The Perceived Urgency and Detection Time of

Multi-Tone and Frequency-Modulated Warning Signals

In Broadband Noise

by

Ellen Carla Haas

Dissertation submitted to the faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctorate of Philosophy

in

Industrial and Systems Engineering

APPROVED:

Dr. John G. Casali, Chairman

Dr. Robert D. Dryden

Prof. Paul T. Kemmerling

Dr. G. Richard Price

Dr. Harry L. Snyder

March, 1993

The Perceived Urgency and Detection Time of

Multi-Tone and Frequency-modulated Warning Signals

In Broadband Noise

by

Ellen Carla Haas

Committee Chairman: John G. Casali
Industrial and Systems Engineering

(ABSTRACT)

In some environments, there is a serious mismatch between the perceived (psychoacoustic) urgency of a warning and its situational urgency. In addition, many auditory warnings are not detectable within their environments. This research examined several prominent pulse parameters which affect the perceived urgency and detection time of auditory warning signals. These elements included pulse format (multitone sequential, multitone simultaneous, and rising sawtooth frequency-modulated pulse formats), pulse level (65 dBC and 79 dBC), and time between pulses (0 ms, 150 ms, and 300 ms). The environments of interest were those settings with steady-state broadband machinery noise. Conditions included a loading task which presented additional attentional demands upon the subject during the signal detection task. Free-modulus magnitude estimation

quantified the relationship between auditory signal parameters and changes in perceived urgency. The method of paired comparisons was used to compare the perceived urgency of the auditory stimuli. Simple reaction time measured signal detectability. Signal effects were analyzed using a multivariate approach.

Results indicated that there was a small but statistically significant relationship between perceived urgency and detection time. As perceived urgency increased, detection time decreased. Both perceived urgency and detection time were influenced by pulse level and format. The higher pulse level resulted in a greater perceived urgency of the signal and shorter detection time. Sequential signals were rated as less urgent than the other pulse formats, and subjects took longer to detect their occurrence. Under most conditions, there was no significant difference in the perceived urgency or detection time of simultaneous and frequency-modulated pulses. Time between pulses (inter-pulse interval) affected only perceived urgency, not detection time. The shorter the time between pulses, the greater the perceived urgency of the signal.

ACKNOWLEDGEMENTS

I would like to express my gratitude to some of the people who contributed to this work. I would first like to thank my Chairman, Dr. John Casali, for his knowledgeable guidance and direction in the development and conduct of this research. I would also like to thank Dr. Richard Price, Chief of the Army Research Laboratory Visual and Auditory Processes Branch, for his wise and patient guidance. Dr. Harry Snyder deserves many thanks for his assistance with statistical and psychophysical matters. I would also like to thank Dr. Robert D. Dryden and Prof. Paul T. Kemmerling for serving on my graduate committee. A special thanks is due to Dr. Mark Hofmann, previous Associate Director of the US Army Human Engineering Laboratory. Without his interest and encouragement, this research would not have been possible.

I would like to express my appreciation to Dr. Joel Kalb, who developed the much of the software and fabricated some of the essential computer interface hardware for my experiment. He also provided invaluable assistance in teaching me the hands-on aspects of laboratory research. I would also like to thank the other members of the US Army Research Laboratory Auditory Processes Team, Ms. Kim Aboucharcra, Mr. Georges Garinther, and Ms. Pamela Mundis, for providing assistance in answering my many acoustical and audiological questions.

Dr. Gary Reid, of the Human Engineering Division of the US Air Force Armstrong Laboratory at Wright-Patterson Air Force Base, and Mr. Mark Crabtree of Logicon Technical Services of Dayton, Ohio, deserve grateful acknowledgement for the Criterion Task Set hardware and software support they provided.

Lastly, I would like to thank my husband, Gary Haas, for his patience, love, and understanding during this work. He gave me the support and incentive to finish this dissertation. I would also like to thank my sons, Jesse and Daniel, and my mother, Marscha Winterfield, for their loving support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Background

Auditory warning signals have found their way into many working environments. They have an advantage over visual warnings and displays in that they usually capture one's attention regardless of where one is looking. When designed correctly, auditory warning signals can improve performance and reduce accidents (Edworthy, 1991).

Despite many experimental studies concerning the detectability of warning signals in different background noises (Adams and Trucks, 1976; Patterson, Lower, Cosgrove and Milroy, 1989; Trucks and Adams, 1975; Wilkins and Martin, 1980), many auditory warnings are not optimally detectable within their environments. Every year, accidents occur in noisy workplaces because a warning signal is not heard (Wilkins and Acton, 1982). In other instances, auditory warnings have been too loud and usually so distracting that people turn them off rather than use them (Patterson, 1982).

In some environments, there is a serious mismatch between the perceived (psychoacoustic) urgency of a warning (the implicit urgency as a function of its sound parameters) and its situational urgency (the degree of urgency which the operator has learned to associate with the warning as a function of the situation itself) (Edworthy, 1991). This mismatch has been demonstrated in aircraft (Patterson, 1982) and in operating and recovery rooms in hospitals (Momtahan and Tansley, 1989).

Patterson (1982) and Edworthy (1991) proposed similar warning signal design methodologies to eliminate some of these problems. Basically, they recommended that warning signal construction be carried out in four stages: the specification of appropriate intensity levels, followed by the design of a small pulse of sound, the incorporation of this pulse into a longer burst of sound, and the formation of a complete warning using bursts of sound.

Issues In the Study

Edworthy (1991) used the methodology described above to construct an urgency matching, in which the effects of combinations of sound parameters on perceived urgency could be used to build a detailed, testable, and usable description of the relationship between auditory warnings and perceived urgency. She found that a wide variety of acoustic pulse and burst parameters had clear and consistent effects on the perceived urgency of auditory warnings and that subjects showed a high level of agreement about the urgency of such warnings.

The Patterson (1982) and Edworthy (1991) studies were by no means exhaustive; the effect of some signal parameters on perceived urgency has not been investigated. Variables which have not been fully explored include but are not limited to pulse format, pulse level, and time between pulses.

It appears that no researcher to date has used the Edworthy (1991) design methodology to explore the detectability of auditory warnings. The effect of combinations of spectral and temporal signal parameters on signal detectability has yet to be determined. The relationship between signal detectability and perceived urgency has not yet been defined.

Listeners in military and industrial environments are often engrossed in alternative tasks and are not specifically attentive to warning signals. To date, there is insufficient evidence in the literature to clearly define the effect of listener attentional demands on signal detectability.

Relevance to Military and Industrial Interests

The detection of various warning signals in industrial and military environments has been investigated by many researchers. Kemmerling, Geiselhart, Thornburn, and Cronburg (1969); Randle, Larson and Williams (1980); Patterson, Milroy and Barton (1980) and Doll and Folds (1986) studied warning signals in civil and military aircraft. Lambert (1980) investigated warning signals in military shipboard machinery noise. Abel, Kunov, and Pichora-Fuller (1985); Laroche (1989); Talamo (1982); and Forshaw (1977) explored the detection of warning signals in industrial environments. To date, no research has focused upon signal detection time and perceived urgency in steady-state ambient noise associated with both military and industrial environments.

Pink noise is a steady-state noise which can be considered broadly representative of machinery in military and industrial environments. Pink noise is a noise for which the spectral pressure density is inversely related to frequency, with a pressure spectral level slope of - 3 dB/octave (American National Standards Institute (ANSI), 1973). The pink noise spectrum contains equal energy in each octave band. It is broadband, not sharply tuned (no predominant frequency), and produced easily. Pink noise, because it includes low frequencies, provides a reasonable approximation of the machinery noise spectrum in many industrial environments (Adams and Trucks, 1976; Botsford, 1970; Johnson and Nixon, 1974; Karplus and Bonvallet, 1953; Michael, 1982). The pink noise spectrum is utilized in the computation of the noise-reduction rating for hearing protection devices because of its high correlation with industrial noise (Michael, 1982).

Intent of the Study

In general, the focus of this study was to examine several prominent elements which affect the perceived urgency and detection time of auditory warning signals. These elements included pulse format, pulse level, and time between pulses. The environments of interest were settings with steady-state broadband machinery noise.

With this purpose in mind, the following literature review contains a discussion of the physical parameters of sound, warning signals and ambient noise, and a study of the design of warning signals used in environments with broadband machinery noise. Factors affecting warning signal detection time and perceived urgency are discussed. Various methods for measurement of signal detection time, perceived urgency, and workload inducement are reviewed. Next, the experimental design and research methodology are presented, followed by the results, discussion, recapitulation, and suggestions for future research.

The Physical Characteristics of Sound Signals

Sound waves. The sensation of sound is defined as auditory sensation that is evoked by an oscillation in pressure or by particle displacement in a medium with elasticity or viscosity (ANSI, 1973). The medium may consist of liquids, solids, air, or other gasses. A disturbance within the medium, such as the abrupt displacement of one part of the medium, can be transmitted or propagated to another part of the medium by means of a "chain reaction," where momentum is transferred from particle to particle, and particles tend to "spring" back to their original positions. The medium as a whole does not move; only the disturbance is passed along, like a wave. Sound waves in air are strictly longitudinal, with particle motion along the axis of the movement of the disturbance. Thus, sound may alternately be defined as a physical disturbance in a medium with elasticity and density, involving an undulatory motion of the particles such that they vibrate back and forth along the axis of propagation (Durrant and Lovrinic, 1984).

Sound may be described in terms of different physical parameters. These parameters include frequency, amplitude, duration, and waveform. Three of these parameters (frequency, waveform, and duration) are concerned with or are limited by time.

Frequency. Frequency, measured in hertz (cycles per second, Hz), is the number of wave cycles completed per unit time (ANSI, 1973). The unit of measure was given its name in honor of the physicist Heinrich Hertz. A pure tone is a waveform consisting of one sinusoidal function or frequency. A complex tone is a sound wave containing multiple simple sinusoidal components of different frequencies. If a complex wave is repetitive, it is called periodic and the repetition rate is the fundamental frequency.

Amplitude. Amplitude describes the magnitude of wave displacement. Amplitude can be described in terms of peak amplitude, which is the difference between the maximum or minimum instantaneous value and equilibrium. Root mean square amplitude is often used as a descriptor of overall magnitude. It is the square root of the mean of the deviation values squared, where the deviations are the instantaneous values above and below the mean (average), which generally is zero (Durrant and Lovrinic, 1984). Sound amplitude is commonly quantified as a ratio of pressures, expressed in decibels.

Duration. The duration of sound is the time interval between the instant the excitation rises above a stated fraction of its maximum (for purposes of this study, 1%) and the instant it decays to this fraction (ANSI, 1973). Rise time is defined as the duration of the waveform envelope's increase from 10% to 90% of its final steady value, and fall time is the duration of the envelope's decrease from 90% to 10% of its final steady value (Richards, 1976). Onset and offset are similar to signal rise and fall time. For purposes of this experiment, onset is the duration from the start of the waveform's increase from zero (the start of the sound) until it reaches maximum (100%) output. Offset is the time during which the sound output falls from maximum output to zero output.

Waveform. The waveform of a sound is the sound pressure of a sound wave plotted as a function of time. The waveform envelope may be described as the curve connecting successive peaks above and below atmospheric pressure. There is an infinite variety of waveforms, including sinusoidal (pure tone), sawtooth, triangular, and white noise.

Neurophysiological Processing of Auditory Warning Signals

Anatomy and physiology of the ear. The normal ear consists of the outer, middle, and inner ear. The outer ear consists of the pinna and the external auditory meatus or auditory canal. These components modify the acoustic wave so that the spectrum of sound impinging on the eardrum is not quite the same as the sound that originally reaches the pinna. Weiner and Ross (1946) found that the head, pinna, and ear canal amplify environmental sounds in the 2-4 kHz region by 10 to 15 decibels (dB).

The middle ear consists of the eardrum (tympanic membrane), the small bones or ossicles of the middle ear (malleus, incus, and stapes), and the air cavity around the ossicles (bulla), which opens to the Eustachian tube. The eardrum vibrates in response to pressure fluctuations of the sound waves in the auditory canal. The vibration is then transmitted through the small bones to the stapes. The stapes footplate fills the oval window at the entrance of the fluid-filled inner ear. The transformer motion allows some of the energy entering the ear via the eardrum to be transmitted to the motion of the stapes, and hence to the stimulation of the inner-ear system.

The inner ear is comprised of the cochlea and the semicircular canals. The auditory portion of the inner ear is the cochlea. It derives its name from its snail-shell shape, having a structure which appears somewhat like a coiled tube. The vibrations of the stapes are transmitted to the oval window, which opens onto the scala vestibuli. Pressure waves in the cochlea propagate to the round window, causing a wave-like displacement of the basilar membrane and the structures attached to it. The pattern of movement of the basilar membrane depends on the frequency of the stimulus.

The organ of Corti, known as the hearing organ, rests on the basilar membrane. The tectorial membrane lies on top of the fine hairs that extend from each hair cell in the organ of Corti. Deflection of the basilar membrane causes a shear force to be developed between the tectorial membrane and the hairs, causing a shearing displacement of the hairs. When this occurs, neural impulses are developed in one or more of the 31,000 sensory neurons that are connected to the hair cells.

The sensory neurons are the fibers of the VIIIth cranial (auditory) nerve to the brain. This represents only the first synaptic stage in the passage of auditory information toward the auditory cortex in the brain. More detailed discussion of the anatomy and physiology of the human ear and the auditory system may be found in books by Durrant and Lovrinic (1986) and Pickles (1988).

<u>The processing of complex auditory signals</u>. The first stage of the analysis of incoming sound is performed by the spatial distribution of the resulting displacements of the hair cells of the cochlea. However, complex sounds must undergo further processing at higher brain centers to provide information regarding signal characteristics. It is not possible within the confines of this dissertation to discuss all the mechanisms which might be involved in the processing of complex sounds. However, it is possible to describe some of the systems which play a major role in the processing of complex auditory stimuli such as warning signals.

The brain stem may accomplish routine stages of stimulus processing, leaving the central processing unit, the auditory cortex, free for higher-level processing. These routine tasks may involve high-speed processing, such as sampling periodicities in the incoming signal. There is also evidence that higher order neurons in the brain stem tend to emphasize peaks of waveforms, which makes it possible to extract or emphasize certain features of complex sounds.

The inferior colliculus and the medial geniculate receive inputs from several high-order fibers. These systems provide alternative pathways which may provide a means by which incoming signals can be autocorrelated, or compared against themselves. By creating these delays, such aspects as the redundancy of incoming information could be evaluated, making it possible to determine if a sound were changing in amplitude over time.

The auditory cortex has a major role in processing complex stimuli, such as speech and complex tones which comprise warning signals. Green (1976) and Durrant and Lovrinic (1986) summarized several studies which describe the effects of cortical destruction on the behavior of animals. They found that bilateral ablation of the auditory cortices caused deficits in the performance of discrimination of tonal pattern, discrimination of sound duration, and localization of sounds in space. However, response to the onset of a sound, as well as response to changes in intensity, frequency, and location, were performed at the subcortical level, without the auditory cortex. Green and Durrant and Lovrinic concluded that the greater the complexity of the sound, and thus the more information contained in it, the greater the extent to which the cortex is likely to be involved in the processing of auditory information.

Pickles (1988) supported this evidence, stating that the auditory cortex may fulfill several functions, including analyzing complex sounds and identifying stimuli on an absolute basis. The auditory cortex is also necessary for the discrimination of auditory temporal patterns and in the formation of short-term memory when one auditory stimulus has to be related to another later in time.

Warning Signals

Wilkins and Martin (1980, 1987) categorized auditory warnings as intentional or incidental. Intentional warnings are generated by devices built for the specific purpose of generating warning sounds. These devices include sirens, bells, and horns. Incidental warnings are not deliberately generated, but arise in conjunction with a dangerous event or process and act as a warning either due to specific context or change in character. These warnings may include sounds made by a malfunctioning piece of machinery, an approaching truck, an impending roof fall in a coal mine ("roof talk"), or the loosening of a die key in drop forging ("the ringing of the keys"). Only intentional warning signals will be discussed here due to their relevance to the issues in this study. For more details on incidental warning signals, the reader is referred to Wilkins and Martin (1980). Intentional warnings may be categorized as speech or non-speech signals.

Speech signals. Because speech is highly redundant, speech signals present a redundant repetition of the warning information, which greatly reduces the possibility of confusion (Patterson, 1982). Speech signals are most desirable when listeners must identify a message source, when they are without special training in coded signals, or if situations of stress might cause them to "forget" the meaning of a code. Speech is preferred when there is a necessity for rapid two-way exchanges of information or when the message deals with a future time requiring some preparation, such as a missile countdown, in which tonal signals could be miscounted (Deatherage, 1972).

Kemmerling, Geiselhart, Thornburn, and Cronburg (1969) found that speech signals, in the form of voice warning systems, are more advantageous than tone warning systems because they provide direct information that enables the listener to take immediate corrective action or to evaluate a failure in terms of mission and safety requirements before acting. Voice warning systems were found to be beneficial under high task-load conditions because they give the listener the flexibility to evaluate a failure before acting.

Speech signals are not advantageous for all environments or applications. Patterson found that although voice warnings are easy to learn and difficult to forget or confuse, they take a relatively long time to present their information. In addition, they require a large dynamic range because vowels are often 30 dB more intense than higher frequency, lower power consonants. For environments where multiple speech warnings are utilized, specific voice warnings may be confused. Patterson (1982) recommended that voice warnings be used to support rather than replace complex tone signals, especially in the case of urgent or immediate-action warnings.

Non-speech signals. Non-speech signals, such as those containing tonal components, should be utilized when greater simplicity is desired, when listeners are trained to understand coded signals, or when immediate action is desired. These signals should be used in conditions unfavorable for receiving speech, such as when speech communication channels are overloaded or if speech will mask other speech signals or annoy listeners for whom the message is not intended (Deatherage, 1972).

Non-speech tonal signals can be categorized as consisting of single or multiple tones. Single-tone signals consist of one pure tone presented during the duration of the signal. Multiple tone signals, also called multitone complexes, consist of two or more tones presented during the signal duration. The tones can have consecutive, simultaneous, or delayed presentations, as well as harmonic relationships.

Frequency-modulated signals are a type of multitone signal in which tones are presented sequentially over a range of frequencies. In frequency modulation, the frequency of a baseline or carrier wave is varied in a manner determined by a modulating wave. The modulating wave defines the manner in which the signal's frequency varies over a specified range. Modulating waveforms include sawtooth, triangular, sine, and square waves.

Patterson (1982) stated that multitone and frequency-modulated signals are superior to single-tone signals in that they can utilize distinctive temporal and spectral patterns to reduce confusion with ambient noise and with other signals. Warning signals with distinctive temporal patterns are less likely to be confused. When the spectral content of the signal contains different components spread throughout the spectrum, it creates a distinctive sound and lessens spectrally based confusions. The spectral content of the complex signal can be defined to permit signal identification and recognition even in changing background noise. In addition, well designed complex signals could be used virtually simultaneously, because the bursts of sound could be interleaved with minimum confusion.

Ambient Noise

Noise is defined as any undesired sound. It may also be described as an erratic, intermittent, or statistically random oscillation which may be steady, nonsteady, or impulsive (ANSI, 1973). Ambient noise is defined as being the all-encompassing noise associated with a given environment, usually a composite of sounds from many sources near and far. Ambient noise may also be categorized as impulse, nonsteady, or steady-state (ANSI, 1973).

Impulse noise. Impulse noise is defined as a noise of transient nature such as that due to impact or explosive bursts (ANSI, 1973). In the United States Department of Defense military standard (MIL-STD-1474C) regarding noise limits for Army materiel, (U.S. Department of Defense, 1979), impulse noise is defined as being a short burst of acoustic energy consisting of either a single impulse or a series of impulses. The pressure history of a single impulse may include a rapid rise to a peak pressure, followed by a somewhat slower decay of the pressure envelope to ambient pressure, both occurring within 1 second. A series of impulses may last longer than 1 second.

Nonsteady noise. Nonsteady noise is noise with or without audible tones, for which the level varies substantially during the period of observation (ANSI, 1973).

Steady noise. Steady-state noise has negligibly small fluctuations of level within the period of observation (ANSI, 1973). Steady-state machinery noise is of most relevance to this experiment.

The Design of Warning Signals Used in Military and Industrial Environments

As mentioned earlier, Patterson (1982) and Edworthy (1991) proposed design principles for warning signal construction. First, they suggested specifying the appropriate loudness level of the signal by assessing the environment in which the signal would occur. Next, they suggested designing a small pulse of sound, which would act as a basic "building block" of the signal. A pulse is a sound contained within one amplitude envelope, which has an onset and offset and a specific duration. This pulse would be repeated several times, with intervals of silence between pulses. The resultant unit is referred to as a burst of sound. The burst forms the basis of a complete warning sound.

There is a paucity of research-based recommendations for the design of warning signals. Most of these are found in Patterson (1982) and in Edworthy (1991). Although Patterson (1982) provided numerous recommendations regarding level, temporal, and spectral factors, he furnished no discussion of research results to support his guidance. Likewise, no discussion of research results has been found in subsequent publications (Patterson, Lower, Cosgrove, and Milroy, 1989). In general, the recommendations provided by Edworthy (1991) are supported by experimental research.

The design of warning signals utilized in military and industrial environments requires an assessment of the background noise spectrum of the environments of interest. The design of the signals themselves involves level, temporal, and spectral factors. These factors are described below.

As was stated above, steady-state machinery noise can be represented in terms of acoustic patterns with common spectral characteristics. A pink noise spectrum (equal sound pressure levels in octave bands) can be considered broadly representative of machinery (Karplus and Bonvallet, 1953; Michael, 1982).

Sound pressure level. The level of the warning signal should be clearly audible above the ambient noise. Patterson (1982) recommended that warnings should be 15 dB or more above masked threshold to ensure that they will be noticed, and no more than 30 dB above masked threshold to avoid disruption of verbal communication. Masked threshold is defined as the threshold of audibility for a specified sound in the presence of another (masking) sound (ANSI, 1973). The International Standard for Auditory Danger Signals for Work Places (ISO, 1986) recommended that the sound pressure level of the signal should exceed the masked threshold by at least 13 dB in one 1/3 octave band or more within the frequency range of the signal, or that the A-weighted sound level of the signal exceeds the level of ambient noise by 15 dB or more. For ambient noise above 110 dBA, danger signals in other modalities (e.g., visual signals) should accompany the auditory signals. To prevent listener startle reactions (alarm, fright, or surprise which may result in a quick involuntary movement), there should be no unexpected steep increase in the sound level of the signal. To this end, Patterson (1982) recommended that rate of rise in the region above threshold not be greater than 1 dB/ms.

Temporal characteristics. The amplitude envelope of the waveform carries what is thought of as the time-related or temporal information of a warning signal. Relevant parameters of the amplitude envelope include rise and fall times, pulse duration, pulse repetition rate, and time between sound pulses.

Pulse rise and fall times ranging from 20 to 30 ms are recommended (Patterson, 1982). This is slow enough to prevent listener startle reaction, and also ensures that the duration of the entire pulse will be as brief as possible, without adding an undue length to total pulse duration. The onset and offset of the pulse should be concave down, or at worst linear, and should allow no overshoot of the steady-state level. A quarter-sine function is recommended for the onset and offset (Patterson, 1982).

The total pulse duration (including onset and offset) should be at least 200 ms to allow the ear enough time to integrate the warning signal. Patterson (1982) recommended that the pulse duration not be overly long to prevent disruption of communication which may be taking place. Patterson did not suggest a specific upper limit for pulse duration.

