Empirical Studies Concerning Aural Alerts for Cockpit Use Leading to an Aural Alerting Signal Categorization Scheme

Jennifer L. Burt

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

Doctor of Philosophy
in
Industrial and Systems Engineering

John G. Casali, Chair
Albert M. Prestrude
Robert C. Williges
Gary S. Robinson
Alan T. Pope

November 11, 1999
Blacksburg, Virginia

Keywords: Attention, Auditory Warnings, Aural Alerts,
Electroencephalogram, Identification, Urgency, Workload

Empirical Studies Concerning Aural Alerts for Cockpit Use Leading to an Aural Alerting Signal Categorization Scheme

Jennifer L. Burt

(ABSTRACT)

The only way to simplify and promote the effective use of an alerting system that must be comprehensive in its coverage of hazardous or non-normal conditions is to convey top level information that provides an indication of criticality and identity. In an attempt to reduce the number of aural alerting signals presented in aircraft flight decks, this investigation pursued advances toward the development of a simple aural alert categorization scheme that provides flight deck function and urgency level information. In Experiment 1, 20 subjects having "normal" hearing threshold levels provided magnitude estimation urgency ratings for a series of aural alerts. These ratings revealed that subjects perceived low, moderate, and high urgency levels within each of four equally urgent aural alerting sets. In Experiment 2, 12 subjects having "normal" hearing threshold levels participated in a brief training session and then performed a sound identification task in conjunction with an automated and manual tracking task. Sound identification data revealed that subjects correctly identified the alerting set (i.e., major flight deck function) and urgency level associated with each of 12 aural alerts in 96.53% of the trials occurring during automated tracking and in 95.83% of the trials occurring during manual tracking; furthermore, subjects correctly identified each alerting set, urgency level, and aural alert equally often during each tracking task condition. Electroencephalogram (EEG) data recorded throughout the performance of each tracking task condition revealed that manual tracking required a significantly higher level of attentional engagement than automated tracking. Subjective assessments of workload collected after the performance of each tracking task condition revealed that a significantly higher level of workload was experienced during the manual condition of the tracking

task than during the automated condition of the tracking task. Collectively, this investigation's results indicated that acoustic parameter manipulations can be used to create four distinctive alerting sets that each convey three levels of urgency and that these alerting sets and urgency levels can be accurately identified when two levels of workload and attentional engagement are experienced.

This research was conducted at NASA Langley Research Center, in conjunction with the Crew/Vehicle Integration Branch's Hazards and Error Management research program, located at Mail Stop 152, Hampton, Virginia 23681-2199, USA. Funding for this research was provided by the NASA Graduate Student Researcher's Program (NGT-1-52107).

DEDICATION

To my family for their many years of encouragement and sacrifice

ACKNOWLEDGMENTS

I would like to gratefully acknowledge the assistance of Dr. John G. Casali, Dr. Albert M. Prestrude, Dr. Gary S. Robinson, and Dr. Robert C. Williges of Virginia Polytechnic Institute and State University; Dr. Alan T. Pope, Dr. J. Raymond Comstock, Jr., and the Flight Dynamics and Control Division of the National Aeronautics and Space Administration (NASA) Langley Research Center; and Mr. Edward H. Bogart of the Unisys Corporation in the design of this experimental investigation. Genuine gratitude is expressed to Mr. Daniel W. Burdette, Capt. Richard B. (Skeet) Gifford (Ret.), and Capt. David A. Simmon (Ret.) of Lockheed-Martin Engineering and Sciences Company; Mrs. Debbie S. Bartolome-Rull formerly of Lockheed-Martin Engineering and Sciences Company; Dr. William H. Rogers of Honeywell, Inc.; Mr. Michael L. Barnes formerly of Virginia Polytechnic Institute and State University; Dr. Kara A. Latorella of NASA Langley Research Center; Dr. Judy Edworthy of the University of Plymouth; and Dr. Elizabeth Hellier of City University for their assistance in the initial conceptualization of this research endeavor which began in the summer of 1995. Great appreciation is also expressed to Mr. Edward H. Bogart, Mr. Michael L. Barnes, Dr. Alan T. Pope, and Dr. J. Raymond Comstock, Jr. for their assistance in the implementation of the current investigation. Additionally, I am very much indebted to Mr. Edward H. Bogart, Mr. Michael L. Barnes, and Dr. Alan T. Pope for their assistance in the analysis of experimental data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Aural Alerts in the Flight Deck

Alerting signals are presented in aircraft flight decks to attract pilot attention and provide pertinent information about current or impending situations that require pilot awareness or action. These signals, or displays, serve as an interface between the pilot and the aircraft thereby helping to make the flight environment manifest to the pilot. In other words, alerting signals facilitate the pilot's situation awareness by conveying information regarding the state of the aircraft's on-board systems as well as the overall conditions of the flight environment.

In addition to communication with other members of the air and ground crew, the auditory channel is also used for the presentation of warnings, cautions, and, in some cases, advisories and messages. Some examples of aural alerting signals used in commercial aircraft include bells, chimes, tones that vary in their pulse and burst characteristics, and synthesized voice messages (Boucek, Po-Chedley, Berson, Hanson, Leffler, and White, 1981). Speech may be appropriate for all types of messages, but the use of tonal signals is recommended for the presentation of qualitative information such as warnings and the indication of status (Kroemer, Kroemer, and Kroemer-Elbert, 1994).

Aural alerts were originally implemented in aircraft as a measure to get pilots' attention because signal lights were considered to be ineffective. Unlike visual inputs of information, aural signals alert the pilot to dangerous or potentially dangerous conditions regardless of head position and direction of gaze, as well as provide sensory inputs that are less disrupted by anoxia and positive G-forces (Doll and Folds, 1985; Doll, Folds, and Leiker, 1984; Edworthy, Loxley, and Dennis, 1991; Munns, 1971). Aural alerting signals also enable the pilot to fly "head-up" for longer periods of time and provide relief from the constant monitoring of visual alerting displays.

Since aural signals reduce the need to scan the instrument panel visually, the pilot's visual workload is decreased, and the probability and speed with which one may react to and correct an emergency situation is increased (Bertone, 1982; Doll and Folds, 1985). Several studies conducted under simulated conditions with experienced pilots have shown that aural alerts lead to faster response times than the visual alerts presented on panel indicators (Reinecke, 1976, 1981; Wheale, 1981, 1982, 1983). In addition to improved levels of performance, pilots have also reported preferences for aural alerts since synthesized voice messages may interfere with other flight deck communications (Axelsson and Stoby, 1991; Boucek, et al., 1981; Doll and Folds, 1985).

While aural alerting signals offer many benefits and have surely reduced the number of aviation mishaps, they have also created special problems. Many of the aural alerts found on older-generation aircraft were described as being too numerous, loud, and disruptive (Cooper, 1977 as cited in Berson, Po-Chedley, Boucek, Hanson, Leffler, and Wasson, 1981; Doll and Folds, 1985; Doll et al., 1984; Folds, 1985; King and Corso, 1993; Marshall, 1987; Munns, 1971; Patterson, 1982, 1989; Rood, 1989; Rood, Chillery, and Collister, 1985; Thorning and Ablett, 1985; Veitengruber, Boucek, and Smith, 1977; Wheale, 1983). Therefore, researchers began to ask if "... the inclusion of new alerts or the reformatting of existing ones [could] help prevent future accidents" (Hanson, Howison, Chikos, and Berson, 1982, p.19).

Aural Alerting Signal Design and Perception

Before discussing the standardization of aircraft alerting systems and the development of an aural alert categorization scheme based on flight deck functionality, two issues will be addressed to promote an understanding of aural alerting signal design and perception. First, the subjective characteristics, or perceptions, associated with various acoustic parameters will be addressed through a review of the feelings evoked by the time and tone at which pulses of sound are generated as well as through the identification of several useful psychoacoustic principles. Second, the perception of aural alerting signals will be addressed by reviewing two conceptual models.

Perceptions associated with acoustic parameters. According to music theory, sound may be manipulated in terms of time (i.e., temporal parameters) and tone (i.e., frequency and spectral parameters) (Lieberman, 1959; Starer, 1969). The organization of time into pulses of sound presented against a steady beat results in the human perception of rhythm, and the organization of various tones within a given rhythmic pattern results in the human perception of melody and harmony. In other words, rhythm creates an "overall pattern" of sound into which melodic and harmonizing tones may be added to produce a unique aural signal (Lieberman, 1959, p.1). A rhythmic pattern can be generated by clapping the hands or tapping a foot, and it may be compared to the dot-dash rhythm of a single letter of Morse code [e.g., •− (A), −• (N), or −•• (D)] as well as to a pattern of accented, or stressed, and unaccented, or unstressed, syllables used in poetry [e.g., ∪/ (an iambic meter in which an unstressed syllable is followed by a stressed syllable), /∪ (a trochaic meter in which a stressed syllable is followed by an unstressed syllable), or /∪∪ (a dactylic meter in which a stressed syllable is followed by two unstressed syllables)]. The ease with which people may categorize (i.e., discriminate among and associate meanings

with) various rhythmic patterns is illustrated by the ability of individuals to identify and decipher Morse Code communications as well as identify and define the construction of various poetic meters.

The tempo, or speed, at which a rhythmic pattern is generated corresponds to the rate at which pulses of sound are presented (i.e., the number of beats presented per unit of time). Tempo may be varied by decreasing or increasing the rhythmic pattern's sound pulse and interpulse interval (i.e., the amount of time between sound pulses) durations. Fast tempos typically give rise to feelings of excitement and turbulence, while slow tempos are often associated with feelings of peace and calm (Lieberman, 1959). Acoustics research regarding perceptions of the "biological siren," also known as the human infant cry sound, reveals that the feelings associated with the rhythm of music are similar to those associated with the rhythm of a baby's cry. Zeskind, Wilhite, and Marshall (1993), for example, found a general monotonic relationship between interpulse intervals and perceptions of urgency, with increasingly shorter pauses between infant cry bursts being perceived as increasingly more urgent. Rhythm, therefore, can be used to design aural signals that are distinguishable from one another (i.e., recognizable), and the tempo at which a rhythmic pattern is presented can be used to convey various levels of urgency.

In addition to the manipulation of sound through the timing of pulses, the tonal properties of sound may also be manipulated. For example, the frequencies of sound pulses may be arranged within a rhythmic pattern to produce distinct tonal patterns. The subjective characteristic of sound related to frequency is pitch, and like rhythm, pitch is a powerful means of expressing feeling. High-pitched sounds are often associated with feelings of excitement and urgency, while low-pitched sounds are typically perceived as being less alarming. Again, research related to affective responses elicited by infant cry sounds reveals that the feelings associated with various

pitches in a musical context are similar to the feelings associated with the pitches of various cry sounds. It has been shown in several studies that as the fundamental frequency of an infant's cry sound increases, the cry is perceived by the listener as being progressively more urgent (Freudenberg, Driscoll, and Stern, 1978; Frodi, 1985; Zeskind and Lester, 1978).

The pitch contour of the sound pulses within a rhythmic pattern can be varied to produce simple ascending or descending, as well as more complex, melodic tonal patterns. According to the acoustics research of Edworthy (1985), pitch contour is a critical component in the recognition of melodic sequences, especially when such sequences are heard in a non-musical context. Therefore, the use of a distinct rhythmic pattern and pitch contour in the design of a sound or alerting signal is beneficial because structured sequences of tones are more easily remembered than random tones (Deutsch, 1980, 1986).

Perceptual theorists have long investigated the psychological quality of various sounds, but it was not until relatively recently that psychoacoustic principles (e.g., structured sequences of tones are more easily recognized, and higher-pitched sounds with faster tempos are perceived as more exciting and urgent) were applied in operational settings such as aircraft flight decks, hospital recovery rooms, and factories (Lazarus and Hoge, 1986; Lower et al., 1986; Meredith and Edworthy, 1994; Patterson, 1982,1989). Patterson (1982, 1989) presented a four step procedure for the construction of nonverbal aural signals having acoustic parameters that can be manipulated to allow for various types of signal design. Patterson's four steps include: 1) determining the appropriate level of loudness, 2) designing a small pulse of sound, 3) incorporating the sound pulse into a longer burst of sound, and 4) forming a complete aural alert using bursts of sound followed by short periods of silence. By using this method, aural alerts may be designed so that they are not startling, are presented at appropriate levels of loudness for the

surrounding environment, are distinctive from one another, and can be presented at varying levels of urgency.

By following the steps of aural alerting signal construction outlined by Patterson (1982, 1989), Judy Edworthy and her colleagues have successfully determined the ways in which the perceived urgency of an aural alert is altered by variations in acoustic parameters. Edworthy et al. (1991) and Edworthy (1994b) constructed alerts that varied in fundamental frequency, harmonic series, amplitude envelope shape, delayed harmonics, number of repeating units (i.e., number of times a sequence of sound pulses is repeated), speed, rhythm, pitch range, pitch contour, and musical structure. All of these factors appeared to have clear and consistent effects on perceived urgency. According to Edworthy (1994a), "[i]t is possible ... to convey an appropriate level of urgency through the warning itself by the way in which the acoustic features of the warning are manipulated" (p.232). Results from Edworthy et al. (1991) and Edworthy (1994b) which characterize the effects of various acoustic parameters on perceived urgency are presented in Table 1.

Table 1. Direction of Effects Found by Edworthy et al. (1991) and Edworthy (1994b)

Parameter	Direction of Effects
Fundamental Frequency	High > Low
Amplitude Envelope Shape	Regular / Slow Onset > Slow Offset
Harmonic Series	Random / 10% Irregular > 50% Irregular > Regular
Delayed Harmonics	No Delayed Harmonics > Delayed Harmonics
Speed	Fast > Moderate > Slow
Number of Repeating Units	4 > 2 > 1
Speed Change	Speeding Up > Regular / Slowing
Rhythm	Regular > Syncopated
Pitch Contour	Random > Down / Up
Pitch Range	Large > Small > Moderate
Musical Structure	Atonal > Unresolved > Resolved

Key: > More urgent than
/ Equally as urgent as

Using Patterson's alerting signal construction technique as well as the findings of Edworthy et al. (1991) and Edworthy (1994b), Hellier, Edworthy, and Dennis (1993) and Edworthy (1994b) applied Stevens' (1957) power law to judgments of perceived urgency in order to examine the relationship between objective manipulations of four parameters (alerting signal speed, fundamental frequency, repetition rate or number of times a sequence of sound pulses is repeated, and harmonic content) and changes in subjective assessments of urgency level. By computing and applying the exponents derived from Stevens' (1957) equation relating objective to subjective values, Hellier et al. (1993) and Edworthy (1994b) demonstrated that systematic changes in acoustic parameter levels result in systematic changes in perceived urgency levels. Derived exponent values greater than 1.0 imply that it takes relatively small parameter changes to produce a unit change in perceived urgency, whereas exponent values less than 1.0 imply that it takes large increments of change in the parameter to produce relatively small increases in perceived urgency. Hellier et al. (1993) and Edworthy (1994b) both determined that speed has an exponential value of 1.35; repetition rate has a value of 0.50; fundamental frequency has a value of 0.38; and inharmonicity has a value of 0.12. These results are depicted in Table 2.

Table 2. Urgency Power Functions Found by Hellier et al. (1993) and Edworthy (1994b)

Parameter	Exponent
Speed	1.35
Number of Repeating Units	0.50
Fundamental Frequency	0.38
Inharmonicity	0.12

Based on these data, Hellier et al. (1993) presented guidelines suggesting how to manipulate the acoustic parameters of an aural alerting signal in order to produce a 50% increase in perceived urgency level, a doubling of perceived urgency level, and a tripling of perceived urgency level. For example, an alerting signal's tempo, or speed, may be manipulated in the following ways to achieve changes in the alert's level of perceived urgency: to increase the alert's level of perceived urgency by 50%, the alert's pulse rate (i.e., sound pulse and interpulse interval durations) must be decreased by a factor of 1.3; to double the alert's level of perceived urgency, the alert's pulse rate must be decreased by a factor of 1.6; to triple the alert's level of perceived urgency, the alert's pulse rate must be decreased by a factor of 2.2.

Aural alerting signal perception. One conceptual model of auditory warning sound perception is presented by Wilkins (1980 as cited in Wilkins and Martin, 1987). This model, which is shown in Figure 1, suggests that the perception of an auditory warning, or aural alerting signal, is determined by the signal's audibility (i.e., whether or not the signal can be heard); the signal's attention demand (i.e., whether or not the signal will be heard); and the signal's recognition (i.e., whether or not the signal's meaning will be understood when it is heard). The presence of noise is also an important consideration because noise may mask warning signals as well as interfere with a listener's attention (Wilkins and Acton, 1982).

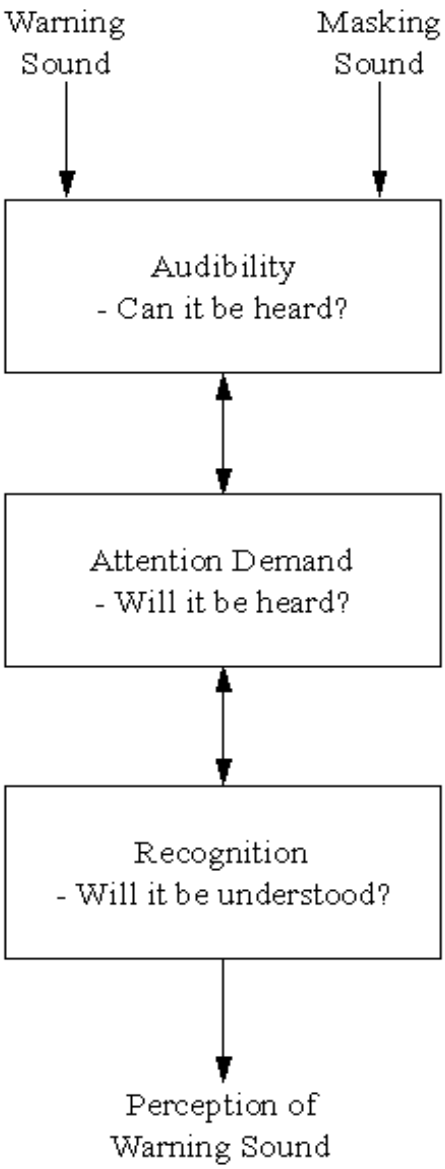


Figure 1. Wilkins' (1980 as cited in Wilkins and Martin, 1987) conceptual model of auditory warning sound perception.

A similar description of aural alerting signal perception is provided by Burt (1996) in the following statement: "... an effective aural alert must be perceived (i.e., heard and attended to) and should also be recognized (i.e., meaningfully identified and interpreted) by the listener" (pp. 23-24). As shown in Figure 2, Burt suggests that an effective aural alert may be conceptualized as being composed of several interrelated independent variables, depicted as ovals, which may be tapped or investigated through the use of dependent variables, depicted as rectangles, associated with subjective assessments. For example, manipulations of alerts' acoustic parameters can be used to influence the alerts' distinctiveness and perceived urgency levels. Then, magnitude estimations, pair comparison ratings, and sound identifications may be collected to assess the ability of listeners to distinguish among and identify the urgency levels of the aural alerts.

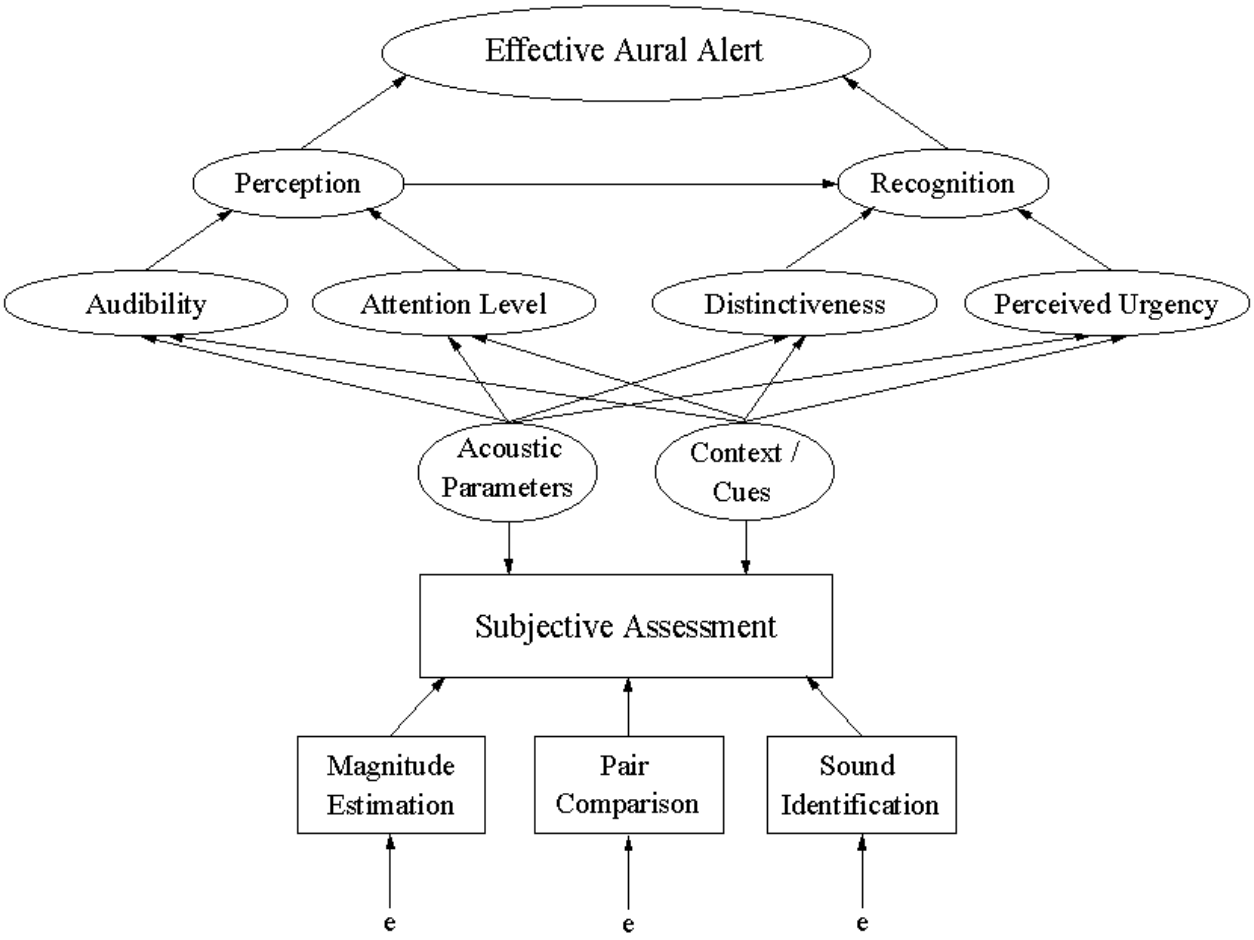


Figure 2. Burt's (1996) conceptualization of an effective aural alert. (NOTE: "e" indicates that some degree of measurement error may occur.)

According to Burt's (1996) conceptualization, perception, or detection, of an aural alert is primarily determined by the audibility of the signal and the attention, or stimulation, level of the listener. To be perceived, an alert should be designed with acoustic parameters that are within the listener's range of hearing and which exceed masked threshold levels by a reasonable margin. Also, the listener should be optimally stimulated so that his or her level of attention allows the aural alert to be detected. Recognition of an aural alert, which may occur if the sound is perceived, is primarily determined by the distinctiveness of the alert and the level of urgency perceived by the listener. To be recognized, an alert should be uniquely identifiable. That is, each alert should be distinctive (i.e., sufficiently different) from every other alert presented in a particular setting. Also, the level of perceived urgency associated with an aural alert facilitates the recognition of an alerting signal by helping the listener interpret and understand the severity of the situation being signaled.

In the diagram depicted in Figure 2, note that a division is made between the effects of an alert's acoustic parameters and the context and cues associated with the alert. This division between independent variables related to acoustic parameter manipulation and independent variables related to the situation in which an alert is presented is important in that it facilitates an understanding of the difference that can exist between experimental manipulation of stimuli and experimental manipulation of the listener. In operational settings, the effectiveness of an aural alert will always be based on both the subjective characteristics associated with the acoustic parameters as well as the context and cues that are associated with the alert. For example, the task that the listener is performing when an aural alerting signal is presented can have a great impact on the perception of the alert. An empirical study conducted by Casali and Wierwille (1983) showed that the attentional demands caused by communications loading affected the

subjects' detection of verbal signals (i.e., aircraft call signs); subjects responded significantly faster to the signals presented during low-load conditions than to the signals presented during high-load conditions. Also, the ways in which the listener learns to interpret aural alerts (i.e., learns to associate alerts with various meanings or categories) can influence the ways in which the alerts are recognized. The work of Burt, Bartolome, Burdette, and Comstock (1995), for example, revealed that the inherent levels of urgency associated with three tonal auditory warning signals were replaced with arbitrary urgency levels assigned to the warnings through verbal instructions and task demands.

Standardization of Flight Deck Alerts

In an attempt to improve aircraft alerting systems, the FAA sponsored a series of studies during the 1970's and 1980's that investigated the visual, aural, and voice alerts presented in the flight deck. As a result of the collective research efforts of Boeing, Lockheed, McDonnell Douglas, and several airline companies, aircraft alerting systems were assessed, and recommendations regarding the improvement, standardization, and simplification of alerting systems were made. An early assessment of alerting systems revealed that discrete alerts were frequently being added to the flight deck and that "[v]ery little standardization had been used by the airframe manufacturers in implementing alerting system elements" (Berson et al., 1981, p.2). Therefore, the "Aircraft Alerting Systems Standardization Study," as this FAA-sponsored research project was called, involved conducting a series of flight tests and analyzing aircraft accident reports in order to identify the primary functions of an advanced alerting system and define guidelines for the design of appropriate flight deck alerts (Boucek et al., 1981).

According to Berson et al. (1981), the Aircraft Alerting Systems Standardization Study produced the following general objectives that should be used to guide the design of aircraft alerting systems:

1. Reduce the total number of discrete visual and aural alerts presented in the flight deck.
2. Conform to a quiet dark flight deck during normal operation.
3. Reduce crew information processing and memory requirements.
4. Minimize the time required for the crew to detect and assess failure conditions and initiate corrective actions.
5. Minimize the distracting effects of the alerting system on other flight crew tasks.
6. Facilitate alerting system standardization among airframe manufacturers, aircraft types, and commercial airplane operations.
7. Provide for alerting system growth in a form that does not necessitate additional components.

The work of Berson et al. (1981) also includes guidelines pertaining specifically to the aural alerts presented in the flight deck. Several of these guidelines are listed below.

1. Aural alerting signals should alert the flight crew to impending or existing conditions that require attention and should advise the crew of the alert urgency level.
2. Three flight deck alerting sounds should be used. One sound should be used to signal high urgency alerts (i.e., warnings that require immediate action); one sound should be used to signal moderate urgency alerts (i.e., cautions that require imminent action); and one sound should be used to signal low urgency alerts (i.e., advisories that require crew awareness).
3. The sounds should be selected to reflect their alert urgency level.
4. Each sound should differ from the other sounds in more than one dimension (e.g., frequency and duration).

5. The frequency of aural alerting signals should be between 250 and 4000 Hz.

These guidelines suggest that a very simple aural alerting system based on a categorization scheme that provides information about an alert's urgency level through the use of three discrete sounds should be implemented in the flight deck. Such a system may be successfully implemented, as demonstrated by Boeing's engine indication and crew alerting system (EICAS) which uses one discrete sound for warnings and one discrete sound for cautions, but other discrete aural alerts are also presented in the flight deck. This continued proliferation of flight deck alerts is probably a result of the development of advanced aircraft sensors capable of providing more detailed information about the flight environment and the subsequent addition of new alerts related to hazardous situations.

Aural Alert Categorization Based on Flight Deck Functionality

While the Aircraft Alerting Systems Standardization Study emphasizes the need to prevent "the proliferation of alerts" in the flight deck, the study also proposes the use of a flight status monitor (FSM) system capable of presenting alerts associated with navigation errors, tire or wheel failure, collision avoidance, aborted takeoff, and windshear (Berson et al., 1981, p.31; Hanson et al., 1982). The researchers acknowledge that "[s]everal available guideline documents recommend that the number of discrete sounds used in an alerting system be limited to 3 to 5" (Berson et al., 1981, p.105); but, the use of a FSM system capable of presenting additional discrete aural alerts is also characterized as having "potential benefit" (Hanson et al., 1982, p.76). It seems, therefore, that the development of a new aural alert categorization scheme in which distinctive signals convey urgency information, as well as some other meaningful identification

information, might more adequately accommodate all alerting functions (both present and future) without requiring that additional aural alerts be introduced in the flight deck.

Rather than use discrete aural alerts to represent various urgency levels or different hazardous situations, it has been suggested that distinctive sounds be used to identify the major flight deck function (i.e., functional category) to which an alert corresponds (Burt, 1996; Burt and Casali, 1997). The four major flight deck functions include: communication, flight control, navigation, and systems management (Swink and Goins, 1992). Communication involves managing the flow of information between each flight deck crew member, air traffic control (ATC), the cabin crew, passengers, and the airline company; flight control involves adjusting or maintaining the flight-path, attitude, and speed of the aircraft relative to the navigation requirements; navigation involves developing a desired plan of flight, positioning the aircraft relative to landmarks, and adjusting the plan of flight as necessary; and systems management involves monitoring and managing the aircraft's systems. Within each of these flight deck functions, low, moderate, and high urgency level situations may occur. Low urgency level situations require crew awareness and may require future action; moderate urgency level situations require some form of action; and high urgency level situations require immediate action.

This characterization of flight deck functionality lends itself nicely to the design of an aural alert categorization scheme in which: 1) a distinctive aural alert is associated with each of the four major flight deck functions, and 2) the acoustic parameters of a given alert can be manipulated in a way that preserves the overall pattern of the signal (thereby preserving the sound's distinctiveness) while conveying low, moderate, and high levels of urgency. Such an aural alerting system would be capable of presenting a total of four discrete alerting sets that convey information about urgency as well as the flight deck function associated with an alert and would

also be able to incorporate any additional alerts that might be deemed to have potential benefit (e.g., windshear and collision avoidance alerts would be subsumed under a single flight control alert).

Adherence to the guidelines set forth by Berson et al. (1981) will be achieved by this aural alerting system for two reasons. First, since the alerting system will present aural alerts having acoustic parameters subjectively described as being distinctive and as conveying appropriate levels of urgency, the signals will alert the flight crew to specific impending or existing conditions that require attention and will advise the crew of the alert urgency level. Second, effective acoustic parameters will be used in conjunction with a simple categorization scheme that will reduce the total number of discrete alerts presented in the flight deck, and all current and future alerting components may be categorized within one of the four major flight deck function alerting sets. As a result of these two developments, the standardization of alerting systems among airframe manufacturers, aircraft types, and commercial airplane operations will be facilitated; and since consistent use of an alerting system with distinctive alerting sets and urgency levels will be promoted, the implementation of this aircraft alerting signal categorization scheme may reduce crew information processing and memory requirements as well as minimize the time required for the crew to detect and assess failure conditions and initiate the appropriate corrective action.

In summary, what will make this type of aural alert categorization scheme so effective is that upon hearing one particular alerting signal, a pilot will not only be able to determine to which flight deck function an alert corresponds, but will also be aware of the alert's urgency level. Since it has been shown that urgency coding can be introduced into an aural alerting system to improve performance without adding to workload (Sorkin, Kantowitz, and Kantowitz, 1988), aural signals can be used to aid in the realization that a change in the flight environment has occurred as well as

to aid in the estimation of the change's level of significance. Although the proposed alert categorization scheme will require training such as that offered by cockpit resource management (CRM) in order for pilots to learn the association between a particular alerting set and a particular flight deck function, the perceived urgency level of the situation will be the result of an inherent response to the alerting signals' sound parameters. This means that the priority level of a situation will be determined and the decision-making process will be assisted by auditory stimuli that do not increase a pilot's workload. Hoge, Schick, Kuwano, Namba, Bock, and Lazarus (1988) found cross-cultural differences in Western European and Asian perceptions of aural alerting signals, but it is still suggested that the implementation of this type of categorization scheme into aircraft alerting systems as an error-reduction measure is a worthwhile endeavor, even though the development of an international standard may be very difficult.

Burt's (1996) Investigation of an Aural Alerting Signal Categorization Scheme

Burt (1996) and Burt and Casali (1997) describe a research study that investigated the aural alerting signal categorization scheme characterized above. This study examined the ability of a population having "normal" hearing to: 1) distinguish among four sets of aural alerting signals having distinctive rhythmic patterns and pitch contours, 2) perceive three urgency levels having distinctive tempos within each alerting set, and 3) associate each alerting set and its related urgency levels with one of the four major flight deck functions. Information regarding Burt's (1996) subjects, experimental design, independent variables, dependent measures, hypotheses, results, conclusions, and recommendations for additional research is provided below.

Since the research endeavor proposed later in this document will make use of the test facilities, test apparatus, test chambers, and test system calibration procedures as well as the

method of stimuli and background noise presentation used in the Burt (1996) study, detailed discussions of this information are provided in the "Method" section of Experiment 1. Furthermore, since the experimental procedures used by Burt to collect magnitude estimation urgency ratings and sound identification data were incorporated into the current investigation, detailed descriptions of these procedures are included in the "Procedure" sections of Experiments 1 and 2. A detailed description of the experimental procedure used by Burt to collect pair comparison ratings of similarity as well as more in-depth information regarding analyses of magnitude estimation, pair comparison, and sound identification data may be found in Burt (1996).

Subjects. Subjects participating in the Burt (1996) study consisted of seven male and five female volunteers from the civil servant population employed at the National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) in Hampton, Virginia. All subjects were at least 18 years old and had auditory thresholds associated with "normal" hearing; that is, all subjects had hearing threshold levels in each ear that were $\leq$ 25 dB at 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz (Davis and Silverman, 1978 as cited in Miller and Wilber, 1991).

Experimental design. The experimental design used for data collection was a 4 (Alerting Set) x 3 (Urgency Level), completely crossed, full factorial, within-subject design. The same 12 subjects were assigned to each experimental cell. The experimental design matrix is shown in Figure 3.

	High	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}
URGENCY LEVEL (Within-Subject)	Moderate	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}
	Low	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}
		I	II	III	IV

ALERTING SET
(Within-Subject)

Figure 3. Experimental design matrix used in the Burt (1996) study.

All independent variables were treated as fixed-effects variables, and subjects were treated as a random-effect variable.

Independent variables. As shown in Figure 3, the two factors for the experimental design were aural alerting set and urgency level. The stimuli were 12 aural alerting signals, each of which belonged to one of four alerting sets and consisted of sound pulses and interpulse intervals having various durations. Most sound pulses included a linear onset time of 20 ms (i.e., the time from the start of the pulse until it reached maximum output) and a linear offset time of 20 ms (i.e., the time during which the pulse output fell from maximum to zero). Sound pulses less than 40 ms in length had linear onset and offset times that peaked at the middle of the pulse. The first harmonic, or fundamental frequency, of each sound pulse was present at 100%; and simultaneously, the second through fifth harmonics were present at 50% of the fundamental frequency's amplitude.

The specific acoustic parameters of the 12 aural stimuli are included in Tables 3 - 6, and graphical representations of the alerts within each set, indicating low, moderate, and high urgency, are depicted in Figures 4 - 15. When examining Tables 3 - 6 and Figures 4 - 15, note that each alert consisted of a given rhythmic pattern played twice; therefore, the total duration of each alert represents the sum of its sound pulse and interpulse interval durations multiplied by a factor of 2. Additionally, each alerting signal was comprised of complex tones made up of a controlled set of harmonics, but harmonic content is not described in Tables 3 - 6.

Table 3. Acoustic Parameters of Aural Stimuli Used by Burt (1996) - Set I

Urgency Level	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
Low	523	100%	80	80	2880
	523	50%	80	80	
	563	50%	80	80	
	563	75%	80	240	
	583	50%	80	80	
	583	100% to 50% logarithmic fade	320	160	
Moderate	523	100%	50	50	1800
	523	50%	50	50	
	563	50%	50	50	
	563	75%	50	150	
	583	50%	50	50	
	583	100% to 50% logarithmic fade	200	100	
High	523	100%	31	31	1122
	523	50%	31	31	
	563	50%	31	31	
	563	75%	31	94	
	583	50%	31	31	
	583	100% to 50% logarithmic fade	125	63	

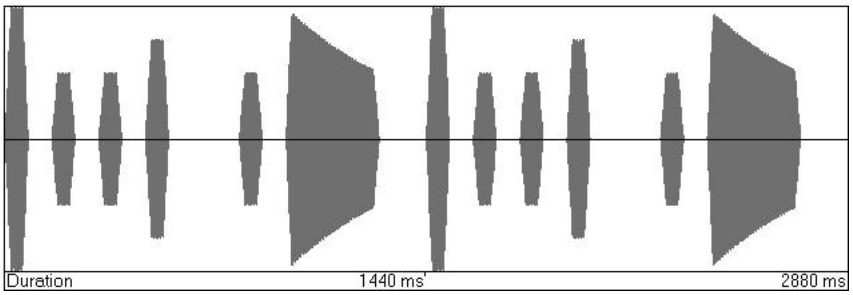


Figure 4. Set I low urgency.

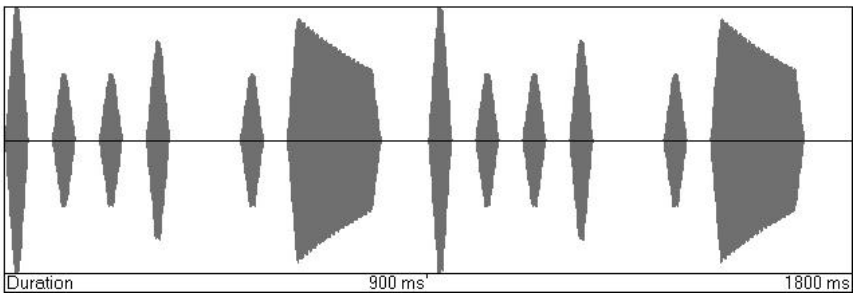


Figure 5. Set I moderate urgency.

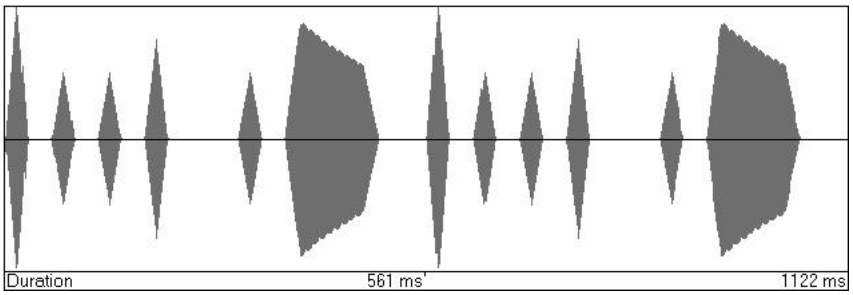


Figure 6. Set I high urgency.

Table 4. Acoustic Parameters of Aural Stimuli Used by Burt (1996) - Set II

Urgency Level	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
Low	593	100%	120	290	2974
	553	50%	60	17	
	593	50%	120	160	
	553	100% to 50% logarithmic fade	480	240	
Moderate	593	100%	75	181	1860
	553	50%	38	11	
	593	50%	75	100	
	553	100% to 50% logarithmic fade	300	150	
High	593	100%	47	113	1164
	553	50%	23	7	
	593	50%	47	63	
	553	100% to 50% logarithmic fade	188	94	

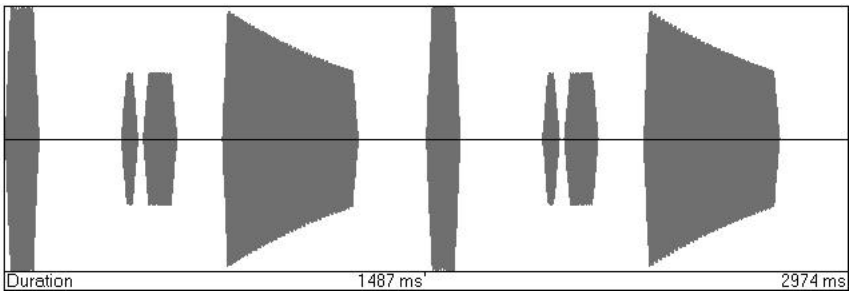


Figure 7. Set II low urgency.

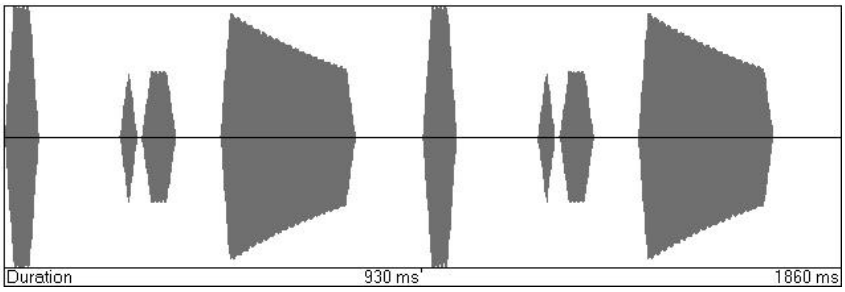


Figure 8. Set II moderate urgency.

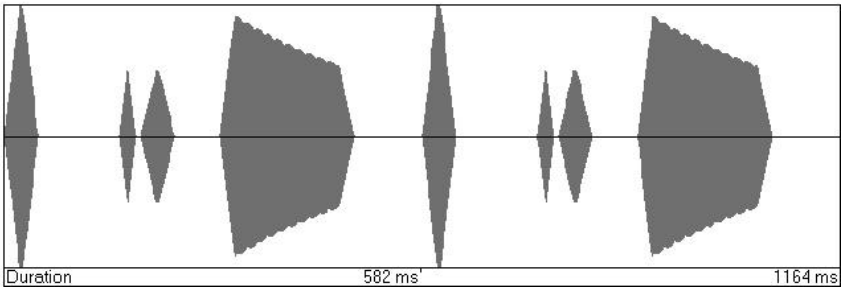


Figure 9. Set II high urgency.

Table 5. Acoustic Parameters of Aural Stimuli Used by Burt (1996) - Set III

Urgency Level	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
Low	675	100%	90	90	4320
	655	50%	90	90	
	635	50%	90	270	
	635	75%	90	90	
	615	50%	90	90	
	598	50%	90	270	
	598	100% to 50% logarithmic fade	180	540	
Moderate	675	100%	56	56	2698
	655	50%	56	56	
	635	50%	56	169	
	635	75%	56	56	
	615	50%	56	56	
	598	50%	56	169	
	598	100% to 50% logarithmic fade	113	338	
High	675	100%	35	35	1686
	655	50%	35	35	
	635	50%	35	106	
	635	75%	35	35	
	615	50%	35	35	
	598	50%	35	106	
	598	100% to 50% logarithmic fade	70	211	

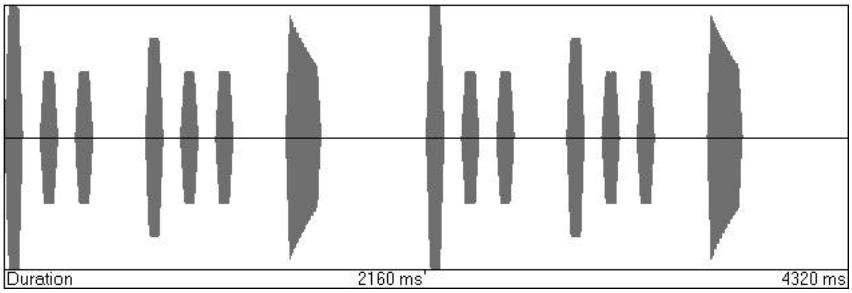


Figure 10. Set III low urgency.

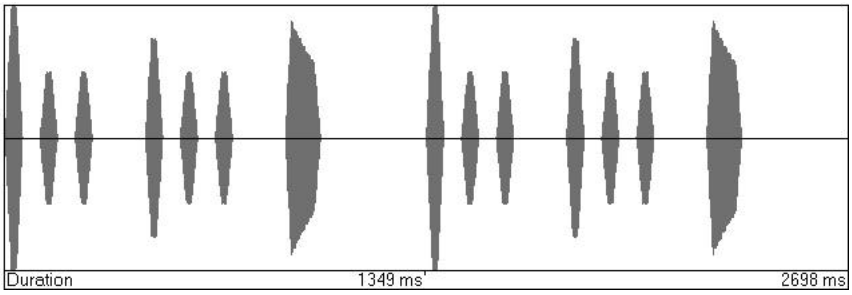


Figure 11. Set III moderate urgency.

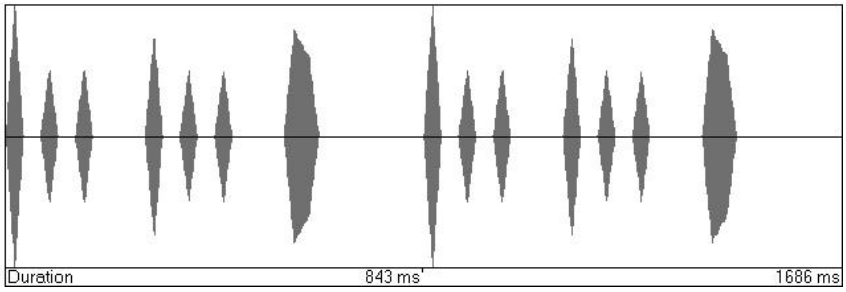


Figure 12. Set III high urgency.

Table 6. Acoustic Parameters of Aural Stimuli Used by Burt (1996) - Set IV

Urgency Level	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
Low	635	100%	90	90	2880
	635	50%	90	270	
	523	50%	90	90	
	523	100% to 50% logarithmic fade	360	360	
Moderate	635	100%	56	56	1798
	635	50%	56	169	
	523	50%	56	56	
	523	100% to 50% logarithmic fade	225	225	
High	635	100%	35	35	1126
	635	50%	35	106	
	523	50%	35	35	
	523	100% to 50% logarithmic fade	141	141	

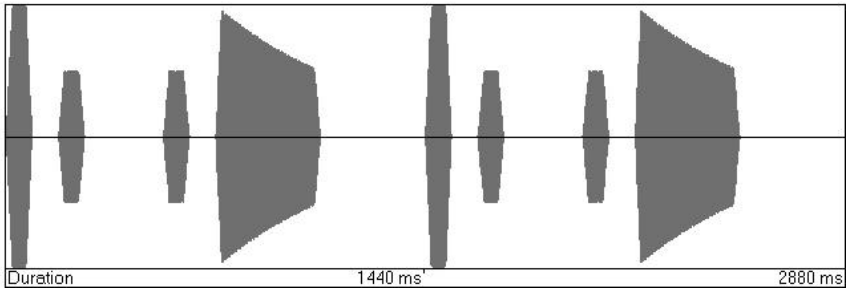


Figure 13. Set IV low urgency.

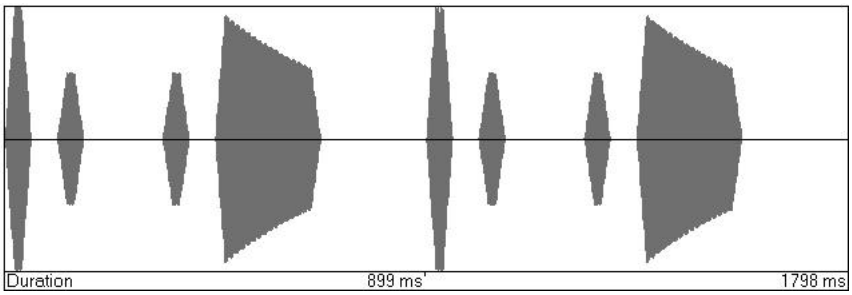


Figure 14. Set IV moderate urgency.

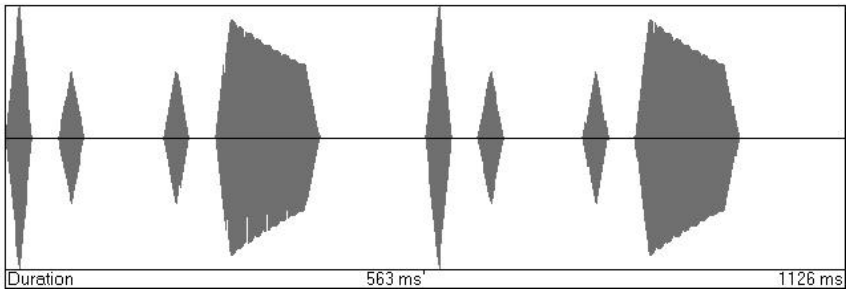


Figure 15. Set IV high urgency.

As shown in Tables 3 - 6, the alerting signals comprising Sets I, II, III, and IV differed from one another in rhythmic pattern and pitch contour as well as in fundamental frequency, pitch range, and duration. Using Patterson's (1982, 1989) guidelines for the creation of nonverbal aural signals as well as the Edworthy et al. (1991) and Edworthy (1994b) guidelines regarding the effects of various acoustic parameter manipulations on the perceived urgency of aural alerts, the construction of the alerting sets began with the development of four distinctive and equally urgent alerting signals or "parent sound bursts."

To develop the parent sound bursts (i.e., Set I Low, Set II Low, Set III Low, and Set IV Low), systematic manipulations of rhythmic pattern and pitch contour were used to create four distinctive aural alerting signals. Then, systematic manipulations of fundamental frequency and pitch range were used to equate the urgency levels of the four alerts since the perception of urgency levels <u>within</u> each alerting set, rather than the overall urgency level differences <u>between</u> alerting sets, was of primary interest. The duration of the alerts varied as a result of rhythmic pattern manipulations. A pretesting endeavor conducted at NASA LaRC revealed that 35 subjects reporting "normal" hearing subjectively rated Set I Low, Set II Low, Set III Low, and Set IV Low as being distinguishable from one another and as having equivalent levels of urgency. Therefore, these alerts were used to create four distinctive aural alerting sets.

The four aural alerting sets were created by constructing two "children sound bursts" from each of the four previously developed parent sound bursts. Hellier et al.'s (1993) guideline regarding the manipulation of an aural alert's speed to double the alert's perceived urgency level was used to create two children sound bursts from each parent sound burst. Therefore, the children sound bursts fundamentally sounded like their respective parent sound burst, but the

children bursts had temporal qualities or speeds (i.e., sound pulse and interpulse durations) intended to convey doublings of urgency level. Each aural alerting set was comprised of one parent sound burst and two children sound bursts: the parent burst was to convey a low level of urgency; one child burst was to convey a moderate level of urgency which, according to the speed manipulation proposed by Hellier et al. (1993), should have been perceived as being twice as urgent as the low urgency level alert; and the other child burst was to convey a high level of urgency which, according to Hellier et al. (1993), should have been perceived as being twice as urgent as the moderate urgency level alert. In effect, the Burt (1996) study attempted to use one of Hellier et al.'s (1993) guidelines regarding the manipulation of physical sound parameters to create four new sets of aural alerts in which the moderate urgency level alert was perceived to be twice as urgent as the low urgency level alert and the high urgency level alert was perceived to be twice as urgent as the moderate urgency level alert.

The stimulus parameters and frequency range corresponded with current research findings (Berson et al., 1981; Boucek et al., 1981; Edworthy, 1994b; Edworthy et al., 1991; Hanson et al., 1983; Hellier et al., 1993; Patterson, 1982, 1989) and design standards (ISO, 1986; SAE, 1993). All alerts were presented over one center, front loudspeaker at 75 dBA, and background noise recorded on the flight deck of NASA 515 (i.e., LaRC's recently retired Boeing 737 research aircraft) during the cruise phase of flight when no conversation or aural alerts were occurring was presented over left and right side loudspeakers at 60 dBA. A signal level of 75 dBA was desirable because it did not endanger the subjects' hearing capabilities and because signal levels that are 15 to 16 dB above a particular masking noise are sufficient for situations involving warning sounds (Fidell, 1978 as cited in Sorkin, 1987; ISO, 1986; Wilkins, 1981 as cited in Wilkins and Acton,

1982). As stated previously, a more detailed account of test facilities and apparatus as well as the presentation of stimuli and background noise is provided in a subsequent "Method" section.

Dependent measures. Three dependent measures were obtained from each of the Burt (1996) study's participants during a single experimental session. These measures included magnitude estimation ratings of aural alert urgency level, pair comparison ratings of aural alert similarity, and identifications of aural alerting set and aural alert urgency level.

