To the Graduate Council:

I am submitting herewith a thesis written by Robert Estel Ford entitled "A Study of Human Performance Using Different Alarming Systems in A Part-Task Simulation of the Clinch River Breeder Reactor Nuclear Control Room." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Industrial Engineering.

J. C. Hungerford

J. C. Hungerford, Major Professor

We have read this thesis
and recommend its acceptance:

Accepted for the Council:

L. Evans Roth

Vice Chancellor
Graduate Studies and Research

A STUDY OF HUMAN PERFORMANCE USING DIFFERENT ALARMING SYSTEMS IN A PART-TASK SIMULATION OF THE CLINCH RIVER BREEDER REACTOR NUCLEAR CONTROL ROOM

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Robert Estel Ford

March 1982

3057955

ABSTRACT

A set of auditory warning signals was developed using general principles and design recommendations found in Human Factors handbooks and texts. A group of university students were tested to determine how people detect, discriminate, and identify auditory alarms that have their priorities for response coded by frequency, amplitude, and modulation. Scaling techniques and statistical analysis were used to determine which of the signals made the "best" warning alarm and to determine how many alarms could be recognized in a multiple-alarm environment. Design recommendations were made for designing auditory signals for warning alarms.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	The Problem	1
	Background to the Problem	1
	Purpose of Study	3
	Research Questions to be Investigated	3
	Scope and Delimitations of the Study	4
	Need for the Research	4
II.	LITERATURE REVIEW	6
	Organization of the Literature Search	6
	Material Acquisition	6
	Human Factors Auditory Alarm Guidelines	6
	Multidimensional Coding of Alarms	12
	Psychoacoustic and Psychophysical Techniques	17
	Paired Comparisons	20
	Integration of Literature	23
III.	DESCRIPTION OF THE RESEARCH	25
	Introduction	25
	Methodology for Paired-Comparison Experiment	26
	Methodology for Multi-Signal Recognition Experiment	26
IV.	EXPERIMENT I. DETERMINATION OF HUMAN PREFERENCES FOR AUDITORY ALARM SIGNALS	29
	Introduction	29
	Experimental Design	30
	Subjects	34
	Subject Instructions	34
	Instrumentation	36
	Experimental Setting	37
	Data Collection	37
	Data Analysis	40
	Results of Experiment I	41
V.	EXPERIMENT II. HUMAN PERFORMANCE IN A SIMULTANEOUS MULTIPLE-ALARM ENVIRONMENT	47
	Introduction	47
	Experimental Design	48
	Subject Selection	49
	Subject Instructions	51
	Instrumentation	52
	Experimental Setting	52
	Data Collection	54
	Data Analysis	54
	Results of Experiment II	56

CHAPTER PAGE

VI. SUMMARY AND CONCLUSIONS 82
Summary 82
Conclusions 84

VII. RECOMMENDATIONS 88
Design Recommendations. 88
Research Recommendations. 89

BIBLIOGRAPHY. 91

APPENDIXES. 94
A. PLOTS AND ANOVA FOR RESPONSE TIME. 95
B. PLOTS AND ANOVA FOR NUMBER OF RESPONSES. 104
C. PLOTS AND ANOVA FOR NUMBER OF INCORRECT RESPONSES. . . . 113
D. HYPOTHETICAL IMPLEMENTATION OF RESEARCH FINDINGS 122

VITA. 132

LIST OF TABLES

TABLE		PAGE
II-1.	Summary of Certain Auditory Coding Methods.	11
II-2.	The Characteristics and Features of Certain Types of Audio Alarms.	16
IV-1.	Physical Parameters for the Alarms Used in Experiment I .	35
IV-2.	Comparison of Rankings.	45
V-1.	Conversion of Alarm Names from Experiment I to Experiment II	49
V-2.	Randomized Schedule of Alarms	50
V-3.	Means for Maximum and Minimum Response Times for Category 1.	57
V-4.	Means for Maximum and Minimum Response Times for Category 2.	59
V-5.	Means for Maximum and Minimum Response Times for Category 3.	59
V-6.	Means for Maximum and Minimum Response Times for Category 4.	60
V-7.	Grand Means for Response Times.	61
V-8.	Grand Means for Number of Responses	63
V-9.	Results of Normality Tests for Each Category on the Number of Responses	64
V-10.	Means for Number of Responses for Category 1.	65
V-11.	Means for Number of Responses for Category 2.	65
V-12.	Means for Number of Responses for Category 3.	66
V-13.	Means for Number of Responses for Category 4.	66
V-14.	Grand Means for Number of Incorrect Responses	68
V-15.	Results of Normality Tests for Each Category on the Number of Incorrect Responses	69

TABLE PAGE

V-16. Means for Number of Incorrect Responses for Category 1. . 69

V-17. Means for Number of Incorrect Responses for Category 2. . 70

V-18. Means for Number of Incorrect Responses for Category 3. . 70

V-19. Means for Number of Incorrect Responses for Category 4. . 71

V-20. Comparison of 1 Alarm System for All Performance Measures. 77

V-21. Comparison of 2 Alarm Systems for All Performance Measures. 78

V-22. Comparison of 3 Alarm Systems for All Performance Measures. 79

V-23. Comparison of 4 Alarm Systems for All Performance Measures. 80

LIST OF FIGURES

FIGURE		PAGE
II-1.	The Just-Noticeable Differences (JNDs) in Sound Intensity for Pure Tones of Selected Frequencies and for Wide-Band Noise.	10
II-2.	Just-Noticeable Differences (JNDs) for Pure Tones at Various Levels Above Threshold.	10
II-3.	Discriminal Process Distributions for Stimuli j and k . .	22
II-4.	Distribution of Pair Differences.	22
III-1.	Overall Research Strategy	27
IV-1.	Research Design	33
IV-2.	Real-time Experimental Design	38
IV-3.	Laboratory Arrangement--Experiment I.	39
IV-4.	Results from the Analysis of Paired-Comparisons of Signal Alternatives (N=38).	42
IV-5.	Ranking of Alarms by Response Time.	44
IV-6.	Physical Characteristics of the Selected Alarms	46
V-1.	Laboratory Arrangement--Experiment II	53
V-2.	Data Analysis Flow Diagram.	55
V-3.	Bar Chart of Grand Means for Response Time.	61
V-4.	Bar Chart of Grand Means for the Number of Responses. . .	63
V-5.	Bar Chart of Grand Means for the Number of Incorrect Responses	68
V-6.	Bar Chart of Response Time, Number of Responses, and Number of Incorrect Responses for Category 1.	73
V-7.	Bar Chart of Response Time, Number of Responses, and Number of Incorrect Responses for Category 2.	74
V-8.	Bar Chart of Response Time, Number of Responses, and Number of Incorrect Responses for Category 3.	75

FIGURE PAGE

V-9. Bar Chart of Response Time, Number of Responses, and Number of Incorrect Responses for Category 4. 76

VI-1. Bar Chart of Grand Means for Response Time, Number of Responses, and Number of Incorrect Responses. 85

VI-2. Flow Diagram of Subject Response. 86

A-1. Means for Maximum and Minimum Response Times for Category 1. 96

A-2. Means for Maximum and Minimum Response Times for Category 2. 97

A-3. Means for Maximum and Minimum Response Times for Category 3. 98

A-4. Means for Maximum and Minimum Response Times for Category 4. 99

A-5. ANOVA for Category 1 Response Time. 100

A-6. ANOVA for Category 2 Response Time. 101

A-7. ANOVA for Category 3 Response Time. 102

A-8. ANOVA for Category 4 Response Time. 103

B-1. ANOVA for Category 1 Number of Responses. 105

B-2. ANOVA for Category 2 Number of Responses. 106

B-3. ANOVA for Category 3 Number of Responses. 107

B-4. ANOVA for Category 4 Number of Responses. 108

B-5. Means for Number of Responses for Category 1. 109

B-6. Means for Number of Responses for Category 2. 110

B-7. Means for Number of Responses for Category 3. 111

B-8. Means for Number of Responses for Category 4. 112

C-1. Means for Number of Incorrect Responses for Category 1. . 114

C-2. Means for Number of Incorrect Responses for Category 2. . 115

C-3. Means for Number of Incorrect Responses for Category 3. . 116

FIGURE PAGE

C-4. Means for Number of Incorrect Responses for Category 4. . 117

C-5. ANOVA for Category 1 Number of Incorrect Responses. . . . 118

C-6. ANOVA for Category 2 Number of Incorrect Responses. . . . 119

C-7. ANOVA for Category 3 Number of Incorrect Responses. . . . 120

C-8. ANOVA for Category 4 Number of Incorrect Responses. . . . 121

D-1. Alarm System 1. 130

D-2. Alarm System 2. 130

D-3. Alarm System 3. 131

D-4. Recommended Design. 131

CHAPTER I

INTRODUCTION

The Problem

The safe operation of a nuclear power plant depends on the interaction of all of the plant's systems and components, including the operating crew. In the event of a transient condition or accident, the operators are to prevent, or limit, the development of hazardous conditions and equipment damage. Their ability to accomplish this depends on their knowing what events are occurring and reacting accordingly.

One of the communication channels used to inform the operators of discrepant conditions in a plant system is auditory alarm signals. These auditory alarms alert the operators and then direct them to the appropriate area of the control console where a visual search of the panel must be made to identify the discrepant condition. Since auditory alarms currently in use do not give an indication of priority or significance, valuable time may be lost, when simultaneous or near-simultaneous alarms occur, in tracking down the most urgent deviate condition for which an alarm is sounding. A non-safety related alarm may achieve priority over an alarm indicating a hazardous condition simply because an operator responded to the non-safety related alarm first. Non-prioritized multiple alarms may cause operator confusion and errors.

Background to the Problem

A nuclear control room is a complex man-machine system. It is common to see hundreds of annunciators, gauges, switches, and controls

in a nuclear control room. The information processing requirements placed upon an operator are demanding during a plant event. The operator must detect and recognize complex patterns of information in order to correctly diagnose events and prescribe corrective actions to restore the plant to a normal operating condition. Occasionally, information coming to the operator is spurious, as was the case with the Three Mile Island accident. The operator's task is even more complicated and stressful in situations where the available information is uncertain.

Historically, nuclear control rooms have often been designed and built by several different subcontractors. The use of human factors techniques have been nonexistent in their design. Consequently, the main control panel (MCP) and back panels lack both integration and standardization. These design shortcomings have made the task of the nuclear control room operator extremely difficult. The usual way in which poor designs have been compensated is by extensive operator training.

Recently, two major related influences have caused designers to re-think their way through the design of nuclear control rooms. First, the accident at Three Mile Island was determined to be a human factors problem. Secondly, the Nuclear Regulatory Commission (NRC) in response to problems at Three Mile Island and other nuclear power plants has required that human factors techniques be implemented into the design and re-design process. Presently, the only new nuclear power under design in the United States is the Clinch River Breeder Reactor Plant at Oak Ridge, Tennessee. The Westinghouse Electric Corporation is the lead reactor designer and manufacturer in the project. The design of

the plant has been continually refined and updated to take advantage of technical improvements and scientific advances. These improvements include the lessons learned from the accident at Three Mile Island in 1979. Extensive use of human factors data and research techniques have been incorporated into the design process. A considerable amount of attention has been given to the standardization and integration of displays and coding on the MCP in order to ease the information processing load of the operator.

In September, 1980 the Industrial Engineering Department of The University of Tennessee began a research project to study the use of coded auditory alarm systems such as might be used in nuclear reactor control rooms. The study was sponsored by the Advanced Power Systems Division of Westinghouse Electric Corporation.

Purpose of Study

The purpose of this study was to determine the capability of people to detect, discriminate, and identify auditory alarms that have their priorities for response coded by changing the dimensions of frequency, amplitude, and modulation. Representative values of these three dimensions were selected for the coded auditory alarms used in this study. A related objective of the study was to develop guidelines and recommendations concerning the coding of auditory alarming systems for use in a nuclear control room.

Research Questions to be Investigated

Several broad research questions are relevant to this research:

1. How do people prioritize alarms that have different frequencies, amplitudes, and modulations?
2. What effect does alarm frequency, amplitude, and modulation have on alarm response time in a paired comparison task?
3. How do people perform in an auditory relative discrimination task when presented with several alarms simultaneously; that is, how does the presence of multiple alarms in a recognition task affect errors, response time, and variability?

Scope and Delimitations of the Study

The study does not evaluate the responses of nuclear control room operators to alarms in a nuclear reactor control room. Rather, the study evaluates the responses of college students in auditory relative discrimination tasks in a laboratory setting. The setting is not an audiometric testing room (controlled environmental chamber) but one that more clearly approximates the ambient conditions of a main control room. To accomplish the purposes of this study a part-task simulation of the auditory environment of a nuclear reactor control room is conducted using real-time computer methodology.

Need for the Research

Human factors handbook information exists that gives broad guidelines for when to use auditory displays and when to use visual displays. Guidelines also exist that recommend general characteristics for a single auditory alarm. Most of these guidelines are in the form of "heuristics" or "rules of thumb." Very little information exists to guide the designer of control rooms in the coding and selection of

auditory alarms for use in a multi-alarm environment. There is virtually no information about how people perform when presented with multiple warning alarms simultaneously.

CHAPTER II

LITERATURE REVIEW

Organization of the Literature Search

The literature search was begun by developing a three-step approach for locating information and by defining the areas of interest. The three steps were:

1. Interviews with professionals in the fields of Human Factors and Psychophysics.
2. Locate and review sources and topics suggested by the interviewees.
3. Locate and review sources and references given in the material located in step 2.

The areas of interest were determined to be guidelines for alarming systems, psychophysics and psychoacoustics, and multidimensional coding of alarms.

Material Acquisition

Material was acquired by doing a computer search of psychological abstracts published from 1965 to 1981, reviewing Human Factors handbooks and texts, reviewing the Human Factors Journals and Ergonomics that are maintained in the Library at The University of Tennessee. A microfiche search was also conducted at the Idaho National Engineering Laboratory Technical Library at Idaho Falls, Idaho.

Human Factors Auditory Alarm Guidelines

The way the human auditory sense functions gives the designer of auditory displays several advantages for the presentation of information.

Comparisons made between the auditory and the visual modality along with other cues from auditory studies have defined certain circumstances in which an auditory display is preferable to a visual display. Some circumstances listed in McCormick (1976) are:

1. When the origin of the signal is itself a sound
2. When the message is simple and short
3. When the message will not be referred to later
4. When the message deals with events in time
5. When sending warnings or when the message calls for immediate action
6. When presenting continuously changing information
7. When the visual system is overburdened
8. When the speech channels are fully employed
9. When illumination limits the use of vision
10. When the receiver moves from one place to another.

The use of such a set of guidelines must be used carefully and not followed blindly so the system transmits maximum information rather than adding to any existing confusion. Particular mention should be made of the desirability of keeping the message short and simple (except in the case of speech), because people do not do well in short-term storage of complex messages.

There are three types of human functions involved in the recognition of auditory signals. These depend on the nature of the signal and are as follows: (1) detection which is determining if a given signal is present or not, such as a warning signal; (2) relative discrimination or differentiating between two or more signals when presented together or close together; and (3) absolute identification

which is the identifying of a particular signal of some class, when only one is presented. Relative discrimination and absolute identification can be made on the basis of several stimulus dimensions, such as intensity, frequency, duration, and direction.

The detection of an auditory signal can be examined in the context of signal detection theory. Signal detection theory indicates the frequency distribution of the signal plus noise should not overlap the distribution of the noise itself. If there is overlap, the signal cannot always be detected from the noise. This confusion is partially a function of masking, that is, when the signal occurs in the presence of noise, the threshold of detectability for the signal is elevated. This higher threshold must be exceeded by the signal if it is to be accurately detected.

The minimum sound level of a signal or tone necessary for detection 50 percent of the time when no noise is present is the absolute auditory threshold. Another threshold, the masked threshold, refers to the minimum sound level of a signal necessary for 50 percent when the signal is masked in a noise field (Hutchingson, 1981). Masking refers to the difference between the absolute threshold and the masked threshold at any given frequency. The degree of masking is stated in terms of a signal-to-noise ratio. This is simply the algebraic difference in the signal intensity and noise intensity at threshold. Wideband noise masks signals more often than pure tones do. For each increase in noise, the signal must be increased an equal number of decibels in order to be detected. This increase in masking is linear and applies to frequencies of 20 Hz and above.

A sound that is approximately 50 dB above absolute threshold is sufficiently above the ambient background noise to be detected in quiet surroundings (McCormick, 1976).

In rather noisy conditions, the signal intensity must be at a level far enough above the noise to insure that it can be detected. This point should be well above the masked threshold of the signal in the presence of the noise and below a safe level of 110 dB to prevent damage to the ear (Huchingson, 1981). A more realistic safe level is probably 85 to 90 dB in light of today's safety standards.

The relative discrimination of auditory signals on the basis of the most commonly used dimensions of intensity and frequency depend on the interaction of these two dimensions. Figure II-1 gives some examples of human ability to discriminate between differences in intensity.

This figure shows the just-noticeable differences (JNDs) for some pure tones and wide-band noise of various sound-pressure levels. The smaller JNDs that can be detected are for intensities that are 60 dB above absolute threshold. For intermediate frequencies (1000 and 7000 Hz), the chart shows that the JNDs of signals above 60 dB are minimal. This is important for auditory signals that are to be identified by intensity.

The ability of people to tell the difference between pure tones of different frequencies is given by the results of an early study by Shower and Biddulph (1931) as shown in Figure II-2. This figure shows the JNDs for pure tones of different frequencies at levels above threshold. The JNDs are smaller for frequencies of 1000 Hz and below, especially at high intensities, but increase sharply for frequencies above 1000 Hz. Therefore, if frequency is to be the basis for signal

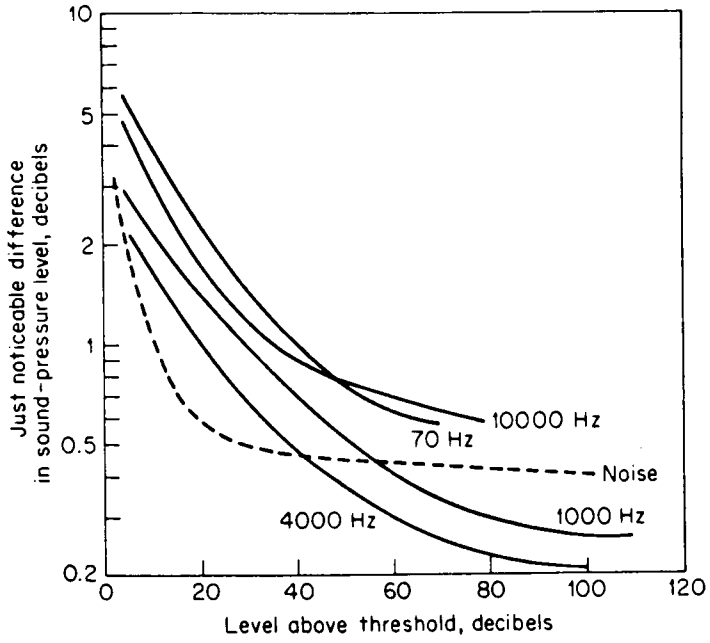


Figure II-1. The Just-Noticeable Differences (JNDs) in Sound Intensity for Pure Tones of Selected Frequencies and for Wide-Band Noise. (From Deatherage, 1972)

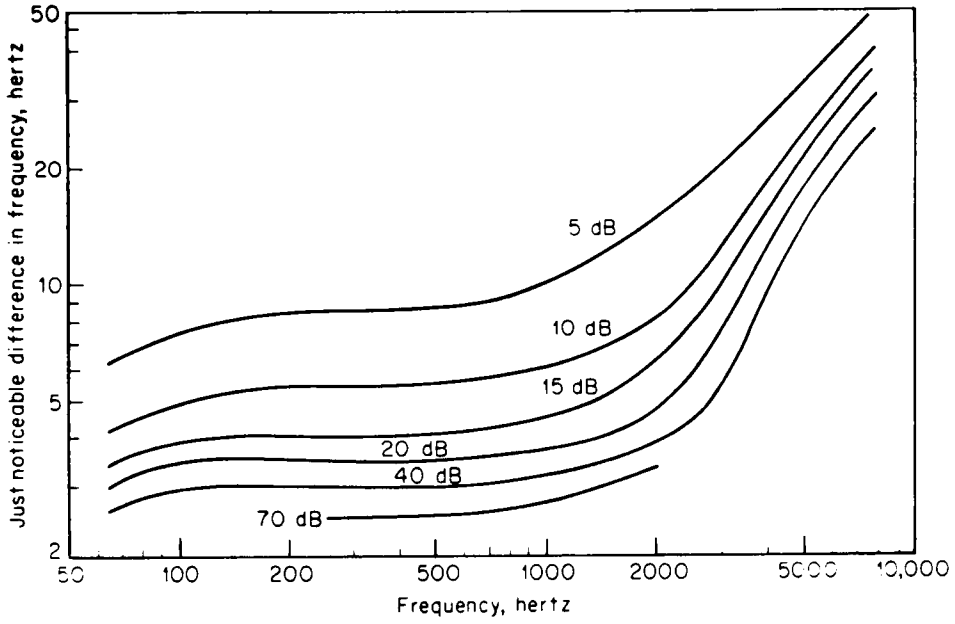


Figure II-2. Just-Noticeable Differences (JNDs) for Pure Tones at Various Levels Above Threshold. (From Shower and Biddulph, 1931)

discrimination it is best to use signals of lower frequencies. However, if there is much ambient noise which is usually made up of lower frequencies, this noise might mask the signals. The figure also shows that the JNDs are smaller for signals of high intensity than for those of low intensity. This suggests that signals probably should be 30 dB above absolute threshold as a minimum (Deatherage, 1972).

The JNDs for a given signal dimension reflects the minimum difference that can be discriminated on a relative basis. But in many cases it is necessary to make an absolute identification of an individual signal. The number of "levels" along a dimension that can be identified is usually quite small. Table II-1 gives some indication of the number of levels that can be identified for some auditory dimensions.

Table II-1. Summary of Certain Auditory Coding Methods

Intensity (pure tones)	30 to 5 levels. Intensity changes easier to detect than single intensities. For pure tones restrict to 1000 to 4000 Hz, but preferably use wide-band noise.
Frequency	Variously estimated at from 4 to 7 levels. Preferably use fewer and space widely apart. Set intensity at least 30 dB above threshold.
Sound duration	Use clear-cut differences, preferably 2 or 3.
Sound direction	Difference in intensity between two ears should be distinct: particularly relevant for directional information (i.e., right versus left).
Intensity and frequency	9 levels.

Source: Deatherage, 1972.

Multidimensional Coding of Alarms

If discriminations are to be made among numerous signals, it is possible to use a multidimensional code system. The study by Pollack and Ficks (1954) showed that various combinations of several dimensions, such as direction, frequency, intensity, repetition rate, on- and off-time, and duration can be used. The use of two or more of these dimensions in combination tends to improve the transfer of information to humans.

Frequency, amplitude, duration, and waveform are the physical parameters of sound. The phase relationships among the components of a complex sound are also of some importance. Each of these physical dimensions has some subjective aspect.

Frequency is expressed in hertzes (Hz, or cycles per second). A pure tone will have a simple sinusoidal waveform, and complex sounds can be analyzed into the set of sinusoids that make them up. These sinusoids are usually expressed mathematically as

$$a_1 \ \mathrm{Sin}\ 2\pi f_1 t + a_2 \ \mathrm{Sin}\ 2\pi f_2 t + \ldots$$

The constants a_1 represent maximum amplitude of particle displacement, and f_1 is frequency. Nothing in this expression denotes duration, but it is necessary to calculate instantaneous amplitude. Neither is there a factor for phase, though there can be. Phase will be discussed later in this section.

Since sounds are pressure waves varying around the normal air pressure, the amplitude (intensity) of these pressure waves is usually expressed in newtons/m^2 or in watts/m^2. Because man can hear sounds over a wide range of sound pressures, it is customary to describe these

quantities by use of logarithmic scales known as sound pressure levels. The most common unit of sound pressure level is the decibel (dB).

The waveform of complex sounds can be either periodic or aperiodic. Periodic sounds tend to be tonal or musical, while aperiodic sounds are impulsive, like clicks or booms, or noisy like the hiss of an air jet.

Waveform is specified in terms of the various frequency components that comprise the sound. Thus, sounds having a precisely periodic waveform are composed of a series of single-frequency components located at discrete frequencies and integrally related. Aperiodic sounds, similarly, might consist of discrete components, or they might be composed of a continuum of frequency components distributed over a broad band; the former are said to have a line spectrum, the latter a continuous spectrum.

Duration is how long a sound lasts in time before an interruption occurs. How long an interruption lasts plus the duration of the sound is the on-and-off time. How many times this on-and-off time cycle is repeated in some unit of time is the repetition rate.

Phase relations among components of periodic complex sounds are useful for identifying sound sources. The utility-of-phase differences seem to lie in the listener's ability to use them in identifying a source of sound.

Complex sounds that are the result of the combination of two or more tones may be altered. How they are altered will depend upon the phase relationships of the components. Loudness will increase or decrease while the pitch remains constant. They can also cancel so that

none can be heard. If the frequencies differ, beats will occur; that is, the loudness rises and falls in a repetitive cycle.

With the principal characteristics and special features of different types of auditory alarm signals thus discussed, the following paragraphs summarize and group these features into guidelines and design recommendations.

Given below are a few guidelines from McCormick (1976) based on research findings and experience from many sources:

I. General principles
   - A. Compatibility: Where feasible, the selection of signal dimensions and their encoding should exploit learned or natural relationships of the users, such as high frequencies being associated with up or high, and wailing signals with emergency.
   - B. Approximation: Two-stage signals should be considered when complex information is to be presented, these stages to consist of:
     1. Attention-demanding signal: to attract attention and identify a general category of information.
     2. Designation signal: to follow the attention-demanding signal and designate the precise information within the general class indicated above.
   - C. Dissociability: Auditory signals should be easily discernible from any ongoing audio input (be it either meaningful input or noise). For example, a warning bell would not stand out if the operator happened to be a Swiss bell ringer. If a person is to listen concurrently to two or more channels, the frequencies of the channels should be different if it is possible to make them so.
   - D. Parsimony: Input signal to the operator should not provide more information than is necessary for him.
   - E. Invariance: The same signal should designate the same information at all times.

II. Principles of presentation
   - F. Avoid extremes of auditory dimensions.
   - G. Establish intensity relative to ambient noise level.
   - H. Use interrupted or variable signals: Where feasible, avoid steady-state signals and, rather, use interrupted or variable signals. This will tend to minimize perceptual adaptation.

III. Principles of installation of auditory displays
   I. Test signals to be used: Such tests should be made with a representative sample of the potential user population, to be sure the signals can be detected by them.
   J. Avoid conflict with previously used signals: Any newly installed signals should not be contradictory in meaning to any somewhat similar signals used in existing or earlier systems.
   K. Facilitate changeover from previous display: Where auditory signals are to replace some other mode of presentation (such as visual), preferably continue both modes for a while, to help people become accustomed to new auditory signals.

The unique features of the auditory system lead auditory displays to special use for signalling warnings and alarms. A summary of such characteristics and features is given in Table II-2. In the selection or design of warning and alarm signals, the following design recommendations have been proposed by Deatherage (1972) and by Mudd (1961), these being in slightly modified form.