Pulsating signals are preferred to signals which are constant over time (ISO, 1986; Patterson, 1982). Pulses should be repeated at a rate in the range of 0.2 to 5 pulses per second (ISO, 1986). Pulse patterns can also affect the distinctiveness of the sound, which influences signal recognition (Patterson, 1982).

The inter-pulse interval is the time between the termination of the offset of one pulse and the beginning of the onset of the next. This is an important factor in warning signal design because it affects the level of interruption of ongoing events in the signal environment, as well as the perceived urgency imposed by the warning sound (Edworthy and Loxley, 1990; Hellier and Edworthy, 1989, 1990). Patterson (1982) recommended that a greater sense of perceived urgency could be communicated by decreasing the temporal spacing between pulses. He recommended that an inter-pulse interval of 0 ms be used to convey a sense of extreme urgency, and that a spacing of 300 ms be used to convey a perception of little or no urgency. Unfortunately, no published research results were used to support these recommendations. He recommended that specific inter-pulse intervals be established empirically for each signal.

Spectral characteristics. Spectral characteristics describe parameters relevant to signal frequency. These include the frequency range, the number of pure-tone components, and the harmonic structure of the components.

Auditory warning signals should incorporate components with frequencies which are efficiently processed by the auditory system. Patterson (1982) recommended that the frequency range should take into account listeners with noise-induced hearing loss, which most often appears in the higher frequencies (i.e., above 3000 Hz), as well as the low-frequency content of many background noises, which can mask signals of low frequencies (i.e., below 500 Hz). Researchers and standards offer slightly different recommendations regarding frequency range. Patterson suggested a range of 500 to 4000 Hz. The International Standard Organization suggests a range of 300 to 3000 Hz (ISO, 1986). The U.S. military human factors engineering standard for Army materiel (MIL-STD-1472C, 1989) suggests a frequency range between 500 and 3000 Hz. This latter standard recommends that the signal should have sufficient energy in the frequency range below 1500 Hz to accommodate persons with hearing loss and persons wearing hearing protectors. Noise Reduction Rating data summarized by Berger (1986) indicate that the average range of real-world attenuation provided by hearing protectors is 10-12 dB for earmuffs, less than 10 dB for non-foam earplugs, and approximately 10 dB for EAR and Decidamp brand foam earplugs (Berger, 1986).

To increase signal recognition, the center frequency of the octave band where the signal has the highest intensity should differ from the center frequency of the octave band where the ambient noise is highest (ISO, 1986). The center frequency of the signal should be lower than that of the noise to prevent upward spread of masking, which occurs when low frequency noise or tones effectively mask high frequency tones (Durrant and Lovrinic, 1984). It is less likely, although not impossible, for high frequency noise to mask low frequency signals (Durrant and Lovrinic, 1984). Remote masking is the phenomenon in which a high intensity band of noise raises the threshold of audibility for sounds lower in frequency (ANSI, 1973).

Patterson noted that the greater the number of components in the warning sound, the greater the potential for making the sound distinctive, which would contribute to better recognition of different warning signals.

Within the frequency range described above, warning signals should have at least four components and these should be harmonically related (Patterson, 1982). Multi-component signals are advantageous because the greater number of components offers greater resistance to masking. Although Patterson performed no published research, he theorized that harmonic tones in a multi-component signal that have a precise pitch will be more cohesive and less resistant to disruption to extraneous maskers. Patterson felt that signals with inharmonic components sound more diffuse and provide less resistance to masking.

Factors Affecting Warning Signal Detection Time and Perceived Urgency

For the purposes of this experiment, signal detection time was operationally defined as the time between the onset or start of an auditory signal (when the amplitude of the signal first rises above zero) and the response recorded from the subject (touching the button on a keypad). An operational definition of perceived urgency is the subject's direct numerical judgement of the sensory magnitude produced by the warning signal, as obtained using free-modulus magnitude estimation. Factors believed to affect warning signal detectability and perceived urgency include signal, noise, and listener characteristics.

<u>Signal characteristics</u>. The detectability and perceived urgency of warning signals can be influenced by temporal and spectral signal characteristics such as level, duration, time between pulses, and format. Other variables include onset and offset, rhythm, and harmonic structure.

One characteristic which may affect both signal detectability and perceived urgency is pulse level. However, there is little research to describe the manner and extent of this relationship. Teichner (1954) found that reaction time to an auditory stimulus becomes shorter, up to an asymptotic value, as the intensity of the stimulus is increased. He did not describe the effects of intensity for signals greater than 200 ms in length. Although Edworthy (1991) stated that there may be a strong relationship between signal intensity and perceived urgency, there is no published literature which describes this relationship.

Bock, Lazarus, and Hoege (1982) investigated the perceived dangerousness of different warning signals as a function of level of signal against background noise. Warning signals were presented at different signal/noise ratios to an ambient pink noise. They found that the perceived dangerousness of certain warning signals, including bell, impulse sound, and square wave signals, was influenced by the level of the signal in the background noise. The temporal and spectral features of the warning signals were not described by the authors.

Edworthy (1991) stated that pulse duration had a "large and consistent effect" on perceived urgency. She found that shorter pulses were perceived as less urgent. However, these data were tentative; she did not specify the pulse durations which produced these effects. Although no data were available, Patterson (1982) suggested that pulse durations of 200, 350, and 500 ms may have a consistent effect on perceived urgency. Research has not yet been conducted to determine to what extent this range of pulse durations is related to the perceived urgency of warning signals.

There is some suggestion in the literature that stimulus duration affects simple reaction time to an auditory signal. Teichner (1954) found that reaction time decreases rapidly as the duration of the stimulus is lengthened from a value approaching zero (e.g., 7 ms) to some small time value (e.g., 200 ms). As signal duration increases above 200 ms, reaction time gradually decreases to become asymptotic to some limit. The specific duration at which reaction time becomes asymptotic was not defined.

Hellier and Edworthy (1989) found that the time between signal pulses had a consistent effect on the perceived urgency of warning signals. The smaller the inter-pulse interval, the greater the perceived urgency of the warning signal. Patterson (1982) recommended than an inter-pulse interval of 0 ms elapsing from the end of the offset of one pulse to the onset of the next, could be used to convey a sense of great urgency, and that an interval of 300 ms be used to convey a perception of little or no urgency. No researcher has as yet investigated the effect of this range of durations on the perceived urgency or detectability of warning signals.

To date, the perceived urgency of some formats of multitone signals (signals consisting of more than one pure tone) and frequency-modulated signals have not been explored or compared. These signals include multitone signals with components presented sequentially (one tone following another in time), as well as pure tones frequency-modulated by sawtooth and triangle waves. The perceived urgency of these signals should be explored because they can potentially offer resistance to masking and because their sound parameters can be easily manipulated (Edworthy, 1991;

Edworthy and Loxley, 1990; Edworthy, Loxley, Geelhoed and Dennis, 1989; Hellier and Edworthy, 1989; Patterson, 1982).

Edworthy (1991) observed that signals containing an onset and offset of relatively short but equal duration are perceived as being more urgent than signals with a relatively longer onset and offset. However, these conclusions are tentative. Edworthy did not clearly define the temporal length of the offset and onset times she used. In addition, the total durations of the signals were different. As was previously stated, pulse duration may influence the perceived urgency of warning signals.

Signals with a regular rhythm, in which all pulses are equally spaced, are perceived as being more urgent than signals in which pulses are syncopated (Edworthy, 1991).

Signals which contain some or all components which are not integer multiples of the fundamental frequency are seen as being more urgent than signals which consist solely of integer multiples of the fundamental frequency. Signals in which all harmonics play for the duration of the pulse are seen as being more urgent than those which contain some delayed harmonics which start at the second half of the pulse (Edworthy, 1991).

The larger the pitch range (i.e., 9 vs. 6 vs. 3 semitones), the greater is the perceived urgency of the signal (Edworthy, 1991).

Noise characteristics. Noise can mask warning sounds. The perception of signals that are disturbed by noise can be best described by the masked threshold (Lazarus, 1983). Theoretically, a warning sound will be audible if any frequency in the sound exceeds the critical signal/noise ratio with respect to the surrounding band of noise. The spectra and intensity of the ambient noise can affect the masked threshold of a signal.

Masking is the process by which the threshold of audibility of one sound is elevated in the presence of another sound. As was defined earlier, the elevated threshold of the signal in the presence of masking noise is called the masked threshold. The amount of masking can be quantified as the difference between the masked threshold and the threshold in quiet (Scharf and Buus, 1986). Since thresholds are specified in dB, masking is also quantified in dB. Masking is simultaneous when the noise and signal occur at the same time and nonsimultaneous when they occur at different times.

The focus of interest in this experiment was the simultaneous masking of a complex auditory warning signal by pink noise. The factors which affect the detectability of this type of signal in the presence of a pink noise masker include the spectrum and level of the masker, as well as the spectrum and duration of the signal (Scharf and Buus, 1986).

When a signal is masked by a broadband noise, only a narrow portion of the noise centered around the signal frequencies is actually effective in masking the signal. Increasing the noise bandwidth beyond each frequency band will not appreciably increase the amount of masking. The theory which explains this occurrence is the critical band concept, originated by Fletcher (1940) and described by Scharf (1970). The critical band concept states that the masking of a tone by a broad-band noise is mainly determined by the frequency band of noise close to the frequency of the tone, which is called the critical band. When a tone is just audible above the noise, the total energy of the tone is equivalent to the energy within the critical band. More detailed discussion of the critical band theory can be found in the book by Scharf (1970).

Hawkins and Stevens (1950) carried out a comprehensive study of pure-tone masking by broad-band white noise. They found that at all measured frequencies (125-9000 Hz), increasing the level of a wide-band masking noise raised the masked threshold for a pure tone by the same dB amount across all frequencies.

Scharf (1959) and Scharf and Buus (1986) reported that for multitone complexes, masking by noise is constant regardless of the number of component tones as long as all added tones fall within a critical band. When components are added outside the critical band, the amount of masking decreases as the number and frequency separation of the components increase.

Garner and Miller (1947) found that signal duration can affect the detectability of a pure tone in noise. Auditory warnings should be temporally long, surpassing the integration time of the ear (approximately 200 ms or longer). For signals with durations below 200 ms, the auditory system integrates acoustic energy linearly over time. Sensitivity increases linearly as duration increases; the signal-to-noise ratio at threshold necessary to maintain signal audibility decreases linearly by approximately 10 dB for every tenfold increase in stimulus duration. Increasing signal duration beyond 200 ms brings a nonlinear increase in signal detectability. Beyond 200 ms, sensitivity increases very slowly, and increasing signal duration beyond 1000 ms brings an extremely small improvement in signal detectability.

As can be seen, the characteristics of the ambient noise present during the occurrence of the warning signal can affect the detectability of that signal. No researcher has investigated the effect of noise parameters on perceived urgency.

Listener attentional demands. Listeners in military and industrial environments are often engrossed in alternative tasks and are not specifically attentive to warning signals. To date, the influence of attentional demands on listener perceived urgency has not been investigated. There is some research which deals with listener attentional demands and warning signal detectability.

Some researchers feel that the perception of warning signals remains unaffected by the lack of attention. Durlach and Colburn (1978) stated that "the fact that nature did not provide us with earlids is probably due to .....the use of the acoustical channel for warning signals, a function to which it is exceptionally well matched". Wilkins and Martin (1980, 1987) supported this philosophy when they found that listener inattention did not reduce the detectability of an unexpected but important warning sound. In their experiment, listener inattention was created by the presence of a loading task as well as uncertainty as to the time of occurrence of the warning sound. They concluded that for an appropriately chosen warning sound, there is normally sufficient spare attentional capacity for the warning to be detected.

Kreezer (1959) and Potter, Fidell, Myles, and Keast (1977) found that attentional demands did affect the detection of warning signals. They found that listener inattention elevated the masked threshold of the signal by an average of up to 9 dB. Inattention was created by the presence of a loading task and uncertainty as to the time of occurrence of the warning. However, methodological differences in determining masked threshold, as well as differences in motivating subjects, could have produced the elevation of threshold in this study (Wilkins and Martin, 1987).

The results presented in these studies may have limited applicability. The majority of the researchers (Potter et al., 1977; Wilkins and Martin, 1980, 1987) used sirens as warning signals, which are distinct and familiar to most listeners. Therefore, the results may not apply to typically indistinct incidental warning sounds or to warnings which consist of a gradual change in an existing sound. These researchers used loading tasks as well as temporal uncertainty of signal occurrence to create operator inattention. Thus, effects due to operator loading alone cannot be discerned.

In a controlled experiment, Casali and Wierwille (1983) found that attentional demands caused by communications loading did affect the subject detection of verbal signals (aircraft call signs). These researchers examined several workload estimation techniques in a simulated flight task which emphasized communications load. Subjects were exposed to three different levels (low, medium, and high) of communications loading. Loading level differed by the rate of call sign presentation (message traffic density) and the number of extraneous call signs presented to the subject. Subjects were instructed to respond to specific "target" call signs by pressing a switch and making a verbal response to acknowledge detection. The detectability measure was the response time measured from the end of the spoken call sign message to the beginning of the subject's correct "now" response. The data indicated that subjects showed a significantly shorter response time in low-load conditions than in high-load conditions. This indicates that the detectability of verbal signals is affected by listener attentional demands.

Listener characteristics. Listener characteristics which influence the perceived urgency and detectability of signals include listener hearing threshold level, occlusion, and experience.

An individual's sensitivity to sound is expressed in terms of "Hearing Threshold Level" (HL), which is the dB level relative to a standard audiometric threshold at which a tone of specified frequency is heard by an ear in a specified number of trials (ANSI, 1989). The sensitivity to sound, or the audibility of the warning signal to the listener, is a basic consideration in the perception of a warning sound in the presence of noise. To date, no researcher has investigated the effect of hearing threshold level on the perceived urgency of warning signals. On the other hand, the effects of hearing threshold level on signal detection have been explored.

Abel, Kunov, Pichora-Fuller, and Alberti (1985) investigated signal detection in quiet and in industrial noise for subjects with normal hearing and noise-induced hearing loss (NIHL). Signal detectability was tested in conditions with and without hearing protection. Subjects included persons with noise-induced hearing loss (NIHL) of 35-85 dB HL at 1 kHz, and at both 1 and 3 kHz. Subjects with NIHL were further categorized into two groups, those with hearing loss due primarily to exposure to steady-state noise and those with hearing loss due primarily to exposure to impulse or intermittent noise. The measure of detectability in this experiment was the detection threshold for a signal. Two types of signal were used; the first consisted of a one-third octave band noise centered at 1 kHz, and second was a one-third octave band noise centered at 3.15 kHz. Specific information regarding the signal spectra was not furnished. Signals were presented in quiet and in backgrounds of industrial noise (mill house or rock drill noise). The detection threshold was defined as the sound pressure level at which the subject was able to identify the presence of the signal 75% of the time in a two-interval forced-choice task.

In quiet conditions, without hearing protectors, subjects with NIHL had a significantly higher detection threshold than did subjects with normal hearing. When no hearing protectors were worn in the presence of industrial noise, subjects with NIHL had the same detection threshold as those with normal hearing. Thus, for unprotected listening in noise, subjects with NIHL performed at the same level as did those with normal hearing. The authors offered no satisfactory explanation for these results. Abel et al. (1985) also found that there was little difference in detection threshold between subjects with NIHL at 1 and 3 kHz and those with NIHL at 1 kHz. There was little difference in detection threshold in NIHL subjects who were exposed to steady-state noise versus those exposed to impulsive or intermittent noise. Abel et al. also showed that there was little difference in detection threshold between older and younger subjects with normal hearing.

When EAR plugs fit binaurally by the experimenter were worn, subjects with NIHL had higher detection thresholds in both noisy and quiet conditions. Subjects wearing hearing protectors in noise demonstrated detection thresholds greater than 100 dBA, which approached the limits of deafness. The combination of NIHL and hearing protection elevated the listeners' hearing threshold level which, in turn, impaired their ability to detect warning sounds.

Lazarus (1983) conducted a similar study, involving a typhon (sic) signal masked by four different noises. The signal and noise spectra were not described by the author. Masked threshold was the pulse level at which 50% of the signals were heard. Lazarus reported that the correlation between the masked threshold and hearing loss was between $\underline{r} = 0.74$ and $\underline{r} = 0.85$. If the HL was less than 20 dB, the masked threshold while wearing plugs was equal to or minimally lower than that for unprotected ears. However, when the HL exceeded 20 dB, the masked threshold was lower when the ear was unprotected. In general, for all four noise conditions, the masked threshold increased with increasing hearing loss when no ear protection was worn. For subjects with hearing loss, the masked threshold decreased more sharply when earplugs were worn.

Listeners in industrial and military environments often wear some type of hearing protection device as part of a hearing conservation program. Lambert (1980), Wilkins and Martin (1987), and Forshaw (1977) reported that many listeners in both military and industrial environments thought that the wearing of hearing protection impaired auditory detection of warning signals. Wilkins and Martin reported one such survey of 80 workmen in industry, in which more than half thought that wearing hearing protectors made it more difficult to hear warning sounds.

Wilkins and Martin (1987) summarized 11 separate studies in which various researchers investigated the audibility of signals by subjects with normal hearing, while wearing hearing protectors. All but two of them reported that for people with normal hearing, the wearing of hearing protectors had no significant adverse effect on the detection of intentional and incidental signals presented in various types of masking noise. At the higher noise levels

(85 dBA and above) for which they are normally required, the protectors actually improved the audibility of the warning sounds by approximately 3 dBA relative to the unoccluded ear condition. The improved audibility is due to the protectors providing a reduction in both the signal and noise intensities at the ear, reducing the overall sound energy, which permits the cochlea to respond without distortion. As a result, detection performance approaches that found in low noise conditions (Lawrence and Yantis, 1956).

Lambert (1980) found that hearing protectors did have a significant effect on the detectability of warning signals in the presence of masking noise. He found that the 9AN/2 earmuffs commonly worn on Navy ships did interfere with the ability to hear incidental warnings (machinery sounds) in a shipboard environment. The earmuffs had a substantial adverse effect in engine room background noise, which had variable temporal characteristics and a strong low frequency content. The reduced ability to hear incidental warnings may have occurred because the muffs allowed the passage of low-frequency noise, causing upward masking of the higher-frequency incidental warnings (although there is no real proof of this occurrence in the study). Other laboratory and real-world data (Berger, 1986) indicate that earmuffs provide poor attenuation in environments with significant low-frequency energy and that foam or premolded earplugs might have provided better attenuation in that particular environment. However, too much attenuation will ultimately reduce detection in such environments.

Listeners with impaired hearing were also studied. Wilkins and Martin (1987) suggested that when hearing protectors are worn by people with an existing noise-induced hearing loss, their elevated absolute threshold may further impair their ability to detect warning sounds. The results of Forshaw (1977) support this, as do some studies of speech intelligibility during the wearing of hearing protection (Abel, Alberti, Haythornthwaite and Riko, 1982; Coles and Rice, 1965). Casali and Horylev (1987) found that for people with moderate (less than 30 dBHL) hearing loss, the use of hearing protectors with higher attenuation (e.g., earplugs rather than earmuffs) reduced speech discrimination by reducing the speech level to near or below threshold, but not below the threshold of those with better hearing. Coles and Rice suggested that the use of hearing protectors in hearing warning signals may be more disadvantageous than is the case with the hearing of speech, because warnings do not contain the amount of redundant information found in speech. They further suggested that persons with hearing loss who wear earplugs will have an impaired ability to perceive indicator sounds, which against a background of noise are usually identified by their high-frequency content.

Listener experience may influence the detectability and perceived urgency of warning signals. Experience may be described in terms of listener familiarity with the signal as well as uncertainty as to time of signal occurrence and expectancy of occurrence. To date, the effect of listener experience on perceived urgency has not been investigated. Research in this area has dealt only with the effect of listener experience on warning signal detectability.

Temporal uncertainty, which is uncertainty as to time of signal occurrence, can affect the detectability (reaction time) to that signal. Wickens (1984) noted that if the occurrence of a stimulus is highly predictable in time, reaction time to that stimulus will be short. However, if the stimulus is not predictable (if there is a long or variable waiting interval before the appearance of the stimulus), then the reaction time to that stimulus will be longer.

Listener expectancy, which is the expectation that a signal will occur, also influences the reaction time to that signal. Drazin (1961) found that reaction time responses which follow long intervals preceding the stimulus presentation tend to be faster than those following short intervals.

Listener familiarity has been shown to have little influence on the detectability of simple signals in noise. Experiments with simple sinusoidal sounds have shown signal detectability to be relatively independent of the listener's knowledge of what the next sound would be (Creelman, 1973; Green 1961). The effect of complex sounds may be different. Experienced listeners appear to discriminate between complex tonal stimuli they have encountered previously, or which they expect or are directed to look for, more accurately than they do with stimuli that are unexpected or unfamiliar (Watson et al., 1976). However, there has been little research dealing with the effect of user familiarity on the detectability of complex tones.

Objective Methods for Measurement of Signal Detectability

Masked threshold. The masked threshold is the threshold of audibility for a specified sound in the presence of another (masking) sound (ANSI, 1973). An operational definition of masked threshold is the minimum stimulus level which will elicit a response half (50%) of the time when the stimulus is presented in the presence of the masking noise.

Engen (1971a) noted that Fechner developed three psychophysical methods to measure absolute and difference thresholds. These are the methods of constant stimuli, limits, and adjustment. Each of these methods consists of an experimental procedure and a mathematical treatment of data. Because this experiment was not concerned with the determination of absolute thresholds, they will not be discussed further.

Subject reaction time. Methods for obtaining masked threshold are concerned with the determination of absolute thresholds in the presence of masking noise. However, in military and industrial environments, signals are often presented above the masked threshold. International design standards commonly require that signals be presented at 15 to 30 dB above masked threshold (ISO, 1961). The U.S. Department of Defense (1989) design standard MIL-STD-1472D requires that signals be presented at least 20 dB above masked threshold. A practical approach to measuring the detectability of signals occurring above masked threshold would be to measure subject reaction time from the onset of the presentation of different auditory signals which are close to the masked threshold. Reaction time should be longer for less detectable signals.

Reaction time, also called response latency, is defined as the latency of the subject's response measured from the onset of the stimulus (Teichner and Krebs, 1972). The major types of reaction time task are simple and choice reaction time tasks. Simple reaction time tasks involve the presentation of one stimulus and require one uniform response with no alternative responses. Choice reaction time tasks involve the presentation of at least two different stimuli, calling for at least two different responses (Teichner, 1954). Because the task of interest in this experiment involves the uniform response to single auditory signals, simple reaction time was the objective measure of interest in this experiment.

Simple reaction time tasks offer many advantages. They are easy to perform and demand little intellectual effort on the part of the subject, and from the experimenter's perspective, are easy to measure. In addition, simple reaction time tasks have been researched extensively. Woodworth and Schlosberg (1954) summarized reaction times to various tasks, including simple response to auditory stimuli after practice. They found that a typical reaction time to an acoustic stimulus was 140 ms for a practiced adult.

Stimulus intensity appears to influence response speed in simple reaction time tasks. In summarizing reaction time studies, Teichner and Krebs (1972) found that simple reaction time is inversely related to the intensity of the stimulus, up to an asymptotic value. This relationship has been demonstrated for signals in many modalities, including the auditory modality, and may be considered to be well established. There are many theoretical explanations of this phenomenon. Grice (1968) presented one response criterion model in which the individual's response criterion is represented by an accumulated number of neural impulses so that the greater the intensity of the stimulus, the greater the accumulation of information and the sooner the criterion is reached.