Based on the recommendation of Stevens (1971) which suggests that it is better to permit observers to choose their own modulus than it is to designate one for them, urgency ratings were obtained using the free modulus magnitude estimation method. Each subject provided two sets of numerical magnitude estimation values for each aural alert. These values were normalized using the methodology described by Engen (1971) and yielded one set of 12 magnitude estimation values for each experimental session.

Similarity ratings were obtained using a pair comparison task in which each stimulus was paired with every other stimulus. However, subjects were asked to rate the pairs in terms of the sounds' similarity to one another rather than indicate which member of a pair had a greater amount of a particular attribute. Subjects used a linear scale that represented a continuum of similarity to make one pair comparison rating for each pair of alerts. These similarity ratings were converted to numerical values ranging from 0 (i.e., completely different) to 100 (i.e., identical) and yielded one set of 66 [i.e., N (N - 1) / 2] ratings.

Sound identifications were obtained using a sound identification task in which subjects were asked to determine the flight deck function to which each sound corresponded (after participating in a brief training session during which instructions were provided regarding the

correspondence between the alerting sets and the major flight deck functions) and to rate each alert as having either a low, moderate, or high urgency level. The sound identification task provided subjects with an opportunity to demonstrate their ability to associate each alerting set with one of the four major flight deck functions and to simultaneously distinguish among and identify three levels of urgency within each alerting set. Each alert was identified and rated twice, and these data yielded frequency counts of correct and incorrect identifications. Since near perfect identifications and urgency ratings are required for the critical functions associated with flying an aircraft, a value of 95% correct identification was set as the criterion for acceptable performance.

Hypotheses. Magnitude estimation ratings of aural alert urgency level were collected to investigate the following hypotheses:

- It was hypothesized that systematic manipulations of aural alerts' fundamental frequency and pitch range could be used to minimize the overall urgency level differences between alerting sets. Therefore, it was hypothesized that the four aural alerting sets would be perceived as being equally urgent. Specifically, subjects were expected to provide the same urgency rating for Sets I, II, III, and IV.
- It was hypothesized that systematic manipulations of a single aural alert's tempo could be used to convey low, moderate, and high levels of urgency. Therefore, it was hypothesized that the low urgency level alerts would be perceived as being less urgent than the moderate urgency level alerts and the high urgency level alerts and that the moderate urgency level alerts would be perceived as being less urgent than the high urgency level alerts. Specifically, subjects were expected to give Set I Low, Set II Low, Set III Low, and Set IV Low lower

urgency ratings than Set I Moderate, Set II Moderate, Set III Moderate, Set IV Moderate, Set I High, Set II High, Set III High, and Set IV High and were expected to give Set I Moderate, Set II Moderate, Set III Moderate, and Set IV Moderate lower urgency ratings than Set I High, Set II High, Set III High, and Set IV High.

- It was hypothesized that the composite manipulation of aural alerts' fundamental frequency, pitch range, and tempo could be used to create alerting sets that convey equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency. Therefore, it was hypothesized that the alerting sets' low urgency level alerts would be perceived as being equally urgent; that the alerting sets' moderate urgency level alerts would be perceived as being equally urgent; and that the alerting sets' high urgency level alerts would be perceived as being equally urgent. Specifically, subjects were expected to provide the same urgency rating for Set I Low, Set II Low, Set III Low, and Set IV Low; provide the same urgency rating for Set I Moderate, Set II Moderate, Set III Moderate, and Set IV Moderate; and provide the same urgency rating for Set I High, Set II High, Set III High, and Set IV High.

Pair comparison ratings of aural alert similarity were collected to investigate the following hypothesis:

- It was hypothesized that systematic manipulations of aural alerts' rhythmic patterns and pitch contours could be used to create four distinctive sets of aural alerting signals. Therefore, it was hypothesized that the three alerts comprising each alerting set would be perceived as being more similar to one another than to any other alert and that each alerting set would be perceived as being different from every other alerting set. Specifically, subjects were expected

to rate the alerts in Set I (i.e., Set I Low, Set I Moderate, and Set I High) as being similar to one another and different from the alerts in Sets II, III, and IV; rate the alerts in Set II (i.e., Set II Low, Set II Moderate, and Set II High) as being similar to one another and different from the alerts in Sets I, III, and IV; rate the alerts in Set III (i.e., Set III Low, Set III Moderate, and Set III High) as being similar to one another and different from the alerts in Set I, II, and IV; and rate the alerts in Set IV (i.e., Set IV Low, Set IV Moderate, and Set IV High) as being similar to one another and different from the alerts in Sets I, II, and III.

Sound identifications of aural alerting set and aural alert urgency level were collected to investigate the following hypothesis:

- It was hypothesized that subjects would be able to: 1) associate each alerting set with one of the four major flight deck functions, and 2) simultaneously recognize a given alerting set as well as identify the correct urgency level within the set. Therefore, it was hypothesized that subjects would identify the correct aural alerting set, the correct urgency level, as well as the correct alerting set <u>and</u> urgency level 95% of the time and that subjects would correctly identify each alerting set, urgency level, and aural alert equally often. Specifically, subjects were expected to choose the correct alerting set in 95% of the trials; choose the correct urgency level in 95% of the trials; and choose the correct alerting set <u>and</u> the correct urgency level in 95% of the trials. Furthermore, subjects were not expected to correctly identify one alerting set, urgency level, or aural alert more often than any other alerting set, urgency level, or aural alert.

Sound identification data were also collected to provide further insight into the ways in which systematic manipulations of acoustic parameters affect aural alert perception and recognition.

Specifically, the ease with which subjects were able to correctly identify Sets I, II, III, and IV provided information regarding the distinctiveness of the alerts' rhythmic patterns and pitch contours; the ease with which subjects were able to correctly identify low, moderate, and high urgency levels provided information regarding the urgency levels associated with various tempos; and the ease with which subjects were able to correctly identify low, moderate, and high urgency levels within Sets I, II, III, and IV provided information regarding the ability of subjects to associate each alerting set with one of the major flight deck functions and to simultaneously identify the correct alerting sets and urgency levels corresponding to various alerts when appropriate acoustic parameters were manipulated.

Results and discussion. Urgency ratings obtained during the magnitude estimation task were analyzed by way of a 4 (Alerting Set) x 3 (Urgency Level) Analysis of Variance (ANOVA). The ANOVA Summary Table is presented in Table 7.

Table 7. ANOVA Summary Table of Magnitude Estimation Data Collected by Burt (1996)

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	31.69				
Within-Subject						
Alerting Set (AS)	3	0.22	0.07	2.74	0.059	
AS X S	33	0.88	0.03			
Urgency Level (UL)	2	4.71	2.36	51.13	0.0001	0.0001
UL X S	22	1.01	0.05			
UL X AS	6	0.12	0.02	1.45	0.21	
UL X AS X S	66	0.91	0.01			
Total	143	39.54				

As shown in Table 7, no significant differences were found among urgency ratings of alerting set, and no significant interaction between urgency ratings of alerting set and urgency level occurred. However, a significant difference existed among urgency ratings of urgency level, even after a Greenhouse-Geisser Epsilon value was used to correct the problems associated with a positively biased F-Test (F [1, 13] = 51.13; $p < 0.05$). Therefore, the null hypothesis (i.e., all urgency levels would receive the same urgency rating) was rejected.

To determine which urgency levels were perceived as being significantly different from one another, a Bonferroni t-Test post-hoc analysis was performed. This statistical procedure was appropriate for evaluating a series of post-hoc comparisons while controlling for inflated alpha error, and it allowed an interpretation of whether or not the low, moderate, and high urgency level alerts were perceived as such by the subjects. Based on this test, a significant difference was found to exist between urgency ratings of the low urgency level alerts and the moderate urgency level alerts as well as between the low urgency level alerts and the high urgency level alerts.

The results of the ANOVA and the Bonferroni t-Test partially support the experimental hypotheses regarding the urgency ratings of the alerting sets and the urgency levels. Specifically, the failure to reject the null hypothesis that each alerting set would receive the same urgency rating supported the expectation that systematic manipulations of aural alerts' fundamental frequency and pitch range could be used to minimize the overall urgency level difference between alerting sets. Also, the rejection of the null hypothesis that all urgency levels would receive the same urgency rating supported the expectation that systematic manipulations of an alert's tempo could be used to convey various levels of urgency. However, the failure to find a significant difference between the ratings provided for the moderate urgency level and high urgency level

alerts does not support the expectation that low, moderate, and high urgency levels would be perceived within each alerting set. Finally, the failure to reject the null hypothesis that each low urgency level alert would receive the same urgency rating, that each moderate urgency level alert would receive the same urgency rating, and that each high urgency level alert would receive the same urgency rating supported the expectation that the composite manipulation of aural alerts' fundamental frequency, pitch range, and tempo could be used to create alerting sets capable of conveying equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency.

Similarity ratings obtained during the pair comparison task were analyzed by way of a Multidimensional Scaling Analysis (MDS). MDS is a quantitative method used to identify the perceived structure in a set of stimuli by geometrically representing stimuli within an n-dimensional spatial configuration (Schiffman, Reynolds, and Young, 1981). This geometric representation is based on measures of perceived similarity between stimuli such that similar stimuli are positioned close to one another in multidimensional space and dissimilar stimuli are positioned far apart. Subsequent examination of relationships among stimuli within the spatial configuration is useful for identifying stimulus properties that affect subjects' perceptual judgements. Therefore, objective numerical procedures are used to create the geometric arrangement of stimuli, and statistical treatment of the similarity ratings determines the number of dimensions (i.e., axes defining the coordinates of a geometrical space) needed to adequately describe the relations among a variety of stimuli (Schiffman et al., 1981).

Once a scaling solution was determined, two measures of fit between similarity data and interstimulus distances in the spatial configurations of various dimensions were examined so that a

decision regarding the number of dimensions that adequately described the relations among the 12 aural stimuli could be made. The two measures of fit included Kruskal stress (i.e., an index with values from 0 to 1.0 that describes the discrepancy between proximity measures derived from raw similarity data and interpoint Euclidian distances among stimuli) and R^2 (i.e., the proportion of variance accounted for). Smaller stress values and larger R^2 values correspond to a better fit or more appropriate solution. Therefore, as a result of a dramatic decrease in stress and dramatic increase in R^2 at two dimensions, a two-dimensional solution was retained for interpretation. In choosing the two-dimensional solution, 79% of the cumulative variance was accounted for when a stress value of 0.177 out of a possible value of 1.0 was assumed.

A plot of the two-dimensional solution is shown in Figure 16.

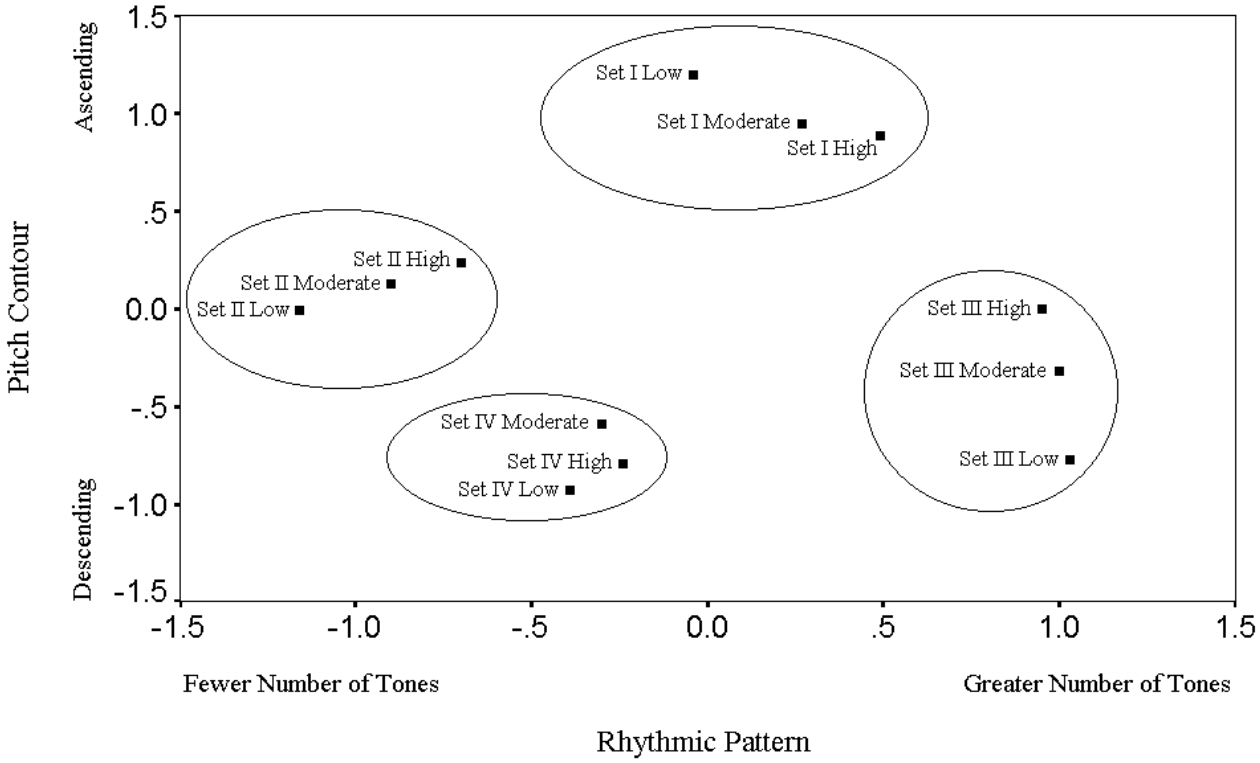


Figure 16. Two-dimensional solution for pair comparison similarity ratings obtained by Burt (1996).

In this plot, the stimulus points refer to the 12 aural alerts, and the axes represent dimensions or stimulus attributes used by the subjects to classify each alert. The alerts are arranged in the spatial configuration according to how they relate to the two stimulus attributes. It was assumed that the perceptions of aural alert similarity would be primarily influenced by the alerts' rhythmic patterns and pitch contours. Therefore, the two-dimensional solution suggested by the stress and R^2 values was also identified as adequately describing the relations among the 12 stimuli since the locations of the stimulus points within the two-dimensional space conformed nicely to these two stimulus properties. Referring to Figure 16, it can be seen that the three aural alerts associated with each of the four alerting sets are positioned closer to one another than to the other aural alerts and that the acoustic characteristics of each alerting set are appropriately represented within the stimulus attribute space. Specifically, the spatial locations of the aural alerts in Set I correspond to the fact that the three alerts comprising Set I consisted of a six tone rhythmic pattern with an ascending pitch contour. The spatial locations of the aural alerts in Set II correspond to the fact that the three alerts comprising Set II consisted of a four tone rhythmic pattern with an ascending and descending pitch contour. The spatial locations of the aural alerts in Set III correspond to the fact that the three alerts comprising Set III consisted of a seven tone rhythmic pattern with a descending pitch contour. The spatial locations of the aural alerts in Set IV correspond to the fact that the three alerts comprising Set IV consisted of a four tone rhythmic pattern with a descending pitch contour.

Since 79% of the cumulative variance was accounted for, the results of the pair comparison similarity ratings generally support the hypothesis that systematic manipulations of aural alerts' rhythmic patterns and pitch contours could be used to create four distinctive sets of

aural alerting signals. Specifically, the predictions that the three alerts comprising each alerting set would be perceived as being more similar to one another than to any other alert and that each alerting set would be perceived as being different from every other alerting set were supported by the spatial locations of Set I Low, Set I Moderate, and Set I High; Set II Low, Set II Moderate, and Set II High; Set III Low, Set III Moderate, and Set III High; and Set IV Low, Set IV Moderate, and Set IV High.

Sound identifications of each aural alert yielded frequency counts of correct and incorrect identifications that were used to determine the extent to which subjects were able to correctly identify the alerting set (i.e., major flight deck function) to which an aural alert corresponded; correctly identify low, moderate, and high urgency levels; and correctly identify the alerting set and urgency level associated with a given alert. These data revealed that subjects correctly identified the alerting set to which a signal corresponded in 95.14% of the trials when identifications were averaged across the three urgency levels (i.e., N = 72); that subjects correctly identified the low, moderate, and high urgency levels in 89.24% of the trials when identifications were averaged across the four alerting sets (i.e., N = 96); and that subjects correctly identified the alerting set and urgency level associated with a given alert in 85.07% of the trials (i.e., N = 24).

Cochran's Q Tests (i.e., nonparametric within-subject tests appropriate for analyzing three or more related samples of nominal data) were performed on the mean percentages of correct identifications to determine if a given alerting set, urgency level, or aural alert was correctly identified more often by the subjects (Norusis, 1992). A significant difference was found to exist among identifications of urgency level (Q [2] = 7.000; $p < 0.05$); however, no significant differences were found among identifications of alerting set or among identifications of alerting

set and urgency level. To determine which urgency level(s) were correctly identified more often by the subjects, McNemar Tests (i.e., nonparametric within-subject tests appropriate for analyzing two related samples of nominal data) were performed (Norusis, 1992). These tests revealed that low urgency level alerts were correctly identified more often than moderate urgency level alerts ($p < 0.05$).

The results of the percentages of correct identifications, the Cochran's Q Tests, and the McNemar Tests partially support the expectation that subjects would be able to associate each alerting set with one of the four major flight deck functions and would also be able to simultaneously recognize a given aural alerting set as well as identify the correct urgency level within the set. The hypothesis that subjects would identify the correct aural alerting set, the correct urgency level, as well as the correct alerting set and urgency level 95% of the time was partially supported by evidence suggesting that subjects identified the correct alerting set in 95.14% of the trials. This hypothesis was not supported, however, by evidence suggesting that subjects identified the correct urgency level in 89.24% of the trials or by evidence suggesting that subjects identified the correct alerting set and urgency level in 85.07% of the trials. The hypothesis that subjects would correctly identify each alerting set, urgency level, and aural alert equally often was partially supported by evidence suggesting that subjects correctly identified each alerting set as well as each aural alert equally often. But, this hypothesis was not supported by evidence suggesting that subjects correctly identified low urgency level alerts more often than moderate urgency level alerts.

The results of the sound identification task also provided partial support for experimental hypotheses regarding the use of acoustic parameter manipulations to create sets of aural alerts

that are identifiable and that are comprised of three alerts that convey low, moderate, and high urgency levels. The hypothesis that rhythmic pattern and pitch contour could be effectively manipulated to create four distinctive alerting sets was supported by evidence suggesting that subjects correctly identified the alerting set to which an alert corresponded in 95.14% of the trials as well as by evidence suggesting that no alerting set was correctly identified more often than any other alerting set by the subjects. The hypothesis that tempo could be effectively manipulated to convey low, moderate, and high urgency levels within each alerting set was partially supported by evidence suggesting that subjects correctly identified low, moderate, and high urgency level alerts in 89.24% of the trials. But, this hypothesis was not supported by evidence suggesting that low urgency level alerts were correctly identified more often than moderate urgency level alerts. Finally, the hypothesis that subjects would be able to associate each alerting set with one of the four major flight deck functions as well as simultaneously identify the correct alerting set to which an alert corresponded and rate the alert as conveying the correct level of urgency when appropriate acoustic parameters are manipulated was partially supported by evidence suggesting that subjects identified the correct alerting set and urgency level associated with a given alert in 85.07% of the trials and by evidence suggesting that no aural alert was correctly identified more often than any other aural alert by the subjects.

Conclusions. As a result of the study conducted by Burt (1996), several advances were made toward the design and development of an aural alerting signal categorization scheme in which a distinctive aural alert is associated with each of the four major flight deck functions and the acoustic parameters of a given alert are manipulated to form an alerting set capable of

conveying low, moderate, and high levels of urgency. In the paragraphs that follow, specific advances as well as related hypotheses and results are discussed.

The first hypothesis investigated by Burt (1996) was that systematic manipulations of aural alerts' rhythmic patterns and pitch contours could be used to create four distinctive sets of aural alerting signals. This hypothesis was supported by pair comparison similarity ratings which revealed that the aural alerts comprising each alerting set (e.g., Set I Low, Set I Moderate, and Set I High) were perceived as being more similar to one another than to any other alert and that each alerting set was perceived as being different from every other alerting set. This hypothesis was also supported by sound identification data which revealed that subjects correctly identified the alerts corresponding to Sets I, II, III, and IV in 95.14% of the trials and correctly identified each alerting set equally often. It is suggested, therefore, that the rhythmic pattern and pitch contour manipulations used to create the aural alerts investigated in this study can be used to form four distinctive alerting sets. However, it is acknowledged that the manipulations of the alerts' fundamental frequency and pitch range used to equate the overall urgency levels of the alerting sets may have also contributed to the distinctiveness of the alerting sets, even though no consistent effects of fundamental frequency and pitch range were evident in the subjects' ratings of alert similarity. But regardless of whether distinctive alerting sets were created through manipulations of rhythmic pattern and pitch contour or through manipulations of rhythmic pattern and pitch contour as well as manipulations of fundamental frequency and pitch range, it is concluded that the acoustic parameters of four distinctive aural alerting sets have been identified.

The second hypothesis investigated by Burt (1996) was that systematic manipulations of aural alerts' fundamental frequency and pitch range could be used to minimize the overall urgency

level differences between alerting sets. This hypothesis was supported by magnitude estimation urgency ratings which revealed that subjects provided the same urgency ratings for Set I, Set II, Set III, and Set IV when ratings were averaged across the three urgency levels. It is acknowledged, however, that the rhythmic pattern and pitch contour manipulations used to create the four distinctive alerting sets may have also contributed to the alerting sets' overall urgency levels, even though such an effect was not hypothesized. But regardless of whether the overall urgency levels of the alerting sets were equated through manipulations of fundamental frequency and pitch range or through manipulations of fundamental frequency and pitch range as well as manipulations of rhythmic pattern and pitch contour, no additional manipulations of acoustic parameters were needed to equate the urgency levels of Sets I, II, III, and IV. Therefore, it is concluded that acoustic parameters capable of equating the overall urgency levels of the alerting sets investigated in this study have been identified.

The third hypothesis investigated by Burt (1996) was that the composite manipulation of aural alerts' fundamental frequency, pitch range, and tempo could be used to create alerting sets capable of conveying equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency. This hypothesis was supported by magnitude estimation urgency ratings which revealed that subjects provided the same urgency ratings for Set I Low, Set II Low, Set III Low, and Set IV Low; provided the same urgency ratings for Set I Moderate, Set II Moderate, Set III Moderate, and Set IV Moderate; and provided the same urgency ratings for Set I High, Set II High, Set III High, and Set IV High. It is acknowledged, however, that the rhythmic pattern and pitch contour manipulations used to create the four distinctive alerting sets may have also contributed to the alerting sets' equivalent levels of low urgency, equivalent levels

of moderate urgency, and equivalent levels of high urgency, even though such an effect was not hypothesized. The alerting sets' equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency may have resulted from the composite manipulation of the aural alerts' fundamental frequency, pitch range, and tempo or from the composite manipulation of the alerts' fundamental frequency, pitch range, tempo, rhythmic pattern, and pitch contour. However, regardless of which composite manipulation was responsible for the results, no additional acoustic parameter manipulations were needed to equate the low urgency levels, the moderate urgency levels, or the high urgency levels conveyed by the alerting sets investigated in this study. It is concluded, therefore, that a composite manipulation of acoustic parameters capable of equating the alerting sets' low urgency level alerts, the alerting sets' moderate urgency level alerts, and the alerting sets' high urgency level alerts has been identified.

The fourth hypothesis investigated by Burt (1996) was that systematic manipulations of a single aural alert's tempo could be used to convey low, moderate, and high levels of urgency. This hypothesis was partially supported by magnitude estimation urgency ratings which revealed that subjects provided significantly different ratings of urgency for the low urgency level alerts and the moderate urgency level alerts as well as for the low urgency level alerts and the high urgency level alerts. However, this hypothesis was not supported by magnitude estimation urgency ratings which revealed that subjects did not provide significantly different ratings of urgency for the moderate urgency level alerts and the high urgency level alerts. Furthermore, this hypothesis was not supported by sound identification data which revealed that subjects correctly identified low, moderate, and high urgency levels in only 89.24% of the trials (i.e., in comparison to the 95%

target) and had particular difficulty with the correct identification of the moderate urgency level alerts when compared to the correct identification of the low urgency level alerts. It is suggested that the Hellier et al. (1993) guideline used to create alerts having tempos, or speeds, intended to convey doublings of urgency level does not provide for an adequate "spread" of urgency-related parameters (i.e., as urgency levels increase, more dramatic parameter changes are needed to produce a unit change in perceived urgency). It is concluded, therefore, that further acoustic manipulations are needed to convey low, moderate, and high urgency levels within each alerting set.

The final hypothesis investigated by Burt (1996) was that subjects would associate each alerting set with one of the four major flight deck functions as well as simultaneously identify the correct alerting set to which an alert corresponded and rate the alert as conveying the correct level of urgency. This hypothesis was supported by sound identification data which revealed that subjects correctly identified the alerting set to which an alert corresponded in 95.14% of the trials and correctly identified each alerting set equally often. This evidence suggests that subjects successfully associated each of the four alerting sets with a major flight deck function and recognized the three alerts within each alerting set. However, subjects correctly identified the urgency level associated with a given alert in only 89.24% of the trials and correctly identified the alerting set <u>and</u> urgency level to which an alert corresponded in only 85.07% of the trials. This evidence suggests that the ability of subjects to simultaneously identify the correct alerting set to which an alert corresponded as well as rate the alert as conveying the correct level of urgency was impeded during the experimental session. The subjects' failure to differentiate between the moderate urgency level alerts and the high urgency level alerts represents the most likely

explanation for the incorrect identification of the urgency level associated with a given alert in 10.76% of the trials and the incorrect identification of the alerting set and urgency level associated with a given alert in 14.93% of the trials. Therefore, it is concluded that the ability of subjects to associate each alerting set with one of the four major flight deck functions and identify the correct alerting set associated with each alert was demonstrated. But, it is suggested that the ability of subjects to identify the correct urgency level associated with each alert as well as the correct alerting set and urgency level associated with each alert will improve once the systematic manipulations needed to differentiate the urgency ratings of the moderate urgency level alerts and the high urgency level alerts are identified and implemented.

Recommendations for additional research. One implication for future research identified through the work of Burt (1996) is the need to investigate how further acoustic manipulations can be used to convey low, moderate, and high urgency levels within each of the proposed alerting sets. The manipulation of an acoustic parameter, such as harmonic content, in addition to tempo could be used to differentiate the moderate urgency level alerts and the high urgency level alerts, but such a manipulation might negatively influence the distinctiveness and overall urgency levels of the alerting sets. The manipulation of the high urgency level alerts' tempos could be used to differentiate the moderate urgency level alerts and the high urgency level alerts, but the perceived distinctiveness of the alerting sets might degrade as the high urgency level alerts' tempos increase. Therefore, it is suggested that additional research be conducted to determine if systematic decreases in the moderate urgency level alerts' tempos (i.e., systematic increases in the moderate urgency level alerts' sound pulse and interpulse interval durations and, consequently, in the moderate urgency level alerts' overall durations) represent appropriate manipulations through

which to differentiate the urgency ratings of the moderate urgency level alerts and the high urgency level alerts. Since it is important that the low urgency level alerts receive subjective urgency ratings that are significantly lower than those given to the moderate urgency level alerts, it is also suggested that requirements for systematic decreases in the tempos of the low urgency level alerts be investigated as well.

A second implication for future research identified by Burt (1996) is the need to investigate the ability of individuals to identify the correct alerting set and urgency level associated with each aural alert when various workload levels and attentional demands are experienced. Since a pilot's level of workload changes within and across the different phases of flight (i.e., takeoff, initial climb, climb, cruise, descent, initial approach, final approach, landing, and taxi), it is important to evaluate the accuracy with which alerting sets and urgency levels may be identified during different loading conditions. Therefore, it is suggested that additional research be conducted to examine the ability of individuals to identify the correct alerting set and urgency level associated with each aural alert while a supervisory task that requires various levels of workload and attentional engagement is performed.

Mental Workload and Attentional Engagement

Mental workload is a multidimensional construct that is assumed to be caused by task demands (i.e., input driven) and related to task accomplishments (i.e., output oriented). It is also assumed that mental workload is responsible for an individual's level of attentional engagement and mental effort or loading (Wierwille, 1995).

The Yerkes-Dodson law states that there is an optimal level of arousal associated with the best performance of any given task. Similarly, optimal levels of mental workload may be achieved. For example, an individual's performance is typically high, and errors are typically low when moderate levels of workload are experienced. But when workload demands are too low or too high, performance typically decreases, and errors are more likely to occur. Assessments of mental workload are primarily used to determine if an individual's workload level is within a desirable range. However, assessments may also be used to examine the level of attentional engagement experienced by an individual when various activities are performed.

EEG as a measure of attentional engagement. In opposition to a strict behaviorist viewpoint, it may be assumed that behavior, such as the act of attending to a stimulus or set of stimuli, is the result of ongoing mental processes. According to Andreassi (1989), "[p]sychophysiology is the study of relations between psychological manipulations and resulting physiological responses, measured in the living organism, to promote an understanding of the relations between mental and bodily processes" (p.2). Therefore, a psychophysiological measure of attentional engagement is appropriate because mental workload is assumed to be linked to the bodily processes of the human being performing a given task. The electroencephalogram (EEG) represents a psychophysiological measure that may be used to assess attentional engagement.

The EEG, or brain wave, represents the electrical activity of the brain, and since the work of Berger (1929), EEG has been used clinically and experimentally as an unobtrusive psychophysiological technique for estimating arousal and its relation to attention and effort (Andreassi, 1989; Kramer, 1991). Unlike the electrocorticogram which measures brain activity from the surface of the cortex, EEG is measured using electrodes attached to the surface of the

scalp. Typically, electrodes are arranged according to the standardized “International 10-20” placement system (Jasper, 1958). When electrodes are arranged in the manner prescribed by this system, as shown using the single plane projection of the head depicted in Figure 17, electrode placement is based on each subject’s head size, and investigators are able to communicate the sites used in EEG research in a standard way.

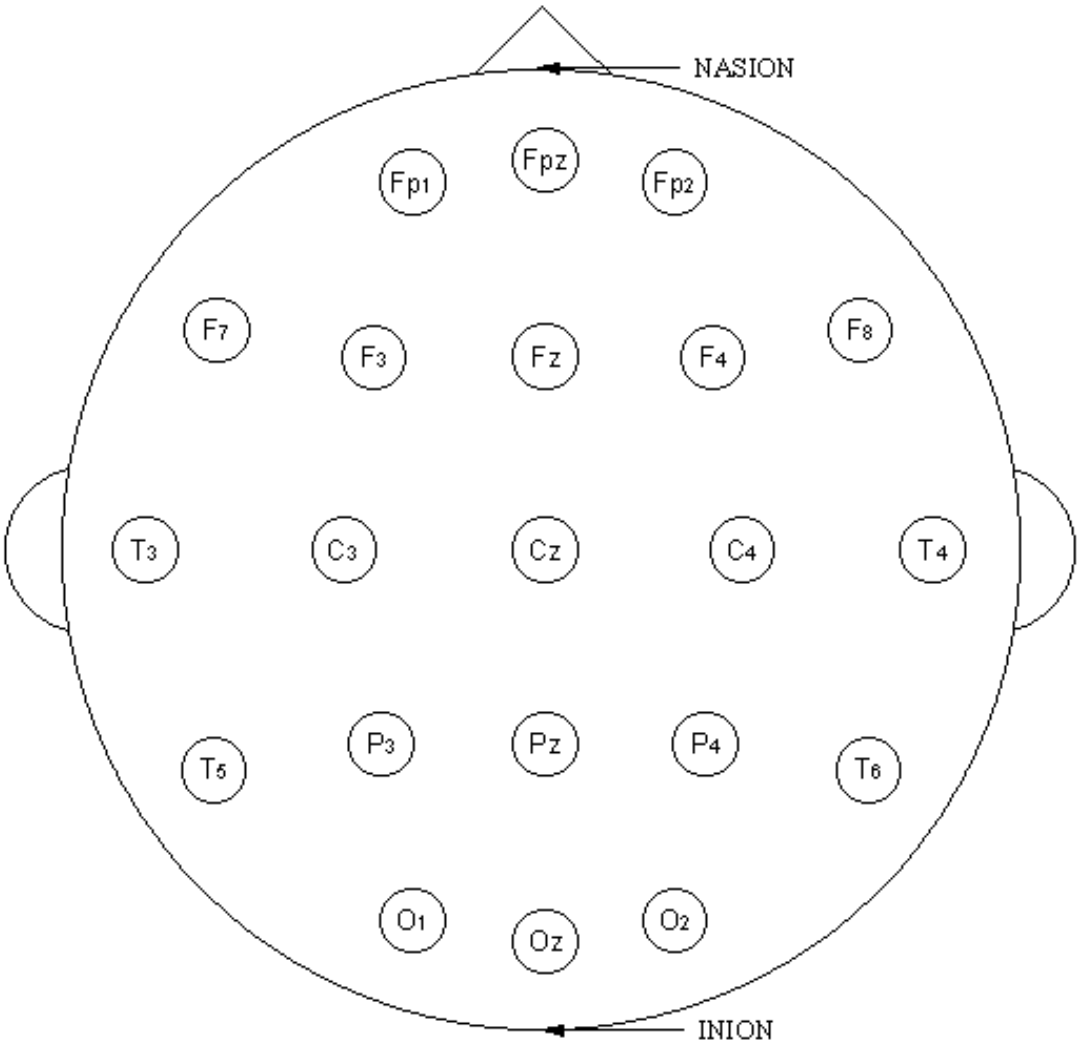


Figure 17. “International 10-20” electrode placement system for EEG examinations.

As shown in Figure 17, EEG data are typically recorded from the occipital (O), parietal (P), central (C), temporal (T), frontal (F), and/or prefrontal (Fp) areas of the cortex. Sites labeled with a "z" are located along the midline between the nasion (i.e., bridge of the nose) and the inion (i.e., occipital protuberance), and site Fpz represents the frontal pole of the brain. Locations on the right side of the head are labeled with even numbers, and locations on the left side of the head are labeled with odd numbers.

In addition to the active electrodes placed on the scalp, a reference electrode is attached to a relatively inactive area such as the earlobe, tip of the nose, or mastoid prominence, and a ground electrode is attached to the scalp, an unused earlobe, an unused mastoid prominence, or the wrist. Once electrodes are making good contact, as demonstrated by the resistance between the active and reference electrodes, the electroencephalograph compares the activity measured from the cortical areas with the activity measured at the reference electrode. The difference between these activity levels represents the EEG record that is generated.

The three most reliable brain wave patterns, in terms of consistency of occurrence, that are found in waking individuals are referred to as theta waves, alpha waves, and beta waves. Theta waves occur at a frequency of 4 - 7 Hz (i.e., four to seven complete cycles each second); alpha waves occur at a frequency of 8 - 13 Hz; and beta waves occur at a frequency of 14 - 30 Hz (Andreassi, 1989). An EEG frequency analysis involves determining the amount of activity associated with the theta, alpha, and beta frequency bands within an EEG segment; and according to Empson (1986), such an analysis can serve as a powerful tool in the investigation of mental processes. An increased amount of theta activity indicates a state of diminished arousal; increased alpha activity indicates a state of relaxed wakefulness; and increased beta activity indicates a state

of alert attentiveness (Alluisi, Coates, & Morgan, 1977; Beatty, Greenberg, Diebler, & O'Hanlon, 1974; Lindsley, 1960; Lubar 1991; O'Hanlon & Beatty, 1979; O'Hanlon, Royal, & Beatty, 1977; Ray, 1990).

Manipulating workload and attentional engagement. Since mental workload is assumed to be caused by task demands, an individual's level of attentional engagement may be manipulated by altering the workload requirements, or demands, associated with the execution of a given task. The Multi-Attribute Task (MAT) Battery developed by Comstock and Arnegard (1992) is a computer program that provides a benchmark set of tasks that can be incorporated into studies of operator workload. The MAT Battery's tasks, which may be presented on a video or computer monitor, are analogous to activities that pilots perform during flight (i.e., monitoring, tracking, communication, and resource management). The option of manual or automated control of tasks is provided, and non-pilot test subjects may participate in investigations making use of these tasks.

The tracking component of the MAT Battery depicted in Figure 18 is a compensatory tracking task that simulates the demands of manual control.

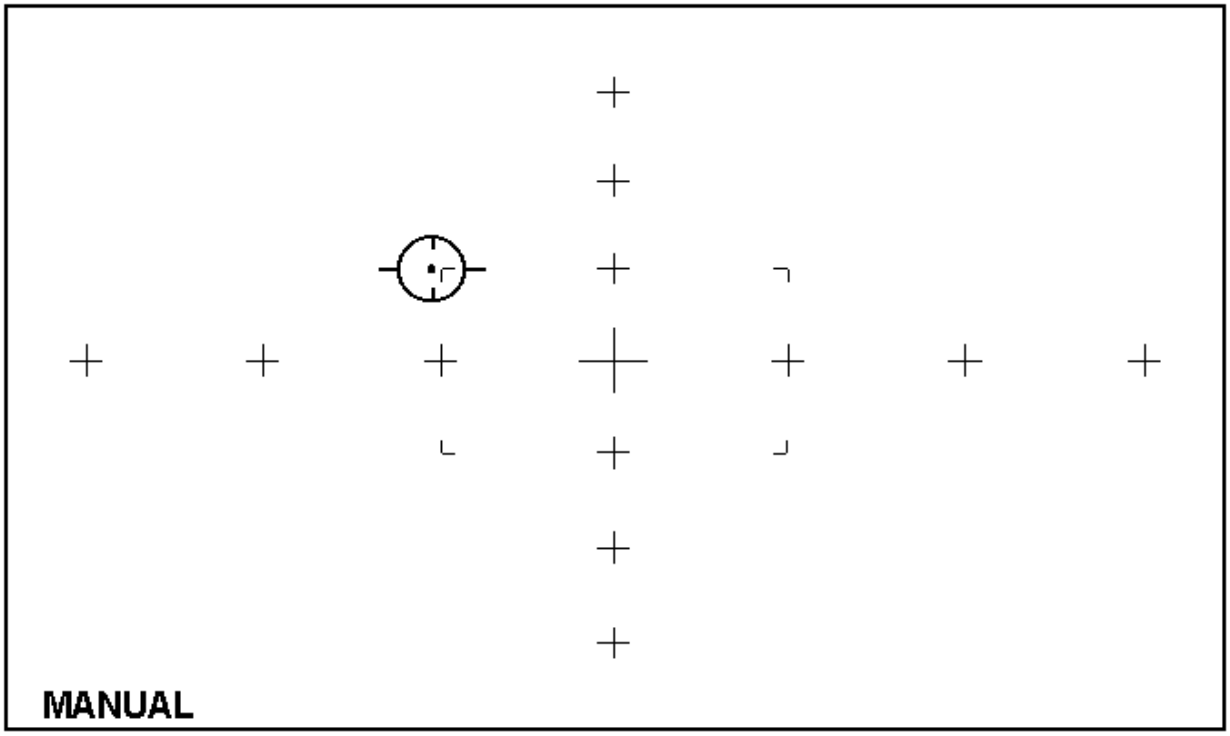


Figure 18. MAT Battery's tracking task display.

The overall purpose of the tracking task is to keep the airplane symbol, represented by the circle, within the dotted rectangular area in the center of the task. This task can be operated in either manual or automatic mode, and the current mode is shown by “MANUAL” or “AUTO” displayed in the lower left corner of the window. In the manual mode, the subject is required to use a joystick to keep the target at the center of the window. In the automatic mode, no action is required of the subject; the task is automated to simulate the reduced manual demands of autopilot.

Burt et al. (1995) used the MAT Battery’s tracking task to manipulate attentional engagement during an investigation of auditory warnings in which EEG data were recorded. Each condition of the tracking task (i.e., manual and automated) lasted 15 min, and each of three tonal warnings was associated with a probability (0.90, 0.50, or 0.10) of tracking system failure (i.e., loss of joystick control of target, or uncontrolled target drift). During the manual condition, subjects were required to use a joystick to manually track the circular airplane target displayed on a computer monitor and were required to press a button on the joystick to resume tracking when joystick control of the target was lost. During the automated condition, subjects were required to monitor computer tracking of the airplane target and were required to provide a button-press response when the target started to drift outside of the dotted rectangular area in the center of the task. Continuous EEG data were recorded during both the manual and automated conditions of the tracking task to assess the ongoing physiological responses of attention to the three auditory warnings and each tracking task condition.

With respect to the conditions of the tracking task, an analysis of EEG frequency band activity revealed that significant brain wave changes occurred in response to the manual and

automated conditions. Based on the work of Makeig, Elliott, Inlow, and Kobus (1990) regarding alertness, EEG data from the vertex region were chosen, and data from sites Cz (midline central), Fz (midline frontal), and F3 (left frontal) were analyzed by determining the amount of activity associated with the theta, alpha, and beta frequency bands during each tracking task condition. Data recorded at site Fz revealed that significantly more alpha activity occurred during the automated condition of the tracking task than during the manual condition of the tracking task (F [1, 5] = 7.26; $p < 0.05$), and data recorded at site F3 revealed that significantly more alpha activity occurred during the automated condition of the tracking task than during the manual condition of the tracking task (F [1, 5] = 7.59; $p < 0.05$). Although analyses involving theta and beta revealed non-significant results at sites Cz, Fz, and F3, previous research suggests that alpha may be the frequency which best reflects attentional differences (Mulholland, 1983). Overall, the results of the EEG data collected by Burt et al. (1995) suggest that the MAT Battery's tracking task may be used to effectively manipulate an individual's level of attentional engagement such that a significantly more relaxed state of awareness is experienced during the automated condition of the tracking task in comparison to the manual condition of the tracking task.

In addition to using EEG data to investigate the ability of the MAT Battery's tracking task to alter attentional engagement, the battery also presents a computerized version of Hart and Staveland's (1988) NASA Task Load Index (NASA TLX) which may be used to obtain a self reported assessment of the level of workload experienced by an individual during task performance. The multi-dimensional rating scale, depicted in Figures 19 and 20, may be presented during or after the completion of the MAT Battery's task(s) using a single computer monitor.

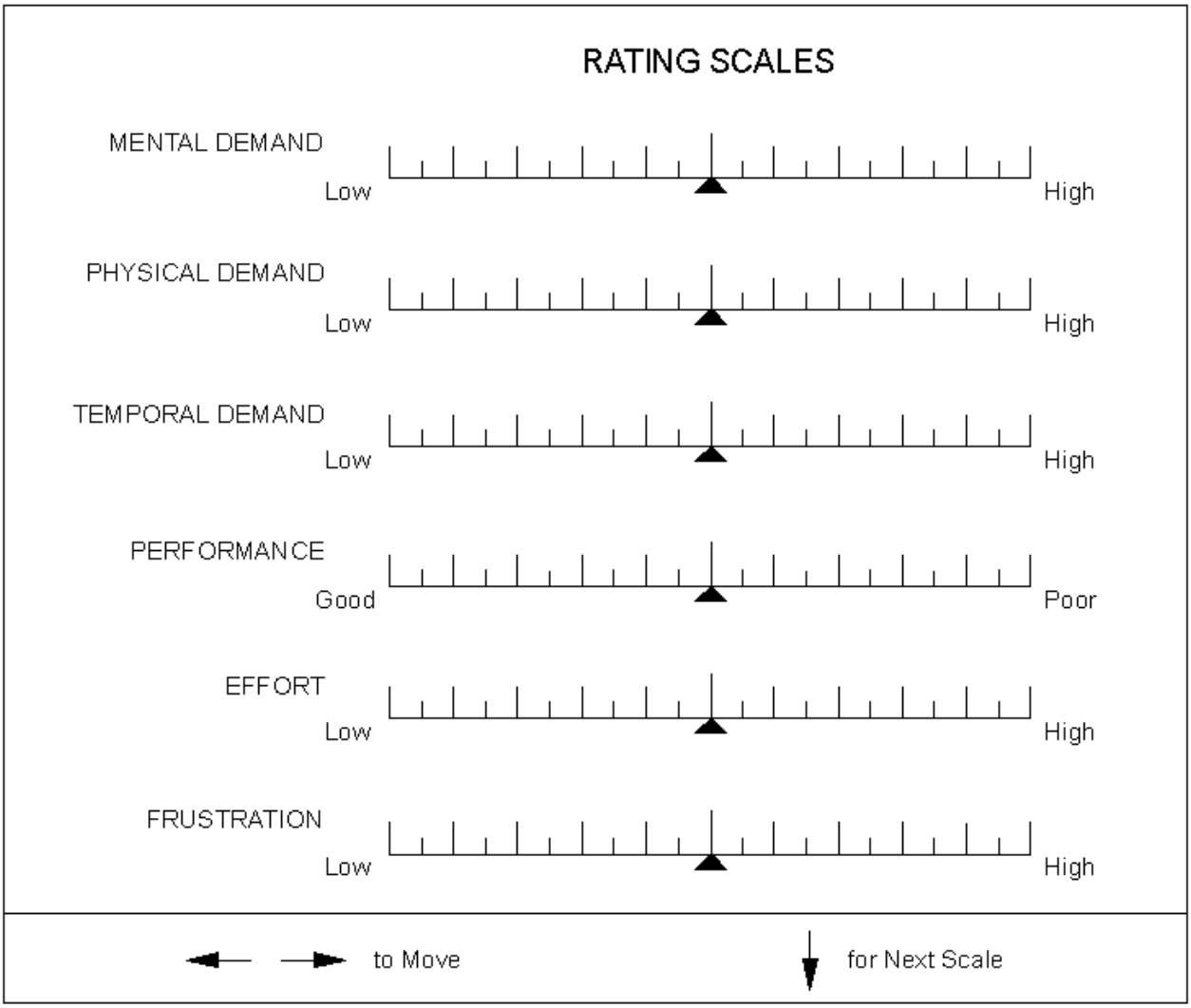


Figure 19. NASA TLX workload rating screen displaying initial instructions.

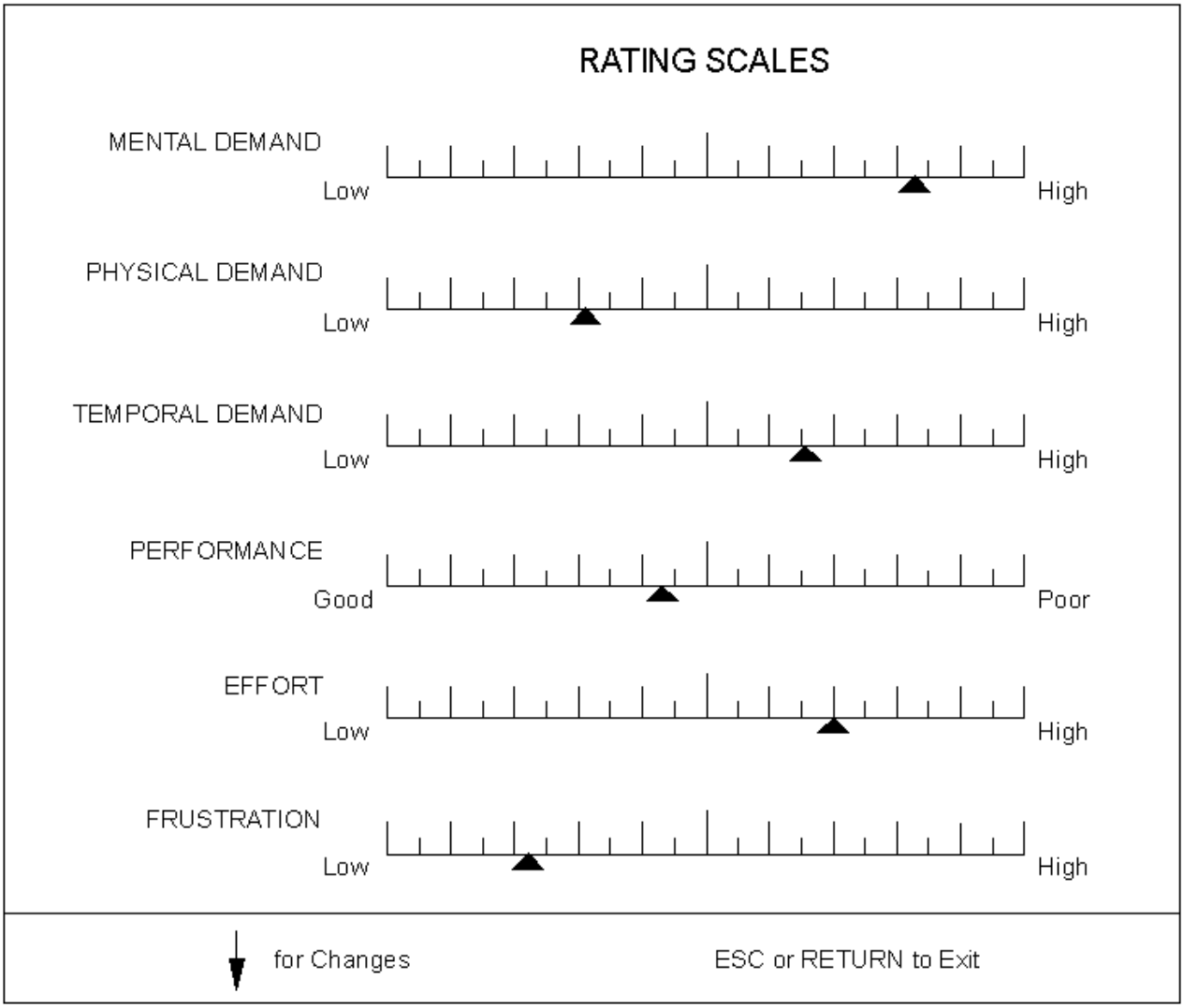


Figure 20. NASA TLX workload rating screen displaying exit instructions.

The NASA TLX is used to obtain a weighted average of ratings on six sub-scales to determine an overall workload rating. Subjects rate their perceived exertion on the Mental Demand, Physical Demand, Temporal Demand, Effort, and Frustration sub-scales using graded scales having endpoints labeled “Low” and “High,” and they provide a rating for the Performance sub-scale using a graded scale having endpoints labeled “Good” and “Poor.” Each gradation on the rating scales represents five points. Since any point along the horizontal line of each rating scale may be selected, scores on each sub-scale can range from 0 to 100.

When the NASA TLX screen depicted in Figure 19 is initially presented, a pointer appears in the middle of the first sub-scale. The subject begins with the Mental Demand sub-scale and uses either a computer keyboard’s left and right arrow keys or a computer mouse’s left to right movement to select a rating value. After providing a rating for the Mental Demand sub-scale, the subject presses either the keyboard’s down arrow key or the mouse’s left button to proceed to the Physical Demand sub-scale. Once the subject presses the keyboard’s down arrow key or the mouse’s left button to progress to the next scale, the pointer of the second scale becomes active (i.e., is illuminated in yellow and become slightly larger), and the pointer of the first scale becomes inactive (i.e., turns gray and become slightly smaller). The keyboard’s left, right, and down arrow keys or the mouse’s left to right movement and left button are used to provide ratings for all six sub-scales.

After a response is provided for the Frustration sub-scale, the command options displayed at the bottom of the screen change as depicted in Figure 20. These commands indicate that the subject may either exit the NASA TLX workload rating screen by pressing the keyboard’s ESCape or Return key, or use the keyboard’s down arrow key to begin altering any of the

responses provided on the ratings scales. Alternately, the subject may exit the NASA TLX workload rating screen by pressing the mouse's right button, or use the mouse's left button to begin altering any of the responses provided on the ratings scales. If, for example, the subject wants to change the response provided on the Effort sub-scale, the keyboard's down arrow key or the mouse's left button would be pressed five times. Five depressions of the keyboard's down arrow key or the mouse's left button would allow the subject to return to the fifth sub-scale without altering any of the scores associated with the first four sub-scales. When the pointer of the fifth rating scale becomes activated, changes can be made by using the keyboard's left and right arrow keys or the mouse's left to right movement. After the desired changes are made, the keyboard's ESCape or Return key or the mouse's right button can used to exit the workload rating screen.

Once the NASA TLX workload ratings are obtained, Comstock and Arnegard's (1992) AFTERMAT program may be used to obtain the sub-scale weightings needed to calculate a subject's mean weighted workload score. The AFTERMAT program individually presents each of the 15 pairs of scale titles (e.g., Mental Demand vs. Physical Demand) on the computer monitor and requires that the subject make factor comparisons of each pair by either: 1) using the keyboard's up and down arrow keys to select the variable that he or she felt was more important to the experience of workload, or 2) verbally reporting the variable that he or she felt was more important to the experience of workload to the experimenter. After a subject's sub-scale ratings and weights are collected, the experimenter is able to calculate the mean weighted workload score associated with the subject's performance of a task by following the procedure described below.