- Use frequencies between 200 and 5000 Hz, and preferably between 500 and 3000 Hz, because the ear is most sensitive to this middle range.
- Use frequencies below 1000 Hz when signals have to travel long distances (over 1000 ft), because high frequencies do not travel as far.
- Use frequencies below 500 Hz when signals have to "bend around" major obstacles or pass through partitions.
- Use a modulated signal (1 to 8 beeps per second or warbling sounds varying from 1 to 3 times per second), since it is different enough from normal sounds to demand attention.
- Use signals with frequencies different from those that dominate any background noise, to minimize masking.
- Preferably use high-intensity sudden-onset signals to alert receiver. Where earphones are used, consider dichotic presentation (alternating signal from one ear to the other).
- If different warning signals are used to represent different conditions, each should be discernible from the others.
- Where feasible, use a separate communication system for warnings, such as loudspeakers, horns, or other devices not used for other purposes.

Table II-2. The Characteristics and Features of Certain Types of Audio Alarms.

Alarm	Intensity	Frequency	Attention-getting ability	Noise-penetration ability
Diaphone (foghorn)	Very high	Very low	Good	Poor in low-frequency noise
Horn	High	Low to high	Good	Good
Whistle	High	Low to high	Good if intermittent	Good if frequency is properly chosen
Siren	High	Low to high	Very good if pitch rises and falls	Very good with rising and falling frequency
Bell	Medium	Medium to high	Good	Good in low-frequency noise
Buzzer	Low to medium	Low to medium	Good	Fair if spectrum is suited to background noise
Chimes and gong	Low to medium	Low to medium	Fair	Fair if spectrum is suited to background noise
Oscillator	Low to high	Medium to high	Good if intermittent	Good if frequency is properly chosen

Source: Deatherage, 1972.

In addition, channel frequencies should be spaced as widely as possible throughout the available spectrum to minimize signals from being masked from one channel by signals from another and to make channel identification easier. Also, the design should avoid simple multiples or submultiples of other frequencies because they are more easily confused.

As more parameters are used to code signals, the number of signals that can be identified becomes very large. The hundreds of voices and thousands of sounds a person hears and recognizes every day are examples of this. Experiments performed by Pollack and Ficks (1954) indicate that it is preferable to use more dimensions with a few steps in each dimension than to use a few dimensions with many steps.

Psychoacoustic and Psychophysical Techniques

Snodgrass (1975) defines psychophysics as:

> . . . a group of methodologies, developed and refined over the last one hundred years, for measuring the effect of some aspect of a stimulus on some aspect of an observer's response. The word psychophysics owes its derivation to the attempt to measure the relation between the psychological and the physical. Traditionally, psychophysics has been concerned with stimuli for which relatively unambiguous physical measures exist, such as lights of given intensities or tones of fixed frequencies. Certain well-defined responses to these stimuli are measured to determine whether the subject detects the presence of the stimulus, knows what the stimulus is, can determine whether it differs from a second stimulus, or can judge how much it differs. The first question, whether the stimulus is there, is the problem of detection; the second question, what the stimulus is, is the problem of recognition; the third question, whether the stimulus is different from a second stimulus, is the problem of discrimination; and the fourth question, how different this stimulus is from that one, is the problem of scaling. Psychophysical measurements are often summarized in a function that relates psychological magnitudes to physical magnitudes of stimuli--the psychophysical function. Psychological magnitude is the subjective intensity of sensation evoked by a stimulus.

Detection, the first of the methodologies defined above, is of little interest in this project. This is because the subjects in the experiments will know that the alarms used have been designed to be easily detected.

Therefore, what is of interest is discrimination, recognition and scaling. It is also desired to learn something (Snodgrass, 1975) of the learned or natural relationships the subjects may have for an alarm system.

In a discrimination experiment, a subject is asked to determine whether one stimulus is different from another. The response may be an equality judgement; that is, he is to say whether one stimulus is greater than, equal to, or less than another stimulus. In other experiments an equality judgement may not be required--only that the two are different. These two cases are absolute discrimination and relative discrimination.

In a recognition experiment, one or more of a fixed number of stimuli are presented to the subject whose task is to identify the stimuli presented. His response may be a number, a name, or even a key-press. Recognition is classified as nominal scaling because even when numbers are used as responses, the numbers serve only to identify the stimuli and do not necessarily reflect ordinal, interval, or ratio characteristics among the stimuli with respect to some property.

In a recognition experiment, the stimuli are usually designed to be easily discriminable from one another when presented simultaneously. Because recognition stimuli are presented successively, the subject must compare each stimulus to some internal standard developed through experience or training with the stimuli. The ability

to do this is the subject of study in recognition experiments.

An important question is what is the number of different stimuli that can be correctly identified as a function of the number of dimensions along which they vary and their relative spacing along those dimensions. It has been determined that the number of unidimensional stimuli that can be identified without error is of the order of seven, plus or minus two, regardless of the dimension involved (Miller, 1956). It is somewhat surprising that this number is so small because people daily recognize a larger number of faces, words, etc. However faces and words and other stimuli in daily life vary on an undetermined but large number of dimensions. The question in recognition experiments is what happens to the number of stimuli that can be recognized when they vary on several dimensions?

Past experiments have shown that subjects can recognize about six tones differing in frequency, and about five tones differing only in loudness. If the subjects were given tones differing in both frequency and loudness and in all possible combinations, it might be expected that the subjects could recognize thirty tones. However Pollack (1952, 1953) found that subjects can recognize only about nine tones without error. The recognition of stimuli that vary on two dimensions simultaneously is an improvement over stimuli of a single dimension but the improvement is not as great as one might expect. The limit in recognition appears to be set by the subject's internal standard and not by the physical characteristics of the stimuli to be recognized.

Pols (1976) states that if absolute identification is required, the number of signal categories should be limited to about four because listeners cannot identify correctly more than a few different intensities, pitches, or interruption rates.

Paired Comparisons

Stimuli can be placed on an ordinal scale if subjects can rank-order them with respect to some attribute. If a subject is to rank a set of n stimuli with respect to some property, for instance sound intensity or a preferred color, he may be given a subset of size k from the original set of n and be asked to rank within the subset. The rank of 1 will be given to the stimulus with the most of the attribute being evaluated, and the rank of k to the stimulus with the least. When $k = 2$, the method is called paired comparisons, and when $k = n$, the method is called ranking.

When $k < n$, the rank of a given stimulus B is obtained by ranking the frequency with which it is chosen over all of the other stimuli. Each stimulus is compared at least once with every other stimulus. The number of sets which must be presented so that each stimulus is compared once with every other stimulus is given by the formula for combinations

$$\binom{n}{k} = \frac{n!}{k!\ (n-k)!}$$

The method of ranking is much faster than the method of paired comparisons. However, the method of paired comparisons has two advantages over the method of ranking. First, the subjects' task is simpler, and second, transitivity of judgements can be tested. Transitivity is simply if one attribute is preferred to a second attribute, and the second is preferred to a third, then the first must be preferred to the

third. When the number of attributes being compared is large, it is rare if the results show perfect transitivity. However this lack of transitivity can be overcome in deriving scale values for the stimuli if Thurstonian scaling is used.

In Thurstonian scaling, it is assumed that each time a stimulus is presented, it is represented by a point along a psychological scale that is determined by an underlying discrimination process which causes a person to react to the stimulus. This discriminal process may be different from one moment to the next because of an inherent variability within a person. The frequency distribution that is produced by this process is assumed to be normal. Therefore when a pair of stimuli are presented to a person, two points are produced along the psychological scale and the differences of normally distributed random variables are also normally distributed. Figure II-3 shows the distributions of the responses given for a pair of stimuli. Thurstone assumed that a person makes a decision about a pair of stimuli based on the value difference of the pair. The distribution of these differences between the two stimuli is shown in Figure II-4. This distribution has a mean = $s_j - s_k$ and a standard deviation = $\sqrt{\sigma_j^2 + \sigma_k^2 - 2r\sigma_j\sigma_k}$. Because it is assumed that $r=0$, and $\sigma_j = \sigma_k = \sigma$, then the standard deviation of the difference distribution is $\sqrt{2}\,\sigma$.

After all pairs of n objects have been presented to the subjects on a comparison task, a dominance matrix is determined which shows the proportion of times j is preferred over k. These proportions are then used to determine an average Z-value. This Z-value can then be used to order the preferences on an interval scale. A high Z-value indicates a

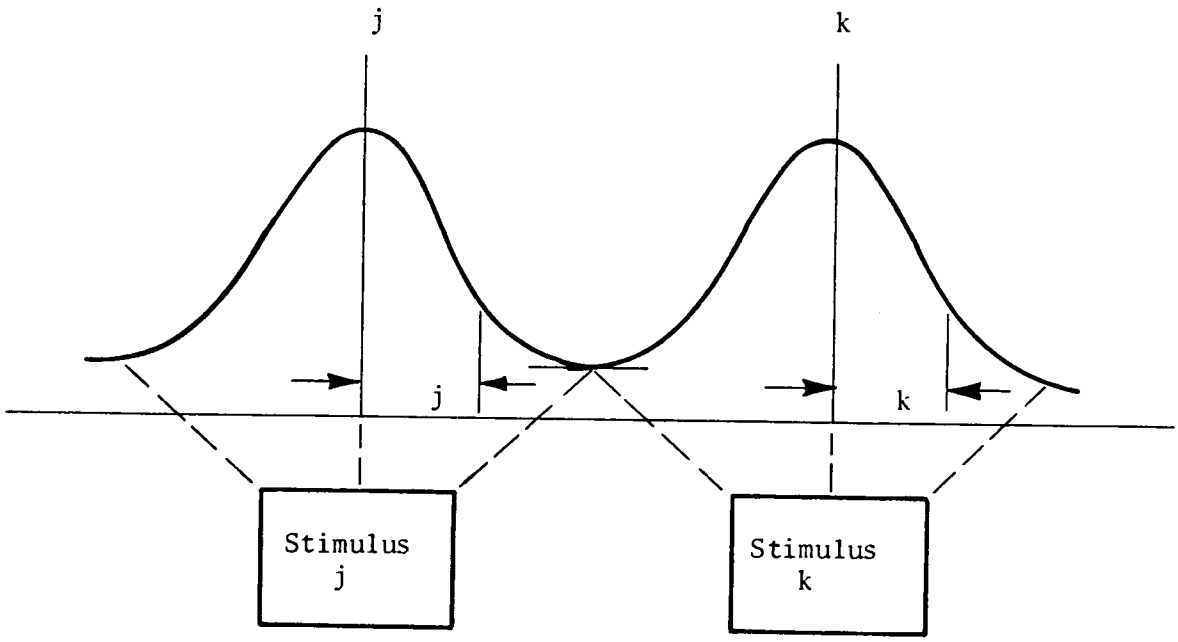


Figure II-3. Discriminal Process Distributions for Stimuli j and k.

Source: Hungerford, 1981.

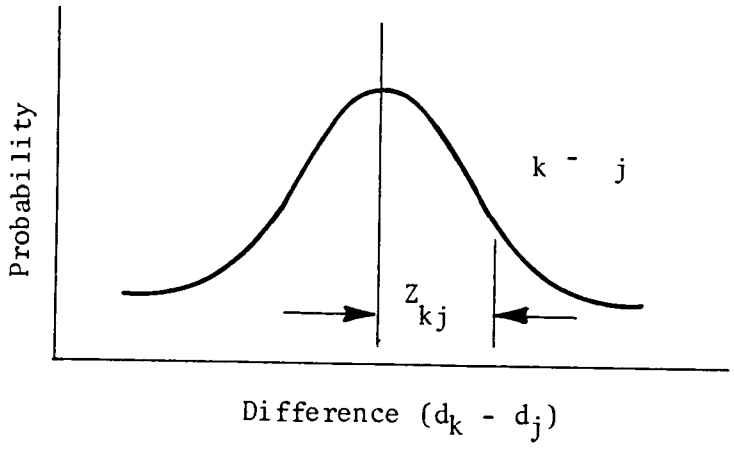


Figure II-4. Distribution of Pair Differences.

Source: Hungerford, 1981.

highly preferred attribute whereas a low Z-value near zero indicates an attribute that was not preferred.

Not only can one determine which response a subject makes to a stimulus, but one can also measure how long it takes him to make that response. The measure of response time is referred to as choice reaction time. This reaction time can also be used to measure subjective stimulus differences. Preference studies using paired-comparison techniques, which also recorded the choice reaction time, consistently found that decision times were longer as distance (defined as the number of ranks separating the two stimuli) decrease; and for a given distance, decision time is longer between the lower-ranked or less-preferred pairs.

For example, Shipley, Coffin, and Hadsell (1945) showed pairs of colors to forty female undergraduates who indicated their preference for a color by pressing a key associated with that color. Six colors were presented in all possible pairs. The relation between reaction time and the ordinal scale distance was linear and negative. In addition, the high reaction times tended to be associated with the low-ranking pairs.

Integration of Literature

The literature has furnished enough information to design an alarm system. It also has given the methodologies to measure the effects an alarm system would have on a person. What has not been found is enough information on human preferences on alarming signals or how many simultaneously sounding alarms a human can recognize.

Therefore two psychophysical experiments have been designed. The first one is to determine if any human preference exists for an alarming system designed using the guidelines. The results of the first experiment define the alarms to be used in the second experiment, which will define the number of alarms that can be correctly recognized in a multiple-alarm situation.

CHAPTER III

DESCRIPTION OF THE RESEARCH

Introduction

After the literature had been collected and reviewed, an overall research strategy was developed to meet the purpose of this research project. The guidelines suggested in the literature were used to design a set of auditory alarm signals. Two experiments were designed to present these alarm signals. The first experiment is a relative discrimination task to measure the subject's preference for one alarm over another. The second experiment is an absolute discrimination task and a recognition task to measure the subject's ability to tell the difference between the alarms and to identify them once the distinction had been made.

In the first experiment the method of paired comparisons was chosen to measure how the subjects prioritized these alarms and to determine how varying the values of the dimensions of frequency, amplitude, and modulation effect the response time to these alarms. This method was chosen because it makes the subject's task simpler and the transitivity of judgements can be evaluated. That is, if signal 1 is judged to be a better alarm than signal 2, and signal 2 is judged to be a better alarm than signal 3, then signal 1 must be a better alarm than signal 3. Also, the more alike two signals are in a paired comparison task, the longer the reaction time is in choosing one of them as the best alarm.

A recognition task was chosen to evaluate how the subjects' performance in identifying alarm signals by name degraded as the number of alarms sounding simultaneously was increased. A recognition task was chosen because the results tell how accurately the subjects can absolutely and relatively identify a set of signals. This overall strategy is outlined in Figure III-1.

Methodology for Paired-Comparison Experiment

The method of paired comparisons was used to determine which signals from the candidate signals were rated by the subjects as the best alarms. Each signal was compared twice with every other signal. The number of times a certain signal was preferred over another signal was tabulated and a probability dominance matrix was developed using Thurstone's Law of Comparative Judgement (Baird and Noma, 1978). This probability was then converted to a scale value and placed on an ordered scale and the highest ranked of the twelve candidates were chosen for the second experiment.

Dependent variables in this experiment were the response and the response time. The independent variables were the dimensions of the alarms, which were tonal frequency, intensity, and modulation.

Methodology for Multi-Signal Recognition Experiment

The second experiment was a recognition task or absolute discrimination task where subjects selected by the same criteria as in the paired comparison experiment. Here all possible combinations of the signals chosen, taken one at a time, two at a time, three at a time, and four at a time were presented to the subjects in a random order.

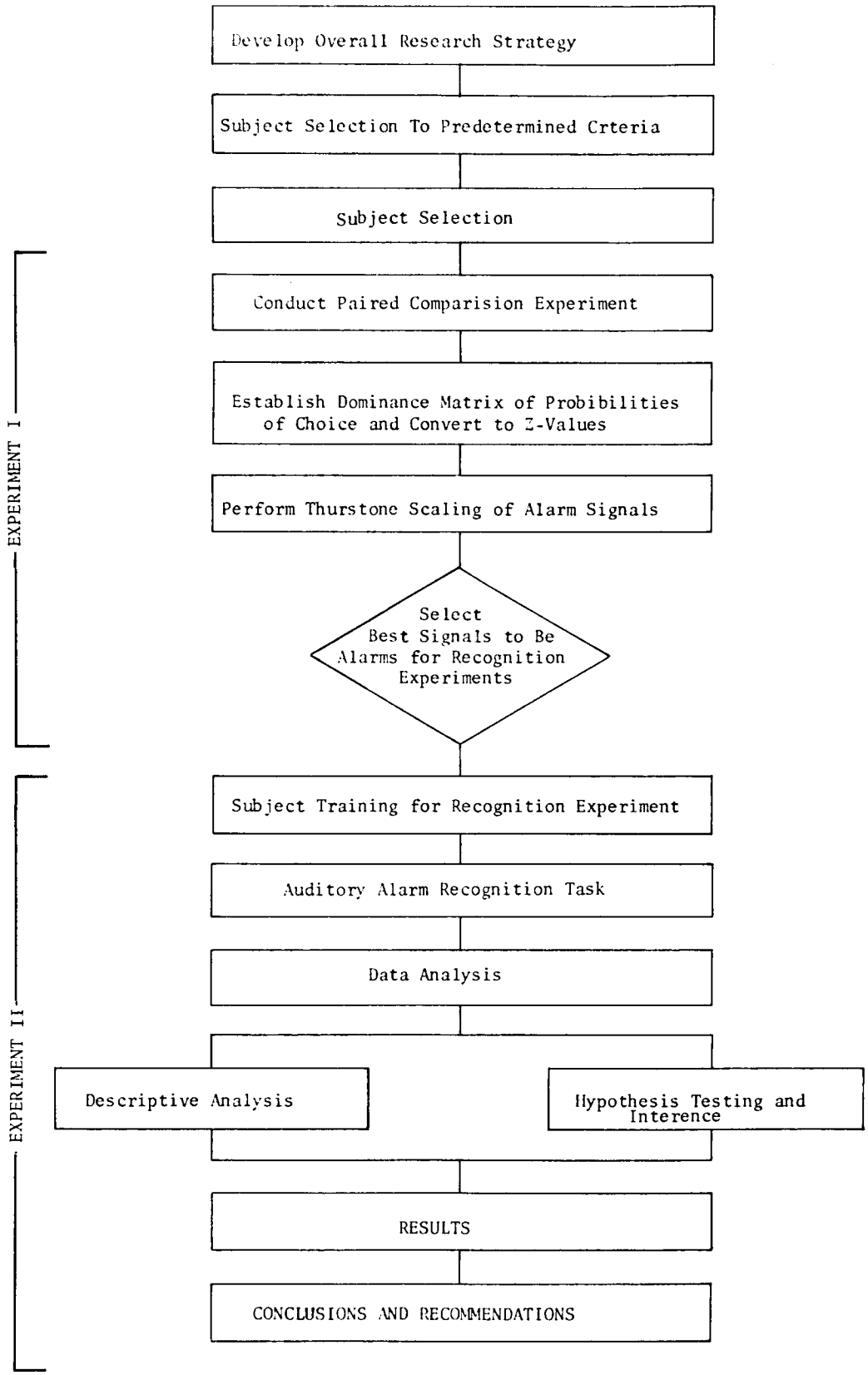


Figure III-1. Overall Research Stategy.

The subjects were asked to identify each alarm sounding for each trial and to reply in priority sequence with the learned names for each alarm identified. Each combination was presented once to each subject.

Each response was recorded whether it was correct or not and the response time of each response was recorded as dependent variables.

The independent variables were the names of the alarms and order in which they were given.

CHAPTER IV

EXPERIMENT I. DETERMINATION OF HUMAN PREFERENCES FOR AUDITORY ALARM SIGNALS

Introduction

Two aspects of auditory information processing, detection and relative discrimination, must be considered in developing coded auditory alarm systems and in measuring the effect such systems have on human subjects. The goal of this experiment was to determine human preference for any single alarm signal from a larger set of alarm signals. The subjects used in the experiment were presented alarm signals that were well above threshold (background) and they were told the signals were alarms from the set of alarms to be measured. The test subjects were asked to differentiate between alarm signals and to make a judgement as to which signal would be the most preferable as an alarm signal using a paired comparison technique. Then by summing how many times each candidate signal was preferred, it was determined which signal from the set of signals would make the best alarm, which would be the second best, and so on, using Thurstone scaling and then converting this scale to an interval scale. A secondary objective of this experiment was to identify which of these would be used in the recognition experiment (Experiment II).

The test subjects were chosen from a uniform population (university undergraduates in engineering) with each candidate being tested to a pre-determined standard before being selected as a subject. The

experiment was controlled and the data taken on a real-time computer system.

Experimental Design

In this experiment the independent variables are the physical parameters of the signals. The physical parameters of sound are frequency, waveform, amplitude, duration, and phase.

Frequency is expressed in hertzes (or cycles per second). A pure tone will have a simple sinusoidal waveform and a complex sound will be a set of sinusoids. Thus the waveform describes the various frequency components that comprise a complex sound. This tonal frequency was an independent variable in this experiment.

The amplitude is the measure of the magnitude of a pressure (sound) wave varying around the normal air pressure. The intensity or loudness is the measure in decibels of a ratio between two sound pressures and is the commonly used measure of the perceived impression on a human. This was the second independent variable in this experiment.

The meaning of duration is simply how long the sound lasts in time before an interruption occurs. When a series of sounds and interruptions occurs at regular intervals this is called modulation. In this experiment modulation was one of the independent variables.

What has been described in the introductory paragraph is the rank-ordering of the candidate signals based upon the degree of preference as to which signal was the best for an alarm signal. This was a dependent variable for this experiment. One way to obtain the

rank of the signals would be to present them, one by one, to the test subjects and ask them to rank them on an ordinal scale, the rank of 1 being assigned to the signal most preferred and the rank of n for the least preferred. This is known as ranking. However a better, although more time-consuming, method is to present the signals in pairs with each signal being compared at least once with every other signal. This is the method of paired comparisons. This method has two advantages that made it the choice for this experiment: first, the test subject's task is simpler, and second, transitivity of judgement can be tested. Since judgements of this type rarely show perfect transitivity another parameter had to be measured to insure a ranking was obtained.

This second dependent variable was the choice reaction time. Preference studies employing the paired-comparison method (Snodgrass 1975) have consistently found that decision time becomes longer as the number of ranks separating two stimuli decreases; and for a given rank separation, decision time is longer between lower-ranked or less-preferred pairs. This reaction time is a measure of how long the test subject takes to discriminate between the pair of signals and then to choose his preference.

With tonal frequency (frequency and waveform), intensity (amplitude), modulation (duration), and direction (phase) established as the independent variables, ranges of values for each of them were defined. These parameters made it possible to use a multidimensional code system for the transmission of relatively large amounts of information. Using the guidelines in Chapter II and the design recommendations proposed by Deatherage (1972) the following general ranges were selected.

- Frequencies between 500 and 3000 Hz, because the ear is most sensitive to this middle range.
- Use of a modulated signal of 1 to 8 beats per second or warbling sounds of 1 to 3 times per second, since these are different enough from normal sounds to demand attention.
- Use of signals with frequencies different from those that dominate any background noise to minimize masking.
- Use of sudden-onset signals of high intensity to alert the receiver.
- Each warning signal used to represent different conditions should be discernible from the others.

Thus for this experiment three tonal frequencies were chosen over the 500- to 3000-Hz range and these were modulated at four and eight beats per second at two intensities of approximately 80 and 100 dB. To prevent any left-to-right preferences from affecting the outcome of the ranking each pair was presented twice in each test, once as a left/right pair and a second time as a right/left pair. This counter-balancing was used to eliminate the data taken on subjects who used the direction of the sound rather than the alarm characteristics in making their choice. The order in which the pairs were presented was a randomized schedule so as to minimize any biasing of the outcome due to expectation or anticipation by the test subjects.

These three tonal frequencies, two modulations, and two intensities, plus the right/left, left/right presentation resulted in 12 different signals (Figure IV-1). Taking the 12 signals two at a time with the left-to-right being different than right-to-left yielded 132

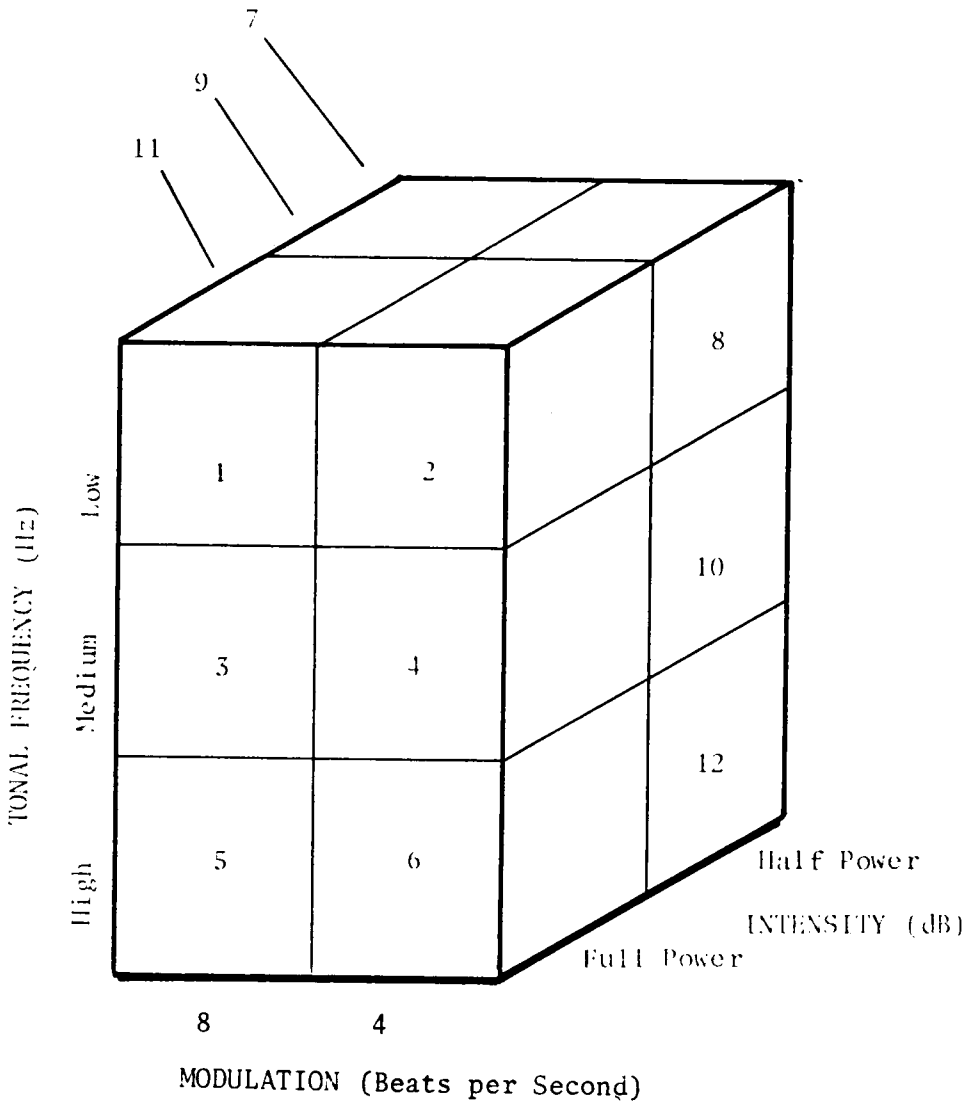


Figure IV-1. Research Design.

different permentations to be presented to each subject. Assuming a test rate of 3 signals a minute, it would require 3/4 of an hour to complete a test cycle. If the dimensions of the test were expanded to 27 different signals and the same test rate were maintained, it would have taken four hours to complete a test cycle. It was impractical for the test subjects to devote this much time to the experiment, so the shorter version was the one used.