Subjective Methods for Measurement of Perceived Urgency

In the domain of psychophysics, numerous scaling methodologies have been invented to provide useful measures of subjective concepts. These methods include magnitude estimation and paired comparisons.

Magnitude estimation. Magnitude estimation is a scaling methodology useful for the determination of ratio scales of apparent magnitude. Ratio scales are useful because they contain the characteristics of order, distance, and origin while retaining maximal correspondence with the number system. In addition, they permit the ratio of one sensation magnitude to another to be specified (Gescheider, 1985).

In magnitude estimation, the observer is required to make direct numerical estimations of the sensory magnitudes produced by various stimuli. The two primary methodologies for applying magnitude estimation are prescribed (fixed) modulus and free modulus magnitude estimation (Stevens, 1957). In fixed modulus estimation, the experimenter defines a modulus by presenting a given stimulus and assigning it some particular value. On subsequent trials, when other stimuli are presented, the observer assigns numbers to his or her sensations relative to the value of the modulus. The observer makes his or her judgements reflect how many times greater one sensation is than another, by judging the ratio between the two sensations.

In free modulus magnitude estimation, the subject is not presented with an experimenter-defined modulus. The stimuli are presented to the observer, who assigns numbers in proportion to the perceived stimulus magnitudes. Gescheider (1985) reported that assigned magnitude and free magnitude scales are in close agreement. Stevens (1971) recommended free modulus estimation because he felt that it was better to permit the observer to choose his or her own modulus than to designate one for him or her. Gescheider

(1985) noted that many investigators agree with this philosophy.

Magnitude estimation is currently one of the most frequently used psychophysical scaling methods (Gescheider, 1985). It has become popular for several reasons. The method is convenient because the subject brings the numbers with him or her, so to speak, and the experimenter needs only to provide the target stimuli to which the numbers are to be matched (Stevens, 1971).

Another advantage is that little training is involved. Stevens (1971) found that untrained, inexperienced college students seem to do as well as those who have had many years of practice. He stated that since there is no right or wrong to the subjects' responses, there is no clear need for training. Stevens suggested, however, that when subjects are to estimate ambiguous stimuli, that subjects first be allowed to perform magnitude estimation on an easier continuum, such as the apparent length of lines or the apparent size of circles.

A third advantage is that judgements can be obtained relatively rapidly. Depending upon the nature of the stimulus, all the data for a single observer can be obtained in one or two sessions. Stevens suggested that stimuli should be presented to the subject for two judgements at the most, because once a subject has learned to recognize a particular stimulus, little or no new information is obtained from subsequent judgements of its repeated presentation. He suggested that biases due to range and spacing of stimuli seem to have less influence when the subject is limited to one or two judgements per stimulus. Because of the need for few observations per data point, magnitude estimation is a valuable tool for extensive experiments which vary several parameters of the stimulus.

Several researchers utilized magnitude estimation in measuring the perceived urgency of auditory warnings (Edworthy, 1991; Edworthy and Loxley, 1990; Edworthy, Loxley, Geelhoed and Dennis, 1989; Hellier and Edworthy, 1989, 1990). Edworthy (1991) found that urgency is a meaningful and salient psychological construct. Subjects using urgency as a construct rated stimuli in a consistent manner both within and between subjects, and had clear ideas about which stimuli were more urgent than others. Hellier and Edworthy (1990) found that Stevens' Power Law was an adequate descriptor of the relationship between perceived urgency and various signal parameters such as speed of signal (systematic change in inter-pulse interval from the start to the end of the burst) and number of signal repetitions.

Free modulus magnitude estimation was used in this study to quantify the relationship between auditory signal parameters and changes in perceived urgency. Because free modulus estimation resulted in subject ratings varying over different number ranges, subject scores were transformed and corrected. One such procedure (Engen, 1971b) is described in the data analysis section. Because the transformed data are ratio data, parametric statistical analyses were performed to determine which signal parameters were statistically significant in producing changes in perceived urgency.

The method of paired comparisons. Fechner (1860) hypothesized that stimulus differences which are detected equally often are subjectively equal, regardless of the physical values involved. Thurstone (1927) hypothesized a mathematical model, the Law of Comparative Judgment, which consists of a set of equations relating the proportion of times any given stimulus is judged greater on a given attribute than any other stimulus to the scale values and discriminal dispersions of the two stimuli on the psychological continuum.

The method of paired comparison is an experimental procedure which utilizes the Law of Comparative Judgment.

Engen (1971b) noted that it was first introduced by Cohn (1894) in his study of color preferences, and then developed further by Thurstone (1927). Because it is often regarded as the most appropriate way of securing value judgments (Engen, 1971b), it was used in this study to compare the perceived urgency of each auditory stimulus with that of every other auditory stimulus.

The method of paired comparisons is a generalization of the two-category case of the method of constant stimuli. In the method of constant stimuli, each stimulus is compared with a single standard. In the method of paired comparisons, each stimulus serves as the standard and each stimulus is paired with every other stimulus. With n stimuli, there are n(n-1)/2 pairs. Each pair is presented to the subject, whose task is to indicate which member of the pair appears to have the greater amount of the attribute to be scaled. The subject must designate one of the pair as greater. No equality judgments are allowed (Torgerson, 1958).

The method of paired comparison can be applied to any stimulus material for which pairs can be presented. The inclusion of too many pairs can make the task tedious for the observer. One can usually overcome these problems by using overlapping ranges (Engen, 1971b).

Torgerson (1958) noted that the paired comparison methodology does not include provision for changes in performance due to fatigue or practice effects, or for judgements based in part on factors other than the relative magnitudes of the discrimination processes. He recommended that most of these factors can be controlled by randomizing the relative positions and order of presentation of stimuli.

Methods for Measurement of Workload

Numerous methods for the measurement of workload have been proposed, devised, and tested. Excellent and thorough reviews of these methods have been offered by many researchers (Chiles and Alluisi, 1979; Ogden, Levine and Eisner, 1979; Williges and Wierwille, 1979). There have been some attempts to combine workload theory and task developments into a unified workload assessment technique. One such effort by the U.S. Air Force resulted in the development of a mental workload metric evaluation tool, the Criterion Task Set (CTS) (Shingledecker, 1984).

The CTS, which is implemented in software on a microcomputer system, was designed to provide a set of standardized loading tasks to evaluate the relative sensitivity, reliability, and intrusiveness of a variety of available workload measures. It is based on a model which represents a synthesis of the multiple resource theory (Wickens, 1984) and processing stage theory (Sternberg, 1969) of the human system. Schlegel and Gilliland (1990) describe the CTS model, which is shown in Figure 1, as involving three primary stages of processing: perceptual input, central processing, and motor output. There are specific mental processing resources associated with the input mode (either visual or auditory), the type of coding during central processing (either spatial/imaginal or abstract/symbolic), and the mode of response output (either manual or vocal). The central processing stage is further divided into working memory functions (encoding, storage, and recall) and central activity functions (information manipulation, reasoning, planning, and scheduling). Shingledecker noted that no universal consensus presently exists regarding an appropriate theory on human performance and that the CTS model was intended to act only as a descriptive summary of state-of-the-art research findings and conceptual approaches.

The CTS battery consists of nine different tasks, each of which is designed to place primary demands upon a single resource within the CTS model. The probability monitoring task was chosen for this experiment because it is a continuous workload task which utilizes the visual modality

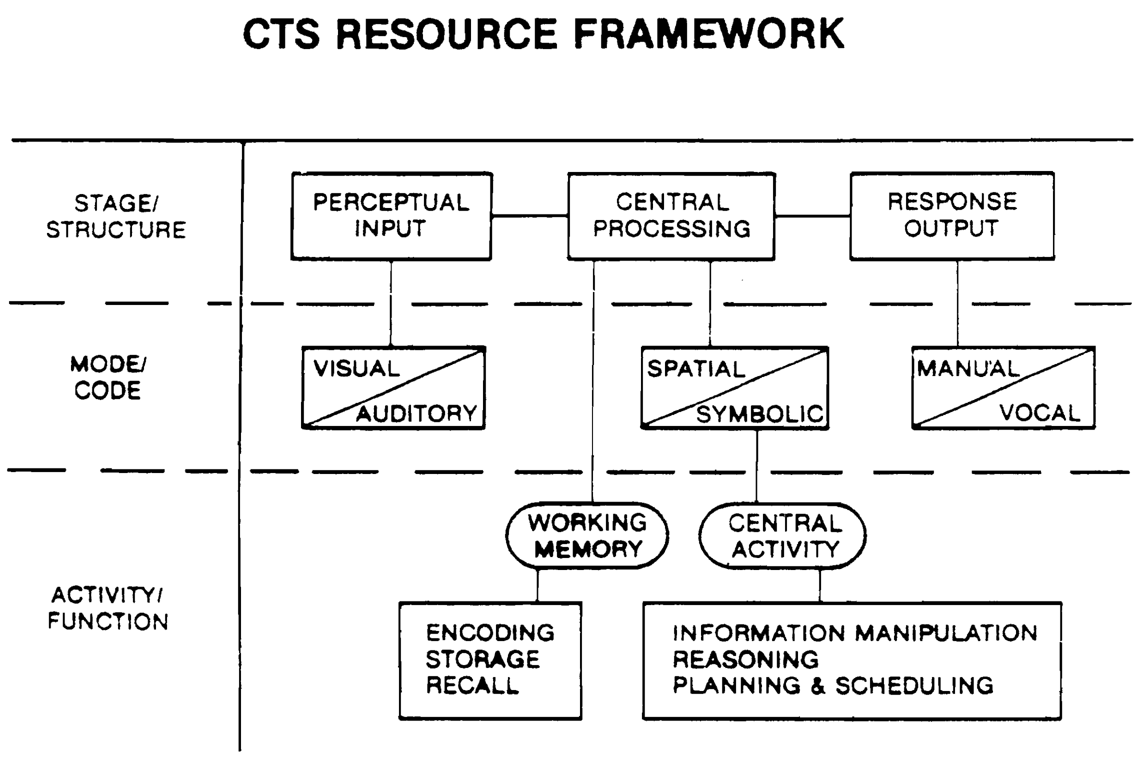


Figure 1. Criterion task set model (adapted from Schlegel and Gilliland, 1990.

and does not require auditory input. The auditory modality is thus preserved exclusively for the detection of acoustic signals.

The probability monitoring task was designed to place demands on the perceptual input stage of processing (Amell, Eggemeier, and Acton, 1987). This task requires the detection of simple visual signals embedded in visual "noise." The subject is required to monitor 1, 2, or 3 simulated dials with moving pointers. Under nonsignal conditions, the pointers move randomly. When a signal occurs, the movement becomes biased so that the pointer spends 95 percent of the time on one side of a center mark on the dial. The subject is instructed to press the appropriate response key to correct the biased dials to a non-bias condition. The performance measures on this test are reaction time to correctly identify signals, number of false positives, and number of misses.

Reaction time to correctly identify signals. This measure is the time required for the subject to respond to the bias by pressing the correct button on a multiple-button keypad when the bias condition is present. Measurement begins at the onset of the bias condition and ends when the subject presses the button on the keypad. Responses are timed to the nearest 0.1 second.

Number of misses. This measure is the number of overlooked signals when the bias is present. A miss occurs when the subject does not press a button on the multiple-button keypad when a bias is present.

Number of false positives. The number of false positives is the number of subject responses to the bias condition when no bias is present. A false positive occurs when the subject presses the keypad when a bias is absent.

Evaluation of performance measures. Research conducted at the Air Force Armstrong Aeromedical Research Laboratory (AAMRL) indicated that the reaction times, missed signals, and subjective workload were significantly different for all three levels. The analysis of the data for the missed signal measure revealed a significant main effect for difficulty level, indicating that the low workload level was significantly different from the high workload level (Amell, Eggemeier, and Acton, 1987).

Research Needs and Objectives of the Study

In summary from the previous literature review, it is clear that research voids exist in defining signal characteristics which affect both the perceived urgency and detection time of warning signals in military and industrial environments.

The Patterson and Edworthy studies were by no means exhaustive; the effect of some signal parameters on perceived urgency have not been investigated. The prominent voids in prior research give rise to a particular need for study of pulse format, time between pulses, and pulse level.

To date, the perceived urgency of most types of multitone signals (signals consisting of more than one pure tone) and frequency-modulated signals has not been explored or compared. These signals include sawtooth wave frequency modulations of sinusoids. Multitone signals with simultaneous and sequential components should be included as a basis for comparison. The perceived urgency of these signals should be determined because, with proper design, they offer resistance to masking and because their sound parameters can be easily manipulated to make the signal sound more urgent (Edworthy, 1991; Edworthy and Loxley, 1990; Edworthy, Loxley, Geelhoed, and Dennis, 1989; Hellier and Edworthy, 1989, 1990; Patterson, 1982).

Hellier and Edworthy (1990) found that the duration between pulses also has a consistent effect on the perceived urgency of warning signals. Patterson recommended that an inter-pulse interval of 0 ms could be used to convey a sense of great urgency and that an interval of 300 ms be used to convey a perception of little or no urgency. No researcher has as yet investigated how this range of durations influences the perceived urgency of warning signals.

Pulse level may affect both signal detectability and perceived urgency. However, there is little research to describe the manner and extent of this relationship. Although Edworthy (1991) stated that there may be a strong relationship between signal intensity and perceived urgency, there is no published literature which describes this relationship. The effect of signal intensity on the detectability of signals greater than 200 ms should be determined.

As was demonstrated in the literature search, no researcher has used the design methodology described by Patterson (1982) and Edworthy (1991) to assess the detectability of auditory warnings. In addition, the relationship between signal detectability and perceived urgency has not yet been defined.

The present study examines some important selected elements which are likely to affect the perceived urgency and detectability of auditory warning signals which occur in military and industrial environments with steady-state ambient noise. The specific research aims were to:

1. Investigate the effect of pulse format, pulse level, and time between pulses on the perceived urgency of warning signals.

2. Investigate the effect of pulse format, pulse level, and time between pulses on the detectability of warning signals.

3. Examine the relationship between the perceived urgency and detectability of warning signals.

The research attempts to distinguish the burst parameters which provide a specific level of perceived urgency and detectability in conditions representative of those encountered in certain military and industrial environments. The purpose of this is to provide a detailed, testable, and usable description of the relationship among auditory warnings, perceived urgency, and detectability. The purpose is also to permit an urgency/detectability mapping of signals, in which the most urgent situations could be signaled by the most urgent and most detectable warning sounds.

Hypotheses

The general hypothesis was that the perceived urgency and detection time of auditory signals would be influenced by temporal, level, and spectral characteristics.

It was hypothesized that there would be at least one independent variable which would affect both perceived urgency and detection time. Pulse sound pressure level would be one such variable. Signals with greater pulse levels would be more detectable and would be rated as more urgent than those with lower pulse levels. It was suggested that this would occur because signals with greater amplitude are more difficult to mask. Therefore, the more conspicuous the signal, the more urgent it would tend to sound.

It was expected that at least one variable would influence perceived urgency. Signals with shorter between-pulse intervals would be rated as more urgent, due to their apparent rapidity. In terms of signal detectability, it was proposed that subjects would rate frequency-modulated signals as more detectable than the multitone signals, because the larger number of tonal components in the modulated signal would provide greater resistance to masking in broadband noise.

It was expected that the paired comparison data would indicate that at least one signal would be rated as more urgent than another. Signals with shorter between-pulse intervals and higher pulse levels were expected to achieve higher rankings than those with longer between-pulse intervals and lower pulse levels. A high, positive correlation between paired comparison rankings and magnitude estimation ratings was also expected.

Finally, it was hypothesized that this research would demonstrate the means by which to construct an urgency/detectability mapping, in which the relationship among auditory warnings, perceived urgency, and detectability could be described. This mapping should provide insights as to which signal parameters provide a definable level of perceived urgency and detectability. This insight would enable the design of appropriate warning signals which provide the desired level of perceived urgency and detectability for specific military and industrial environments.

EXPERIMENTAL METHOD AND DESIGN

Subjects

Thirty-six male and female military personnel employed by the Department of the Army at Aberdeen Proving Ground, Maryland, constituted the paid subject population for the study. Eighteen males and 18 females were used. Subject ages ranged from 18 to 22 years, with a mean age of 20 years.

Subjects had "unimpaired hearing," which is defined as an audiometric hearing threshold level (HL) of less than or equal to 15 dB HL in both ears (Nicolosi, Harryman, and Kresheck, 1989) at pure tone frequencies of 500 to 8000 Hz in octave steps. Also, subjects had no bilateral hearing differences of more than 15 dB. These qualifications were verified using pure-tone audiometry.

Test Facility and Apparatus

The test facility was housed in the Auditory Performance Laboratory (Building 520) within the Human

Research and Engineering Directorate of the U. S. Army Research Laboratory at Aberdeen Proving Ground, Maryland.

Test booths. An audiometric booth was used for audiometric screening. A semi-reverberant chamber was used for warning signal testing.

The pure-tone audiometric screening tests were conducted in an Industrial Acoustics Corporation (IAC) 1200-A Series sound-attenuated booth with double-walled construction. The booth meets the specifications found within ANSI (1978), "Methods for Manual Pure-Tone Audiometry."

An IAC 400-A Series semi-reverberant chamber was used as the test booth for all warning signal testing. The inside dimensions of the chamber are 4.67 x 3.45 x 2.49 m. Chamber reverberation time (T_{60}) was measured (Table 1), using an integrated impulse decay methodology (Schroeder, 1965) programmed into a Type 830 Norwegian Electronics Real-Time Analyzer. Chamber ambient noise levels are presented in Table 2. A Realistic Minimus-7 loudspeaker was positioned at the ear height of a 50th percentile subject (U.S. Department of Defense, 1989), 0.762 m from the floor at the right rear corner of the chamber, facing the center of the chamber. The speaker generated a nearly diffuse

TABLE 1

Reverberation Times (T_{60}) for the Semi-Reverberant Chamber

1/3 Octave Band Center Frequency (Hz)	Reverberation Time (s)
63	1.44
80	1.50
100	1.68
125	1.16
160	1.54
200	1.02
250	1.10
315	1.24
400	1.34
500	0.98
630	1.06
800	1.14
1000	1.08
1250	1.14
1600	1.02
2000	1.16
2500	1.24
3150	1.14
4000	1.22
5000	1.10
6300	1.22
8000	1.16
10000	1.20
12500	1.18
dBA	1.16
dBC	1.20
dB (LINEAR)	1.32

TABLE 2

Ambient Noise Levels for the Semi-Reverberant Chamber

1/3 Octave Band Center Frequency (Hz)	Ambient Noise Level (Leq)*
31.5	30.8
40	27.6
50	28.6
63	29.8
80	25.7
100	17.4
125	20.0
160	7.4
200	0.7
250	- 3.1
315	- 1.1
400	- 2.2
500	- 2.2
630	- 1.4
800	- 0.8
1000	0.1
1250	1.2
1600	4.7
2000	3.2
2500	3.7
3150	5.1
4000	5.6
5000	6.8
6300	7.8
8000	8.6
10000	8.9
12500	9.6
dBA	17.3
dB (LINEAR)	35.9

*Measures obtained using Bruel and Kjaer 4144 pressure response microphone, real-time analyzer set at fast response, Leq sample period of 0.05 seconds. Negative ambient levels were those below the reference level (20 u Pa) sound field (a volume in which the sound pressure level assumes a nearly constant value, +/- 1.0 dB) in the vicinity of the subject's head. The subject was seated within this field, in the center of the chamber (approximately 2.82 m away from the loudspeaker), facing away from the loudspeaker. A plumb bob suspended from the ceiling was used to provide a reference for maintaining the listener's head in a constant position.

<u>Test apparatus</u>. Audiograms were administered using a Teledyne Avionics Model TA-20 Automatic Audiometer. Subjects wore Telephonics TDH-50P earphones with MX41-AR cushions. The printer in the audiometer system plotted dB hearing threshold level as a function of test frequency.

Warning signal functions were generated using Data Analysis and Digital Signal Processing Software by the DSP Development Corporation. An Ariel Corporation Signal to Disk Interface (SDI) system with digital signal processing capabilities was used to store and play the signals. The hardware and software was integrated onto two Zenith Model Z-248PC series computers.

The apparatus used to produce the warning signals and masking noise was a General Radio Corporation Model 1450 Decade Attenuator, an Altec-Lansing Model 1653A Graphic

Equalizer, an Altec-Lansing Model 1692A Mixer/Preamplifier, and an Altec/Lansing Model 1270B Power Amplifier. Warning signals and noise were presented to the subject using a Realistic Minimus-7 loudspeaker.

The 68 dBC pink masking noise was generated by a Type 830 Norwegian Electronics Real-Time Analyzer. The frequency range of the noise was 125 Hz to 12.5 kHz. The noise had equal sound pressure levels in all octave bands, as per ANSI (1973).

A one-button response keypad (Acton and Crabtree, 1985) was used to record subject response latency to the warning signal stimulus. The keypad was integrated onto a Zenith model Z-248PC series computer. Subject reaction time was measured to 0.001 s, and was recorded in a file on the Zenith computer.

Version 2.1 of the Criterion Task Set (CTS) software released by the Air Force Aerospace Medical Research Laboratory (Acton and Crabtree, 1985; Shingledecker, 1984) was used to generate the Probability Monitoring workload task. In order to provide a demanding workload task, the "high workload" task level was used. The software was run on a Commodore 64 computer, integrated with a Commodore 1541 disk drive and a Commodore 1802 color monitor. A multiple-button response keypad was used to record subject responses to the workload task. Probability monitoring task data were recorded on the Commodore computer.

The Modified Cooper-Harper (MCH) workload rating (Appendix A) was used to provide a general indication of the workload imposed by the CTS loading task. As suggested by Wierwille and Casali (1983), the MCH scale was reproduced such that the horizontal dimension was approximately 26.67 cm. The scale was fit on a standard 21.59 cm by 27.94 cm sheet of paper and the print on the scale was fully legible. Subject ratings were recorded directly on the sheet of paper containing the scale.

Test system calibration. The audiometer was calibrated with a Bruel and Kjaer Type 4153 artificial ear with a headphone coupler and a Bruel and Kjaer Type 1625 sound level meter with one-inch microphone. A Bruel and Kjaer Type 4230 standard source calibrator was used for sound level meter calibration at 93.8 dB.

The sound levels at the subject's ear height in the test chamber were calibrated using a Norwegian Electronics Type 830 Real-Time Analyzer, with a Bruel and Kjaer Model 1465 one-half inch microphone. 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.

Experimental Design

The experimental design (Figure 2) used for data collection and to structure the primary data analysis was a 3 x 3 x 2, full factorial, within-subjects (repeated measures) design. The 36 subjects were assigned to each experimental cell with an equal male/female split in each cell. All independent variables were treated as fixed-effect variables, and subjects were treated as a random-effect variable.

Independent variables. The three independent variables for the experimental design were pulse format, time between pulses, and pulse level.

Three **pulse formats** were used, all incorporating a frequency range of 500 to 3000 Hz. These were a multitone signal with four simultaneous components, a multitone signal with four sequential components, and a "sawtooth" frequency-modulated signal.