1. The rating that a subject provided for each sub-scale is multiplied by its respective weight. For example, the rating that a subject provided for the Mental Demand sub-scale using the NASA TLX rating screen is multiplied by the weight (i.e., the total number of times that the subject chose the Mental Demand sub-scale as the largest contributor to workload) provided for the Mental Demand sub-scale using the AFTERMAT program.
2. The sum of the products of the sub-scales' ratings and weights is calculated.
3. The sum of the products of the sub-scales' ratings and weights is divided by the total number of weights (i.e., 15), and this quotient represents the subject's mean weighted workload score for a particular task.

An example calculation of the mean weighted workload score associated with a task is presented in Table 8.

Table 8. Example Calculation of the NASA TLX Mean Weighted Workload Score Associated with a Task

Sub-Scale	Rating		Weight		Product
Mental Demand	40	x	3	=	120
Physical Demand	74	x	1	=	74
Temporal Demand	24	x	2	=	48
Performance	76	x	3	=	228
Effort	31	x	4	=	124
Frustration	74	x	2	=	148
			Sum	=	742
			Weights (Total)	=	15
			Mean Weighted Workload Score	=	49.47

When two separate mean weighted workload scores are calculated (i.e., one score is calculated for the manual tracking task condition, and one score is calculated for the automated tracking task condition), these data may be used to determine the extent to which different workload levels were perceived to be experienced during the two conditions of the MAT Battery's tracking task. It is suggested that subjective assessments of workload obtained using the NASA TLX may be evaluated in conjunction with information provided by EEG data to gain a more comprehensive understanding of the levels of workload and attentional engagement associated with the MAT Battery's manual and automated tracking task conditions.

Intent of Current Research

The goal of the research endeavor discussed in the remainder of this document was to advance the development of an aural alerting system based on a categorization scheme in which a distinctive aural alert was associated with each of the four major flight deck functions. The acoustic parameters of a given alert were manipulated in a way that preserved the overall pattern of the signal (thereby preserving the sound's distinctiveness) while conveying low, moderate, and high levels of urgency. This goal was achieved by pursuing recommendations for additional research identified through the work of Burt (1996). Specifically, the current research endeavor investigated: 1) if systematic decreases in the tempos of Burt's (1996) moderate urgency level alerts and low urgency level alerts represented appropriate manipulations through which to differentiate the urgency ratings of the moderate urgency level alerts and the high urgency level alerts while maintaining differences between the urgency ratings provided for the low urgency level alerts and the moderate urgency level alerts; and 2) if individuals were able to identify the

correct alerting set and urgency level associated with each aural alert 95% of the time while performing a tracking task that required two levels of workload and attentional engagement.

To investigate the improvement of the aural alerts developed by Burt (1996), data were collected during two experiments. The first experiment employed a magnitude estimation task to assess the aural alerting signal urgency ratings provided by a population having "normal" hearing threshold levels. Magnitude estimation urgency ratings were obtained to determine if systematic decreases of Burt's moderate urgency level alerts' tempos and systematic decreases of Burt's low urgency level alerts' tempos resulted in the perception of low, moderate, and high urgency levels within each alerting set. The second experiment employed a sound identification task as well as the automated and manual conditions of the MAT Battery's tracking task to assess the ability of individuals having "normal" hearing threshold levels to associate each alerting set with one of the four major flight deck functions and then identify a given alerting set and urgency level 95% of the time while performing a task requiring two levels of workload and attentional engagement.

In general, the hypotheses examined during this research endeavor were as follows:

- It was hypothesized that systematic manipulations of a single aural alert's tempo could be used to convey low, moderate, and high levels of urgency.
- It was hypothesized that a composite manipulation of aural alerts' fundamental frequency, pitch range, rhythmic pattern, and pitch contour could be used to minimize the overall urgency level differences between distinctive alerting sets.
- It was hypothesized that a composite manipulation of aural alerts' fundamental frequency, pitch range, rhythmic pattern, pitch contour, and tempo could be used to create signals within

distinctive alerting sets that conveyed equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency.

- It was hypothesized that subjects would be able to associate each alerting set with one of the four major flight deck functions; identify the correct aural alerting set, the correct urgency level, as well as the correct alerting set and urgency level 95% of the time; and correctly identify each alerting set, urgency level, and aural alert equally often while performing the MAT Battery's tracking task in automatic and manual modes. Interactions between alerting sets, urgency levels, and/or tracking conditions were not expected to occur among sound identifications, but significant interactions would have been interpreted in terms of individual variables' effects had such interactions been found to exist.
- It was hypothesized that subjects would experience higher levels of workload and attentional engagement during the manual tracking task condition than during the automated tracking task condition.

More explicit statements of the hypotheses are presented in the "Method" sections.

EXPERIMENT 1

Purpose

The purpose of Experiment 1 was to determine if systematic decreases in the tempos of Burt's (1996) moderate urgency level alerts and low urgency level alerts represented appropriate manipulations through which to differentiate the urgency ratings of the moderate urgency level alerts and the high urgency level alerts while maintaining differences between the urgency ratings provided for the low urgency level alerts and the moderate urgency level alerts. During this

experiment, magnitude estimation tasks were used to collect estimates of line length as well as ratings of aural alerting signal urgency level from a population having "normal" hearing threshold levels. Line length estimates were used to assess the ability of subjects to assign numerical values to target stimuli, and urgency ratings were used to determine if systematic manipulations of aural alerts' tempos could be used to convey low, moderate, and high urgency levels within each of four equally urgent aural alerting sets. Based on the results of analyses performed on the urgency rating data collected in Experiment 1, a decision regarding the aural stimuli investigated in Experiment 2 was made.

Method

Subjects. Subjects consisted of 14 male and six female volunteers from the NASA LaRC civil servant population. All subjects had auditory thresholds associated with "normal" hearing; that is, each subject had hearing threshold levels in each ear that were $\leq$ 25 dB at 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz (Davis and Silverman, 1978 as cited in Miller and Wilber, 1991). All subjects were at least 18 years old and were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (APA, 1992). The proposal of recording magnitude estimations of lines' lengths and aural alerting signals' urgency levels was approved by the Virginia Polytechnic Institute and State University (V. P. I. & S. U.) Institutional Review Board (IRB) for Research Involving Human Subjects and by the NASA LaRC IRB prior to the collection of any data.

Test facilities. The primary test facility was housed in the Crew Hazards and Error Management (CHEM) Laboratory (Building 1268A) within the Crew/Vehicle Integration Branch

of the Flight Dynamics and Control Division at NASA LaRC, Hampton, Virginia. Audiometric testing took place at the NASA LaRC Medical Center (Building 1149).

Test apparatus. Ambient sound pressure levels within the CHEM Laboratory experimental chamber were measured prior to data collection using a 1.25 cm Bruel & Kjaer condenser microphone (Model No. 4133) that was attached to a Bruel & Kjaer preamplifier (Model No. 2619) and a Bruel & Kjaer microphone power supply (Model No. 2801), a Stanford Research Systems FFT spectrum analyzer (Model No. SR760), and a Philips analog storage oscilloscope (Model No. PM3266). A Bruel & Kjaer pistonphone (Model No. 4220) was used to calibrate the octave band spectrum analyzer prior to acoustic measurements. The experimental chamber's reverberation times were measured using a Hewlett Packard synthesizer/function generator (Model No. 3325A), a Realistic omnidirectional dynamic microphone (Model No. 33-985F), an Acoustic Research powered partner loudspeaker (Model No. PP-570), a Stanford Research Systems FFT spectrum analyzer (Model No. SR760), and a Philips analog storage oscilloscope (Model No. PM3266). A Coulbourn Instruments adjustable gain amplifier (S79-02) and contour following integrator (S76-01) were also used during the reverberation time measurements to adjust the gain as well as integrate the microphone signal.

Pure tone audiogram testing was administered by NASA LaRC Medical Center personnel, all of whom are certified through the Council for Accreditation in Occupational Hearing Conservation, using a TRACOR Microprocessor Audiometer (Model No. RA 400). Subjects wore Telephonics Corp. earphones (Model No. PN510C017-1). The audiometer system was used in an automatic mode which presents pure tones in 5 dB increments over a possible range of -10 dB to +90 dB at 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz. The

audiometer used a total of three correct responses at a given hearing level as its criterion for determining threshold. The printer in the audiometer system plotted dB hearing threshold level as a function of test frequency, and hearing threshold levels in each ear that were $\leq$ 25 dB at 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz were considered to be "normal" as suggested by Davis and Silverman (1978 as cited in Miller and Wilber, 1991).

A paper-and-pencil questionnaire was used to obtain subject information prior to the experimental session. Specifically, information regarding the subjects' musical experience, recent noise exposure, and use of medications that may affect sound perception were obtained. This questionnaire is included in Appendix A. Subjects also estimated the lengths of five lines in order to practice free-modulus magnitude estimation prior to the collection of experimental data. These lines are shown in Appendix B.

The aural alerting signals were constructed on a 586 PC microcomputer using Cool Edit 95 software (Syntrillium Software Corporation, 1995) and an Acer Magic FX-3D soundcard. Each signal was saved in a 16-bit mono 22 kHz pulse code modulation (PCM) format. Although the alerting signals were constructed digitally, the alerts and the task instructions were recorded onto analog cassettes (i.e., one cassette was made for each subject) and were presented using a Tascam stereo cassette tape deck (Model No. 103) and a center, front Acoustic Research powered partner loudspeaker (Model No. PP-570). The flight deck background noise against which all of the stimuli were heard and subjectively rated was recorded using a Marantz digital-to-analog stereo cassette tape deck (Model No. PMD430) and was continuously looped using an Ensoniq advanced sampling recorder (Model No. ASR-10) and Alchemy software (Passport Designs, Inc., 1989). The background noise was recorded onto analog tape with an Onkyo stereo

double cassette tape deck (Model No. TA-RW99) and was presented using an Onkyo Integra stereo cassette tape deck (Model No. TA-2058) and left and right side Acoustic Research powered partner loudspeakers (Model No. PP-570).

An intercom system consisting of microphones and loudspeakers in the test chamber and in the control room allowed the experimenter and subject to communicate with each other. The subject used a Realistic microphone (Model No. PZM 33-1090B) to speak to the experimenter; the experimenter heard the subject's voice through a lightweight headset; the experimenter used the headset and a Realistic stereo mixing console (Model No. 32-1200B) to speak to the subject; and the subject heard the experimenter's voice through the center, front loudspeaker. The experimenter visually monitored the subject throughout the experimental session using a Javelin Electronics Inc. television camera (Model No. JE2071IRGNA) with a 6.12 mm focal length (i.e., wide angle) lens (Model No. C60607 H612A) and a JVC 35.5 cm (diagonal) color video monitor (Model No. TM 1400SU).

Test chambers. The NASA LaRC Medical Center used two free-standing Industrial Acoustics Company, Inc. audiometric test chambers. One of the test chambers (Model No. 1201-A-W/FV) had 60 cm walls with an 11.25 cm air space, and the other test chamber (Model No. 401-A-SE) had 35 cm walls. Ambient noise levels inside each audiometric test chamber are included in Appendix C.

All aural alerting signal testing was conducted in a sound attenuated chamber located in the CHEM Laboratory. The inside dimensions of the chamber were 2.25 m x 1.5 m x 2.4 m, and the interior walls of the chamber were treated with 11.88 m^2 of 2.5 cm thick sound absorbing panels having a noise reduction coefficient of 0.80 in order to model the acoustical characteristics

of an aircraft flight deck. As shown in Figure 21, the chamber was furnished with a table and a chair that served as the subject's workstation, and three loudspeakers were mounted on the chamber's interior walls.

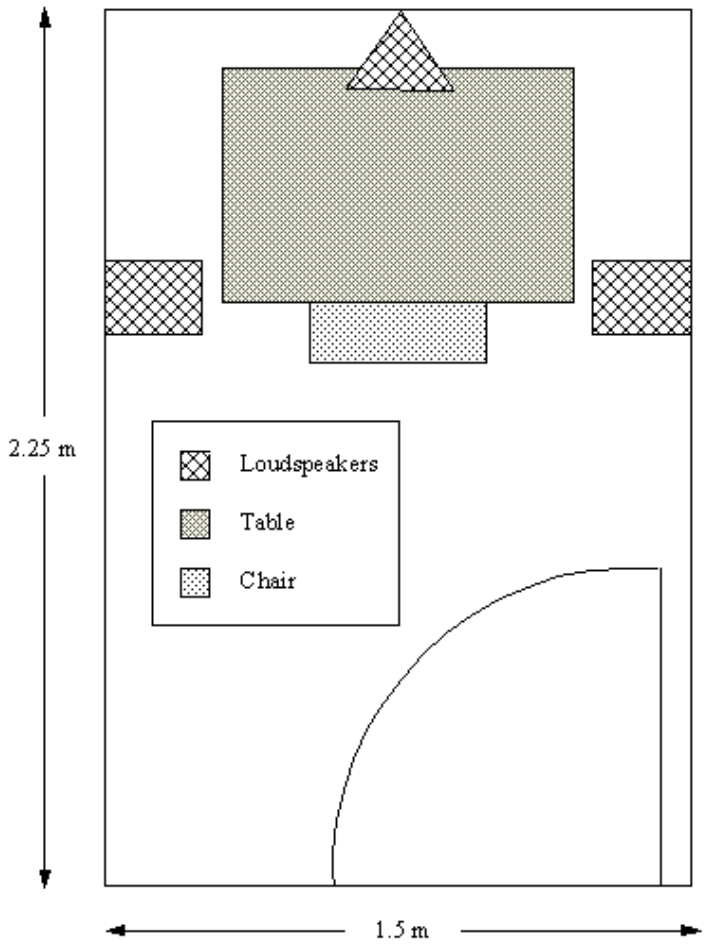


Figure 21. Physical layout of the experimental chamber.

One loudspeaker was mounted on the wall to the left of the workstation; one loudspeaker was mounted on the wall to the right of the workstation; and one loudspeaker was mounted on the wall in front of the workstation. The loudspeakers on either side of the workstation were mounted 1.83 m above the floor, 1.13 m apart from one another, and 95 cm from the approximate head position of the subjects. The loudspeaker in front of the workstation was mounted 1.05 m above the floor, 1.08 m from each of the other two loudspeakers, and 50 cm from the approximate head position of the subjects. The axes of all three loudspeakers were aimed directly toward the approximate head position of the subjects.

The ambient sound pressure levels within the experimental chamber and the chamber reverberation times were obtained prior to the collection of experimental data through acoustical measurements. To measure the chamber's ambient sound pressure levels, a pistonphone was used to generate a 124 dB calibration level that registered on an octave band spectrum analyzer as 0 dBV at 250 Hz with an A-weighting. After establishing the calibration reference, the pistonphone was removed from the test microphone, and ambient sound pressure levels were measured. Specifically, 10 averages of 15 one-third octave bands were taken in the root mean square (RMS) linear mode with a 50% overlap, and a 2.3V peak-to-peak value was displayed on an analog storage oscilloscope. The results of the experimental chamber's ambient sound pressure level measurements are included in Appendix D and were deemed to be appropriate for the purpose of the current investigation.

In order to measure the experimental chamber's reverberation times, one-third octave band warble tones having a 0.05 Hz period were presented at 105 dBA, and the time that it took a given warble tone to decrease to a sound pressure level 60 dBA below its original level, or 45

dBA, was measured. An analog storage oscilloscope displayed a 120 mV peak-to-peak value as well as an integrated value of the sound oscillations (i.e., 20 ms), and five trials were performed at each center frequency. The results of the experimental chamber's reverberation time measurements as well as the reverberation times found on the flight deck of a modern commercial jet aircraft are included in Appendix E. Due to a request for confidentiality, the identity of the aircraft manufacturer that provided the flight deck reverberation times included in Appendix E cannot be revealed.

To further model the acoustical characteristics of an aircraft flight deck, the background noise present on a flight deck during the cruise phase of flight when no conversation or aural alerts were occurring was also presented in the experimental chamber. Specifically, a 10 sec analog recording of the background noise present on the flight deck of NASA 515, LaRC's recently retired Boeing 737 research aircraft, was digitized, continuously looped, recorded onto analog tape, and then presented as a constant background noise against which all aural stimuli were heard and rated by the subjects. The specific sound pressure levels of the flight deck background noise are included in Appendix F, and a spectral plot of the background noise is included in Appendix G. In Appendices F and G, - 84.00 dBV referenced to 1.00V RMS serves as the 0 dBA point.

Test system calibration. The NASA LaRC Medical Center's audiometer used to assess each subject's hearing threshold levels prior to the collection of experimental data was last calibrated using an artificial ear with an earphone coupler in August 1998, approximately 9 months prior to data collection for Experiment 1. The audiometer system also underwent daily biological calibrations with an octave monitor. Within the CHEM Laboratory's test chamber

where experimental data were collected, the sound pressure levels at the subject's ear height were calibrated prior to each experimental session using a Simpson sound level meter (Model No. 866).

Experimental design. The experimental design used for data collection was a 4 (Alerting Set) x 5 (Urgency Level), completely crossed, full factorial, within-subject design. The same 20 subjects were assigned to each experimental cell. The experimental design matrix is shown in Figure 22.

	High	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}
	Moderate #1	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}
URGENCY LEVEL (Within-Subject)	Moderate #2	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}
	Low #1	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}
	Low #2	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}	S_1 - S_{20}
		I	II	III	IV

ALERTING SET
(Within-Subject)

Figure 22. Experimental design matrix used in Experiment 1.

All independent variables were treated as fixed-effects variables, and subjects were treated as a random-effect variable.

Independent variables. As shown in Figure 22, the two factors used in the experimental design were aural alerting set and urgency level. The stimuli were 20 aural alerting signals, each of which belonged to one of four alerting sets and consisted of sound pulses and interpulse intervals having various durations. Most sound pulses included a linear onset time of 20 ms and a linear offset time of 20 ms, and sound pulses less than 40 ms in length had linear onset and offset times that peaked at the middle of the pulse. The first harmonic, or fundamental frequency, of each sound pulse was present at 100%; and simultaneously, the second through fifth harmonics were present at 50% of the fundamental frequency's amplitude.

Specific acoustic parameters of the 20 aural stimuli are included in Tables 9 - 13. As described above, each aural alerting signal was comprised of complex tones made up of a controlled set of harmonics; however, harmonic content is not described in these tables. Since the graphical representations of the alerts within each set are identical to those presented in Figures 4 - 15, with the exception of the low urgency level alerts' and the moderate urgency level alerts' durations, additional figures are not provided here. When examining Tables 9 - 13, note that each alert consisted of a given rhythmic pattern played twice; therefore, the total duration of each alert represents the sum of its sound pulse and interpulse interval durations multiplied by a factor of 2.

Table 9. Acoustic Parameters of Aural Stimuli Used in Experiment 1 - High Urgency Level Alerts

Alerting Set	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
I	523	100%	31	31	1122
	523	50%	31	31	
	563	50%	31	31	
	563	75%	31	94	
	583	50%	31	31	
	583	100% to 50% logarithmic fade	125	63	
II	593	100%	47	113	1164
	553	50%	23	7	
	593	50%	47	63	
	553	100% to 50% logarithmic fade	188	94	
III	675	100%	35	35	1686
	655	50%	35	35	
	635	50%	35	106	
	635	75%	35	35	
	615	50%	35	35	
	598	50%	35	106	
	598	100% to 50% logarithmic fade	70	211	
IV	635	100%	35	35	1126
	635	50%	35	106	
	523	50%	35	35	
	523	100% to 50% logarithmic fade	141	141	

Table 10. Acoustic Parameters of Aural Stimuli Used in Experiment 1 - Moderate #1 Urgency Level Alerts

Alerting Set	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
I	523	100%	65	65	2340
	523	50%	65	65	
	563	50%	65	65	
	563	75%	65	195	
	583	50%	65	65	
	583	100% to 50% logarithmic fade	260	130	
II	593	100%	98	235	2418
	553	50%	49	14	
	593	50%	98	130	
	553	100% to 50% logarithmic fade	390	195	
III	675	100%	73	73	3512
	655	50%	73	73	
	635	50%	73	220	
	635	75%	73	73	
	615	50%	73	73	
	598	50%	73	220	
	598	100% to 50% logarithmic fade	147	439	
IV	635	100%	73	73	2342
	635	50%	73	220	
	523	50%	73	73	
	523	100% to 50% logarithmic fade	293	293	

Table 11. Acoustic Parameters of Aural Stimuli Used in Experiment 1 - Moderate #2 Urgency Level Alerts

Alerting Set	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
I	523	100%	80	80	2880
	523	50%	80	80	
	563	50%	80	80	
	563	75%	80	240	
	583	50%	80	80	
	583	100% to 50% logarithmic fade	320	160	
II	593	100%	120	290	2974
	553	50%	60	17	
	593	50%	120	160	
	553	100% to 50% logarithmic fade	480	240	
III	675	100%	90	90	4320
	655	50%	90	90	
	635	50%	90	270	
	635	75%	90	90	
	615	50%	90	90	
	598	50%	90	270	
	598	100% to 50% logarithmic fade	180	540	
IV	635	100%	90	90	2880
	635	50%	90	270	
	523	50%	90	90	
	523	100% to 50% logarithmic fade	360	360	

Table 12. Acoustic Parameters of Aural Stimuli Used in Experiment 1 - Low #1 Urgency Level Alerts

Alerting Set	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
I	523	100%	104	104	3744
	523	50%	104	104	
	563	50%	104	104	
	563	75%	104	312	
	583	50%	104	104	
	583	100% to 50% logarithmic fade	416	208	
II	593	100%	156	377	3866
	553	50%	78	22	
	593	50%	156	208	
	553	100% to 50% logarithmic fade	624	312	
III	675	100%	117	117	5616
	655	50%	117	117	
	635	50%	117	351	
	635	75%	117	117	
	615	50%	117	117	
	598	50%	117	351	
	598	100% to 50% logarithmic fade	234	702	
IV	635	100%	117	117	3744
	635	50%	117	351	
	523	50%	117	117	
	523	100% to 50% logarithmic fade	468	468	

Table 13. Acoustic Parameters of Aural Stimuli Used in Experiment 1 - Low #2 Urgency Level Alerts

Alerting Set	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
I	523	100%	128	128	4608
	523	50%	128	128	
	563	50%	128	128	
	563	75%	128	384	
	583	50%	128	128	
	583	100% to 50% logarithmic fade	512	256	
II	593	100%	192	464	4758
	553	50%	96	27	
	593	50%	192	256	
	553	100% to 50% logarithmic fade	768	384	
III	675	100%	144	144	6912
	655	50%	144	144	
	635	50%	144	432	
	635	75%	144	144	
	615	50%	144	144	
	598	50%	144	432	
	598	100% to 50% logarithmic fade	288	864	
IV	635	100%	144	144	4608
	635	50%	144	432	
	523	50%	144	144	
	523	100% to 50% logarithmic fade	576	576	

As shown in the tables above, the alerting signals comprising Sets I, II, III, and IV differed from one another in rhythmic pattern and pitch contour as well as in fundamental frequency, pitch range, and duration. The high urgency level alerts were identical to those used in the Burt (1996) study. The tempos of Burt's (1996) moderate urgency level alerts and low urgency level alerts were systematically manipulated to create the moderate #1 urgency level alerts, moderate #2 urgency level alerts, low #1 urgency level alerts, and low #2 urgency level alerts described in Tables 10 - 13.

As stated previously, Burt (1996) attempted to create four sets of aural alerts in which moderate urgency level alerts were perceived to be twice as urgent as low urgency level alerts, and high urgency level alerts were perceived to be twice as urgent as moderate urgency level alerts. Hellier et al.'s (1993) guideline regarding the alerting signal speed manipulation needed to produce a doubling of an alert's perceived urgency level was used to decrease the sound pulse and interpulse interval durations of Burt's low urgency level alerts by a factor of 1.6 to create moderate urgency level alerts, and the sound pulse and interpulse interval durations of Burt's moderate urgency level alerts were decreased by a factor of 1.6 to create high urgency level alerts. Since magnitude estimation data revealed that subjects did not provide significantly different ratings of urgency for the moderate urgency level alerts and the high urgency level alerts, Burt concluded that additional research should be conducted to determine if further manipulations of the moderate urgency level alerts' tempos represent appropriate means through which to differentiate the urgency ratings of the moderate urgency level alerts and the high urgency level alerts. Burt also concluded that manipulations of the low urgency level alerts' tempos should be

investigated to ensure that differences between the urgency ratings provided for the low urgency level alerts and the moderate urgency level alerts are maintained.

Based on these conclusions, Hellier et al.'s (1993) guidelines regarding manipulations of aural alerting signals' speeds were used to systematically manipulate the perceived urgency level of Burt's (1996) low urgency level alerts and moderate urgency level alerts. To create the moderate #1 urgency level alerts described in Table 10, the sound pulse and interpulse interval durations of Burt's moderate urgency level alerts were increased by a factor of 1.3 (i.e., the total duration of each of Burt's moderate urgency level alerts was increased by a factor of 1.3). The moderate #2 urgency level alerts described in Table 11 were created by increasing the sound pulse and interpulse interval durations of Burt's moderate urgency level alerts by a factor of 1.6; the low #1 urgency level alerts described in Table 12 were created by increasing the sound pulse and interpulse interval durations of Burt's low urgency level alerts by a factor of 1.3; and the low #2 urgency level alerts described in Table 13 were created by increasing the sound pulse and interpulse interval durations of Burt's low urgency level alerts by a factor of 1.6. Tempo manipulations were used to create two groups of moderate urgency level alerts and two groups of low urgency level alerts in an attempt to ensure that low, moderate, and high urgency levels within each alerting set were identified through Experiment 1.

As with the alerts used by Burt (1996), the stimulus parameters and frequency range corresponded with current research findings (Berson et al., 1981; Boucek et al., 1981; Edworthy, 1994b; Edworthy et al., 1991; Hanson et al., 1983; Hellier et al., 1993; Patterson, 1982, 1989) and design standards (ISO, 1986; SAE, 1993). A measurement of the ambient noise levels found in the flight deck of the NASA 515 (i.e., LaRC's recently retired Boeing 737 research aircraft)

during “level-flight” when no conversation or aural alerts were occurring revealed sound pressure levels of approximately 80 dBA. Flight crew members typically wear circumaural headsets which provide 15-20 dBA attenuation of ambient noise; therefore, subjects in this investigation, who did not wear any type of hearing protection device, were presented with simulated flight deck background noise at a level of 60 dBA. The aural alerts were presented at a level of 75 dBA because signal levels of 75 dBA do not cause any aural damage or discomfort, and signal levels of 15 to 16 dB above a particular masking noise are sufficient for situations involving warning sounds (Fidell, 1978 as cited in Sorkin, 1987; ISO, 1986; Wilkins, 1981 as cited in Wilkins and Acton, 1982).

Dependent measure. The dependent measure obtained in this experiment was the magnitude estimation rating of aural alert urgency level. Urgency ratings were obtained using the free modulus magnitude estimation method. Rather than present subjects with an experimenter-defined modulus, free modulus magnitude estimation was used based on the recommendation of Stevens (1971) which suggests that it is better to permit observers to choose their own modulus than it is to designate one for them. Each subject provided two sets of numerical magnitude estimation values for each aural alert. These values were normalized using the methodology described by Engen (1971) and yielded one set of 20 magnitude estimation values for each experimental session.

Hypotheses. Magnitude estimation ratings were collected to investigate the hypothesis that systematic manipulations of a single aural alert’s tempo could be used to convey low, moderate, and high levels of urgency. Stated simply, it was hypothesized that the high urgency level alerts would be perceived as being more urgent than one or both groups of moderate

urgency level alerts; that the high urgency level alerts would be perceived as being more urgent than both groups of low urgency level alerts; and that one or both groups of moderate urgency level alerts would be perceived as being more urgent than one or both groups of low urgency level alerts. A more explicit, yet complicated, statement of this hypothesis may be made however since tempo manipulations were employed to create moderate #2 urgency level alerts that were intended to be perceived as being less urgent than moderate #1 urgency level alerts and low #2 urgency level alerts that were intended to be perceived as being less urgent than low #1 urgency level alerts. Specifically, it was hypothesized that the high urgency level alerts would be perceived as being more urgent than either the moderate #2 urgency level alerts or the moderate #1 urgency level alerts and the moderate #2 urgency level alerts; that the high urgency level alerts would be perceived as being more urgent than the low #1 urgency level alerts and the low #2 urgency level alerts; and that either the moderate #2 urgency level alerts or the moderate #1 urgency level alerts and the moderate #2 urgency level alerts would be perceived as being more urgent than either the low #2 urgency level alerts or the low #1 urgency level alerts and the low #2 urgency level alerts. Said another way, it was expected that subjects would give the low #1 urgency level alerts and the low #2 urgency level alerts lower urgency ratings than the high urgency level alerts; would give either the low #2 urgency level alerts or the low #1 urgency level alerts and the low #2 urgency level alerts lower urgency ratings than either the moderate #1 urgency level alerts or the moderate #1 urgency level alerts and the moderate #2 urgency level alerts; and would give either the moderate #2 urgency level alerts or the moderate #1 urgency level alerts and the moderate #2 urgency level alerts lower urgency ratings than the high urgency level alerts.

Since the current study did not address the issue of whether one flight deck function and its associated aural alerting set should be perceived as being more or less urgent than another flight deck function alerting set, magnitude estimation ratings were also collected to investigate the hypothesis that systematic manipulations of aural alerts' fundamental frequency, pitch range, rhythmic pattern, and pitch contour could be used to minimize the overall urgency level differences between alerting sets. It was hypothesized that Sets I, II, III, and IV would be perceived as being equally urgent. In other words, it was expected that subjects would give the same urgency rating to Sets I, II, III, and IV.

Finally, since it was hypothesized that: 1) systematic manipulations of fundamental frequency, pitch range, rhythmic pattern, and pitch contour could be used to equate the overall urgency levels of the alerting sets, and 2) systematic manipulations of tempo could be used to convey low, moderate, and high levels of urgency, magnitude estimation ratings were also collected to investigate a third hypothesis – the four aural alerting sets would convey equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency. It was hypothesized that Set I Low #1, Set II Low #1, Set III Low #1, and Set IV Low #1 would be perceived as being equally urgent; that Set I Low #2, Set II Low #2, Set III Low #2, and Set IV Low #2 would be perceived as being equally urgent; that Set I Moderate #1, Set II Moderate #1, Set III Moderate #1, and Set IV Moderate #1 would be perceived as being equally urgent; that Set I Moderate #2, Set II Moderate #2, Set III Moderate #2, and Set IV Moderate #2 would be perceived as being equally urgent; and that Set I High, Set II High, Set III High, and Set IV High would be perceived as being equally urgent. Said another way, it was expected that subjects would give the same urgency rating to Set I Low #1, Set II Low #1, Set III Low #1, and

Set IV Low #1; would give the same urgency rating to Set I Low #2, Set II Low #2, Set III Low #2, and Set IV Low #2; would give the same urgency rating to Set I Moderate #1, Set II Moderate #1, Set III Moderate #1, and Set IV Moderate #1; would give the same urgency rating to Set I Moderate #2, Set II Moderate #2, Set III Moderate #2, and Set IV Moderate #2; and would give the same urgency rating to Set I High, Set II High, Set III High, and Set IV High.

Procedure. In this experiment, each subject individually participated in a pre-experimental session and one data collection session during a single visit to the CHEM Laboratory. Before arriving at the CHEM Laboratory, all subjects confirmed that their hearing threshold levels had been assessed at the NASA LaRC Medical Center at some point within the preceding six months. The pre-experimental session involved gathering preliminary information from the subjects. During the data collection session, subjects practiced performing free modulus magnitude estimation by providing numerical estimates of line length and then provided numerical estimates of aural alerting signal urgency level. The protocol and duration for the entire experiment, including audiometric testing at the NASA LaRC Medical Center, is shown in Table 14.

Table 14. Protocol and Duration for Experiment 1

Session		Duration
1. Audiometric testing		30 min
2. Pre-experimental session		15 min
3. Magnitude estimation data collection session		15 min
4. Subject debriefing		15 min
	Total:	1 hr 15 min

As shown in Table 14, the total duration of this experiment was approximately 1 hr and 15 min.

During the experimental session, the same five lines were presented to each subject in the same order. Line lengths within a range of 2.5 to16.5 cm and the order in which the lines were presented were selected randomly. Each line was presented on a separate sheet of paper; however, all five lines are shown on a single sheet of paper in Appendix B. The same 20 aural stimuli were presented to each subject, and, as shown in Table 15, all aural stimuli were presented in partially counterbalanced order.

Table 15. Balanced Latin Square Ordering of Stimuli Presentation Used in Experiment 1

Subject Number																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	1
20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	1	2
19	20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	1	2	3
18	19	20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	1	2	3	4
17	18	19	20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	1	2	3	4	5
16	17	18	19	20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
7	8	9	10	11	12	13	14	15	16	17	18	19	20	1	2	3	4	5	6
15	16	17	18	19	20	1	2	3	4	5	6	7	8	9	10	11	12	13	14
8	9	10	11	12	13	14	15	16	17	18	19	20	1	2	3	4	5	6	7
14	15	16	17	18	19	20	1	2	3	4	5	6	7	8	9	10	11	12	13
9	10	11	12	13	14	15	16	17	18	19	20	1	2	3	4	5	6	7	8
13	14	15	16	17	18	19	20	1	2	3	4	5	6	7	8	9	10	11	12
10	11	12	13	14	15	16	17	18	19	20	1	2	3	4	5	6	7	8	9
12	13	14	15	16	17	18	19	20	1	2	3	4	5	6	7	8	9	10	11
11	12	13	14	15	16	17	18	19	20	1	2	3	4	5	6	7	8	9	10

NOTE: The sequence of numbers listed vertically under each subject number represents the order in which each sound was presented to each subject. For example, Subject 1 heard Sound 1 (i.e., Set I Low Urgency), then Sound 2 (i.e., Set I Moderate Urgency), then Sound 20 (i.e., Set IV High Urgency), etc.

Prior to the collection of data, subjects signed the Informed Consent Form included in Appendix H, and the experimenter was provided with written documentation from the NASA LaRC Medical Center that subjects had “normal” hearing (i.e., hearing threshold levels in each ear that were $\leq$ 25 dB at 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz). The subjects then completed the questionnaire (Appendix A) regarding their musical experience, recent noise exposure, and types and doses of medications currently being taken which may have affected their perception of sound. After the collection of this preliminary information, subjects were seated in the test chamber. After a subject entered the test chamber, the chamber’s door was closed, and all communication between the subject and the experimenter took place through an intercom system. The protocol for the pre-experimental session is shown in Table 16.

Table 16. Protocol for the Pre-Experimental Session Used in Experiment 1

1. The experimenter greeted the subject.
2. The subject read and signed the Informed Consent Form (Appendix H).
3. The subject provided written documentation of “normal” hearing threshold levels from the NASA LaRC Medical Center.
4. The subject completed the Preliminary Questionnaire (Appendix A).
5. The subject was seated in the test chamber, and subsequent experimenter-subject communications took place through an intercom system.

The experimental session immediately followed the pre-experimental session. The protocol for this session is shown in Table 17.

Table 17. Protocol for Magnitude Estimation Data Collection Session

1. The subject silently read the instructions for the line length magnitude estimation task while the experimenter read the instructions aloud.
2. The subject practiced performing free-modulus magnitude estimation by providing numerical estimates of line length. The same five lines were presented in the same order to each subject.
3. The subject silently read the instructions for the aural alert urgency level magnitude estimation task while the experimenter read the instructions aloud.
4. The subject practiced providing magnitude estimation urgency ratings for aural alerts during four practice trials.
5. The subject provided two magnitude estimation urgency ratings for each of the 20 aural alerts. Alerts were presented in partially counterbalanced order.
6. At the close of the session, the subject was debriefed.

At the beginning of the magnitude estimation data collection session, each subject silently read the following instructions, adapted from the work of Engen (1971), while the experimenter read the instructions aloud:

> You will be presented with a series of lines. These lines will be presented to you one at a time, each on a separate sheet of paper. Your task is to estimate the length of each line by assigning the line a number and writing the number in a space like the one shown below.
>
> Line A __(your #)__
>
> When you see the first line, give its length a number - any number greater than or equal to zero that you think is appropriate. Then, write that number in the appropriate space on the response form. You will then turn to the next sheet of paper and view the next line. Do the same thing - give the length of the second line a number, then, write that number in the appropriate space on the response form. You will do the same thing with all of the lines that you view.
>
> When you estimate the lengths of the lines that you will be viewing, try to make the number that you assign to a line proportional to the length of that line. For example, if the length of a line is twice as long as the one before it, give it a number twice as high. Remember that you can assign any number equal to or greater than zero, and there is no limit to the number you assign.
>
> There are no right or wrong answers. I want to know how you judge the lengths of the lines. Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to interact with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

> During this task, please be certain that you view and provide a numerical estimate of length for every line. You may now turn to the next page and view the first line.

The five lines were then presented to each subject in the same order. Subjects recorded their numerical magnitude estimation values on the Magnitude Estimation Task Response Form located in Appendix I. After a numerical estimate of line length was provided for each line, each subject silently read the following instructions, adapted from the work of Engen (1971), while the experimenter read the instructions aloud:

> You will now hear a series of sounds presented against a background of ambient 737 cockpit noise. Your task is to determine how urgent each sound is by assigning the sound a number and writing the number in a space like the one shown below.
>
> Sound A __(your #)__
>
> You should be concerned only with the urgency of the sounds - <u>not</u> with the urgency of the background noise.
>
> When you have heard the first sound, give its urgency a number - any number greater than or equal to zero that you think is appropriate. Then, write that number in the appropriate space on the response form. You will then hear the next sound. Do the same thing - give the urgency of the second sound a number, then, write that number in the appropriate space on the response form. You will do the same thing with all of the sounds that you hear.
>
> For our purposes, "urgency" will be defined as "the quality or state of being important, insistent, or pressing." A sound that is perceived as having a "low" level of urgency gives the impression that awareness is required and that future action may be necessary. A sound that is perceived as having a "moderate" level of urgency gives the impression that some form of action is required. A sound that is perceived as having a "high" level of urgency gives the impression that immediate action is required.
>
> When you rate the urgency of the sounds that you will be hearing, try to make the number that you assign to a sound proportional to the urgency of that sound. For example, if the urgency of a sound is twice as high as the one before it, give it a number twice as high.

Remember that you can assign any number equal to or greater than zero, and there is no limit to the number you assign.

There are no right or wrong answers. I want to know how you judge the urgency of the sounds. Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to speak with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

I will now give you four examples of the sounds that you will be hearing. Each sound will consist of a rhythmic pattern played twice, and the sounds will vary by rhythm, pitch, duration, and tempo. Here are the examples; please use them to familiarize yourself with the sounds and the rating procedure, and be sure to record a numerical rating of urgency for every example.

The experimenter presented Practice Trial A; 3 sec later, Practice Trial B was presented; 3 sec later, Practice Trial C was presented; and 3 sec later, Practice Trial D was presented.

Do you have any final questions?

If the subject asked a question, the experimenter used the intercom system to respond to the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

Although it will take a great deal of very focused attention, please be certain that you listen to and provide a numerical rating of urgency for every sound. I will now present the sounds.

The 20 aural alerts were then presented to each subject twice in partially counterbalanced order. Subjects recorded their numerical magnitude estimation values on the Magnitude

Estimation Task Response Form located in Appendix J. Following this, the data collection session was terminated, and each subject was debriefed.

Results and Discussion

Line length estimation. Subjects practiced the task of free-modulus magnitude estimation by providing five estimates of line length (i.e., each subject provided one estimate of length for each of five lines prior to the collection of experimental data). In order to compare magnitude estimation judgements among subjects who responded using varied number ranges, every score for each subject was corrected. The procedure described by Engen (1971) was used to eliminate inter-observer variance caused by different choices of moduli and to eliminate intra-observer variability. The steps in the procedure used to correct the line length estimation data are outlined below.

1. Convert each response value to its logarithm.
2. Plot these logarithmic responses in a table in which subjects are listed by row and lines are listed by column.
3. Obtain the arithmetic mean of the logarithmic responses in each row. This is equal to the logarithm of the geometric mean of each observer's responses to all the lines.
4. Obtain the arithmetic mean of all the values obtained in step 3. This is equal to the logarithmic value of the grand mean of all the responses for all observers to all the lines in the original data matrix.
5. Subtract the value obtained in step 4, the grand mean log response, from each of the arithmetic individual mean log responses determined in step 3.

6. Add the value obtained in step 5 to the row of values obtained for each observer in step 1.

After each line length estimate was corrected, the distribution of the original line length estimates provided by all 20 subjects was graphically depicted in histogram format, and a normal probability density function with the same mean and standard deviation was superimposed on the histogram to determine if the data were skewed. Then, the distribution of the corrected line length estimates provided by all 20 subjects was graphically depicted in histogram format, and a normal probability density function with the same mean and standard deviation was superimposed on this histogram to determine if the data appear to be less skewed (i.e., more normally distributed). As shown in Figures 23, the distribution of the original line length estimates is positively skewed, while the distribution of the corrected line length estimates appears to be more normally distributed. These results indicate that biases present in subjects' magnitude estimates were eliminated through the use of Engen's (1971) "correction" procedure.

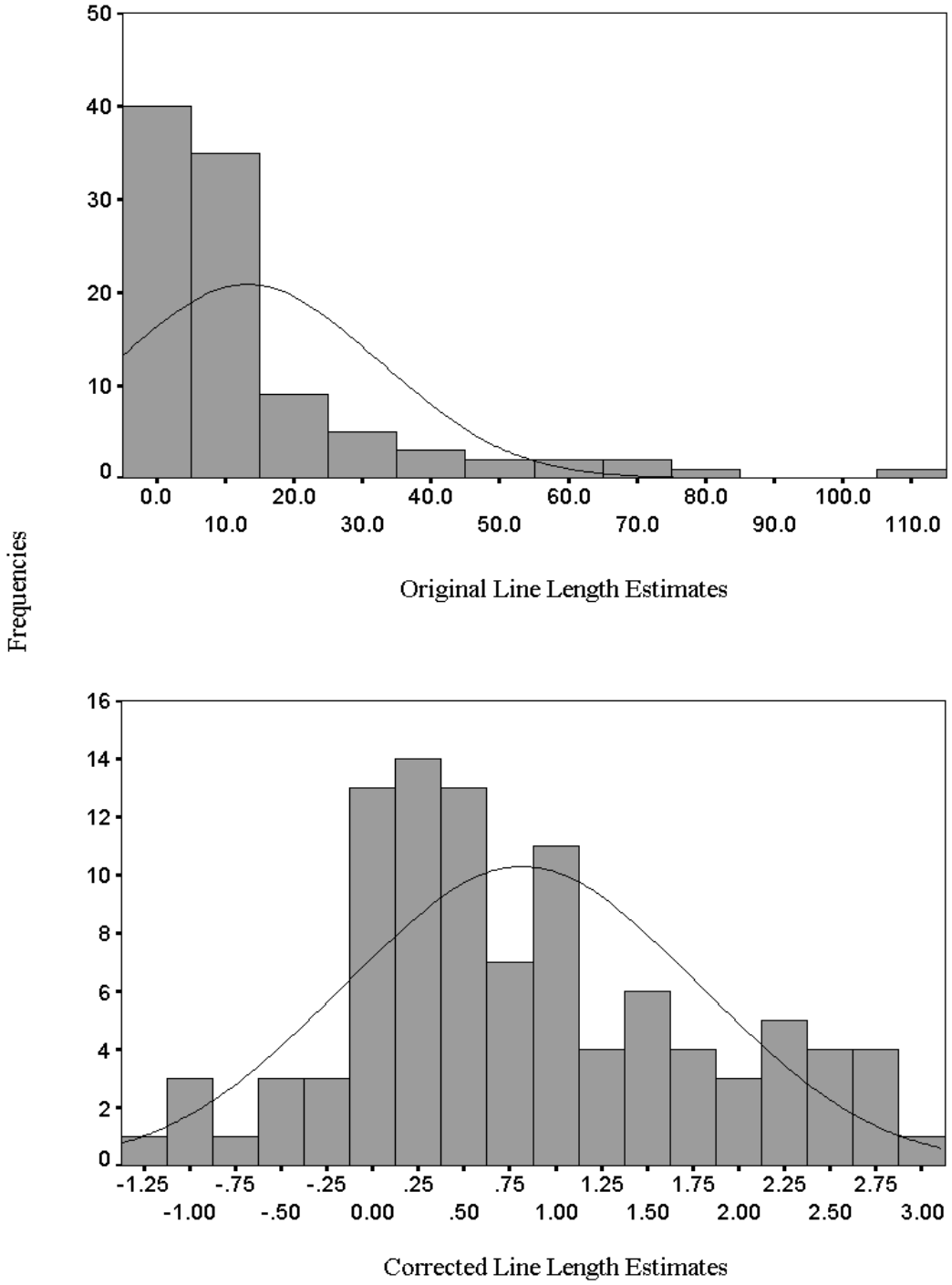


Figure 23. Distributions of original line length estimates and corrected line length estimates in comparison to normal distributions.

A linear regression analysis in which log line length served as the predictor and the log of the arithmetic mean of subjects' corrected line length estimates served as the response was also performed. Since logarithmic values were used in this analysis, the slope of the resulting regression equation is equivalent to the exponent in the power function relating estimated (i.e., perceived) line length to actual line length. According to Gescheider (1985), it is "... widely accepted that the sensory experience of line length is directly proportional to actual length" (p.196). The results of previous experiments (e.g., Zwislocki, 1983 as cited in Gescheider, 1985) have revealed that magnitude estimation of line length is proportional to actual line length and can be described by a power function having an exponent of approximately 1.0. The regression equation yielded by data collected in the current study is $Y' = 0.3225 + 1.0631X$ ($R^2 = 0.994$; R^2-adjusted = 0.992); hence, a power exponent of 1.0631 for magnitude estimation of line length and actual line length was obtained. This exponential value is essentially equal to a power exponent of 1.0; therefore, the current study's corrected line length estimation data indicate that subjects were able to accurately assign numbers to the magnitude of target stimuli's attributes. Figure 24 depicts the line produced by the regression equation superimposed on the graphical representation of subjects' corrected line length estimates.

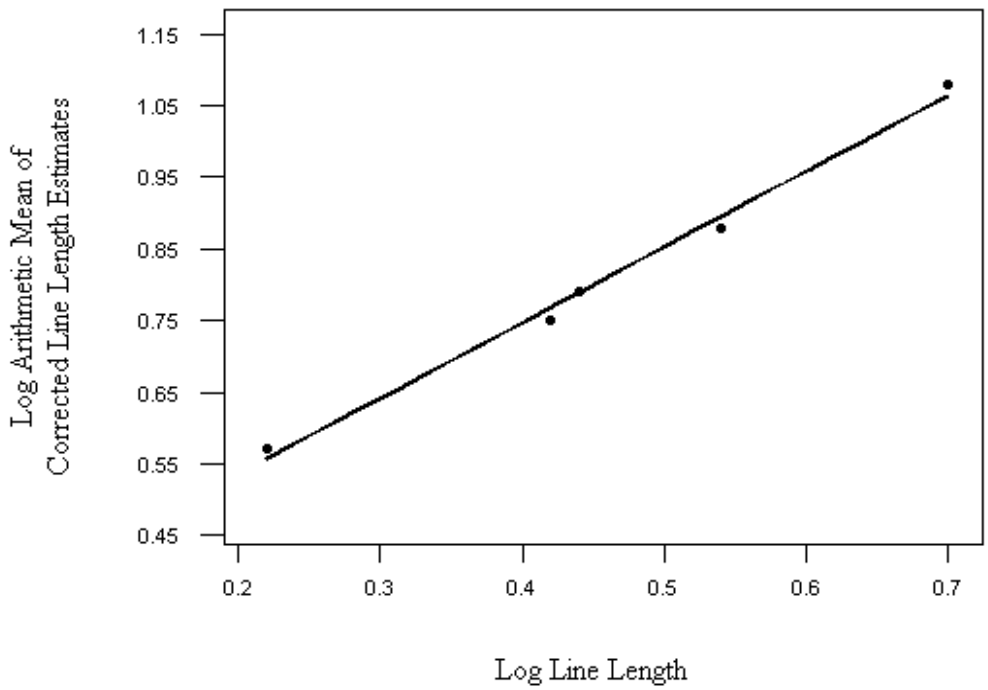


Figure 24. Plot of the power function relating estimated line length to actual line length.

Aural alert urgency level estimation. Each subject provided 40 magnitude estimation ratings of aural alert urgency level; that is, each subject provided two urgency ratings for each of 20 aural alerts. In order to compare magnitude estimation judgements among subjects who responded using varied number ranges, every score for each subject was corrected. The procedure described by Engen (1971) was used to eliminate inter-observer variance caused by different choices of moduli and to eliminate intra-observer variability. The procedure used in conjunction with the aural alert urgency level estimation data is as follows:

1. Convert each response value to its logarithm.
2. Calculate the arithmetic mean of the logarithms of the two responses made by each observer to each sound. This value is equal to the logarithm of the geometric mean of the observer's responses to each stimulus.
3. Plot the means in a table in which subjects are listed by row and sounds are listed by column.
4. Obtain the arithmetic mean of the logarithmic responses in each row. This is equal to the logarithm of the geometric mean of each observer's responses to all the sounds.
5. Obtain the arithmetic mean of all the values obtained in step 4. This is equal to the logarithmic value of the grand mean of all the responses for all observers to all the sounds in the original data matrix.
6. Subtract the value obtained in step 5, the grand mean log response, from each of the arithmetic individual mean log responses determined in step 4.
7. Add the value obtained in step 6 to the row of values obtained for each observer in step 2.

After each urgency rating was corrected, the distribution of the original aural alert urgency level ratings provided by all 20 subjects was graphically depicted in histogram format, and a

normal probability density function with the same mean and standard deviation was superimposed on the histogram to determine if the data were skewed. Then, the distribution of the corrected aural alert urgency level ratings provided by all 20 subjects was graphically depicted in histogram format, and a normal probability density function with the same mean and standard deviation was superimposed on this histogram to determine if the data appeared to be more normally distributed. As shown in Figures 25, the distribution of the original aural alert urgency level ratings is positively skewed, while the distribution of the corrected urgency ratings is less skewed and more normally distributed. These results indicate that biases present in subjects' magnitude estimates were eliminated through the use of Engen's (1971) "correction" procedure. A simple linear regression analysis was not performed on these data since exponents of psychophysical power functions for comparison do not exist.