Table IV-1 shows the various values assigned to each of the 12 alarms.

Subjects

The test subjects were university students and the sample was large enough to determine a population preference rather than an individualized preference. Each student completed a questionnaire to determine if he had some impairment such as tone deafness, that would make his response different from that of the other students. Also an attempt was made to assess their experience with alarm systems. Each student also took an audiometric test and a vision test. The information on the questionnaire and the results of the audiometric and vision test were reviewed to determine if any of the candidates had an impairment that could influence the outcome of the experiment. Data taken by any one who had an impairment was not used in the data analysis.

Subject Instructions

Each subject was acquainted with the purpose of the experiment, the experimental procedure, and the equipment being used. Sample signal

Table IV-1. Physical Parameters for the Alarms Used in Experiment I.

Alarm Number	Tonal Frequencies in Hz			Intensity (Loudness) in dB	Modulation Beats/sec.
	Frequency 1	Frequency 2	Frequency 3		
1	522.7	688.0	--	80	8
2	522.7	688.0	--	80	4
3	1177.5	1553.6	1864.3	80	8
4	1177.5	1553.6	1864.3	85	4
5	1869.6	2110.6	--	78	8
6	1869.6	2110.6	--	78	4
7	522.7	688.0	--	74	8
8	522.7	688.0	--	74	4
9	1177.5	1553.6	1864.3	82	8
10	1177.5	1553.6	1864.3	82	4
11	1869.6	2110.6	--	72	8
12	1869.6	2110.6	--	72	4

pairs were demonstrated, and the subjects were allowed to operate the system. Instructions given to the subjects were as follows:

> I would like for you to listen to 132 sets of auditory signals. The signals will be presented in pairs. The first signal will come either from the speaker on your left or from the speaker on your right. Then the second of the pair will come from the other speaker. Listen to the two signals and then tell me which of the pair you think is the better signal to use as a warning device. Indicate your choice by pressing the right-hand button for the signal from the right or the left-hand button for the signal from the left. The signals will continue to alternate between speakers until you make a choice. Even if you are not sure in some cases, I want you to tell me which one you <u>think</u> is the better alarm signal. When you have indicated your preference the next pair of signals will be presented. Do you have any questions?

These instructions appeared on the CRT of the terminal at the beginning of the experiment and were also available on paper at all times during the experiment.

Instrumentation

The actual presentation of the signal pairs was controlled by a Digital Corp. PDP-11/34A minicomputer coupled to a digitally controlled sound generation system. Two keys on the terminal keyboard were used to input the subjects' response to the minicomputer which

also served as the data recorder. A sound level meter was used to determine the ambient noise level and to monitor the output intensities of the sound generator. Figure IV-2 shows a block diagram of the system.

Experimental Setting

The experiments were conducted in a laboratory where background noise, disruptions, and other distractions could be minimized. The subjects did not have any task to perform other than responding to the given signal pairs. The environment was maintained to a nominal comfort zone to reduce stress and maintain alertness. The computer software was written so that the subjects could self-pace themselves in the experiment and even take a short break without having to start the experiment over. The arrangement of the laboratory is shown in Figure IV-3.

Data Collection

After a test subject was given his instruction for the test and acquainted with the equipment, he was logged on to the PDP-11 computer to take the paired-comparison experiment. Each subject was assigned a code number, which he then entered into the computer. This created a data file for that subject on a storage device. Thus, as the subjects proceeded through each trial of the experiment, information about each trial was recorded on the storage device. The information recorded consisted of the trial number, the names of the signals making up that trial, the name of the signal chosen by the subject and the elapsed time (response time) since the trial began. The data thus stored was analyzed to determine if there was a preference for one or more of the alarm signals.

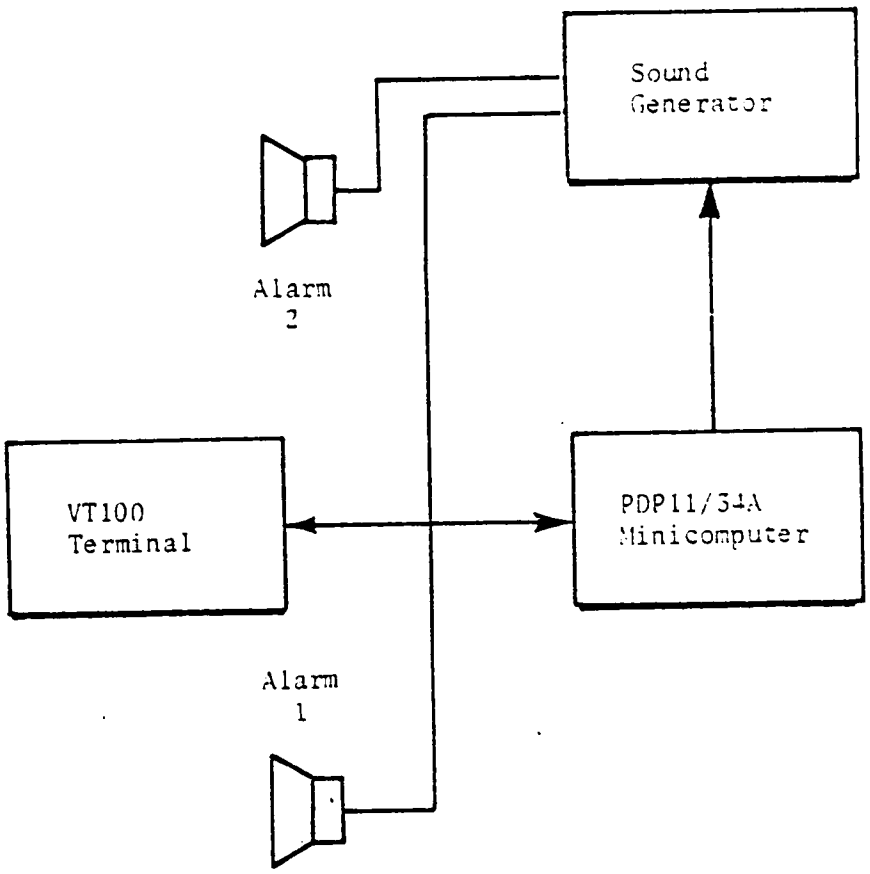


Figure IV-2. Real-time Experimental Design.

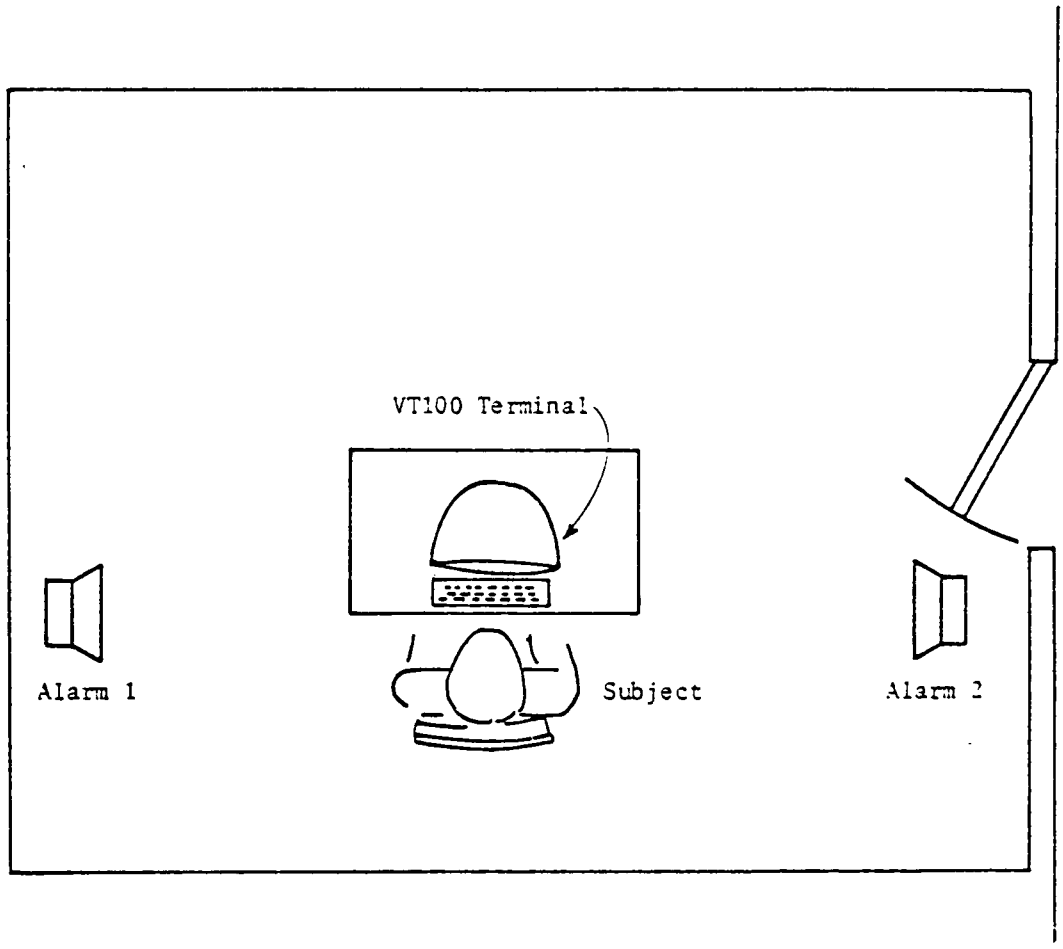


Figure IV-3. Laboratory Arrangement--Experiment I.

Data Analysis

After the data were collected, each subject's data were reviewed for completeness and errors. The data were also examined for a directional preference; any data found to have a significant preference (60 percent) for a given direction were not used in the analysis. When this integrity review was completed, the individual data sets were combined into a single data set and transmitted to the University's Digital Corporation DEC-10 computer. The DEC-10 was then interfaced with the University's IBM 360/370 computer where SAS (Statistical Analysis System) was used to analyze the data.

SAS is a computer system for data analysis, developed by the SAS Institute, Inc., of Cary, North Carolina. SAS provides for storage and retrieval, file handling, programming, statistical analysis, and report writing.

SAS was used to produce a summary of the very large data set that was created when the individual data files were combined. This summary provided the frequency with which one alarm is chosen over another. It also calculated the mean reaction time and the variance for each of these comparisons.

A probability dominance matrix was developed by using Thurstone's Law of Comparative Judgement (Baird and Noma, 1978). This matrix was transferred into a dominance matrix of Z-values. From these Z-values a scale value was derived and placed on an ordered scale where the order reflects the importance placed on that signal by the subjects.

Results of Experiment I

The subjects' task in this experiment was to select the auditory signal of the pair of signals being considered that would be the best signal for an auditory alarm. Of the fifty-two candidates for this experiment forty-six became subjects. Thirty-eight data sets were chosen for analysis. The eight not chosen were eliminated for various reasons such as directional preference, data errors, or because they did not fit the population criteria, as explained in the subject selection and data analysis sections. The paired comparison technique used yielded a dominance matrix which is the probability of one signal being chosen over another signal. This dominance matrix was then transformed to a dominance matrix of Z-values. From these Z-values an ordered scale value for each system was derived. The Z-value represents the importance each signal has to the test subjects. The scaled values derived for each signal are shown in Figure IV-4.

Signal 2 (low tonal frequency at 4 beats/sec and at full intensity) received the highest value (.77). The majority of the subjects thought the low tonal frequency, 4 beats/sec, full intensity signal made the best alarm signal. The next signal chosen had an intermediate tonal frequency, 4 beats/sec, and full intensity. The signal order in descending order is 2, 4, 1, 10, 3, 8, 9, 6, 7, 12, 5, 11. The majority of the top six alarms were split evenly by tonal frequency between low and intermediate tones and 4 of these were at 4 beats/sec and 4 at full intensity. The Z-value of .77 suggests that 78 percent of the subjects chose this signal when it was compared to another signal. A score of zero indicates that the subjects were equivocal

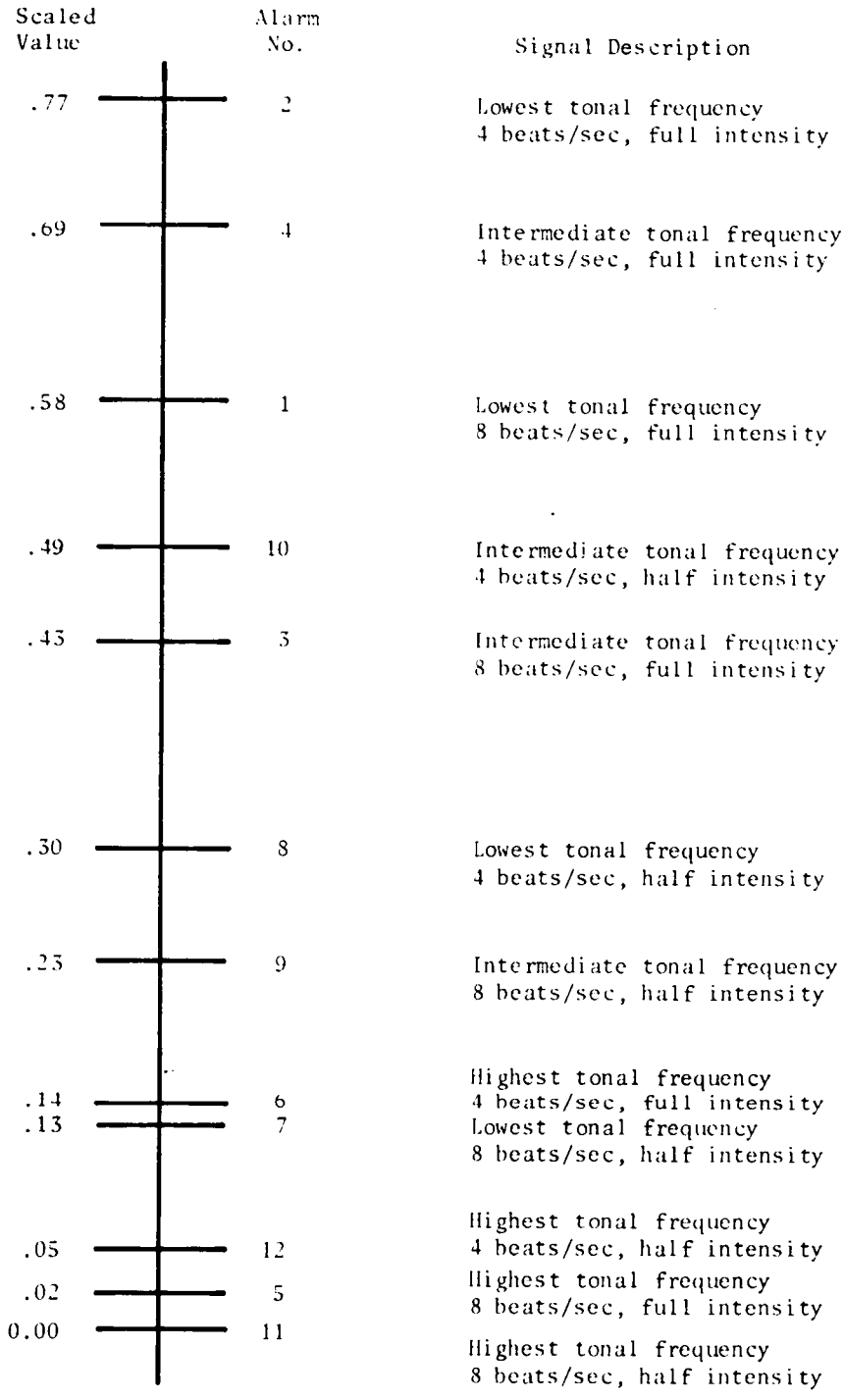


Figure IV-4. Results from the Analysis of Paired-Comparisons of Signal Alternatives (N=38).

towards the signal being presented for comparison, that is half the time they would prefer some other signal when that signal was presented.

An evaluation of the response times gave another dominance matrix which is the mean reaction time when one signal is chosen over another. From this dominance matrix an ordered scale was obtained based on the mean response time for each alarm. The alarms were ranked from 1 to 12 on this scale with the rank of 1 corresponding to the smallest mean response time. This scale is shown in Figure IV-5.

A comparison of this ranking to the one obtained from the preferences is shown in Table IV-2. A Spearman rank correlation coefficient was calculated for these two rankings. The correlation coefficient, r_s, was found to be -0.213.

Since the secondary objective of this experiment was to identify which of these alarms would be used in the recognition experiment, alarms 2, 4, 1, 10, 3, and 8, the six most preferred alarms, were chosen. These six were chosen because they are very evenly spaced on the ordinal scale and form an alarm set of a size that could be reasonably tested. A comparison of the physical characteristics for these six alarms is shown in Figure IV-6. The alarms were not chosen on the basis of mean response time because what would have been the 3, 4, and 5 alarms had essentially the same response times.

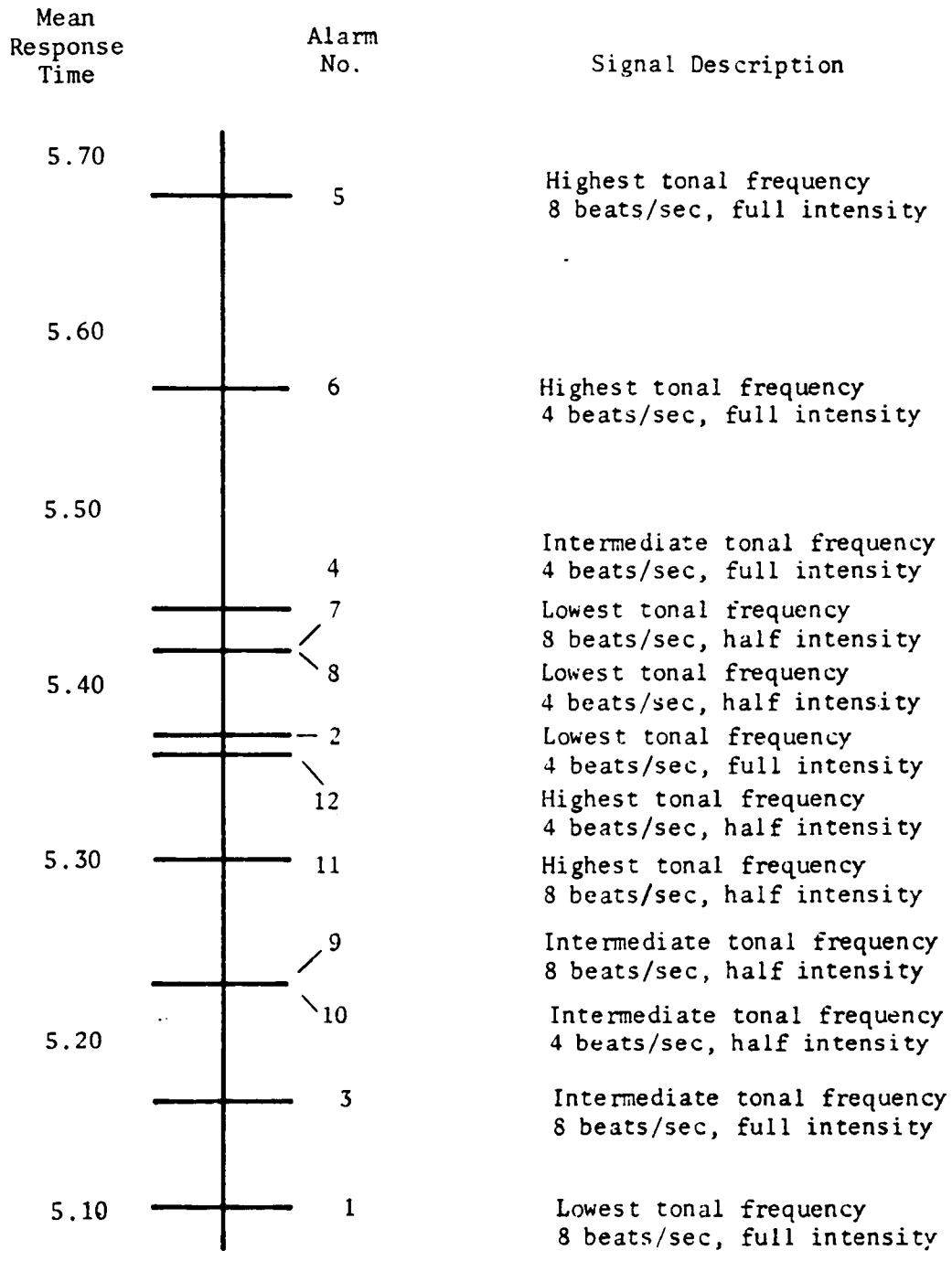


Figure IV-5. Ranking of Alarms by Response Time.

Table IV-2. Comparison of Rankings.

Rank	Alarm Name	
	Paired-Comparison	Response Time
1	2	5
2	4	6
3	1	4
4	10	7
5	3	8
6	8	2
7	9	12
8	6	11
9	7	9
10	12	10
11	5	3
12	11	1

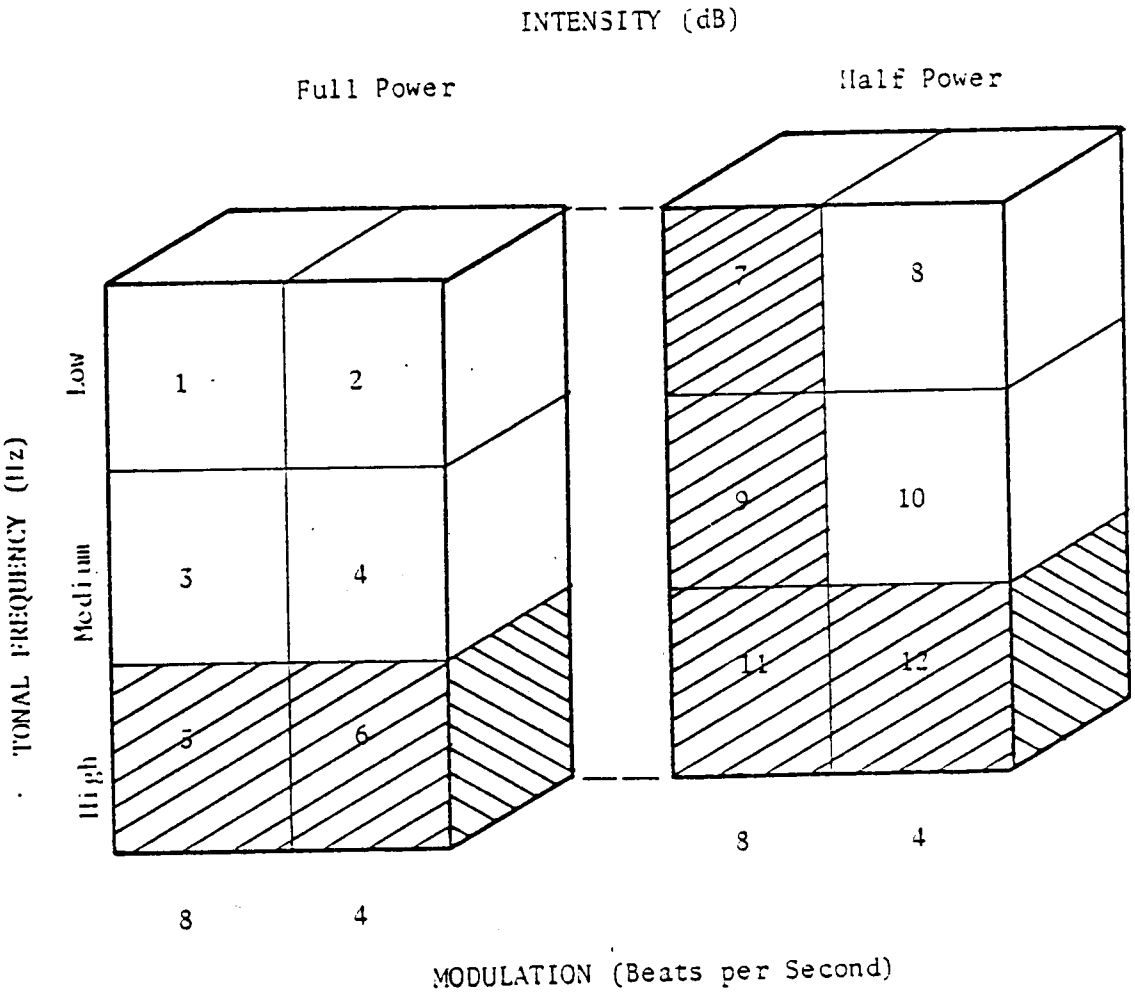


Figure IV-6. Physical Characteristics of the Selected Alarms.

CHAPTER V

EXPERIMENT II. HUMAN PERFORMANCE IN A SIMULTANEOUS MULTIPLE-ALARM ENVIRONMENT

Introduction

The aspects of and information processes considered in this experiment are absolute discrimination and recognition. In this task the problem for the test subject was to determine how many alarms were being presented in a trial. The number of alarms being presented varied from trial to trial in a random manner. Also, the subject was to recognize and identify each of the alarms by name for each trial. The subject was then to respond by giving the names of the alarms in a predetermined priority sequence. As the number of alarms to be detected and then identified increases, the subject's ability to discriminate between them decreases. Also, his ability to correctly identify the alarms, even if he can discriminate between them, decreases as the number being presented increases. The question to be answered is how does the presence of multiple alarms in an absolute discrimination/recognition task affect errors, response time, and variability? The subjects used in this task were taught to identify the six alarms chosen in the paired-comparison experiment by name and rank. The subjects were then asked to listen to sets of these alarms where each set consisted of either one single alarm, or two, three, or four alarms sounding simultaneously. The task was to identify the alarms sounding by name and priority. Error and response times can be evaluated to determine performance with one-, two-, three-, or four-alarm systems.

Experimental Design

In this absolute discrimination/recognition experiment there are five independent variables. These independent variables are frequency, waveform, amplitude, duration, and phase. These are the same as in Experiment I.

There are two dependent variables in this absolute discrimination/recognition experiment. These two variables are the responses that are given when the alarms in a trial have been identified, whether or not the identifications are correct, and the response times for each response.

Each response was the name that had been learned to be associated with a given sound. There were six names to be learned and associated with the physical parameters of the alarms. In reviewing the literature it was found that the dependent variable, response time, increased for signals that are similar or hard to discriminate between, and decreased for signals that are easy. Also, response time increases because several stimuli are being presented, each requiring its own response. Here response time is increased largely because of the additional time required to make a decision on what the correct responses should be (McCormick, 1976).

For this experiment six alarms were used. The alarms were named and ranked as shown in Table V-1. These rankings were determined from Experiment I. These alarms were presented to the subjects in sets in all possible combinations of alarms where the alarms were taken one at a time, two at a time, three at a time, and four at a time. This resulted in 56 sets or trials to be tested. However, since the alarms had to be responded to in a pre-defined order, any combination that

Table V-1. Conversion of Alarm Names from Experiment I to Experiment II

Rank	Experiment I	Experiment II
1	2	1
2	4	2
3	1	3
4	10	4
5	3	5
6	8	6

ended with alarms 5 and 6 made it extremely easy to respond correctly with a 6 after the 5 alarm had been identified. Therefore all trials that ended with alarms 5 and 6 were eliminated. This resulted in an experiment with 45 trials. Each subject was given these 45 trials by the same randomized sequence, as shown in Table V-2.