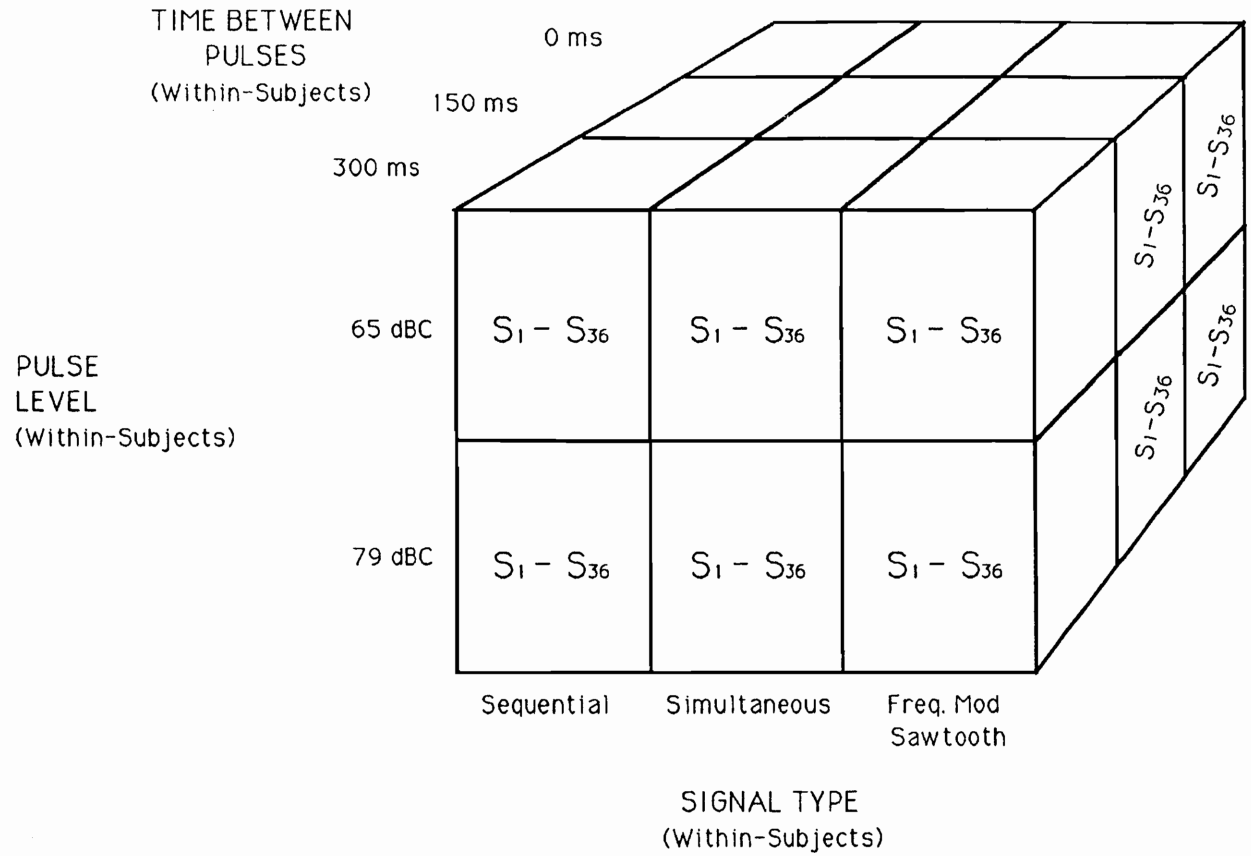


Figure 2. Experimental design matrix.

The multitone signal with four simultaneous components ("simultaneous pulse format" or "sim") was a signal consisting of four pure tones at 500, 1000, 2000, and 3000 Hz, presented simultaneously (concurrently) during one pulse duration. The amplitude of each component was equal. Because this type of signal had been used in other studies exploring perceived urgency (Edworthy, 1991), it was included in this experiment for purposes of comparison.

The multitone signal with four sequential components ("sequential pulse format" or "seq") was a signal consisting of pure tones at 500, 1000, 2000, and 3000 Hz, presented sequentially during one pulse duration. The amplitude of each component was equal. Each tone was one-quarter the duration of the pulse. This signal was included in this experiment because pilot study data indicated that listeners perceived this signal as relatively non-urgent.

The "sawtooth" frequency-modulated signal ("sawtooth frequency-modulated pulse format" or "saw") was a 500-Hz pure tone carrier frequency-modulated over one pulse duration by a positive sawtooth function. This produced a signal which travelled from 500 to 3000 Hz during one pulse duration. This signal was included in this study because pilot study data indicated that listeners perceived this signal as relatively urgent.

Time between pulses was the time elapsed from the end of the offset of one pulse to the onset of the next. The times were 0, 150, and 300 ms. Pulse onset was the time from the waveform's increase from zero amplitude until it reached maximum output. Pulse offset was the time during which the pulse output fell from maximum, or near maximum, to zero.

Pulse level was the rms sound pressure level of the pulse (in dBC) at the subject's ear, measured with the Norwegian Electronics real-time analyzer on "slow mode". The pulse levels were 65 and 79 dBC SPL, to ensure that subjects were able to discriminate the signals over the pink masking noise. Pilot study data indicated that these pulse levels corresponded to mean pulse sensation levels of 5 and 19 dB, in reference to the pulse's average threshold of audibility (masked threshold), in the presence of the 68 dBC pink noise masker.

The stimuli were 18 auditory signals, each signal consisting of eight pulses. Each pulse had a duration of 350 ms, including an onset time of 25 ms and an offset of 25 ms. The rise time and fall time of each pulse was 16 ms (measured from 10% to 90% of pulse amplitude). The stimulus parameters and frequency ranges corresponded with current research findings and design recommendations (Edworthy, 1991; Patterson, 1982) and design standards (ISO, 1961; U.S. Department of Defense, 1989).

Dependent measures. Three dependent measures were obtained. These were the magnitude estimation rating of the urgency of each signal stimulus, subject reaction time to warning signals, and the paired-comparison ranking of warning signal urgency. Because the study was not aimed at assessing changes in loading task performance due to changes in signal parameters, subject performance on the loading task was recorded to track workload changes, but were not included as a dependent variable.

Magnitude estimation data were obtained using the previously discussed free-modulus magnitude estimation method (Stevens, 1957). Two sets of magnitude estimation values were collected for each signal and ultimately averaged, using the methodology described by Engen (1971b). This procedure yielded one set of 18 magnitude estimation values per subject.

Subject reaction time data were obtained using the previously discussed simple reaction time methodology. Two sets of subject reaction times were collected for each auditory signal and ultimately averaged. This yielded one set of 18 reaction time values per subject.

Paired comparison rankings were obtained using the previously discussed paired-comparison methodology. Subject rankings were obtained for each pair of auditory signals. This yielded one set of 153 rankings per subject. The paired comparison data transformation (described in the forthcoming data analysis section) yielded one set of 18 signal rankings (one ranking per signal) for the experiment.

Experimental Procedure

Each participant underwent four sessions, all in a single visit to the laboratory. The first session was the screening session, in which the subject received an audiogram. This session entailed the qualification test for the subject's participation in the experiment. The next three sessions entailed experimental data collection for magnitude estimation, reaction time, and paired comparison data. After the third experimental data collection session, the subject received a final audiometric test. The protocol sequence and duration for the entire experiment appear in Table 3. The total duration of this experiment was 3.75 hours, including breaks, per subject.

TABLE 3

Protocol for the Experiment

	EVENT	ESTIMATED DURATION
1.	Screening session, including audiometry	30 min
2.	Magnitude estimation data collection session	30 min
3.	Break	10 min
4.	Reaction time data collection session	40 min
5.	Break	30 min
6.	Paired comparison data collection session	60 min
7.	Subject debriefing	10 min
8.	Follow-up audiometric test	15 min
	TOTAL:	3.75 hrs

The same 18 warning signal stimuli were used in all three data collection sessions. The order of presentation of signals was randomized independently across subjects and across sessions.

Screening session protocol. The screening session followed the order shown in Table 4. Each candidate for participation was welcomed and asked to read a brief overview description of the experiment. He or she was then asked to read and sign an informed consent document (Appendix B), indicating his or her willingness to participate. The experimenter then briefly explained the experimental procedures. The instructions and explanations for each data collection session are contained in Appendix C.

Next, each candidate filled out the Case History Form, on which he or she provided information regarding noise exposure history and history of hearing problems. This form is contained in Appendix D.

The candidate then had an audiometric screening test in accordance with ANSI (1978), using the audiometer in the IAC audiometric booth. The audiometer was used in an automatic modified Houghson-Westlake mode. For each ear, the mean hearing threshold at each of the pure-tone frequencies of 500, 1000, 2000, 4000, 6000, and 8000 Hz were determined by

TABLE 4

Protocol for the Screening Session

1. The experimenter read the Description of Experiment to the candidate. The candidate read and signed the Informed Consent document (Appendix B).

2. The experimenter read the Instructions for Audiometry (Appendix C) and administered the Case History Form (Appendix D). The candidate was seated in the audiometric chamber.

3. The candidate was audiometrically tested in both ears using pure-tone stimuli. A pure-tone automatic audiometer was used to administer an audiogram, using the Houghson-Westlake threshold testing procedure (explained in text).

4. If of "qualified hearing," the candidate (hereafter "subject") was asked to attend the experimental data collection sessions.

a Houghson-Westlake, or "5 dB up, 10 dB down" procedure (Morrill, 1984). In this procedure, from a reference level (typically 30 dB), the audiometer decreased the intensity in 10 dB steps until the candidate did not respond. When the candidate no longer responded, the audiometer then increased the intensity in 5 dB steps until he or she responded. Threshold at each frequency was taken as the level at which the candidate responded, with agreement on at least two of three trials. The candidate (hereafter "subject") was asked to participate in the experimental data collection sessions if he or she met the audiometric criteria explained previously in the "subject" section.

Magnitude estimation data collection session. The magnitude estimation data collection session immediately followed the screening session. The protocol for this session appears in Table 5.

At the beginning of the magnitude estimation data collection session, the subject was seated in the test chamber by the experimenter. The experimenter read instructions for the magnitude estimation rating task (Appendix C). After reading the instructions, the experimenter exited the chamber, and then presented the sample stimuli to the subject.

TABLE 5

Protocol for Magnitude Estimation Data Collection Session

1.	Subject was seated in the test chamber by experimenter. Experimenter read instructions for the magnitude estimation task.
2.	Magnitude estimation sample stimuli 1 through 10 were presented. Sample stimuli 6, 7, 8, 9, and 10 (explained in text) were played in the presence of the pink noise.
3.	Subject provided magnitude estimation ratings for all 18 auditory signals. Signals were presented in random order in the presence of the pink noise.
4.	Subject was given a five-minute break.
5.	Subject provided magnitude estimation ratings for all 18 auditory signals. Signals were presented in a different random order in the presence of the masking noise.
6.	At the close of session, the subject was given a 10-minute break.

The sample stimuli provided the subject with a concept of the range of variable values of the signal stimuli. At the same time, because the parameters of the sample pulses were different from the warning signal stimuli, the subject was prevented from obtaining a preconceived idea of the perceived urgency of the variables, or combinations of variables, which comprised the signal stimuli.

Sample stimuli 1 and 2 demonstrated time between pulses. Sample stimulus 1 consisted of four 1000-Hz tone pulses, 400 ms in duration, with 0 ms between pulses. Sample stimulus 2 consisted of four 1000-Hz tone pulses, 400 ms in duration, with 300 ms between pulses. The amplitude of each pulse was 79 dBC. The pulse onset and offset were each 25 ms in duration. To ensure that all subjects heard all inter-pulse intervals, these stimuli were presented without the pink noise.

Sample stimuli 3, 4, and 5 demonstrated pulse format. These stimuli were a multitone simultaneous signal, a multitone sequential signal, and a sawtooth frequency-modulated signal, respectively. Each sample stimulus consisted of four pulses. Each pulse was 500 ms in duration, with 300 ms between pulses, and a frequency range of 500-3000 Hz. The amplitude of each pulse was 79 dBC. The pulse onset and offset were 25 ms. To ensure that all subjects heard the signal pulses, these stimuli were presented without the pink noise.

Sample stimuli 6 and 7 demonstrated pulse level. Both samples were four, 1000-Hz tone pulses, 400 ms in duration, with 0 ms between pulses. To introduce the subject to the ambient noise conditions under which the experiment would be run, these stimuli were presented in the presence of pink noise.

Sample stimuli 8, 9, and 10 were utilized to determine whether the subject would hear the pulses at their lowest amplitude (65 dBC) when the pulses were presented over the masking noise of 68 dBC. These stimuli were a multitone simultaneous signal, a multitone sequential signal, and a sawtooth frequency-modulated signal, respectively. Each stimulus consisted of four pulses. Each pulse was 500 ms in duration, with 300 ms between pulses and a frequency range of 500-3000 Hz. The amplitude of each pulse was 65 dBC.

After the sample stimuli signals were presented, the subject verbally provided a magnitude estimation rating for each of the 18 experimental signal stimuli. The experimental signals were presented in random order, after which a 5-minute break ensued.

Following the break, the subject again provided magnitude estimation ratings for each of the 18 signal stimuli. The stimuli were presented in a different random order, as previously described. Following this, the magnitude estimation data collection session was terminated and the subject had a 10-minute break.

Reaction time data collection session protocol. The protocol for the reaction time data collection session which occurred next is shown in Table 6. Subject instructions for this session are presented in Appendix C.

At the beginning of the reaction time data collection session, the subject was seated at a table in the test chamber directly in front of the Commodore 1802 Color Monitor. The one-button keypad was placed on the table at the subject's dominant hand. The multiple-button keypad was placed on the table at the subject's non-dominant hand. Figure 3 depicts the placement of this apparatus.

The experimenter read instructions for the workload task. After hearing the instructions, the subject performed five 3-minute workload practice periods in the presence of the pink noise. The Criterion Task Set software presented approximately 10 bias conditions within each three-minute period (Amell et al., 1987). When the subject completed all

TABLE 6

Protocol for Reaction Time Data Collection Session

1. Subject entered the test chamber, was seated by the experimenter. Experimenter determined which hand was the subject's dominant hand, and used this information to position the multiple-button and one-button keypads.

2. Experimenter read instructions for the workload task. Subject performed five 3-minute workload task practice trials.

3. Experimenter read instructions for the reaction time task. Subject performed 10 reaction time trials.

4. Experimenter read instructions for simultaneous workload/reaction time task. Subject performed one 3-minute practice workload task in conjunction with 18 reaction time trials.

5. Subject performed workload and reaction time tasks simultaneously. Subject performed one 3-minute workload task and provided reaction time data for all 18 auditory signals. Signals were presented in random order in the presence of the pink noise.

6. Subject was given a five-minute break.

7. Subject performed workload and reaction time tasks simultaneously. Subject performed one 3-minute workload task and provided reaction time data for all 18 auditory signals. Signals were presented in random order in the presence of the pink noise.

8. Immediately following the performance of the workload and reaction time tasks, the subjects completed the Modified Cooper-Harper worksheet.

9. At the close of session, the subject was given a 30-minute break.

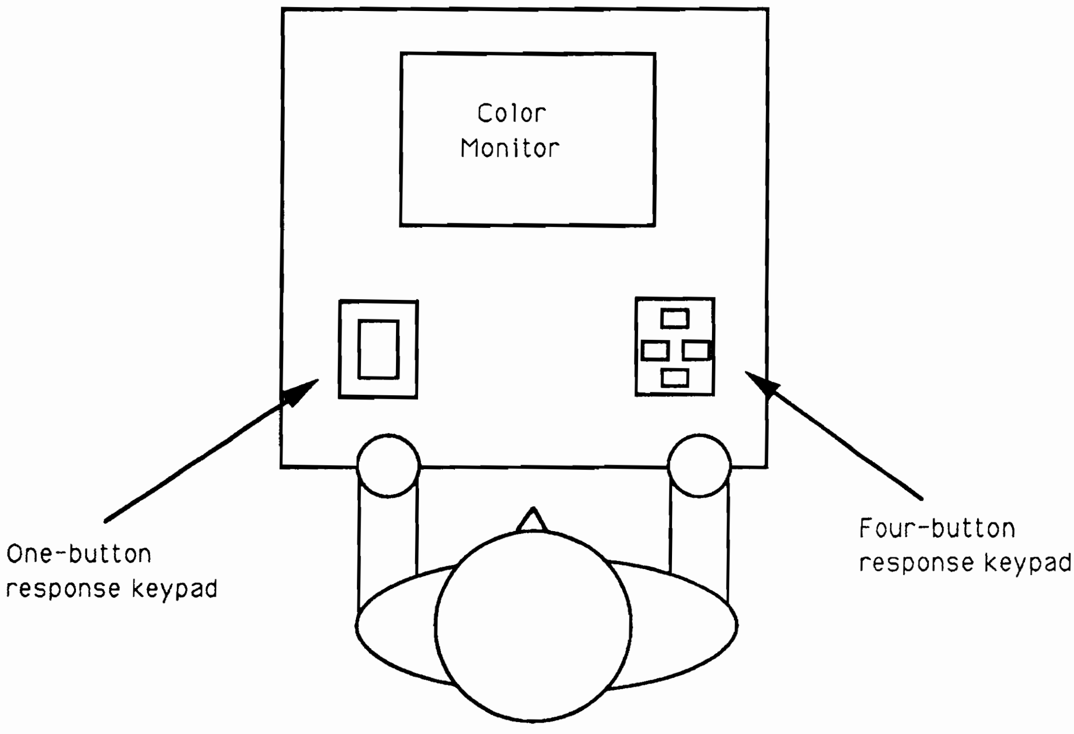


Figure 3. Placement of reaction time data collection apparatus (assuming dominant left hand).

five practice periods, the experimenter stopped the pink noise, re-entered the chamber, and presented instructions for the reaction time task.

After reading the instructions for the reaction time task (Appendix C), the experimenter then exited the chamber and activated the pink noise generator. The experimenter then presented 10 auditory signals at random intervals as practice trials. As was previously described, the subject provided reaction time measures by pressing the one-button response keypad with his or her dominant hand after the onset of the auditory warning signal. Reaction time measurement began at the start of the first pulse of the warning signal (when the amplitude of the first pulse was greater than zero) and ended when the subject pressed the key of the one-button response keypad.

After the subject completed the reaction time practice trials, the experimenter turned off the pink noise, re-entered the chamber and read instructions for the simultaneous performance of the workload and reaction time tasks (Appendix C). After the instructions were read, the experimenter exited the chamber and turned on the pink noise. The subject then performed a practice trial, consisting of one 3-minute workload task performed in conjunction with 10 auditory signal reaction time tasks. As before, the Criterion Task Set software presented approximately 10 bias conditions at random intervals per each three-minute period.

After the subject completed the practice trials, the experimenter turned off the pink noise, re-entered the chamber, and asked if there were any further questions. Once all questions were answered, the experimenter turned on the pink noise generator, and presented the workload and reaction time tasks in the same manner as for the aforementioned practice task. The Criterion Task Set software presented approximately 10 bias conditions at random intervals per each three-minute period. Approximately 18 auditory stimuli were presented at random intervals within the same three-minute period (the number is approximate because the randomness of the inter-signal time intervals did not always permit the presentation of all 18 stimuli in every three-minute period). All auditory signals not presented in the three-minute workload task were presented during an additional workload task presentation, run immediately afterwards. The workload task was halted immediately after the subject responded to the final auditory signal. After the subject completed the reaction time task for all 18 signal stimuli, he or she had a five-minute break.

Following the break, the subject again performed the workload and reaction time tasks described above. The subject provided reaction time data for each of the 18 signals, presented in a different random order from before.

Immediately after the completion of the workload and reaction time tasks, the subject provided a numerical rating of the mental workload involved in the detection time task (responding to both visual bias conditions and auditory signals) by completing the Modified Cooper-Harper workload scale. Subject instructions emphasized the use of the MCH decision tree in the rating process (Wierwille and Casali, 1983). Then the subject was given a 30-minute break.

As indicated in Table 6, each subject provided 36 reaction times (two for each signal). For each subject, the mean of the two reaction time values as calculated. In this manner, one mean reaction time for each of 18 signals was obtained for each subject.

Paired comparison data collection session protocol. The paired comparison data collection session took place immediately following the break, the protocol for which appears in Table 7. At the beginning of the paired comparison session, the subject was seated in the test chamber and the experimenter read instructions

TABLE 7

Protocol for Paired Comparison Data Collection Session.

---

1. Subject was seated in the test chamber by the experimenter, who then read the instructions for the paired comparison task.

2. Two paired comparison practice trials were presented.

3. Subject provided paired comparison data for 153 signal pairs. The order of signal presentation and the relative position within each signal pair were randomized. The signals were played in the presence of the pink masking noise.

4. At the close of session, the subject was debriefed.

(Appendix C) for the paired comparison task.

The experimenter then exited the chamber, turned on the pink noise generator, and presented two practice trials (two pairs of signals) to the subject. After finishing the practice trials, the subject then verbally provided paired comparison data for each of the 153 stimulus pairs (n(n-1)/2) for the 18 signal stimuli described in the Experimental Design section. The order of signal presentation and the relative position within each signal pair were randomized to control for biases due to fatigue, practice, or factors other than the relative magnitudes of the discriminal processes, as per the guidelines of Torgerson (1958). For each pair, the subject was instructed to verbally indicate which of the two signals was the most urgent. In the event that both signals sounded equally urgent to the subject, he or she was instructed to choose one or the other. After the subject provided data for all stimulus pairs, the paired comparison data collection session was terminated. The subject was then debriefed.

Follow-up audiometry protocol. To verify that the subject experienced no threshold shift during the course of the experiment, the subject received a second audiogram following the debriefing. The procedure for the second audiogram was identical to that used in the screening session. Pre-determined criteria for the occurrence of threshold shift were as follows. Subjects were re-tested with a pulsed tone if a hearing threshold level (HTL) greater than 10 dB occurred in any test frequency. If the audiogram indicated that the increase persisted, the subject received another pulsed-tone audiogram in 24 hours. If the HTL increase persisted after 24 hours, the subject was referred to an audiologist, who administered comprehensive audiological tests. The follow-up audiometry indicated that no subject experienced an HTL increase greater than 10 dB; thus, no experimental subject required a second follow-up audiogram or the intervention of an audiologist.

DATA ANALYSIS

Magnitude Estimation Data Transformation

Each subject provided 36 magnitude estimation ratings (two for each signal). To compare magnitude estimation judgements among subjects who responded using varied number ranges, every score for each subject required correction. A procedure originated by Lane, Catania, and Stevens (1961) with corrections added by Kalikow (1967), and described by Engen (with an error) (1971b) and Snyder (with a correction to the error) (H.L. Snyder, personal communication, July 31, 1991), eliminated inter-subject variance caused by differing choices of moduli and eliminated intra-subject variability. The procedure 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 signal stimulus. 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 signal parameters 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 stimuli.

(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 stimuli 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) Subtract the value obtained in step 6 from the row of values obtained for each observer in step 2. These values represent the logarithmic perceived urgency of each signal.

(8) Obtain the antilog of every value obtained in step 7. This effectively makes the unit of measure more meaningful and more straightforward to interpret (perceived urgency rather than logarithmic perceived urgency).

A very small number of magnitude estimation ratings (approximately 25 out of a total of 1,296) were "zero," which were impossible to convert to logarithms in the magnitude estimation transformation. Where one of the two ratings made by a subject for the same signal stimulus was zero, both ratings were replaced with the arithmetic mean of the two ratings, prior to the transformation. Where both stimulus ratings made by a subject for the same signal stimulus were zero, both ratings were replaced with a small positive number (1.0), prior to the transformation.