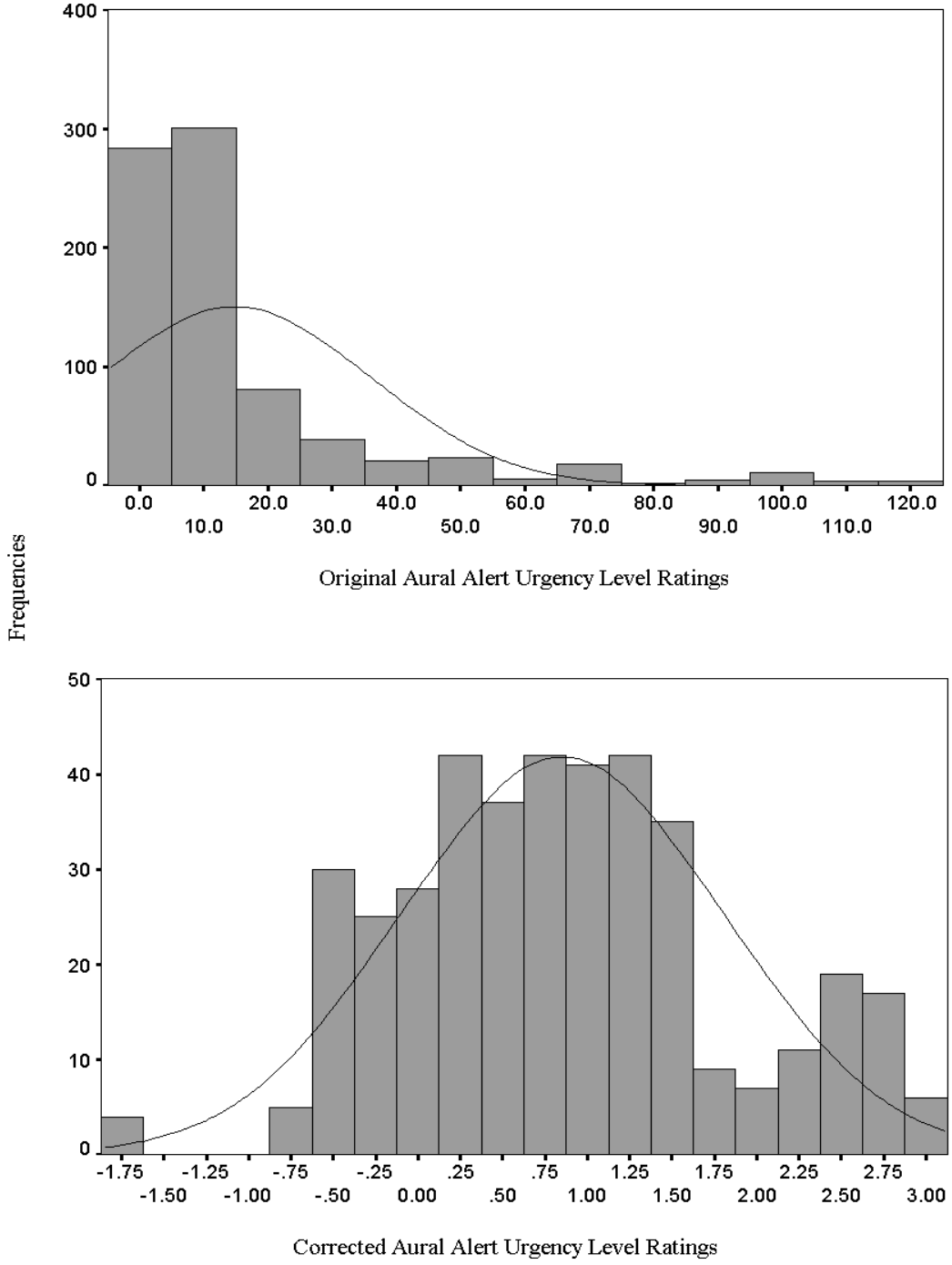


Figure 25. Distributions of original aural alert urgency level ratings and corrected aural alert urgency level ratings in comparison to normal distributions.

The corrected aural alert urgency level ratings were analyzed by way of a 4 (Alerting Set) x 5 (Urgency Level) ANOVA to determine the extent to which subjects perceived low, moderate, and high urgency levels within four equally urgent alerting sets. As shown by the ANOVA Summary Table presented in Table 18, a significant difference existed among subjective ratings of urgency level (F [4, 76] = 75.79; $p < 0.05$). However, no significant differences were found among aural alerting set urgency ratings, and no significant interaction between alerting set and urgency level ratings occurred.

Table 18. ANOVA Summary Table of Magnitude Estimation Data

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	19	329.06				
Within-Subject						
Alerting Set (AS)	3	0.02	0.01	2.25	0.092	
AS X S	57	0.15	0.00			
Urgency Level (UL)	4	25.66	6.41	75.79	0.0001	0.0001
UL X S	76	6.43	0.08			
UL X AS	12	0.03	0.00	1.43	0.154	
UL X AS X S	228	0.40	0.00			
Total	399	361.75				

Although analyses of within-subject designs are performed using the distribution assumptions of normality, homogeneity of within-treatment variances, and independence, the correlations between multiple observations obtained from the same subjects, as well as the associated homogeneity of covariance assumption, must also be considered (Keppel, 1982). Since a violation of the homogeneity of covariance assumption in within-subject design analyses results in the use of "... a more 'lenient' significance level than we had set originally" (Keppel, 1982, p.469), it was necessary to conduct a test of sphericity to determine if positive bias present in the F-Test was responsible for the significant difference found among ratings of urgency level. A Mauchly's test of sphericity was used to determine if "... the covariance matrix of the transformed variables had a constant variance on the diagonal and zeros off the diagonal" (Norusis, 1992, p. 134). This statistical procedure produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [9] = 84.98; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.33741 be used to correct the problems associated with a positively biased F-Test. Therefore, the usual ANOVA was performed, but the observed F ratio that was found to be significant with the uncorrected F-Test was evaluated against a new F table critical value determined by reduced numerator and denominator degrees of freedom. Since a significant difference existed among the ratings of urgency level (F [1, 25] = 75.79; $p < 0.05$) even after the conservative Greenhouse-Geisser correction was applied, the null hypothesis (i.e., all urgency levels received the same urgency rating) was rejected.

To determine which urgency levels were perceived as being significantly different from one another, a Bonferroni t-Test post-hoc analysis was performed. This statistical procedure was

appropriate for evaluating a series of post-hoc comparisons while controlling for inflated alpha error, and it allowed an interpretation of whether or not the low, moderate, and high urgency level alerts were perceived as such by the subjects. As shown in Table 19 (and as depicted in Figure 27 on p.120), the post-hoc analysis revealed that subjects rated the low #2 urgency level alerts as being significantly less urgent than the moderate #2, moderate #1, and high urgency level alerts; that subjects rated the low #1 urgency level alerts as being significantly less urgent than the moderate #1 and high urgency level alerts; that subjects rated the moderate #2 urgency level alerts as being significantly less urgent than the high urgency level alerts; and that subjects rated the moderate #1 urgency level alerts as being significantly less urgent than the high urgency level alerts.

Table 19. Bonferroni t-Test Summary Table of Magnitude Estimation Data

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order				
	Low #2 (0.49)	Low #1 (0.69)	Moderate #2 (0.86)	Moderate #1 (0.97)	High (1.24)
Low #2 (0.49)	-	0.20	0.37*	0.48*	0.75*
Low #1 (0.69)		-	0.17	0.28*	0.55*
Moderate #2 (0.86)			-	0.11	0.38*
Moderate #1 (0.97)				-	0.27*
High (1.24)					-

* $p \leq 0.05$

Mean urgency ratings and 95% confidence intervals (CI) calculated from the corrected magnitude estimation data are displayed in Figures 26, 27, and 28. Figure 26 depicts the mean urgency ratings associated with Sets I, II, III, and IV when ratings were averaged across the five urgency levels (i.e., N = 100). Figure 27 depicts the mean urgency ratings associated with the low #1, low #2, moderate #1, moderate #2, and high urgency levels when ratings were averaged across the four alerting sets (i.e., N = 80). Figure 28 depicts the mean urgency ratings associated with each urgency level within each alerting set (i.e., N = 20).

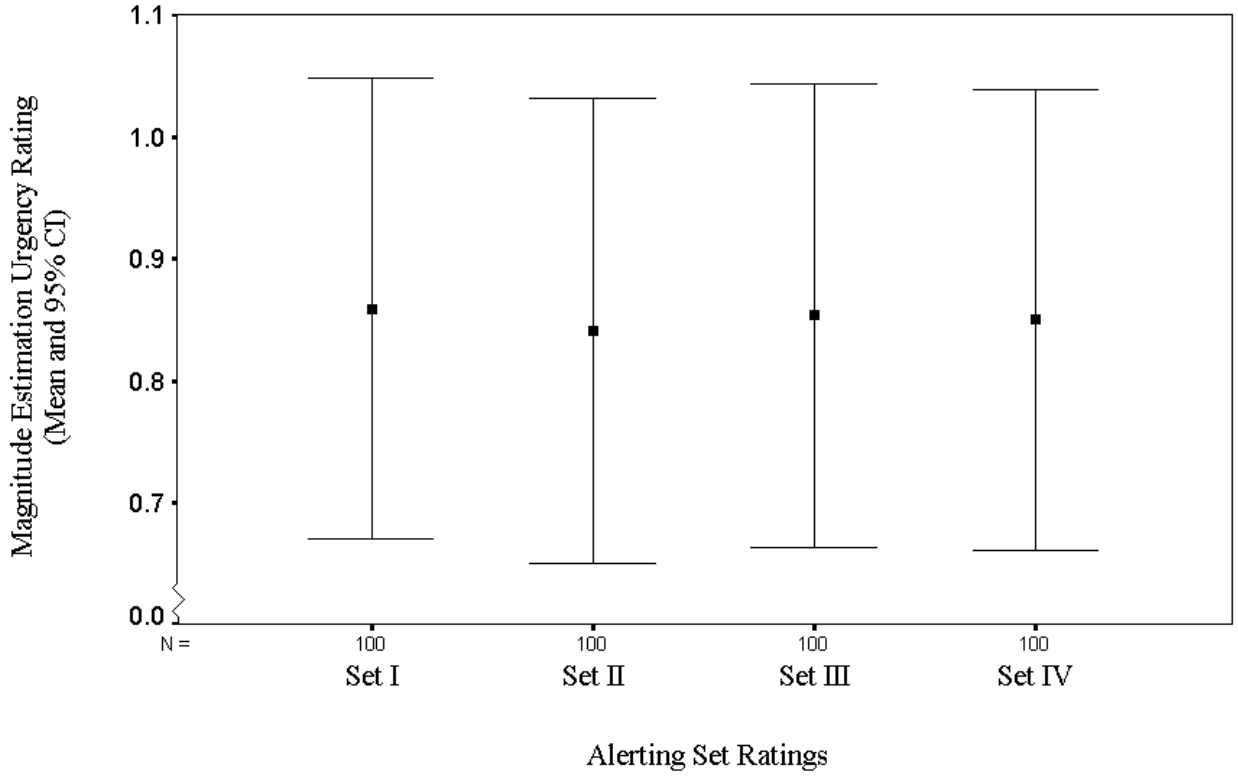


Figure 26. Magnitude estimation urgency ratings of alerting set collapsed across urgency level.

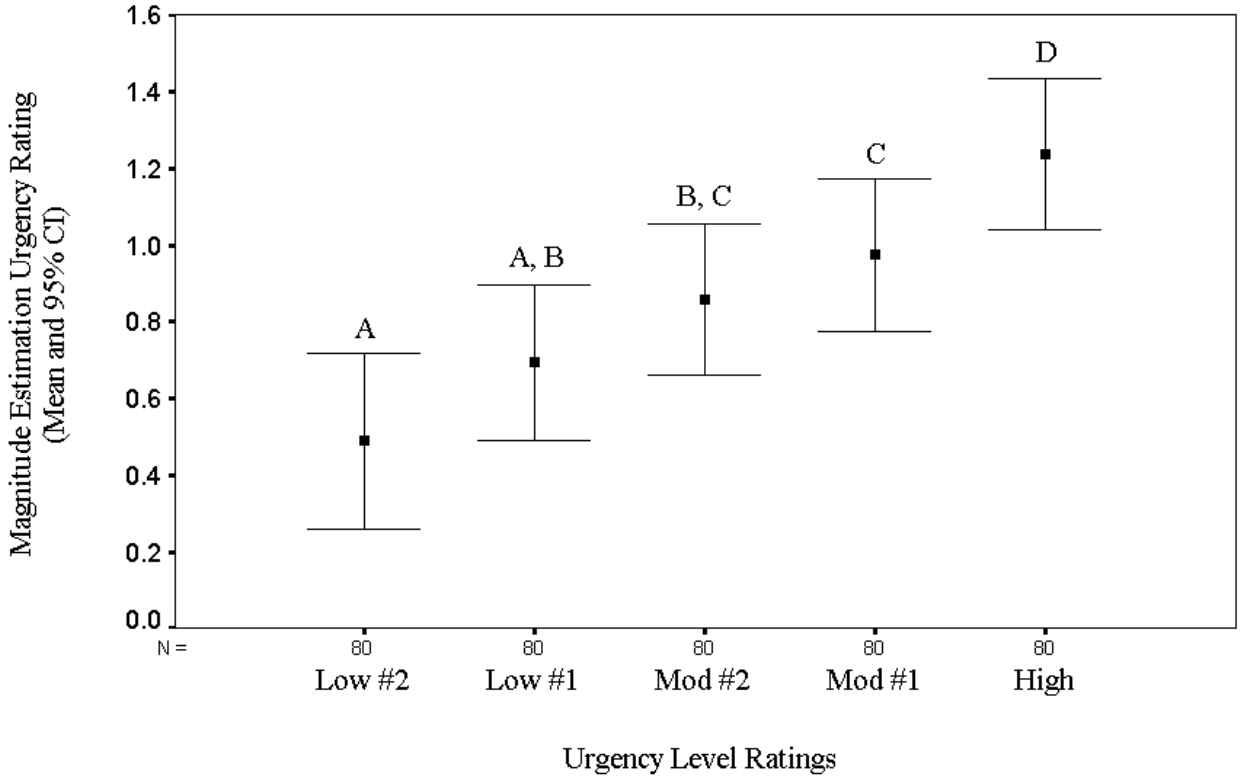


Figure 27. Magnitude estimation urgency ratings of urgency level collapsed across alerting set. (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

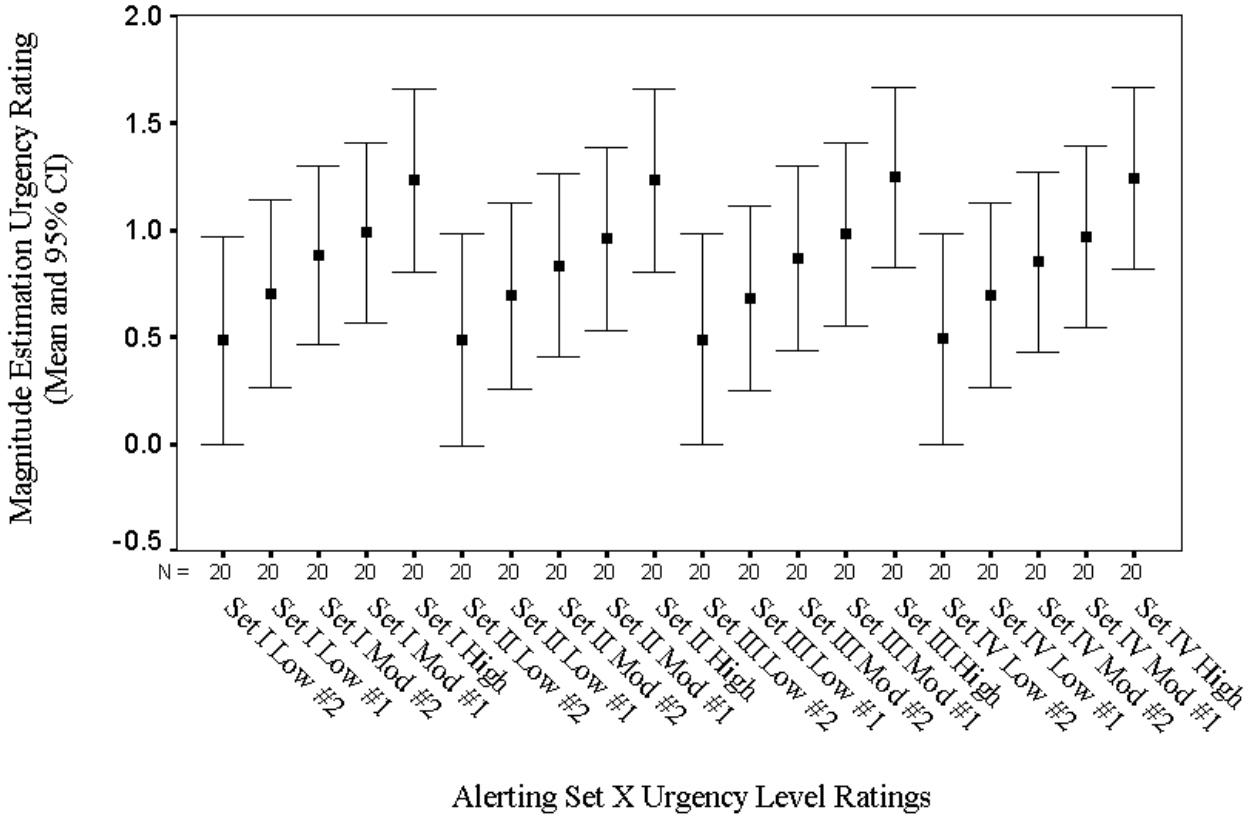


Figure 28. Magnitude estimation urgency ratings of alerting set by urgency level.

The results of the ANOVA and the Bonferroni t-Test support the experimental hypotheses regarding the urgency ratings of the alerting sets and the urgency levels. The failure to reject the null hypothesis that each alerting set would receive the same urgency rating supported the expectation that a composite manipulation of aural alerts' fundamental frequency, pitch range, rhythmic pattern, and pitch contour could be used to minimize the overall urgency level differences between distinctive alerting sets. Sets I, II, III, and IV were perceived as being equally urgent.

The rejection of the null hypothesis that all urgency levels would receive the same urgency rating supported the expectation that systematic manipulations of a single aural alert's tempo could be used to convey low, moderate, and high levels of urgency. The high urgency level alerts were perceived as being more urgent than both groups of moderate urgency level alerts as well as more urgent than both groups of low urgency level alerts; the moderate #1 urgency level alerts were perceived as being more urgent than both groups of low urgency level alerts; and the moderate #2 urgency level alerts were perceived as being more urgent than the low #2 urgency level alerts. Hence, the high urgency level alerts were rated as being significantly more urgent than at least one group of moderate urgency level alerts, and this same group of moderate urgency level alerts was rated as being significantly more urgent than at least one group of low urgency level alerts.

The failure to reject the null hypothesis that each group of low urgency level alerts would receive the same urgency rating, that each moderate urgency level alert would receive the same urgency rating, and that each high urgency level alert would receive the same urgency rating supported the expectation that a composite manipulation of aural alerts' fundamental frequency,

pitch range, rhythmic pattern, pitch contour, and tempo could be used to create distinctive alerting sets capable of conveying equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency. Each low #1 urgency level alert was perceived as being equally urgent; each low #2 urgency level alert was perceived as being equally urgent; each moderate #1 urgency level alert was perceived as being equally urgent; each moderate #2 urgency level alert was perceived as being equally urgent; and each high urgency level alert was perceived as being equally urgent.

Through Experiment 1, progress was made toward the ultimate goal of developing an aural alerting system in which the tempos of four distinctive and equally urgent alerts are manipulated to convey three levels of urgency: low, moderate, and high. However, Experiment 1's high urgency level alerts were perceived as being more urgent than the moderate #1 and the moderate #2 urgency level alerts as well as more urgent than the low #1 and the low #2 urgency level alerts; the moderate #1 urgency level alerts were perceived as being more urgent than the low #1 and the low #2 urgency level alerts; and the moderate #2 urgency level alerts were perceived as being more urgent than the low #2 urgency level alerts. Therefore, a decision had to be made regarding which group of moderate urgency level alerts and which group of low urgency level alerts would be retained for further investigation.

Selection of Stimuli Investigated in Experiment 2

To determine which 12 aural stimuli would be further investigated in a subsequent experiment, the following *a priori* decision rule was considered: The alerts that have the shortest total durations and are perceived as having significantly different urgency levels will be selected

for inclusion in Experiment 2. As stated previously, subjects rated the low #2 urgency level alerts as being significantly less urgent than the moderate #2, moderate #1, and high urgency level alerts; rated the low #1 urgency level alerts as being significantly less urgent than the moderate #1 and high urgency level alerts; rated the moderate #2 urgency level alerts as being significantly less urgent than the high urgency level alerts; and rated the moderate #1 urgency level alerts as being significantly less urgent than the high urgency level alerts. The low #1 urgency level alerts have shorter total durations than the low #2 urgency levels alerts, and the moderate #1 urgency level alerts have shorter total durations than the moderate #2 urgency level alerts. Therefore, Set I Low Urgency #1, Set I Moderate Urgency #1, Set I High Urgency, Set II Low Urgency #1, Set II Moderate Urgency #1, Set II High Urgency, Set III Low Urgency #1, Set III Moderate Urgency #1, Set III High Urgency, Set IV Low Urgency #1, Set IV Moderate Urgency #1, and Set IV High Urgency were investigated in Experiment 2.

EXPERIMENT 2

Purpose

The purpose of Experiment 2 was to determine if individuals were able to identify the correct alerting set and urgency level associated with each of 12 aural alerts 95% of the time while performing a tracking task that required two levels of workload and attentional engagement. During this experiment, a sound identification task was used to assess the ability of individuals having “normal” hearing threshold levels to associate each alerting set with one of the four major flight deck functions and then identify a given alerting set and urgency level while performing the MAT Battery’s tracking task in automatic and manual modes. Subjective assessments of

workload were collected to determine the extent to which different workload levels were perceived to be experienced during the automated and manual conditions of the tracking task, and EEG data were collected to assess the ongoing physiological responses of attention associated with each tracking task condition.

Method

Subjects. Subjects consisted of nine male and three female volunteers from the NASA LaRC civil servant population. Individuals who served as subjects in Experiment 1 were not eligible to participate in Experiment 2. All subjects were right handed; had "normal" (i.e., 20/20 or better) or corrected-to-normal vision; and had no history of neurological problems that could have interfered with the recording of EEG. Auditory threshold and age requirements as well as ethical considerations were identical to those defined in conjunction with Experiment 1. The proposal of manipulating subjects' workload and attentional engagement levels through the performance of a tracking task; eliciting identifications of aural alerting sets and urgency levels as well as subjective assessments of workload; and recording EEG data was approved by the V. P. I. & S. U. IRB for Research Involving Human Subjects and by the NASA LaRC IRB prior to the collection of any data.

Test facilities. The test facilities used in this experiment were identical to those used in Experiment 1. Audiometric testing took place at the NASA LaRC Medical Center (Building 1149), and experimental data were collected in the NASA LaRC CHEM Laboratory (Building 1268A).

Test apparatus. The test apparatus used in this experiment was identical to that used in Experiment 1; however, additional equipment was employed to: 1) allow the use of the MAT Battery's (Comstock and Arnegard, 1992) tracking task and computerized version of the NASA TLX (Hart and Staveland, 1988), and 2) record EEG data. The MAT Battery's tracking task, NASA TLX, and AFTERMAT programs were run using a Pentium class computer with Microsoft's Windows 98 operating system. The tracking task was displayed to the subject on a Gateway Vivitron 35.5 cm (diagonal) color computer monitor (Model No. CPD-15F13) located approximately 61 cm from the subject, and the tracking task was coupled with an Advanced Gravis Computer Tech Ltd. isotonic spring-return joystick (Model No. 3194). The MAT Battery's version of the NASA TLX was displayed to the subject on the same computer monitor, and a Microsoft two-button mouse (Model No. 52695) was used for data entry. The MAT Battery's tracking task and computerized version of the NASA TLX was displayed to the experimenter on a Gateway Vivitron 35.5 cm (diagonal) color computer monitor (Model No. CPD-15F23), and a Gateway ANYKEY keyboard (Model No. 2191011) was used by the experimenter to record each subject's NASA TLX sub-scale weightings.

EEG data were collected with an Electro-Cap International lycra head cap [i.e., a head cap consisting of 22 recessed electrodes arranged according to the standardized "International 10-20" placement system (Jasper, 1958)]; a reference electrode attached to the left earlobe; and a ground electrode attached to the left mastoid prominence. Data were collected at sites Cz (midline central), Pz (parietal central), P3 (left parietal), and P4 (right parietal) via four of the head cap's 22 electrodes. Sites Cz, Pz, P3, and P4 were selected because data collected at these sites were

less likely to be contaminated by artifacts associated with eye and eye muscle movements as well as facial and neck muscle movements.

Continuous EEG data were recorded using the Crew Response Evaluation Window (CREW) system (Bogart, 1999). The CREW system is a modular system capable of acquiring, analyzing, combining, and storing physiological data, environmental sensor data, binary status data, and National Television Standards Committee (NTSC) video streams. The specifications and operational characteristics of the instantiation of the CREW system used in the current investigation are described below.

Four channels of EEG voltages (i.e., one channel for each of the four recording sites) were amplified in individual BIOPAC Systems EEG100A preamplifiers and digitized at 1024 samples per second in a BIOPAC Systems MP100 data concentrator. The four channels of digitized EEG data were combined and sent via a serial interface to a PowerMac 8500/150. The data stream was read from the serial port by a LabVIEW Virtual Instrument (VI) that comprises the CREW software.

The CREW VI de-commutated the data stream into four individual data streams (i.e., one data stream for each of the EEG channels). Each data stream was appended to previous data that were kept in a ring buffer which held two seconds worth of data (i.e., each time new data were added to the ring buffer, the oldest data were removed so that the buffer always contained two seconds worth of data). An EEG analysis VI selected the newest 1024 data samples (i.e., the data vector) and computed a power spectrum using a standard Fast Fourier Transform (FFT) algorithm.

To compute a power spectrum, the CREW VI performed the following steps:

1. The mean of the vector was subtracted from the vector to eliminate the DC component of the signal. This eliminated offset errors introduced by the electrodes, wires, and preamplifiers.
2. The vector was smoothed using a triangle window (i.e., a standard windowing technique that modified the input array so that it began at zero amplitude, increased linearly to full amplitude at midpoint, and then decreased linearly to zero at the end). Windowing was necessary because continuous data, rather than a discrete length vector, must be used in conjunction with an FFT. (Although windowing reduced the magnitude of spectral powers, it did not alter their relative values; additionally, windowing eliminated "edge artifacts" that resulted when the FFT attempted to create outputs that described sudden onsets and offsets of the signal that were not actually present in the continuous signal.)
3. An FFT was used to calculate a spectral power vector that described the input data. Each value in the vector (i.e., a bin) represented the total power over a finite range of frequencies; therefore, as an example, the tenth bin represented the total power in the frequency range of 9.5 to 10.49 Hz.

EEG band powers were computed by summing the individual powers in the individual frequency bins. Theta power (i.e., 4 - 7 Hz) was the sum of bins 4 - 7; alpha power (i.e., 8 - 13 Hz) was the sum of bins 8 - 13; and beta power (i.e., 14 - 30 Hz) was the sum of bins 14 - 30. These were the values recorded in the actual data file. Although the data records were written every half second, each value recorded in the data file represented the EEG band power associated with the past one second "epoch." EEG data are typically recorded in this manner as it has the advantage of diluting short length noise events.

Test chambers. The test chambers used in this experiment were identical to those used in Experiment 1. Audiometric testing took place in the NASA LaRC Medical Center's free-standing Industrial Acoustics Company, Inc. audiometric test chambers, and aural alerting signal testing was conducted in the CHEM Laboratory's sound attenuated chamber.

Test system calibration. The NASA LaRC Medical Center's audiometer used to assess each subject's hearing threshold levels prior to the collection of experimental data was last calibrated using an artificial ear with an earphone coupler in August 1998, approximately 11 months prior to data collection for Experiment 2. The audiometer system also underwent daily biological calibrations with an octave monitor. Within the CHEM Laboratory's test chamber where experimental data were collected, the sound pressure levels at the subject's ear height were calibrated prior to each experimental session using a Simpson sound level meter (Model No. 866). Additionally, the joystick used by subjects during the manual condition of the MAT Battery's tracking task was calibrated using the Windows 98 calibration routine.

Experimental design. The experimental design used for data collection was a 4 (Alerting Set) x 3 (Urgency Level) x 2 (Tracking Condition), completely crossed, full factorial, within-subject design. The experimental design matrix is shown in Figure 29, and as shown in Figure 30, the same 12 subjects were assigned to each experimental cell.

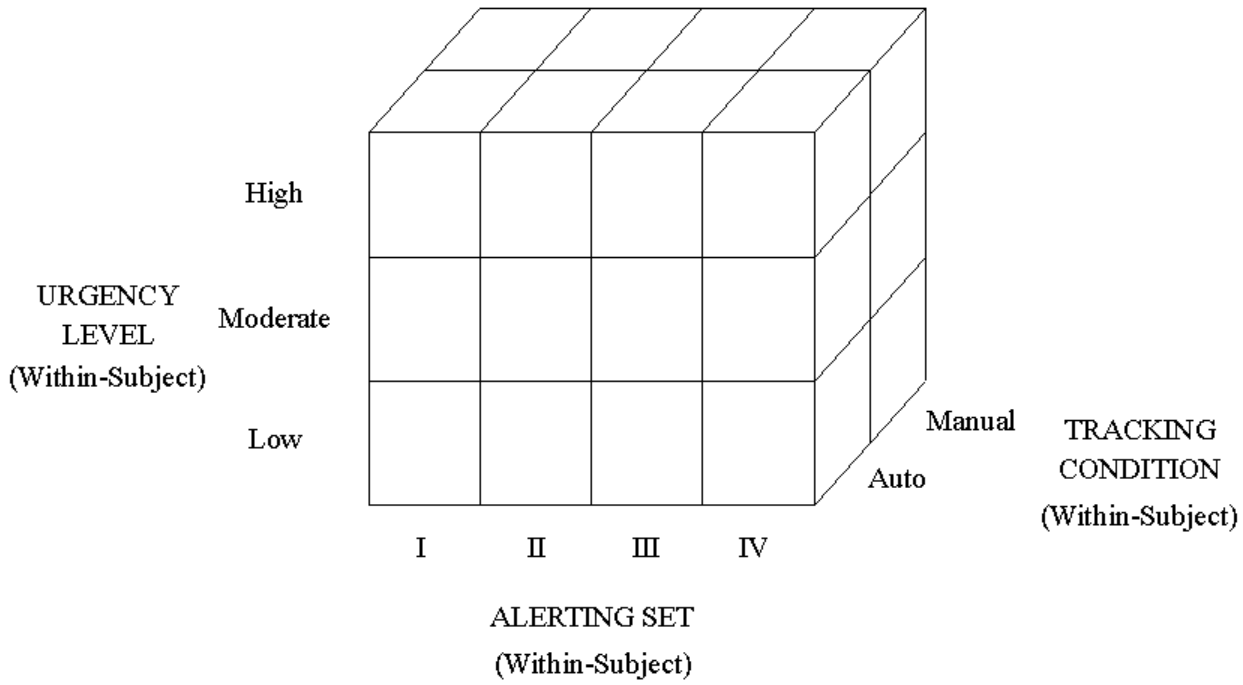


Figure 29. Experimental design matrix used in Experiment 2.

AUTOMATED TRACKING CONDITION

URGENCY LEVEL (Within-Subject)	I	II	III	IV
High	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}
Moderate	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}
Low	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}

ALERTING SET
(Within-Subject)

MANUAL TRACKING CONDITION

URGENCY LEVEL (Within-Subject)	I	II	III	IV
High	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}
Moderate	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}
Low	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}	S_1 - S_{12}

ALERTING SET
(Within-Subject)

Figure 30. Assignment of same 12 subjects to each experimental cell in Experiment 2.

All independent variables were treated as fixed-effects variables, and subjects were treated as a random-effect variable.

Independent variables. The three factors used in the experimental design were aural alerting set, urgency level, and tracking condition. The stimuli were 12 aural alerting signals, each of which belonged to one of four alerting sets and consisted of sound pulses and interpulse intervals having various durations. Most sound pulses included a linear onset time of 20 ms and a linear offset time of 20 ms, and sound pulses less than 40 ms in length had linear onset and offset times that peak at the middle of the pulse. The first harmonic, or fundamental frequency, of each sound pulse was present at 100%; and simultaneously, the second through fifth harmonics were present at 50% of the fundamental frequency's amplitude.

Specific acoustic parameters of the 12 aural stimuli are included in Tables 20 - 23. As described above, each aural alerting signal was comprised of complex tones made up of a controlled set of harmonics; however, harmonic content is not described in these tables. When examining Tables 20 - 23, note that each alert consisted of a given rhythmic pattern played twice; therefore, the total duration of each alert represents the sum of its sound pulse and interpulse interval durations multiplied by a factor of 2.

Table 20. Acoustic Parameters of Aural Stimuli Used in Experiment 2 - Set I

Urgency Level	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
Low	523	100%	104	104	3744
	523	50%	104	104	
	563	50%	104	104	
	563	75%	104	312	
	583	50%	104	104	
	583	100% to 50% logarithmic fade	416	208	
Moderate	523	100%	65	65	2340
	523	50%	65	65	
	563	50%	65	65	
	563	75%	65	195	
	583	50%	65	65	
	583	100% to 50% logarithmic fade	260	130	
High	523	100%	31	31	1122
	523	50%	31	31	
	563	50%	31	31	
	563	75%	31	94	
	583	50%	31	31	
	583	100% to 50% logarithmic fade	125	63	

Table 21. Acoustic Parameters of Aural Stimuli Used in Experiment 2 - Set II

Urgency Level	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
Low	593	100%	156	377	3866
	553	50%	78	22	
	593	50%	156	208	
	553	100% to 50% logarithmic fade	624	312	
Moderate	593	100%	98	235	2418
	553	50%	49	14	
	593	50%	98	130	
	553	100% to 50% logarithmic fade	390	195	
High	593	100%	47	113	1164
	553	50%	23	7	
	593	50%	47	63	
	553	100% to 50% logarithmic fade	188	94	

Table 22. Acoustic Parameters of Aural Stimuli Used in Experiment 2 - Set III

Urgency Level	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
Low	675	100%	117	117	5616
	655	50%	117	117	
	635	50%	117	351	
	635	75%	117	117	
	615	50%	117	117	
	598	50%	117	351	
	598	100% to 50% logarithmic fade	234	702	
Moderate	675	100%	73	73	3512
	655	50%	73	73	
	635	50%	73	220	
	635	75%	73	73	
	615	50%	73	73	
	598	50%	73	220	
	598	100% to 50% logarithmic fade	147	439	
High	675	100%	35	35	1686
	655	50%	35	35	
	635	50%	35	106	
	635	75%	35	35	
	615	50%	35	35	
	598	50%	35	106	
	598	100% to 50% logarithmic fade	70	211	

Table 23. Acoustic Parameters of Aural Stimuli Used in Experiment 2 - Set IV

Urgency Level	Sound Pulse Fundamental Frequency (Hz)	Sound Pulse Amplitude	Sound Pulse Duration (ms)	Interpulse Interval Duration (ms)	Total Duration of Alert (ms)
Low	635	100%	117	117	3744
	635	50%	117	351	
	523	50%	117	117	
	523	100% to 50% logarithmic fade	468	468	
Moderate	635	100%	73	73	2342
	635	50%	73	220	
	523	50%	73	73	
	523	100% to 50% logarithmic fade	293	293	
High	635	100%	35	35	1126
	635	50%	35	106	
	523	50%	35	35	
	523	100% to 50% logarithmic fade	141	141	

The alerting signals comprising Sets I, II, III, and IV differed from one another in rhythmic pattern and pitch contour as well as in fundamental frequency, pitch range, and duration. The high urgency level alerts were identical to those used in the Burt (1996) study as well as in Experiment 1. The low urgency level alerts and moderate urgency level alerts were selected based upon the results of Experiment 1 and were constructed according to Burt's (1996) recommendations. Details regarding the construction of the alerting signals may be found in previous sections of this document.

As with the alerts investigated by Burt (1996) and those investigated in Experiment 1, the stimulus parameters and frequency range corresponded with current research findings (Berson et al., 1981; Boucek et al., 1981; Edworthy, 1994b; Edworthy et al., 1991; Hanson et al., 1983; Hellier et al., 1993; Patterson, 1982, 1989) and design standards (ISO, 1986; SAE, 1993). Furthermore, the alerts were presented at 15 dBA above the 60 dBA masked threshold created by "level-flight" flight deck noise (i.e., all alerts were presented at 75 dBA). It is suggested that the aural alerting signals presented to subjects participating in this investigation were clearly audible; were sufficiently different from other sounds in the environment; and, as a result of the experimental manipulations of various sound parameters as well as a brief subject training session, had unambiguous meanings (ISO, 1986).

The conditions of the MAT Battery's tracking task (i.e., automated and manual) were used to manipulate workload and attentional engagement. As described previously, the manual condition of the tracking task required subjects to use a joystick to keep a circular target within a rectangular boundary. The automated condition of the tracking task did not require subjects to perform any action; the tracking task was automated to simulate the reduced manual demands

associated with the use of autopilot. In this experiment, the manual tracking task condition was used to demand a relatively high level of subject workload and attentional engagement, and the automated tracking task condition was used to demand a relatively low level of subject workload and attentional engagement.

Dependent measures. Three dependent measures were obtained in this experiment. The first dependent measure consisted of identifications of aural alerting sets and aural alert urgency levels made while performing the MAT Battery's automated and manual tracking task conditions. The second dependent measure consisted of subjective assessments of the level of workload experienced during each tracking task condition. The third dependent measure consisted of EEG data recorded throughout the performance of each tracking task condition.

Sound identifications were obtained through a sound identification task performed in conjunction with each of the MAT Battery's tracking task conditions. During the sound identification task, subjects were asked to determine the flight deck function to which each alert corresponded and to rate each alert as having either a low, moderate, or high urgency level. The purpose of this task was to provide subjects with an opportunity to demonstrate their ability to associate each alerting set with one of the four major flight deck functions and to simultaneously distinguish among, as well as identify three levels of urgency within, each alerting set while an additional task was performed. Subjects verbally identified the flight deck function and urgency level corresponding to each alert twice, and these data yielded frequency counts of correct and incorrect identifications. Since near perfect identifications of alerts and their urgency levels are required for the critical functions associated with flying an aircraft, a value of 95% correct identification was set as the criterion for acceptable performance.

Subjective assessments of workload were obtained using the MAT Battery's computerized version of the NASA TLX. The NASA TLX was used to collect weighted ratings on six sub-scales, and these data were used to calculate two mean weighted workload scores for each subject. After performing each condition of the MAT Battery's tracking task, subjects rated their perceived exertion on the Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration sub-scales and made factor comparisons of each of the 15 pairs of scale titles in order to select the variable that he or she felt was more important to the experience of workload. Mean weighted workload scores associated with the automated condition of the tracking task and mean weighted workload scores associated with the manual condition of the tracking task were used to determine the extent to which different levels of workload were perceived to be experienced during the two tracking conditions.

Continuous EEG data were recorded throughout the performance of the automated and manual conditions of the MAT Battery's tracking task. These data were used to assess the ongoing physiological responses of attention associated with each tracking task condition. EEG data were also evaluated in conjunction with the subjective NASA TLX workload ratings in order to obtain a more comprehensive understanding of the levels of workload and attentional engagement experienced by subjects during each tracking task condition.

Hypotheses. Sound identification data were collected to investigate hypotheses regarding the ability of subjects to: 1) associate each alerting set with one of the four major flight deck functions, and 2) simultaneously recognize a given aural alerting set and identify the correct urgency level within the set while performing a tracking task requiring two levels of workload and attentional engagement. It was hypothesized that subjects would identify the correct aural

alerting set, the correct urgency level, as well as the correct alerting set and urgency level 95% of the time while performing the MAT Battery's tracking task in automatic and manual modes. In other words, it was expected that subjects would choose the correct alerting set in 95% of the trials occurring during the automated tracking condition and in 95% of the trials occurring during the manual tracking condition; would choose the correct urgency level in 95% of the trials occurring during the automated tracking condition and in 95% of the trials occurring during the manual tracking condition; and would choose the correct alerting set and the correct urgency level in 95% of the trials occurring during the automated tracking condition and in 95% of the trials occurring during the manual tracking condition.

Since no single alerting set, urgency level, or aural alert was expected to be particularly easy or difficult to identify, sound identification data were also collected to investigate the hypothesis that subjects would correctly identify each alerting set, urgency level, and aural alert equally often while performing the MAT Battery's tracking task in automatic and manual modes. Said another way, it was not expected that subjects would correctly identify one alerting set, urgency level, or aural alert more often than any other alerting set, urgency level, or aural alert during the automated condition of the tracking task and/or the manual condition of the tracking task. Interactions between alerting sets, urgency levels, and/or tracking conditions were not expected to occur among sound identifications. However, if, for example, low, moderate, and/or high urgency level alerts presented during the automated tracking task condition had been correctly identified more often than low, moderate, and/or high urgency level alerts presented during the manual tracking task condition, this Urgency Level x Tracking Condition interaction would have been evaluated.

Additionally, sound identification data were collected to provide further insight into the ways in which systematic manipulations of acoustic parameters affect aural alert perception and recognition when a task that required two levels of workload and attentional engagement was performed. For example, the ease with which subjects were able to correctly identify Sets I, II, III, and IV during each condition of the tracking task provided information regarding the distinctiveness of the alerts' rhythmic patterns, pitch contours, fundamental frequencies, and pitch ranges when two levels of workload and attentional engagement were experienced. Similarly, the ease with which subjects were able to correctly identify low, moderate, and high urgency levels during each condition of the tracking task provided information regarding the urgency levels associated with various tempos when two levels of workload and attentional engagement were experienced. Finally, the ease with which subjects were able to correctly identify low, moderate, and high urgency levels within Sets I, II, III, and IV during each condition of the tracking task provided information regarding the ability of subjects to associate each alert with one of the major flight deck functions as well as their ability to simultaneously identify the correct alerting set and urgency level corresponding to each alert when appropriate acoustic parameters were manipulated and two levels of workload and attentional engagement were experienced.

Subjective assessments of workload were obtained to investigate the hypothesis that a higher level of workload would be perceived to be experienced by subjects during the manual condition of the tracking task. It was hypothesized that the NASA TLX mean weighted workload scores that subjects provided for the manual tracking task condition would be higher than the NASA TLX mean weighted workload scores that they provided for the automated tracking task condition.

EEG data were recorded to investigate the hypothesis that a higher level of attentional engagement would be experienced by subjects during the manual condition of the tracking task. It was hypothesized that more beta activity (14 - 30 Hz), less theta activity (4 - 7 Hz), and less alpha activity (8 - 13 Hz) would be present in the EEG data recorded during the performance of the manual tracking task condition when compared with the EEG data recorded during the performance of the automated tracking task condition.

Procedure. During a single visit to the CHEM Laboratory, each subject individually participated in a pre-experimental session, a training session, a subject feedback period, two data collection sessions, and a post-experimental session. Before arriving at the CHEM Laboratory, all subjects confirmed that their hearing threshold levels had been assessed at the NASA LaRC Medical Center within the preceding six months. The pre-experimental session involved gathering preliminary information from the subjects. During the training session, subjects were presented with aural alerting signals and were asked to: 1) associate functional categories (i.e., major flight deck functions) with four aural alerting sets, and 2) identify the flight deck function and urgency level corresponding to each of 12 alerts. During the subject feedback period, the experimenter and the subjects reviewed the accuracy of the flight deck function and urgency level identifications made during the training session. During the first data collection session, subjects were presented with aural alerting signals and were asked to simultaneously perform the MAT Battery's tracking task, in either the automatic or manual mode, as well as provide flight deck function and urgency level identifications for alerts while EEG data were recorded. During the second data collection session, subjects were presented with the same aural alerting signals and were asked to simultaneously perform the tracking task, in whichever mode they had not yet experienced, as

well as provide flight deck function and urgency level identifications for alerts while EEG data were recorded. Additionally, after completing each condition of the tracking task, subjects were asked to use the MAT Battery's computerized version of the NASA TLX to provide subjective workload ratings, and the experimenter recorded subjects' NASA TLX sub-scale weightings. The post-experimental session involved asking the subjects to complete a paper-and-pencil questionnaire (Appendix N), removing the electrodes required to record EEG data, and debriefing the subjects. The protocol and duration for the entire experiment, including audiometric testing, breaks, and electrode attachment, is shown in Table 24.

Table 24. Protocol and Duration for Experiment 2

Session	Duration
1. Audiometric testing	30 min
2. Pre-experimental session	15 min
3. Sound identification training session	30 min
4. Break	5 min
5. Subject feedback period	5 min
6. Electrode attachment	15 min
7. Sound identification data collection session #1	45 min
8. Break	5 min
9. Sound identification data collection session #2	45 min
10. Post-experimental session	15 min
Total:	3 hrs 30 min

As shown in Table 24, the total duration of this experiment was approximately three and a half hours.

The same 12 stimuli were used in the training session and both data collection sessions. During the training session, all stimuli were presented to each subject in the same random order. During the data collection sessions, all stimuli were presented in the partially counterbalanced order shown in Table 25.

Table 25. Balanced Latin Square Ordering of Stimuli Presentation Used in Experiment 2's Data Collection Sessions

Subject Number											
1	2	3	4	5	6	7	8	9	10	11	12
1	2	3	4	5	6	7	8	9	10	11	12
2	3	4	5	6	7	8	9	10	11	12	1
12	1	2	3	4	5	6	7	8	9	10	11
3	4	5	6	7	8	9	10	11	12	1	2
11	12	1	2	3	4	5	6	7	8	9	10
4	5	6	7	8	9	10	11	12	1	2	3
10	11	12	1	2	3	4	5	6	7	8	9
5	6	7	8	9	10	11	12	1	2	3	4
9	10	11	12	1	2	3	4	5	6	7	8
6	7	8	9	10	11	12	1	2	3	4	5
8	9	10	11	12	1	2	3	4	5	6	7
7	8	9	10	11	12	1	2	3	4	5	6

NOTE: The sequence of numbers listed vertically under each subject number represents the order in which each sound was presented to each subject. For example, Subject 1 heard Sound 1 (i.e., Set I Low Urgency), then Sound 2 (i.e., Set I Moderate Urgency), then Sound 12 (i.e., Set IV High Urgency), etc.

Prior to the collection of data, subjects signed the Informed Consent Form included in Appendix K, and the experimenter was provided with written documentation from the NASA LaRC Medical Center that subjects had “normal” hearing (i.e., hearing threshold levels in each ear that were $\leq$ 25 dB at 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz). Subjects also confirmed that they were right handed, had “normal” (i.e., 20/20 or better) or corrected-to-normal vision, and had no history of neurological problems that could interfere with the recording of EEG. Subjects then completed the questionnaire (Appendix A) regarding their musical experience, recent noise exposure, and types and doses of medications currently being taken which may have affected their perception of sound. After the collection of this preliminary information, subjects were seated in the test chamber. After a subject entered the test chamber, the chamber’s door was closed, and all communication between the subject and the experimenter took place through an intercom system. The protocol for the pre-experimental session is shown in Table 26.

Table 26. Protocol for the Pre-Experimental Session Used in Experiment 2

1. The experimenter greeted the subject.
2. The subject read and signed the Informed Consent Form (Appendix K).
3. The subject provided written documentation of “normal” hearing threshold levels from the NASA LaRC Medical Center.
4. The subject confirmed that he or she was right handed, had “normal” or corrected-to-normal vision, and had no history of neurological problems that could interfere with the recording of EEG.
5. The subject completed the Preliminary Questionnaire (Appendix A).
6. The subject was seated in the test chamber, and subsequent experimenter-subject communications took place through an intercom system.

The sound identification training session immediately followed the pre-experimental session. The protocol for the sound identification training session is shown in Table 27.

Table 27. Protocol for Sound Identification Training Session

1. The subject silently read the instructions for the sound identification training task while the experimenter read the instructions aloud.
2. The 12 alerts were presented, and the subject was instructed as to which alerting set corresponded to each of the four major flight deck functions.
3. Twelve to twenty sound identification practice trials were presented.
4. The subject provided two sound identifications for each of the 12 aural alerts. Alerts were presented to each subject in the same random order.
5. At the close of the session, the subject was given a 5 min break.

At the beginning of the sound identification training session, each subject silently read the following instructions while the experimenter read the instructions aloud:

You will hear four sets of sounds presented against a background of ambient 737 cockpit noise. Each set is comprised of a distinctive sound that is presented at three different speeds. As shown in the table below, the first, second, and third sounds that you will hear will form Set I; the fourth, fifth, and sixth sounds that you will hear will form Set II; the seventh, eighth, and ninth sounds that you will hear will form Set III; and the tenth, eleventh, and twelfth sounds that you will hear will form Set IV.

SET I	**SET II**	**SET III**	**SET IV**
Sound 1	Sound 4	Sound 7	Sound 10
Sound 2	Sound 5	Sound 8	Sound 11
Sound 3	Sound 6	Sound 9	Sound 12

Your first task is to make a "mental note" of each set's basic rhythm and pitch. As shown in the table below, all of the sounds in SET I - that is, Sounds 1, 2, and 3 - have a rhythm composed of six tones that increase in pitch. All of the sounds in SET II - that is, Sounds 4, 5, and 6 - have a rhythm composed of four tones that increase and decrease in pitch. All of the sounds in SET III - that is, Sounds 7, 8, and 9 - have a rhythm composed of seven tones that decrease in pitch. All of the sounds in SET IV - that is Sounds 10, 11, and 12 - have a rhythm composed of four tones that decrease in pitch.

SET I	**SET II**	**SET III**	**SET IV**
Sound 1	Sound 4	Sound 7	Sound 10
Sound 2	Sound 5	Sound 8	Sound 11
Sound 3	Sound 6	Sound 9	Sound 12
Six Tone Rhythm	Four Tone Rhythm	Seven Tone Rhythm	Four Tone Rhythm
Increasing Pitch	Increasing and Decreasing Pitch	Decreasing Pitch	Decreasing Pitch

Each set of sounds has a unique rhythmic pattern and pitch contour, and, as mentioned before, each set's distinctive sound is presented at three different speeds. Listen to Set I and notice how similar Sounds 1, 2, and 3 are - they only differ in their speed of presentation.

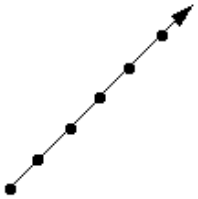
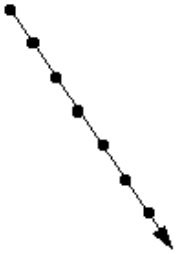
SET I	**SET II**	**SET III**	**SET IV**
Sound 1	Sound 4	Sound 7	Sound 10
Sound 2	Sound 5	Sound 8	Sound 11
Sound 3	Sound 6	Sound 9	Sound 12
Six Tone Rhythm	Four Tone Rhythm	Seven Tone Rhythm	Four Tone Rhythm
Increasing Pitch	Increasing and Decreasing Pitch	Decreasing Pitch	Decreasing Pitch

The experimenter presented Set I Low; 1.5 sec later, Set I Moderate was presented; and 1.5 sec later, Set I High was presented.

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to interact with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

I will now present the four sets to you twice in the same order. Please listen carefully to each sound and attempt to "mentally group" the sounds into the appropriate sets. The pitch contour of each set has been pictorially represented in the table below to help you remember the sounds.

SET I	SET II	SET III	SET IV
Sound 1	Sound 4	Sound 7	Sound 10
Sound 2	Sound 5	Sound 8	Sound 11
Sound 3	Sound 6	Sound 9	Sound 12
Six Tone Rhythm	Four Tone Rhythm	Seven Tone Rhythm	Four Tone Rhythm
Increasing Pitch 	Increasing and Decreasing Pitch 	Decreasing Pitch 	Decreasing Pitch 

The experimenter presented the 12 aural alerts twice in the following order: Set I Low, Set I Moderate, Set I High, Set II Low, Set II Moderate, Set II High, Set III Low, Set III Moderate, Set III High, Set IV Low, Set IV Moderate, and Set IV High.

> Would you like to have the four sets repeated, or do you feel confident that you can identify each set?

If the subject requested that the four sets be repeated, the experimenter presented each of the 12 alerts once in the following order: Set I Low, Set I Moderate, Set I High, Set II Low, Set II Moderate, Set II High, Set III Low, Set III Moderate, Set III High, Set IV Low, Set IV Moderate, and Set IV High.

> Your next task is to associate each of the four sets of sounds with one major flight deck function. The four flight deck functions that we will be concerned with include: 1) communication, 2) flight control, 3) navigation, and 4) systems management.

Communication involves managing the flow of information between each flight deck crew member, air traffic control (ATC), the cabin crew, passengers, and the airline company. If you were flying an aircraft and heard a sound from the "communication" set, the sound could, for example, advise you that the Datalink has been lost, caution you that the cockpit passenger address handset is inoperative, or warn you that an ATC radio is inoperative.

Flight control involves adjusting or maintaining the flight-path, attitude, and speed of the aircraft relative to the navigation requirements. If you were flying an aircraft and heard a sound from the "flight control" set, the sound could, for example, advise you that the automatic speedbrake is inoperative, caution you that the automatic throttle has been inadvertently disconnected, or warn you that the aircraft's center of gravity is grossly out of limits.