Subject Selection

Most of the subjects used in this second experiment were participants in the preference experiment, Experiment I. The few "new" subjects from the student body were given the same questionnaire, hearing, and vision tests that the first subjects were given.

However, since this is an absolute recognition experiment and also a relative discrimination experiment, the subjects had to be trained to recognize each alarm and to respond by giving the correct name that was assigned to each alarm. The training was divided into two steps. The first training step consisted of a routine on the computer with which the subject could ask for an alarm by name and the computer system displayed that alarm for as long as the subject desired. The subject

Table V-2. Randomized Schedule of Alarms.

Trial* No.	Alarms in** Trial (By Name)	Trial* No.	Alarms in** Trial (By Name)	Trial* No.	Alarms in** Trial (By Name)
1	1 5	18	1 6	33	1 3 4
2	1 2 4	19	2 3 6	34	1 2 4 5
3	3 4 5	20	3 4	35	1 2 3
4	1 2 3 5	21	4 5	36	1 3 6
5	2 6	22	1	37	3 4 6
6	1 4 6	23	1 2	38	1 3 5
7	2 3 4 5	24	3 6	39	1 2 4 6
8	2 3 5	25	1 2 6	40	3
9	1 2 5	26	3 5	41	2 3
10	4 6	27	1 3 4 6	42	2 5
12	2 3 4 6	28	2	44	1 3
13	5	29	2 4 6	45	2 4 5
15	1 2 3 6	30	6	46	2 4
16	1 4 5	31	1 2 3 4	47	1 4
17	2 3 4	32	1 3 4 5	48	4

*There are no trials numbered 11, 14, or 43.

**For the dimensions of each alarm see Table IV-1.

could ask for any of the six alarms, in any order, for as many times as was necessary for the subject to learn them.

Once the subject convinced himself that he could recognize the six alarms, he used the computer system to give himself a short recognition test. This was the second step in the training. This recognition test consisted of a randomized schedule of the six alarms with each alarm being given twice for a total of twelve trials. Only one alarm was given at a time. To prevent the subject from memorizing this short sequence, the sequence was changed very often. The computer system displayed the subject's responses and what the correct responses were at the end of each test. A perfect score was required before the subject was allowed to continue. When a less-than-perfect score resulted, the subject was asked to return to the first step and practice with the alarms some more. When the twelve-trial test was passed, one more test was given to the subject to make sure he could recognize all six of the alarms.

This test consisted of eighteen trials with each alarm being presented three times by a randomized schedule. Again the alarms were given one at a time. Only two incorrect responses were allowed on this test and they could not be the same alarm. If the subject failed this test he was either asked to go back to step one or was dismissed as a subject. Only three subjects declined to participate when asked to repeat the learning steps.

Subject Instructions

As in the earlier experiment the subject was acquainted with the purpose of the experiment, the experimental procedure, and the equipment

to be used. Sample sets of alarms were demonstrated and the subjects were allowed to operate the system. In addition to this, the subjects were told that the alarm named one had the highest rank, the alarm named two had the next highest rank, and so on to the alarm named six having the lowest rank.

The computer system displayed these instructions to the subject at the beginning of each experiment:

> I would like for you to listen to several sets of auditory alarm signals. The signals will be presented in sets that contain from 1 to 4 alarms in each set. You are to identify each alarm given in a set by name and to enter the name into the computer in order by rank. To enter a name press the numeric key on the keyboard that corresponds to the name.

Instrumentation

The presentation of the alarm sets was presented by the PDP-11/34A minicomputer coupled to the sound generation system used in the first experiment. However, in this experiment all four channels were utilized rather than just two. Also, the numeric keys were used to input the subjects' responses to the minicomputer.

Experimental Setting

The setting was the same as described for the previous experiment with only the use of four sound channels rather than two, necessitating a different physical arrangement of the components. Figure V-1 shows this arrangement.

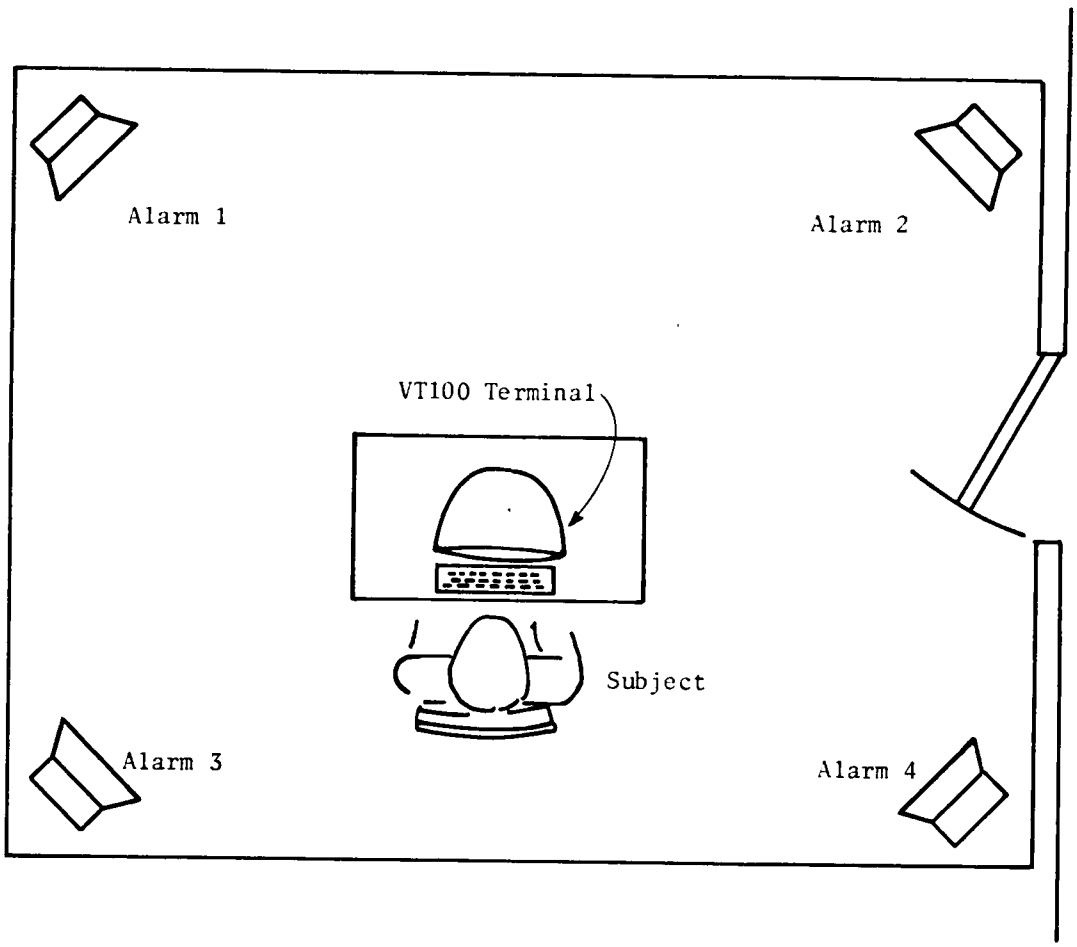


Figure V-1. Laboratory Arrangement--Experiment II.

Data Collection

As in the paired-comparison experiment, each subject began the recognition experiment by entering an assigned identification number into the computer to create a data file on the storage device. As the experiment progressed, information about each step of the experiment was recorded. The information recorded consisted of the trial number, the names of the alarms presented to the subject, the responses given by the subject, the time since the start of the trial to each response, and whether or not it was a correct response.

Data Analysis

After all of the subjects had completed the experiment, their data were reviewed for completeness and errors. Any data file that was found to have errors or missing data was not used in the analysis. When this data integrity review was completed, the individual data files were combined into a single data set after being transmitted to the University's Digital Corporation DEC-10 computer. The DEC-10 was then linked to the IBM 360/370 computer where SAS was used to analyze the data in a manner similar to the analysis used in Experiment I. A flow diagram of the data analysis is shown in Figure V-2. Response time, number of responses, and the number of incorrect responses for each trial for each subject were recorded and analyzed. These measures were also divided into four categories, where the category is the number of alarms sounding for a trial. These measures consist of description statistics by trial and by category. With the descriptive statistics compiled the data used to compile them was tested for the null hypothesis that the data came from a random sample with a normal distribution. A

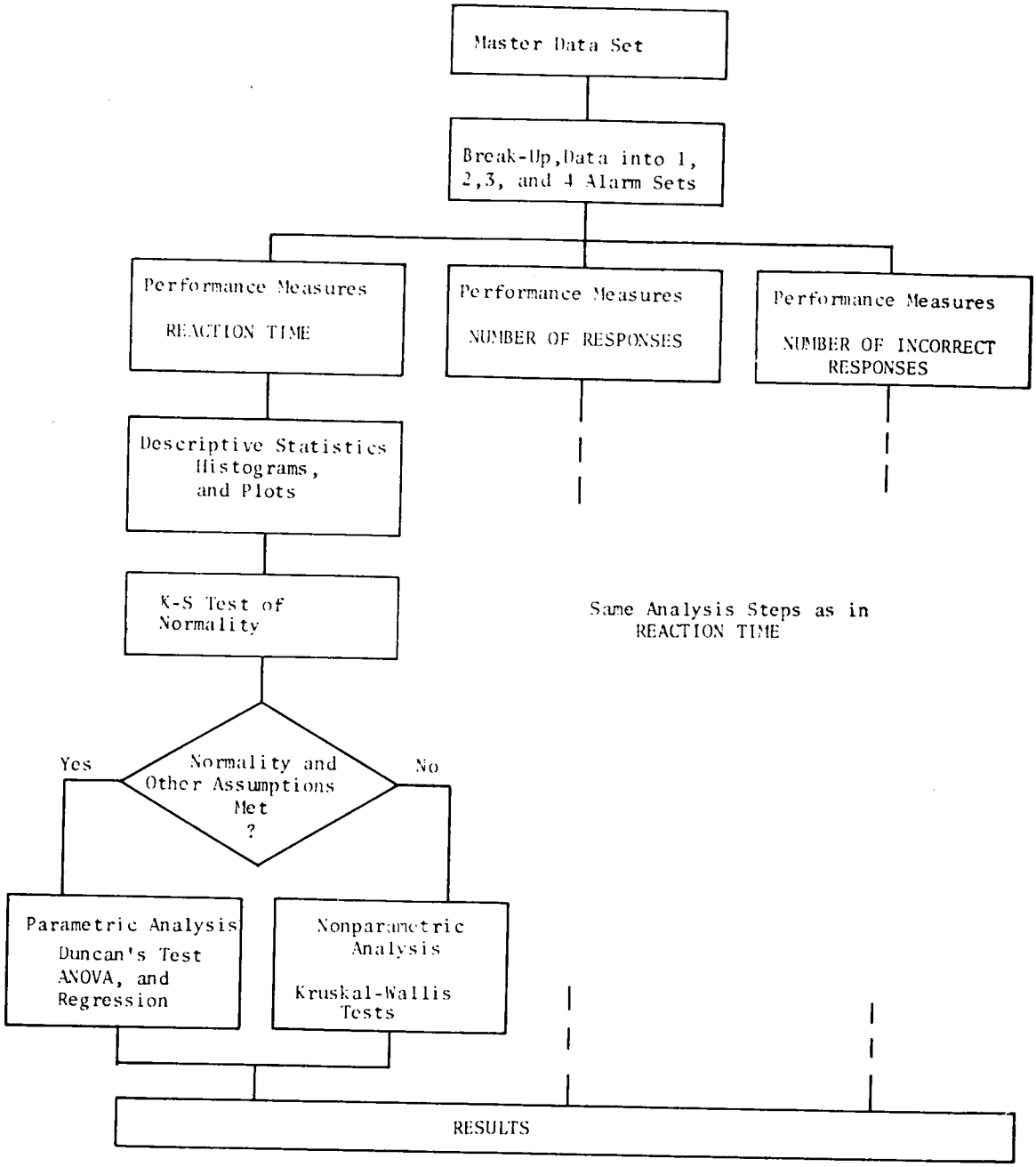


Figure V-2. Data Analysis Flow Diagram.

Kolomogorov-Smirnov D-statistic was calculated and used to decide whether a parametric analysis or nonparametric analysis was to be used. This analysis was then performed to determine how people perform in an auditory relative discrimination task when presented with several alarms simultaneously. Also, the data were used to determine which alarm system within a one-, two-, three-, or four-alarm category seems to be the best. Next, several mathematical models were developed and tested with the data from each of the three areas for which measures of performance were obtained in order to find the best model to predict how a person's performance might change if the alarm system were changed.

Results of Experiment II

In this experiment the test subjects were asked to perform a relative discrimination task. The task consisted of listening to sets of alarms and attempting to identify the alarms by name and rank. Twenty subjects participated in the experiment and eighteen sets of data were chosen for analysis. Two data sets were rejected because of problems encountered in transmitting the data between computers. None of the data was rejected in the integrity review.

The first of the performance measures to be analyzed was the maximum response time for each trial by category. Maximum response time is defined as the elapsed time from the beginning of a trial until the last alarm being presented has been correctly identified. If all of the alarms were not correctly identified in 20 seconds the trial was terminated by the computer. These data are shown in Table V-3. An

Table V-3. Means for Maximum and Minimum Response Times for Category 1.

TRIAL	MAXMEAN[a]	MINMEAN[b]	MAXSTD[c]	MINSTD[d]
13	4.82778	3.72778	3.02632	2.13150
22	4.60000	3.41111	3.26677	1.49505
28	4.66667	3.15556	2.12021	0.96481
30	2.73889	1.93889	2.24188	0.93314
40	4.50000	3.65000	2.38426	1.93824
48	4.05000	2.95000	3.74437	1.66883

[a]MAXMEAN = Mean for the Maximum Response Time.

[b]MINMEAN = Mean for the Minimum Response Time.

[c]MAXSTD = Standard deviation for the Maximum Response Time.

[d]MINSTD = Standard deviation for the Minimum Response Time.

examination of this data shows that response times increase as the number of alarms increases.

The distribution of response times was examined for normality using a Kolmogorov-Smirnov (K-S) test statistic. The K-S test statistic, D(n) is defined at the maximum absolute deviation between a cumulative sample function (S(x)) and a cumulative normal function (F(x)). In this instance it was found that the response time sample functions for 1, 2, or 3 alarms were normal at the alpha < 0.01 level. The sample distribution for response times for four alarms was normal at the alpha < 0.085 level. This latter distribution was truncated because of the 20-second time-out capability built into the experimental system on each trial.

With the data's normality established a one-way analysis of variance was performed to determine if any of the categories were significantly different from any of the others at the 0.5 alpha level. The categories were significantly different because the F-value is 114.11 as compared to a critical value of 2.6 and the probability of getting this F-value equal to 114.11 or larger is 0.0001 under the null hypothesis that the categories are all the same. A Duncan test showed that all of the alarms were significantly different at the alpha = 0.05 level.

The next question to be answered is, which trial or alarm system within each category would make the best alarm system? It would appear that the trial with the lowest mean response time would be the best alarm system. The times for the category 1 trials appear in Table V-3. For category 1, this would be trial 30 followed by trial 48 and so on to trial 13 being the least desirable. Likewise, an examination of the data in Table V-4 for category 2, Table V-5 for category 3, and Table V-6 for category 4 yielded similar results to those obtained for category 1. This data is also shown graphically in Appendix A.

In order to determine if there is a significant difference between the trials, a one-way analysis of variance with Duncan's multiple range test was performed at the 0.05 alpha level within each category. For category 1, no differences were found between the 6 trials examined. In other words, any one of the six alarms could be considered to be equally effective. In category 2 the 14 trials were found to be divided into 3 groups, category 3 divided into 5 groups, and category 4 divided into 3 groups. The ANOVA tables with Duncan's Test results are in Appendix A.

Table V-4. Means for Maximum and Minimum Response Times for Category 2.

TRIAL	MAXMEAN	MINMEAN	MAXSTD	MINSTD
1	11.1833	5.07778	4.32071	3.08269
5	9.1556	4.11667	4.79402	2.73006
10	9.2778	3.60000	3.27083	1.82724
18	7.2111	3.60556	4.34672	1.89006
20	7.6833	3.13333	2.53870	2.30906
21	7.8222	3.73333	3.51744	2.20587
23	7.7722	3.03333	4.82394	2.44396
24	6.4833	2.75000	2.50488	1.64433
26	6.8500	3.42222	2.61989	1.84823
41	6.0500	2.70000	2.03101	1.58300
42	7.5222	4.33889	3.57862	2.39242
44	6.8556	2.91667	4.21332	1.75943
46	6.2111	3.21667	2.45785	1.17486
47	6.6000	2.72778	3.64482	1.29923

Table V-5. Means for Maximum and Minimum Response Times for Category 3.

TRIAL	MAXMEAN	MINMEAN	MAXSTD	MINSTD
2	12.9722	3.15556	4.49991	1.26362
3	14.7444	4.27222	4.37065	2.57678
6	12.8333	4.06111	4.10409	2.09288
8	10.2111	3.63889	2.59794	1.63963
9	11.1222	3.55000	3.56413	2.05176
16	11.5667	3.68333	4.48435	2.56452
17	11.8722	3.51667	4.68989	2.17641
19	9.9278	3.15556	3.64059	1.76876
25	9.3500	2.65000	3.93928	1.40807
29	8.7667	2.81111	3.75280	1.20776
33	11.2556	3.45556	4.21262	1.06452
35	10.8278	2.86111	3.54481	1.37462
36	8.8444	3.11667	4.35744	1.45935
37	9.8778	3.22222	3.42286	2.00310
38	9.1222	3.24444	3.43966	1.72396
45	8.5167	3.10556	2.88225	1.39177

Table V-6. Means for Maximum and Minimum Response Times for Category 4.

TRIAL	MAXMEAN	MINMEAN	MAXSTD	MINSTD
4	14.1444	3.58889	5.01342	1.56276
7	12.9778	3.25556	3.34093	1.61581
12	13.2278	3.57222	4.14783	1.52385
15	14.3222	4.28889	4.76946	3.07607
27	12.0111	2.93889	4.86438	2.30808
31	13.0778	2.93333	4.45389	1.96499
32	9.6778	2.81111	4.11404	1.35468
34	9.1111	2.36111	3.87342	1.39036
39	10.5556	2.93333	2.93282	1.69671

In order to determine if there is a significant difference between the trials, a one-way analysis of variance with Duncan's multiple range test was performed at the 0.05 alpha level within each category. For category 1, no differences were found between the 6 trials examined. In other words, any one of the six alarms could be considered to be equally effective. In category 2 the 14 trials were found to be divided into 3 groups, category 3 divided into 5 groups, and category 4 divided into 3 groups. The ANOVA tables with Duncan's Test results are in Appendix A.

An examination of the mean minimum reaction times in Tables V-3 through V-6 shows that they are all approximately the same. This can also be seen in the bar charts in Appendix A. The grand means, shown in Table V-7 and in the bar chart, Figure V-3, for each category support this conclusion. Only the category 2 grand mean is high and this is because of the high minimum time that occurred in trial number 1. Almost every subject had high times for this first trial--apparently

Table V-7. Grand Means for Response Times.

CATEGORY	GMAXT	GMINT	MAXDEV	MINDEV	MAXVAR	MINVAR
1	4.2306	3.13883	2.88115	1.66300	8.3010	2.76558
2	7.6198	3.45516	3.75350	2.13204	14.0888	4.54559
3	10.7582	3.34375	4.15132	1.79633	17.2334	3.22860
4	12.1228	3.18704	4.49446	1.93240	20.2002	3.73418

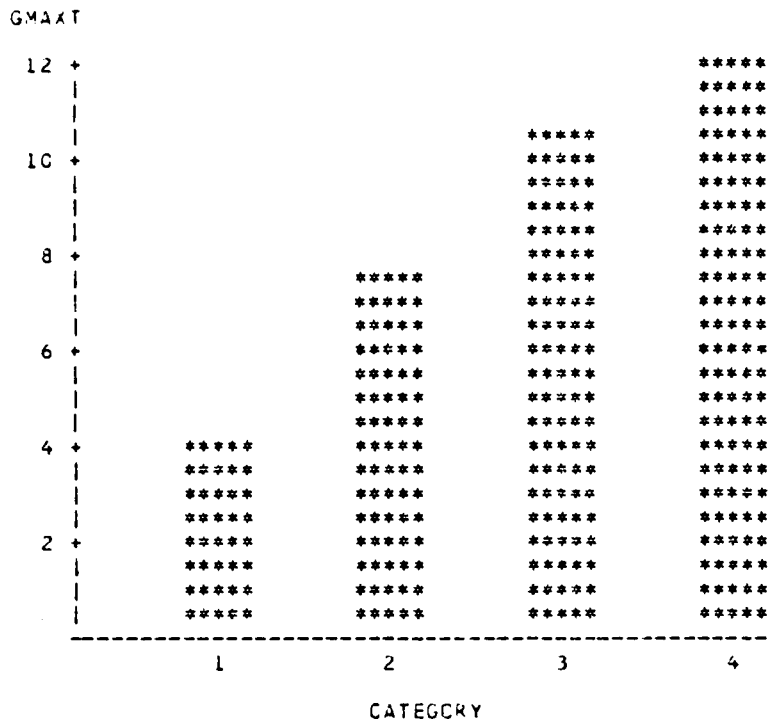


Figure V-3. Bar Chart of Grand Means for Response Time.

because it was the first trial. This relatively constant minimum response time for each category indicates that the subjects were making their first attempted identification in the same amount of time regardless of the number of alarms being presented.

For the mean response time measure of performance the functional relationship between mean response time and number of alarms suggested that three mathematical models might fit the data. These models are of the form $y = \alpha + \beta x$, a linear model; $y = \alpha + \beta \log x$, a logarithmic model, and $\frac{1}{y} = \alpha - \beta/x$, a reciprocal model. The last, or the reciprocal, model yielded the best fit. The F-ratio for this model was 563.98 with a probability of a larger F being 0.0001 and R-squared being 0.2937. The value of α was 0.0002 and the value of β was 0.336.

The second performance measure that was analyzed was the number of responses for each trial by category. If each subject responded correctly 100 percent of the time the data would be represented by a straight line with a slope of 1. What is expected is for the number of responses to increase somewhat more quickly; that is, the slope of a fitted curve would be greater than 1, as the number of alarms presented is increased. The data shown in Table V-8 and Figure V-4 support this and appear to be linear.

The data were examined for normality by the same procedures used to test the response times. In each case the data were found to be normally distributed. The D-statistic and other related statistics are shown in Table V-9.

The one-way analysis of variance that was performed on the reaction time for each trial between the categories yielded an F-value of

Table V-8. Grand Means for Number of Responses.

CATEGORY	GMEANR	STDDEV	VARIANCE
1	1.87963	1.57502	2.48070
2	3.46032	1.47840	2.18567
3	5.13750	2.08184	4.33406
4	6.35062	3.00683	9.04102

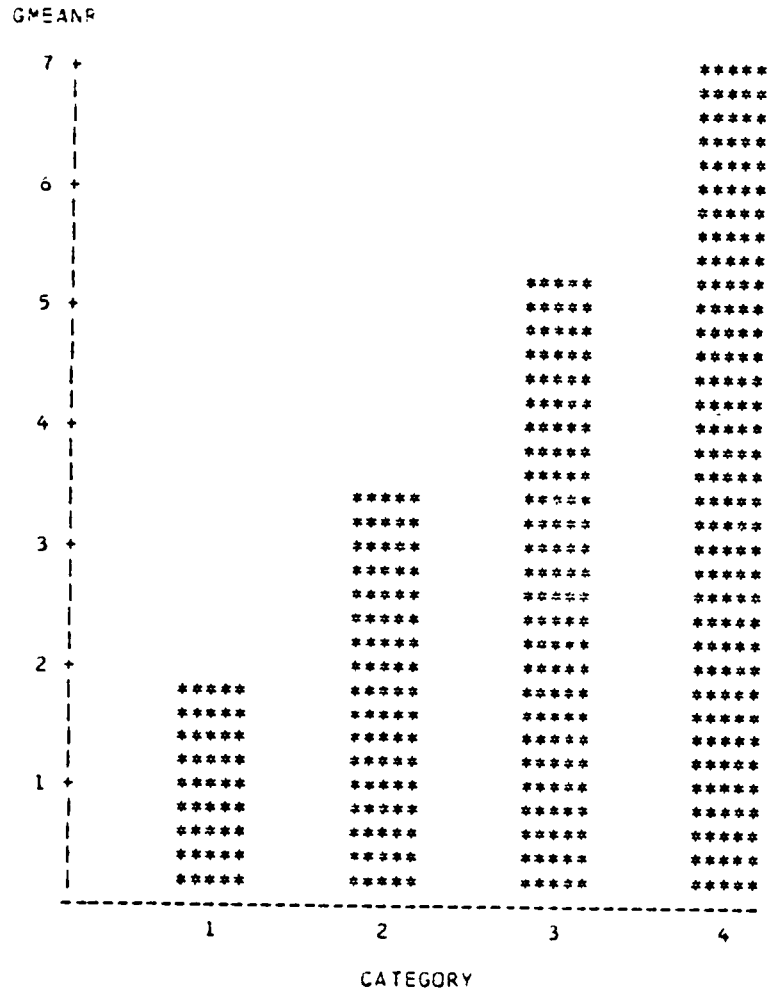


Figure V-4. Bar Chart of Grand Means for the Number of Responses.

Table V-9. Results of Normality Tests for Each Category on the Number of Responses

Category	N	T:Mean =0	Prob > |T|	D:Normal	Prob > 0
1	108	12.402	0.0001	0.3229	0.01
2	252	37.156	0.0001	0.2214	0.01
3	288	42.287	0.0001	0.1921	0.01
4	162	29.422	0.0001	0.2228	0.01

162.10 with a p-value of 0.0001. Thus the hypothesis that they are all the same is rejected and that at least two categories are different. The results of Duncan's multiple range test show that each category is significantly different from all other categories at the 0.05 alpha level.

A one-way analysis of variance was performed on the trials within each category. The data used in these analyses is summarized in Table V-10 through Table V-13. These data are also shown as bar charts in Appendix B.

The results of these one-way analyses of variance tests performed on each of the categories are shown in Appendix B. The analyses for category 1 indicate that the trials are not significantly different. However, categories 2 and 3 show a significant difference and Duncan's test places the trials into 4 groupings for each category. Category 4 has an F-value of 2.06 as compared to a critical value of 1.94. Duncan's test shows only one trial is significantly different from the other eight. A logarithmic model of the form $y = \alpha + \beta \ln x$ was chosen for the average number of responses. The model yielded a value of 0.153 for α and a value of 0.457 for β. The F-ratio was 843.53 with a p-value

Table V-10. Means for Number of Responses for Category 1.

TRIAL	AVGRSP	STDDEV
13	1.94444	1.51356
22	1.61111	1.14475
28	2.27778	1.99427
30	1.50000	0.92355
40	2.00000	1.32842
48	1.94444	2.23534

Table V-11. Means for Number of Responses for Category 2.

TRIAL	AVGRSP	STDDEV
1	4.11111	1.60473
5	3.38889	1.61387
10	4.38889	1.24328
18	3.00000	1.64496
20	3.88889	1.07861
21	4.22222	1.55509
23	3.66667	2.19625
24	2.94444	0.99836
26	3.16667	1.24853
41	3.22222	0.94281
42	3.27778	1.27443
44	3.33333	1.81497
46	2.88889	1.07861
47	2.94444	1.16175

Table V-12. Means for Number of Responses for Category 3.