Only one subject gave zero magnitude estimation ratings for a large proportion (approximately 25%) of his responses, in most cases for the same signal stimulus. Because the range of magnitude estimation ratings for the remaining signals was small (0.5 to 4.0), it was feared that replacement of zero responses would artificially reduce the range of this subject's responses. Therefore, the magnitude estimation data for this subject were removed from the data analysis prior to the transformation, and replaced with those of a new subject of the same sex. The replacement data contained no zero ratings, and reduced the total number of zero ratings to 17.

The transformed data were analyzed in a multivariate analysis of variance (MANOVA) for repeated measures, described in the Data Analysis section below.

Paired Comparison Data Reduction

Each subject provided ratings for 153 pairs of signals. A procedure described in Gescheider (1985) was used in the paired comparison data reduction. The purpose of the data reduction was to obtain a preference score for each stimulus, in z units, which represents a point on a perceptual continuum. The procedure is as follows:

(1) For each pair of signal ratings provided by each subject, assign the signal judged more urgent a value of 1. Assign the signal judged less urgent a value of 0.

(2) For each subject, plot the values obtained in step 1 into a data matrix in which each stimulus is listed in order, once per row and once per column.

The cell (row and column) in which the values are placed are determined by the members of the pair. For the (A,B) pair, the value for signal A is placed in row A, column B. The value for signal B is placed in row B, column A. Because signals are not compared with themselves, the diagonals are left blank.

(3) Sum the values in each cell across subjects. Divide these values by the number of subjects in the experiment. This forms the proportion matrix, which represents the proportion of times one stimulus was perceived as most urgent, in a specific stimulus pair.

(5) Convert the proportions to z scores, using a table of the normal curve.

(6) Add each z score across the row of the data matrix, then divide by the number of stimulus pairs. This is the mean preference score for each stimulus, in z units.

Because the transformation eliminated all subject effects, the design was no longer a repeated measures design, and therefore could not be analyzed in a multivariate analysis with the magnitude estimation data. As described in the Data Analysis section, an ANOVA was used to analyze the paired comparison data.

Detection Time and Workload Data Reduction

Detection time. Each subject provided 36 detection times (two for each signal). The arithmetic mean of the two responses made by each observer to each signal stimulus was obtained. These were the scores utilized in a multivariate analysis of variance for repeated measures, as described in the Analysis of Data section.

Modified Cooper-Harper workload rating. Each subject provided one numerical rating of the mental workload involved in the detection time task (responding to both visual bias conditions and auditory signals). Because CTS data is ordinal, the median response was obtained to give a general indication of the workload imposed by the CTS loading task.

Data Analysis: Rationale and Procedures

Experimental significance level. This experiment utilized inferential statistical analyses from three different sets of data. For purposes of this experiment, effects showing a prob value p greater than 0.05 were not considered statistically significant. Statistically nonsignificant trends of general interest or practical value are also reported where appropriate.

Repeated measures data (magnitude estimation and detection time). Traditionally, repeated measures (within-subjects) designs have been analyzed through univariate analysis of variance. In general, ANOVA applied to repeated measures designs carries a requirement known as sphericity or circularity, which exists if and only if the variance of all pairwise differences between repeated measurements is constant. This is commonly referred to as homogeneity of covariance. The consequence of such a violation is positively biased or liberal tests, resulting in the likelihood of a Type I error exceeding the alpha level set by the user. Because sphericity assumptions are not made in multivariate analysis of variance (MANOVA), the multivariate test is considered to be "exact" for repeated measures designs while the univariate approach can only be considered "approximate" (Vasey and Thayer, 1987) and requires corrections for violations of the assumptions. Vasey and Thayer (1987) noted that the experimenter may adopt a great degree of confidence in the validity of the MANOVA results, since MANOVA makes the least unrealistic assumptions regarding sphericity. In addition, MANOVA, like ANOVA, is quite robust to violations of normality. For these reasons, a MANOVA was used in the analysis of the magnitude estimation and detection time data, which are repeated measures data.

Several steps were involved in the analysis of the repeated measures data. First, a correlational analysis was performed between the transformed magnitude estimation and detection time data to determine the direction and strength of relationship between the two variables. Because the correlation was small in magnitude, these variables were analyzed in separate MANOVAs to avoid potential reduction in statistical power.

Next, repeated-measures MANOVA was performed as an initial overall test of significance for the magnitude estimation and detection time data. The factors in the MANOVA that were shown to be significant were further explored through the use of post-hoc tests. The Newman-Keuls post-hoc procedure was selected because it distributes the alpha error efficiently across the range of all possible comparisons, saving the most power for those comparisons with the smallest differences, while controlling the overall alpha error. A simple-effects F-test was considered as a post-hoc procedure for significant interactions, because it is generally less conservative. The use of this test was rejected because it would contribute no useful information, due to the relatively small number of means involved in the interactions.

To correct for violations of sphericity in the post-hoc tests, the error degrees of freedom for variables with three or more treatment levels were adjusted using a Huynh-Feldt correction (Vasey and Thayer, 1987), in which the error degrees of freedom of all applicable significant effects were multiplied by a constant, known as epsilon (e). The Huynh-Feldt correction was chosen because it is less conservative than the Greenhouse-Geisser correction. Huynh-Feldt e is accurate when nonsphericity is relatively mild and is positively biased when nonsphericity is moderate to severe (Vasey and Thayer, 1987). Use of the correction factor results in little reduction of power if sphericity exists, since e will be near 1 and will protect the tests against bias under nonsphericity. Epsilon corrections were not made for variables with only two treatment levels, because the assumption of sphericity is always fulfilled (Vasey and Thayer, 1987).

Paired comparison data. The paired comparison data were not repeated measures data and thus had no potential for violating sphericity assumptions. Therefore, an ANOVA was utilized as a test of significance on the *z* scores. As discussed previously, the paired comparison data reduction procedure eliminated subject effects, so the denominator for all *F*-tests was the highest-order interaction. Effects which were shown to be significant were further explored through the use of Newman-Keuls post-hoc tests.

Response surface procedures. Response surface procedures (Myers, 1983) were performed where applicable to determine what values of the significant continuous independent variables (pulse level and/or time between pulses) were optimum (greatest) for each dependent variable. Where pulse format was significant, three separate response surfaces were generated, because the response surface would be different for each pulse format. Response surface procedures for each dependent variable involved performing a multiple regression on the continuous variables of interest, examining the response surface through graphical procedures, and then reporting the optimum point on the response surface.

RESULTS

Introduction

All analyses were computed on a Northgate personal computer using Version 4.0 of the Statistical Package for the Social Sciences, SPSS/PC+ (SPSS Inc., 1990). ANOVA and post-hoc tests were checked through recomputation using Release 7 of MINITAB (MINITAB, INC., 1989) to ensure accuracy.

Correlational Analyses

A correlational analysis was performed between the magnitude estimation and detection time data to determine the direction and strength of relationship between the two variables. The correlation is small in magnitude (r = -0.164), but is statistically significant (t = -4.23, p < 0.01). The large sample size (N = 648), which gave high statistical power to the t test, contributed to the significance obtained. Despite the statistical significance, because the magnitude of the correlation is very small, these variables were analyzed in separate MANOVAs to avoid potential reduction in statistical power.

A correlational analysis between paired comparison and magnitude estimation data was performed, after the magnitude estimation data for each signal were collapsed across subjects and then averaged. The correlation is large in magnitude ($\underline{r}$ = 0.95), and is statistically significant ($\underline{t}$ = 11.60, $\underline{p}$ < .01).

Magnitude Estimation

Table 8 contains the MANOVA results on the magnitude estimation data, which used Wilk's criterion $\underline{U}$ as the test statistic. The significant interactions are pulse format and time between pulses ($\underline{U}$ = 0.749, $\underline{p}$ = 0.050), pulse format and pulse level ($\underline{U}$ = 0.479, $\underline{p}$ < 0.001), and time between pulses and pulse level ($\underline{U}$ = 0.686, $\underline{p}$ = 0.002). The significant main effects are pulse format ($\underline{U}$ = 0.591, $\underline{p}$ < 0.001), time between pulses ($\underline{U}$ = 0.399, $\underline{p}$ < 0.001), and pulse level ($\underline{F}$ = 114.890, $\underline{p}$ < 0.001).

Magnitude estimation ANOVA data are presented in Table 1 in Appendix E, only to supplement information provided in the MANOVA. In the presence of Huynh-Feldt (H-F) e corrections, in which actual decimal values were used, almost all sources of variance which are significant in the MANOVA are also significant in the ANOVA. The exception is the format x time interaction, which is significant in the

TABLE 8

Magnitude Estimation Data MANOVA Summary Table*

Source of Approx. Variance	dv	dfH	dfE	U	F	p
Between-Subjects						
Subjects (S)	1	35				
Within-Subjects						
Pulse Format (F)	1	2	34	0.591	11.768	<0.001
F X S						
Time Between Pulses (T)	1	2	34	0.399	25.638	<0.001
T X S						
Pulse Level (L)	1	1	35	**	114.890	<0.001
L X S						
F X T	1	4	32	0.749	2.676	0.050
F X T X S						
F X L	1	2	34	0.479	18.480	<0.001
F X L X S						
T X L	1	2	34	0.686	7.788	0.002
T X L X S						
F X T X L	1	4	32	0.949	0.430	0.786
F X T X L X S						

where:
dv = number of dependent measures
dfH = degrees of freedom for treatment effect
dfE = degrees of freedom for error effect
U = Wilk's likelihood ratio statistic
p = significance of approx. F

* Denominators used for each source of variance in the U tests appear as the second term in each grouping in the table.

** When dfH = 1, the MANOVA statistic yields exact univariate (F) test results only.

MANOVA ($\underline{U}$ = 0.749, $\underline{p}$ = 0.050), but not in the ANOVA ($\underline{F}$ = 2.39, $\underline{p}$ = 0.055), because the ANOVA significance exceeds the predetermined experimental significance level by $\underline{p}$ = 0.005. Because the difference in significance is so small, it should not be of concern. Overall, where the significance achieved in the ANOVA differs from that of the MANOVA, it is due to the more "approximate" nature of the univariate approach in the presence of nonsphericity.

Table 9 contains the post hoc-test results for the interaction of pulse format and time between pulses on magnitude estimation. The means are plotted in Figure 4. The perceived urgency of the pulse formats decreases differentially as inter-pulse interval increases. At inter-pulse intervals of 0 ms (no time between pulses) and 150 ms, the perceived urgency of the frequency-modulated pulse is greater than that of the simultaneous pulse (at 0 ms, Δ = 0.731; at 150 ms, Δ = 0.191). This difference is significant at 0 ms, but is not significant at 150 ms. At 300 ms, the perceived urgency of the simultaneous pulse is greater than that of the frequency-modulated pulse, (Δ = 0.233) but this difference is not significant. The perceived urgency of the sequential pulse is significantly less than that of the frequency-modulated and simultaneous pulses, at all inter-pulse intervals (Δ = 2.320, 2.063, 1.876 between frequency-modulated and sequential pulses at 0

TABLE 9

Newman-Keuls Test: Magnitude Estimation Data, Pulse Format x Time Between Pulses

Pulse Format	Time Between Pulses (ms)	Mean Magnitude Estimate	S.D.	Significance*
Sequential	0	5.983	3.006	A
Sequential	150	5.037	2.525	B
Sequential	300	4.516	2.419	C
Simultaneous	0	7.572	4.054	D
Simultaneous	150	6.909	4.101	E F
Simultaneous	300	6.625	3.752	E F
Sawtooth Freq. Mod.	0	8.303	4.377	G
Sawtooth Freq. Mod.	150	7.100	3.869	F
Sawtooth Freq. Mod.	300	6.392	3.533	A E

*Levels of mean magnitude estimate with the same letters are not significantly different from each other at the 0.05 level. An H-F epsilon correction of 0.834 was used to compute the error degrees of freedom used in this test.

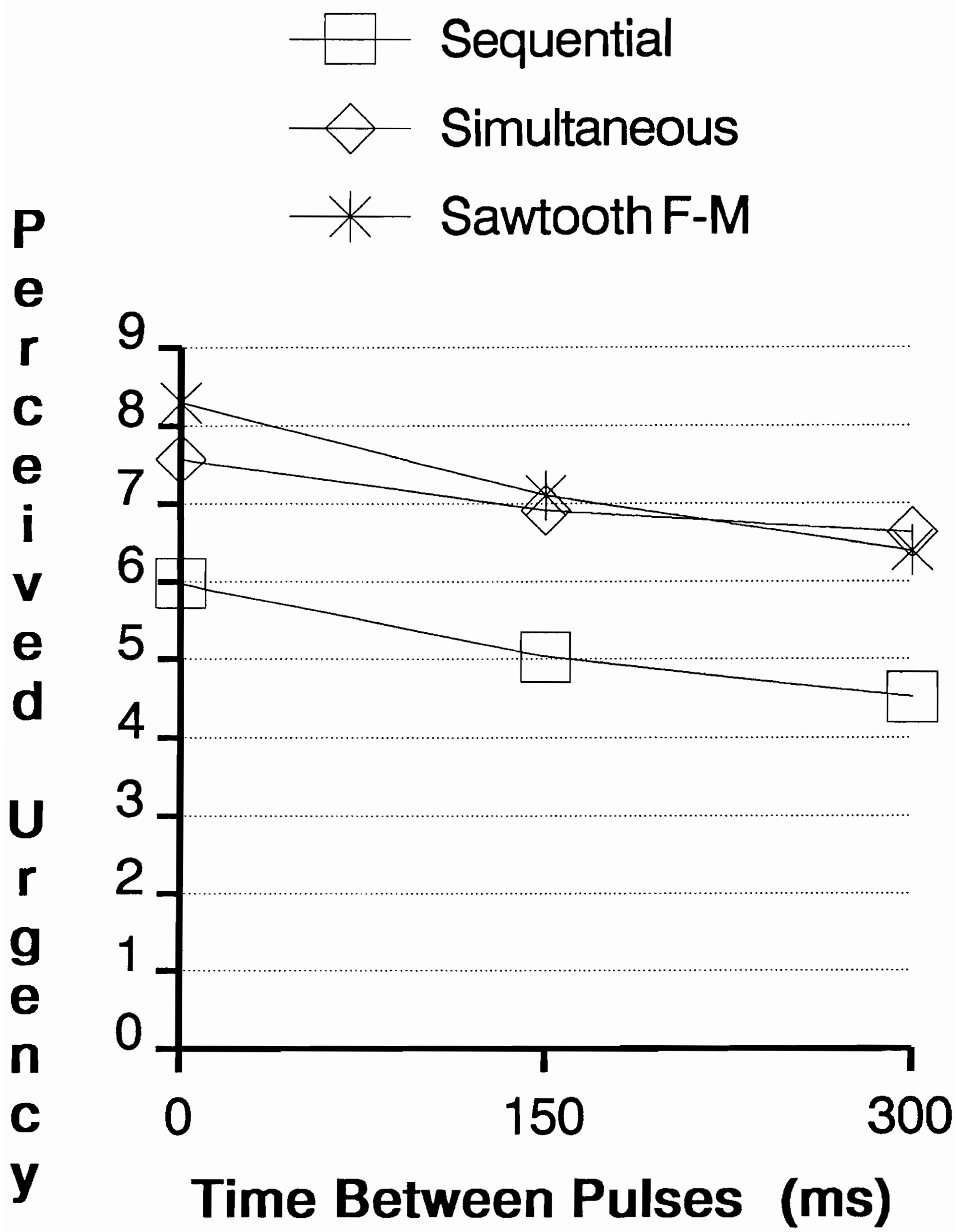


Figure 4. Mean perceived urgency for the pulse format x time between pulses interaction.

ms, 150 ms, and 300 ms, respectively; and $\triangle$= 1.589, 1.872, 2.109 between simultaneous and sequential pulses at 0 ms, 150 ms, and 300 ms, respectively).

The interaction between pulse format and time between pulses was analyzed to determine if it would preclude the interpretation of the pulse format main effect. As seen in Figure 4, only the sequential pulse format is ordinal (relative positioning did not change) with respect to the other pulse formats. No other ordinal relationships exist for this interaction. The main effect of pulse format can be unambiguously interpreted between categories in which there is an ordinal relationship. Careful interpretation should be used where disordinality exists. Therefore, this interaction allows the unambiguous interpretation of the signal type main effect for any pulse format, and caution is used in interpreting the main effects for simultaneous and frequency-modulated pulses.

This interaction was also analyzed to determine if it would preclude the interpretation of the main effect of time between pulses. As can be seen in Figure 5, all inter-pulse intervals are ordinal with respect to all others. This particular interaction allows the unambiguous interpretation of the main effect of time between pulses.

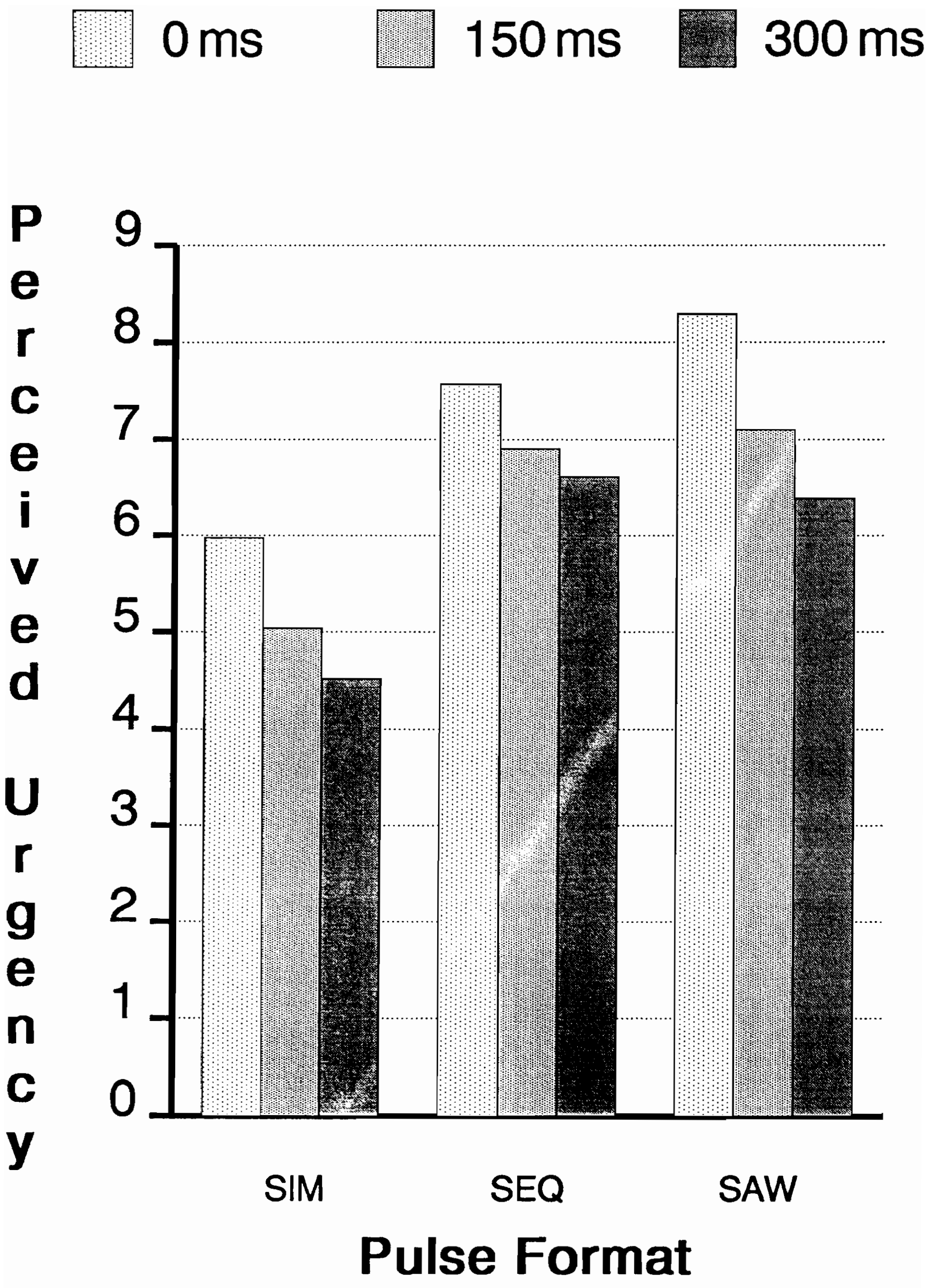


Figure 5. Graph of the pulse format x time between pulses interaction.

Table 10 contains the post-hoc test results for the interaction of pulse format and pulse level on magnitude estimation. The means are plotted in Figure 6. The perceived urgency of all pulse formats increases differentially as pulse level increases. At 65 dBC, the perceived urgency of frequency-modulated pulses is significantly greater than that of sequential pulses ($\triangle$ = 0.577) and simultaneous pulses ($\triangle$ = 1.271). At 79 dBC, frequency-modulated pulses are perceived as more urgent than simultaneous pulses ($\triangle$ = 0.207), but this difference is not significant. The perceived urgency of the sequential pulse is significantly less than that of the frequency-modulated and simultaneous pulses at all pulse levels ($\triangle$ = 1.271, 2.901 between frequency-modulated and sequential pulses at 65 dBC and 79 dBC, respectively; and $\triangle$ = 0.694, 3.108 between simultaneous and sequential pulses at 65 dBC and 79 dBC, respectively).

The interaction between pulse format and pulse level was analyzed to determine if it would allow the interpretation of the pulse format main effect. As seen in Figure 6, only the sequential pulse format is ordinal with respect to the other pulse formats. No other ordinal relationships exist for this interaction. Therefore, this interaction does not preclude the unambiguous interpretation of the signal type main effect for any pulse format,

TABLE 10

Newman-Keuls Test: Magnitude Estimation Data, Pulse Format x Pulse Level

Pulse Format	Pulse Level (dBC)	Mean Magnitude Estimate	S.D.	Significance*
Sequential	65	3.479	1.418	A
Sequential	79	6.879	2.648	B
Simultaneous	65	4.173	1.899	C
Simultaneous	79	9.897	3.403	D
Sawtooth Freq. Mod.	65	4.750	2.877	E
Sawtooth Freq. Mod.	79	9.780	3.337	D

*Levels of mean magnitude estimate with the same letters are not significantly different from each other at the 0.05 level. An H-F epsilon correction of 0.945 was used to compute the error degrees of freedom used in this test.