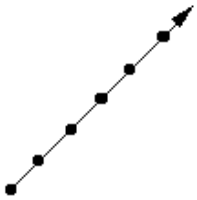
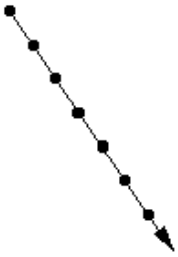
Navigation involves developing a desired plan of flight, positioning the aircraft relative to landmarks, and adjusting the plan of flight as necessary. If you were flying an aircraft and heard a sound from the "navigation" set, the sound could, for example, advise you that a marker beacon is inoperative, caution you that the First Officer's instrument landing system (ILS) is inoperative, or warn you that the aircraft is grossly off course.

Systems management involves monitoring the aircraft's systems. If you were flying an aircraft and heard a sound from the "systems management" set, the sound could, for example, advise you that the fuel temperature is low, caution you that one engine has failed, or warn you that the cabin altitude is dangerously high.

For our purposes, each of these four flight deck functions will be considered to be equally important.

As shown in the table below, SET I will be associated with communication; SET II will be associated with flight control; SET III will be associated with navigation; and SET IV will be associated with systems management. In order to perform well on the sound identification tasks, it is only necessary to identify the flight deck function to which a sound corresponds; you do not

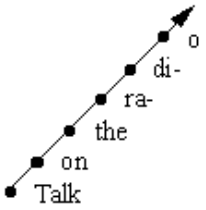
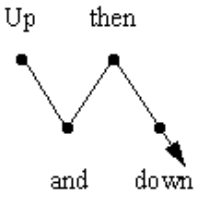
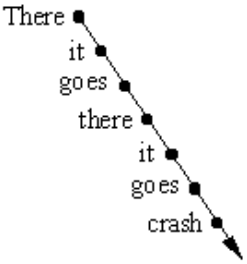
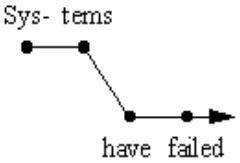
need to be concerned with the specific types of failures that may be associated with a given flight deck function.

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
SET I	**SET II**	**SET III**	**SET IV**
Sound 1	Sound 4	Sound 7	Sound 10
Sound 2	Sound 5	Sound 8	Sound 11
Sound 3	Sound 6	Sound 9	Sound 12
Six Tone Rhythm	Four Tone Rhythm	Seven Tone Rhythm	Four Tone Rhythm
Increasing Pitch 	Increasing and Decreasing Pitch 	Decreasing Pitch 	Decreasing Pitch 

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to respond to the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

I will now present the four sets to you twice in the same order. Please listen carefully to the sounds and attempt to associate each set of sounds with the appropriate flight deck function. You may find it helpful to associate the rhythm of each set with a group of words such as those shown in the examples below. Please feel free to take notes in the space provided on this page to familiarize yourself with the characteristics of each of the sound sets, but you will not be able to use such notes during the sound identification tasks.

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
SET I	**SET II**	**SET III**	**SET IV**
Sound 1	Sound 4	Sound 7	Sound 10
Sound 2	Sound 5	Sound 8	Sound 11
Sound 3	Sound 6	Sound 9	Sound 12
Six Tone Rhythm	Four Tone Rhythm	Seven Tone Rhythm	Four Tone Rhythm
Increasing Pitch 	Increasing and Decreasing Pitch 	Decreasing Pitch 	Decreasing Pitch 
"Talk on the radio"	"Up and then down"	"There it goes there it goes crash"	"Systems have failed"

NOTES:

The experimenter presented the 12 aural alerts twice in the following order: Set I Low, Set I Moderate, Set I High, Set II Low, Set II Moderate, Set II High, Set III Low, Set III Moderate, Set III High, Set IV Low, Set IV Moderate, and Set IV High.

Since the task of associating a given flight deck function with a set of sounds may be very difficult, I will now either: 1) repeat the four sets for you, or 2) give you some time to review your notes so that you may strengthen your association of the flight deck functions with the sets of sounds. Which would you prefer?

If the subject requested that the four sets be repeated, the experimenter presented each of the 12 alerts once in the following order: Set I Low, Set I Moderate, Set I High, Set II Low, Set II Moderate, Set II High, Set III Low, Set III Moderate, Set III High, Set IV Low, Set IV Moderate, and Set IV High. If the subject requested some time to review his or her notes, the reading of the task instructions did not resume until the subject acknowledged that he or she was ready to proceed.

You will now hear a series of sounds presented against a background of ambient 737 cockpit noise. After you hear each sound, please tell me whether the sound corresponds with:

COMMUNICATION	**FLIGHT CONTROL**	**NAVIGATION**	**SYSTEMS MANAGEMENT**

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to converse with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

I will now present you with the sounds.

The experimenter presented the subject with eight practice trials in which two randomly chosen alerts from each of the four sets served as the stimuli. After the presentation of each practice stimulus, the subject stated whether the alert corresponded with communication, flight control, navigation, or systems management, and the experimenter stated whether the subject's answer was correct or incorrect. The correct answer was always provided to the subject after an incorrect answer was given. If a subject failed to correctly identify all of the practice stimuli, he or she was given the opportunity to identify the practice stimuli a second time; however, a third series of the practice trials was not offered to, or in fact needed by, any of the subjects. After a

subject correctly identified all of the practice stimuli, the following instructions were read silently by the subject and were read aloud by the experimenter:

You will now perform a preliminary sound identification task. Please read the following instructions silently while I read them to you aloud:

You will hear a series of sounds presented against a background of ambient 737 cockpit noise. Your task is to identify the flight deck function to which a sound corresponds and to rate the sound as having either a low, moderate, or high level of urgency by making a check mark in the appropriate category like the one shown below.

COMMUNICATION	**FLIGHT CONTROL**	**NAVIGATION**	**SYSTEMS MANAGEMENT**
Urgency Level	**Urgency Level**	**Urgency Level**	**Urgency Level**
Low___ Mod___ High___	**Low___ Mod___ High___**	**Low___ Mod___ High✓**	**Low___ Mod___ High___**

You should only be concerned with the sounds - <u>not</u> with the background noise.

When you have heard the first sound, decide whether the sound corresponds to communication, flight control, navigation, or systems management and rate the sound as having either a low, moderate, or high level of urgency. Then, record your answer by making a check mark in the appropriate category on the response form. You will then hear the next sound. Do the same thing - decide to which flight deck function the sound corresponds and rate the level of urgency; then, record your answer by making a check mark in the appropriate category on the response form. You will do the same thing with all of the sounds that you hear.

For our purposes, “urgency” will be defined as “the quality or state of being important, insistent, or pressing.” A sound that is perceived as having a “low” level of urgency gives the impression that awareness is required and that future action may be necessary. A sound that is perceived as having a “moderate” level of urgency gives the impression that some form of action is required. A sound that is perceived as having a “high” level of urgency gives the impression that immediate action is required.

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to interact with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

I will now present you with four practice trials. Please use these practice trials to familiarize yourself with the rating procedure, and be sure to provide an identification and urgency rating for every sound.

The experimenter presented Practice Trial A; 6 sec later, Practice Trial B was presented; 6 sec later, Practice Trial C was presented; and 6 sec later, Practice Trial D was presented.

Do you have any final questions?

If the subject asked a question, the experimenter used the intercom system to speak with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

Please be certain that you listen to and provide an identification and urgency rating for every sound. I will now present the sounds.

The aural alerts were then presented to each subject twice in the same random order. Subjects recorded their decisions as to which flight deck function each signal corresponded as well as their perceptions of urgency level on the Sound Identification Training Task Response Form located in Appendix L. After a subject provided two sound identifications for each of the 12 alerts, he or she was given a 5 min break.

The subject feedback period immediately followed the break provided after the sound identification training session was completed. The protocol for the subject feedback period is shown in Table 28.

Table 28. Protocol for the Subject Feedback Period

1. The subject exited the test chamber and was seated at a table with the experimenter.
2. The experimenter and the subject reviewed the accuracy of the flight deck function and urgency level identifications made during the training session.
3. After subject feedback was provided, the electrodes required to record EEG data were attached to the subject's head as previously described.
4. Once the electrodes were in place, the subject was seated in the test chamber, and subsequent experimenter-subject communications took place through an intercom system.

During the break provided after the sound identification training session, the experimenter reviewed the subject's Sound Identification Training Task Response Form and calculated the subject's percentages of correct and incorrect flight deck function and urgency level identifications. After the 5 min break, the experimenter asked the subject to exit the test chamber and be seated at a table. Then, the experimenter and the subject reviewed the accuracy of the flight deck function and urgency level identifications made during the training session.

After the accuracy of the subject's identifications was reviewed, the electrodes required to record EEG data were attached. Once skin surfaces were cleaned with an alcohol pad, conductive gel and a self-adhering ground electrode were placed on the subject's left mastoid prominence, and conductive gel and a clip electrode were placed on the subject's left earlobe. Then, the frontal pole of the subject's brain was located through a measurement of head size; each subject's frontal pole corresponded to the point above his or her nasion (i.e., the bridge of the nose) that equaled 10% of the total distance between his or her nasion and inion (i.e., occipital protuberance). A lycra head cap consisting of 22 recessed electrodes was positioned on the subject's head according to his or her frontal pole, and conductive gel was applied to sites Cz (midline central), Pz (parietal central), P3 (left parietal), and P4 (right parietal). After the electrodes were making good contact (i.e., the impedance levels at the four recording sites and earlobe were reduced below 30 kOhms), the subject was seated in the test chamber, and subsequent experimenter-subject communications took place through an intercom system.

The first sound identification data collection session began after the electrodes were attached and the subject was comfortably seated in the test chamber. The protocol for the first sound identification data collection session is shown in Table 29.

Table 29. Protocol for Sound Identification Data Collection Session #1

1. A 3 min baseline EEG measurement was recorded while the subject's eyes were open.
2. A 3 min baseline EEG measurement was recorded while the subject's eyes were closed.
3. The subject silently read the instructions for the first sound identification task while the experimenter read the instructions aloud.
4. The subject engaged in 3 min of practice with the first condition of the tracking task, and four sound identification practice trials were presented.
5. The subject performed the first condition of the tracking task for 30 min while he or she provided verbal identifications of the flight deck function and urgency level that corresponded to each aural alert and his or her EEG data were recorded. Tracking task conditions were presented in counterbalanced order; alerts were presented in partially counterbalanced order at randomly occurring intervals of time; and two sound identifications were provided for each of the 12 alerts.
6. The experimenter provided the subject with verbal instructions regarding how to perform the NASA TLX rating of workload level.
7. The subject recorded his or her perception of the level of workload experienced while performing the first condition of the tracking task using a computerized version of the NASA TLX.
8. At the close of the session, the subject was given a 5 min break.

During the first sound identification data collection session, subjects were instructed to keep their facial muscles and jaw as relaxed as possible and to refrain from excessive eye blinking while their EEG data were recorded. Then, a 3 min baseline EEG measurement was recorded while each subject sat quietly with his or her eyes open, and a 3 min baseline EEG measurement was recorded while each subject sat quietly with his or her eyes closed. After baseline measurements were recorded, each subject silently read the instructions provided below, which were adapted in part from those developed by Comstock and Arnegard (1992), while the experimenter read the instructions aloud. Note that in this document the instructions associated with the automated condition of the tracking task are provided in conjunction with the first sound identification data collection session, and the instructions associated with the manual condition of the tracking task are provided in conjunction with the second sound identification data collection session. The tracking task conditions were presented to subjects, however, in counterbalanced order. That is, half of the subjects performed the automated condition of the tracking task during the first sound identification data collection session and performed the manual condition of the tracking task during the second sound identification data collection session; and the other half of the subjects performed the manual condition of the tracking task during the first sound identification data collection session and performed the automated condition of the tracking task during the second sound identification data collection session.

> The purpose of this portion of the study is to assess the ability of individuals to identify the correct flight deck function and urgency level associated with each of 12 aural alerting signals while performing a task that requires a certain level of engagement or "busyness." The task that is displayed before you on the computer monitor is a simulation of one kind of task that pilots perform: tracking. All of the information that you will need to perform the tracking task is

currently displayed on the monitor. The overall purpose of the tracking task is to keep the airplane symbol, represented by the circle, within the dotted rectangular area in the center of the task.

During this experimental session, you will perform the tracking task in an automatic or "AUTO" mode, as indicated in the lower left corner of the window. Since the tracking task is in "AUTO" mode, the computer will control the airplane symbol. It is your responsibility to visually monitor the computer's tracking performance. You are not responsible for controlling the airplane symbol in any way.

While the computer is controlling the airplane symbol, you will hear a series of aural alerts presented against a background of ambient 737 cockpit noise. You should only be concerned with the alerting signals - not with the background noise. The alerts that you will hear are identical to the alerts that you associated with the four major flight deck functions in the sound identification training session. Whenever you hear an alert, verbally identify the flight deck function and urgency level that corresponds to the alert as quickly as you can. For example, if you hear an alert that is associated with communication, and the alert conveys a low level of urgency, please respond by saying "communication low." Similarly, if you hear an alert that is associated with flight control, and the alert conveys a moderate level of urgency, please respond by saying "flight control moderate."

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to interact with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

You will perform the tracking task in "AUTO" mode and verbally identify aural alerts presented at random time intervals for a total of 30 minutes. After you complete the automated tracking condition, the tracking task will disappear from the computer monitor, and a series of six rating scales will be displayed. You will use these rating scales to report the level of

workload that you experienced while performing the automated condition of the tracking task; however, once the rating scales appear on the monitor, please wait for my instructions before you attempt to enter your ratings.

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to speak with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

Before the actual tracking task begins, you will have a three minute practice period. During this practice period, visually monitor the computer's tracking performance; listen carefully for the alerting signals; and verbally identify the flight deck function and urgency level that corresponds to each alert as quickly as you can. Do you have any final questions?

If the subject asked a question, the experimenter used the intercom system to respond to the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

Are you ready to begin the practice period?

Once the subject indicated that he or she was ready to being the practice period, the experimenter provided the following instructions:

Remember to visually monitor the computer's tracking performance; listen carefully for the alerting signals; and verbally identify the flight deck function and urgency level that corresponds to each alert as quickly as you can. The practice period will now begin.

The subject practiced performing the automated condition of the tracking task for 3 min. During this practice period, four sound identification trials were presented; one randomly chosen alert from each of the four sets served as the stimuli. After the presentation of each practice stimulus, the subject stated whether the alert corresponded with communication, flight control, navigation,

or systems management and stated whether the alert corresponded to a low, moderate, or high urgency level. After the practice period was completed, the following instructions were read aloud by the experimenter:

> You will now perform the automated condition of the tracking task for 30 minutes. Please try to keep your facial muscles and jaw as relaxed as possible and try to refrain from excessive eye blinking; this will help ensure that a good recording of your brain waves is obtained. Also, please wait for my instructions before you attempt to enter your ratings on the scales that will appear on the computer monitor once you complete the tracking task.
>
> Do you have any final questions?

If the subject asked a question, the experimenter used the intercom system to speak with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

> Throughout this task, please be certain that you visually monitor the computer's tracking performance; listen carefully for the alerting signals; and verbally identify the flight deck function and urgency level that corresponds to each alert as quickly as you can. Are you ready to begin the tracking task?

Once the subject indicated that he or she was ready to being the tracking task, the automated condition of the tracking task began. The aural alerts were presented at random time intervals, and each alert was presented twice in partially counterbalanced order. Subjects verbally reported the flight deck function and urgency level that corresponded to each alert, and the experimenter recorded each response on the Sound Identification Data Collection Form located in Appendix M. No subject ever failed to provide a verbal identification for an alert.

When the 30 min automated tracking task condition was completed, the experimenter read the following instructions to the subject:

You have now completed the automated tracking condition. You will use the six rating scales displayed on the computer monitor in front of you to report the level of workload that you experienced while performing the automated condition of the tracking task. Please listen to the following instructions before you attempt to enter your ratings.

The objective of this part of the study is to capture your perceived workload level. The concept of workload is hard to define specifically, as it is composed of many different aspects. Workload may refer, in part, to the physical demands of a task, the time pressure involved, your expended effort, or your resulting stress or frustration levels. I hope to understand the workload associated with the automated condition of the tracking task by asking you to describe various feelings and perceptions that you experienced while performing the task. Since many factors may be involved, I'd like for you to tell me about several individual factors rather than provide me with one overall workload score.

The set of six rating scales that you now see before you was developed at NASA Ames Research Center and has been used in a wide variety of tasks. As you can see, there are six scales on which you will be asked to provide a rating score: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The first scale, Mental Demand, refers to the level of mental activity like thinking, deciding, and looking that was required during the task. You will rate this scale from low (on the left side) to high (on the right side). The second scale, Physical Demand, involves the amount of physical activity required of you, such as controlling or activating. The third scale, Temporal Demand, refers to the time pressure that you experienced during the task. In other words, was the pace of the task slow and leisurely or rapid and frantic? The fourth scale, Performance, involves your perceptions about your own performance level. Your rating here should reflect your satisfaction with your performance in accomplishing the goals of the task. Notice that this scale ranges from good (on the left side) to poor (on the right side). All of the other scales range from low to high. The fifth scale, Effort, inquires as to how hard you had to work (both mentally and physically) in order to achieve your

level of performance. Finally, the sixth scale, Frustration, is an index of how secure and relaxed (low frustration) versus stressed and discouraged (high frustration) you felt during the task.

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to converse with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

Now, I will explain the method you will use to rate your experiences with these six scales. The pointer on the first scale is currently illuminated in yellow. You must select the score that best reflects your perceptions about that scale from low to high by using the computer mouse located on the desk in front of you. Move the mouse to the left or right to place the pointer at any point along the rating scale's horizontal line; the position of the pointer is only restricted by the rating scale's two endpoints. After you have decided upon your score, use the mouse's left button to move to the next scale. You will notice that the pointer on the first scale will turn to a gray color, and the pointer on the active second scale will turn yellow. The yellow pointer always indicates the scale that is active or available for change.

After you have entered the sixth score, you may either exit the rating screen by pressing the mouse's right button, or, if you wish to change your scores, you may do so by pressing the mouse's left button. If, for example, you want to change the response provided on the Effort sub-scale, you will press the mouse's left button five times. Five depressions of the left button will allow you to return to the fifth sub-scale without altering any of the scores entered on the first four sub-scales. The pointer of the fifth rating scale will turn yellow, and you can adjust your rating by moving the mouse to the right or left. After you make any desired changes, you will use the mouse's right button to exit the workload rating screen.

When you exit the workload rating screen, a pair of rating scale titles will appear on the computer monitor. Please wait for additional instructions when you reach this screen.

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to interact with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

> Give your responses thoughtful consideration, but do not spend too much time deliberating over them. Your first responses will probably accurately reflect your feelings and experiences. Be sure that you consider the tracking task as well as the task of providing verbal sound identifications when you make your workload ratings. Please begin entering your responses on the rating scales, and feel free to ask any questions that you like.

If the subject asked a question while entering his or her workload ratings, the experimenter used the intercom system to speak with the subject.

Once the subject exited the workload rating screen, the experimenter read the following instructions to the subject:

> You will now be presented with several pairs of scale titles similar to the one currently displayed on the computer monitor in front of you. For each pair of scale titles, please tell me the title of the scale, or variable, that you feel was more important to the level of workload that you experienced during the automated tracking task condition.

After the subject made factor comparisons of each pair of scale titles, the subject was given a 5 min break.

The second sound identification data collection session immediately followed the break provided after the completion of the first sound identification data collection session. The protocol for the second sound identification data collection session is shown in Table 30.

Table 30. Protocol for Sound Identification Data Collection Session #2

1. The subject silently read the instructions for the second sound identification task while the experimenter read the instructions aloud.
2. The subject engaged in 3 min of practice with the second condition of the tracking task, and four sound identification practice trials were presented.
3. The subject performed the second condition of the tracking task for 30 min while he or she provided verbal identifications of the flight deck function and urgency level corresponding to each aural alert and his or her EEG data were recorded. Tracking task conditions were presented in counterbalanced order; alerts were presented in partially counterbalanced order at randomly occurring intervals of time; and two sound identifications were provided for each of the 12 alerts.
4. The experimenter provided the subject with verbal instructions regarding how to perform the NASA TLX rating of workload level.
5. The subject recorded his or her perception of the level of workload experienced while performing the second condition of the tracking task using a computerized version of the NASA TLX.

During the second sound identification data collection session, each subject silently read the instructions provided below, which were adapted in part from those developed by Comstock and Arnegard (1992), while the experimenter read the instructions aloud. In this document, the instructions associated with the manual condition of the tracking task are presented in conjunction with the second sound identification data collection session; however, the tracking task conditions were presented to subjects in counterbalanced order.

> As stated previously, the purpose of this portion of the study is to assess the ability of individuals to identify the correct flight deck function and urgency level associated with each of 12 aural alerting signals while performing a task that requires a certain level of engagement or "busyness." The task that is displayed before you on the computer monitor is a simulation of one kind of task that pilots perform: tracking. All of the information that you will need to perform the tracking task is currently displayed on the monitor. The overall purpose of the tracking task is to keep the airplane symbol, represented by the circle, within the dotted rectangular area in the center of the task.
>
> During this experimental session, you will perform the tracking task in a "MANUAL" mode, as indicated in the lower left corner of the window. Since the tracking task is in "MANUAL" mode, you will be responsible for controlling the airplane symbol with the joystick located on the desk in front of you. You must control the airplane symbol with movements of the joystick. To move the airplane symbol to a different area of the screen, you will move the joystick in that identical direction. If you do not control the airplane symbol with the joystick, the plane will drift away from the center. Basically, you must compensate for this random drifting by pulling the plane back to center with corresponding movements with the joystick. For example, if the plane is drifting to the right, moving the joystick to the left will return the plane to center. Most of the time, however, you will be working in two dimensions: horizontal <u>and</u> vertical; so you will be making many diagonal movements. If the plane is away from the center,

you must make rather large movements to return it. If the plane is already in the center, smaller movements will be required. Remember, the overall purpose of this task is to keep the plane in the center rectangular area. Try to maintain this at all times. If the plane leaves the rectangular area, try to return the plane to center as quickly as possible.

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to converse with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

While you are controlling the airplane symbol, you will hear a series of aural alerts presented against a background of ambient 737 cockpit noise. You should only be concerned with the alerting signals - not with the background noise. The alerts that you will hear are identical to the alerts that you associated with the four major flight deck functions in the sound identification training session. Whenever you hear an alert, verbally identify the flight deck function and urgency level that corresponds to the alert as quickly as you can. For example, if you hear an alert that is associated with communication, and the alert conveys a low level of urgency, please respond by saying "communication low." Similarly, if you hear an alert that is associated with flight control, and the alert conveys a moderate level of urgency, please respond by saying "flight control moderate."

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to speak with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

You will perform the tracking task in "MANUAL" mode and verbally identify aural alerts presented at random time intervals for a total of 30 minutes. After you complete the manual tracking condition, the tracking task will disappear from the computer monitor, and a

series of six rating scales will be displayed. You will use these rating scales to report the level of workload that you experienced while performing the manual condition of the tracking task; however, once the rating scales appear on the computer monitor, please wait for my instructions before you attempt to enter your ratings.

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to respond to the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

Before the actual tracking task begins, you will have a three minute practice period. During this practice period, use the joystick to keep the airplane symbol in the center rectangular area; listen carefully for the alerting signals; and verbally identify the flight deck function and urgency level that corresponds to each alert as quickly as you can. Do you have any final questions?

If the subject asked a question, the experimenter used the intercom system to interact with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

Are you ready to begin the practice period?

Once the subject indicated that he or she was ready to being the practice period, the experimenter provided the following instructions:

Remember to use the joystick to keep the airplane symbol in the center rectangular area; listen carefully for the alerting signals; and verbally identify the flight deck function and urgency level that corresponds to each alert as quickly as you can. The practice period will now begin.

The subject practiced performing the manual condition of the tracking task for 3 min. During this practice period, four sound identification trials were presented; one randomly chosen alert from each of the four sets served as the stimuli. After the presentation of each practice stimulus, the

subject stated whether the alert corresponded with communication, flight control, navigation, or systems management and stated whether the alert corresponded to a low, moderate, or high urgency level. When the practice period was completed, the following instructions were read aloud by the experimenter:

> You will now perform the manual condition of the tracking task for 30 minutes. Please try to keep your facial muscles and jaw as relaxed as possible and try to refrain from excessive eye blinking; this will help ensure that a good recording of your brain waves is obtained. Also, please wait for my instructions before you attempt to enter your ratings on the scales that will appear on the computer monitor once you complete the tracking task.
>
> Do you have any final questions?

If the subject asked a question, the experimenter used the intercom system to speak with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

> Throughout this task, please be certain that you keep the plane in the center rectangular area; listen carefully for the alerting signals; and verbally identify the flight deck function and urgency level that corresponds to each alert as quickly as you can. Are you ready to begin the tracking task?

When the subject indicated that he or she was ready to being the tracking task, the manual condition of the tracking task began. The aural alerts were presented at random time intervals, and each alert was presented twice in partially counterbalanced order. Subjects verbally reported the flight deck function and urgency level corresponding to each alert, and the experimenter recorded each response on the Sound Identification Data Collection Form located in Appendix M. No subject ever failed to provide a verbal identification for an alert.

Once the 30 min manual tracking task condition was completed, the experimenter read the following instructions to the subject:

> You have now completed the manual tracking condition. You will use the six rating scales displayed on the computer monitor in front of you to report the level of workload that you experienced while performing the manual condition of the tracking task. Please listen to the following instructions before you attempt to enter your ratings.
>
> The objective of this part of the study is to capture your perceived workload level. The concept of workload is hard to define specifically, as it is composed of many different aspects. Workload may refer, in part, to the physical demands of a task, the time pressure involved, your expended effort, or your resulting stress or frustration levels. I hope to understand the workload associated with the manual condition of the tracking task by asking you to describe various feelings and perceptions that you experienced while performing the task. Since many factors may be involved, I'd like for you to tell me about several individual factors rather than provide me with one overall workload score.
>
> The set of six rating scales that you now see before you was developed at NASA Ames Research Center and has been used in a wide variety of tasks. As you can see, there are six scales on which you will be asked to provide a rating score: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The first scale, Mental Demand, refers to the level of mental activity like thinking, deciding, and looking that was required during the task. You will rate this scale from low (on the left side) to high (on the right side). The second scale, Physical Demand, involves the amount of physical activity required of you, such as controlling or activating. The third scale, Temporal Demand, refers to the time pressure that you experienced during the task. In other words, was the pace of the task slow and leisurely or rapid and frantic? The fourth scale, Performance, involves your perceptions about your own performance level. Your rating here should reflect your satisfaction with your performance in accomplishing the goals of the task. Notice that this scale ranges from good (on the left side) to

poor (on the right side). All of the other scales range from low to high. The fifth scale, Effort, inquires as to how hard you had to work (both mentally and physically) in order to achieve your level of performance. Finally, the sixth scale, Frustration, is an index of how secure and relaxed (low frustration) versus stressed and discouraged (high frustration) you felt during the task.

Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to converse with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

Now, I will explain the method you will use to rate your experiences with these six scales. The pointer on the first scale is currently illuminated in yellow. You must select the score that best reflects your perceptions about that scale from low to high by using the computer mouse located on the desk in front of you. Move the mouse to the right or left to place the pointer at any point along the rating scale's horizontal line; the position of the pointer is only restricted by the rating scale's two endpoints. After you have decided upon your score, use the mouse's left button to move to the next scale. You will notice that the pointer on the first scale will turn to a gray color, and the pointer on the active second scale will turn yellow. The yellow pointer always indicates the scale that is active or available for change.

After you have entered the sixth score, you may either exit the rating screen by pressing the mouse's right button, or, if you wish to change your scores, you may do so by pressing the mouse's left button. If, for example, you want to change the response provided on the Effort sub-scale, you will press the mouse's left button five times. Five depressions of the left button will allow you to return to the fifth sub-scale without altering any of the scores entered on the first four sub-scales. The pointer of the fifth rating scale will turn yellow, and you can adjust your rating by moving the mouse to the right or left. After you make any desired changes, you will use the mouse's right button to exit the workload rating screen.

> When you exit the workload rating screen, a pair of rating scale titles will appear on the computer monitor. Please wait for additional instructions when you reach this screen.
>
> Do you have any questions?

If the subject asked a question, the experimenter used the intercom system to speak with the subject. Once the question(s) asked and the answer(s) given were understood by both the subject and the experimenter, the reading of the task instructions resumed.

> Give your responses thoughtful consideration, but do not spend too much time deliberating over them. Your first responses will probably accurately reflect your feelings and experiences. Be sure that you consider the tracking task as well as the task of providing verbal sound identifications when you make your workload ratings. Please begin entering your responses on the rating scales, and feel free to ask any questions that you like.

If the subject asked a question while entering his or her workload ratings, the experimenter used the intercom system to respond to the subject.

Once the subject exited the workload rating screen, the experimenter read the following instructions to the subject:

> You will now be presented with several pairs of scale titles similar to the one currently displayed on the computer monitor in front of you. For each pair of scale titles, please tell me the title of the scale, or variable, that you feel was more important to the level of workload that you experienced during the manual tracking task condition.

The second sound identification data collection session was completed when the subject finished making factor comparisons of each pair of scale titles.

The post-experimental session immediately followed the second sound identification data collection session. The protocol for the post-experimental session is shown in Table 31.

Table 31. Protocol for the Post-Experimental Session

1. The subject completed the Post-Experiment Questionnaire (Appendix N).
2. The subject exited the experimental chamber, and the experimenter removed all of the electrodes required to record EEG data.
3. The subject was debriefed.

As shown in Table 31, the post-experimental session involved requesting that the subject complete the paper-and-pencil questionnaire included in Appendix N; removing the lycra head cap from the subject's head; removing the reference electrode attached to the subject's left earlobe; removing the ground electrode attached to the subject's left mastoid prominence; and debriefing the subject.

Results and Discussion

Sound identification data. During the sound identification training session, subjects associated each alerting set with a major flight deck function and then provided 24 sound identifications (i.e., each subject provided two identifications for each of 12 aural alerts). These data yielded frequency counts of correct and incorrect identifications and were used to determine the extent to which subjects were able to correctly identify the alerting set (i.e., major flight deck function) to which an aural alert corresponded; correctly identify low, moderate, and high urgency levels; and correctly identify the alerting set and urgency level associated with each alert.

Frequency counts and percentages of correct identifications made during the sound identification training session are shown in Tables 32, 33, and 34. Table 32 reveals that subjects correctly identified the alerting set to which a signal corresponded in 96.88% of the trials when identifications were averaged across the three urgency levels (i.e., N = 72). Table 33 reveals that subjects correctly identified the low, moderate, and high urgency levels in 90.97% of the trials when identifications were averaged across the four alerting sets (i.e., N = 96). Table 34 reveals that subjects correctly identified the alerting set and urgency level associated with a given alert in 88.54% of the trials (i.e., N = 24).

Table 32. Alerting Set Identifications Made During the Sound Identification Training Session

Alerting Set	Number of Correct Identifications	Number of Incorrect Identifications
I	69	3
II	70	2
III	70	2
IV	70	2
Total	279	9
Grand Total	288	
Percentage of Correct Identifications	96.88	

Table 33. Urgency Level Identifications Made During the Sound Identification Training Session

Urgency Level	Number of Correct Identifications	Number of Incorrect Identifications
Low	87	9
Moderate	86	10
High	89	7
Total	262	26
Grand Total	288	
Percentage of Correct Identifications	90.97	

Table 34. Alerting Set and Urgency Level Identifications Made During the Sound Identification Training Session

Aural Alert	Number of Correct Identifications	Number of Incorrect Identifications
Set I Low	20	4
Set I Moderate	20	4
Set I High	22	2
Set II Low	21	3
Set II Moderate	21	3
Set II High	22	2
Set III Low	22	2
Set III Moderate	22	2
Set III High	20	4
Set IV Low	22	2
Set IV Moderate	22	2
Set IV High	21	3
Total	255	33
Grand Total	288	
Percentage of Correct Identifications	88.54	

Cochran's Q Tests (i.e., nonparametric within-subject tests appropriate for analyzing three or more related samples of nominal data) were performed on mean percentages of correct identifications to determine if a given alerting set, urgency level, or aural alert was correctly or incorrectly identified more often by the subjects (Norusis, 1992). Based on these tests, no significant differences were found among identifications of alerting set (Q [3] = 0.5294; p = 0.9124); among identifications of urgency level (Q [2] = 0.9333; p = 0.6271); or among identifications of aural alerts (Q [11] = 5.8235; p = 0.8849). Therefore, each alerting set, urgency level, and aural alert was correctly identified equally often.

Mean percentages of correct identifications and 95% confidence intervals are displayed in Figures 31, 32 and 33. Figure 31 depicts the mean percentages of correct identifications associated with Set I, Set II, Set III, and Set IV when identifications were averaged across the three urgency levels (i.e., N = 72). Figure 32 depicts the mean percentages of correct identifications associated with the low, moderate, and high urgency levels when identifications were averaged across the four alerting sets (i.e., N = 96). Figure 33 depicts the mean percentages of correct identifications associated with each urgency level within each alerting set (i.e., N = 24).

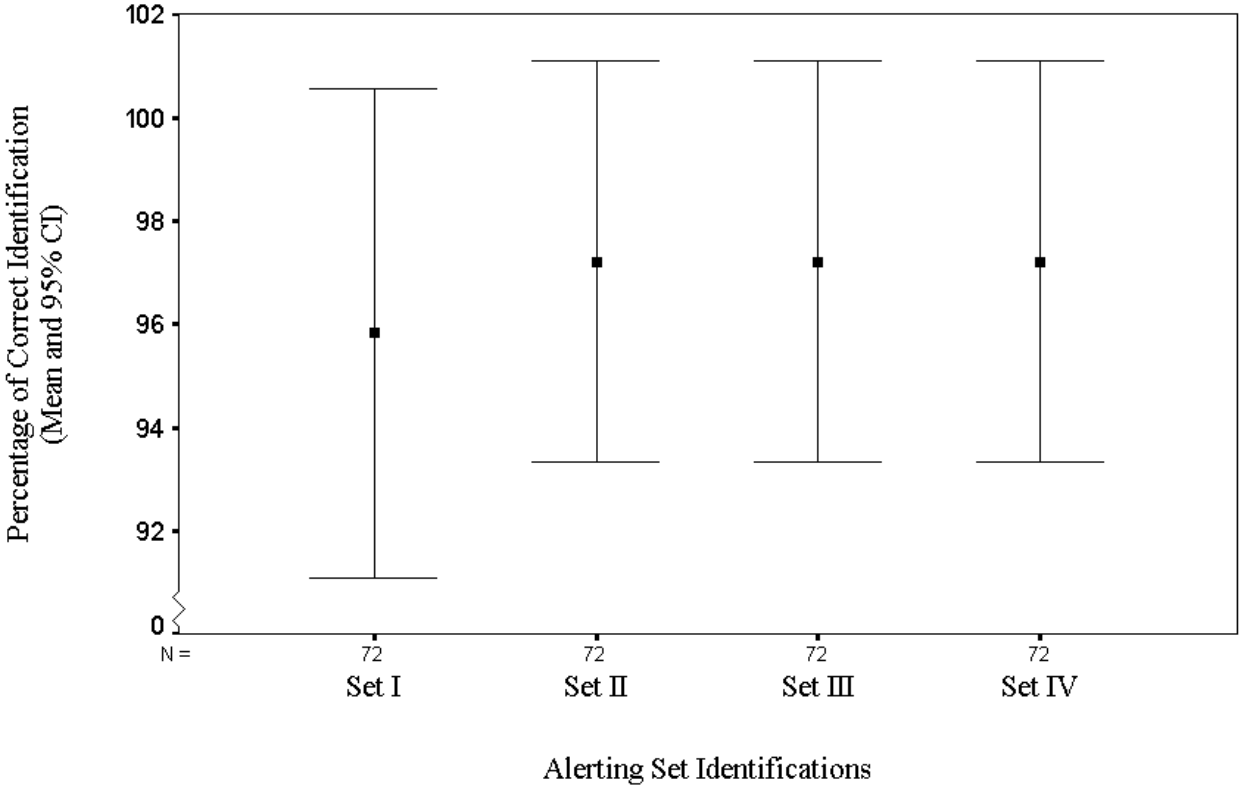


Figure 31. Percentages of correct alerting set identifications made during the sound identification training session collapsed across urgency level.

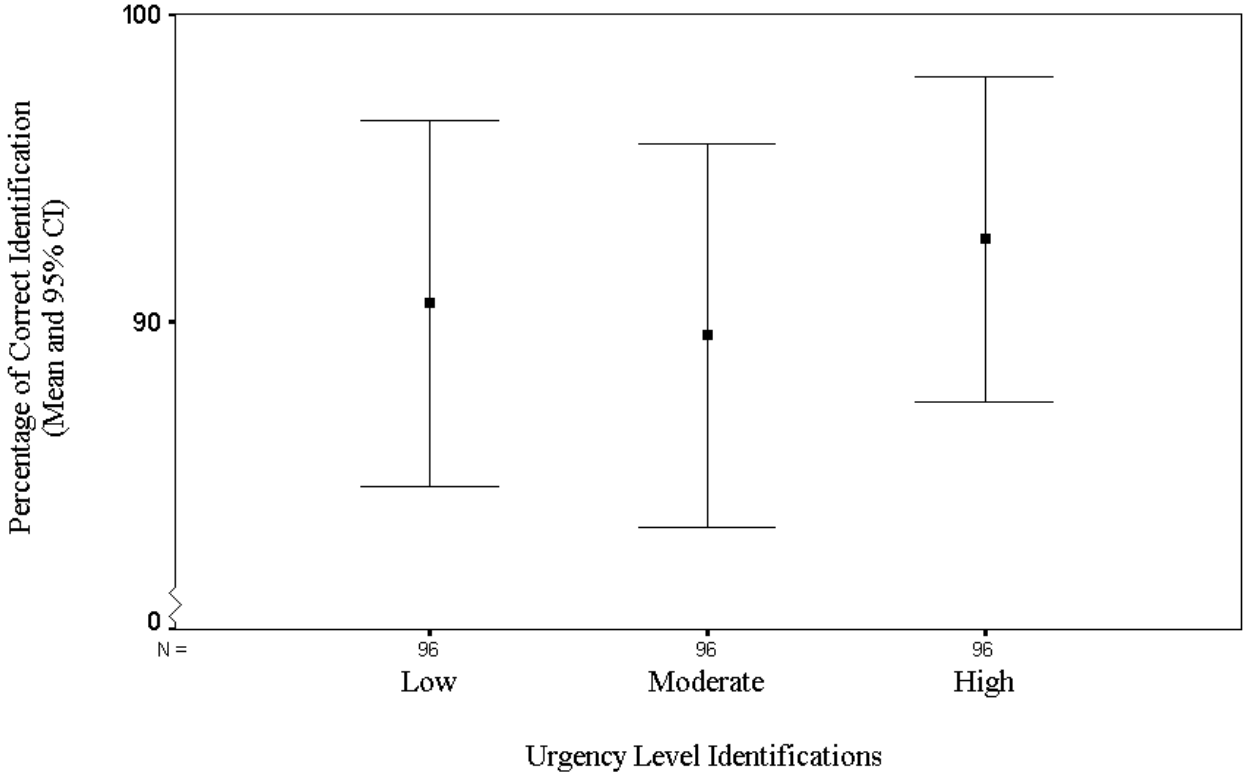


Figure 32. Percentages of correct urgency level identifications made during the sound identification training session collapsed across alerting set.

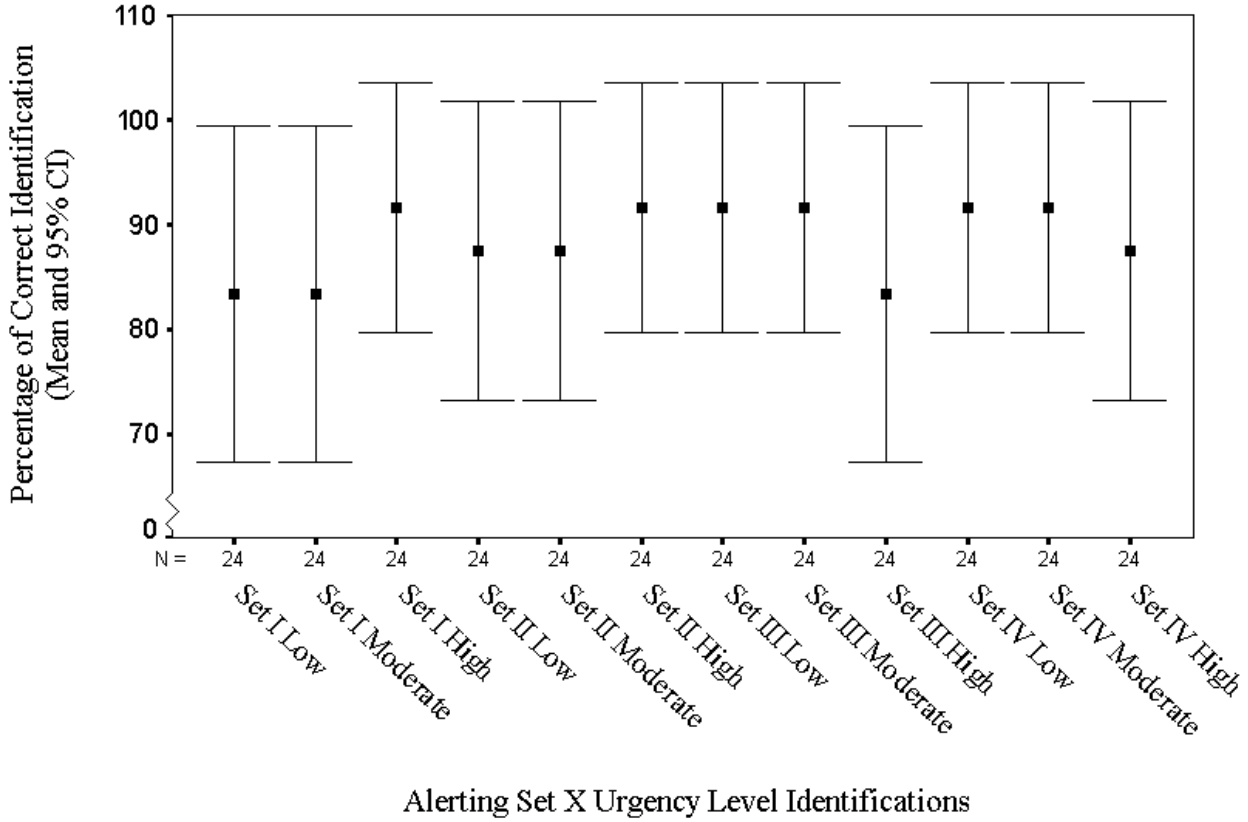


Figure 33. Percentages of correct alerting set and urgency level identifications made during the sound identification training session.

The percentages of correct identifications made during the sound identification training session indicate that subjects were able to associate each of four distinctive aural alerting sets with a major flight deck function and were then able to correctly identify the flight deck function as well as the urgency level corresponding to each alert 88.54% of the time after completing a brief instructional period. Due to the saliency of acoustic parameter manipulations, as well as the necessity for near perfect identifications of flight deck alerts and their urgency levels, subjects were expected to be able to correctly identify each alerting set, urgency level, and aural alert at least 95% of the time while performing each condition of the MAT Battery's tracking task. Therefore, the subject feedback period, which followed the sound identification training session and preceded the performance of the tracking task conditions, was used to review the accuracy of identifications made during the training session so that subjects' questions regarding the alerts could be answered and any confusion regarding the alerts' meanings could be eliminated. Additionally, the acoustic parameter manipulations used to differentiate the alerting sets as well as convey a sense of urgency were explained to the subjects.

During each tracking task condition, each subject provided 24 sound identifications (i.e., two identifications were provided for each aural alert during the automated tracking task condition, and two identifications were provided for each aural alert during the manual tracking task condition). These data yielded frequency counts of correct and incorrect identifications and were used to determine the extent to which subjects were able to correctly identify the alerting set to which an aural alert corresponded; correctly identify low, moderate, and high urgency levels; and correctly identify the alerting set <u>and</u> urgency level associated with each alert while performing a tracking task requiring two levels of workload and attentional engagement.

For each condition of the tracking task, the percentage of trials in which subjects correctly identified the alerting set to which a signal corresponded was calculated; the percentage of trials in which subjects correctly identified the low, moderate, and high urgency levels was calculated; and the percentage of trials in which subjects correctly identified the alerting set and urgency level associated with a given alert was calculated. Additionally, the percentage of trials in which subjects correctly identified the alerting set to which a signal corresponded during the automated and the manual tracking task conditions was calculated; the percentage of trials in which subjects correctly identified the low, moderate, and high urgency levels during the automated and the manual tracking task conditions was calculated; and the percentage of trials in which subjects correctly identified the alerting set and urgency level associated with a given alert during the automated and the manual tracking task conditions was calculated. Percentages of correct identifications made during each as well as both tracking task condition(s) are shown in Table 35.

Table 35. Percentages of Correct Identifications Made During Tracking Task Conditions

	Alerting Set	Urgency Level	Aural Alert
Automated Tracking Task	97.92%	97.92%	96.53%
Manual Tracking Task	97.57%	97.92%	95.83%
Automated *and* Manual Tracking Tasks	97.74%	97.92%	96.18%

Table 35 reveals that subjects correctly identified the alerting set to which a signal corresponded in 97.92% of the trials when identifications were averaged across the three urgency levels (i.e., N = 72); correctly identified the low, moderate, and high urgency levels in 97.92% of the trials when identifications were averaged across the four alerting sets (i.e., N = 96); and correctly identified the alerting set and urgency level associated with a given alert in 96.53% of the trials (i.e., N = 24) during the automated condition of the tracking task. During the manual condition of the tracking task, subjects correctly identified the alerting set to which a signal corresponded in 97.57% of the trials when identifications were averaged across the three urgency levels (i.e., N = 72); correctly identified the low, moderate, and high urgency levels in 97.92% of the trials when identifications were averaged across the four alerting sets (i.e., N = 96); and correctly identified the alerting set and urgency level associated with a given alert in 95.83% of the trials (i.e., N = 24). During the automated and manual conditions of the tracking task, subjects correctly identified the alerting set to which a signal corresponded in 97.74% of the trials when identifications were averaged across the three urgency levels as well as across the two tracking task conditions (i.e., N = 144); correctly identified the low, moderate, and high urgency levels in 97.92% of the trials when identifications were averaged across the four alerting sets as well as across the two tracking task conditions (i.e., N = 192); and correctly identified the alerting set and urgency level associated with a given alert in 96.18% of the trials when identifications were averaged across the two tracking task conditions (i.e., N = 48).

Cochran's Q Tests (i.e., nonparametric within-subject tests appropriate for analyzing three or more related samples of nominal data) were performed on the means associated with the percentages of correct identifications made during the automated tracking task condition and the

means associated with the percentages of correct identifications made during the manual tracking task condition to determine if a given alerting set, urgency level, or aural alert was correctly or incorrectly identified more often by the subjects during either tracking task condition (Norusis, 1992). Based on these tests, no significant differences were found among identifications of alerting set (Q [7] = 1.2353; p = 0.9901); among identifications of urgency level (Q [5] = 1.3043; p = 0.9345); or among identifications of aural alerts (Q [23] = 13.3852; p = 0.9431) made during either tracking task condition. Therefore, each alerting set, urgency level, and aural alert was correctly identified equally often during both conditions of the tracking task.

Mean percentages of correct identifications and 95% confidence intervals calculated for the identifications made during the automated versus the manual condition of the tracking task are shown in Figures 34 - 36. Figure 34 depicts the mean percentages of correct identifications made for Set I, Set II, Set III, and Set IV during each tracking task condition when identifications were averaged across the three urgency levels (i.e., N = 72). Figure 35 depicts the mean percentages of correct identifications made for the low, moderate, and high urgency levels during each tracking task condition when identifications were averaged across the four alerting sets (i.e., N = 96). Figure 36 depicts the mean percentages of correct identifications made for each urgency level within each alerting set during each tracking task condition (i.e., N = 24).

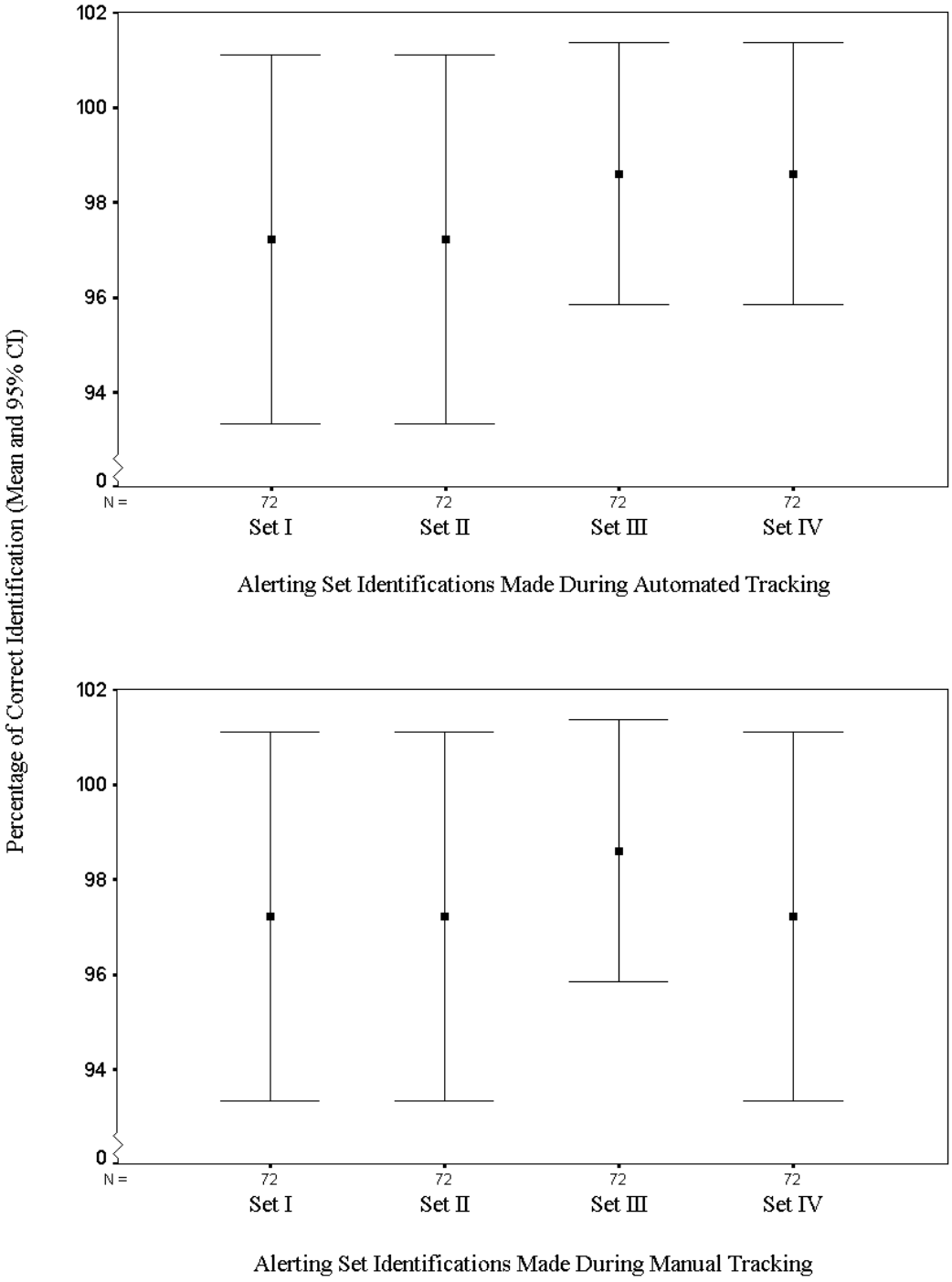


Figure 34. Percentages of correct alerting set identifications made during each tracking condition collapsed across urgency level.