TRIAL	AVGRSP	STDDEV
2	5.27778	1.96456
3	6.16667	2.40710
6	5.16667	1.75734
8	4.72222	1.99427
9	5.72222	1.87257
16	5.88889	1.52966
17	5.16667	1.29479
19	4.88889	2.13896
25	4.44444	1.85416
29	4.16667	0.98518
33	5.88889	3.19722
35	6.61111	3.18339
36	4.27778	1.74240
37	5.00000	0.97014
38	4.55556	1.78958
45	5.05556	2.12747

Table V-13. Means for Number of Responses for Category 4.

TRIAL	AVGRSP	STDDEV
4	6.22222	3.04004
7	6.72222	1.74240
12	6.77778	2.62467
15	7.16667	3.12956
27	6.05556	1.51356
31	9.44444	5.05525
32	6.61111	2.40438
34	6.66667	3.04863
39	6.88889	2.32351

of 0.0001. The R-square value is 0.511. These values indicate a reasonable fit for the data and explain 51 percent of the variation.

The final measure of performance was the average number of incorrect responses for each trial in each category. These data are given in Table V-14 and shown in the form of a bar chart in Figure V-5. These data show the means increasing as the number of alarms sounding increases as expected from Hick's Law. The variances for categories 1 and 2 are small and approximately the same. However, the variance increases by a factor of 2 for the 3-alarm category and by a factor of 5 for the 4-alarm category. The higher mean values and more variable performance indicates that the subjects experienced a larger degree of difficulty in identifying the alarms in categories 3 and 4 than they did for categories 1 and 2.

Again the data were found to be normally distributed for each category by the K-S test. The results of these tests are given in Table V-15.

The one-way analysis of variance performed on the trials between all categories indicate that at least two of the categories are significantly different with an F-value of 30.58 and critical value of 2.60. The p-value is 0.0001. Duncan's multiple range test shows that each category is significantly different from any other category at the 0.05 alpha level.

A one-way analysis of variance was also performed on the trials within each category. The data used in these analyses are summarized in Tables V-16 through V-19 and also shown in the form of bar charts in Appendix C.

Table V-14. Grand Means for Number of Incorrect Responses.

CATEGORY	GMEAN1	STDDEV	VARIANCE
1	0.87963	1.57502	2.48070
2	1.48810	1.53701	2.36241
3	2.25347	2.14175	4.58710
4	3.12346	3.10832	9.66163

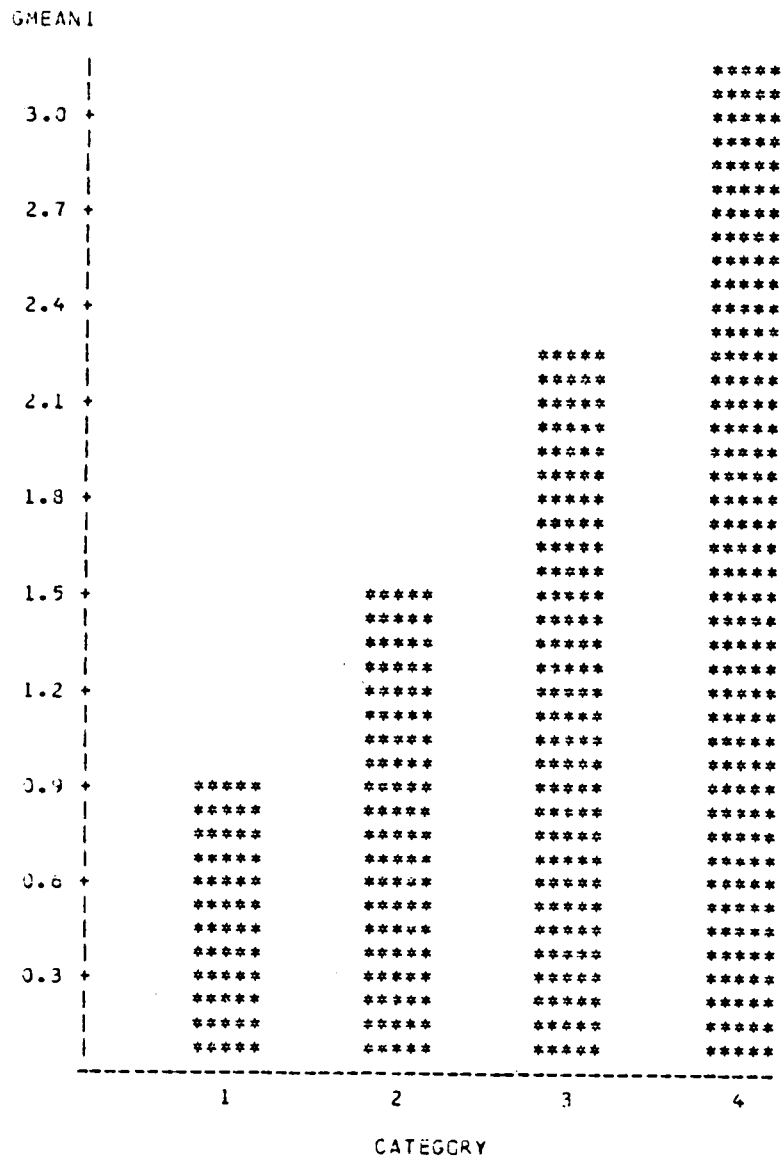


Figure V-5. Bar Chart of Grand Means for the Number of Incorrect Responses.

Table V-15. Results of Normality Tests for Each Category on the Number of Incorrect Responses.

Category	N	Mean	T:Mean = 0	Prob > |T|	Prob > 0
1	108	0.880	5.804	0.0001	>0.01
2	252	1.488	15.369	0.0001	>0.01
3	288	2.253	17.856	0.0001	>0.01
4	162	3.123	12.790	0.0001	>0.01

Table V-16. Means for Number of Incorrect Responses for Category 1.

TRIAL	AVGINC	STDDEV
13	0.94444	1.51356
22	0.61111	1.14475
28	1.27778	1.99427
30	0.50000	0.92355
40	1.00000	1.32842
48	0.94444	2.23534

Table V-17. Means for Number of Incorrect Responses for Category 2.

TRIAL	AVGINC	STDDEV
1	2.33333	1.90973
5	1.44444	1.68810
10	2.38889	1.24328
18	1.00000	1.64496
20	1.88889	1.07861
21	2.22222	1.55509
23	1.72222	2.27015
24	0.94444	0.99836
26	1.16667	1.24853
41	1.22222	0.94281
42	1.27778	1.27443
44	1.38889	1.97451
46	0.88889	1.07861
47	0.94444	1.16175

Table V-18. Means for Number of Incorrect Responses for Category 3.

TRIAL	AVGINC	STDDEV
2	2.44444	2.20220
3	3.44444	2.54887
6	2.38889	2.00408
8	1.72222	1.99427
9	2.77778	1.89599
16	3.00000	1.45521
17	2.33333	1.41421
19	1.88889	2.13896
25	1.44444	1.85416
29	1.16667	0.98518
33	2.88889	3.19722
35	3.61111	3.18339
36	1.33333	1.90973
37	2.00000	0.97014
38	1.55556	1.73958
45	2.05556	2.12747

Table V-19. Means for Number of Incorrect Responses for Category 4.

TRIAL	AVGINC	STDDEV
4	2.88889	3.90910
7	2.88889	1.77859
12	2.83333	2.77064
15	3.50000	3.03412
27	2.22222	1.80051
31	5.61111	4.91264
32	2.61111	2.40433
34	2.66667	3.04863
39	2.88889	2.32351

The one-way analysis of variance performed on the trials within each category indicated that there was no significant difference between the six trials for category 1. The F-value was 0.57 and the critical value was 2.21 and the p-value was 0.729. Duncan's multiple range test verified this result. An F-value of 2.32 and a critical value of 1.72 was obtained for the 14 category 2 alarms. The p-value was 0.0066. Therefore at least two of the trials are different. Duncan's test placed the trials into four groups. These groupings are in Appendix C. Category 3 had 16 trials and the analysis of variance test yielded an F-value of 2.30 and a critical value of 1.67 with a p-value of 0.0043. Thus at least two of the trials are significantly different. Duncan's test grouped these trials into five groups with very little separation between any pair of groupings. The groupings are given in Appendix C. The final category, category 4, had an F-value of 1.91 and a critical value of 1.91. The p-value was 0.0161 and the R-square was 0.0910. This means there is no significant difference

between the nine trials but a probability of 0.062 of having at least one trial being different. Duncan's multiple range test did find one trial, trial 31, to be significantly different at the 0.05 alpha level.

An examination of the data as presented in Figure V-5 suggested the same model be used for the number of incorrect responses as was used in the total number of responses evaluation. This was a logarithmic model of the form $y = \alpha + \beta \ln x$. The results of the regression analysis using this model were an F-ratio of 111.9 and with a p-value of 0.0001. The value of R-square was 0.122. The values for α and β were 0.215 and 0.249, respectively.

A side-by-side comparison of the means of these three performance measures is given in Figures V-6 through V-9. These figures show the magnitude of these means increasing as the number of signals increases. It also shows the ratio of the number of incorrect responses to the number of responses increasing as the number of signals increases.

The responses the subjects made to the alarms in each category will also indicate which alarm or set of alarms would make the best alarm system for each category. The results from the one-way analysis of variance procedure and Duncan's multiple range test have been summarized in Tables V-20 through V-23. The alarm systems have been ranked for each of the performance measures for each category. Within each measure the alarm systems are grouped together if they are not significantly different. The system that consistently ranks the best in each measure would be chosen as the best system for that category. By examining the characteristics of the alarms in the chosen system, some indication of what makes the best alarms can be determined.

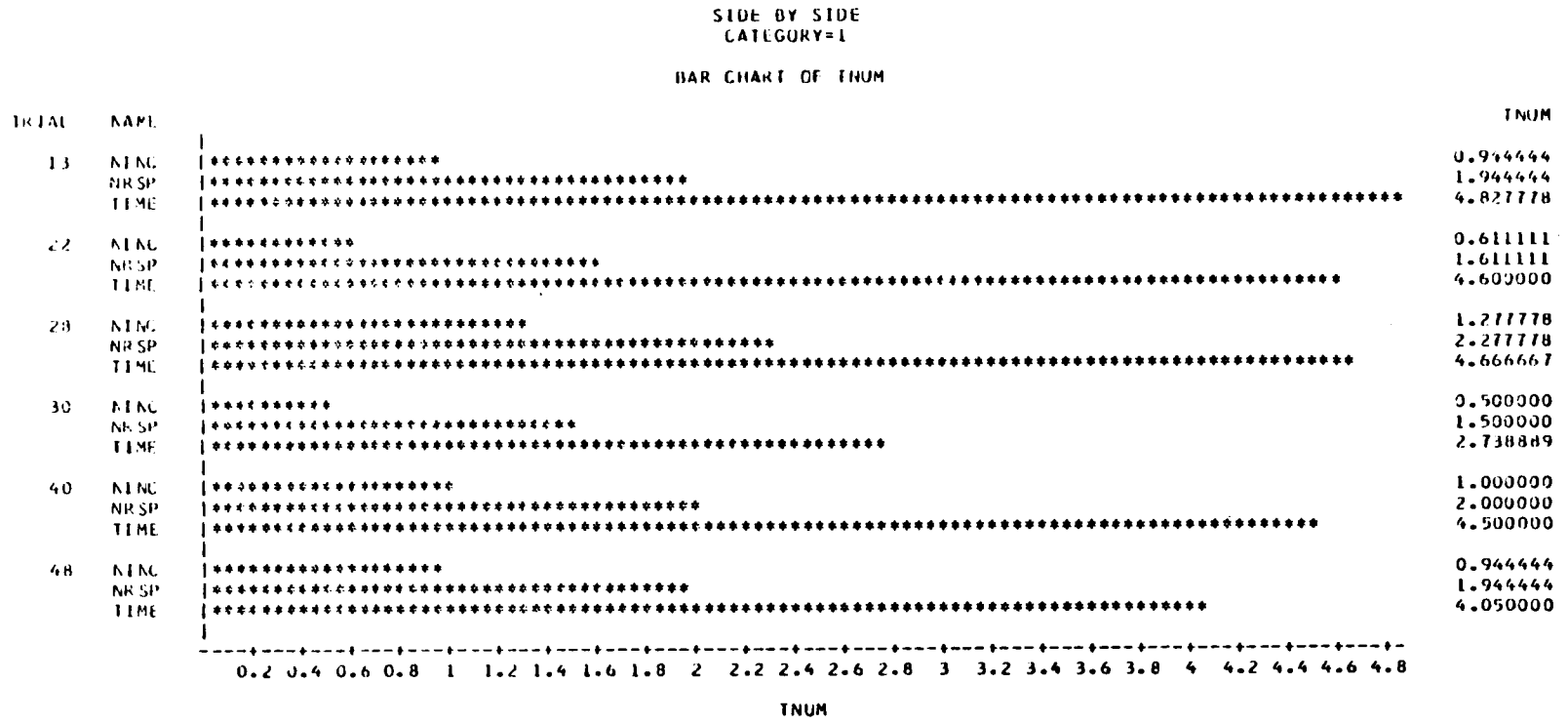


Figure V-6. Bar Chart of Response Time, Number of Responses, and Number of Incorrect Responses for Category 1.

SIDE BY SIDE
CATEGORY=2

BAR CHART OF TNUM

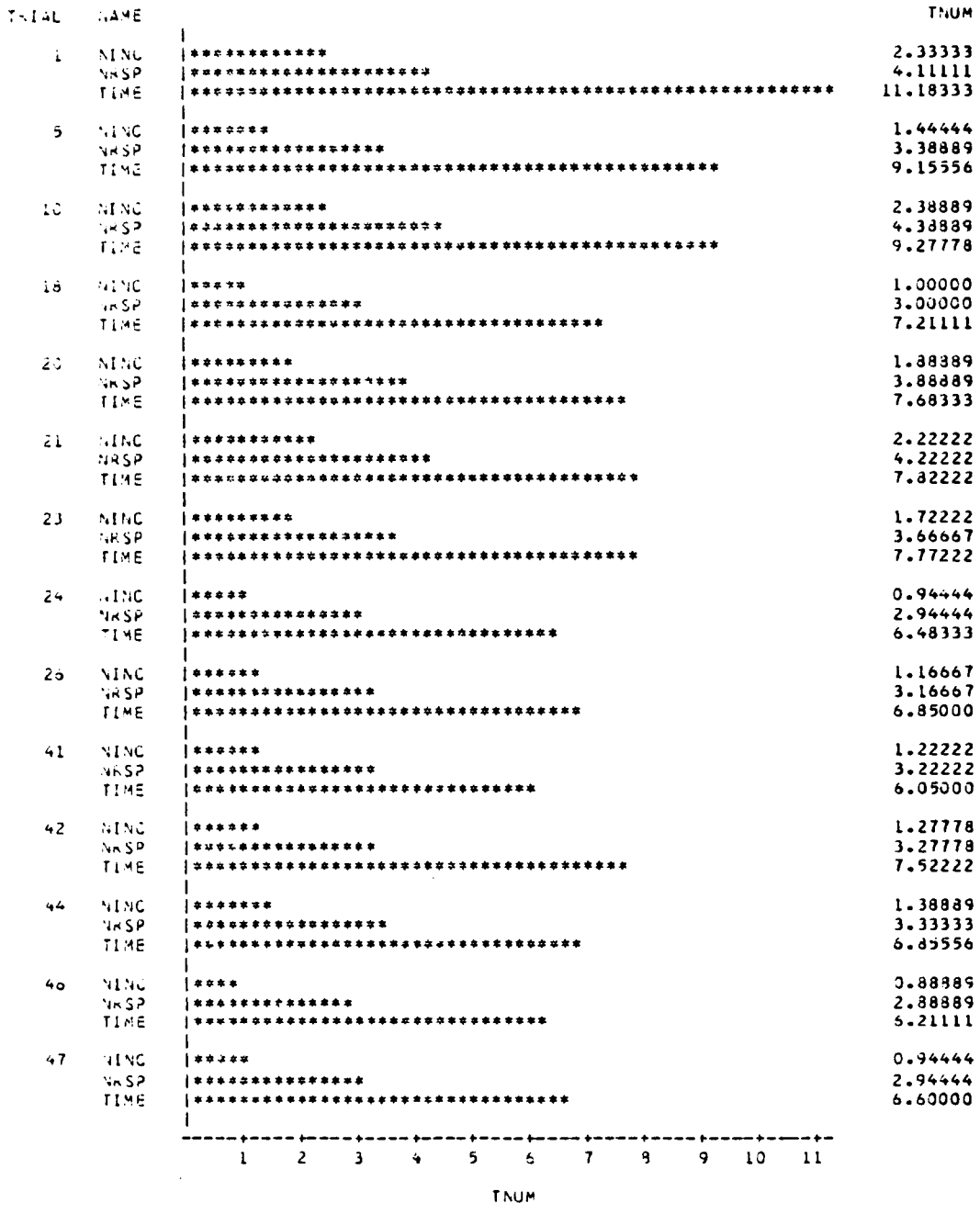


Figure V-7. Bar Chart of Response Time, Number of Responses, and Number of Incorrect Responses for Category 2.

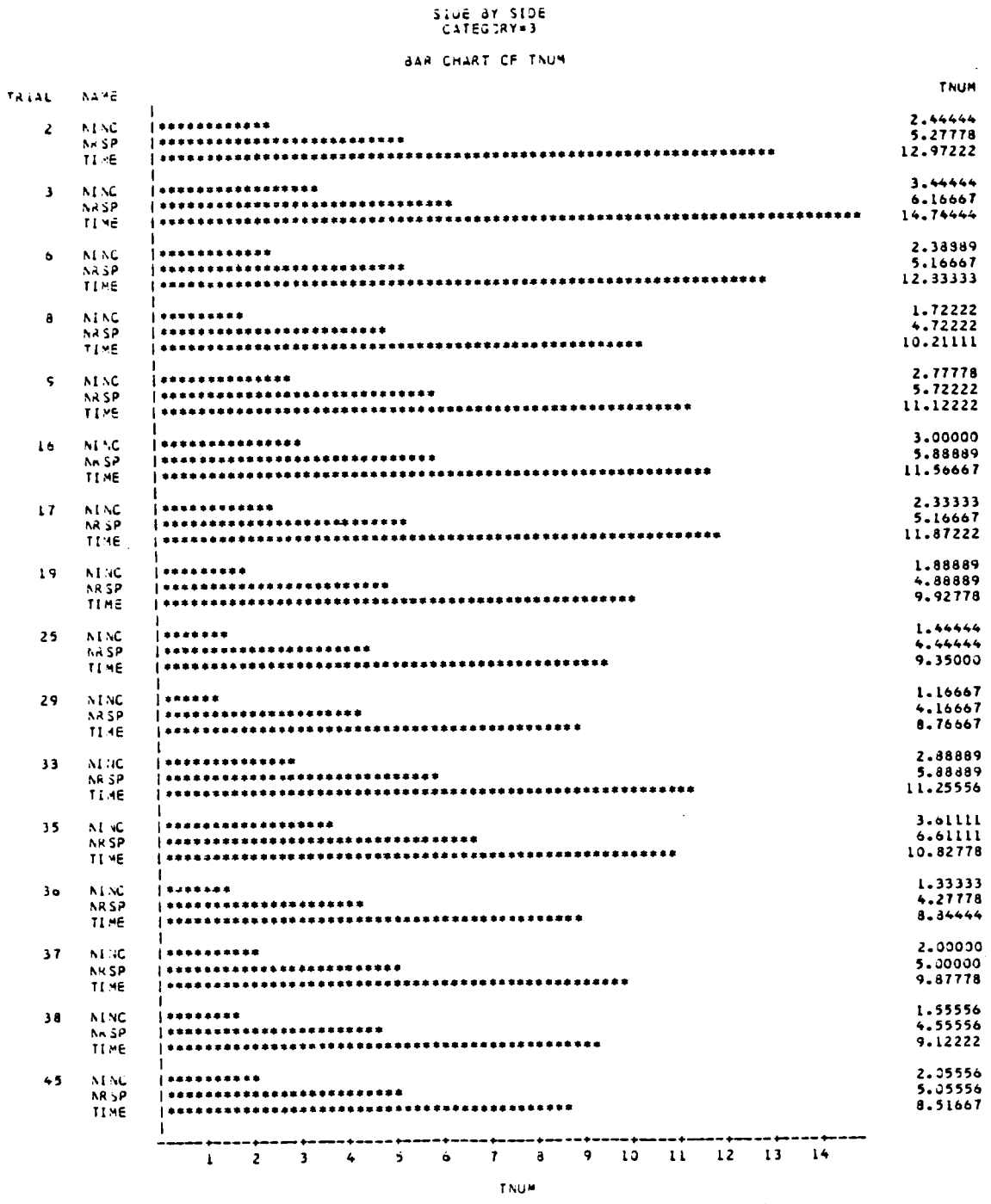


Figure V-8. Bar Chart of Response Time, Number of Responses, and Number of Incorrect Responses for Category 3.

SIDE BY SIDE
CATEGORY=4

BAR CHART OF TNUM

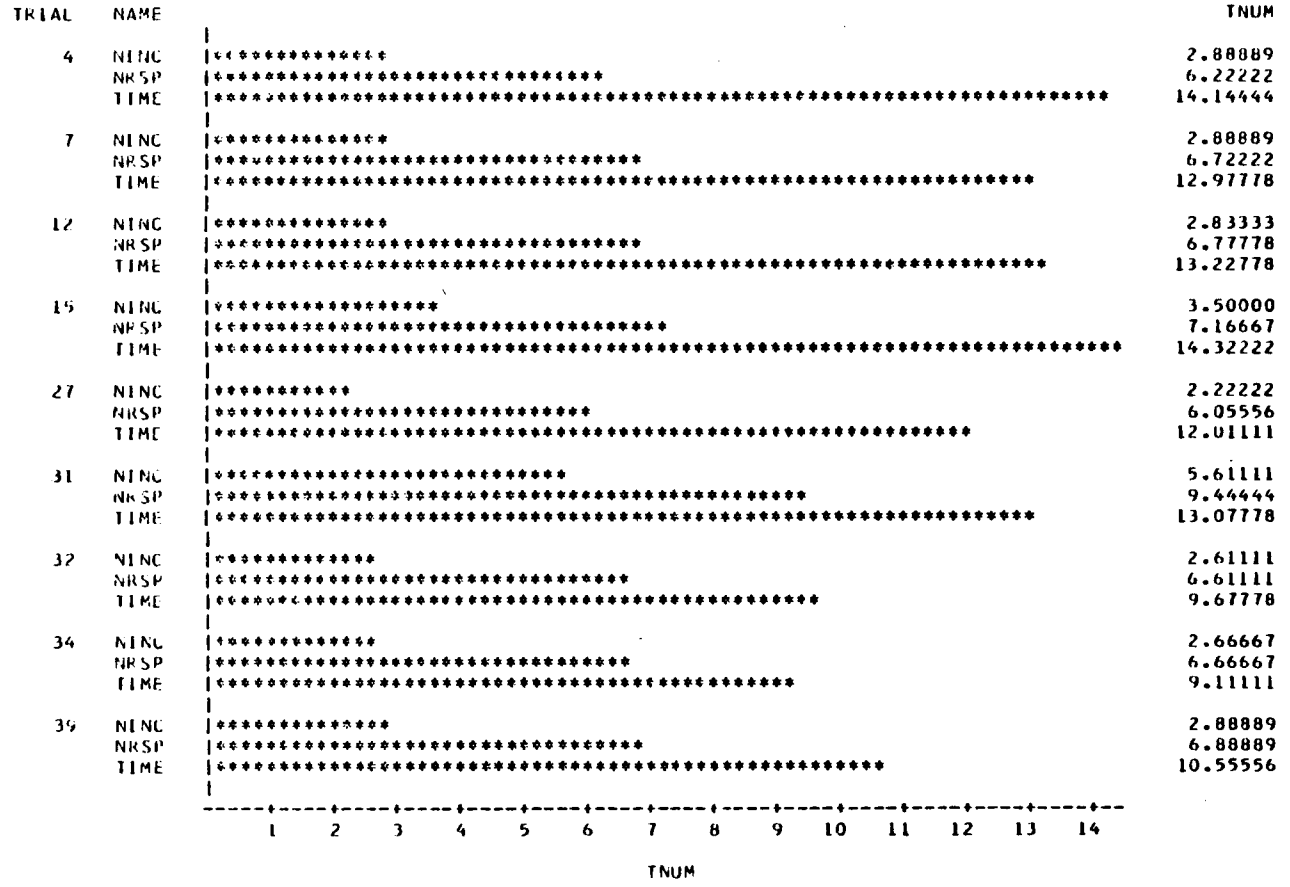


Figure V-9. Bar Chart of Response Time, Number of Responses, and Number of Incorrect Responses for Category 4.

Table V-20. Comparison of 1 Alarm System for All Performance Measures.

Reaction Time				Total Number of Responses				Total Number of Incorrect Responses			
Trials	Alarms	Mean	Grouping	Trials	Alarms	Mean	Grouping	Trials	Alarms	Mean	Grouping
30	6	2.74		30	6	1.50		30	6	.50	
48	4	4.05		22	1	1.61		22	1	.61	
40	3	4.50		48	4	1.94		48	4	.94	
22	1	4.60		13	5	1.94		13	5	.94	
28	2	4.67		40	3	2.00		40	3	1.00	
13	5	4.83		28	2	2.78		28	2	1.28	

Table V-21. Comparison of 2 Alarm Systems for All Performance Measures.

Reaction Time				Total Number of Responses				Total Number of Incorrect Responses			
Trials	Alarms	Mean	Grouping	Trials	Alarms	Mean	Grouping	Trials	Alarms	Mean	Grouping
41	2,3	6.05		46	2,4	2.89		46	2,4	.89	
46	2,4	6.21		47	1,4	2.94		47	1,4	.94	
24	3,6	6.48		24	3,6	2.94		24	3,8	.94	
47	1,4	6.60		18	1,6	3.00		18	1,6	1.00	
26	3,5	6.85		26	3,5	3.16		26	3,5	1.17	
44	1,3	6.86		41	2,3	3.22		41	2,3	1.22	
18	1,6	7.21		42	2,5	3.28		42	2,5	1.28	
42	2,5	7.52		44	1,3	3.33		44	1,3	1.39	
20	3,4	7.68		5	2,6	3.38		5	2,6	1.44	
23	1,2	7.77		23	1,2	3.67		23	1,2	1.72	
21	4,5	7.82		20	3,4	3.88		20	3,4	1.89	
5	2,6	9.16		1	1,5	4.11		21	4,5	2.22	
10	4,6	9.28		21	4,5	4.22		1	1,5	2.23	
1	1,5	11.18		10	4,6	4.38		10	4,6	2.39	

Table V-22. Comparison of 3 Alarm Systems for All Performance Measures.