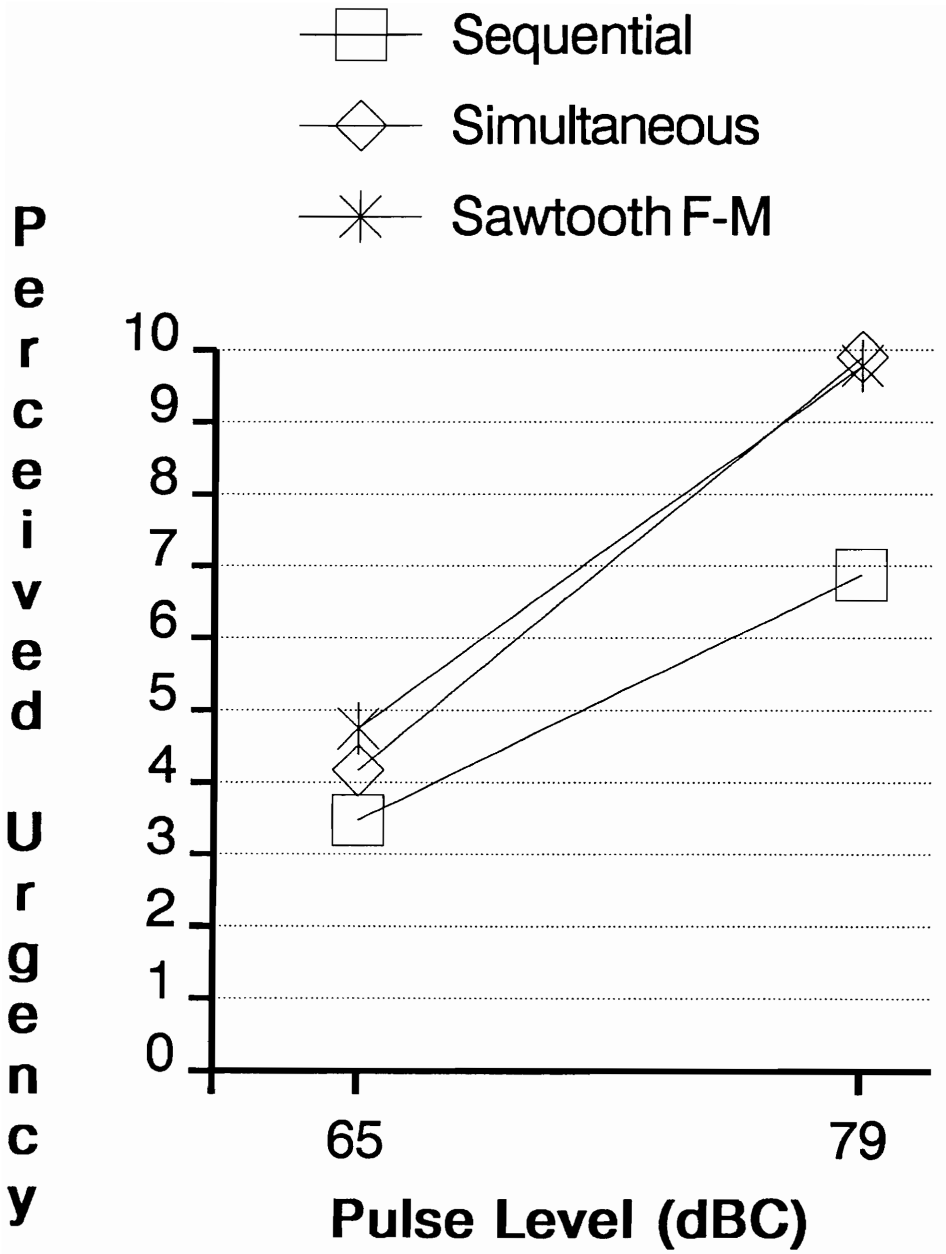


Figure 6. Mean perceived urgency for the pulse format x pulse level interaction.

although caution is used in interpreting the main effects for simultaneous and frequency-modulated pulses.

This interaction was also analyzed to determine if it precludes the interpretation of the main effect of pulse level. As can be seen in Figure 7, all pulse levels are ordinal with respect to all others. The interaction between pulse format and pulse level allows the unambiguous interpretation of the main effect of time between pulses.

Table 11 contains the post-hoc test results for the interaction of time between pulses and pulse level on magnitude estimation. The means are plotted in Figure 8. The perceived urgency of all inter-pulse intervals increases differentially with a decrease in pulse level. At 65 dBC, the perceived urgency of the 0 ms inter-pulse interval is significantly greater than that of the 150 ms interval ($\triangle = 0.550$) and the 300 ms interval ($\triangle = 0.340$). The perceived urgency of the 150 ms interval is significantly greater than that of the 300 ms interval ($\triangle = 0.384$). At 79 dBC, the perceived urgency of the 0 ms interval is significantly greater than the 150 ms interval ($\triangle = 1.324$) and the 300 ms interval ($\triangle = 1.949$). The perceived urgency of the 150 ms interval is significantly greater than that of the 300 ms interval ($\triangle = 0.625$). As pulse level increases, the 0 ms inter-pulse interval shows the greatest increase in

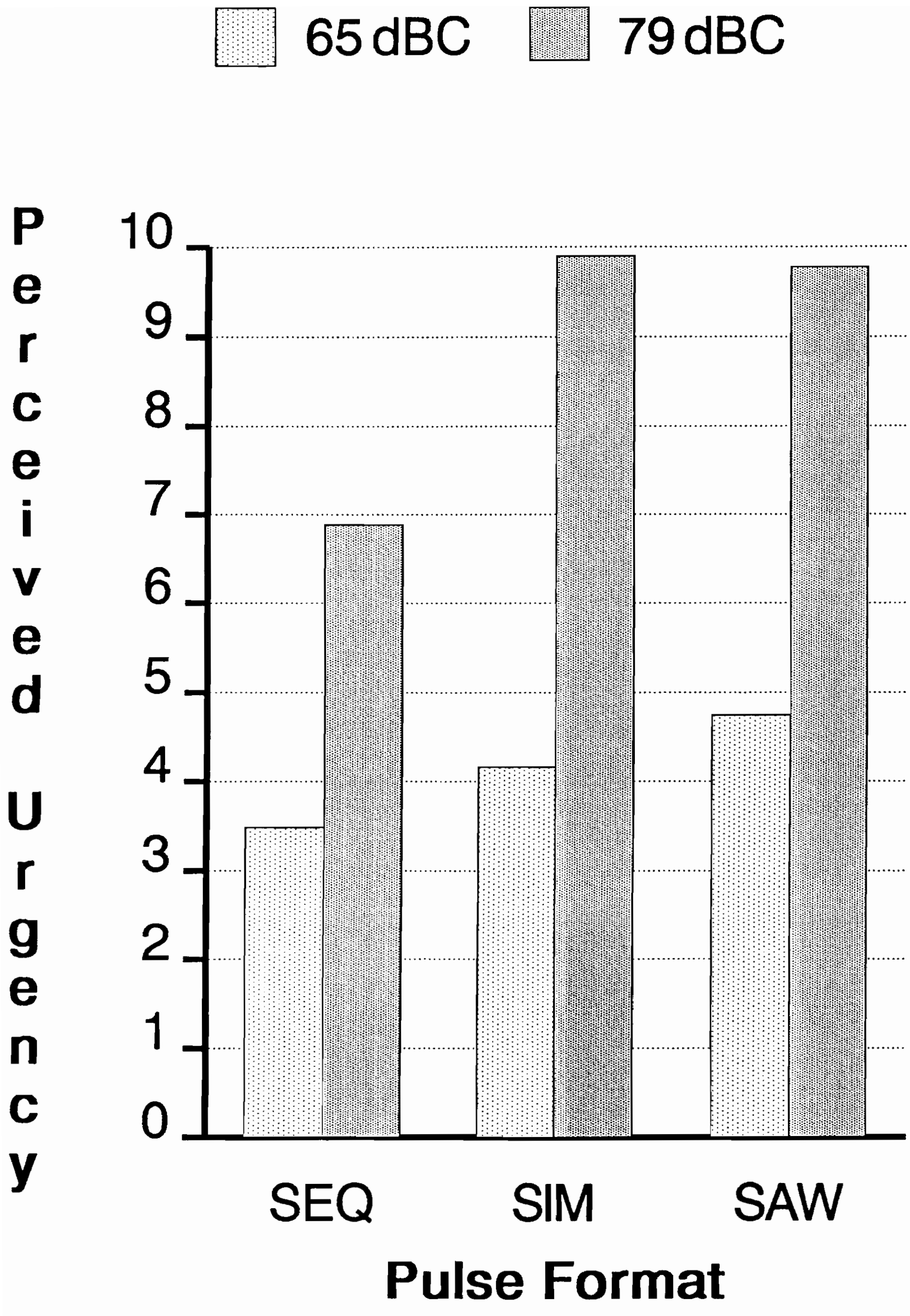


Figure 7. Graph of the pulse format x pulse level interaction.

TABLE 11
Newman-Keuls Test: Magnitude Estimation Data, Time Between Pulses x Pulse Level

Time Between Pulses (ms)	Pulse Level (dBC)	Mean Magnitude Estimate	S.D.	Significance*
0	65	4.629	2.243	A
0	79	9.943	3.497	B
150	65	4.079	2.241	C
159	79	8.619	3.413	D
300	65	3.695	2.055	E
300	79	7.994	3.123	F

*Levels of mean magnitude estimate with the same letters are not significantly different from each other at the 0.05 level. An H-F epsilon correction of 0.906 was used to compute the error degrees of freedom used in this test.

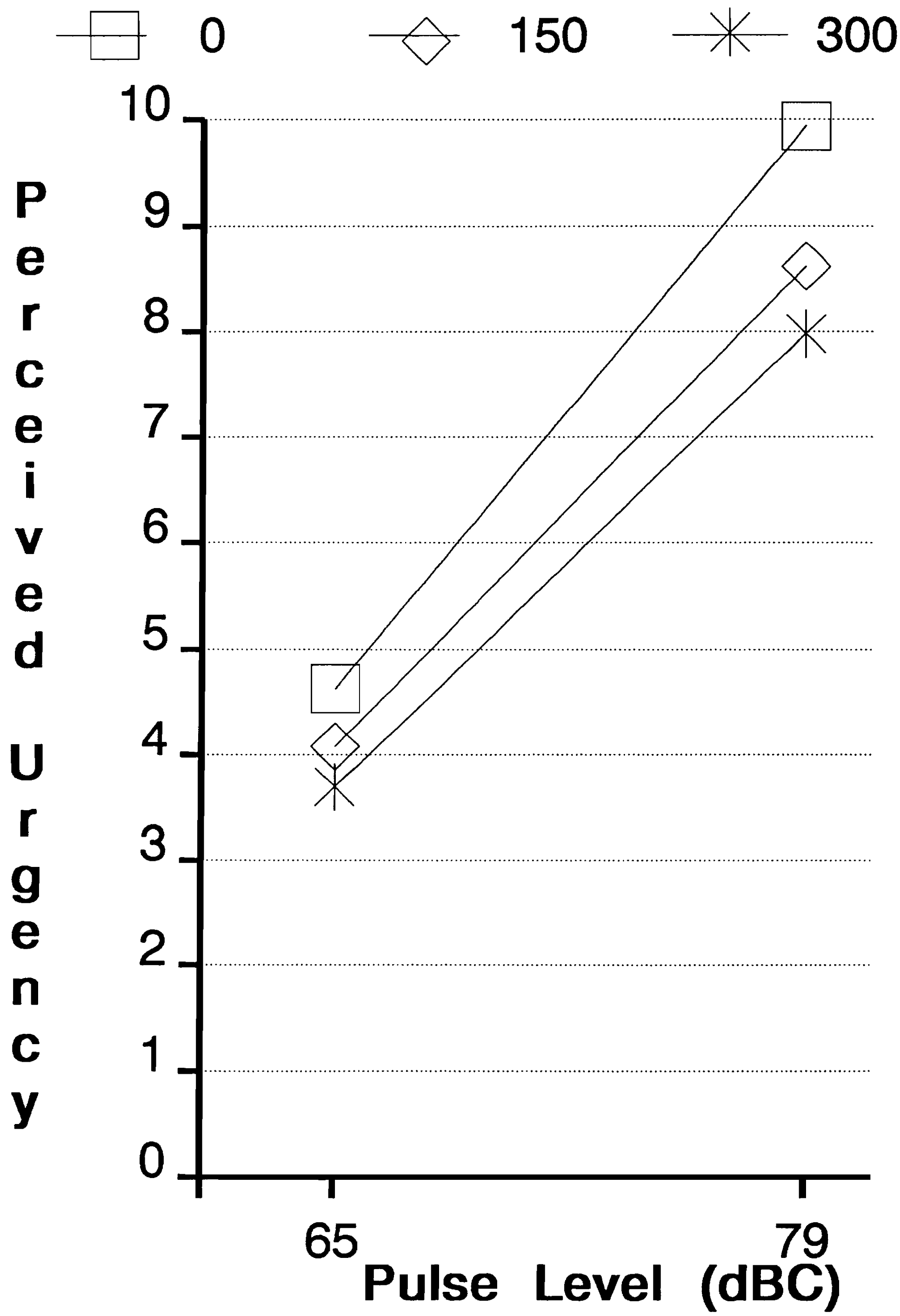


Figure 8. Mean perceived urgency for the time between pulses x pulse level interaction.

perceived urgency.

The interaction between time between pulses and pulse level was analyzed to determine if it would allow the interpretation of the time between pulses and the pulse level main effects. The nature of the interaction is relatively weak; every level of each variable is ordinal. As seen in Figure 8, all inter-pulse intervals are ordinal with respect to the others. All pulse levels are ordinal with respect to the others. Therefore, this interaction allows the unambiguous interpretation of the time between pulses and the pulse level main effects.

The Newman-Keuls post-hoc test results for the magnitude estimation of the pulse format main effect are contained in Table 12, with means and confidence intervals plotted in Figure 9. These results should not be interpreted independently of the pulse format x time and the format x level interactions, which indicate that the perceived urgency of pulse format differs depending upon time between pulses and pulse level. Only sequential pulses are completely independent of these factors. The interaction data permit the conclusion that subjects rated signals with sequential pure tones as sounding significantly less urgent than those containing simultaneous or frequency-modulated tones. Under most conditions, there

TABLE 12

Newman-Keuls Test: Magnitude Estimation Data, Pulse Format

Pulse Format	Mean Magnitude Estimate	S.D.	Significance*
Sequential	5.179	2.719	A
Simultaneous	7.035	3.973	B
Sawtooth Freq. Mod.	7.265	4.002	B

*Levels of mean magnitude estimate with the same letters are not significantly different from each other at the 0.05 level. An H-F epsilon correction of 0.968 was used to compute the error degrees of freedom used in this test.

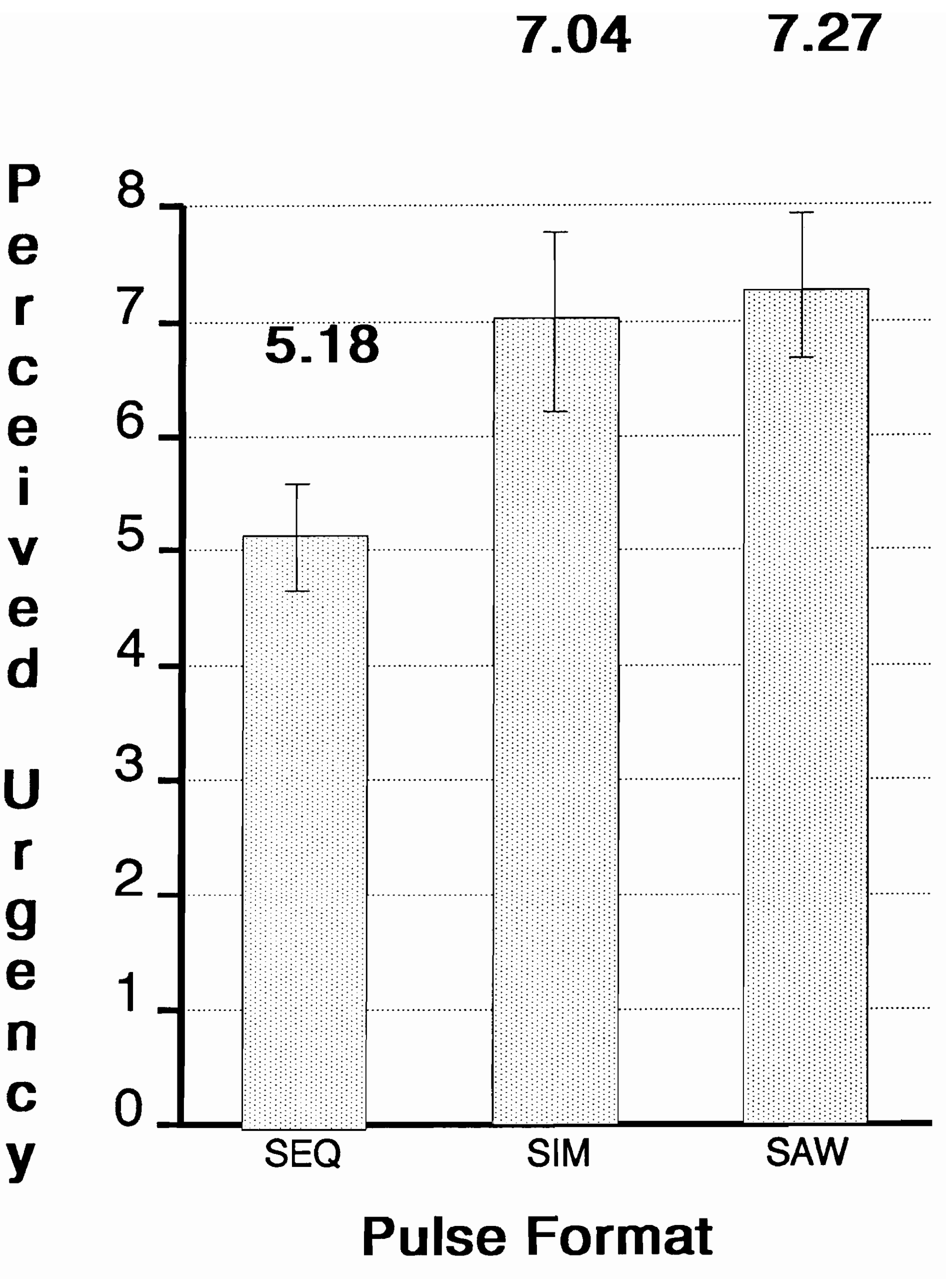


Figure 9. Mean perceived urgency and 95% confidence intervals for pulse format.

is no significant difference in perceived urgency between simultaneous and sequential signals. Subjects rated the frequency-modulated signals as significantly more urgent than the simultaneous signal at the 0 ms inter-pulse interval and at the 65 dBC pulse level.

The post-hoc test results for the time between pulses main effect using the magnitude estimation data are contained in Table 13. The means and confidence intervals are plotted in Figure 10. Subjects rated signals with an inter-pulse interval of 0 ms as sounding significantly more urgent than those containing intervals of 150 ms and 300 ms ($p \leq 0.05$). There is no significant difference in the perceived urgency of signals with inter-pulse intervals of 150 ms and 300 ms. These results can be interpreted independently of the pulse format x time between pulses and the time between pulses x pulse level interactions.

The means and standard deviations for the magnitude estimation of the pulse level main effect indicate that subjects rated signals with greater pulse levels as sounding significantly more urgent. The 79 dBC pulse level has a mean magnitude estimation of 8.852, while the 65 dBC pulse level has a mean of 4.134. The standard deviations of the 65 dBC and 79 dBC pulses are 3.436 and 2.208, respectively. These results can be interpreted independently of the pulse

TABLE 13

Newman-Keuls Test: Magnitude Estimation Data, Time Between Pulses

Time Between Pulses, ms	Mean Magnitude Estimate	S.D.	Significance*
0.0	7.286	3.960	A
150.0	6.349	3.670	B
300.0	5.844	3.406	C

*Levels of mean magnitude estimate with the same letters are not significantly different from each other at the 0.05 level. An H-F epsilon correction of 0.694 was used to compute the error degrees of freedom used in this test.

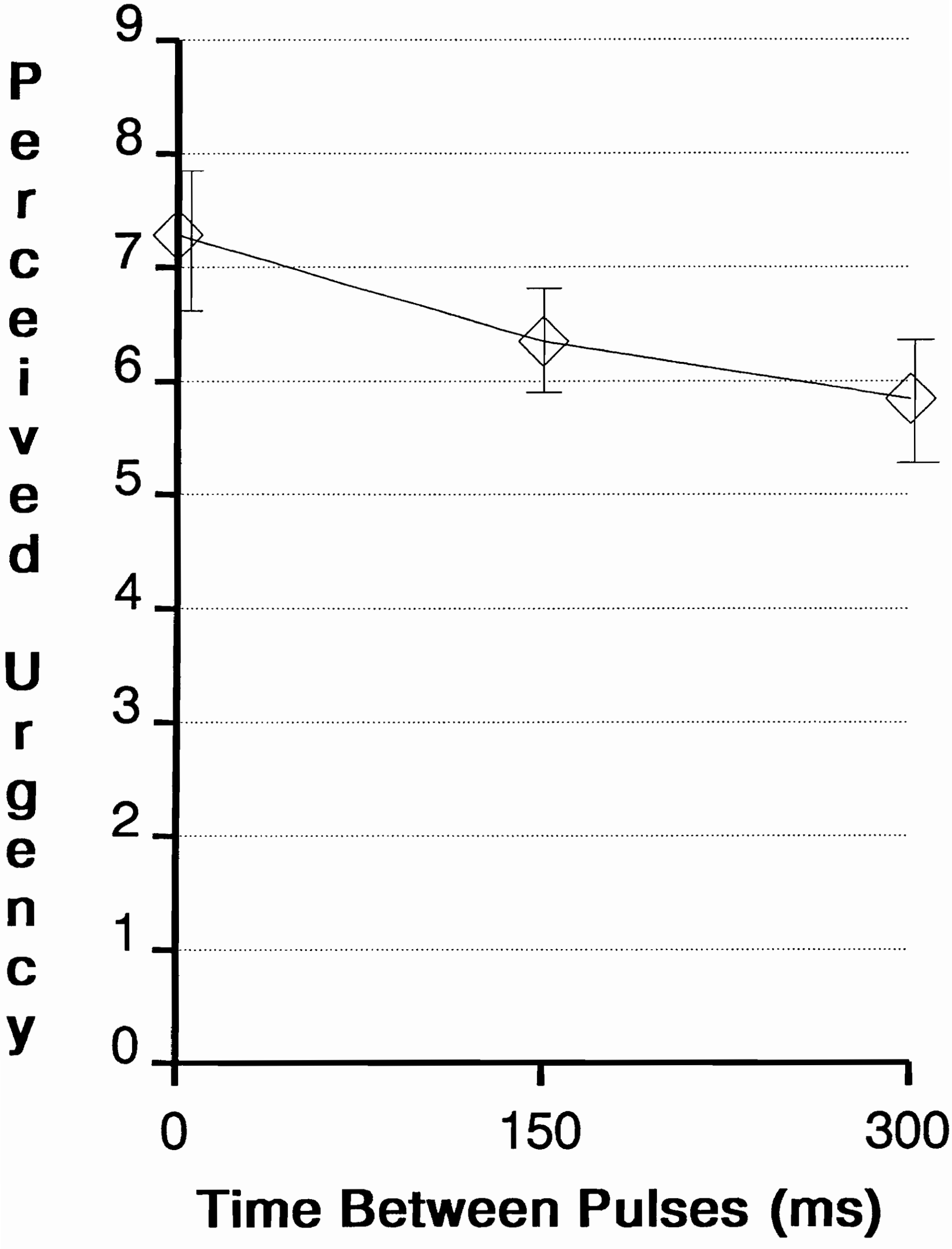


Figure 10. Mean perceived urgency and 95% confidence intervals for time between pulses.

format x pulse level and the time between pulses x pulse level interactions.

Response surface procedures (Myers, 1983) were performed to determine what levels of the significant continuous independent variables provided the greatest (most urgent) magnitude estimation values. Both pulse level, time between pulses, and the pulse level x time between pulses interaction were included, because they are significant in the magnitude estimation MANOVA. The significant MANOVA for pulse format indicated that the response surface is different for each pulse type. Therefore, a separate response surface was generated for each pulse format.