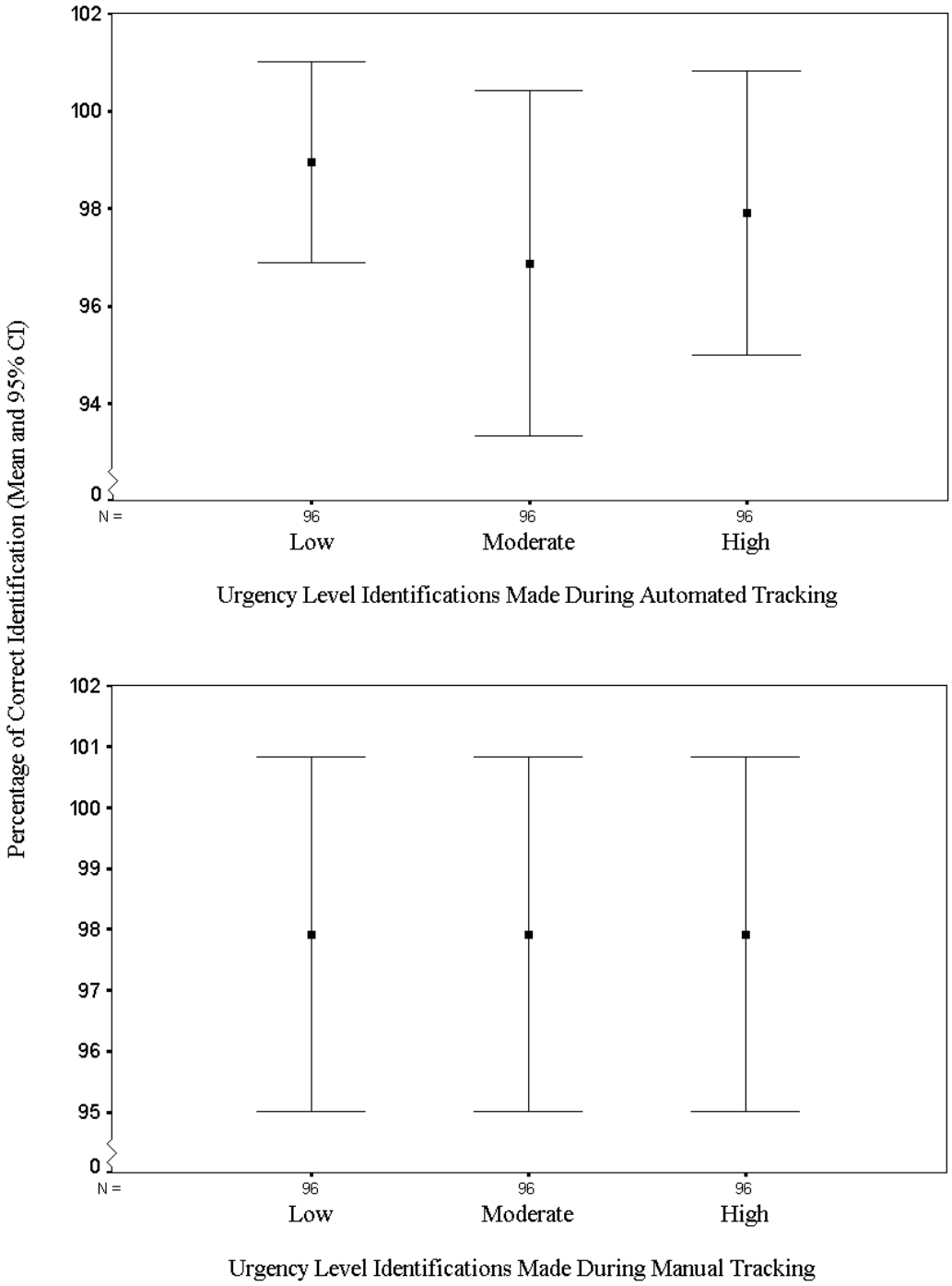


Figure 35. Percentages of correct urgency level identifications made during each tracking condition collapsed across alerting set.

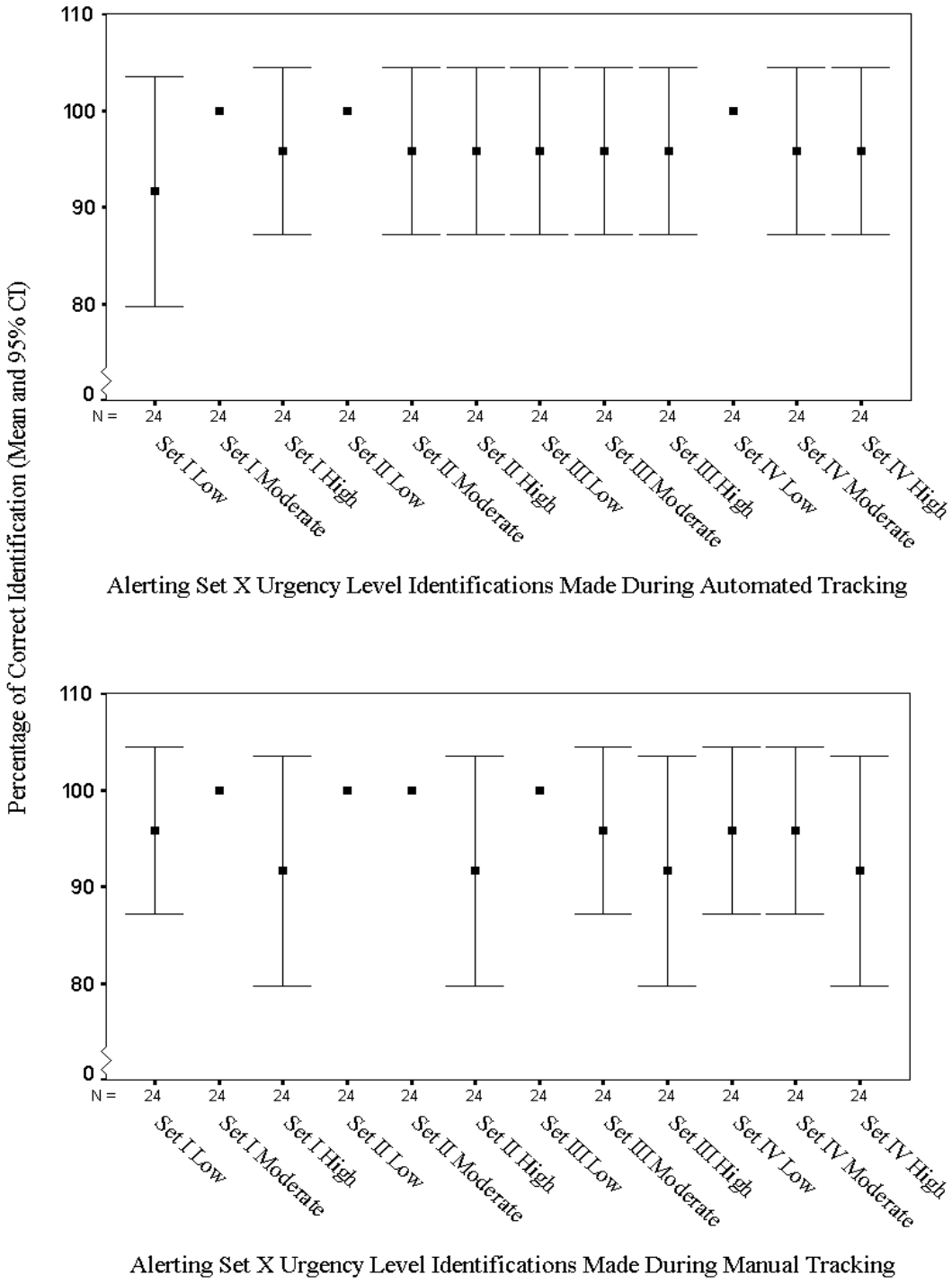


Figure 36. Percentages of correct alerting set and urgency level identifications made during each tracking condition.

Mean percentages of correct identifications and 95% confidence intervals calculated for the identifications made during the automated and manual tracking task conditions are shown in Figures 37 - 39. Figure 37 depicts the mean percentages of correct identifications made for Set I, Set II, Set III, and Set IV when identifications were averaged across the three urgency levels and both tracking task conditions (i.e., N = 144). Figure 38 depicts the mean percentages of correct identifications made for the low, moderate, and high urgency levels when identifications were averaged across the four alerting sets and both tracking task conditions (i.e., N = 192). Figure 39 depicts the mean percentages of correct identifications made for each urgency level within each alerting set when identifications were averaged across both tracking task conditions (i.e., N = 48).

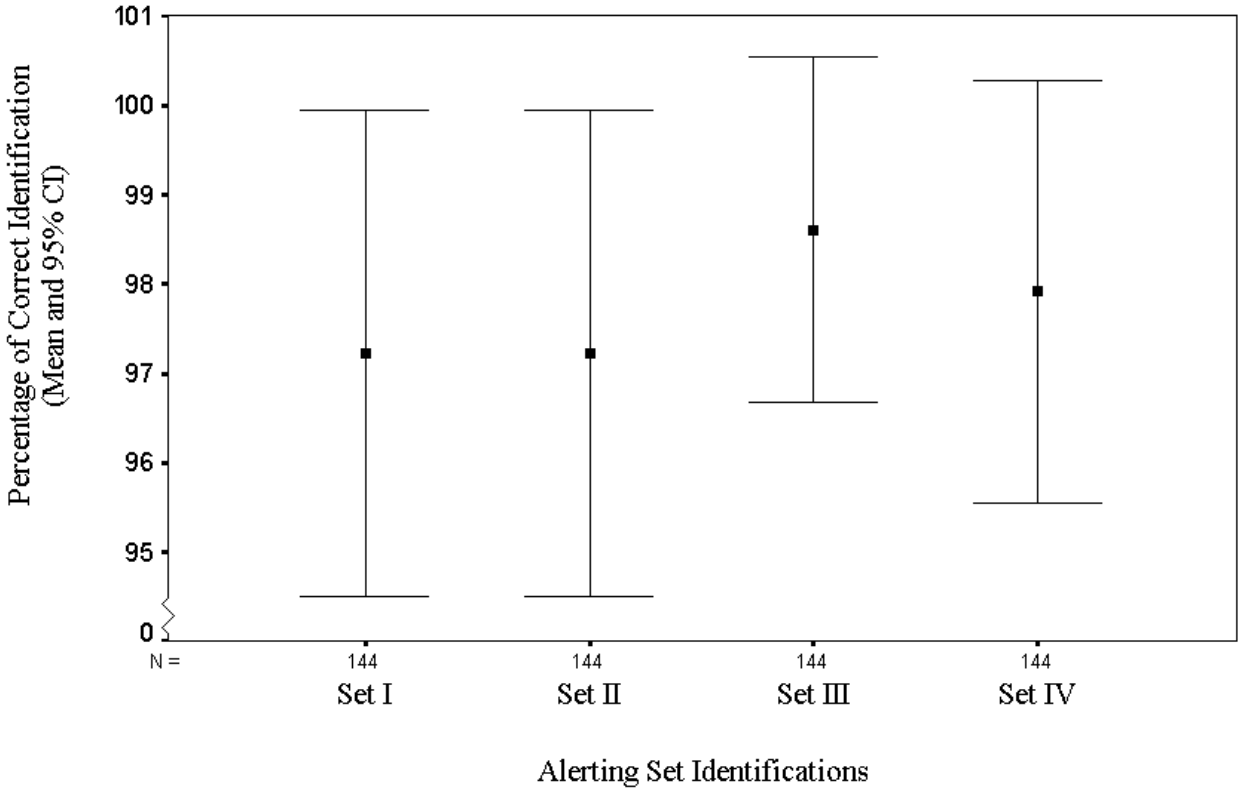


Figure 37. Percentages of correct alerting set identifications collapsed across urgency level and tracking condition.

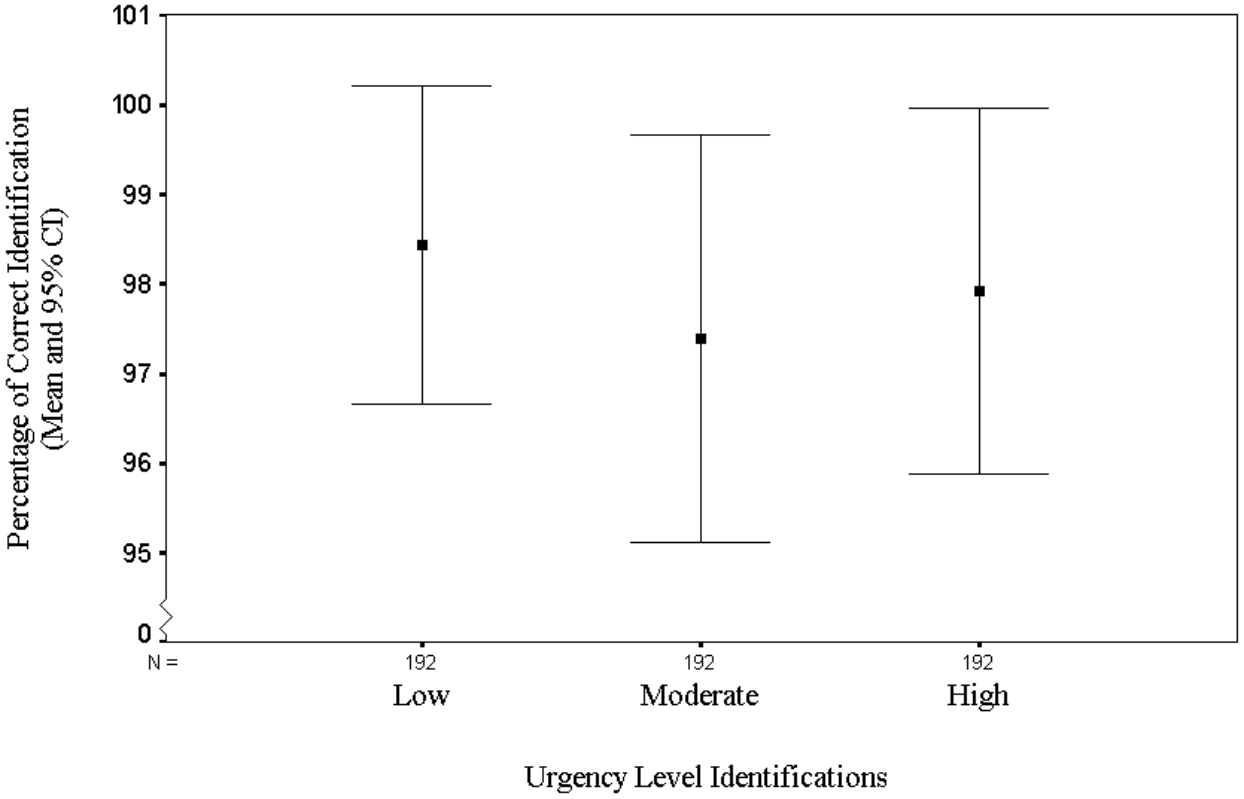


Figure 38. Percentages of correct urgency level identifications collapsed across alerting set and tracking condition.

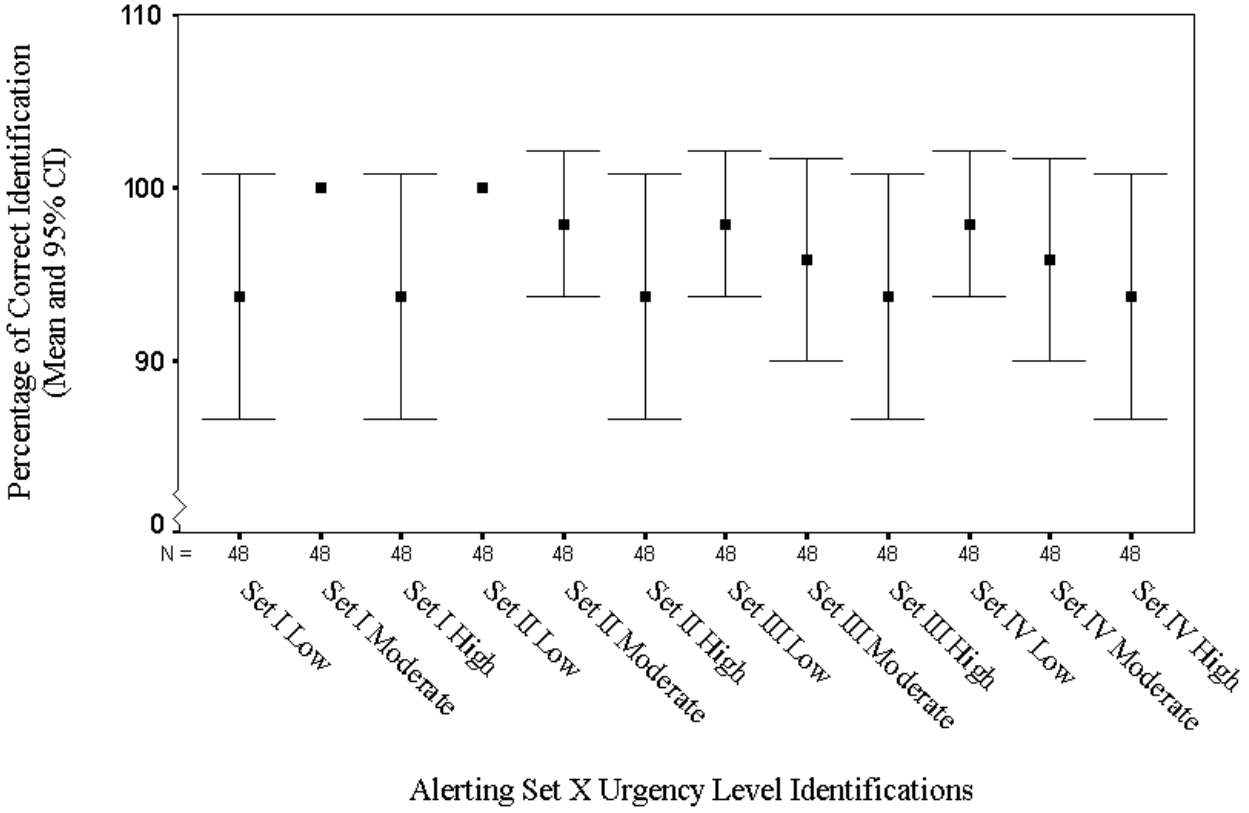


Figure 39. Percentages of correct alerting set and urgency level identifications collapsed across tracking condition.

The percentages of correct sound identifications and the results of the Cochran's Q Tests support the expectation that subjects would be able to associate each alerting set with one of the four major flight deck functions and would also be able to simultaneously recognize a given aural alerting set as well as identify the correct urgency level within the set while performing a task requiring two levels of workload and attentional engagement. It was hypothesized that subjects would identify the correct aural alerting set, the correct urgency level, as well as the correct alerting set and urgency level 95% of the time while performing the MAT Battery's tracking task in automatic and manual modes. This hypothesis was supported by evidence suggesting that subjects chose the correct alerting set in 97.92% of the trials occurring during the automated tracking task condition and in 97.57% of the trials occurring during the manual tracking task condition; chose the correct urgency level in 97.92% of the trials occurring during the automated tracking task condition and in 97.92% of the trials occurring during the manual tracking task condition; and chose the correct alerting set and the correct urgency level in 96.53% of the trials occurring during the automated tracking task condition and in 95.83% of the trials occurring during the manual tracking task condition.

It was also hypothesized that subjects would correctly identify each alerting set, urgency level, and aural alert equally often while performing the MAT Battery's tracking task in automatic and manual modes. This hypothesis was supported by evidence suggesting that interactions between alerting sets, urgency levels, and/or tracking conditions did not occur among sound identifications.

The sound identification data also provided insight into the ways in which systematic manipulations of acoustic parameters affected aural alert perception and recognition when a task

that required different levels of workload and attentional engagement was performed. For example, the ease with which subjects correctly identified Sets I, II, III, and IV during each tracking task condition supported the expectation that rhythmic pattern, pitch contour, fundamental frequency, and pitch range manipulations could be used to create alerting sets that were distinctive when two levels of workload and attentional engagement were experienced. Subjects chose the correct alerting set in 97.92% of the trials occurring during the automated tracking task condition as well as in 97.57% of the trials occurring during the manual tracking task condition. This evidence suggests that the alerts' rhythmic patterns, pitch contours, fundamental frequencies, and pitch ranges were perceived to be distinctive under two levels of workload and attentional engagement.

The ease with which subjects correctly identified low, moderate, and high urgency levels during each tracking task condition provided information regarding the urgency levels associated with various tempos when two levels of workload and attentional engagement were experienced. Since subjects chose the correct urgency level in 97.92% of the trials occurring during the automated tracking task condition as well as in 97.92% of the trials occurring during the manual tracking task condition, it is suggested that appropriate urgency levels were perceived to be associated with various tempos under two levels of workload and attentional engagement.

Finally, the ease with which subjects correctly identified low, moderate, and high urgency levels within Sets I, II, III, and IV during each tracking task condition provided information regarding the ability of subjects to associate alerting sets with functional categories; distinguish among and recognize alerting sets; and perceive intended urgency levels when rhythmic pattern, pitch contour, fundamental frequency, pitch range, and tempo were manipulated and two levels of

workload and attentional engagement were experienced. Subjects chose the correct alerting set and the correct urgency level in 96.53% of the trials occurring during the automated tracking task condition as well as in 95.83% of the trials occurring during the manual tracking task condition. This evidence suggests that individuals were able to associate each alert with one of the major flight deck functions and simultaneously identify the correct alerting sets and urgency levels corresponding to various alerts when appropriate acoustic parameters were manipulated and two levels of workload and attentional engagement were experienced.

Subjective workload ratings. Two NASA TLX mean weighted workload scores were calculated for each subject using the procedure outlined previously in Table 8. One score corresponded to the overall level of workload that a subject reported experiencing during the automated tracking task condition, and the other score corresponded to the overall level of workload that the subject reported experiencing during the manual tracking task condition. All subjects provided workload ratings using the same linear scales, and a number range of 0 to 100 was applied to responses made along the rating scales' horizontal lines. Since responses could be made anywhere along the continuum of "low" to "high" as well as along the continuum of "good" to "poor," the assumption of normally distributed data was met.

Workload rating data were analyzed using a one-way ANOVA to determine if subjects reported experiencing a significantly higher level of workload during the manual condition of the tracking task. As shown by the ANOVA Summary Table presented in Table 36, a significant difference was found to exist between subjective workload ratings of tracking condition (F [1, 11] = 12.68; $p < 0.05$). Since a Greenhouse-Geisser correction can not be used to correct the problems associated with a positively biased F-Test when only two levels of a factor are

investigated, such a correction was not employed in the analysis of workload ratings collected after the performance of the automated and manual tracking task conditions.

Table 36. ANOVA Summary Table of Subjective Workload Rating Data

Source	df	SS	MS	F	p
Between-Subjects					
Subjects (S)	11	9385.54			
Within-Subject					
Tracking Condition (TC)	1	1486.13	1483.13	12.68	0.004
TC X S	11	1286.25	116.93		
Total	23	12154.92			

Mean workload ratings and 95% confidence intervals calculated for each of the tracking task conditions (i.e., N = 12) are shown in Figure 40. As shown in this figure, subjects reported experiencing a mean weighted workload score of 34.41 during the automated condition of the tracking task and reported experiencing a mean weighted workload score of 50.13 during the manual condition of the tracking task.

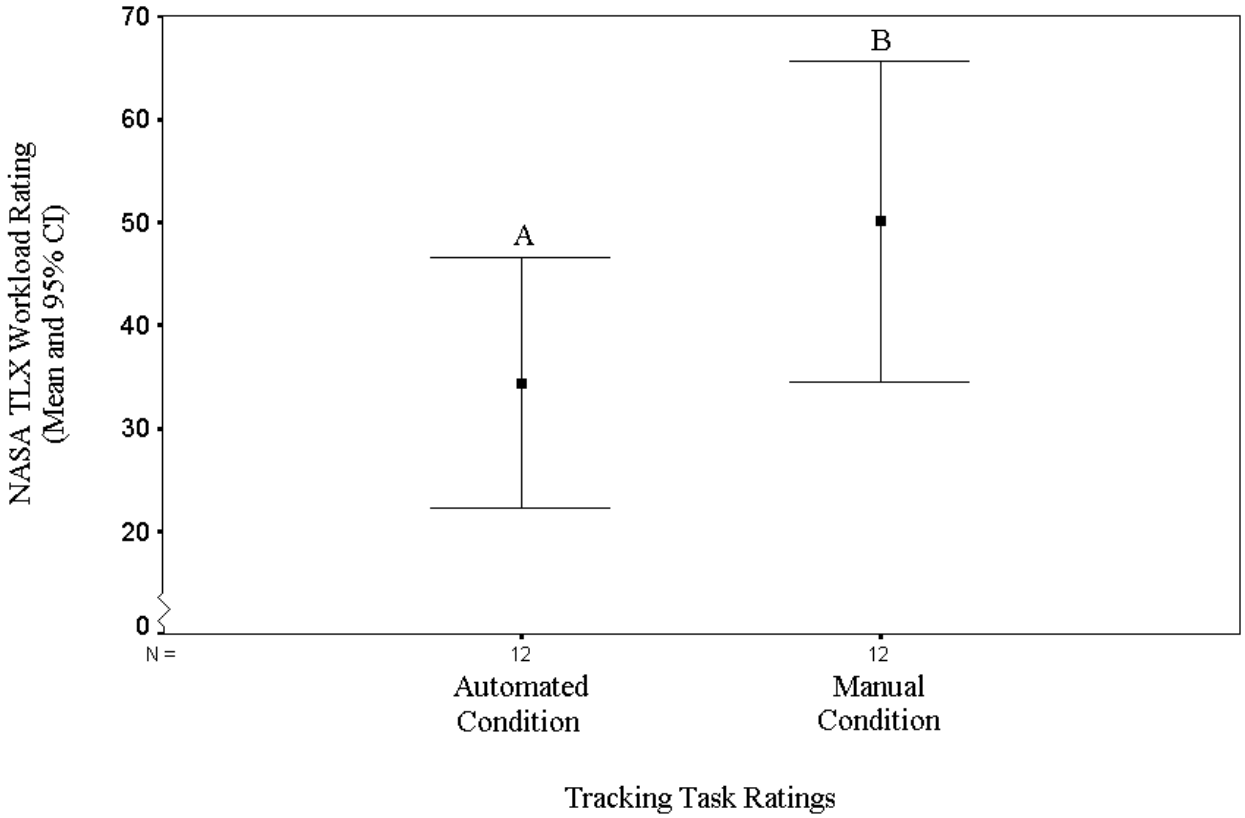


Figure 40. NASA TLX workload ratings of tracking task conditions. (NOTE: Means with different letters are significantly different in an ANOVA at $p < 0.05$.)

The results of the ANOVA support the hypothesis that the mean weighted workload scores that subjects provided for the manual tracking task condition would be higher than the mean weighted workload scores that they provided for the automated tracking task condition. As expected, the null hypothesis that the same overall workload rating would be provided for the automated and manual conditions of the tracking task was rejected. This evidence suggests that two different levels of workload were experienced by subjects while performing the automated and manual conditions of the MAT Battery's tracking task.

EEG data. Continuous EEG data were recorded during two baseline conditions as well as during the automated and manual conditions of the tracking task. The powers (i.e., one measure of activity) associated with the theta (4 - 7 Hz), alpha (8 - 13 Hz), and beta (14 - 30 Hz) frequency bands were collected at sites Cz, Pz, P3, and P4 as explained previously.

Although EEG band power data are not normally distributed, the underlying EEG voltage used to calculate band power data is. Therefore, continuous EEG data were analyzed after the band power data computed by Bogart's (1999) CREW system were transformed into normally distributed micro Volt data. According to E. H. Bogart (personal communication, August 5, 1999), CREW's band power data may be transformed into micro Volt data by dividing each original band power data point by a conversion factor of 200 and then taking the square root of each resulting value.

In Figure 41, a sample of Subject 1's theta band power data (i.e., the theta band power data recorded at site Cz during the 3 min eyes open baseline condition) is graphically depicted in histogram format. The normal probability density function superimposed on these data reveals that the band power data are positively skewed. Figure 41 also depicts a histogram of these same

data after they were transformed into micro Volt data. The normal probability density function superimposed on the micro Volt data reveals that they are more normally distributed.

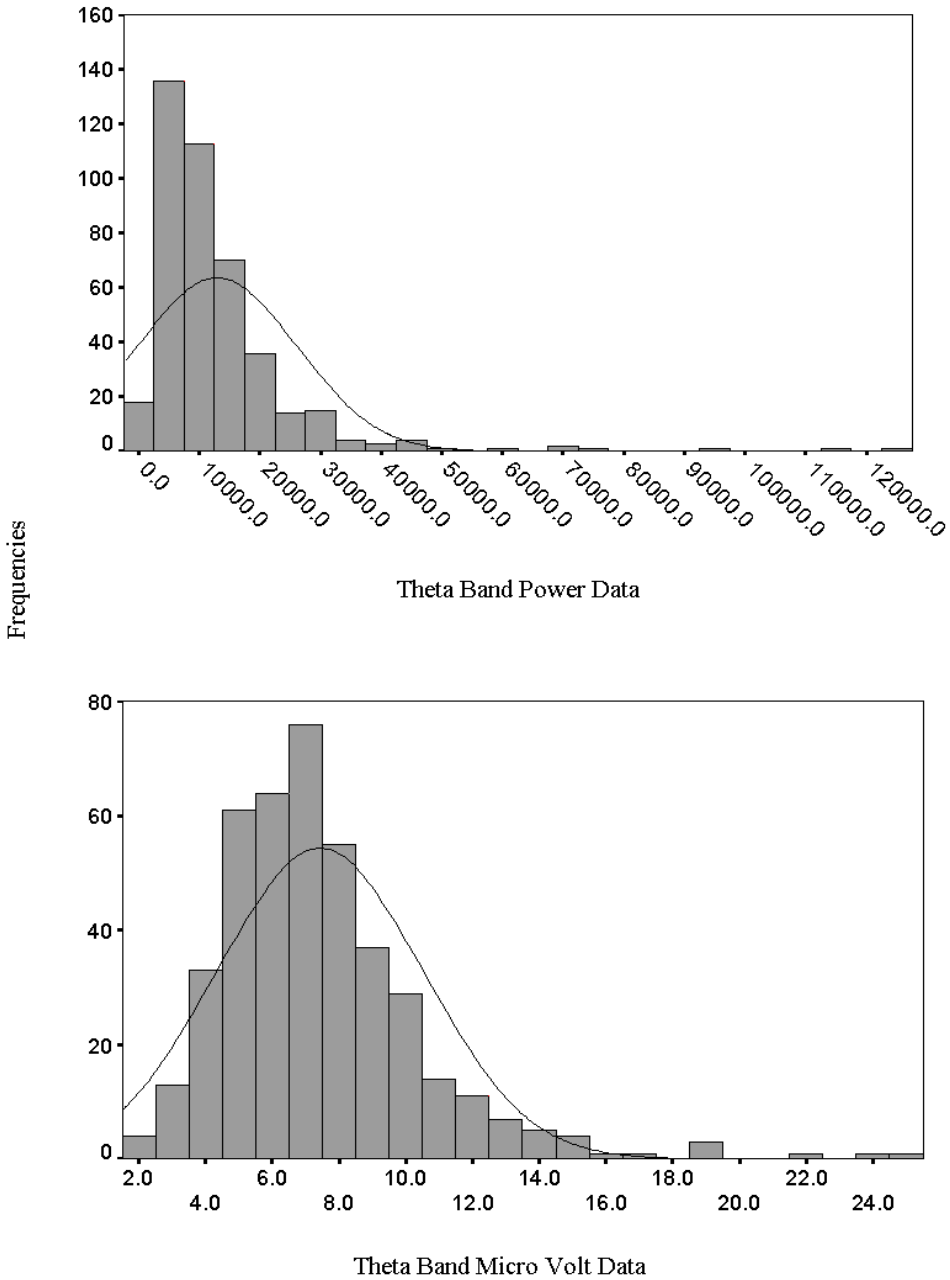


Figure 41. Distributions of Subject 1's theta band power data sample and theta band micro Volt data sample in comparison to normal distributions. (NOTE: The scale of measurement for theta band power data is analog-to-digital converter output units. Output units are related to inputs, or actual voltage at the scalp, via the following formula: actual voltage at the scalp^2 x 200.)

After each subject's band power data were transformed into micro Volt data, artifacts associated with eye blinks, eye and eye muscle movements, and facial and neck muscle movements were rejected. The following procedure was used to reject artifacts from each subject's EEG frequency band micro Volt data:

1. The mean activity associated with each frequency band (i.e., theta, alpha, and beta) recorded at each site (i.e., Cz, Pz, P3, and P4) was calculated for the data collected during the eyes open baseline condition. For example, the mean theta activity recorded at site Cz during the eyes open baseline was calculated; the mean theta activity recorded at site Pz during the eyes open baseline was calculated; etc.
2. Data associated with a given frequency band recorded at a given site were considered to be artifactual (and were rejected) if they fell outside ± 2 standard deviations about the mean calculated from those data.
3. Steps 1 and 2 were repeated using the data recorded during the eyes closed baseline condition; steps 1 and 2 were repeated using the data recorded during the automated condition of the tracking task; and steps 1 and 2 were repeated using the data recorded during the manual tracking task condition.

To further clarify this procedure, consider the distribution of the theta band micro Volt data sample that was depicted in Figure 41 (i.e., Subject 1's theta activity recorded at site Cz during the eyes open baseline condition). The mean activity associated with these data was calculated, and then data falling outside ± 2 standard deviations about the mean were rejected from the data file. The distributions of the original theta band micro Volt data sample and the theta band micro Volt data sample that remained after artifacts were rejected are graphically depicted in histogram

format in Figure 42. The normal probability density functions superimposed on the distributions highlight the normal distribution of the post-artifact rejection data.

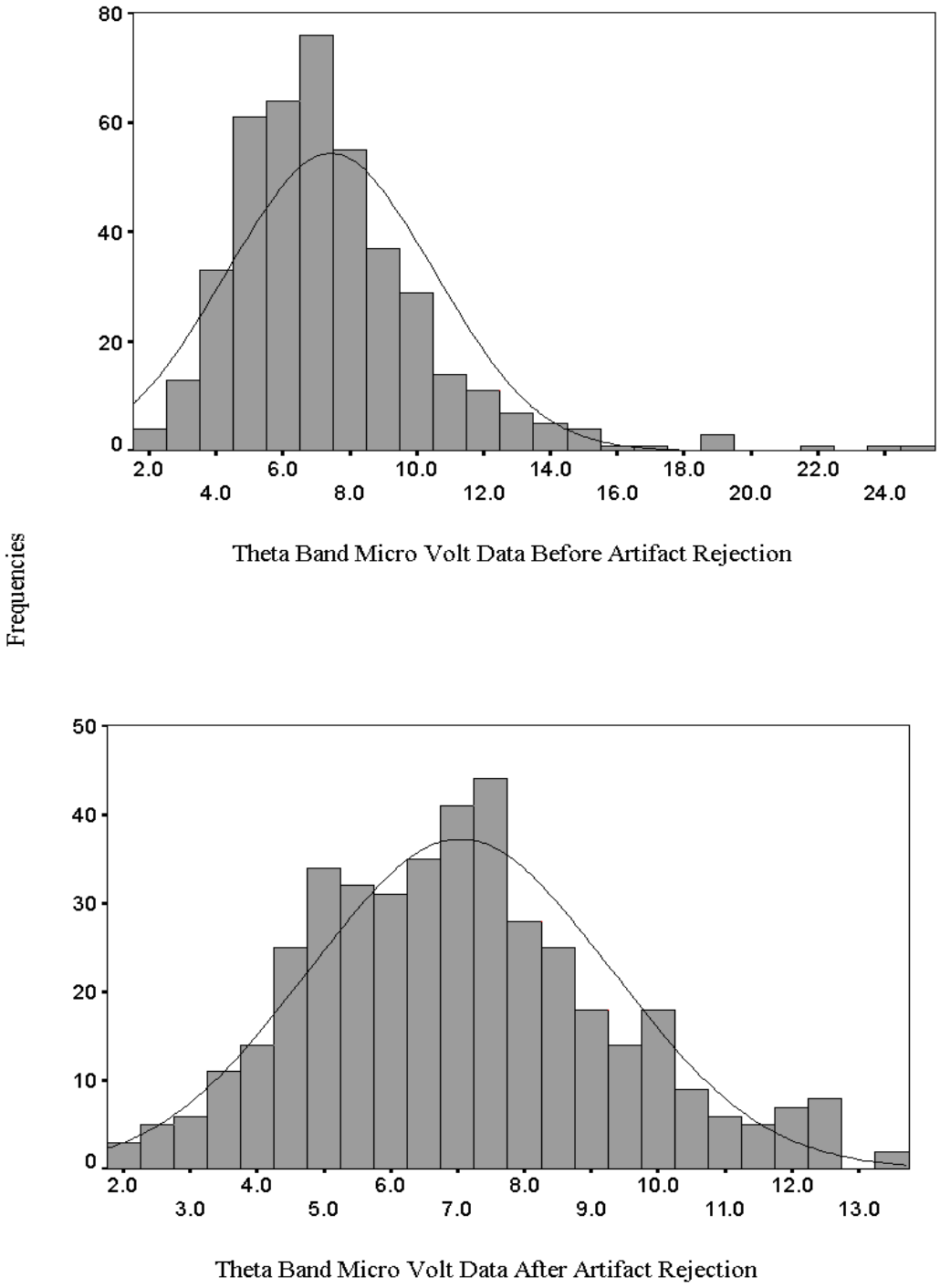


Figure 42. Distributions of Subject 1's theta band micro Volt data sample before and after artifact rejection in comparison to normal distributions.

As described previously, continuous EEG baseline data were recorded during a 3 min condition in which subjects sat quietly with their eyes open (i.e., an eyes open baseline condition) as well as during a 3 min condition in which subjects sat quietly with their eyes closed (i.e., an eyes closed baseline condition). These data were collected so that a general assessment of the overall behavior of the subjects' brain wave activity at sites Cz, Pz, P3, and P4 could be made.

A series of 2 (Baseline Condition) x 3 (EEG Frequency Band) ANOVAs were performed on the baseline data to determine if significant brain wave changes occurred during the eyes open and eyes closed conditions. Specifically, the mean activity recorded at site Cz was analyzed by way of a 2 x 3 ANOVA; the mean activity recorded at site Pz was analyzed by way of a second 2 x 3 ANOVA; the mean activity recorded at site P3 was analyzed by way of a third 2 x 3 ANOVA; and the mean activity recorded at site P4 was analyzed by way of a fourth 2 x 3 ANOVA. A single 2 (Baseline Condition) x 3 (EEG Frequency Band) x 4 (EEG Recording Site) ANOVA was not feasible because the EEG source that is strongest at one site will also be recorded to a lesser extent at contiguous sites. This means that EEG voltage variations at contiguous sites are highly correlated. In this investigation, the site to site correlations for data recorded within each of the three frequency bands at sites Cz, Pz, P3, and P4 ranged from 0.46 ($p < 0.05$) to 0.99 ($p < 0.05$); therefore, each site's frequency band data were analyzed using a separate ANOVA to avoid violating the ANOVA's assumption of independence.

As shown by the sample ANOVA Summary Table presented in Table 37, the baseline condition main effect, the frequency band main effect, and the Baseline Condition x Frequency Band interaction were investigated in each ANOVA. However, the interaction effect was of primary interest since attentional differences experienced by subjects during the two baseline

conditions may be identified by comparing the theta band activity occurring during the eyes open baseline condition with the theta band activity occurring during the eyes closed baseline condition; by comparing the alpha band activity occurring during the eyes open baseline condition with the alpha band activity occurring during the eyes closed baseline condition; and by comparing the beta band activity occurring during the eyes open baseline condition with the beta band activity occurring during the eyes closed baseline condition.

Table 37. Sample ANOVA Summary Table of EEG Data Recorded During Baseline Conditions

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	SS_{S}				
Within-Subject						
Baseline Condition (BC)	1	SS_{BC}	MS_{BC}	$MS_{BC} / MS_{BC \times S}$		
BC X S	11	$SS_{BC \times S}$	$MS_{BC \times S}$			
Frequency Band (FB)	2	SS_{FB}	MS_{FB}	$MS_{FB} / MS_{FB \times S}$		
FB X S	22	$SS_{FB \times S}$	$MS_{FB \times S}$			
BC X FB	2	$SS_{BC \times FB}$	$MS_{BC \times FB}$	$MS_{BC \times FB} / MS_{BC \times FB \times S}$		
BC X FB X S	22	$SS_{BC \times FB \times S}$	$MS_{BC \times FB \times S}$			
Total	71	SS_{total}				

The complete ANOVA Summary Tables and Bonferroni t-Test Summary Tables that accompany the analyses of the baseline EEG data are located in Appendix O. A Mauchley's test of sphericity was performed in conjunction with each ANOVA, and, when necessary, Greenhouse-Geisser Epsilon values were used to correct the problems associated with positively biased F-Tests. Additionally, two Bonferroni t-Test post-hoc analyses were also performed in conjunction with each ANOVA. These analyses were appropriate for evaluating a series of post-hoc comparisons while controlling for inflated alpha error.

As shown by the tables included in Appendix O, analyses of the theta, alpha, and beta frequency band data recorded at sites Cz, Pz, P3, and P4 revealed that significantly more EEG activity occurred during the eyes closed baseline condition than during the eyes open baseline condition (i.e., the baseline condition main effect was significant at $p < 0.05$ in all instances); that significantly more alpha activity and theta activity occurred than beta activity (i.e., the frequency band main effect was significant at $p < 0.05$ in all instances); and that more alpha activity occurred during the eyes closed baseline condition than during the eyes open baseline condition (i.e., the Baseline Condition x Frequency Band interaction was significant at $p < 0.05$ in all instances). The Bonferroni t-Test post hoc analyses revealed that other significant differences occurred between the activity levels of various frequency bands recorded during the two baseline conditions. But, as stated previously, differences between theta activity during the eyes open and eyes closed baseline conditions, differences between alpha activity during the eyes open and eyes closed baseline conditions, and differences between beta activity during the eyes open and eyes closed baseline conditions were of primary interest.

Means and 95% confidence intervals (CI) calculated for: 1) the combined theta, alpha, and beta activity recorded at each site during each of the baseline conditions; 2) the activity associated with each frequency band recorded at each site; and 3) the activity associated with each frequency band recorded at each site during each baseline condition are shown in Figures 43 - 48. Figure 43 depicts the mean activity recorded at each site during the eyes open and eyes closed baseline conditions when activity was averaged across the three frequency bands (i.e., N = 36). Figure 44 depicts the mean activity associated with each frequency band recorded at each site when activity was averaged across the two baseline conditions (i.e., N = 24). Figures 45 - 48 depict the mean activity associated with each frequency band recorded at each site during each baseline condition (i.e., N = 12).

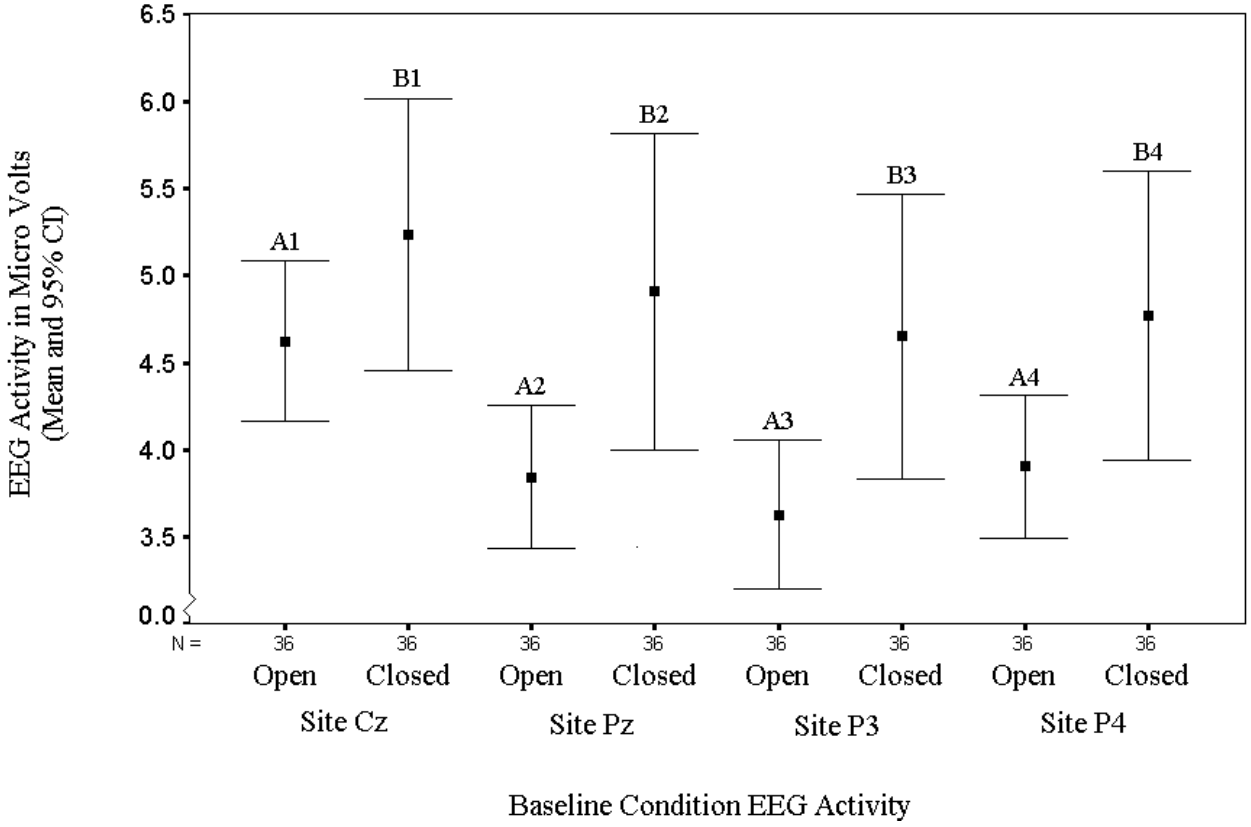


Figure 43. Combined theta, alpha, and beta activity recorded at each site during each baseline condition (i.e., eyes open baseline condition versus eyes closed baseline condition). (NOTE: Means with different letters are significantly different in an ANOVA at $p < 0.05$. Numbers associated with each grouping of letters indicate the results of individual analyses since the mean EEG activity recorded at each site was analyzed by way of a separate ANOVA.)

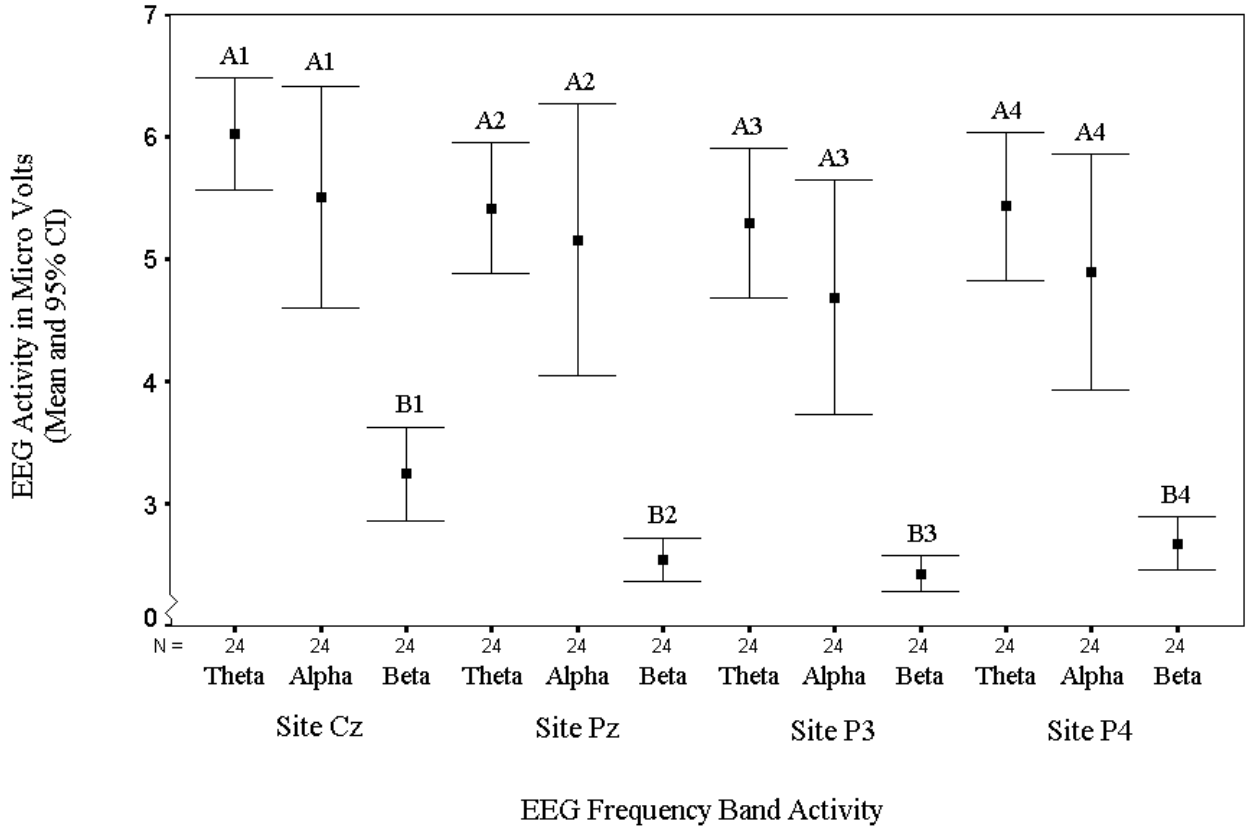


Figure 44. Baseline EEG activity recorded at each site. (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$. Numbers associated with each grouping of letters indicate the results of individual post-hoc analyses since the mean EEG activity recorded at each site was analyzed by way of a separate ANOVA.)