Reaction Time				Total Number of Responses				Total Number of Incorrect Responses			
Trials	Alarms	Mean	Grouping	Trials	Alarms	Mean	Grouping	Trials	Alarms	Mean	Grouping
45	2,4,5	8.52		29	2,4,6	4.17		29	2,4,6	1.17	
29	2,4,6	8.77		36	1,3,6	4.28		36	1,3,6	1.33	
36	1,3,6	8.84		25	1,2,6	4.44		25	1,2,6	1.44	
38	1,3,5	9.12		38	1,3,5	4.56		38	1,3,5	1.56	
25	1,2,6	9.35		8	2,3,5	4.72		8	2,3,5	1.72	
37	3,4,6	9.88		19	2,3,6	4.89		19	2,3,6	1.89	
19	2,3,6	9.93		37	3,4,6	5.00		37	3,4,6	2.00	
8	2,3,5	10.21		45	2,4,5	5.06		45	2,4,5	2.06	
35	1,2,3	10.82		17	2,3,4	5.17		17	2,3,4	2.33	
9	1,2,5	11.12		6	1,4,6	5.17		6	1,4,6	2.39	
33	1,3,4	11.26		2	1,2,4	5.28		2	1,2,4	2.44	
16	1,4,5	11.57		9	1,2,5	5.72		9	1,2,5	2.78	
17	1,3,4	11.87		33	1,3,4	5.89		33	1,3,4	2.89	
6	1,4,6	12.83		16	1,4,5	5.89		16	1,4,5	3.00	
2	1,2,4	12.97		3	3,4,5	6.17		3	3,4,5	3.44	
3	3,4,5	14.74		35	1,2,3	6.61		35	1,2,3	3.61	

Table V-23. Comparison of 4 Alarm Systems for All Performance Measures.

Reaction Time				Total Number of Responses				Total Number of Incorrect Responses			
Trials	Alarms	Mean	Grouping	Trials	Alarms	Mean	Grouping	Trials	Alarms	Mean	Grouping
34	1,2,4,5	9.11		27	1,3,4,6	6.06		27	1,3,4,6	2.22	
32	1,3,4,5	9.69		4	1,2,3,5	6.22		32	1,3,4,5	2.61	
39	1,2,4,6	10.56		32	1,2,4,5	6.61		34	1,2,4,5	2.67	
27	1,3,4,6	12.01		34	1,2,4,5	6.67		12	2,3,4,6	2.83	
7	2,3,4,5	12.98		7	2,3,4,5	6.72		39	1,2,4,6	2.89	
31	1,2,3,4	13.08		12	2,3,4,6	6.78		7	2,3,4,5	2.89	
12	2,3,4,6	13.23		39	1,2,4,6	6.89		4	1,2,3,5	2.89	
4	1,2,3,5	14.14		15	1,2,3,6	7.17		15	1,2,3,6	3.50	
15	1,2,3,6	14.32		31	1,2,3,4	9.44		31	1,2,3,4	5.61	

For category 1 trial number 30 was the best for all three performance measures. This trial contained alarm 6, which was a 500 to 700 Hz.-74 dB-4 beats/sec. signal. Category 2 yielded trial number 46 as the best system. The alarms in this trial were alarm 4 and alarm 6. They differed in frequency and intensity. The frequencies were 1100 to 1900 Hz. and 500 to 700 Hz. respectively for the two alarms and the intensities were 82 dB and 74 dB. Trial 29 was the best for category 3. The alarms in this category were numbers 2, 4, and 6. Alarm 2 was a 1100 to 1900 Hz.-85 dB-4 beats/sec. signal. The only difference between it and alarm 4 was the intensity. The final category, category 4, had trial number 27 for the best system. The alarms in this system were 1, 3, 4, and 6. Alarms 4 and 6 were described above. Alarms 1 and 3 are both 500 to 700 Hz.-80 dB signals. Alarm 1 was 4 beats/sec. and alarm 3 was 8 beats/sec. It appears that the characteristics that varied the most in these best ranked systems were the frequency and the intensity. Modulation was a common value for all but one of the alarms.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

One of the questions to be answered in this project was how do people prioritize alarms that have different frequencies, amplitudes, and modulations? This was accomplished by conducting a paired-comparison test. The results of this test were used to develop a dominance matrix of Z-values. An ordered scale value was obtained for each alarm from these Z-values. This scale reflects the importance the test subjects placed on each alarm. By examining the physical characteristics of each of the alarms, it can be seen that alarms in the highest tonal frequencies were not chosen. The next observation is that the alarms with the lower intensities were rarely chosen. There was only a slight preference for alarms with 8 beats/sec. as compared to 4 beats/sec. Overall, the alarms with a low to intermediate tonal frequency, with a high intensity, and 4 beats/sec. were preferred. This translates into a frequency range of 500 to 2000 Hz with intensities in the 75 to 90 dB range and a modulation rate of 4 beats/sec. Therefore if preference for alarms were to be the criteria for selecting alarms, the alarms should have these characteristics.

A second question was what effect does alarm frequency, amplitude and modulation have on alarm response time in a paired comparison task? The alarms were ranked on a time scale where the alarm with the smallest overall mean response time derived from a dominance matrix of the mean response times for each pair across all subjects.

An examination of the physical characteristics of the alarms after they were ranked by mean response time shows the highest tonal frequency, highest intensity, fastest modulated alarm has the smallest response time. The second highest response time was an alarm that differed only in modulation. The next five alarms were well separated from the first two alarms and reasonably well from the bottom five. There is no trend or common characteristic for these alarms but they all have a response time within the 5.36 to 5.44 second range.

The balance of the alarms are evenly spaced on the lower end of the time scale. Each has either a high modulation coupled to a high intensity and an intermediate or low tonal frequency, or a high or intermediate tonal frequency paired to a low intensity. These observations suggest that sudden, intense alarms get a quick response as do those with a low rate of modulation. Alarms with a high rate of modulation and any combination of frequency and intensity seem to take longer for a choice to be made. Thus if response time were to be the criteria for selecting alarms, the alarms should be composed of frequencies in the 1700 to 3000 Hz range with intensities of 75 to 90 dB and a modulation rate of 4 beats/sec.

The third and last question to be investigated was how does the presence of multiple alarms in a recognition task affect errors, response time, and variability? This was done by conducting a relative discrimination task where 2, 3, or 4 simultaneous alarms were presented to a subject. Single alarms were also presented. The subject was to recognize which alarm or alarms were sounding and identify them by name in a priority sequence. The analysis of these data shows people

have increasing difficulty in identifying alarms as the number of alarms being presented increases. This was true for both of the dependent variables, responses, and response times. The number of responses, the number of errors, and the response times all increased as the number of alarms increased. This is illustrated in Figure VI-1 as the number of responses between each category in the response time for the same categories is decreasing. This can be explained by an examination of the number of times the 20 sec. time limit is reached for each category. The time limit was never reached for category 1 and for category 2 the limit was reached 5 times. This limit was reached 14 times in category 3 and 15 times in category 4. Also the variances for category 1 and two are the same but show rapid increase for categories 3 and 4.

The combination of increasing number of errors along with the rapid increases in their variances and the leveling or rollover of the response time indicates that the subjects were using trial and error to respond to the alarms in categories 3 and 4. Figure VI-2 represents what is believed to be the reactions of the subjects in this task.

The models that were tested for the data support this conclusion. The number of responses and the number of incorrect responses were best represented by a logorithmic model while a reciprocal model was the best for the response times.

Conclusions

It can be concluded that the presence of multiple alarms in a recognition task causes errors and response times to

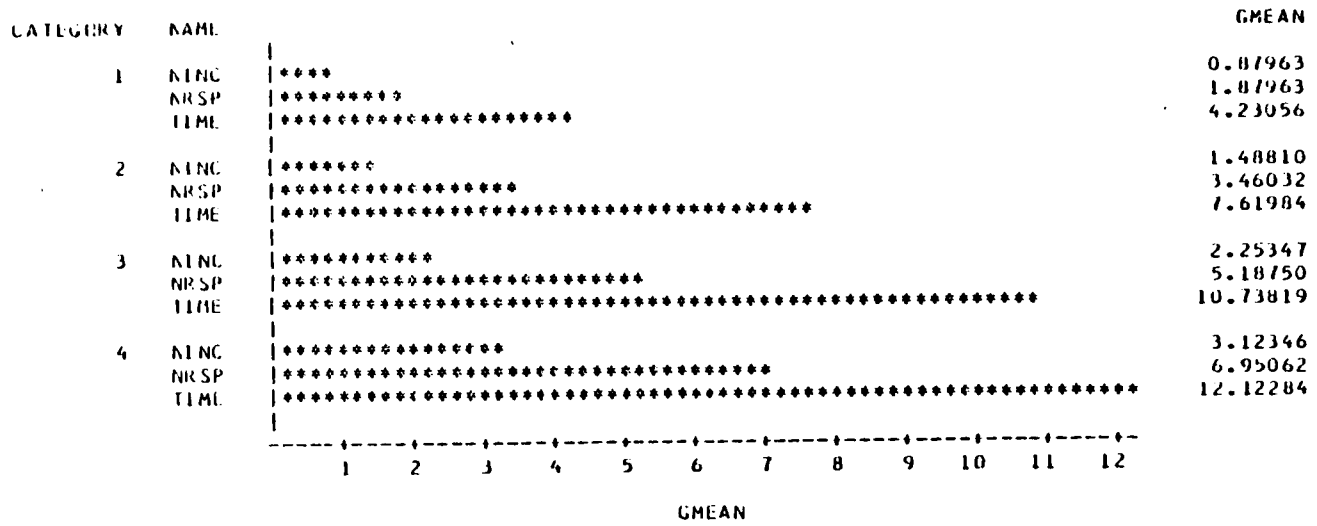


Figure VI-1. Bar Chart of Grand Means for Response Time, Number of Responses, and Number of Incorrect Responses.

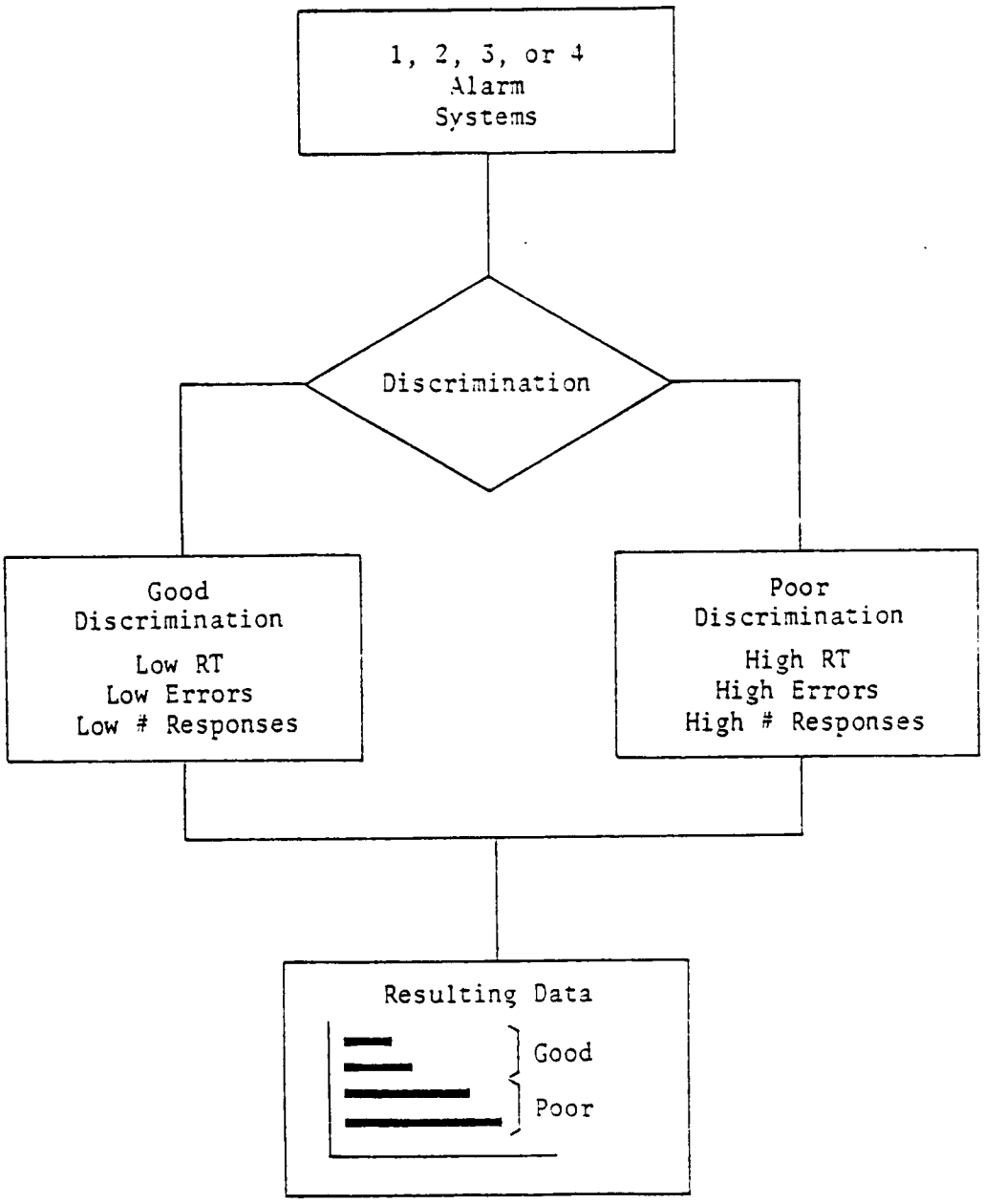


Figure VI-2. Flow Diagram of Subject Response.

increase as the number of alarms presented increase. The variability with which the subject responded also increased as the number of alarms increased. Because of this it was not possible to identify the alarms or the characteristics the alarms have to develop a system with more than two alarms. Furthermore, the use of multiple alarms with the coding that was used in these tasks does not appear to be a good way to present priority information to control room operators.

CHAPTER VII

RECOMMENDATIONS

Design Recommendations

The general principles and design recommendations that were given and discussed in Chapter II were found to be valid for designing auditory signals for a variety of applications. These principles and recommendations should be given careful study and consideration when designing auditory signals for warning alarms. A number of specific recommendations to adapt these principles and recommendations for a simultaneous alarm situation are given below:

- Restrict the number of alarms to two for alarms coded by frequency, intensity, and on-and-off modulation when simultaneous alarms could occur because people cannot adequately identify more alarms in a control room type situation.
- Use complex tones composed of two or more sinusoidal waveforms to have alarms that are easily recognizable as warning signals.
- Keep the frequencies as high as possible in the 500 to 3000 Hz range to command attention and obtain a fast response.
- Restrict intensities to the 70 to 95 dB range for safety and comfort provided the intensity is at least 30 dB above background.

- If on-and-off modulation is used, restrict the rate to near 4 beats/second. Rates of around 8 beats/second tend to irritate and confuse rather than aid in recognition.
- Use a minimum of three dimensions and vary at least 2 of the dimensions to code each alarm to make them very easy to recognize.

Research Recommendations

If it is desirable to have more than two coded multidimensional alarms, the following recommendations are made for future research:

- Perform more research into the basic effects that auditory alarms, coded by frequency, intensity, and modulation have on the ability of people to recognize and identify alarms in a control room environment.
- Use different shaped envelopes for frequency modulation and/or amplitude modulation for some alarms in an attempt to increase the number of coded alarms that can be effectively used in a control room.
- Obtain more information on the additive and subtractive effects of sound waves to insure that the alarms are distinct and recognizable. This will prevent one signal from being masked by another signal as opposed to a signal being masked by random noise.
- Try making use of location coding to enhance the discrimination and recognition of the alarms.
- Try redundancy in coding, such as combining a visual display to obtain more control room realism in the experiments.

. Add a diagnostic task to simulate the control room with still more fidelity.

BIBLIOGRAPHY

BIBLIOGRAPHY

Baird, J. C., and Norma E. Fundamentals of Scaling and Psychophysics. John Wiley and Sons, Inc., New York, 1978.

Deatherage, B. H. Auditory and other sensory forms of information presentation, in H. P. VanCott and R. G. Kinkade (eds.), Human Engineering Guide to Equipment Design. U. S. Government Printing Office, Washington, D. C., 1972, Chapter 4.

Hutchingson, R. D. New Horizons for Human Factors in Design. McGraw-Hill Book Company, New York, 1981.

Hungerford, J. C., and K. W. Heathington. Conceptualization and Subjective Evaluation of Innovative Railroad-Highway Crossing Active Warning Devices. The University of Tennessee, Knoxville, Transportation Center, FHWA Report, September 1981.

McCormick, E. J. Human Factors in Engineering and Design. McGraw-Hill Book Company, New York, 1976.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, Vol. 63, pp. 81-97.

Mudd, S. A. The Scaling and Experimental Investigation of Four Dimensions of Pure Tone and Their Use in an Audio-Visual Monitoring Problem. Unpublished Ph.D. thesis, Purdue University, Lafayette, Ind., 1961.

Pollack, I. Comfortable listening levels for pure tones in quiet and noise, Journal of the Acoustical Society of America, 1952, Vol. 24, p. 158.

Pollack, I. The information of elementary auditory displays, Journal of the Acoustical Society of America, 1953, Vol. 25, p. 745.

Pollack, I., and L. Ficks. Information of elementary multidimensional auditory displays, Journal of the Acoustical Society of America, 1954, Vol. 26, pp. 155-158.

Pols, L. C. W. Hearing, speech and auditory displays, in Kraiss and Moraal, Introduction to Human Engineering. Verlag TUV Rheinland GmbH, Koln, Germany, 1976, p. 170.

Shipley, W. C., J. J. Coffin, and K. C. Hadsell. Affective distance and other factors determining reaction time in judgments of color preference, Journal of Experimental Psychology, 1945, Vol. 35, pp. 206-215.

Shower, E. G., and R. Biddulph. Threshold of hearing and equal-loudness relations for pure tones, and the loudness function, Journal of the Acoustical Society of America, 1931, Vol. 3, pp. 257-287.

Snodgrass, J. G. Psychophysics, in Bertran Scharf and G. S. Reynolds (eds.). Experimental Sensory Psychology. Scott, Foresman and Company, Glenview, Ill., 1975, Chapter 2.

APPENDIXES

APPENDIX A

PLOTS AND ANOVA FOR RESPONSE TIME

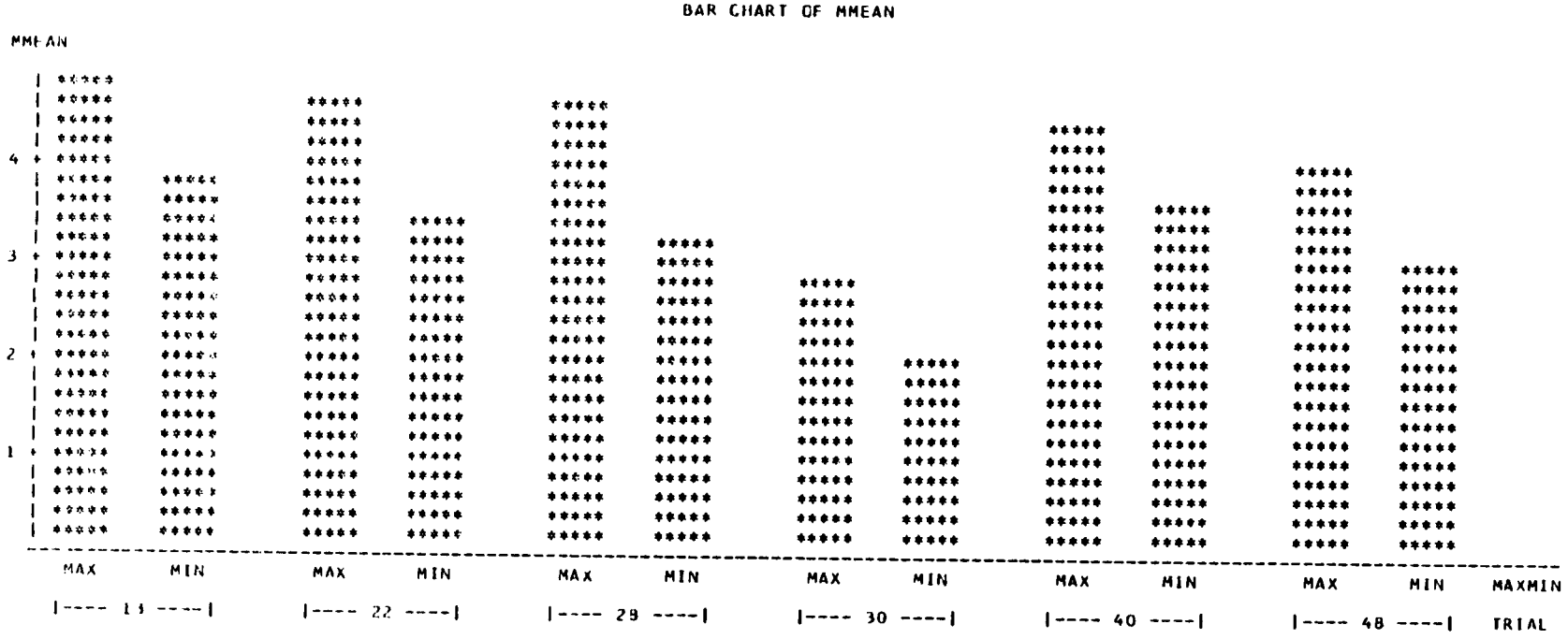


Figure A-1. Means for Maximum and Minimum Response Times for Category 1.

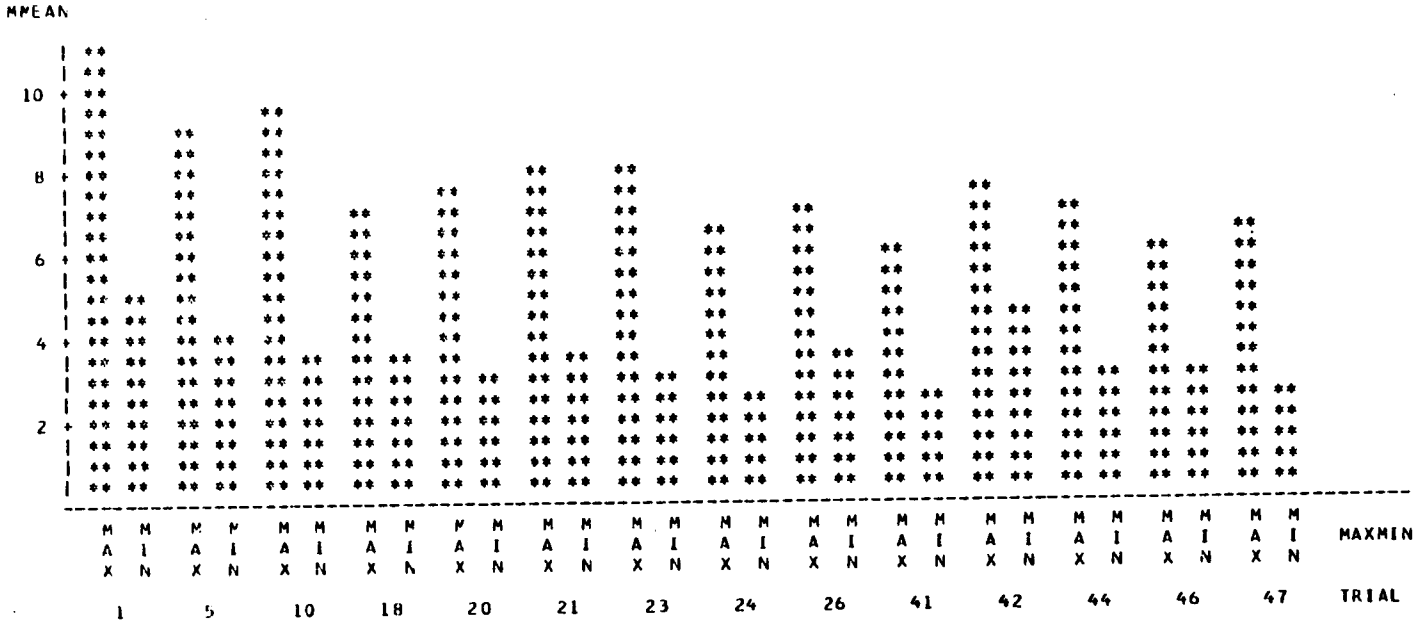


Figure A-2. Means for Maximum and Minimum Response Times for Category 2.

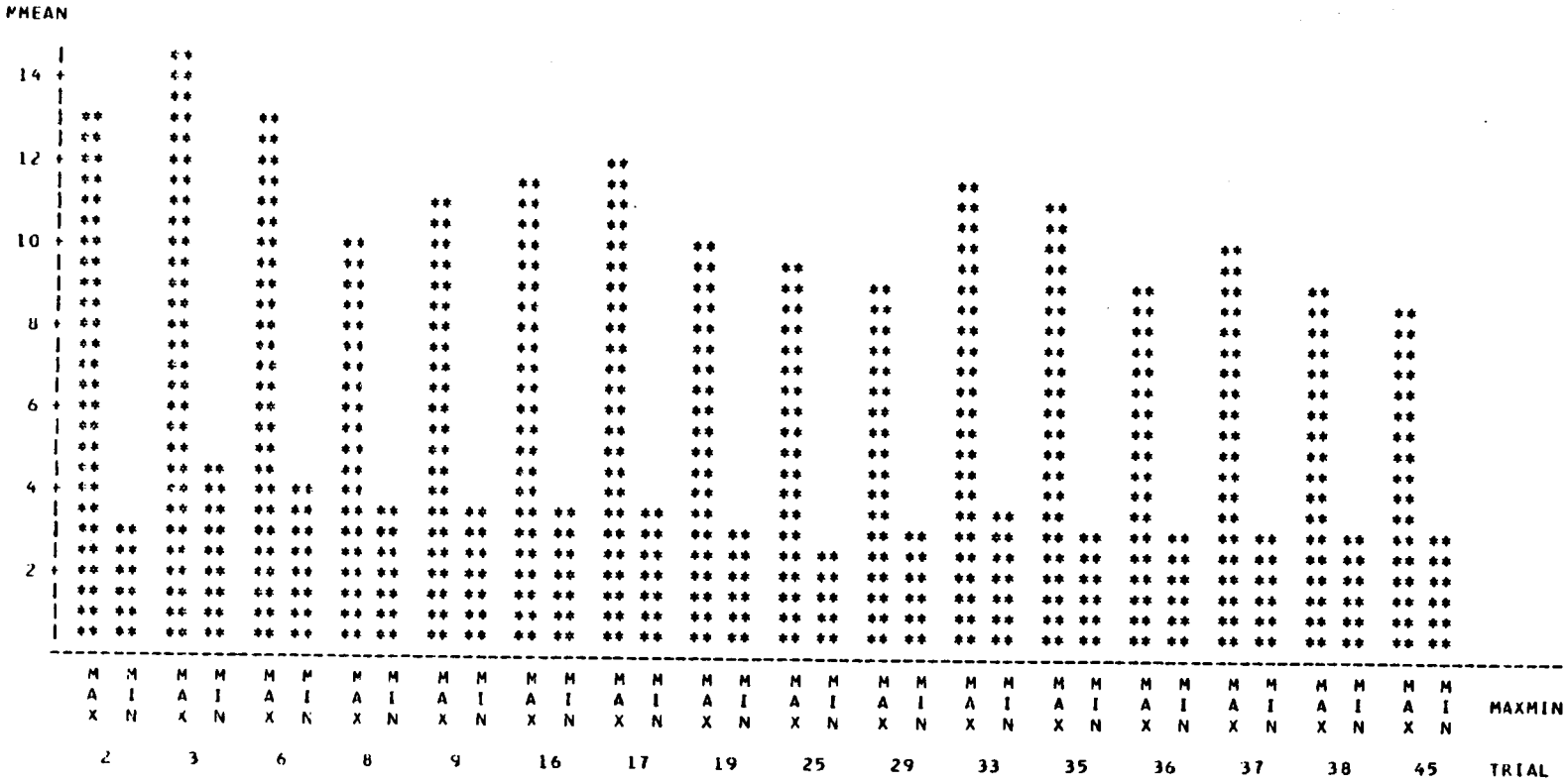


Figure A-3. Means for Maximum and Minimum Response Times for Category 3.

