchers should have answered any questions you may have had about the research. If you have any questions later on, then the researchers should be able to answer them:

Dr. James P. Bliss at (757) 683-4051

If at any time you feel pressured to participate, or if you have any questions about your rights or this form, then you should call Dr. David Swain, the current IRB chair, at (757) 683-6028, or the Old Dominion University Office of Research, at 757-683-3460.

APPENDIX F (continued)

And importantly, by signing below, you are telling the researcher YES, that you agree to participate in this study. The researcher should give you a copy of this form for your records.

------------------------------------	------------------------------------	----------------------
Participant's Name	Participant's Signature	Date

INVESTIGATOR'S STATEMENT
I certify that I have explained to this subject the nature and purpose of this research, including benefits, risks, costs, and any experimental procedures. I have described the rights and protections afforded to human subjects and have done nothing to pressure, coerce, or falsely entice this subject into participating. I am aware of my obligations under state and federal laws, and promise compliance. I have answered the subject's questions and have encouraged him/her to ask additional questions at any time during the course of this study. I have witnessed the above signature(s) on this consent form.

------------------------------------	------------------------------------	----------------------
Investigator's Name	Investigator's Signature	Date

APPENDIX G

Monte Carlo Study Syntax

```
COMPUTE hi = RV.UNIFORM(0,1) .
EXECUTE .

COMPUTE fa = RV.UNIFORM(0,1) .
EXECUTE .

COMPUTE a = .5 + .5*hi - .5*fa .
EXECUTE .

COMPUTE b =  .5*hi + .5*fa .
EXECUTE .

COMPUTE d = IDF.NORMAL(hi,0,1) - IDF.NORMAL(fa,0,1) .
EXECUTE .

COMPUTE c = (-1) * (.5*IDF.NORMAL(hi,0,1) + .5*IDF.NORMAL(fa,0,1)) .
EXECUTE .
```

APPENDIX H

Restriction of *a* Values Syntax

```
USE ALL.
COMPUTE filter_$=(a  >=  .5  & a < .6).
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

GRAPH
 /SCATTERPLOT(BIVAR)=a WITH b
 /MISSING=LISTWISE .

GRAPH
 /SCATTERPLOT(BIVAR)=d WITH c
 /MISSING=LISTWISE .

USE ALL.
COMPUTE filter_$=(a  >=  .6  & a < .7).
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

GRAPH
 /SCATTERPLOT(BIVAR)=a WITH b
 /MISSING=LISTWISE .

GRAPH
 /SCATTERPLOT(BIVAR)=d WITH c
 /MISSING=LISTWISE .

USE ALL.
COMPUTE filter_$=(a  >=  .7  & a < .8).
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

GRAPH
 /SCATTERPLOT(BIVAR)=a WITH b
 /MISSING=LISTWISE .

GRAPH
 /SCATTERPLOT(BIVAR)=d WITH c
 /MISSING=LISTWISE .
```

APPENDIX H (continued)

```
USE ALL.
COMPUTE filter_$=(a  >=  .8  & a < .9).
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

GRAPH
 /SCATTERPLOT(BIVAR)=a WITH b
 /MISSING=LISTWISE .

GRAPH
 /SCATTERPLOT(BIVAR)=d WITH c
 /MISSING=LISTWISE .

USE ALL.
COMPUTE filter_$=(a  >=  .9  & a <=1).
VALUE LABELS filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

GRAPH
 /SCATTERPLOT(BIVAR)=a WITH b
 /MISSING=LISTWISE .

GRAPH
 /SCATTERPLOT(BIVAR)=d WITH c
 /MISSING=LISTWISE .
```

VITA

Ernesto A. Bustamante, Ph.D.
Old Dominion University
Norfolk, VA 23529-0267
Phone: (757) 288-4646
E-mail: ebustama@odu.edu

EDUCATION

2005 - 2007	Ph. D. - Psychology Major - Human Factors Area of Concentration - Quantitative Methods
2003 - 2005	M. S. - Experimental Psychology
1997 - 2001	B. S. - Psychology (Summa Cum Laude) Minor : Management Old Dominion University Psychology Department Norfolk, VA 23529-0267

PROFESSIONAL EXPERIENCE

2007 - Present	Assistant Professor University of Idaho Moscow, ID 83843
2006	Research Associate Smart Information Flow Technologies Minneapolis, MN 55401
2002 - 2003	Consultant University of Los Andes Psychological Research Center Merida, Venezuela

SELECTED PUBLICATIONS

Bustamante, E. A., Bliss, J. P., & Anderson, B. L. (in press). Effects of varying the threshold of alarm systems and workload on human performance. *Ergonomics*.

Bustamante, E. A., Fallon, C. K., Bliss, J. P., Bailey, W. R., & Anderson, B. L. (2005). Pilots' workload, situation awareness, and trust during weather events as a function of time pressure, role assignment, pilots' rank, weather display, and weather system. *International Journal of Applied Aviation Studies*, *5*(2), 347-367.