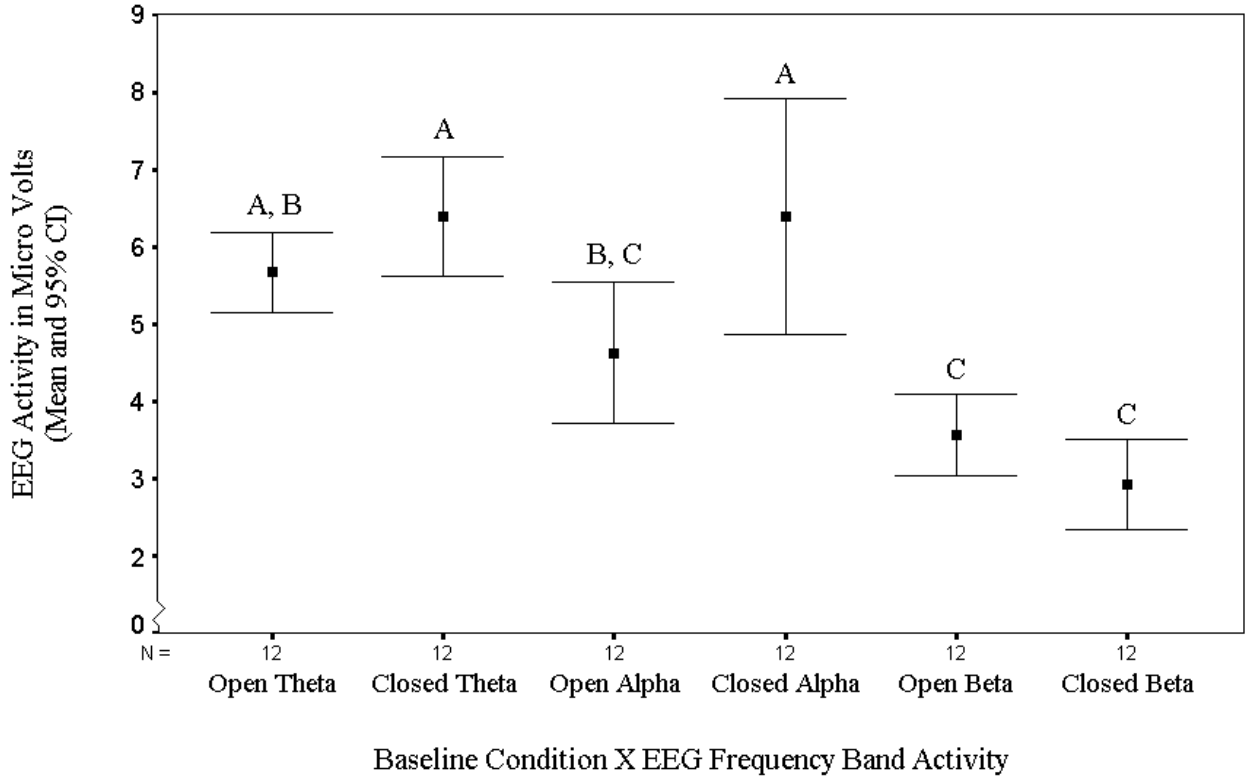


Figure 45. Theta, alpha, and beta activity recorded at site Cz during each baseline condition (i.e., eyes open baseline condition versus eyes closed baseline condition). (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

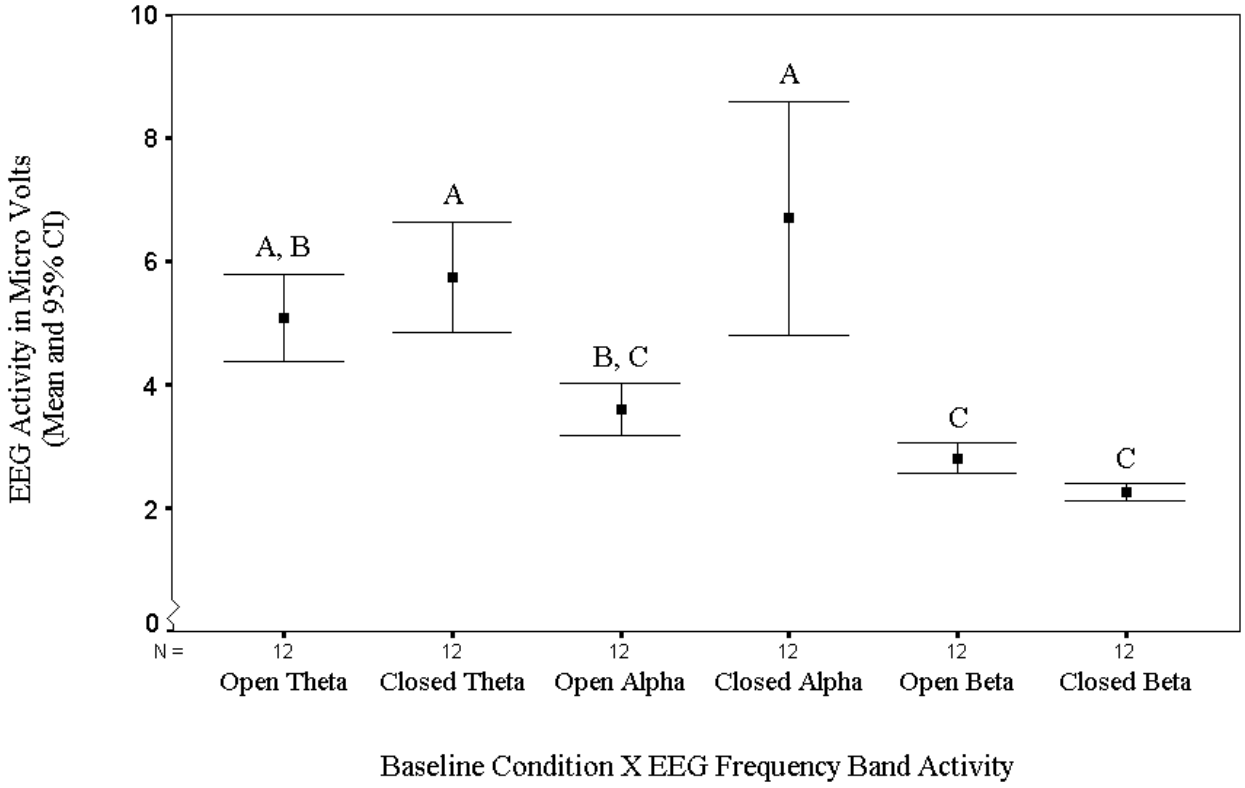


Figure 46. Theta, alpha, and beta activity recorded at site Pz during each baseline condition (i.e., eyes open baseline condition versus eyes closed baseline condition). (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

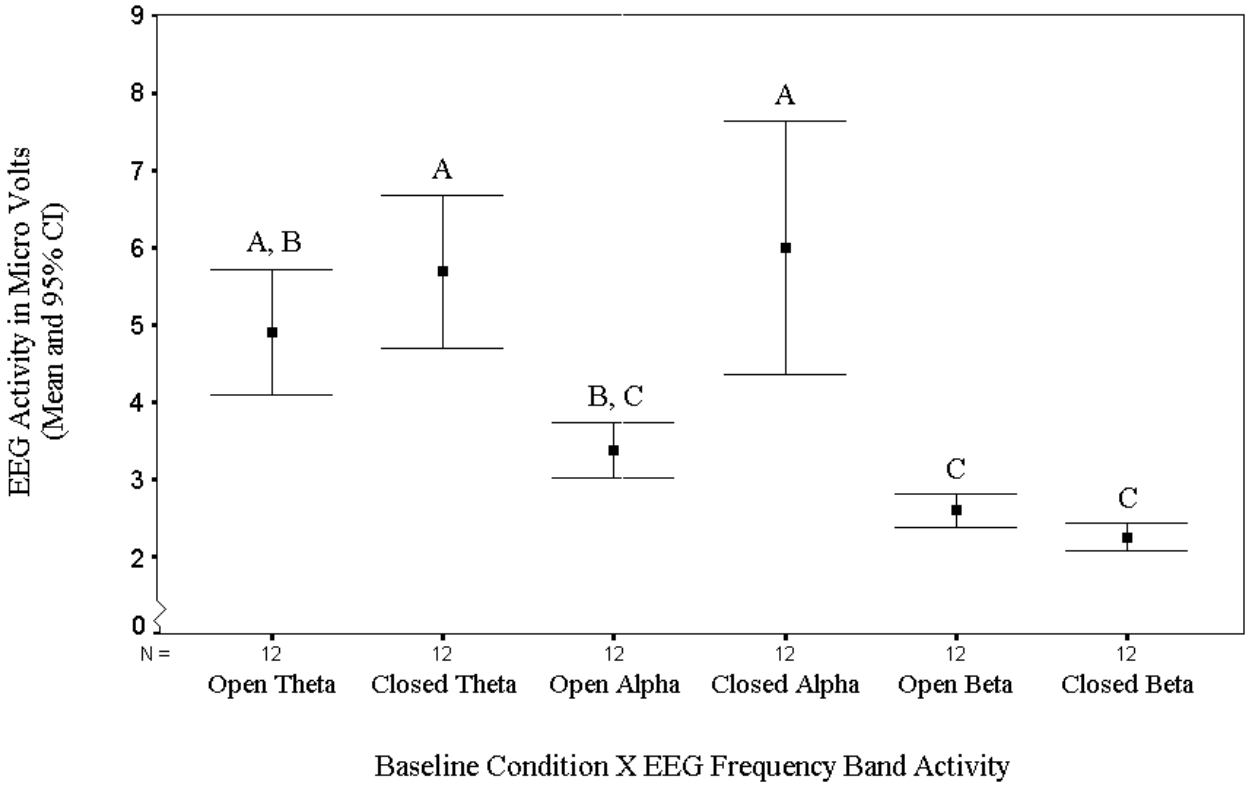


Figure 47. Theta, alpha, and beta activity recorded at site P3 during each baseline condition (i.e., eyes open baseline condition versus eyes closed baseline condition). (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

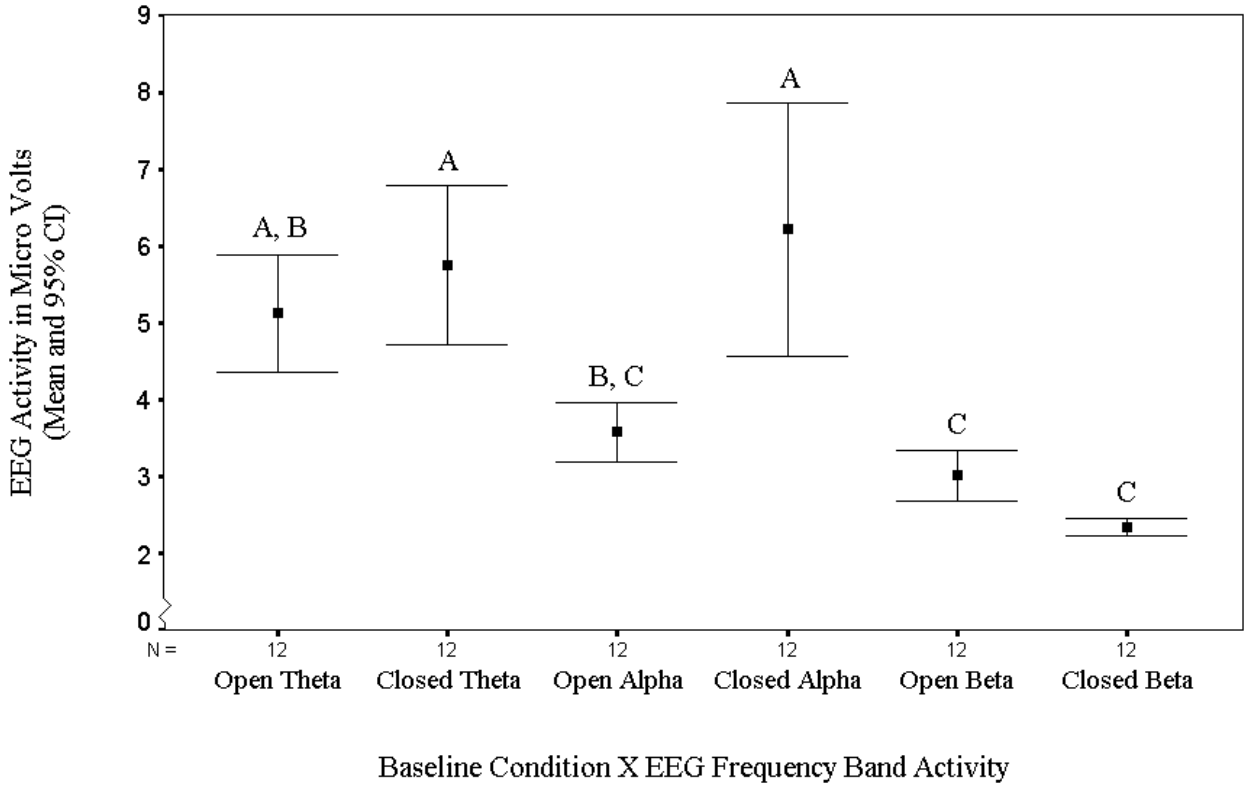


Figure 48. Theta, alpha, and beta activity recorded at site P4 during each baseline condition (i.e., eyes open baseline condition versus eyes closed baseline condition). (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

With respect to the main effect of frequency band, it is suggested that more alpha and theta activity occurred than beta activity because subjects were merely required to sit quietly during the baseline conditions. Since subjects were not engaged in a task during either baseline condition, they exhibited brain wave activity associated with a relaxed state of diminished arousal (i.e., relatively high levels of alpha and theta activity and relatively little beta activity) during both baseline conditions. With respect to the main effect of baseline condition, it is suggested that more activity was recorded during the eyes closed baseline than during the eyes open baseline because alpha and theta activity increased and beta activity decreased when visual stimulation from external surroundings was eliminated. With respect to the Baseline Condition x Frequency Band interaction, it is suggested than more alpha activity occurred during the eyes closed baseline condition than during the eyes open baseline condition because subjects were able to achieve significantly higher levels of relaxation after closing their eyes. Overall, the behavior of subjects' brain waves were consistent with general expectations.

Thirty minutes of continuous EEG data were recorded from each subject during the automated condition of the MAT Battery's tracking task, and 30 min of continuous EEG data were recorded from each subject during the manual condition of the MAT Battery's tracking task. These data were collected to determine if significant changes in brain wave activity occurred at sites Cz, Pz, P3, and P4 in response to each tracking task condition.

A series of 2 (Tracking Condition) x 3 (EEG Frequency Band) ANOVAs were performed on the continuous EEG data to determine if significant brain wave changes occurred during the two conditions of the tracking task. The mean activity recorded at site Cz was analyzed by way of a 2 x 3 ANOVA; the mean activity recorded at site Pz was analyzed by way of a second 2 x 3

ANOVA; the mean activity recorded at site P3 was analyzed by way of a third 2 x 3 ANOVA; and the mean activity recorded at site P4 was analyzed by way of a fourth 2 x 3 ANOVA.

As shown by the sample ANOVA Summary Table presented in Table 38, the tracking condition main effect, the frequency band main effect, and the Tracking Condition x Frequency Band interaction were investigated in each ANOVA. However, the interaction effect was of primary interest since attentional differences experienced by subjects during the two tracking task conditions may be identified by comparing the theta band activity occurring during automated tracking with the theta band activity occurring during manual tracking; by comparing the alpha band activity occurring during automated tracking with the alpha band activity occurring during manual tracking; and by comparing the beta band activity occurring during automated tracking with the beta band activity occurring during manual tracking.

Table 38. Sample ANOVA Summary Table of EEG Data Recorded During Tracking Task Conditions

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	SS_{S}				
Within-Subject						
Tracking Condition (TC)	1	SS_{TC}	MS_{TC}	$MS_{TC} / MS_{TC \times S}$		
TC X S	11	$SS_{TC \times S}$	$MS_{TC \times S}$			
Frequency Band (FB)	2	SS_{FB}	MS_{FB}	$MS_{FB} / MS_{FB \times S}$		
FB X S	22	$SS_{FB \times S}$	$MS_{FB \times S}$			
TC X FB	2	$SS_{TC \times FB}$	$MS_{TC \times FB}$	$MS_{TC \times FB} / MS_{TC \times FB \times S}$		
TC X FB X S	22	$SS_{TC \times FB \times S}$	$MS_{TC \times FB \times S}$			
Total	71	SS_{total}				

The complete ANOVA Summary Tables and Bonferroni t-Test Summary Tables that accompany the analyses of the EEG data recorded during the tracking tasks are located in Appendix P. A Mauchley's test of sphericity was performed in conjunction with each ANOVA, and, when necessary, Greenhouse-Geisser Epsilon values were used to correct the problems associated with positively biased F-Tests. Additionally, two Bonferroni t-Test post-hoc analyses were also performed in conjunction with each ANOVA. These analyses were appropriate for evaluating a series of post-hoc comparisons while controlling for inflated alpha error.

As shown by the tables included in Appendix P, analyses of the theta, alpha, and beta frequency band data recorded at sites Cz, Pz, P3, and P4 revealed that significantly more EEG activity occurred during the automated tracking task condition than during the manual tracking condition (i.e., the tracking condition main effect was significant at $p < 0.05$ in all instances) and that significantly more theta activity occurred than alpha activity and beta activity (i.e., the frequency band main effect was significant at $p < 0.05$ in all instances). Data recorded at sites Pz, P3, and P4 revealed that more theta activity, more alpha activity, and less beta activity occurred during the automated tracking task condition than during the manual tracking task condition (i.e., the Tracking Condition x Frequency Band interaction was significant at $p < 0.05$ in all instances). However, data recorded at site Cz only revealed that more theta activity occurred during the automated tracking task condition than during the manual tracking task condition. The Bonferroni t-Test post hoc analyses revealed that other significant differences occurred between the activity levels of various frequency bands recorded during the two tracking conditions. But, as stated previously, differences between theta activity occurring during the automated and manual tracking task conditions, differences between alpha activity occurring during the

automated and manual tracking task conditions, and differences between beta activity occurring during the automated and manual tracking task conditions were of primary interest.

Means and 95% confidence intervals (CI) calculated for: 1) the combined theta, alpha, and beta activity recorded at each site during each of the tracking task conditions; 2) the activity associated with each frequency band recorded at each site; and 3) the activity associated with each frequency band recorded at each site during each tracking task condition are shown in Figures 49 - 54. Figure 49 depicts the mean activity recorded at each site during the automated and manual tracking task conditions when activity was averaged across the three frequency bands (i.e., N = 36). Figure 50 depicts the mean activity associated with each frequency band recorded at each site when activity was averaged across the two tracking task conditions (i.e., N = 24). Figure 51 - 54 depict the mean activity associated with each frequency band recorded at each site during each tracking task condition (i.e., N = 12).

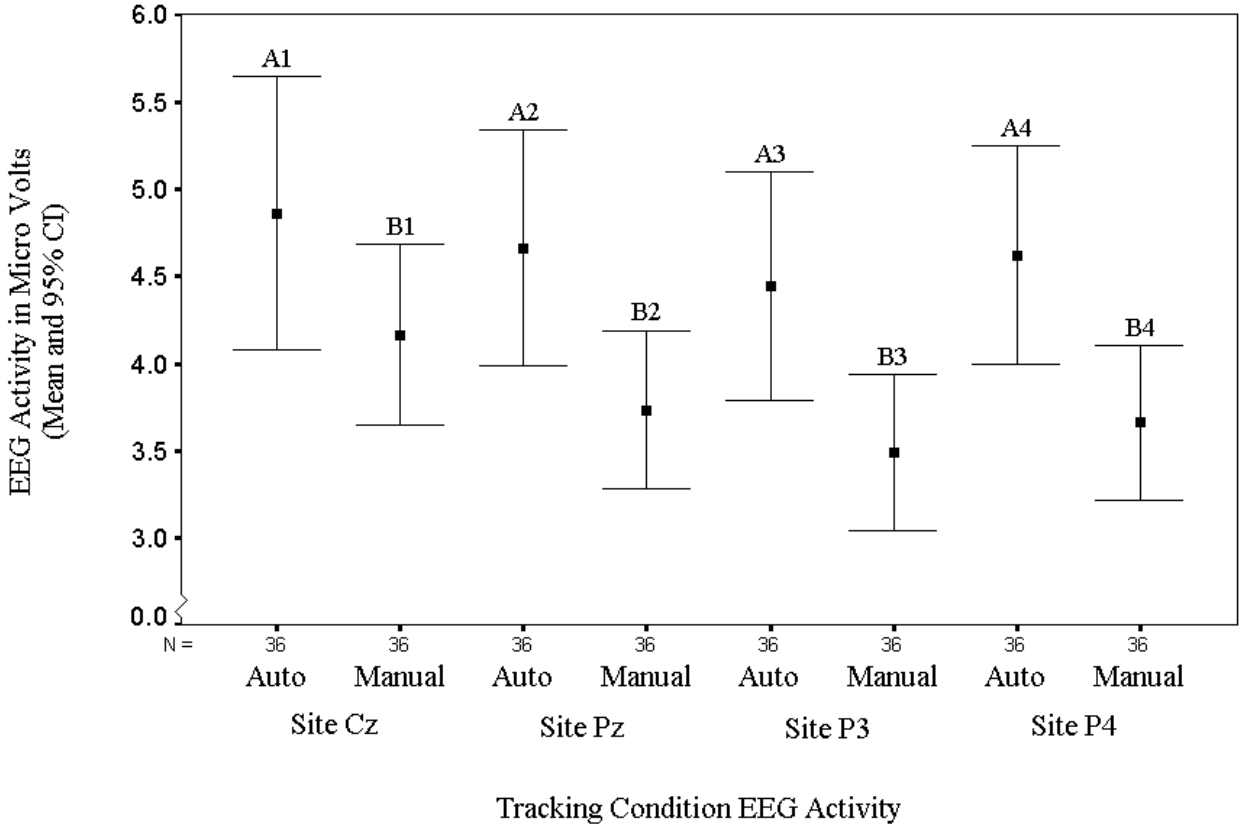


Figure 49. Combined theta, alpha, and beta activity recorded at each site during each tracking task condition. (NOTE: Means with different letters are significantly different in an ANOVA at $p < 0.05$. Numbers associated with each grouping of letters indicate the results of individual analyses since the mean EEG activity recorded at each site was analyzed by way of a separate ANOVA.)

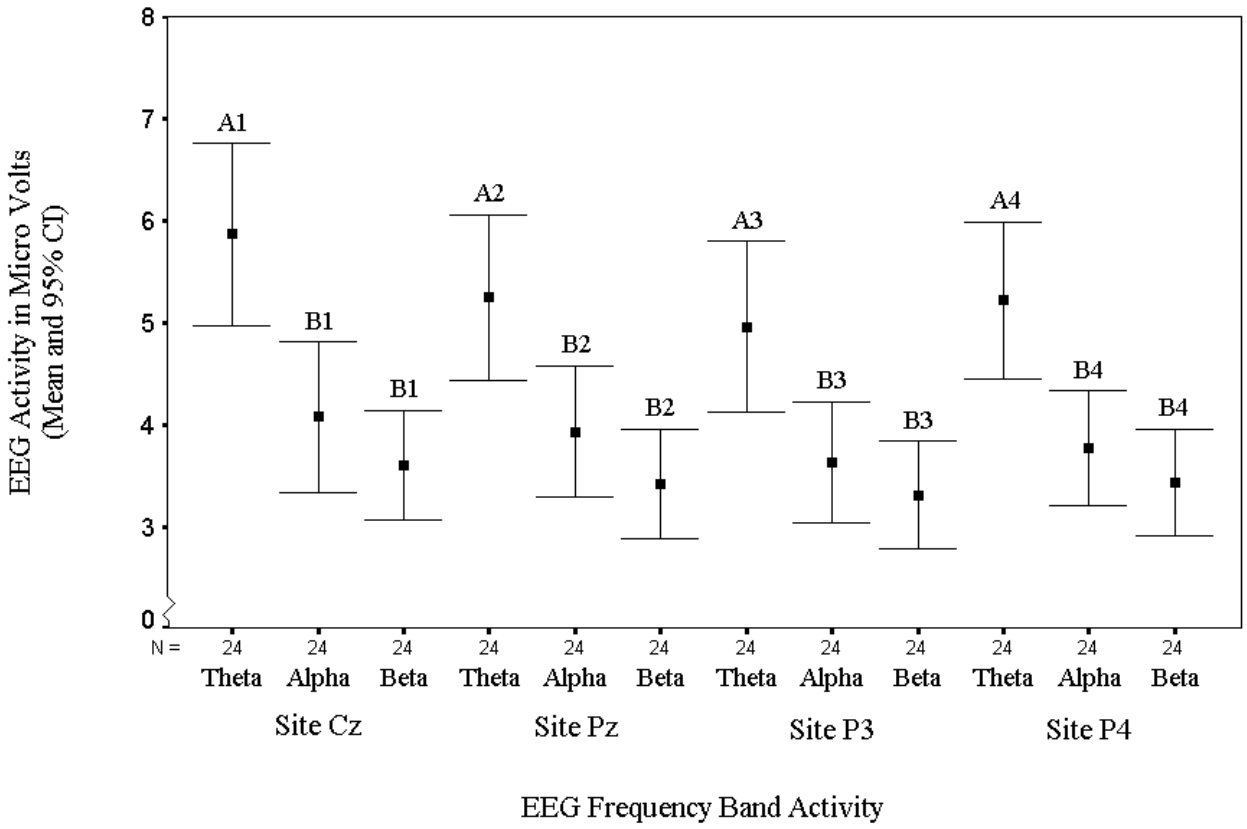


Figure 50. EEG activity recorded at each site during tracking tasks. (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$. Numbers associated with each grouping of letters indicate the results of individual post-hoc analyses since the mean EEG activity recorded at each site was analyzed by way of a separate ANOVA.)

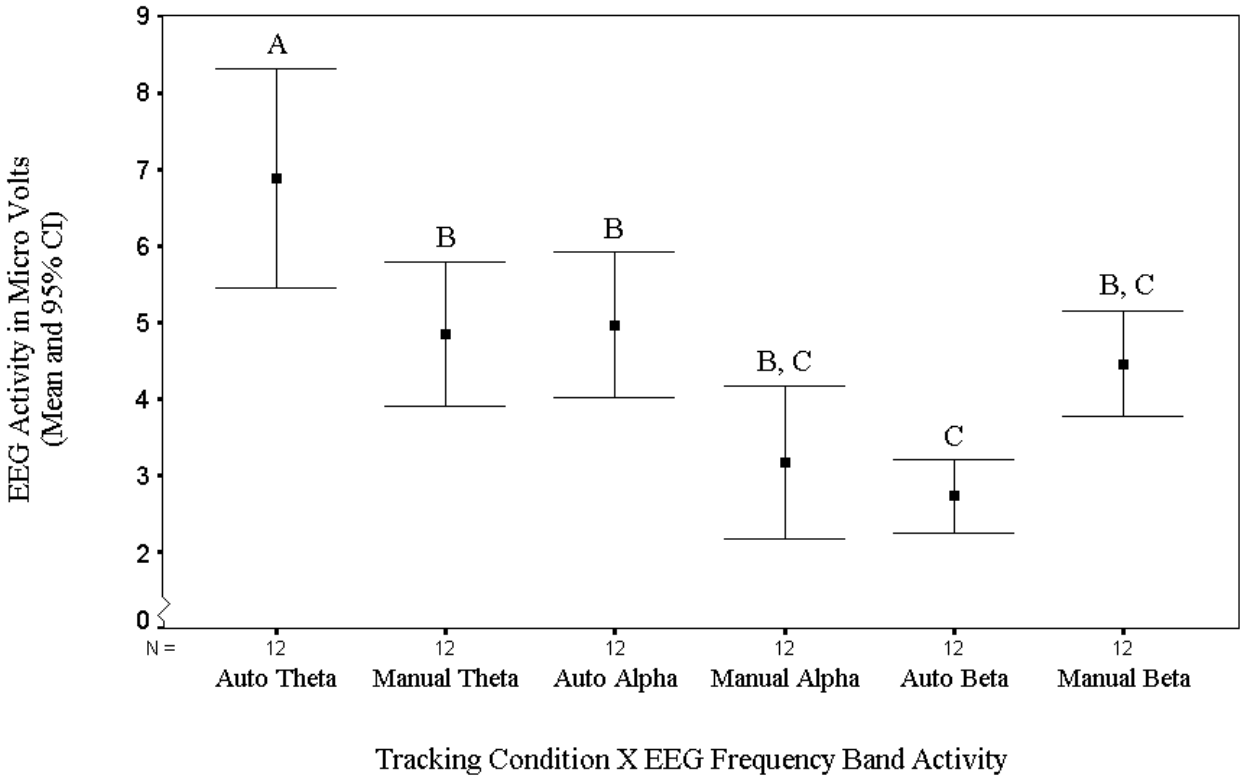


Figure 51. Theta, alpha, and beta activity recorded at site Cz during each tracking task condition. (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

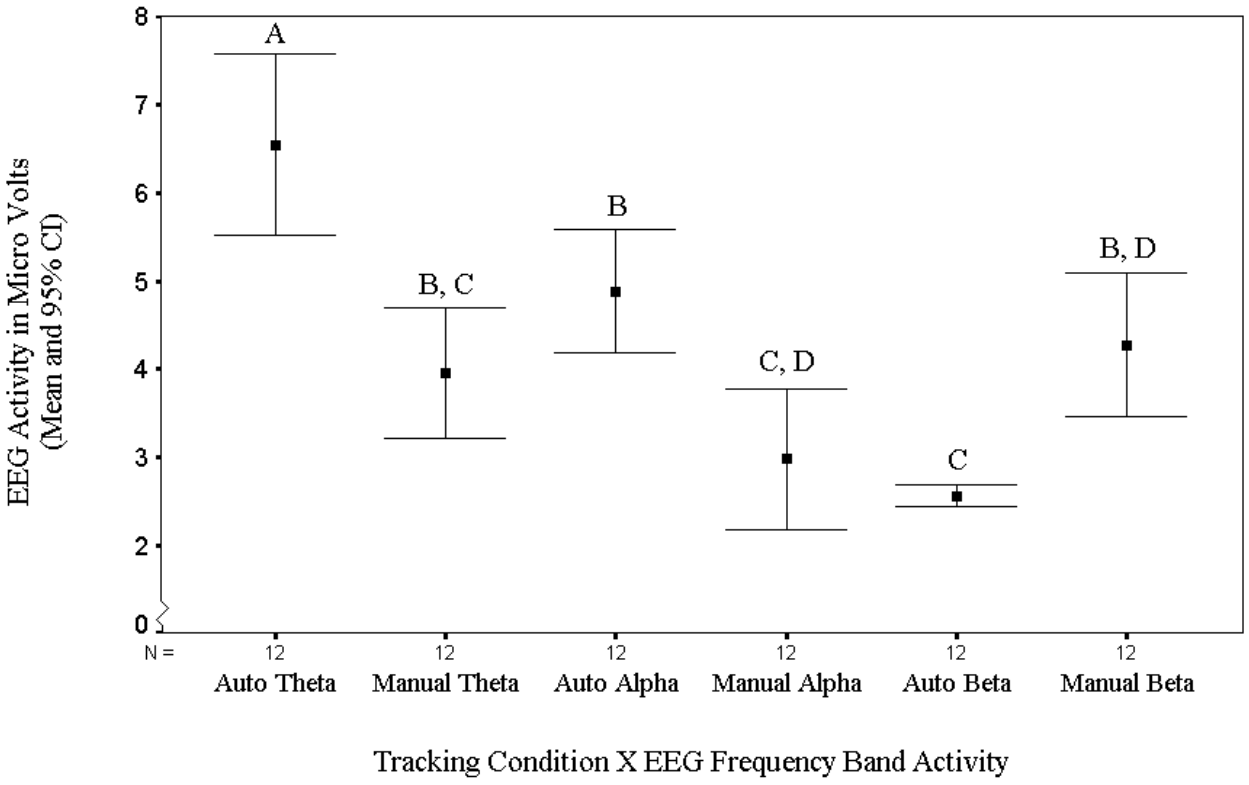


Figure 52. Theta, alpha, and beta activity recorded at site Pz during each tracking task condition. (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

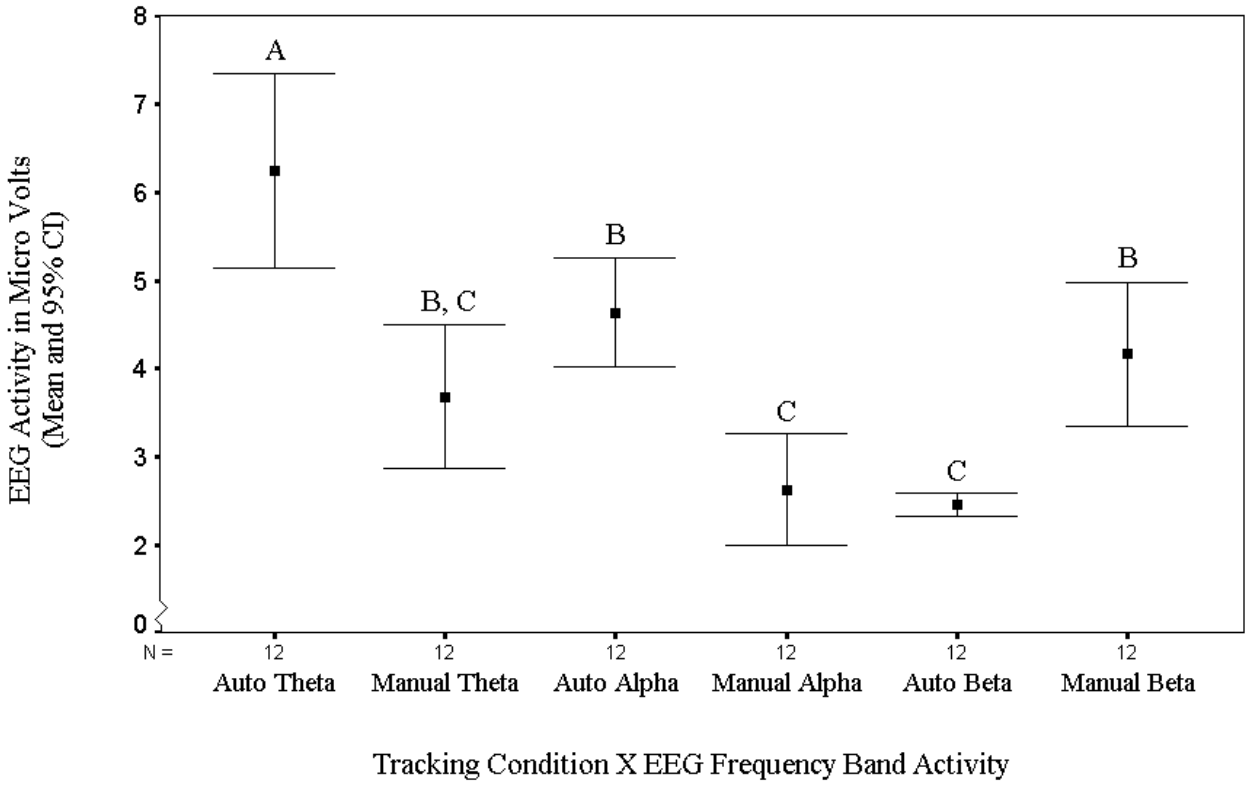


Figure 53. Theta, alpha, and beta activity recorded at site P3 during each tracking task condition. (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

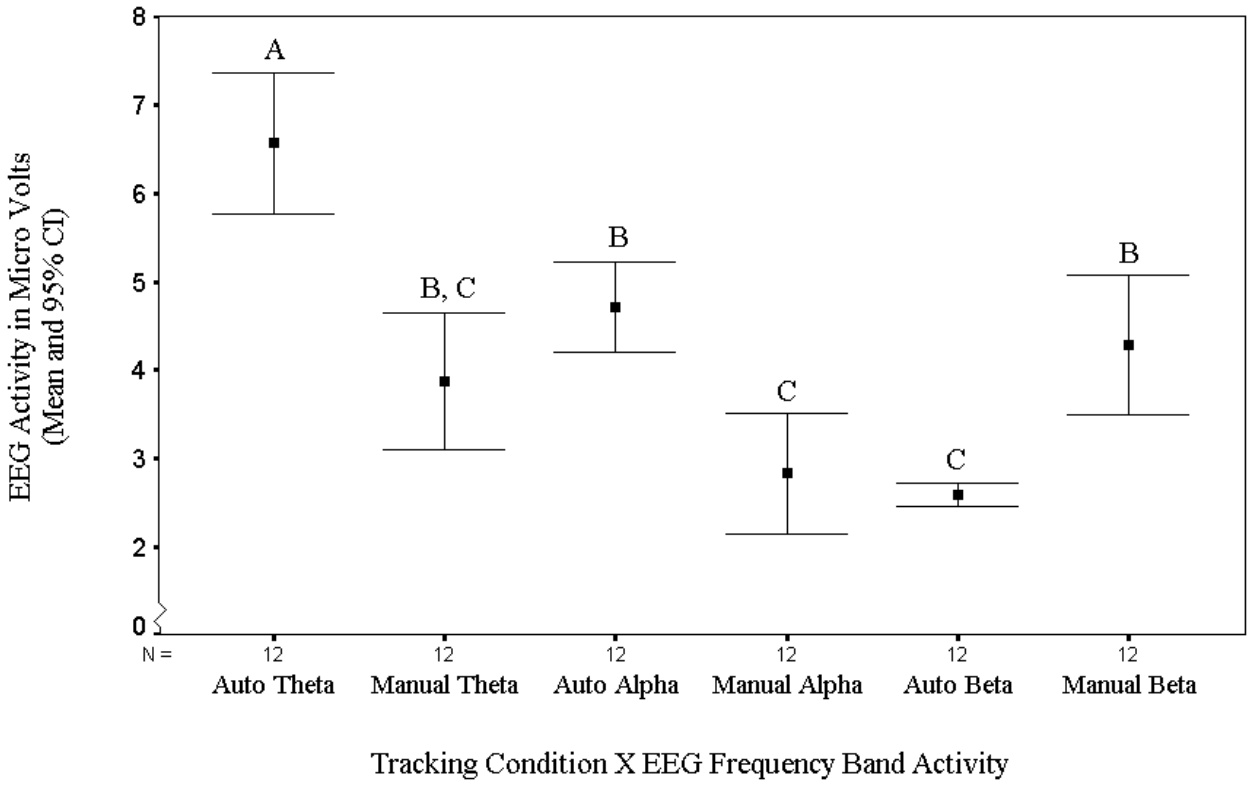


Figure 54. Theta, alpha, and beta activity recorded at site P4 during each tracking task condition. (NOTE: Means with different letters are significantly different in a Bonferroni t-Test post-hoc analysis at $p < 0.05$.)

With respect to the main effect of frequency band, evidence suggesting that more theta activity occurred than alpha activity and beta activity may indicate that subjects exhibited brain wave activity associated with a diminished state of arousal during both tracking conditions. However, if this were the case, a relatively high level of alpha activity in comparison to beta activity would also be expected. Therefore, these results are somewhat difficult to explain. Since alpha and beta activity levels were not significantly different from one another, these data may indicate that subjects experienced a level of attentional engagement that was "balanced" between a state of relaxed wakefulness and a state of alertness. With respect to the main effect of tracking condition, it is suggested that more activity in the three frequency bands examined was recorded during the automated condition of the tracking task than during the manual condition of the tracking task because the highest levels of theta activity as well as the highest levels of alpha activity occurred in conjunction with automated tracking as a result of the automated tracking task condition's relatively low task demands.

With respect to the Tracking Condition x Frequency Band interaction found at sites Pz, P3, and P4, it is suggested that lower levels of theta activity, lower levels of alpha activity, and higher levels of beta activity were exhibited during manual tracking because the manual condition of the tracking task required that subjects use a joystick to keep an airplane symbol within the center rectangular area of a tracking task; listen carefully for aural alerting signals; and verbally identify the flight deck function and urgency level corresponding to each of 12 aural alerts as quickly as possible. The automated condition of the tracking task required that subjects listen carefully for aural alerting signals and verbally identify the flight deck function and urgency level corresponding to each of 12 aural alert as quickly as possible, but subjects were not required to

perform manual tracking. Since visually monitoring a computer's tracking performance represents a relatively lower level of workload, subjects experienced a reduced level of attentional engagement during automated tracking.

With respect to the Tracking Condition x Frequency Band interaction found at site Cz, evidence suggesting that a higher level of theta activity was exhibited during the automated condition of the tracking task than during the manual condition of the tracking task indicates that a lower level of attentional engagement was experienced during automated tracking. As with the interaction effect found at sites Pz, P3, and P4, these results suggest that a reduced level of attentional engagement occurred during automated tracking as a result of relatively lower task demands. Neither a significant difference between alpha activity recorded during automated versus manual tracking nor a significant difference between beta activity recorded during automated versus manual tracking was found to exist. However as shown in Figure 51, even though the differences were not statistically significant, higher levels of alpha activity and lower levels of beta activity were exhibited by subjects during the automated condition of the tracking task.

The results associated with data recorded at sites Pz, P3, and P4 support the experimental hypothesis that more beta activity, less theta activity, and less alpha activity would occur during the performance of the manual tracking task condition and that more theta activity, more alpha activity, and less beta activity would occur during the performance of the automated tracking task condition. The results of data recorded at site Cz partially support this same hypothesis since more theta activity was found to have occurred during the performance of the automated tracking task condition and less theta activity was found to have occurred during the performance of the

manual tracking task condition. It is suggested, therefore, that continuous EEG data indicate that subjects experienced a higher level of attentional engagement during the manual condition of the tracking task than during the automated condition of the tracking task.

GENERAL DISCUSSION

The first hypothesis under test was that systematic manipulations of a single aural alert's tempo could be used to convey low, moderate, and high levels of urgency. In this study, subjects provided significantly different urgency ratings for the low urgency level alerts, the moderate urgency level alerts, and the high urgency levels. Furthermore, subjects correctly identified low, moderate, and high urgency levels 97.92% of the time and identified the low, moderate, and high urgency levels equally often during the automated and manual conditions of the tracking task. These results indicate that systematic manipulations of a single aural alert's tempo can be used to convey low, moderate, and high urgency levels which can be accurately perceived even when a task requiring two levels of workload and attentional engagement is performed. Additionally, these results also indicate that systematic decreases in the tempos of Burt's (1996) low urgency level and moderate urgency level alerts represent appropriate manipulations through which to differentiate the urgency ratings of moderate urgency level and high urgency level alerts while maintaining differences between the urgency ratings of low urgency level and moderate urgency level alerts. Therefore, it does not appear that any additional manipulations are needed to convey low, moderate, and high levels of urgency within the alerting sets investigated in this study.

The second hypothesis under test was that a composite manipulation of aural alerts' fundamental frequency, pitch range, rhythmic pattern, and pitch contour could be used to

minimize the overall urgency level differences between distinctive alerting sets. In this study, subjects provided the same urgency rating for Set I, Set II, Set III, and Set IV when ratings were averaged across the three urgency levels. Therefore, it does not appear that any additional manipulations of acoustic parameters are needed to equate the urgency levels of the alerting sets investigated in this study.

The third hypothesis under test was that a composite manipulation of aural alerts' fundamental frequency, pitch range, rhythmic pattern, pitch contour, and tempo could be used to create signals within distinctive alerting sets that convey equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency. In this study, subjects provided the same urgency rating for Set I Low #1, Set II Low #1, Set III Low #1, and Set IV Low #1; provided the same urgency rating for Set I Low #2, Set II Low #2, Set III Low #2, and Set IV Low #2; provided the same urgency rating for Set I Moderate #1, Set II Moderate #1, Set III Moderate #1, and Set IV Moderate #1; provided the same urgency rating for Set I Moderate #2, Set II Moderate #2, Set III Moderate #2, and Set IV Moderate #2; and provided the same urgency rating for Set I High, Set II High, Set III High, and Set IV High. As a result of these urgency ratings, it does not appear that any additional manipulations of acoustic parameters are needed to equate the alerting sets' low urgency level alerts, the alerting sets' moderate urgency level alerts, and the alerting sets' high urgency level alerts.

The fourth hypothesis under test was that subjects would be able to: 1) associate each alerting set with one of the four major flight deck functions; 2) identify the correct aural alerting set, the correct urgency level, as well as the correct alerting set <u>and</u> urgency level associated with each alert 95% of the time; and 3) correctly identify each alerting set, urgency level, and aural

alert equally often while performing the MAT Battery's tracking task in automatic and manual modes. In this study, subjects chose the correct alerting set in 97.92% of the trials occurring during automated tracking and in 97.57% of the trials occurring during manual tracking. Subjects chose the correct urgency level in 97.92% of the trials occurring during automated tracking and in 97.92% of the trials occurring during manual tracking. Subjects chose the correct alerting set and the correct urgency level in 96.53% of the trials occurring during automated tracking and in 95.83% of the trials occurring during manual tracking. Additionally, subjects correctly identified each alerting set, urgency level, and aural alert equally often during each tracking task condition. Thus, it appears that subjects successfully recognized and were able to identify the three alerts associated with each alerting set when a task requiring two levels of workload and attentional engagement was performed.

The final hypothesis under test was that subjects would experience higher levels of workload and attentional engagement while performing the manual condition of the MAT Battery's tracking task. In this study, subjects provided higher NASA TLX mean weighted workload ratings for the manual condition of the tracking task than the automated condition of the tracking task. Furthermore, subjects exhibited brain wave activity associated with a higher level of attentional engagement (i.e., more beta activity, less theta activity, and less alpha activity) during the manual condition of the tracking than during the automated condition of the tracking task. Therefore, it appears that workload and attentional engagement levels experienced by subjects participating in this study were effectively manipulated through the use of the MAT Battery's automated and manual tracking task conditions.

CONCLUSIONS

As a result of the current research endeavor, several advances were made toward the design and development of an aural alerting signal categorization scheme in which one distinctive alert was associated with each of the four major flight deck functions, and the acoustic parameters of a given alert were manipulated to form an alerting set capable of conveying low, moderate, and high levels of urgency. First, validation of the claim that the four aural alerting sets were distinctive from one another was supported by sound identification data which revealed that subjects identified the correct alerting set to which an alert corresponded 97.74% of the time while performing a task requiring two levels of workload and attentional engagement. Second, acoustic parameters capable of equating or prioritizing the overall urgency levels of the alerting sets were identified. Third, a composite manipulation of acoustic parameters capable of equating the alerting sets' low urgency level alerts, the alerting sets' moderate urgency level alerts, and the alerting sets' high urgency level alerts was identified. Fourth, tempo manipulations capable of conveying three levels of urgency within each alerting set were identified. Fifth, the ability of subjects to identify the correct alerting set and urgency level associated with each alert when a task that required two different levels of workload and attentional engagement was performed was demonstrated.

In general, this research suggests that acoustic parameter manipulations can be used to create distinctive alerting sets that each convey multiple levels of urgency. As shown in Table 39, the distinctiveness of the four aural alerting sets investigated during the current research endeavor was achieved through rhythmic pattern and pitch contour manipulations and, to a lesser degree, through fundamental frequency and pitch range manipulations. The overall urgency level of each

set was equated through fundamental frequency and pitch range manipulations and, to a lesser degree, through rhythmic pattern and pitch contour manipulations. Three levels of urgency were conveyed within each alerting set through tempo manipulations. The final result of four distinctive alerting sets capable of conveying equivalent levels of low urgency, equivalent levels of moderate urgency, and equivalent levels of high urgency was achieved through a composite manipulation of rhythmic pattern, pitch contour, fundamental frequency, pitch range, and tempo.

Table 39. Summary of Acoustic Parameter Manipulations Associated with Proposed Aural Alerting Signal Categorization Scheme

Aural Alerting Set Characteristics	Acoustic Parameter Manipulations				
	Rhythmic Pattern	Pitch Contour	Fundamental Frequency	Pitch Range	Tempo
Distinctiveness	X	X	x	x	
Equivalent Overall Urgency Levels	x	x	X	X	
Within Set Urgency Levels					X
Equivalent Within Set Urgency Levels Across Sets	x	x	x	x	x

Key: X Primary contributor
x Contributor

It is suggested that rhythmic pattern be used to design distinctive alerts because, as explained earlier, rhythm creates an "overall pattern" of sound into which melodic and harmonizing tones may be added to produce a unique aural signal (Lieberman, 1951, p.1). Pitch contour manipulations are also recommended in the design of distinctive alerts based on evidence suggesting that pitch contour is a critical component in the recognition of melodic sequences especially when such sequences are heard in a non-musical context (Edworthy, 1985). Next, it is suggested that fundamental frequency and pitch range be manipulated according to the guidelines provided by Edworthy (1991, 1994b) to equate or prioritize the urgency levels of distinctive alerts as well as to add further distinguishing acoustic characteristics. Finally, the tempo of each distinctive alert can be manipulated to convey different levels of perceived urgency. Hellier et al.'s (1991) guidelines may be used to this end, but as demonstrated by the work of Burt (1996), additional tempo manipulations may be required to convey the desired levels of urgency.

Designers should realize that limiting the number of alerting sets (i.e., discrete sounds) to three to five will help ensure the effectiveness of the aural alerting signal categorization scheme. Additionally, it should also be realized that time constraints governing the total duration of aural alerts may limit the feasibility of using tempo manipulations to convey more than three levels of urgency within a given alerting set.

Practical Implications

The aural alerting signal categorization scheme investigated by the current research endeavor may be used in any environment where multiple aural alerts are presented (e.g., aircraft flight decks, hospital intensive care units and/or operating rooms, and industrial settings). If this

alerting scheme is implemented within aircraft flight decks, adherence to the guidelines set forth by Berson et al. (1981) will be achieved for two reasons. First, since the alerting system will present aural alerts having acoustic parameters subjectively described as being distinctive and as conveying appropriate levels of urgency, the signals will alert the flight crew to specific impending or existing conditions that require attention and will advise the crew of the alert urgency level. Second, effective acoustic parameters will be used in conjunction with a simple categorization scheme that will reduce the total number of discrete alerts presented in the flight deck, and all current and future alerting components may be categorized within one of the four major flight deck function alerting sets. As a result of these two developments, the standardization of alerting systems among airframe manufacturers, aircraft types, and commercial airplane operations will be facilitated; and since consistent use of an alerting system with distinctive alerting sets and urgency levels will be promoted, the implementation of this aircraft alerting signal categorization scheme may reduce crew information processing and memory requirements as well as minimize the time required for the crew to detect and assess failure conditions and initiate the appropriate corrective action.

Although the proposed aural alert categorization scheme will require that individuals be trained in order to learn the association between a particular alerting set and a particular functional category, the perceived urgency levels of situations signaled by alerts will be the result of inherent responses to the alerting signals' sound parameters. This means that the priority level of a situation will be determined and the decision-making process will be assisted by auditory stimuli that do not increase listeners' workload levels. As stated previously, Hoge et al. (1988) found cross-cultural differences in Western European and Asian perceptions of aural alerting

signals, but it is still suggested that the implementation of this type of categorization scheme as an error-reduction measure is a worthwhile endeavor, even though the development of an international standard may be very difficult.

Recommendations for Additional Research

Before this aural alerting signal categorization scheme is implemented within any operational setting, additional research must be conducted. As suggested by Burt (1996) and, more importantly, by the work of Wilkins (1980 as cited in Wilkins and Martin, 1987), research must be conducted to determine if each aural alerting signal will be heard and attended to as well as recognized (i.e., meaningfully identified and interpreted) by an intended population of listeners who are performing tasks within a particular environment. An aural alert categorization scheme will only serve as an effective error-reduction method within a particular application environment if a representative sample of listeners performing a set of representative tasks within a given environment perceive each alerting set as being distinctive from every other alerting set and perceive appropriate levels of urgency within each alerting set.

Additionally, if the overall urgency levels of the alerting sets need to be prioritized within a particular operational setting, additional research will have to be conducted in order to determine which acoustic parameter(s) should be manipulated. Empirical investigations will have to be conducted to ensure that the distinctiveness of each alerting set is not compromised and that urgency levels within each alerting set are effectively conveyed.

REFERENCES

Alchemy [Computer software]. (1989). Half Moon Bay, CA: Passport Designs, Inc.

Alluisi, E. A., Coates, G. D., and Morgan, B. B. (1977). Effects of temporal stressors on vigilance and information processing. In R. R. Mackie (Ed.), Vigilance: Theory, operational performance, and physiological correlates (pp. 361-421). New York: Plenum Press.

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. American Psychologist, 47, 1597-1611.

Andreassi, J. L. (1989). Psychophysiology: Human behavior and physiological response (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Axelsson, R., and Stoby, F. (1991). Investigation of speech systems for ground collision warning in military aircraft, Report No. FFA-TN-1991-05. Stockholm: The Aeronautical Research Institute.

Beatty, J., Greenberg, A., Diebler, W. P., and O'Hanlon, J. F. (1974). Operant control of occipital theta rhythm affects performance in a radar monitoring task. Science, 183, 871-873.

Berger, H. (1929). On the electroencephalogram of man. Archives of Psychiatry and Nervous Diseases, 87, 511-570.

Berson, B. L., Po-Chedley, D. A., Boucek, G. P., Hanson, D. C., Leffler, M. F., and Wasson, R. L. (1981). Aircraft alerting systems standardization study, volume II: Aircraft alerting systems design guidelines, Report No. DOT/FAA/RD-81/38/II. Springfield, VA: National Technical Information Service.

Bertone, C. M. (1982). Human factors considerations in the development of a voice warning system for helicopters. In C. W. Conner (Chair), Behavioral Objectives in Aviation Automated Systems Symposium, Proceedings No. P-114 (pp. 133-142). Anaheim, CA: Aerospace Congress & Exposition.

Bogart, E. H. (1999). Crew Response Evaluation Window (CREW) system [Computer software]. Hampton, VA: Unisys Corporation.

Boucek, G. P., Po-Chedley, D. A., Berson, B. L., Hanson, D. C., Leffler, M. F., and White, R. W. (1981). Aircraft alerting systems standardization study, volume I: Candidate system validation and time-critical display evaluation, Report No. DOT/FAA/RD-81/38/I. Springfield, VA: National Technical Information Service.

Burt, J. L. (1996). An evaluation of the urgency, similarity, and identification of aural alerts with implications for flight deck use. Unpublished master's thesis, Virginia Polytechnic Institute and State University, Blacksburg.

Burt, J. L., Bartolome, D. S., Burdette, D. W., and Comstock, J. R. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. Ergonomics, 38(11), 2327-2340.

Burt, J. L., and Casali, J. G. (1997). An evaluation of the urgency, similarity, and identification of aural alerts with implications for flight deck use. In B. Bidanda & S. Jagdale (Eds.), 1997 IERC Proceedings: The Sixth Annual Industrial Engineering Research Conference (pp. 731-736). Norcross, GA: Institute of Industrial Engineers.

Casali, J. G., and Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. Human Factors, 25(6), 623-641.

Comstock, J. R., and Arnegard, R. J. (1992). Multi-attribute task battery, Technical Memorandum 104174. Hampton, VA: NASA Langley Research Center.

Cool Edit 95 [Computer software]. (1995). Phoenix, AZ: Syntrillium Software Corporation.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. Perception and Psychophysics, 28(5), 381-389.

Deutsch, D. (1986). Auditory pattern recognition. In K. R. Boff, L. Kaufman & P. Thomas (Eds.), Handbook of perception and human performance (pp. 32.1-32.49). New York: Wiley.