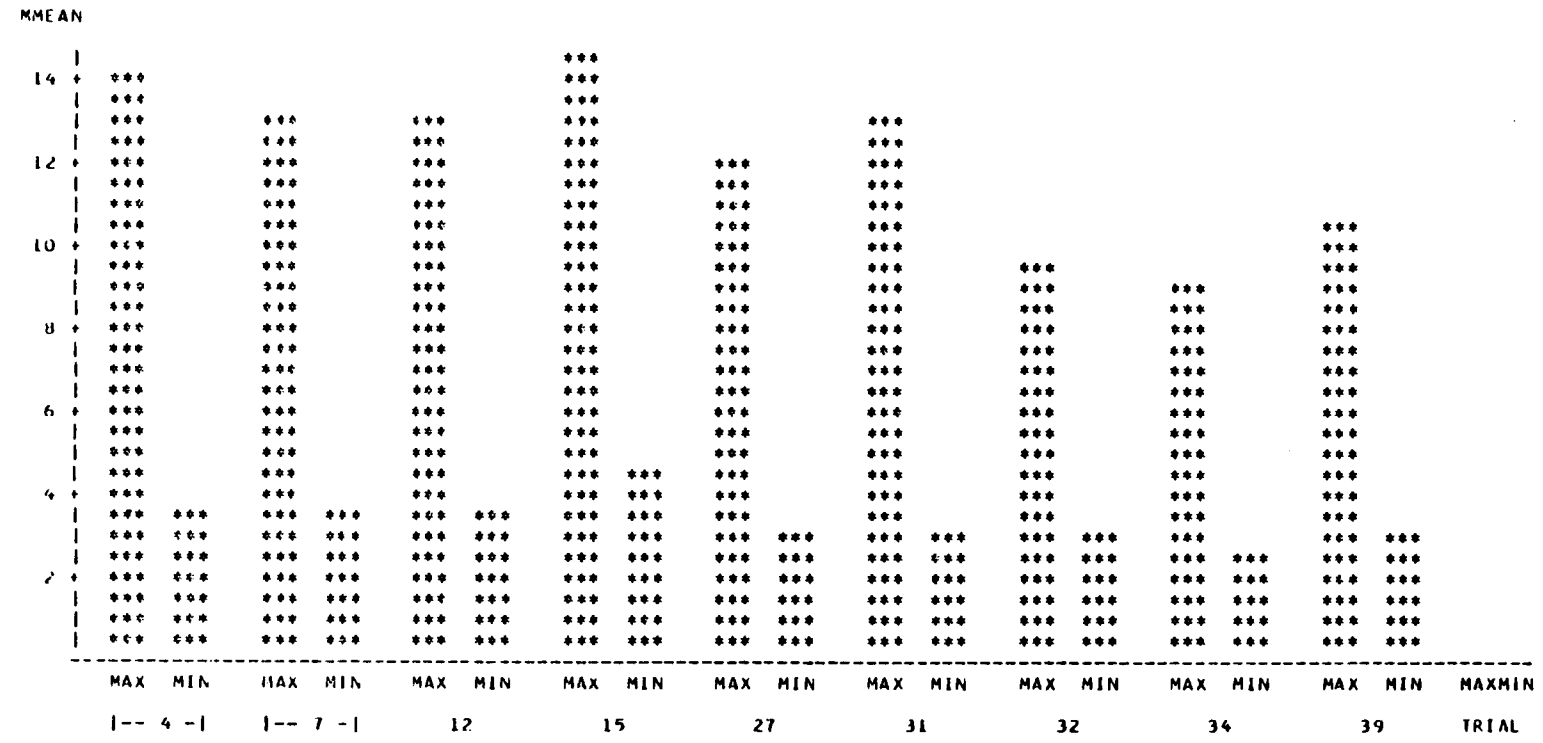


Figure A-4. Means for Maximum and Minimum Response Times for Category 4.

DEPENDENT VARIABLE: MAXTIME

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	54.24527778	10.84905556	1.33	0.2580	0.061073	67.5890
ERROR	102	833.96388889	8.17611656		STD DEV		MAXTIME MEAN
CORRECTED TOTAL	107	888.20916667			2.85939094		4.23055556

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	5	54.24527778	1.33	0.2580

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	6	13 22 28 30 40 48

NUMBER OF OBSERVATIONS IN BY GROUP = 108

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE MAXTIME

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=102 MS=8.17612

GROUPING	MEAN	N	TRIAL
A	4.827778	18	13
A			
A	4.666667	18	28
A			
A	4.600000	18	22
A			
A	4.500000	18	40
A			
A	4.050000	18	48
A			
A	2.738889	18	30

Figure A-5. ANOVA for Category 1 Response Time.

DEPENDENT VARIABLE: MAXTIME

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	13	468.14190476	36.01091575	2.79	0.0010	0.132383	47.1198
ERROR	238	3068.13888889	12.89133987		STD DEV		MAXTIME MEAN
CORRECTED TOTAL	251	3536.28079365			3.59045121		7.61784127

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	13	468.14190476	2.79	0.0010

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	14	1 5 10 18 20 21 23 24 26 41 42 44 46 47

NUMBER OF OBSERVATIONS IN BY GROUP = 252

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE MAXTIME

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=238 MS=12.8913

GROUPING	MEAN	N	TRIAL
A	11.183333	18	1
A			
B A	9.277778	18	10
B A			
B A	9.155556	18	5
B			
B C	7.822222	18	21
B C			
B C	7.772222	18	23
B C			
B C	7.683333	18	20
B C			
B C	7.522222	18	42
B C			
B C	7.211111	18	18
B C			
B C	6.855556	18	44
B C			
B C	6.850000	18	26
B C			
B C	6.600000	18	47
B C			
B C	6.483333	18	24
C			
C	6.211111	18	46
C			
C	6.050000	18	41

Figure A-6. ANOVA for Category 2 Response Time.

DEPENDENT VARIABLE: MAXTIME

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	15	836.06097222	55.73739815	3.69	0.0001	0.169038	36.1994
ERROR	272	4109.93888889	15.11006944		STD DEV		MAXTIME MEAN
CORRECTED TOTAL	287	4945.99986111			3.88716728		10.73819444

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	15	836.06097222	3.69	0.0001

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	16	2 3 6 8 9 16 17 19 25 29 33 35 36 37 38 45

NUMBER OF OBSERVATIONS IN BY GROUP = 288

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE MAXTIME

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=272 MS=15.1101

GROUPING	MEAN	N	TRIAL
A	14.744444	18	3
A			
B A	12.972222	18	2
B A			
B A C	12.833333	18	6
B C			
B D C	11.872222	18	17
B D C			
B E D C	11.566667	18	16
B E D C			
B F D C	11.255556	18	33
B E D C			
B E D C	11.122222	18	9
B E D C			
B E D C	10.827778	18	35
B E D C			
B E D C	10.211111	18	8
E D C			
E D C	9.927778	18	19
E D C			
E D C	9.877778	18	37
E D			
E D	9.350000	18	25
E D			
E D	9.122222	18	38
E D			
E D	8.844444	18	36
	8.766667	18	29
	8.516667	18	45

Figure A-7. ANOVA for Category 3 Response Time.

DEPENDENT VARIABLE: MAXTIME

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	8	527.50049383	65.93756173	3.70	0.0006	0.162197	34.8106
ERROR	153	2724.72500000	17.80866013		STD DEV		MAXTIME MEAN
CORRECTED TOTAL	161	3252.22549383			4.22033082		12.12283951

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	8	527.50049383	3.70	0.0006

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	9	4 7 12 15 27 31 32 34 39

NUMBER OF OBSERVATIONS IN BY GROUP = 162

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE MAXTIME

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=153 MS=17.8087

GROUPING	MEAN	N	TRIAL
A	14.322222	18	15
A	14.144444	18	4
B A	13.227778	18	12
B A	13.077778	18	31
B A	12.977778	18	7
B A	12.011111	18	27
B C	10.555556	18	39
C	9.677778	18	32
C	9.111111	18	34

Figure A-8. ANOVA for Category 4 Response Time.

APPENDIX B

PLOTS AND ANOVA FOR NUMBER OF RESPONSES

DEPENDENT VARIABLE: NRSP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	7.15740741	1.43148148	0.57	0.7287	0.026965	84.6586
ERROR	102	258.27777778	2.53213508		STD DEV		NRSP MEAN
CORRECTED TOTAL	107	265.43518519			1.59126839		1.87962963

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	5	7.15740741	0.57	0.7287

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	6	13 22 28 30 40 48

NUMBER OF OBSERVATIONS IN BY GROUP = 108

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE NRSP

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=102 MS=2.53214

GROUPING	MEAN	N	TRIAL
A	2.277778	18	28
A			
A	2.000000	18	40
A			
A	1.944444	18	13
A			
A	1.944444	18	48
A			
A	1.611111	18	22
A			
A	1.500000	18	30

Figure B-1. ANOVA for Category 1 Number of Responses.

DEPENDENT VARIABLE: NRSP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	13	60.49206349	4.65323565	2.27	0.0079	0.110266	41.3861
ERROR	238	488.11111111	2.05088702		STD DEV		NRSP MEAN
CORRECTED TOTAL	251	548.60317460			1.43209183		3.46031746

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	13	60.49206349	2.27	0.0079

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	14	1 5 10 18 20 21 23 24 26 41 42 44 46 47

NUMBER OF OBSERVATIONS IN BY GROUP = 252

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE NRSP

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=238 MS=2.05089

GROUPING	MEAN	N	TRIAL
A	4.388889	18	10
A	4.222222	18	21
A	4.111111	18	1
A C	3.888889	18	20
A C	3.666667	18	23
A C	3.388889	18	5
A C	3.333333	18	44
C	3.277778	18	42
C	3.222222	18	41
C	3.166667	18	26
C	3.000000	18	18
D	2.944444	18	24
D	2.944444	18	47
D	2.888889	18	46

Figure B-2. ANOVA for Category 2 Number of Responses.

CATEGORY=3

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: NRSP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	15	133.98611111	8.93240741	2.19	0.0071	0.107717	38.9401
ERROR	272	1109.88888889	4.08047386		STD DEV		NRSP MEAN
CORRECTED TOTAL	287	1243.87500000			2.02001828		5.18750000

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	15	133.98611111	2.19	0.0071

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	16	2 3 6 8 9 16 17 19 25 29 33 35 36 37 38 45

NUMBER OF OBSERVATIONS IN BY GROUP = 288

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE NRSP

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=272 MS=4.08047

GROUPING	MEAN	N	TRIAL
A	6.611111	18	35
B A	6.166667	18	3
B A C	5.888889	18	16
B A C	5.888889	18	33
B D A C	5.722222	18	9
B D A C	5.277778	18	2
B D A C	5.166667	18	6
B D A C	5.166667	18	17
B D A C	5.055556	18	45
B D C	5.000000	18	37
B D C	4.888889	18	19
B D C	4.722222	18	8
D C	4.555556	18	38
D C	4.444444	18	25
D	4.277778	18	36
D	4.166667	18	29

Figure B-3. ANOVA for Category 3 Number of Responses.

DEPENDENT VARIABLE: NRSP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	8	141.82716049	17.72839506	2.06	0.0426	0.097435	42.1592
ERROR	153	1313.77777778	8.58678286		STD DEV		NRSP MEAN
CORRECTED TOTAL	161	1455.60493827			2.93032129		6.95061728

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	8	141.82716049	2.06	0.0426

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	9	4 7 12 15 27 31 32 34 39

NUMBER OF OBSERVATIONS IN BY GROUP = 162

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE NRSP

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=153 MS=8.58678

GROUPING	MEAN	N	TRIAL
A	9.444444	18	31
B	7.166667	18	15
B	6.888889	18	39
B	6.777778	18	12
B	6.722222	18	7
B	6.666667	18	34
B	6.611111	18	32
B	6.222222	18	4
B	6.055556	18	27

Figure B-4. ANOVA for Category 4 Number of Responses.

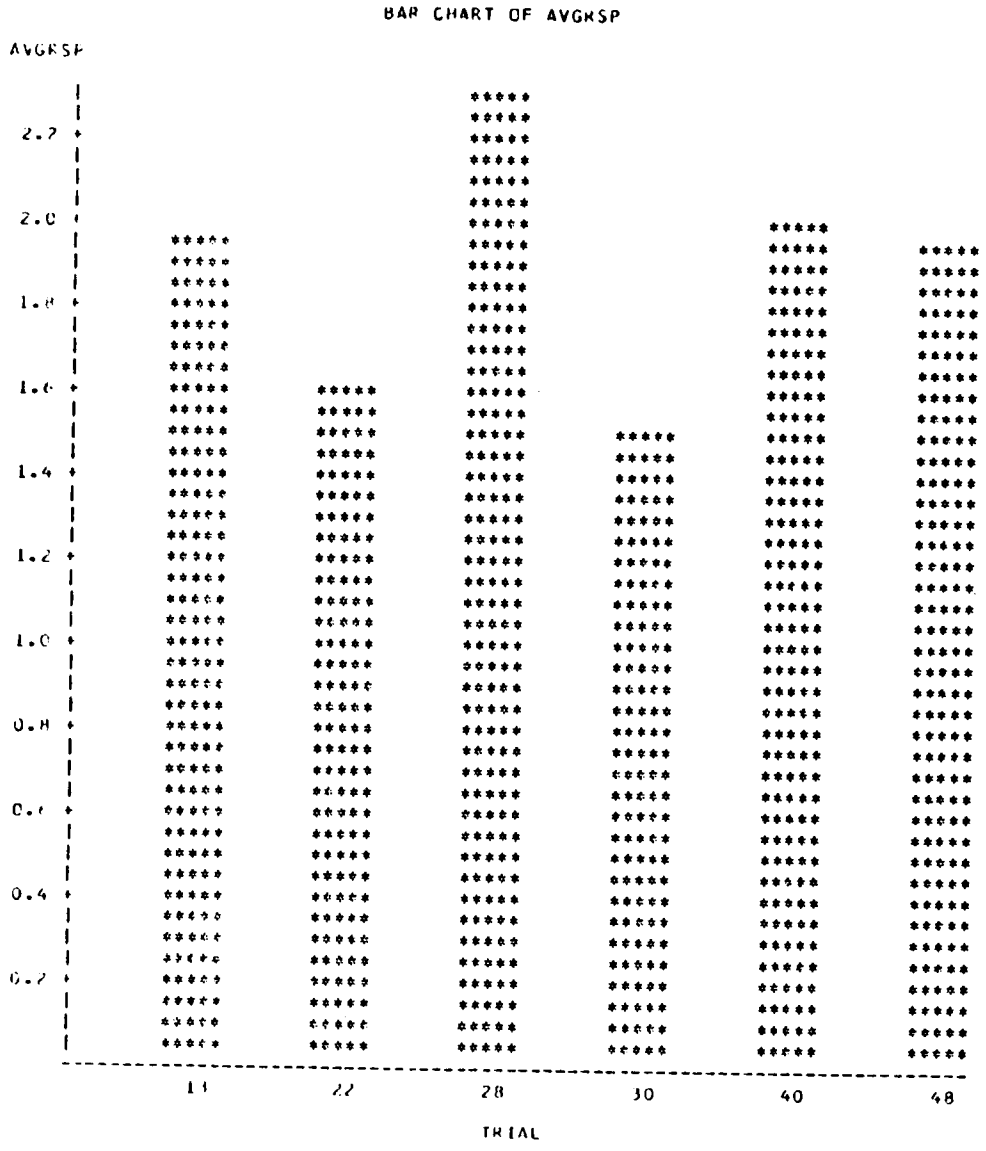


Figure B-5. Means for Number of Responses for Category 1.

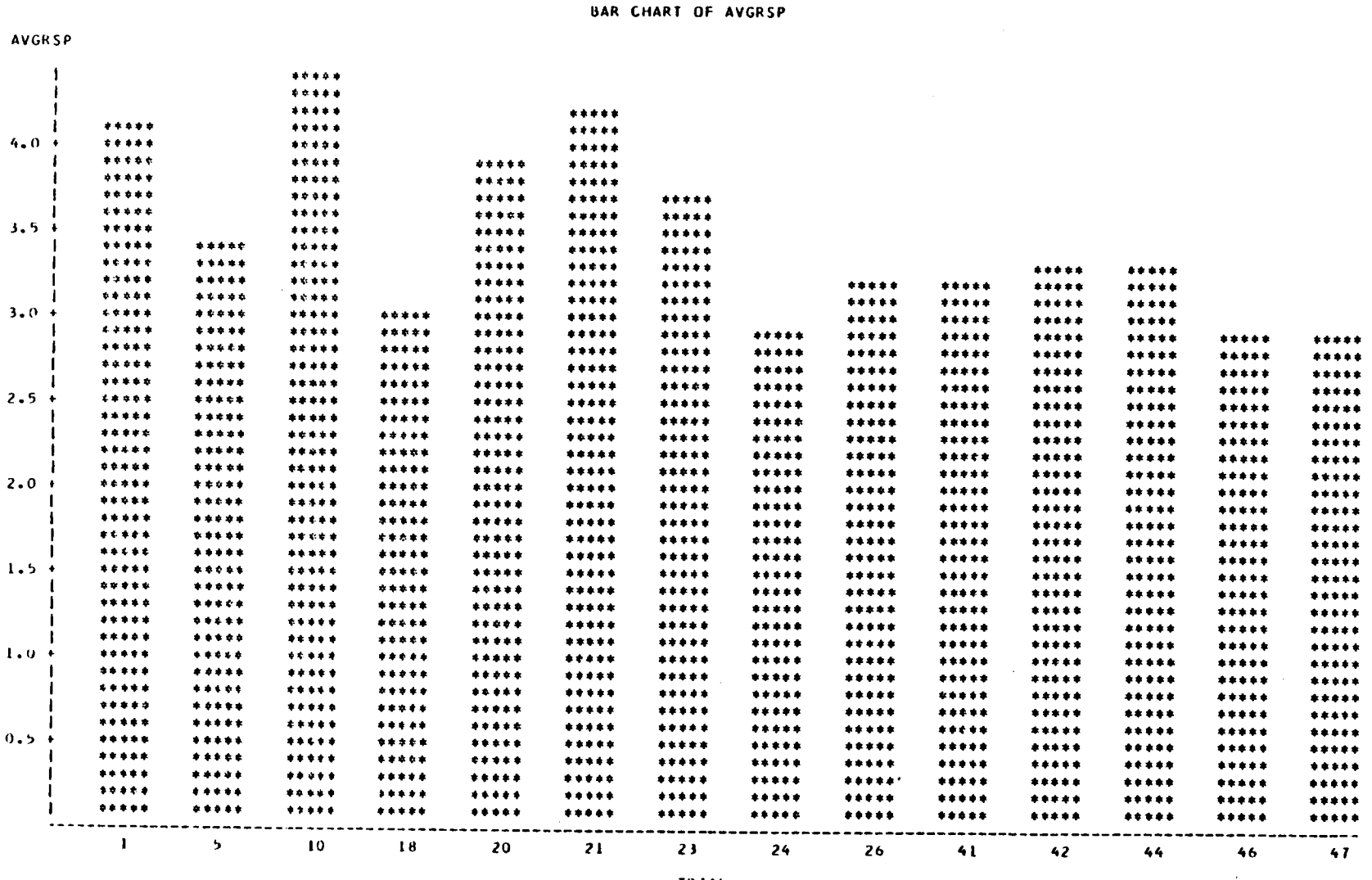


Figure B-6. Means for Number of Responses for Category 2.

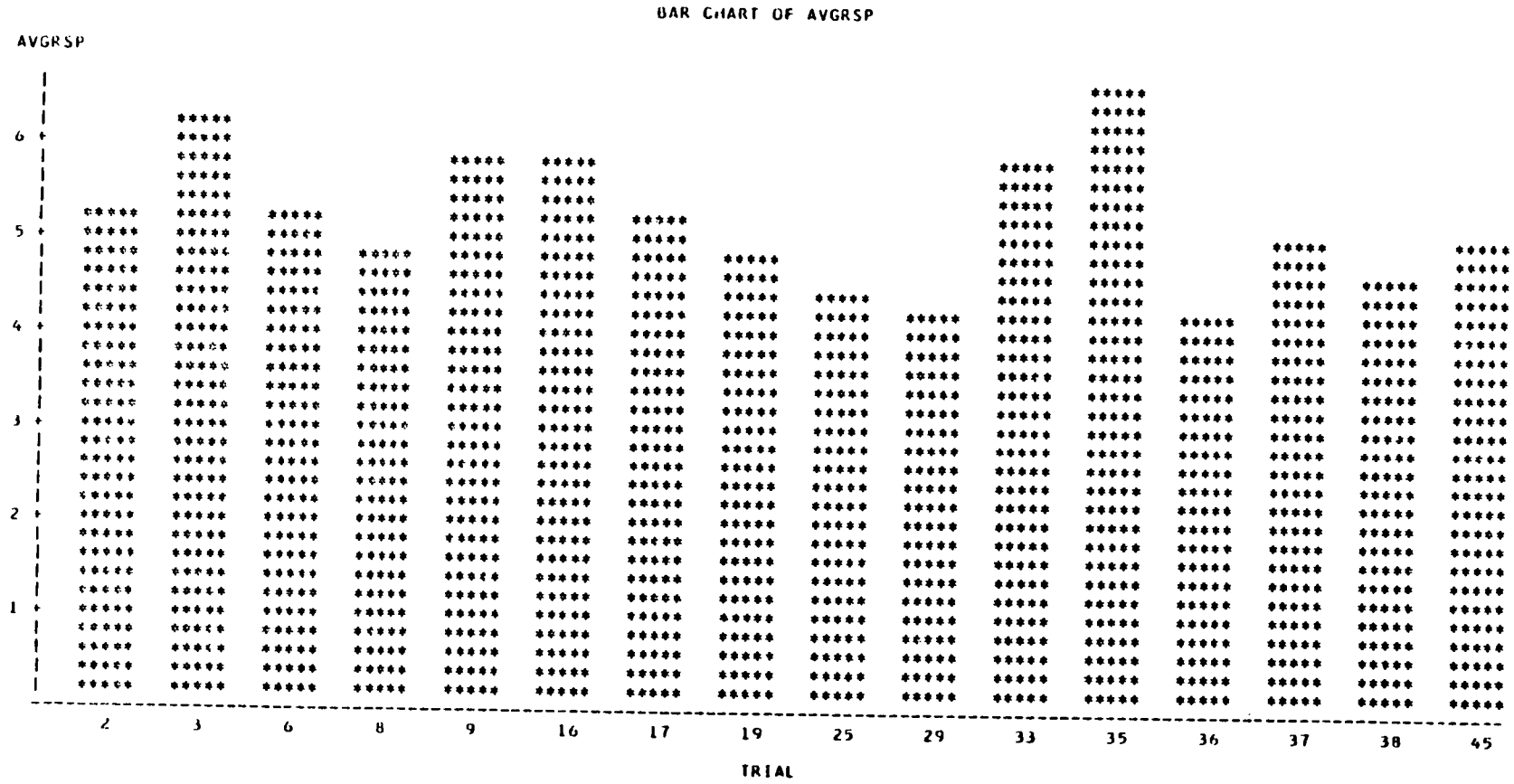


Figure B-7. Means for Number of Responses for Category 3.



Figure B-8. Means for Number of Responses for Category 4.

APPENDIX C

PLOTS AND ANOVA FOR NUMBER OF INCORRECT RESPONSES

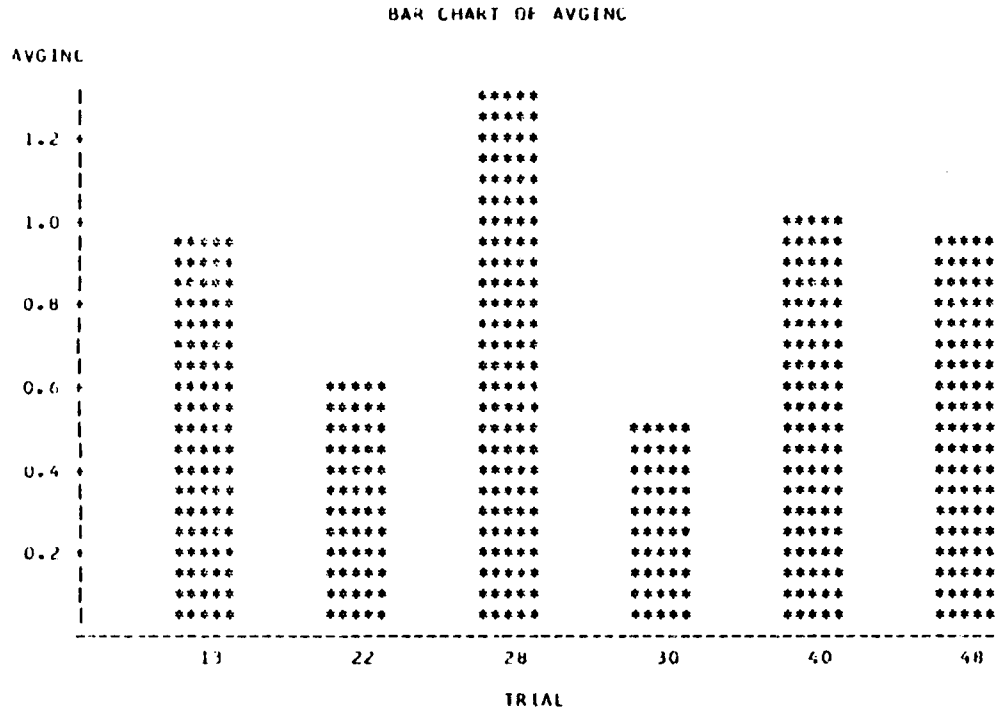


Figure C-1. Means for Number of Incorrect Responses for Category 1.