The multiple regression ANOVA for the sequential pulse data is contained in Table 14, and the multiple regression variable summary in Table 15. The regression is significant ($F = 57.698$, $p < 0.001$). The beta values for pulse level ($t = 9.148$, $p < 0.001$) and the constant (intercept) ($t = -6.488$, $p < 0.001$) are significant. The beta values for time between pulses ($t = 1.342$, $p = 0.181$), as well as for the time x level interaction ($t = -1.770$, $p = 0.0782$) are not significant. The nonsignificant variables were included in the regression equation because they are continuous variables of interest, and are significant in the MANOVA.

TABLE 14

Sequential Pulse Magnitude Estimation Data, Multiple Regression ANOVA Summary Table

Source of Variance	df	MS	*F*	*p*
Regression	3	238.161	57.698	<0.001
Residual	212	4.128		

TABLE 15

Sequential Pulse Magnitude Estimation Data, Multiple Regression Variable Summary Table

Variable	Beta	S.E. Beta	*t*	*p*
Pulse Level (L)	0.286	0.031	9.148	<0.001
Time Between Pulses (T)	0.016	0.012	1.342	0.181
L X T	0.000	0.000	-1.770	0.078
Constant (Intercept)	-14.654	2.259	-6.488	<0.001

The regression function in Table 15 is plotted in Figure 11. The function is linear. An increase in pulse level produces the largest increase in perceived urgency. Perceived urgency is greatest when pulse level is highest (79 dBC) and when time between pulses is smallest (0 ms between pulses.

The multiple regression ANOVA for the simultaneous pulse data is contained in Table 16 and the multiple regression variables in Table 17. The regression is significant ($F = 80.265$, $p < 0.001$). The beta values for pulse level ($t = 10.253$, $p < 0.001$) and the constant (intercept) ($t = -7.738$, $p < 0.001$) are significant. The beta values for time between pulses ($t = 0.489$, $p = 0.626$), as well as for the time x level interaction ($t = -0.693$, $p = 0.489$) are not significant. The nonsignificant variables were included in the regression equation because they are continuous variables of interest and are significant in the MANOVA.

The regression function in Table 17 is plotted in Figure 12. The function is linear. As with the sequential pulse format, an increase in pulse level produces the largest increase in perceived urgency. A comparison of Figures 11 and 12 indicates that the level of perceived urgency for the simultaneous signals is greater than that

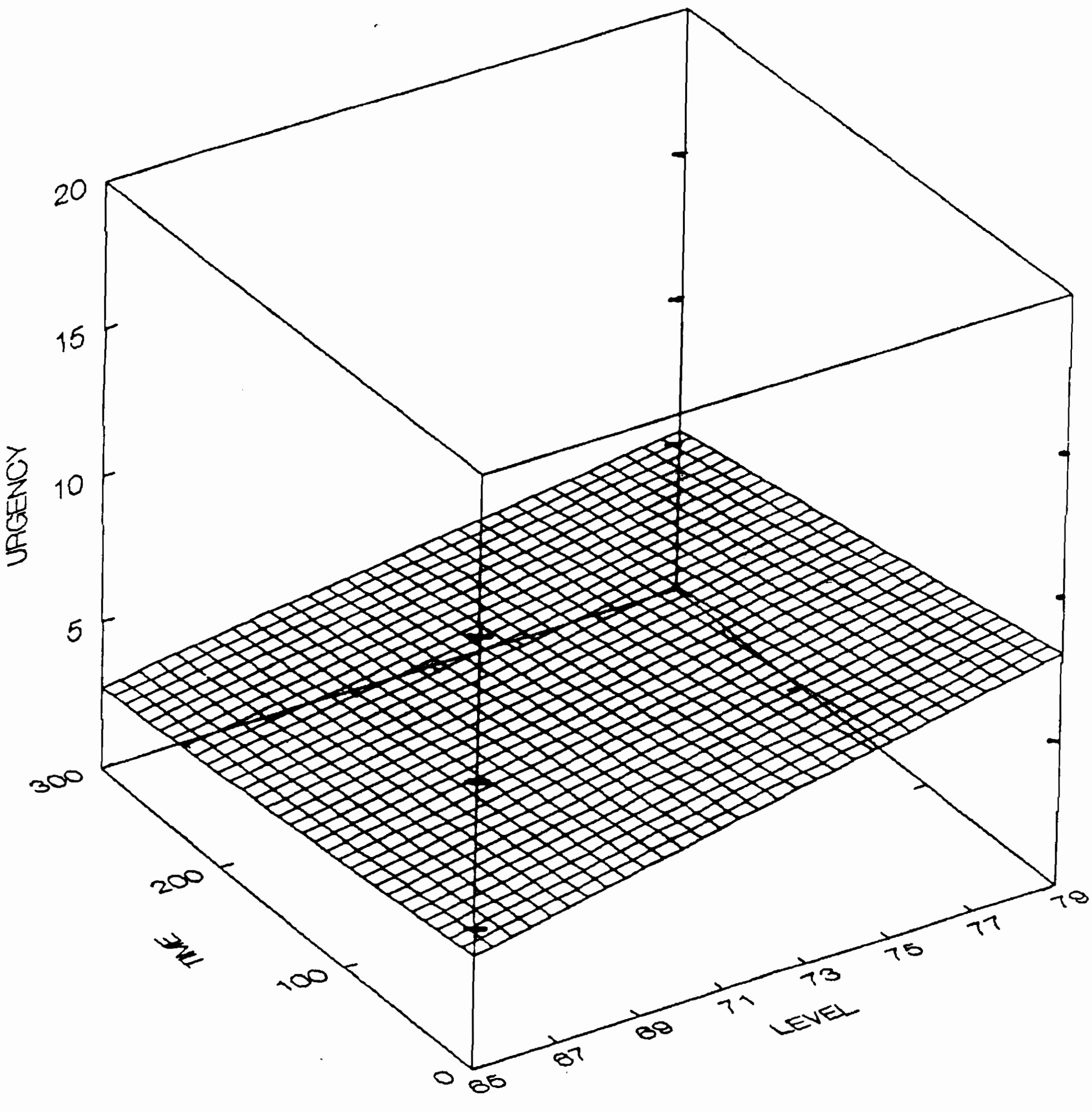


Figure 11. Perceived urgency response surface for sequential pulse format.

TABLE 16

Simultaneous Pulse Magnitude Estimation Data, Multiple Regression ANOVA Summary Table

Source of Variance	df	MS	*F*	*p*
Regression	3	601.695	80.265	<0.001
Residual	212	7.496		

TABLE 17

Simultaneous Pulse Magnitude Estimation Data, Multiple Regression Variable Summary Table

Variable	Beta	S.E. Beta	*t*	*p*
Pulse Level (L)	0.431	0.042	10.253	<0.001
Time Between Pulses (T)	0.008	0.016	0.489	0.626
L X T	0.000	0.000	-0.693	0.489
Constant (Intercept)	-23.554	3.044	-7.738	<0.001

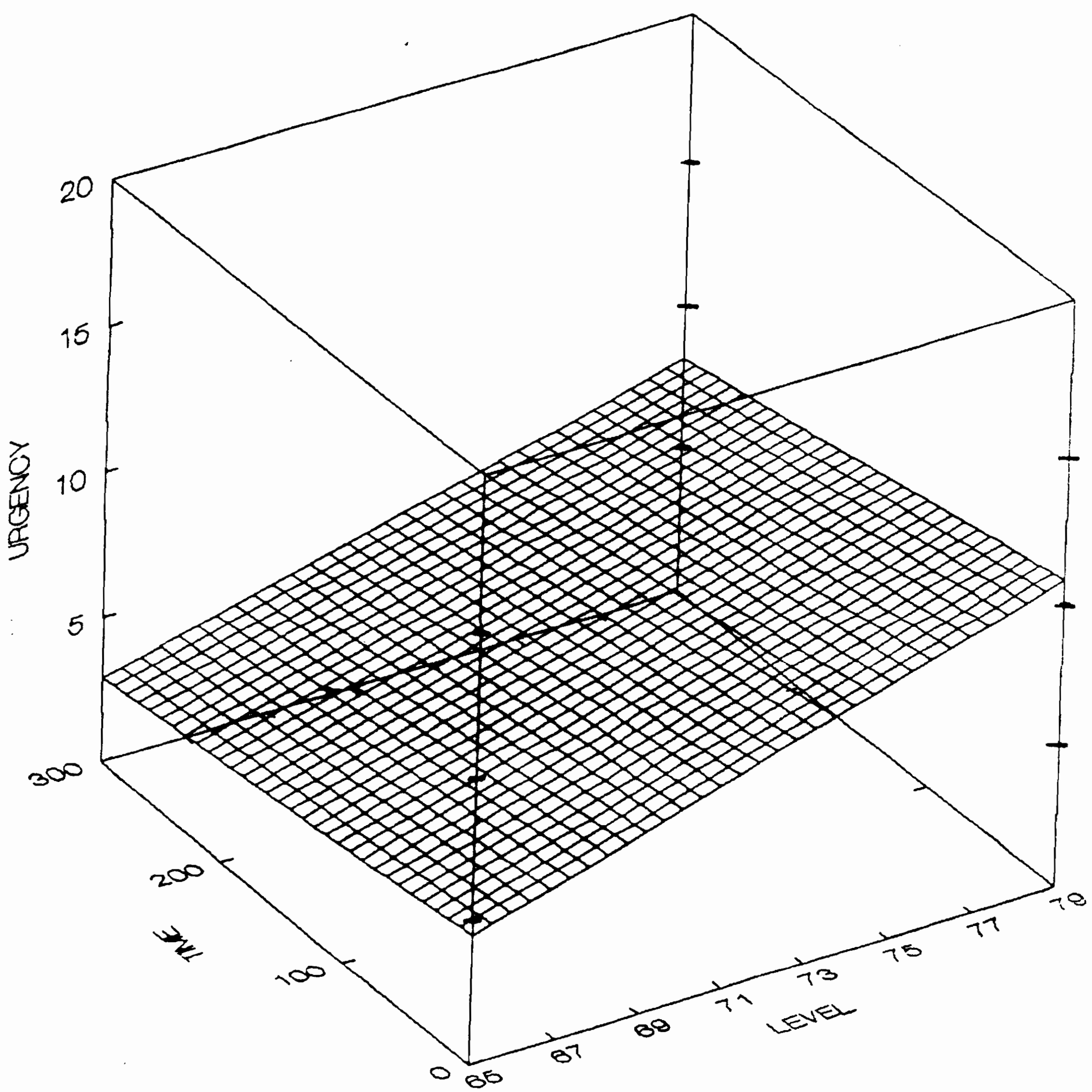


Figure 12. Perceived urgency response surface for simultaneous pulse format.

for the sequential signals at all levels of all continuous variables. As with sequential pulses, the perceived urgency of simultaneous pulses is greatest when pulse level is highest (79 dBC) and when time between pulses is smallest (0 ms between pulses).

The multiple regression ANOVA for the sawtooth frequency-modulated pulse data appears in Table 18 and the regression variables in Table 19. The regression is significant (F = 55.233, $p < 0.001$). The beta values for pulse level (t = 8.674, $p < 0.001$) and the constant (intercept) (t = - 6.185, $p < 0.001$) are significant. The beta values for time between pulses (t = 0.831, p = 0.407), as well as for the time between pulses x pulse level interaction (t = - 1.204, p = 0.230) are not significant. The nonsignificant variables were included in the regression equation because they are continuous variables of interest, and are significant in the MANOVA.

The regression function in Table 19 is plotted in Figure 13. The function is linear. As with the simultaneous and sequential pulse formats, an increase in pulse level produces an increase in perceived urgency. Perceived urgency is greatest when pulse level is highest (79 dBC) and when time between pulses is smallest (0 ms between pulses).

TABLE 18

Frequency-modulated Pulse Magnitude Estimation Data,
Multiple Regression ANOVA Summary Table

Source of Variance	df	MS	*F*	*p*
Regression	3	503.579	55.233	<0.001
Residual	212	9.117		

TABLE 19

Frequency-modulated Pulse Magnitude Estimation Data, Multiple Regression Variable Summary Table

Variable	Beta	S.E. Beta	*t*	*p*
Pulse Level (L)	0.403	0.046	8.674	<0.001
Time Between Pulses (T)	0.014	0.017	0.831	0.407
L X T	0.000	0.000	-1.204	0.230
Constant (Intercept)	-20.763	3.357	-6.185	<0.001

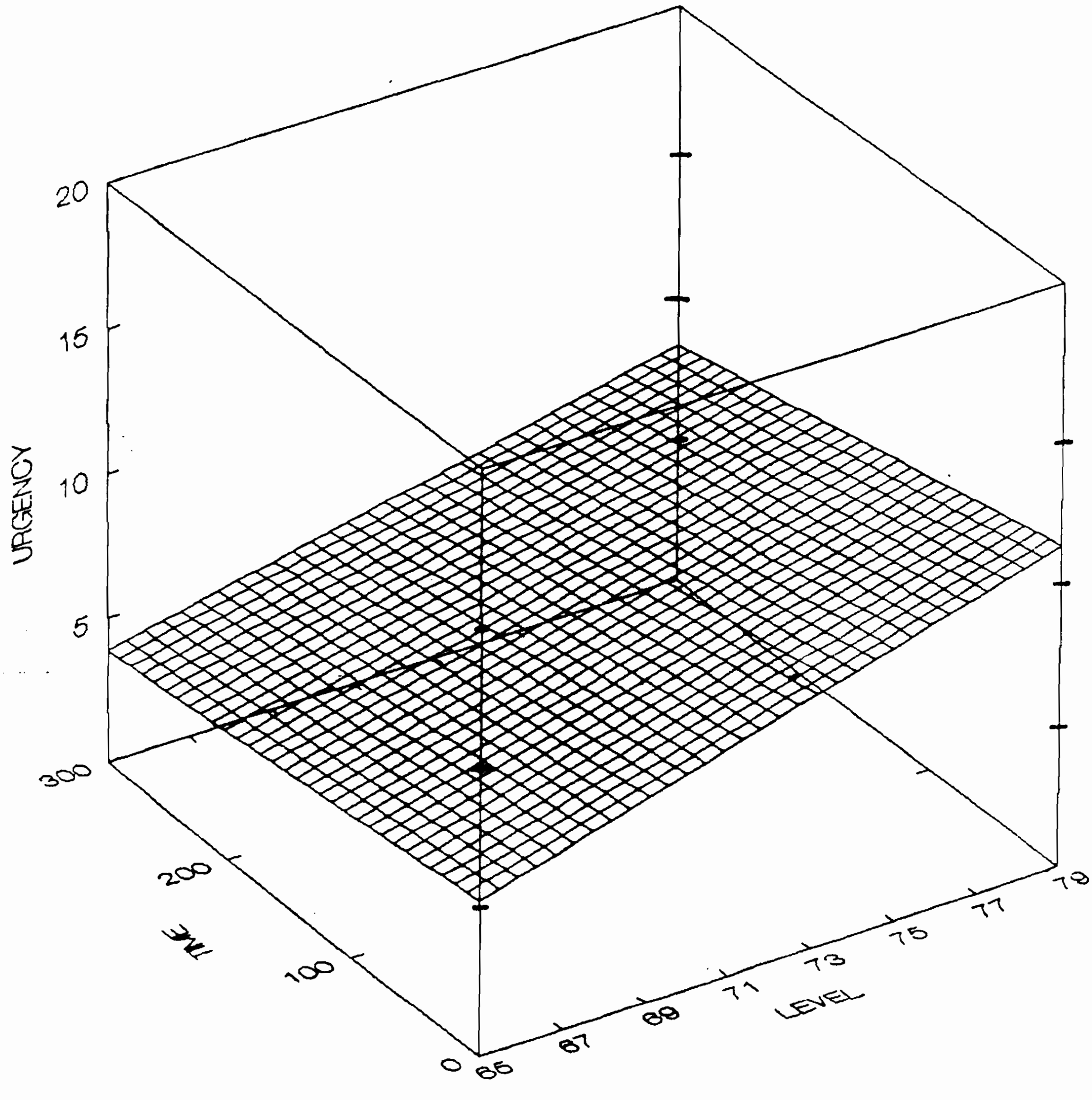


Figure 13. Perceived urgency response surface for frequency modulated pulse format.

A comparison of all response surfaces (Figures 11, 12, and 13) indicates that the sequential signals are perceived to be less urgent than the simultaneous and frequency-modulated pulses at all levels of all continuous variables. This accurately reflects the findings presented thus far. For all variables, perceived urgency is greatest when pulse level is highest (79 dBC) and when time between pulses is smallest (0 ms between pulses). The perceived urgency of the frequency-modulated pulses appears to be greater than that of the other pulse types at all levels of all continuous variables. This seemingly contradicts the results presented in Table 9 and Figure 4, which indicates that mean perceived urgency of the simultaneous pulses is greater than that of the frequency-modulated pulses at longest interval between pulses (300 ms), although this difference is not significant. The difference between the post-hoc and the multiple regression data may have occurred because the regression line suffers from some small degree of lack of fit. The lack of fit could be due to the approximate nature of the prediction of the least-square approach utilized in the multiple regression.

Although the data for the three-way interaction is not significant in the MANOVA, the slopes for the three response surfaces may appear to be different, but not significantly so. The apparent difference may have occurred because each multiple regression (each response surface) used only a subset of pulse format data. As a result, the response surface model does not distribute the alpha error across all levels of the three-way interaction, and the three-way interaction is not represented in its entirety.

Paired Comparisons

The paired comparison data resulted in the ANOVA of Table 20. The interaction between pulse format and pulse level (F = 7.711, p = 0.042) is significant. Significant main effects are pulse format (F = 210.367, p < 0.001), time between pulses (F = 141.583, p < 0.001), and pulse level (F = 940.916, p < 0.001).

An index of association, omega squared (Keppel, 1973), was performed on the paired comparison ANOVA data to obtain the proportion of the total variability accounted for by each of the experimental treatments. These data indicate that main effects alone account for approximately 98 percent of the variability. Pulse level accounts for 56 percent, pulse format explains 25 percent, and time between pulses explains 17 percent of the total variability. Interactions of the main effects account for the remaining 2 percent.

TABLE 20

Paired Comparison Data ANOVA Summary Table

Source of Variance	df	MS	*F**	*p*
Within-Subjects				
Pulse format (F)	2	1.584	210.367	<0.001
F X T X L	4	0.008		
Time Between Pulses (T)	2	1.066	141.583	<0.001
F X T X L	4	0.008		
Pulse Level (L)	1	7.085	940.916	<0.001
F X T X L	4	0.008		
F X T	4	0.012	1.612	0.327
F X T X L	4	0.008		
F X L	2	0.058	7.711	0.042
F X T X L	4	0.008		
T X L	2	0.015	1.938	0.258
F X T X L	4	0.008		
Total	17			

* Denominators used for each source of variance in the *F* tests appear as the second term in each grouping in the table.

Table 21 contains the post-hoc test results for the interaction between pulse format and pulse level. The means are plotted in Figure 14. The perceived urgency of all pulse formats increases differentially as pulse level increases. The nature of the interaction is relatively weak; every level of each variable is ordinal. At 65 dBC, the mean preference score for frequency-modulated pulses is significantly greater than that of the sequential pulses ($\triangle$ = 0.759 _z_ units) and simultaneous pulse formats ($\triangle$= 0.083 _z_ units). At 79 dBC, frequency-modulated pulses are perceived as more urgent than simultaneous pulses ($\triangle$ = 0.065 _z_ units), but this difference is not significant. The perceived urgency of the sequential pulse is significantly less than that of the frequency-modulated and simultaneous pulses at all pulse levels ($\triangle$ = 1.424, 1.359 _z_ units between frequency-modulated and sequential pulses at 65 dBC and 79 dBC, respectively; and $\triangle$= 0.694, 3.108 _z_ units between simultaneous and sequential pulses at 69 dBC and 79 dBC, respectively).

The interaction between pulse format and pulse level was analyzed to determine if it would preclude the interpretation of the pulse format and pulse level main effects. As seen in Figure 14, each pulse format is ordinal with respect to the other pulse formats. Each pulse level is ordinal to the other pulse levels. This interaction

TABLE 21

Newman-Keuls Test: Paired Comparison Data, Pulse Format x Pulse Level

Pulse Format	Pulse Level (dBC)	Mean Preference Score (z units)	S.D. (z units)	Significance**
Sequential	65	0.397	0.452	A
Sequential	79	1.092	0.299	B
Simultaneous	65	1.073	0.438	B
Simultaneous	79	2.450	0.278	C
Sawtooth Freq. Mod.	65	1.156	0.524	B
Sawtooth Freq. Mod.	79	2.515	0.375	C

* Higher preference scores indicate greater preference as an urgent signal

**Levels of the mean preference score with the same letters are not significantly different from each other at the 0.05 level.

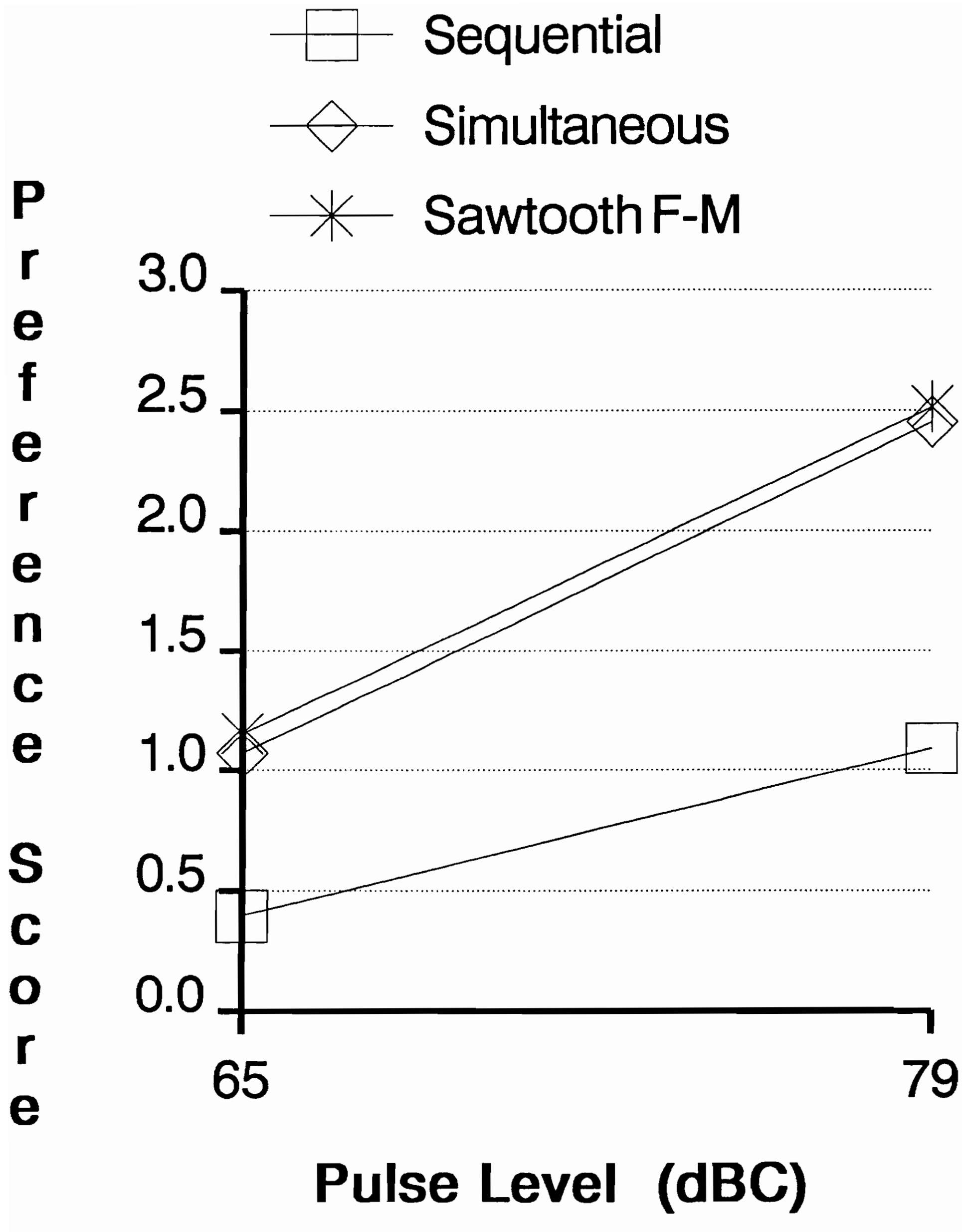


Figure 14. Mean preference scores (z units) for the pulse format x pulse level interaction.

permits the unambiguous interpretation of the pulse type and pulse level main effects.