Doll, T. J., and Folds, D. J. (1985). Auditory signals in military aircraft: Ergonomic principles versus practice. In R. S. Jensen & J. Adrion (Eds.), Proceedings of the Third Symposium on Aviation Psychology (pp. 111-125). Columbus, OH: Ohio State University Department of Aviation.

Doll, T. J., Folds, D. J., and Leiker, L. A. (1984). Auditory information systems in military aircraft: Current configurations versus the state of the art, Report No. USAFSAM-TR-84-15. Atlanta, GA: Georgia Institute of Technology.

Edworthy, J. (1985). Interval and contour in melody processing. Music Perception, 2, 375-388.

Edworthy, J. (1994a). Auditory warning design. In M. Vallet (Ed.), Noise as a public health problem (pp. 232-237). Arcueil, France: INRETS.

Edworthy, J. (1994b). Urgency mapping in auditory warning signals. In N. Stanton (Ed.), Human factors in alarm design (pp. 15-30). Bristol, PA: Taylor & Francis, Inc.

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

Empson, J. (1986). Human brainwaves: The psychological significance of the electroencephalogram. New York: Stockton Press.

Engen, T. (1971). Psychophysics II: Scaling methods. In J. W. Kling & L. A. Riggs (Eds.), Woodworth and Schlosberg's experimental psychology (3rd ed., pp.47-86). New York: Holt, Rinehart and Winston.

Folds, D. J. (1985). Development of a comprehensive voice message vocabulary for tactical aircraft. In R. S. Jensen & J. Adrion (Eds.), Proceedings of the Third Symposium on Aviation Psychology (pp. 103-110). Columbus, OH: Ohio State University Department of Aviation.

Freudenberg, R. P., Driscoll, J. W., and Stern, G. S. (1978). Reactions of adult humans to cries of normal and abnormal infants. Infant Behavior and Development, 1, 224-227.

Frodi, A. (1985). When empathy fails: Aversive crying and child abuse. In B. M. Lester & C. F. Z. Boukydis (Eds.), Infant crying: Theoretical and research perspectives (pp. 263-278). New York: Plenum Press.

Gescheider, G. A. (1985). Psychophysics: Method, theory, and application (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hanson, D. C., Howison, W. W., Chikos, S. F., and Berson, B. L. (1982). Aircraft alerting systems standardization study, phase IV: Accident implications on system design, Report No. DOT/FAA/RD-82/26. Springfield, VA: National Technical Information Service.

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload. Amsterdam: North Holland Press.

Hellier, E. J., Edworthy, J., and Dennis, I. (1993). Improved auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. Human Factors, 35(4), 693-706.

Hoge, H., Schick, A., Kuwano, S., Namba, S., Bock, M., and Lazarus, H. (1988). Are there invariants of sound interpretation? The case of danger signals. In Noise as a public health problem, hearing, communication, sleep, and nonauditory physiological effects (pp. 253-258). Stockholm: Swedish Council for Building Research.

ISO (1986). Danger signals for work places - auditory danger signals, ISO 7731-1986(E). Geneva, Switzerland: International Organization for Standardization.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology, 10, 371-375.

Keppel, G. (1982). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall, Inc.

King, R. A., and Corso, G. M. (1993). Auditory displays: If they are so useful, why are they turned off? Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting (pp. 549-553). Seattle, WA.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. Damos (Ed.), Multiple-task performance (pp.279-328). Bristol, PA: Taylor & Francis Inc.

Kroemer, K. H. E., Kroemer, H. B., and Kroemer-Elbert, K. E. (1994). Ergonomics: How to design for ease and efficiency. Englewood Cliffs, NJ: Prentice Hall.

Lazarus, H., and Hoge, H. (1986). Industrial safety: Acoustic signals for danger situations in factories. Applied Ergonomics, 17, 41-46.

Lieberman, M. (1959). Ear training and sight singing. New York: W. W. Norton & Company, Inc.

Lindsley, D. B. (1960). Attention, consciousness, sleep and wakefulness. In J. Field, H. W. Magoun, & V. E. Hall (Eds.), Handbook of physiology, neurophysiology III (pp. 1553-1593). Washington, D. C.: American Physiological Society.

Lower, M., Wheeler, P., Patterson, R., Edworthy, J., Shailer, M., Milroy, R., Rood, G., and Chillery, J. (1986). The design and production of auditory warnings for helicopters 1: The Sea King, ISVR Report No. AC527A.

Lubar, J. F. (1991). Discourse on the development of EEG diagnostics and biofeedback treatment for attention-deficit/hyperactivity disorders. Biofeedback and Self-Regulation, 15, 201-225.

Makeig, S., Elliott, F., Inlow, M., and Kobus, D. (1990). Lapses in alertness: Brain-evoked responses to task-irrelevant auditory probes, Report No. NHRC-90-39. San Diego, CA: Naval Health Research Center.

Marshall, R. (1987). Automatic voice alert devices (AVAD). In Recent advances in cockpit aids for military operations (pp. 70-73). London: The Royal Aeronautical Society.

Meredith, C., and Edworthy, J. (1994). Sources of confusion in intensive therapy unit alarms. In N. Stanton (Ed.), Human factors in alarm design (pp. 207-219). Bristol, PA: Taylor & Francis, Inc.

Miller, M. H., and Wilber, L. A. (1991). Hearing evaluation. In C. M. Harris (Ed.), Handbook of acoustical measurements and noise control (pp. 19.1-19.7). New York: McGraw-Hill, Inc.

Mulholland, T. (1983). Attention and regulation of EEG alpha-attenuation responses. Biofeedback and Self-Regulation, 8(4), 585-600.

Munns, M. (1971). Ways to alarm pilots, Aerospace Medicine, 42, 731-734.

Norusis, M. J. (1992). SPSS/PC+ advanced statistics, version 5.0. Chicago: SPSS Inc.

O'Hanlon, J. F., and Beatty, J. (1979). Concurrence of electroencephalographic and performance changes during a simulated radar watch and some implications for the arousal theory of vigilance. In R. R. Mackie (Ed.), Vigilance: Theory, operational performance, and physiological correlates (pp. 189-201). New York: Plenum Press.

O'Hanlon, J. F., Royal, J. W., and Beatty, J. (1977). EEG theta regulation and radar vigilance performance of professional radar operators and air traffic controllers under laboratory and controlled field conditions. In J. Beatty & H. Legewie (Eds.), Biofeedback and Behavior (pp. 147-165). New York: Plenum Press.

Patterson, R. D. (1982). Guidelines for auditory warning systems on civil aircraft, Report No. 82017. London: Civil Aviation Authority.

Patterson, R. D. (1989). Guidelines for the design of auditory warning sounds, Proceedings of the Institute of Acoustics, 11(5), 17-25.

Ray, W. (1990). The electrocortical system. In J. T. Cacioppo & L. G. Tassinary (Eds.), Principles of psychophysiology: Physical, social and inferential elements (pp. 385-412). Cambridge, England: Cambridge University Press.

Reinecke, M. (1976). Experimentelle erkenntnisse zur eignung von stimmwarnsystemen in flugzeugen [Experimental cognition for qualification of voice warning systems in aircraft], Report No. DGLR PAPER 76-211. West Germany: Flugmedizinisches Institut, Ingolstadt. (Abstract from NASA RECON Document Reproduction Service, Display No. 77A16531)

Reinecke, M. (1981). Voice warning systems: Some experimental evidence concerning application. In K. E. Money (Ed.), AGARD Conference Proceedings No. 311, Aural Communication in Aviation (pp. 9, 1-8). London: Technical Editing and Reproduction, Ltd.

Rood, G. M. (1989), Auditory warnings for fixed and rotary wing aircraft, Proceedings of the Institute of Acoustics, 11(5), 59-71.

Rood, G. M., Chillery, J. A., and Collister, J. B. (1985). Requirements and application of auditory warnings to military helicopters. Ergonomics International, 85, 169-172.

SAE International (1993). Surface vehicle standard, SAE J994 Revised AUG93. Warrendale, PA: Society of Automotive Engineers, Inc.

Schiffman, S. S., Reynolds, M. L., and Young, F. L. (1981). Introduction to multidimensional scaling. New York: Academic Press.

Sorkin, R. (1987). Design of auditory and tactile displays. In G. Salvendy (Ed.), Handbook of human factors (pp.549-576). New York: John Wiley & Sons.

Sorkin, R., Kantowitz, B., and Kantowitz, S. (1988). Likelihood alarm displays. Human Factors, 30, 455-459.

Starer, R. (1969). Rhythmic training. Melville, NY: MCA Music, a Division of MCA, Inc.

Stevens, S. S. (1957). On the psychophysical law. Psychological Review, 64, 153-181.

Stevens, S. S. (1971). Issues in psychophysical measurement. Psychological Review, 78(5), 426-450.

Swink, J. R., and Goins, R. T. (1992). Identification of high-level functional/system requirements for future civil transports, Contractor Report No. 189561. Hampton, VA: NASA Langley Research Center.

Thorning, A. G., and Ablett, R. M. (1985). Auditory warning systems on commercial transport aircraft, Ergonomics International, 85, 166-168.

Veitengruber, J. E., Boucek, G. P., and Smith, W. D. (1977). Aircraft alerting systems criteria study, Vol. I: Collation and analysis of aircraft alerting systems data, Report No. AD-A042 328. Seattle, WA: Boeing Commercial Airplane.

Wheale, J. L. (1981), The speed of response to synthesized voice messages. In K. E. Money (Ed.), AGARD Conference Proceedings No. 311, Aural Communication in Aviation (pp. 7, 1-11). London: Technical Editing and Reproduction, Ltd.

Wheale, J. L. (1982). Performance decrements associated with reaction to voice warning messages. In G. H. Hunt (Chair), AGARD Conference Proceedings No. 329, Advanced Avionics and the Military Aircraft Man/Machine Interface (pp. 18, 1-12). London: Technical Editing and Reproduction, Ltd.

Wheale, J. L. (1983). Evaluation of an experimental central warning system with a synthesized voice component. Aviation, Space, and Environmental Medicine, 54, 517-523.

Wierwille, W. W. (1995, Fall). Industrial and Systems Engineering 5605: Human Factors Systems Design I. Unpublished lecture notes, Virginia Polytechnic Institute and State University.

Wilkins, P. A., and Acton, W. I. (1982). Noise and accidents - A review. Annals of Occupational Hygiene, 25(3), 249-260.

Wilkins, P. A., and Martin, A. M. (1987). Hearing protection and warning sounds in industry - A review. Applied Acoustics, 21, 267-293.

Zeskind, P. S., and Lester, B. M. (1978). Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications. Child Development, 49, 580-589.

Zeskind, P. S., Wilhite, A., and Marshall, T. R. (1993). Infant cry as a graded signal: Experimental modifications of durations of pauses and expiratory sounds alter mothers' perceptions. Early Development and Parenting, 2(2), 99-105.

LIST OF APPENDICES

APPENDIX A. Preliminary Questionnaire

PRELIMINARY QUESTIONNAIRE

(1) Do you find it difficult to understand whispered or faint speech during your everyday communications?

(2) Are you currently taking any of the following medications? If yes, please circle the medication(s) you are taking and list an approximate dosage.

Salicylates (e.g., aspirin)

Aminoglycosides (e.g., streptomycin, neomycin)

Cisplatin

(3) Do you experience a ringing in your ears (i.e., tinnitus)? If yes, how frequently does this occur?

(4) To what types of noise have you been exposed during the last 24 hours (e.g., recreational noise such as that produced by music, auto/motorcycle racing, and gunfire; industrial noise such as that produced by power tools and heavy construction)?

(5) Would you describe any of the noise to which you have been exposed during the last 24 hours as being louder than the noise produced by a lawnmower or food blender? If yes, how long were you exposed to this noise?

(6) Please list any formal musical training you have had (e.g., private lessons, college education, etc.).

(7) Do you play an instrument? If yes, what instrument(s) do you play, and how long have you played each instrument?

(8) Do you sing alone or in a group for your own enjoyment? If yes, how long have you done this?

(9) What is your gender?

(10) What is your age?

APPENDIX B. Line Length Estimation Stimuli

LINE LENGTH ESTIMATION TASK

Line A:

Line B:

Line C:

Line D:

Line E:

NOTE: Each line was presented to subjects on a separate sheet of paper.

APPENDIX C. Audiometric Testing Chamber Noise Levels

Model #	Wall Thickness	Frequency	Ambient Noise Levels
1201-A-W/FV	60 cm walls with 11.25" air space	125 Hz	24.1 dB
		250 Hz	14.2 dB
		500 Hz	9.5 dB
		1000 Hz	9.5 dB
		2000 Hz	12.1 dB
		4000 Hz	7.2 dB
		8000 Hz	5.7 dB
401-A-SE	35 cm walls	125 Hz	34.5 dB
		250 Hz	15.5 dB
		500 Hz	10.0 dB
		1000 Hz	3.2 dB
		2000 Hz	4.0 dB
		4000 Hz	3.0 dB
		8000 Hz	---- dB

APPENDIX D. Experimental Chamber Ambient Sound Pressure Level Measurements

One-Third Octave Bands				
Lower Frequency (Hz)	Center Frequency (Hz)	Upper Frequency (Hz)	dBV*	dBA**
224	250	280	- 92.3	31.8
280	315	355	- 90.9	33.2
355	400	450	- 90.0	34.0
450	500	560	- 88.5	35.5
560	630	710	- 86.9	37.1
710	800	900	- 83.2	40.8
900	1000	1120	- 83.7	40.3
1120	1250	1400	- 83.1	40.9
1400	1600	1800	- 84.7	39.4
1800	2000	2240	- 82.4	41.6
2240	2500	2800	- 82.3	41.7
2800	3150	3550	- 81.5	42.5
3550	4000	4500	- 80.4	43.6
4500	5000	5600	- 80.5	43.5
5600	6300	7100	- 79.7	44.3

* dBV re. 0 dBV
** dBA re. 124 dBA

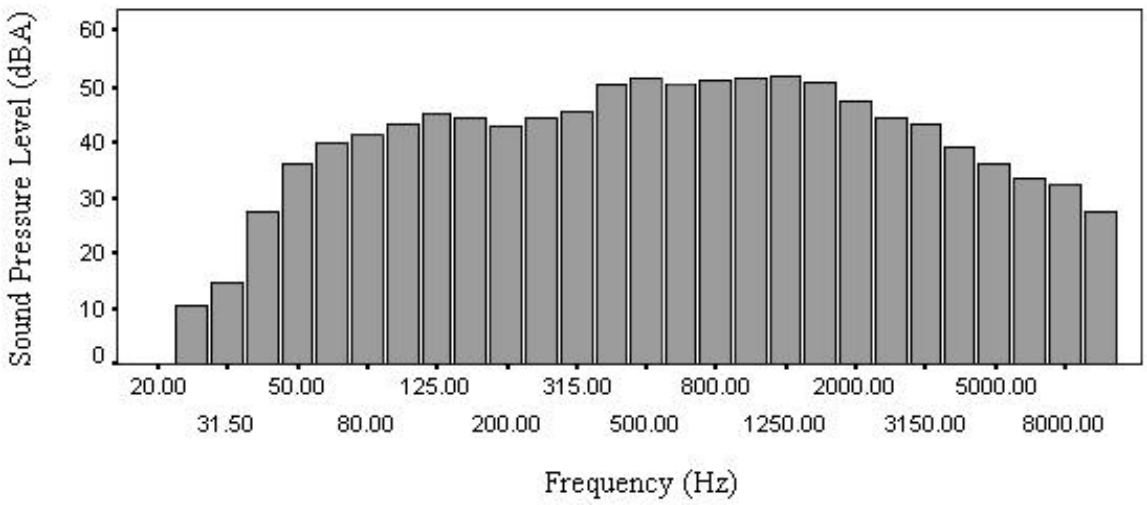
APPENDIX E. Experimental Chamber and Aircraft Flight Deck Reverberation Time Measurements

One-Third Octave Bands				
Lower Frequency (Hz)	Center Frequency (Hz)	Upper Frequency (Hz)	Chamber Reverberation Time (ms)	Flight Deck Reverberation Time (ms)
224	250	280	160	150
280	315	355	110	140
355	400	450	80	130
450	500	560	70	130
560	630	710	50	100
710	800	900	50	80
900	1000	1120	60	90
1120	1250	1400	50	80
1400	1600	1800	60	110
1800	2000	2240	60	100
2240	2500	2800	60	100
2800	3150	3550	60	100
3550	4000	4500	60	100
4500	5000	5600	60	110
5600	6300	7100	60	120

APPENDIX F. Flight Deck Background Noise Sound Pressure Level Measurements

One-Third Octave Band			
Lower Frequency (Hz)	Center Frequency (Hz)	Upper Frequency (Hz)	dBA
—	20	---	0
22.4	25	28	10.7
28	31.5	35.5	14.8
35.5	40	45	27.7
45	50	56	36.0
56	63	71	39.9
71	80	90	41.2
90	100	112	43.1
112	125	140	45.2
140	160	180	44.1
180	200	224	42.8
224	250	280	44.3
280	315	355	45.6
355	400	450	50.3
450	500	560	51.5
560	630	710	50.3
710	800	900	51.2
900	1000	1120	51.5
1120	1250	1400	51.8
1400	1600	1800	51.0
1800	2000	2240	47.4
2240	2500	2800	44.4
2800	3150	3550	43.4
3550	4000	4500	39.3
4500	5000	5600	35.8
5600	6300	7100	33.6
7100	8000	9000	32.4
9000	10000	11200	27.8

APPENDIX G. Spectral Plot of Flight Deck Background Noise



APPENDIX H. Informed Consent Form for Experiment 1

INFORMED CONSENT FOR PARTICIPANTS OF INVESTIGATIVE PROJECTS

Title of Project: Continued Empirical Studies Concerning Aural Alerts for Cockpit Use: Doctoral Dissertation - Experiment 1

Investigators: Jennifer L. Burt, Dr. John G. Casali, and Dr. Alan T. Pope

I. The Purpose of this Research

You are invited to participate in a study concerning the perception of aural alerting signal urgency level. This study involves experimentation for the purpose of identifying sound parameters that are subjectively described as having low, moderate, and high levels of perceived urgency. This study involves 19 participants other than yourself.

II. Procedures

During a single visit to the Crew Hazards and Error Management Laboratory, located at NASA Langley Research Center in Building 1268A, Room 1139, you will be asked to: complete a preliminary paper-and-pencil questionnaire regarding your musical experience, recent noise exposure, and use of medications that may affect sound perception; view a series of lines and record numerical estimates of the lines' lengths using a paper-and-pencil response form; and listen to a series of aural alerting signals presented against a background of ambient 737 cockpit noise and record your subjective perceptions of alerting signal urgency using a paper-and-pencil response form. The alerting signals and background noise will be presented to you over three loudspeakers at sound pressure levels below that which causes any damage or discomfort to the ears.

To participate in this study, you must have "normal" hearing threshold levels as determined by the NASA Langley Research Center Medical Center (Building 1149) within the last six months, and you must not have been exposed to any excessively loud sounds (e.g., any sound louder than that produced by a power lawnmower) during the past 24 hours.

Participation in this study will require approximately one hour and fifteen minutes of your time. This time requirement includes the approximately 30 minutes spent having your hearing assessed at the NASA Langley Research Center Medical Center.

III. Risks

Since the alerting signals and background noise to which you will listen will be presented at sound pressure levels below that which causes any damage or discomfort to the ears, there are no apparent risks to you from participation in this study. All alerting signals will be presented at a sound pressure level below that produced by a garbage disposal or very busy traffic (i.e., signals will be presented at a level of 75 dBA), and background noise will be presented at a sound pressure level equivalent to that produced by an ordinary two-person conversation (i.e., background noise will be presented at a level of 60 dBA).

IV. Benefits of this Project

Your participation in this project will provide information that may be helpful in the design of a safer and more effective aural alerting system.

No promise or guarantee of benefits has been made to encourage you to participate.

If you would like to receive a synopsis or summary of this research when it is completed, please notify Jennifer Burt.

V. Extent of Anonymity and Confidentiality

The results of this study will be kept strictly confidential. At no time will the researchers release your results to anyone other than individuals working on the project without your written consent. The information you provide will have your name removed, and only a participant number (i.e., Participant #1, Participant #2, etc.) will identify you during analyses and any written reports of the research.

VI. Compensation

No financial compensation will be offered to you for participation in this project.

VII. Freedom to Withdraw

You are free to withdraw from this study at any time without penalty. Furthermore, you are free not to answer any questions or respond to any experimental situations that you choose without penalty.

VIII. Approval of Research

This research project has been approved, as required, by the Institutional Review Board (IRB) for Research Involving Human Participants at Virginia Polytechnic Institute and State University, by the Virginia Polytechnic Institute and State University Department of Industrial and Systems Engineering, and by the IRB for Research Involving Human Participants at NASA Langley Research Center.

IX. Participant's Responsibilities

I voluntarily agree to participate in this study. I understand that I have the following responsibilities:

1. To have undergone an audiometric evaluation at the NASA Langley Research Center Medical Center within the last six months
2. To complete a preliminary paper-and-pencil questionnaire regarding my musical experience, recent noise exposure, and use of medications that may affect sound perception
3. To examine a series of lines and record numerical estimates of the lines' lengths using a paper-and-pencil response form
4. To listen to a series of aural alerting signals presented against a background of ambient 737 cockpit noise and record my subjective perceptions of aural alerting signal urgency using a paper-and-pencil response form

X. Participant's Permission

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.

If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

_____

Signature Date

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

(Principle Investigator) Jennifer L. Burt	(757) 864-8304
(Faculty Advisor, Virginia Tech) Dr. John G. Casali	(540) 231-5073
(Chair, Virginia Tech IRB Research Division) H. T. Hurd	(540) 231-9359
(Supervisor, NASA LaRC) Dr. Alan T. Pope	(757) 864-6642
(Chair, NASA LaRC IRB) Dr. Alan W. Wilhite	(757) 864-2982

NOTE: Each participant signed **two** copies of this Informed Consent Form. One form was given to the participant, and the other form was retained by the investigator.

APPENDIX I. Magnitude Estimation Task Response Form: Line Length

MAGNITUDE ESTIMATION RATING TASK: LINE LENGTH

Line A _____

Line B _____

Line C _____

Line D _____

Line E _____

APPENDIX J. Magnitude Estimation Task Response Form: Aural Alert Urgency Level

MAGNITUDE ESTIMATION RATING TASK: AURAL ALERT URGENCY LEVEL

urgency (*n.*) - the quality or state of being important, insistent, or pressing.

Practice Trial A _____

Practice Trial B _____

Practice Trial C _____

Practice Trial D _____

Sound A _____

Sound B _____

Sound C _____

Sound D _____

Sound E _____

Sound F _____

-

-

-

Sound AL _____

Sound AM _____

Sound AN _____

APPENDIX K. Informed Consent Form for Experiment 2

INFORMED CONSENT FOR PARTICIPANTS OF INVESTIGATIVE PROJECTS

Title of Project: Continued Empirical Studies Concerning Aural Alerts for Cockpit Use: Doctoral Dissertation - Experiment 2

Investigators: Jennifer L. Burt, Dr. John G. Casali, and Dr. Alan T. Pope

I. The Purpose of this Research

You are invited to participate in a study concerning the identification of various aural alerting sets and urgency levels. This study involves experimentation for the purpose of validating sound parameters that have been subjectively described as being distinct and as having low, moderate, and high levels of perceived urgency. This study involves 11 participants other than yourself.

II. Procedures

During a single visit to the Crew Hazards and Error Management (CHEM) Laboratory, located at NASA Langley Research Center in Building 1268A, Room 1139, you will be asked to: complete a preliminary paper-and-pencil questionnaire regarding your musical experience, recent noise exposure, and use of medications that may affect sound perception; listen to a series of aural alerting signals presented against a background of ambient 737 cockpit noise; associate each of four aural alerting sets with one of the four major flight deck functions (i.e., communication, flight control, navigation, and systems management); and record your identifications of various alerting sets and urgency levels via a paper-and-pencil response form. The alerting signals and background noise will be presented to you over three loudspeakers at sound pressure levels below that which causes any damage or discomfort to the ears.

During the same visit to the CHEM Laboratory, you will also be asked to simultaneously: perform a computer-based tracking task involving two conditions (i.e., automated tracking and manual tracking) while your electroencephalogram (EEG), or brain wave, data are recorded; listen to a series of aural alerting signals presented against a background of ambient 737 cockpit noise; and verbally identify various alerting sets and urgency levels whenever alerting signals are presented. In the automated condition of the tracking task, you will be asked to monitor computer tracking of a circular target displayed on a computer monitor. In the manual condition of the tracking task, you will be asked to use a joystick to keep a circular target within a rectangular boundary displayed on a computer monitor. Your brain wave data will be collected

II. Procedures (continued)

using a lycra head cap consisting of 22 recessed electrodes, a ground electrode attached to your left mastoid prominence (i.e., the small protrusion located behind your ear), and a reference electrode attached to your left earlobe; this is a non-invasive procedure routinely employed in clinical practice and experimental research. Again, the alerting signals and background noise will be presented to you over three loudspeakers at sound pressure levels below that which causes any damage or discomfort to the ears.

After completing the first condition of the tracking task (i.e., either the automated or manual condition), you will be asked to record your subjective assessment of the level of workload that you experienced while performing this tracking task condition. After completing the second condition of the tracking task, you will be asked to record your subjective assessment of the level of workload that you experienced while performing this tracking task condition. Subjective assessments of workload will be recorded using a computerized version of the NASA Task Load Index (TLX). The NASA TLX rating screens will be displayed on a computer monitor, and a computer mouse will be used for data entry.

After recording your subjective assessment of the workload level experienced while performing the second tracking task condition, you will be asked to complete a post-experiment paper-and-pencil questionnaire.

To participate in this study, you must be right handed; have "normal" (i.e., 20/20 or better) or corrected-to-normal vision; have no history of neurological problems that could interfere with the recording of brain wave data; and have "normal" hearing threshold levels as determined by the NASA Langley Research Center Medical Center (Building 1149) within the last six months. Furthermore, you must not have been exposed to any excessively loud sounds (e.g., any sound louder than that produced by a power lawnmower) during the past 24 hours.

Participation in this study will require approximately three and a half hours of your time. This time requirement includes the approximately 30 minutes spent having your hearing assessed at the NASA Langley Research Center Medical Center.

III. Risks

Since the alerting signals and background noise to which you will listen will be presented at sound pressure levels below that which causes any damage or discomfort to the ears and the recording of brain wave data is a non-invasive procedure routinely employed in clinical practice and experimental research, there are no apparent risks to you from participation in this study.

All alerting signals will be presented at a sound pressure level below that produced by a garbage disposal or very busy traffic (i.e., signals will be presented at a level of 75 dBA), and

III. Risks (continued)

background noise will be presented at a sound pressure level equivalent to that produced by an ordinary two-person conversation (i.e., background noise will be presented at a level of 60 dBA).

Brain wave data will be recorded using a hospital grade electroencephalograph that is designed to prevent the occurrence of electrical shock, and the investigator responsible for recording brain wave data is well trained in the safe operation of all equipment.

IV. Benefits of this Project

Your participation in this project will provide information that may be helpful in the design of a safer and more effective aural alerting system.

No promise or guarantee of benefits has been made to encourage you to participate.

If you would like to receive a synopsis or summary of this research when it is completed, please notify Jennifer Burt.

V. Extent of Anonymity and Confidentiality

The results of this study will be kept strictly confidential. At no time will the researchers release your results to anyone other than individuals working on the project without your written consent. The information you provide will have your name removed, and only a participant number (i.e., Participant #1, Participant #2, etc.) will identify you during analyses and any written reports of the research.

VI. Compensation

No financial compensation will be offered to you for participation in this project.

VII. Freedom to Withdraw

You are free to withdraw from this study at any time without penalty. Furthermore, you are free not to answer any questions or respond to any experimental situations that you choose without penalty.

VIII. Approval of Research

This research project has been approved, as required, by the Institutional Review Board (IRB) for Research Involving Human Participants at Virginia Polytechnic Institute and State University, by the Virginia Polytechnic Institute and State University Department of Industrial and Systems Engineering, and by the IRB for Research Involving Human Participants at NASA Langley Research Center.

IX. Participant's Responsibilities

I voluntarily agree to participate in this study. I understand that I have the following responsibilities:

1. To have undergone an audiometric evaluation at the NASA Langley Research Center Medical Center within the last six months
2. To complete a preliminary paper-and-pencil questionnaire regarding my musical experience, recent noise exposure, and use of medications that may affect sound perception
3. To listen to a series of aural alerting signals presented against a background of ambient 737 cockpit noise
4. To associate each of four aural alerting sets with one of the four major flight deck functions and record my identifications of aural alerting sets and urgency levels using a paper-and-pencil response form
5. To perform a computer-based tracking task involving automated and manual control conditions and verbally identify various aural alerting sets and urgency levels while my brain wave data are recorded
6. To record my subjective assessments of workload experienced while performing each tracking task condition using a computerized version of the NASA Task Load Index
7. To complete a post-experiment paper-and-pencil questionnaire

X. Participant's Permission

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.

X. Participant's Permission (continued)

If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

_____

Signature Date

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

(Principle Investigator) Jennifer L. Burt	(757) 864-8304
(Faculty Advisor, Virginia Tech) Dr. John G. Casali	(540) 231-5073
(Chair, Virginia Tech IRB Research Division) H. T. Hurd	(540) 231-9359
(Supervisor, NASA LaRC) Dr. Alan T. Pope	(757) 864-6642
(Chair, NASA LaRC IRB) Dr. Alan W. Wilhite	(757) 864-2982

NOTE: The participant signed **two** copies of this Informed Consent Form. One form was given to the participant, and the other form was retained by the investigator.

APPENDIX L. Sound Identification Training Task Response Form

SOUND IDENTIFICATION TRAINING TASK

Practice Trial A:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

Practice Trial B:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

Practice Trial C:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

Practice Trial D:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

Sound A:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

Sound B:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

-

-

-

Sound X:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

APPENDIX M. Sound Identification Data Collection Form

SOUND IDENTIFICATION DATA COLLECTION FORM

Tracking condition: _____

Practice Trial A:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

Practice Trial B:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

-

-

-

Practice Trial H:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

Sound A:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

Sound B:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

-

-

-

Sound X:

COMMUNICATION	FLIGHT CONTROL	NAVIGATION	SYSTEMS MANAGEMENT
Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___	Urgency Level Low___ Mod___ High___

APPENDIX N. Post-Experiment Questionnaire

POST-EXPERIMENT QUESTIONNAIRE

(1) Please rate how difficult it was for you to associate each set of sounds (i.e., SET I, SET II, SET III, and SET IV) with one flight deck function (i.e., communication, flight control, navigation, and systems management) by making a mark anywhere along the horizontal line of the rating scale shown below:

| | | | | | | | | |

Extremely Easy — Extremely Difficult

Additional comments:

_____

_____

(2) What characteristics of the sounds (e.g., rhythm, pitch, etc.) did you use to "mentally group" the three sounds found in each set?

_____

_____

_____

_____

(3) What is your overall reaction to the sounds that you listened to and identified during the experimental sessions?

_____

_____

_____

(4) Did you find it difficult to keep your attention focused on the sounds during the experimental sessions?

_____

_____

_____

(5) Would you consider volunteering to participate in a similar study sometime in the future?

_____

_____

_____

(6) Any additional comments that you would like to share:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

APPENDIX O. ANOVA Summary and Bonferroni t-Test Summary Tables for EEG Data Recorded During Baseline Conditions

ANOVA Summary Table of Baseline EEG Data Recorded at Site Cz

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	65.64				
Within-Subject						
Baseline Condition (BC)	1	6.77	6.77	28.20	0.0001	
BC X S	11	2.64	0.24			
Frequency Band (FB)	2	105.10	52.55	21.78	0.0001	
FB X S	22	53.08	2.41			
BC X FB	2	17.36	8.68	33.55	0.0001	0.0001
BC X FB X S	22	5.69	0.26			
Total	71	256.28				

NOTE: With respect to the Baseline Condition x Frequency Band interaction, a Mauchly's test of sphericity produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [2] = 7.74; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.64978 be used to correct the problems associated with a positively biased F-Test. A significant interaction effect was found to exist, however, even after the conservative Greenhouse-Geisser correction was applied (F [1, 14] = 33.55; $p \leq 0.05$).

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site Cz During Baseline Conditions

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order		
	Beta Activity (3.25)	Alpha Activity (5.51)	Theta Activity (6.03)
Beta Activity (3.25)	-	2.26*	2.78*
Alpha Activity (5.51)		-	0.52
Theta Activity (6.03)			-

* $p \leq 0.05$

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site Cz During Each Baseline Condition

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order					
	Beta Activity Eyes Closed (2.93)	Beta Activity Eyes Open (3.57)	Alpha Activity Eyes Open (4.63)	Theta Activity Eyes Open (5.67)	Alpha Activity Eyes Closed (6.39)	Theta Activity Eyes Closed (6.39)
Beta Activity Eyes Closed (2.93)	-	0.64	1.70	2.74*	3.46*	3.46*
Beta Activity Eyes Open (3.57)		-	1.06	2.10*	2.82*	2.82*
Alpha Activity Eyes Open (4.63)			-	1.04	1.76*	1.76*
Theta Activity Eyes Open (5.67)				-	0.72	0.72
Alpha Activity Eyes Closed (6.39)					-	0
Theta Activity Eyes Closed (6.39)						-

* $p \leq 0.05$

ANOVA Summary Table of Baseline EEG Data Recorded at Site Pz

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	47.06				
Within-Subject						
Baseline Condition (BC)	1	20.38	20.38	19.63	0.001	
BC X S	11	11.43	1.04			
Frequency Band (FB)	2	121.75	60.88	23.66	0.0001	
FB X S	22	56.61	2.57			
BC X FB	2	41.44	20.72	18.99	0.0001	0.0001
BC X FB X S	22	24.01	1.09			
Total	71	322.68				

NOTE: With respect to the Baseline Condition x Frequency Band interaction, a Mauchly's test of sphericity produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [2] = 24.06; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.52360 be used to correct the problems associated with a positively biased F-Test. A significant interaction effect was found to exist, however, even after the conservative Greenhouse-Geisser correction was applied (F [1, 11] = 18.99; $p \leq 0.05$).

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site Pz During Baseline Conditions

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order		
	Beta Activity (2.54)	Alpha Activity (5.16)	Theta Activity (5.42)
Beta Activity (2.54)	-	2.62*	2.88*
Alpha Activity (5.16)		-	0.26
Theta Activity (5.42)			-

* $p \leq 0.05$

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site Pz During Each Baseline Condition

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order					
	Beta Activity Eyes Closed (2.27)	Beta Activity Eyes Open (2.82)	Alpha Activity Eyes Open (3.62)	Theta Activity Eyes Open (5.10)	Theta Activity Eyes Closed (5.75)	Alpha Activity Eyes Closed (6.71)
Beta Activity Eyes Closed (2.27)	-	0.55	1.35	2.83*	3.48*	4.44*
Beta Activity Eyes Open (2.82)		-	0.80	2.28*	2.93*	3.89*
Alpha Activity Eyes Open (3.62)			-	1.48	2.13*	3.09*
Theta Activity Eyes Open (5.10)				-	0.65	1.61
Theta Activity Eyes Closed (5.75)					-	0.96
Alpha Activity Eyes Closed (6.71)						-

* $p \leq 0.05$

ANOVA Summary Table of Baseline EEG Data Recorded at Site P3

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	42.27				
Within-Subject						
Baseline Condition (BC)	1	18.81	18.81	19.35	0.001	
BC X S	11	10.69	0.97			
Frequency Band (FB)	2	109.55	54.77	23.15	0.0001	
FB X S	22	52.06	2.37			
BC X FB	2	26.92	13.46	16.40	0.0001	0.0001
BC X FB X S	22	18.05	0.82			
Total	71	278.35				

NOTE: With respect to the Baseline Condition x Frequency Band interaction, a Mauchly's test of sphericity produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [2] = 29.41; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.51357 be used to correct the problems associated with a positively biased F-Test. A significant interaction effect was found to exist, however, even after the conservative Greenhouse-Geisser correction was applied (F [1, 11] = 16.40; $p \leq 0.05$).

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site P3 During Baseline Conditions

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order		
	Beta Activity (2.43)	Alpha Activity (4.69)	Theta Activity (5.30)
Beta Activity (2.43)	-	2.26*	2.87*
Alpha Activity (4.69)		-	0.61
Theta Activity (5.30)			-

* $p \leq 0.05$

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site P3 During Each Baseline Condition

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order					
	Beta Activity Eyes Closed (2.26)	Beta Activity Eyes Open (2.60)	Alpha Activity Eyes Open (3.38)	Theta Activity Eyes Open (4.91)	Theta Activity Eyes Closed (5.69)	Alpha Activity Eyes Closed (6.00)
Beta Activity Eyes Closed (2.26)	-	0.34	1.12	2.65*	3.43*	3.74*
Beta Activity Eyes Open (2.60)		-	0.78	2.31*	3.09*	3.40*
Alpha Activity Eyes Open (3.38)			-	1.53	2.31*	2.62*
Theta Activity Eyes Open (4.91)				-	0.78	1.09
Theta Activity Eyes Closed (5.69)					-	0.31
Alpha Activity Eyes Closed (6.00)						-

* $p \leq 0.05$

ANOVA Summary Table of Baseline EEG Data Recorded at Site P4

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	45.36				
Within-Subject						
Baseline Condition (BC)	1	13.43	13.43	23.13	0.001	
BC X S	11	6.39	0.58			
Frequency Band (FB)	2	102.93	51.47	22.11	0.0001	
FB X S	22	51.22	2.33			
BC X FB	2	33.39	16.70	15.75	0.0001	0.0001
BC X FB X S	22	23.32	1.06			
Total	71	276.04				

NOTE: With respect to the Baseline Condition x Frequency Band interaction, a Mauchly's test of sphericity produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [2] = 17.22; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.54904 be used to correct the problems associated with a positively biased F-Test. A significant interaction effect was found to exist, however, even after the conservative Greenhouse-Geisser correction was applied (F [1, 12] = 15.75; $p \leq 0.05$).

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site P4 During Baseline Conditions

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order		
	Beta Activity (2.68)	Alpha Activity (4.90)	Theta Activity (5.44)
Beta Activity (2.68)	-	2.22*	2.76*
Alpha Activity (4.90)		-	0.54
Theta Activity (5.44)			-

* $p \leq 0.05$

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site P4 During Each Baseline Condition

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order					
	Beta Activity Eyes Closed (2.34)	Beta Activity Eyes Open (3.02)	Alpha Activity Eyes Open (3.58)	Theta Activity Eyes Open (5.12)	Theta Activity Eyes Closed (5.76)	Alpha Activity Eyes Closed (6.22)
Beta Activity Eyes Closed (2.34)	-	0.68	1.24	2.78*	3.42*	3.88*
Beta Activity Eyes Open (3.02)		-	0.56	2.10*	2.74*	3.20*
Alpha Activity Eyes Open (3.58)			-	1.54	2.18*	2.64*
Theta Activity Eyes Open (5.12)				-	0.64	1.10
Theta Activity Eyes Closed (5.76)					-	0.46
Alpha Activity Eyes Closed (6.22)						-

* $p \leq 0.05$

APPENDIX P. ANOVA Summary and Bonferroni t-Test Summary Tables for EEG Data Recorded During Tracking Task Conditions

ANOVA Summary Table of Tracking Task EEG Data Recorded at Site Cz

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	69.65				
Within-Subject						
Tracking Condition (TC)	1	8.73	8.73	8.32	0.015	
TC X S	11	11.54	1.05			
Frequency Band (FB)	2	68.76	34.38	29.69	0.0001	
FB X S	22	25.48	1.16			
TC X FB	2	53.01	26.50	13.46	0.0001	0.0001
TC X FB X S	22	43.31	1.97			
Total	71	280.48				

NOTE: With respect to the Tracking Condition x Frequency Band interaction, a Mauchly's test of sphericity produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [2] = 11.65; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.59237 be used to correct the problems associated with a positively biased F-Test. A significant interaction effect was found to exist, however, even after the conservative Greenhouse-Geisser correction was applied (F [1, 13] = 13.46; $p \leq 0.05$).

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site Cz During Tracking Task Conditions

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order		
	Beta Activity (3.60)	Alpha Activity (4.08)	Theta Activity (5.87)
Beta Activity (3.60)	-	0.48	2.27*
Alpha Activity (4.08)		-	1.80*
Theta Activity (5.87)			-

* $p \leq 0.05$

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site Cz During Each Tracking Task Condition

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order					
	Beta Activity Auto Tracking (2.74)	Alpha Activity Manual Tracking (3.18)	Beta Activity Manual Tracking (4.46)	Theta Activity Manual Tracking (4.86)	Alpha Activity Auto Tracking (4.97)	Theta Activity Auto Tracking (6.88)
Beta Activity Auto Tracking (2.74)	-	0.44	1.72	2.12*	2.23*	4.14*
Alpha Activity Manual Tracking (3.18)		-	1.28	1.68	1.79	3.70*
Beta Activity Manual Tracking (4.46)			-	0.40	0.51	2.42*
Theta Activity Manual Tracking (4.86)				-	0.11	2.02*
Alpha Activity Auto Tracking (4.97)					-	1.91*
Theta Activity Auto Tracking (6.88)						-

* $p \leq 0.05$

ANOVA Summary Table of Tracking Task EEG Data Recorded at Site Pz

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	14.51				
Within-Subject						
Tracking Condition (TC)	1	15.54	15.54	26.67	0.0001	
TC X S	11	6.41	0.58			
Frequency Band (FB)	2	42.76	21.38	13.01	0.0001	
FB X S	22	36.16	1.64			
TC X FB	2	64.39	32.20	19.68	0.0001	0.0001
TC X FB X S	22	35.99	1.64			
Total	71	215.76				

NOTE: With respect to the Tracking Condition x Frequency Band interaction, a Mauchly's test of sphericity produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [2] = 22.52; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.52775 be used to correct the problems associated with a positively biased F-Test. A significant interaction effect was found to exist, however, even after the conservative Greenhouse-Geisser correction was applied (F [1, 11] = 19.68; $p \leq 0.05$).

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site Pz During Tracking Task Conditions

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order		
	Beta Activity (3.42)	Alpha Activity (3.93)	Theta Activity (5.25)
Beta Activity (3.42)	-	0.51	1.83*
Alpha Activity (3.93)		-	1.32*
Theta Activity (5.25)			-

* $p \leq 0.05$

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site Pz During Each Tracking Task Condition

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order					
	Beta Activity Auto Tracking (2.56)	Alpha Activity Manual Tracking (2.98)	Theta Activity Manual Tracking (3.95)	Beta Activity Manual Tracking (4.28)	Alpha Activity Auto Tracking (4.88)	Theta Activity Auto Tracking (6.55)
Beta Activity Auto Tracking (2.56)	-	0.42	1.39	1.72*	2.32*	3.99*
Alpha Activity Manual Tracking (2.98)		-	0.97	1.30	1.90*	3.57*
Theta Activity Manual Tracking (3.95)			-	0.33	0.93	2.60*
Beta Activity Manual Tracking (4.28)				-	0.60	2.27*
Alpha Activity Auto Tracking (4.88)					-	1.67*
Theta Activity Auto Tracking (6.55)						-

* $p \leq 0.05$

ANOVA Summary Table of Tracking Task EEG Data Recorded at Site P3

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	17.42				
Within-Subject						
Tracking Condition (TC)	1	16.42	16.42	24.80	0.0001	
TC X S	11	7.28	0.66			
Frequency Band (FB)	2	36.82	18.41	14.45	0.0001	
FB X S	22	28.02	1.27			
TC X FB	2	64.85	32.42	18.58	0.0001	0.0001
TC X FB X S	22	38.40	1.75			
Total	71	209.21				

NOTE: With respect to the Tracking Condition x Frequency Band interaction, a Mauchly's test of sphericity produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [2] = 21.47; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.53103 be used to correct the problems associated with a positively biased F-Test. A significant interaction effect was found to exist, however, even after the conservative Greenhouse-Geisser correction was applied (F [1, 11] = 18.58; $p \leq 0.05$).

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site P3 During Tracking Task Conditions

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order		
	Beta Activity (3.31)	Alpha Activity (3.63)	Theta Activity (4.96)
Beta Activity (3.31)	-	0.32	1.65*
Alpha Activity (3.63)		-	1.33*
Theta Activity (4.96)			-

* $p \leq 0.05$

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site P3 During Each Tracking Task Condition

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order					
	Beta Activity Auto Tracking (2.46)	Alpha Activity Manual Tracking (2.63)	Theta Activity Manual Tracking (3.68)	Beta Activity Manual Tracking (4.17)	Alpha Activity Auto Tracking (4.64)	Theta Activity Auto Tracking (6.25)
Beta Activity Auto Tracking (2.46)	-	0.17	1.22	1.71*	2.18*	3.79*
Alpha Activity Manual Tracking (2.63)		-	1.05	1.54*	2.01*	3.62*
Theta Activity Manual Tracking (3.68)			-	0.49	0.96	2.57*
Beta Activity Manual Tracking (4.17)				-	0.47	2.08*
Alpha Activity Auto Tracking (4.64)					-	1.61*
Theta Activity Auto Tracking (6.25)						-

* $p \leq 0.05$

ANOVA Summary Table of Tracking Task EEG Data Recorded at Site P4

Source	df	SS	MS	F	p	G-Gp
Between-Subjects						
Subjects (S)	11	7.48				
Within-Subject						
Tracking Condition (TC)	1	16.57	16.57	28.67	0.0001	
TC X S	11	6.36	0.58			
Frequency Band (FB)	2	43.12	21.56	20.92	0.0001	
FB X S	22	22.67	1.03			
TC X FB	2	65.65	32.82	21.16	0.0001	0.0001
TC X FB X S	22	34.13	1.55			
Total	71	195.98				

NOTE: With respect to the Tracking Condition x Frequency Band interaction, a Mauchly's test of sphericity produced an observed significance level based on a X^2 approximation that led to the rejection of the hypothesis of sphericity (X^2 [2] = 22.35; $p < 0.05$) and suggested that a Greenhouse-Geisser Epsilon value of 0.52827 be used to correct the problems associated with a positively biased F-Test. A significant interaction effect was found to exist, however, even after the conservative Greenhouse-Geisser correction was applied (F [1, 11] = 21.16; $p \leq 0.05$).

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site P4 During Tracking Task Conditions

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order		
	Beta Activity (3.44)	Alpha Activity (3.77)	Theta Activity (5.22)
Beta Activity (3.44)	-	0.33	1.78*
Alpha Activity (3.77)		-	1.45*
Theta Activity (5.22)			-

* $p \leq 0.05$

Bonferroni t-Test Summary Table for EEG Frequency Band Data Recorded at Site P4 During Each Tracking Task Condition

Differences Among Treatment Means in Increasing Order	Differences Among Treatment Means in Increasing Order					
	Beta Activity Auto Tracking (2.59)	Alpha Activity Manual Tracking (2.83)	Theta Activity Manual Tracking (3.87)	Beta Activity Manual Tracking (4.29)	Alpha Activity Auto Tracking (4.71)	Theta Activity Auto Tracking (6.57)
Beta Activity Auto Tracking (2.59)	-	0.24	1.28	1.70*	2.12*	3.98*
Alpha Activity Manual Tracking (2.83)		-	1.04	1.46*	1.88*	3.74*
Theta Activity Manual Tracking (3.87)			-	0.42	0.84	2.70*
Beta Activity Manual Tracking (4.29)				-	0.42	2.28*
Alpha Activity Auto Tracking (4.71)					-	1.86*
Theta Activity Auto Tracking (6.57)						-

* $p \leq 0.05$

VITA

JENNIFER L. BURT

ACADEMIC EXPERIENCE

August 1994 - present	Enrolled as a graduate student at Virginia Polytechnic Institute and State University, Blacksburg, Virginia
	Ph.D. candidate, Industrial and Systems Engineering Concentration: Human Factors Engineering Dissertation topic: Empirical studies concerning aural alerts for cockpit use leading to an aural alerting signal categorization scheme
	M.S., Industrial and Systems Engineering, December 1996 Concentration: Human Factors Engineering Thesis topic: An evaluation of the urgency, similarity, and identification of aural alerts with implications for flight deck use
August 1990 - May 1994	Enrolled as an undergraduate student at Christopher Newport University, Newport News, Virginia
	B.A., Psychology, May 1994 Research topic: A psychophysiological evaluation of the perceived urgency of auditory warning signals

PROFESSIONAL EXPERIENCE

July 1999 - present	Crew-Automation Integration Technologist Crew/Vehicle Integration Branch NASA Langley Research Center Hampton, Virginia
July 1995 - June 1999	Graduate Student Researchers Program Participant Crew/Vehicle Integration Branch NASA Langley Research Center Hampton, Virginia
August 1994 - May 1995	Graduate Teaching Assistant Department of Psychology Virginia Polytechnic Institute and State University Blacksburg, Virginia

PROFESSIONAL EXPERIENCE (continued)

June 1994 - August 1994	Langley Aerospace Research Summer Scholars Program Participant Crew/Vehicle Integration Branch NASA Langley Research Center Hampton, Virginia
September 1993 - June 1994	Engineering Aid Crew/Vehicle Integration Branch NASA Langley Research Center Hampton, Virginia
September 1993 - May 1994	Teaching Assistant Department of Psychology Christopher Newport University Newport News, Virginia
June 1993 - September 1993	Volunteer Service Program Participant Crew/Vehicle Integration Branch NASA Langley Research Center Hampton, Virginia
February 1993 - September 1993	Medical Research Service Participant Department of Veterans Affairs Medical Center Hampton, Virginia

GRANT SUPPORT

July 1995 - June 1999	NASA Graduate Student Researchers Program Training Grant NGT-1-52107

PUBLICATIONS

Burt, J. L., Bartolome-Rull, D. S., Burdette, D. W., and Comstock, J. R. (1999). A psychophysiological evaluation of the perceived urgency of auditory warning signals. In N. A. Stanton & J. Edworthy (Eds.), Human Factors in Auditory Warnings. Brookfield, VT: Ashgate Publishing.

Burt, J. L., and Casali, J. G. (1997). An evaluation of the urgency, similarity, and identification of aural alerts with implications for flight deck use. In B. Bidanda & S. Jagdale (Eds.), 1997 IERC Proceedings: The Sixth Annual Industrial Engineering Research Conference (pp. 731-736). Norcross, GA: Institute of Industrial Engineers.

PUBLICATIONS (continued)

Burt, J. L., Bartolome, D. S., Burdette, D. W., and Comstock, J. R. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. Ergonomics, 38(11), 2327-2340.

CONFERENCE PRESENTATIONS

*Burt, J. L., and Casali, J. G. (May, 1997). An evaluation of the urgency, similarity, and identification of aural alerts with implications for flight deck use. Paper presented at the 6th Industrial Engineering Research Conference, Miami Beach, Florida.

Burt, J. L., Casali, J. G., and Prestrude, A. M. (April, 1997). An evaluation of the urgency, similarity, and identification of aural alerts with implications for flight deck use. Poster session presented at the 9th International Symposium on Aviation Psychology, Columbus, Ohio.

*Burt, J. L., Bartolome, D. S., Burdette, D. W., and Comstock, J. R. (September, 1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. Paper presented at the Human Factors in Alarm Design II Conference, Southampton, United Kingdom.

*Burt, J. L., Bartolome, D. S., Burdette, D. W., and Comstock, J. R. (May, 1994). A psychophysiological evaluation of the perceived urgency of auditory warning signals. Paper presented at the 1994 Annual Meeting of the Virginia Academy of Science, Harrisonburg, Virginia.

Burt, J. L. (April, 1994). Possible ramifications of the 21st Century's impending demographic explosion from a psychological perspective. Paper presented at the 1994 Annual Meeting of the Virginia Collegiate Honors Council, Ashland, Virginia.

Reig, T. S., Burt, J. L., Contakes, L. M., Chiou, A. L., and Aravich, P. F. (April, 1994). Central serotonergic depletion exacerbates activity-based anorexia. Poster session presented at the 1994 Annual Meeting of the Eastern Psychological Association, Providence, Rhode Island.

*Associated printed works listed previously.

NATIONAL HONORARY SOCIETIES

Alpha Chi National Collegiate Honor Society
Alpha Pi Mu National Honor Society for Industrial Engineering
Psi Chi National Honor Society for Psychology