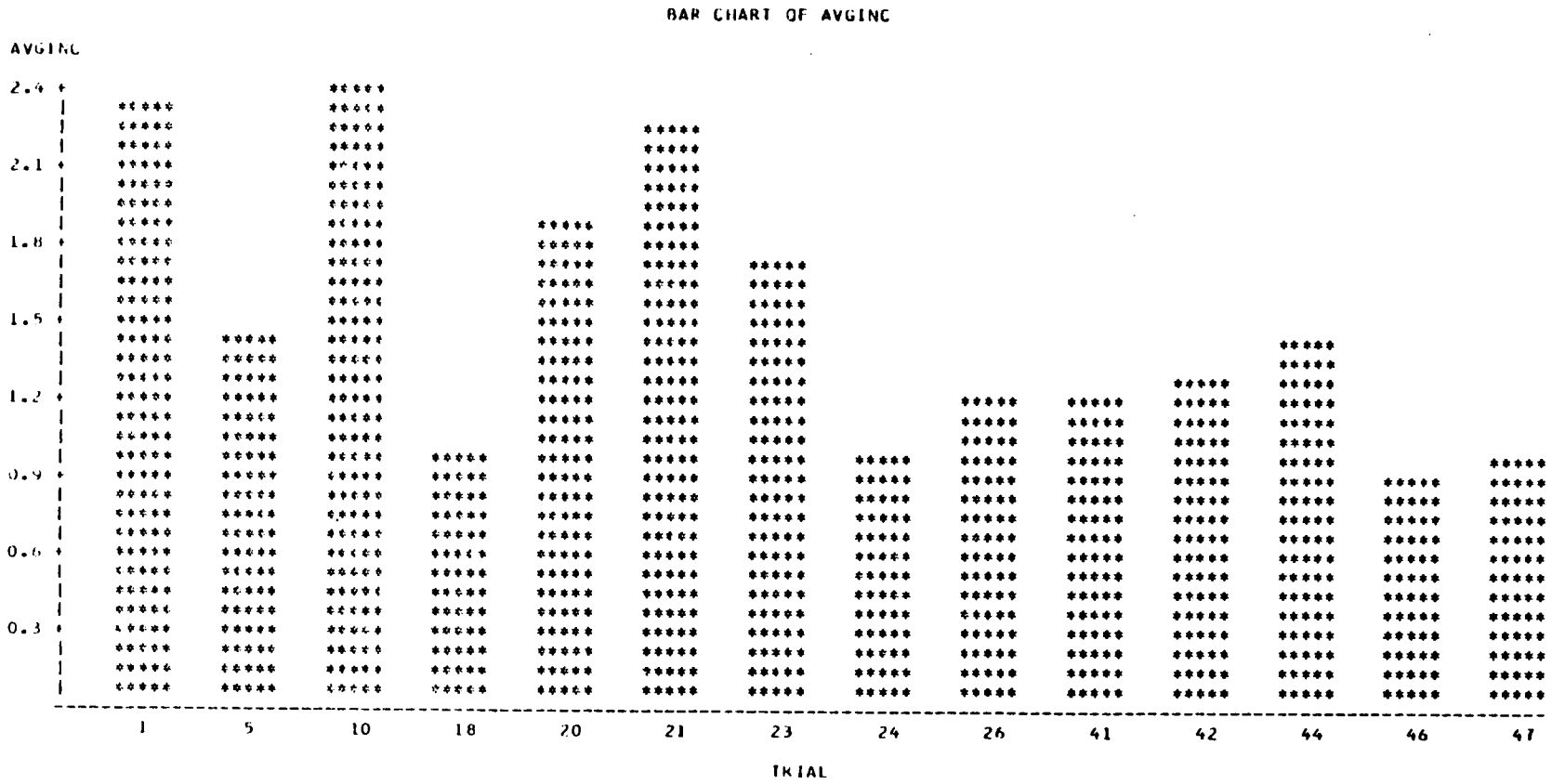


Figure C-2. Means for Number of Incorrect Responses for Category 2.

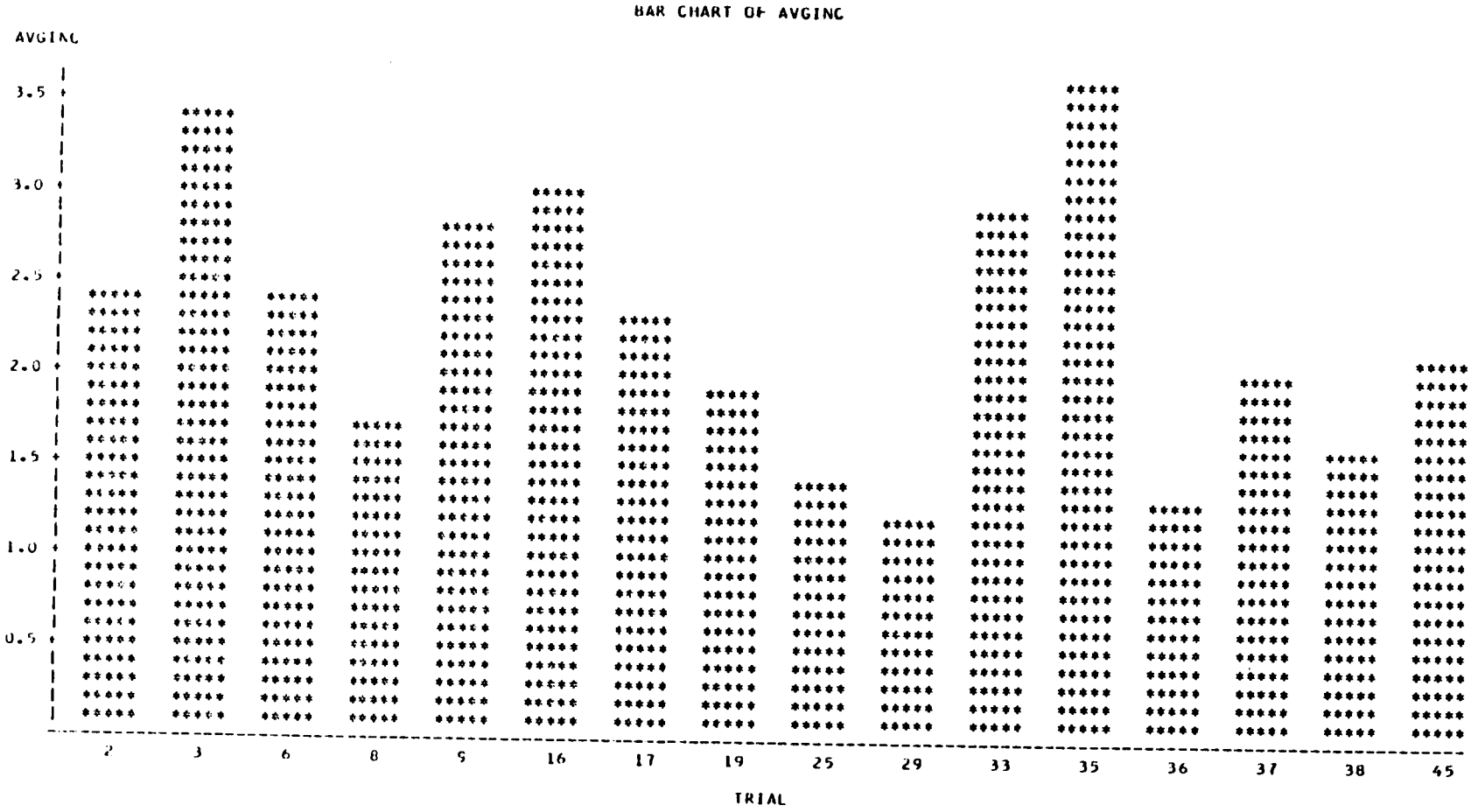


Figure C-3. Means for Number of Incorrect Responses for Category 3.

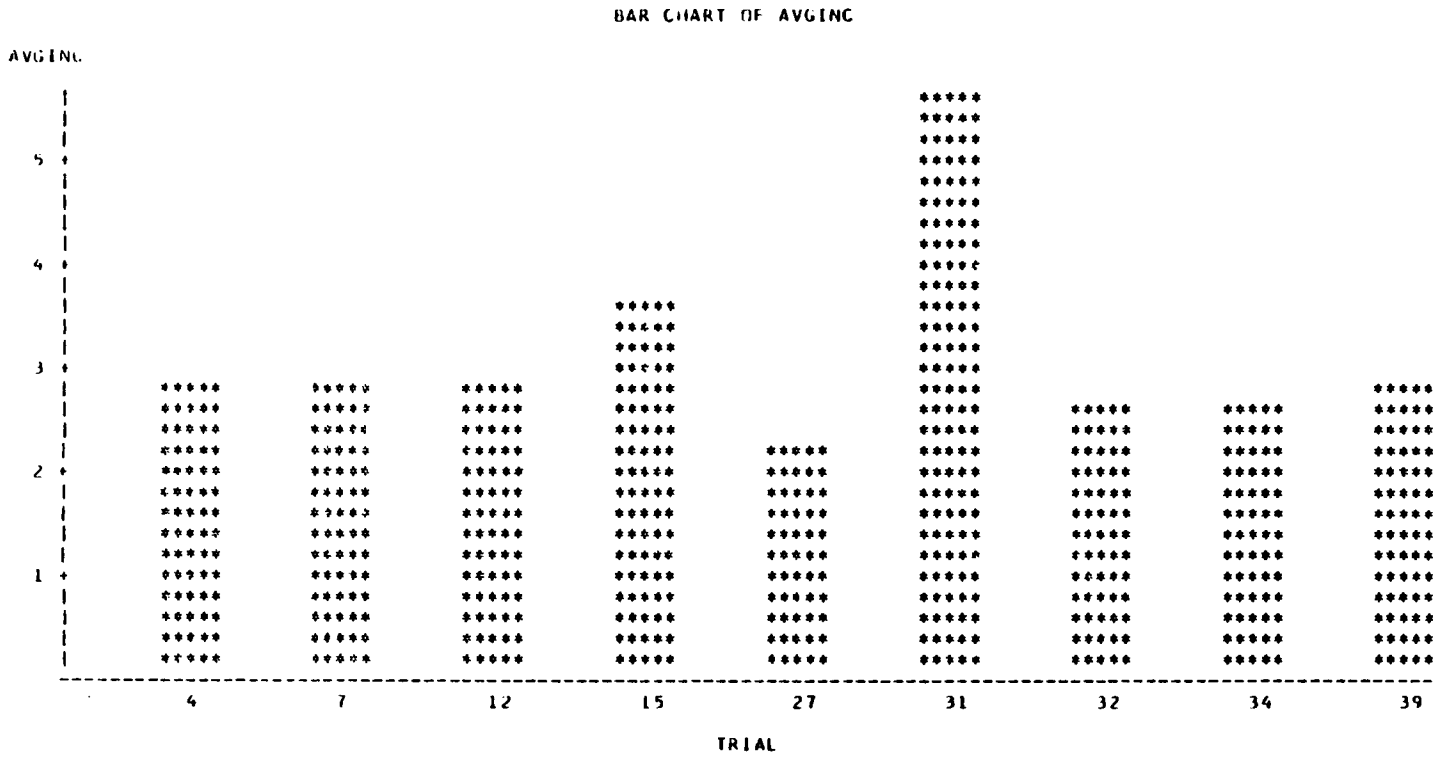


Figure C-4. Means for Number of Incorrect Responses for Category 4.

DEPENDENT VARIABLE: NINC

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	7.15740741	1.43148148	0.57	0.7287	0.026965	180.9021
ERROR	102	258.27777778	2.53213508		STD DEV		NINC MEAN
CORRECTED TOTAL	107	265.43518519			1.59126839		0.87962963

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	5	7.15740741	0.57	0.7287

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	6	13 22 28 30 40 48

NUMBER OF OBSERVATIONS IN BY GROUP = 108

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE NINC

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=102 MS=2.53214

GROUPING	MEAN	N	TRIAL
A	1.277778	18	28
A	1.000000	18	40
A	0.944444	18	13
A	0.944444	18	48
A	0.611111	18	22
A	0.500000	18	30

Figure C-5. ANOVA for Category 1 Number of Incorrect Responses.

DEPENDENT VARIABLE: NINC

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	13	66.57539683	5.12118437	2.32	0.0066	0.112276	99.9308
ERROR	238	526.38888889	2.21171802		STD DEV		NINC MEAN
CORRECTED TOTAL	251	592.96428571			1.48718460		1.48809524

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	13	66.57539683	2.32	0.0066

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	14	1 5 10 18 20 21 23 24 26 41 42 44 46 47

NUMBER OF OBSERVATIONS IN BY GROUP = 252

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE NINC

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=238 MS=2.21172

GROUPING	MEAN	N	TRIAL
A	2.388889	18	10
A	2.333333	18	1
A	2.222222	18	21
A	1.888889	18	20
A	1.722222	18	23
A	1.444444	18	5
A	1.388889	18	44
A	1.277778	18	42
B	1.222222	18	41
B	1.166667	18	26
B	1.000000	18	18
B	0.944444	18	24
B	0.944444	18	47
B	0.888889	18	46

Figure C-6. ANOVA for Category 2 Number of Incorrect Responses.

DEPENDENT VARIABLE: NINC

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	15	148.44097222	9.89606481	2.30	0.0043	0.112755	91.9592
ERROR	272	1168.05555556	4.29432190		STD DEV		NINC MEAN
CORRECTED TOTAL	287	1316.49652778			2.07227457		2.25347222

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	15	148.44097222	2.30	0.0043

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	16	2 3 6 8 9 16 17 19 25 29 33 35 36 37 38 45

NUMBER OF OBSERVATIONS IN BY GROUP = 288

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE NINC

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=272 MS=4.29432

GROUPING	MEAN	N	TRIAL
A	3.611111	18	35
A	3.444444	18	3
A	3.000000	18	16
B A	2.888889	18	33
B C A	2.777778	18	9
B C D A	2.444444	18	2
B C D A	2.388889	18	6
B C D A	2.333333	18	17
B C D A	2.055556	18	45
B C D A	2.000000	18	37
B C D	1.888889	18	19
C D	1.722222	18	8
C D	1.555556	18	38
C D	1.444444	18	25
D	1.333333	18	36
	1.166667	18	29

Figure C-7. ANOVA for Category 3 Number of Incorrect Responses.

DEPENDENT VARIABLE: NINC

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	8	141.53086420	17.69135802	1.91	0.0616	0.090986	97.3293
ERROR	153	1414.00000000	9.24183007		STD DEV		NINC MEAN
CORRECTED TOTAL	161	1555.53086420			3.04003784		3.12345679

SOURCE	DF	ANOVA SS	F VALUE	PR > F
TRIAL	8	141.53086420	1.91	0.0616

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRIAL	9	4 7 12 15 27 31 32 34 39

NUMBER OF OBSERVATIONS IN BY GROUP = 162

DUNCAN'S MULTIPLE RANGE TEST FOR VARIABLE NINC

MEANS WITH THE SAME LETTER ARE NOT SIGNIFICANTLY DIFFERENT.

ALPHA LEVEL=.05 DF=153 MS=9.24183

GROUPING	MEAN	N	TRIAL
A	5.611111	18	31
B	3.500000	18	15
B	2.388889	18	4
B	2.888889	18	7
B	2.888889	18	39
B	2.833333	18	12
B	2.666667	18	34
B	2.611111	18	32
B	2.222222	18	27

Figure C-8. ANOVA for Category 4 Number of Incorrect Responses.

APPENDIX D

HYPOTHETICAL IMPLEMENTATION OF RESEARCH FINDINGS

N. O. Nonsense, the chief engineer at the Manywatts Utility Company's newest nuclear power plant, Turkey Number One, which is under construction at Turkey Island, has called for a meeting with five of his systems engineers. These five engineers have responsibility for the design of the main control room of the new plant. "N.O." has been notified that the current design for the auditory alarm system for the control room is no longer acceptable because of some recent changes in regulations. A new alarm system must be designed and approved before construction of the control room can proceed and construction is already behind schedule.

Charlie Ezout immediately puts forth his pet design of a single warning signal that consists of a loud obnoxious horn or claxton. He maintains that it is simple, cheap, and very effective in getting the operating crew's attention. Not to be outdone, the team's newest engineer, Joseph C. Noseawl, proposes a more complex scheme. His proposal is to have a unique and separate alarm for each component in the plant that must be monitored. He recalls that one of his professors in a recently completed college course had stated that a person can distinguish and recognize thousands of sounds like voices and musical tones because each is made up of many components or dimensions. "N.O.," with his usual tact and understanding, carefully sidesteps the two extreme ideas and casually suggests the team get busy working on the problem.

After brainstorming for a couple of hours, the team has reached agreement for some of the design parameters for an auditory alarm system and has identified some problem areas. They have also bounced

around some ideas of how to apply an auditory alarm system to the main control room if they can come up with an acceptable alarm system.

The team has agreed that the auditory alarm system must supplement the visual displays; provide short prioritized messages that convey severity; have concise and meaningful structured phrasing; and be clear and distinct from all areas of the control but they must not startle the operators. They also agreed that the alarms should be multidimensional for the reasons Joe C. had put forth in his earlier suggestion.

In addition to setting these design parameters, one of the better suggestions for an alarm system was rejected. This suggestion had been to have only one alarm from a fairly large set of alarms sound at a time. The alarm that would sound would be the most urgent alarm as determined by an analysis of the various plant systems and components and arranging them in a prioritized hierarchy. The control of the system would have been a microprocessor or computer monitoring the input signals and then triggering the highest corresponding alarm and holding the others until the first alarm had been responded to. The idea was rejected because no one in the utilities or in the regulating agencies had ever been able to agree on what this hierarchy of systems should be. Therefore, they concluded their system would have to have multiple simultaneous alarms that would convey severity.

This led to some heated conversation over how many alarms would be needed and if the operators could handle the number of alarms that some of the ideas being suggested would require.

This discussion also identified another problem area. This was how to section or divide the plant and on what basis should this

division take place if an effective alarm system could be organized. The division could take place by system, function, priority, or even by location. Most of the group leans toward division by system, while at least one member wants to prioritize everything by a combination of system and function, and Charlie has been arguing for division by location.

The chief engineer finally divides the team into two groups. The first group is to determine how coded auditory alarm systems can be logically and reasonably designed for use in the reactor's main control room. The second group is to gather information from design handbooks and other technical information sources on human performance with coded auditory alarm systems in order to evaluate the designs the first group will develop.

After several days of diligent work, the two groups meet to review the ideas of the design group and the results of the research group.

The design group has come up with three possible ways of dividing the plant and then assigning coded auditory alarms to the divisions. These divisions could be ranked or ordered by priority, by system, by function, or by location, thus resolving some of the earlier arguments.

The first proposed design is to divide the plant into k divisions and assign each monitored phenomena or component to one of the divisions depending upon the importance of that phenomena or component. Then the "best" of k alarms would be assigned to the most important division and so on down to the least important division getting the least ranked alarm.

This proposal is similar to the one Joe put forth the day of the first meeting. The difference is that there are k alarms for k divisions rather than an alarm for each component or phenomena.

The second proposal is to divide the plant into j divisions and to assign all k alarms to each division. Each division would then be further divided into k subdivisions with one alarm assigned to each subdivision.

The last proposal is to again divide the plant into j divisions but then rank the divisions as to importance. Each division would be assigned n alarms and each division would receive the same number of alarms. Thus the number of divisions would be limited to k divided by n divisions. The n alarms assigned to the most important division would be the easiest to identify and those assigned to the least important division would be those alarms that are ranked the lowest as being easy to identify. This method will limit the number of divisions but will allow some control over the depth of the coding in the divisions.

The group looking into the areas of human performance had not had as much success as had the group working on the design problems. They had found that if only one dimension is used for coding no more than four or five alarms could be identified all the time. Four is the limit if the dimension is intensity and five is the limit if the dimension is frequency. However if two dimensions are used for coding the alarms, the number of absolute identifications possible increases to eight alarms. The literature review also indicates that if additional dimensions are used and only a few steps of each dimension are

used the number of alarms that can be identified can be very large. This was what Joe C. has been referring to in his first suggestion. A study done at The University of Tennessee had found that if more than two alarms are sounded simultaneously, where each alarm is coded by changes in frequency, intensity, or on-and-off modulation, the person attempting to identify them will make a large number of mistakes in identifying the alarms.

With the three proposed designs and the performance information presented, the team now had to evaluate the designs in order to find a workable system.

One problem was immediately found with the first design. Turkey #1 had four cooling loops and secondary systems. No matter how the plant was divided, there would be divisions of equal importance. It was suggested that the alarms chosen for the equal divisions could be made different in only one dimension, say intensity, thus giving each division its own alarm set which would be different but similar to the sets for the other divisions. This was rejected because of the very good chance of confusing an operator and having him make a mistake when several of the alarms sounded simultaneously. Besides, if only two alarms could be used, then the plant could be divided into two divisions with a large number of each of the two alarms installed for the various phenomena and components to be monitored.

The second design looked better than the first. Here there can be many divisions with at least 2 levels of coding in each division. Each system, function, or location would be assigned one of the two alarms.

The last design will become the same as the first design if k = 2 and n = 1 and this proposal has already been rejected. If k = 2 and n = 2 then the last design has only one division with 2 levels of coding. This is really the same as the second design where each system, function, or location would be assigned one of the two alarms. But, if n could be equal to one or two then there could be many divisions with some divisions having two levels of coding and some having only one. This was a design not originally proposed.

Thus, this newest design of having many divisions of the plant with either one or two levels of coding for each division appears to be the best system to use because it is the most versatile when only two alarms are available. What happens if there were 3 alarms available? Design one still suffers from the same flaw as before. Design two simply achieves another level of coding for each division. The third design is still no good if k equals 3 and n equals one. Furthermore if n equals two, the impossible situation of having a fraction of a division occurs. The third design is again the same as the second if n equals three. The new design is still more versatile if three alarms are available, with each division having from one to three levels of coding. This design will continue to gain versatility as the number of alarms available increases.

In order to increase the number of alarms beyond two they must become easier to tell apart when sounded simultaneously. Thus the systems engineers conclude they need someone to do some more research and testing of subjects in the following areas:

1. Increase the number of dimensions and reduce the number of steps in each dimension used to code the alarms to increase absolute discrimination and recognition by reducing the masking of one signal by another.
2. Try using periodic and aperiodic sounds as signals in the same alarm systems to see if this will increase the number of alarms that can be correctly identified.
3. Try combining an auditory system with a visual system to perhaps improve subject performance.

If any of the above suggestions will allow more coded auditory alarms to be used, the use of multiple simultaneous alarms in the control room will greatly improve this important communication channel.

However, if the research does not increase the number of alarms that can be used, then the system that was rejected in the planning stage should be pursued. This was the microprocessor or computer controlled system that would present only the most urgent alarm if several occurred simultaneously.

		DIVISION NUMBER					
		1	2	3	4		k
ALARM NUMBER	1	X_1					
	2		X_2				
	3			X_3			
	4				X_4		
	. . .						
	k						X_k

Figure D-1. Alarm System 1.

		DIVISION NUMBER					
		1	2	3	4		j
ALARM NUMBER	1	X_1	X_2	X_1	X_1		X_1
	2	X_2	X_2	X_2	X_2		X_2
	3	X_3	X_3	X_3	X_3		X_3
	4	X_4	X_4	X_4	X_4		X_4
	. . .						
	k	X_k	X_k	X_k	X_k		X_k

Figure D-2. Alarm System 2.

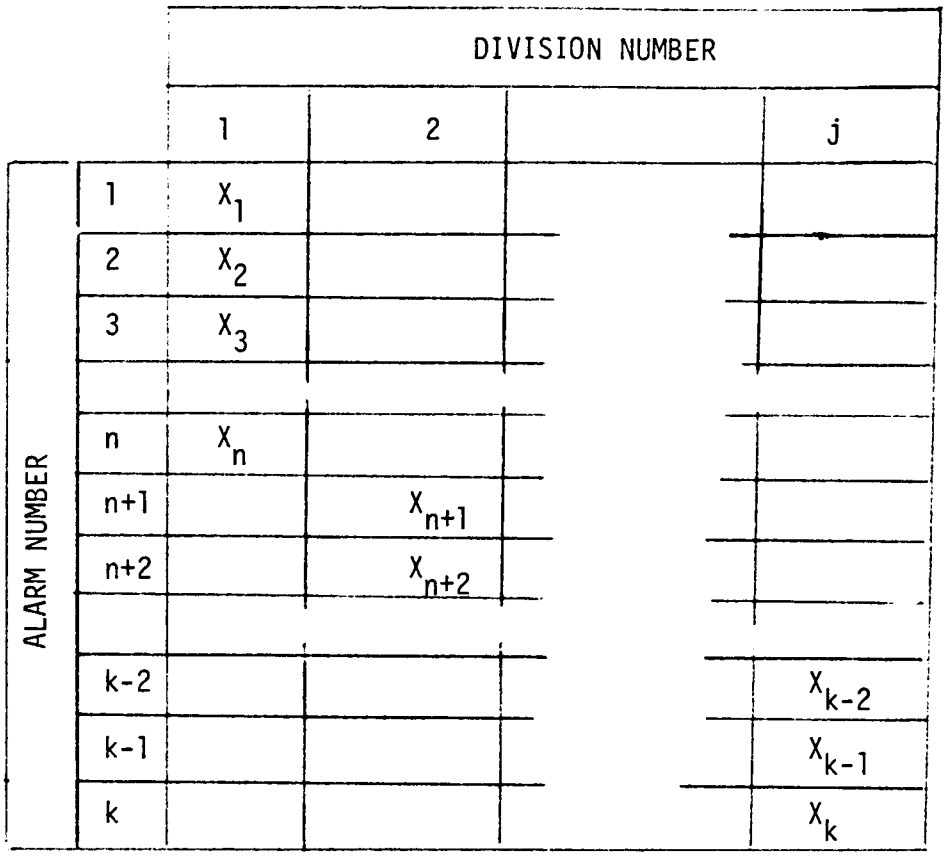


Figure D-3. Alarm System 3.

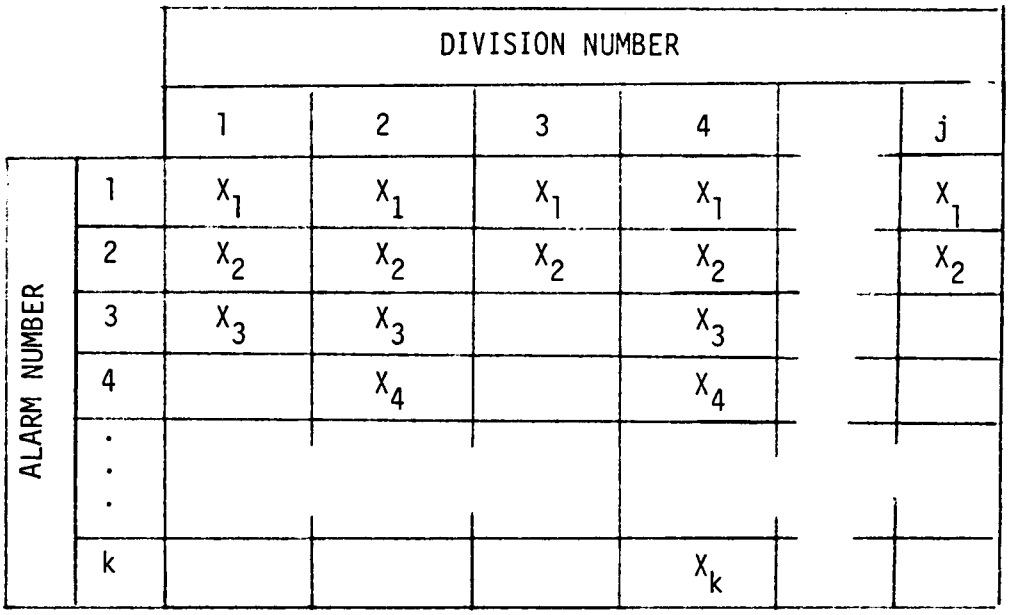


Figure D-4. Recommended Design.

VITA

Robert Estel Ford was born in Spring City, Tennessee on September 3, 1943. He attended elementary school in that city and was graduated from Spring City High School in June 1961. In March 1969 he received a Bachelor of Science degree in Engineering Physics from The University of Tennessee, Knoxville.

After pursuing a career in industry, he entered the Graduate School of The University of Tennessee, Knoxville, in the fall of 1980. He received a Master of Science degree with a major in Industrial Engineering in March of 1982.

The author is a member of the American Institute of Industrial Engineers. Mr. Ford will be employed with EG&G Idaho after graduation.

He is married to the former Frances Price of Knoxville, Tennessee. He has two sons, David and Jason.