Table 22 shows the results of the Newman-Keuls post-hoc test for the paired comparison ratings of pulse format. The means and confidence intervals are plotted in Figure 15. Subjects rated sequential pure-tone signals as sounding significantly less urgent than either simultaneous pure tones or the sawtooth frequency-modulated signal ($p \leq 0.05$). There is no significant difference in perceived urgency between the simultaneous and the frequency-modulated signal ($p > 0.05$).

The Newman-Keuls post-hoc test results for the paired comparison ratings of time between pulses are contained in Table 23. The means and confidence intervals are plotted in Figure 16. As in the magnitude estimation, subjects rated signals with an inter-pulse interval of 0 ms (no time between pulses) as sounding significantly more urgent than those containing inter-pulse intervals of 150 and 300 ms ($p \leq 0.05$). Signals with inter-pulse intervals of 150 ms were perceived as being significantly more urgent than those with intervals of 300 ms ($p \leq 0.05$).

The paired comparison ratings data for pulse level indicate that signals with higher pulse levels (79 dBC SPL)

TABLE 22

Newman-Keuls Test: Paired Comparison Data, Pulse Format

Pulse Format	Mean Preference Score* (z Units)	S.D. (z Units)	Significance**	
Sequential Pure Tones	0.91	0.70	A	
Simultaneous Pure Tones	1.76	0.82		B
Sawtooth Freq. Mod.	1.84	0.85		B

* Higher preference scores indicate greater preference as an urgent signal

**Levels of the mean preference score with the same letters are not significantly different from each other at the p = 0.05 level.

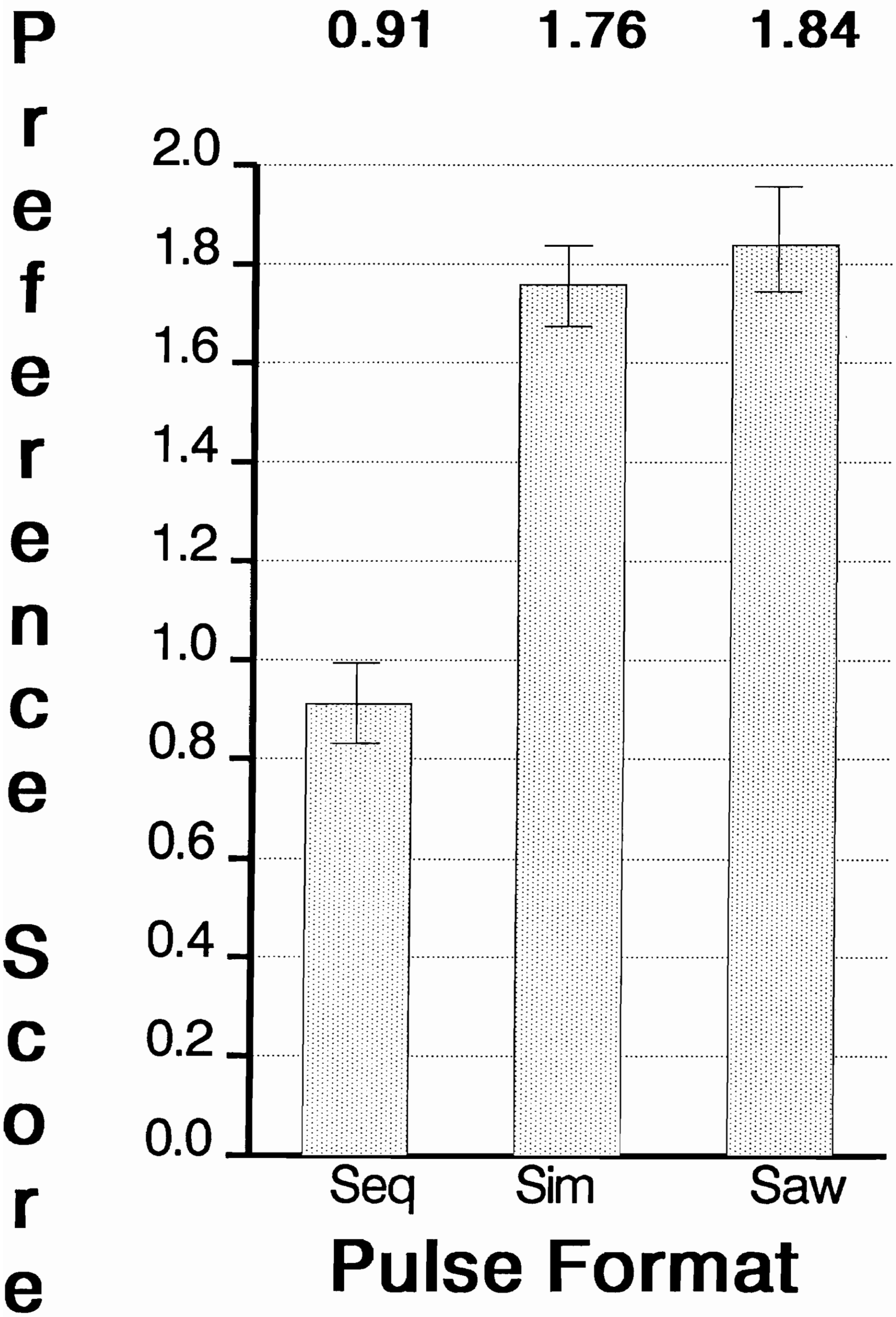


Figure 15. Mean preference scores and 95% confidence intervals (z units) for pulse format.

TABLE 23

Newman-Keuls Test: Paired Comparison Data, Time Between Pulses

Time Between Pulses (ms)	Mean Preference Score (z Units)	S.D. (z Units)	Significance**		
0.0	1.92	0.75	A		
150.0	1.51	0.87		B	
300.0	1.08	0.88			C

* Higher preference scores indicate greater preference as an urgent signal

**Levels of the mean preference score with the same letters are not significantly different from each other at the p = 0.05 level.

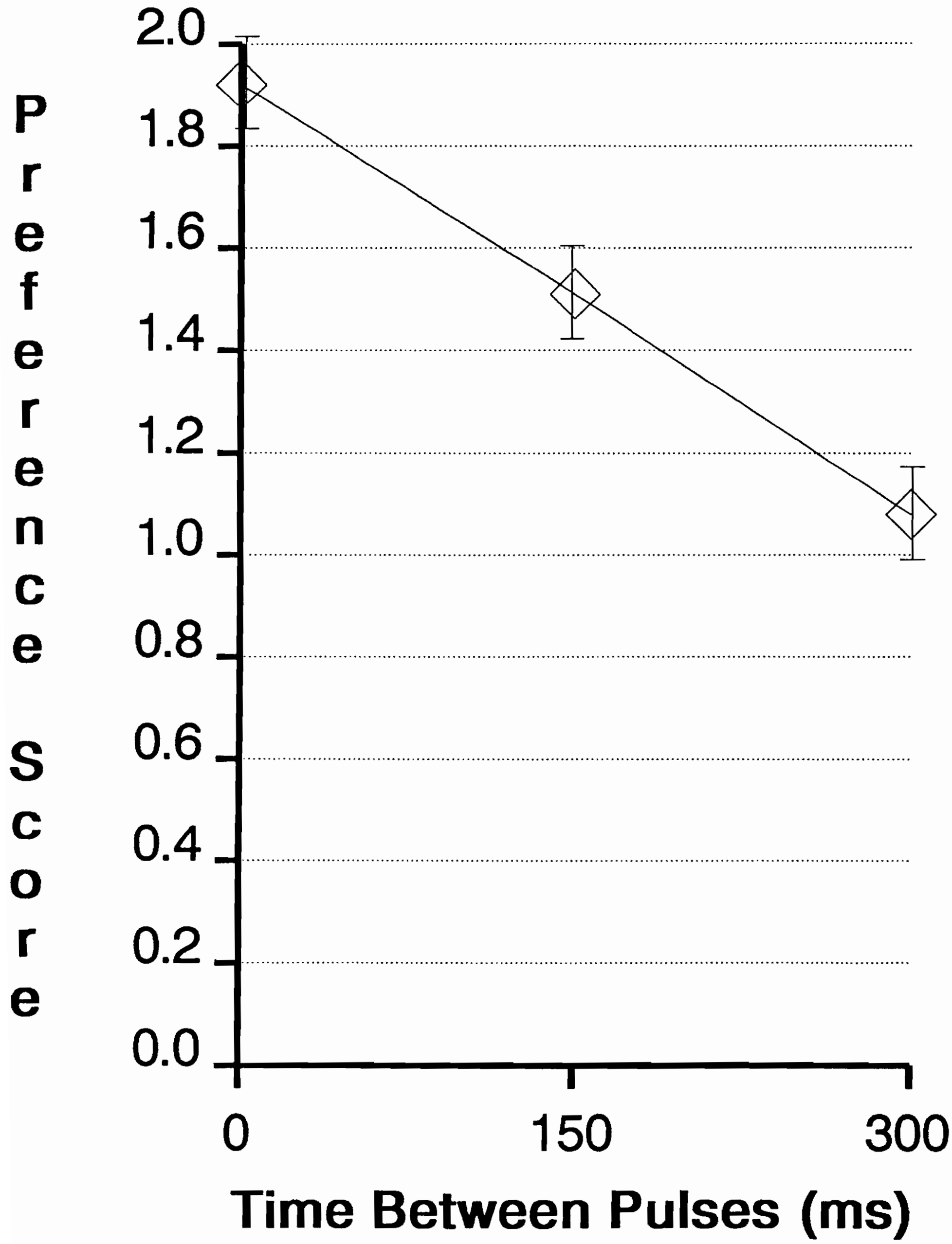


Figure 16. Mean preference scores (z units) and 95% confidence intervals for time between pulses.

received greater mean preference scores than those with lower pulse levels (65 dBC SPL). The mean preference scores (in $\underline{z}$ units) for the 65 dBC and 79 dBC pulse levels are 0.88 and 2.13, respectively. The standard deviations (in $\underline{z}$ units) for the 65 dBC and 79 dBC pulse levels are 0.55 and 0.63, respectively.

Response surface procedures (Myers, 1983) could not be performed on the paired comparison data because there were too few data points from which to plot each response surface. As previously stated, the paired comparison transformation reduced the data points to a final total of eighteen (18). Only six data points were available to define each response surface, because a separate response surface is required for each pulse format (the ANOVA for pulse format was significant). As a result, the statistical power of the regression would be very low due to the small number of data points.

Detection Time

Table 24 contains the MANOVA for the detection time data. Again, Wilk's criterion $\underline{U}$ was chosen as the test statistic. Only pulse format ($\underline{U} = 0.762$, $\underline{p} = 0.010$) and pulse level ($\underline{F} = 34.24$, $\underline{p} < 0.001$) are significant. Time

TABLE 24

Detection Time Data MANOVA Summary Table*

Source of Approx. Variance	dv	dfH	dfE	U	F	p
Between-Subjects						
Subjects	1	35				
Within-Subjects						
Pulse format (F) F X S	1	2	34	0.762	5.317	0.010
Time Between Pulses (T) T X S	1	2	34	0.878	2.361	0.110
Pulse Level (L) L X S	1	1	35	**	34.240	<0.001
F X T F X T X S	1	4	32	0.871	1.183	0.337
F X L F X L X S	1	2	34	0.968	0.558	0.579
T X L T X L X S	1	2	34	0.965	0.610	0.549
F X T X L F X T X L X S	1	4	32	0.929	0.612	0.657

where:
dv = number of dependent measures
dfH = degrees of freedom for treatment effect
dfE = degrees of freedom for error effect
U = Wilk's likelihood ratio statistic
p = significance of approx. F

* Denominators used for each source of variance in the U tests appear as the second term in each grouping in the table.

** When dfH = 1, the MANOVA statistic yields exact univariate (F) test results only.

between pulses, which is significant in both paired comparison and magnitude estimation measures, is not a significant effect in detection time ($U = 0.878$, $p = 0.110$). The interaction of pulse format and pulse level, which is significant in the perceived urgency ratings, is also not a significant effect in detection time ($U = 0.968$, $p = 0.579$).

Detection time ANOVA statistics are presented in Table 2 in Appendix E only to supplement information provided in the MANOVA. In the presence of Huynh-Feldt df corrections (actual decimal values were used) for applicable significant effects, all variables which are significant in the ANOVA are also significant in the MANOVA. Again, the "greater" ANOVA significance levels are due to the more "approximate" nature of the univariate approach in the presence of sphericity.

Table 25 shows the results of the Newman-Keuls post-hoc test for the detection times associated with pulse format. The means and confidence intervals are plotted in Figure 17. Subjects have significantly longer detection times with sequential pure tones than with simultaneous pure tones or sawtooth frequency-modulated signals ($p \leq 0.05$). There are no other significant differences.

Detection time statistics associated with pulse level

TABLE 25

Newman-Keuls Test: Detection Time Data, Pulse Format

Pulse Format	Mean Detection Time (ms)	S.D. (ms)	Significance*	
Sequential Pure Tones	492	152	A	
Simultaneous Pure Tones	452	119		B
Sawtooth Freq. Mod.	465	130		B

*Levels of mean detection time with the same letters are not significantly different from each other at the $p = 0.05$ level. An H-F epsilon correction of 0.780 was used to compute the error degrees of freedom used in this test.

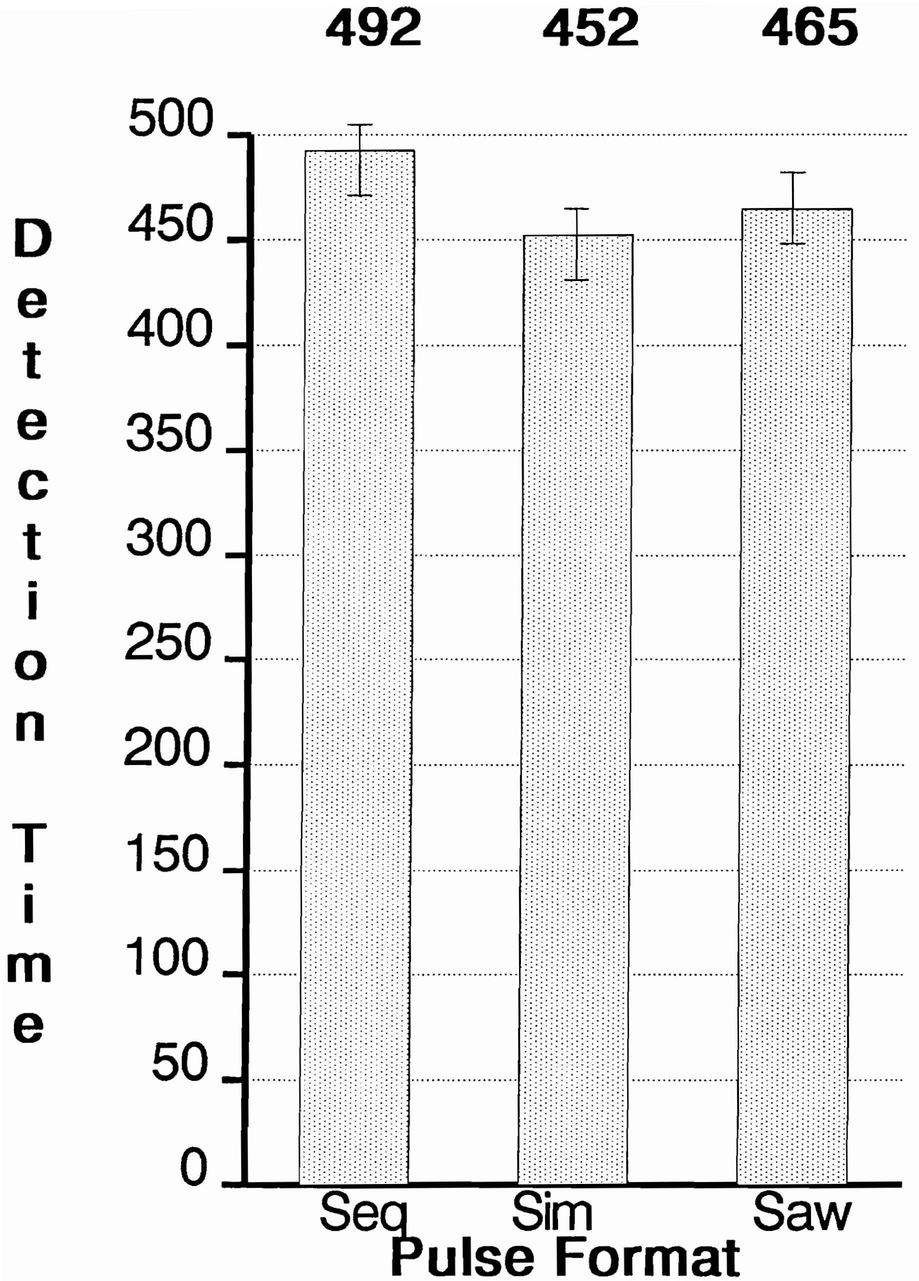


Figure 17. Mean detection times (seconds) and 95% confidence intervals for pulse format.

indicate that subjects have significantly shorter mean detection times for signals with higher pulse levels. The mean detection times for the 65 dBC and 79 dBC pulse levels are 500 ms and 440 ms, respectively. The standard deviations for the 65 dBC and 79 dBC pulse levels are 130 ms and 134 ms, respectively.

Response surface methodology was not applied for the detection time data. The response surface would incorporate only pulse level, because it is the only significant continuous variable in the detection time MANOVA. Because pulse level has only two levels, the presence of curvilinearity or any second-order effects could not be detected because three or more levels of an independent variable are required to detect nonlinearity. In this case, information regarding the optimum values of detection time for each pulse level is best provided by means and standard deviations. As was previously described, detection time statistics associated with pulse level indicate that subjects have significantly shorter mean detection times for signals with higher pulse levels. Detection time is clearly shortest when pulse level is greatest (79 dBC).

Responses to the Modified Cooper-Harper mental workload rating scale indicated that subjects rated the level of mental workload for the detection time task (i.e., responding to both visual bias conditions and auditory signals) as being acceptable. Subjects reported that the task could be completed with fair to mild difficulty (the median difficulty rating is 3.0, with the scores ranging from 1 to 10). Of the 34 subjects who completed the scale, 18 responded that the mental workload was acceptable, with a difficulty level that was either very easy, easy, or fair to mild. The remaining 16 respondents indicated that mental workload was high and should be reduced, but indicated that the detection time task presented minor or moderately objectionable difficulty. No subjects indicated that the task presented major difficulty. From this metric of workload, it can be concluded that the CTS provided a low to moderate attentional demand for half of the subjects and was somewhat difficult for the other half.

DISCUSSION

The objectives of this research were threefold. The first was to investigate the effect of the independent variables (pulse format, signal level, and time between pulses) on the perceived urgency of warning signals. The second was to investigate the effect of these variables on the detectability of warning signals. The third was to examine the relationship between the perceived urgency and detectability of warning signals. Each of these objectives was met and is discussed in further detail.

The Effect of the Independent Variables on the Perceived Urgency of Warning Signals

Both magnitude estimation and paired comparison data were applied as measures of perceived urgency. As originally hypothesized, perceived urgency was influenced by temporal, level, and spectral characteristics.

Pulse level. The effect of pulse level on perceived urgency had not been explored prior to this study. Edworthy (1991) suggested that there may be a strong relationship between signal level and perceived urgency. The results of this study confirm Edworthy's hypothesis. Pulse level is significant in both the magnitude estimation and paired comparison measures. Post-hoc tests indicate that the higher the pulse level, the greater is the perceived urgency of the signal. Pulse level is the strongest effect in the paired comparison ANOVA, accounting for more than half (56%) of the total variance.

Future experiments should investigate even higher pulse levels but should exercise caution for the effects of over-exposing subjects. Standards and research recommendations state that auditory warnings should exceed the level of ambient noise by at least 15 to 30 dB (Patterson, 1982), or should have an A-weighted sound level which exceeds the level of ambient noise by 15 to 30 dB (ISO, 1986). The pulses used in this study were somewhat lower than was recommended. The 79 dBC pulse exceeded the 68 dBC ambient noise level by only 11 dBC. The level of the ambient noise exceeded the 65 dBC pulse by 3 dBC. These low levels presented no detectability problems whatsoever in this study. For both pulse levels, the positive signal/noise ratios in at least one 1/3 octave band permitted subjects to detect and respond to every signal each time it occurred. There were no occurrences in which subjects did not detect and respond to a signal. To examine the perceived urgency and detectability of signals which approach the recommended levels, future experiments should incorporate pulse levels which range from 15 to 30 dB above the level of ambient noise.

Time between pulses. Hellier and Edworthy (1989) found that the interval between signal pulses had a consistent effect on the perceived urgency of warning signals. Their results were replicated in this study, which indicate that, as was hypothesized, signals with shorter inter-pulse intervals were rated as significantly more urgent. Signals with no time between pulses (0 ms) were rated as most urgent of all. The paired comparison data indicate that this effect is relatively weak, accounting for a relatively small proportion (16 percent) of the total variance in the ANOVA.

This study did not explore a wide practical range of inter-pulse intervals. The intervals used in this experiment (0 to 300 ms) are relatively short, which conform to recommendations made by Patterson (1982). While this range permitted the exploration of shorter intervals, which are rated as more urgent, it did not permit an examination of the relationship between perceived urgency and relatively long inter-pulse intervals. Future experiments should include longer between-pulse intervals, such as those ranging from 500 ms to 2 s, which would allow for quiet periods in which speech communication could take place between signal pulses.

Response-surface methodology was used to determine what values of pulse level and time between pulses produced the optimum perceived urgency, as measured by magnitude estimation. For each signal type, perceived urgency is greatest when pulse level is greatest and when time between pulses is smallest. For response surface procedures to be used to greater advantage, future experiments should employ three or more levels of each continuous, quantitative independent variable to allow detection of the presence of second-order effects. A fractional factorial central composite design (Myers, 1976) should be considered as a potential data collection technique because it permits an economy of data collection while permitting the derivation of higher-order polynomial approximations of all factors in the response surface. When fractional factorials are incorporated into the central-composite design, at least one factorial effect, the defining contrast, is lost entirely. However, if the effect used as the defining contrast is a higher-order interaction that seldom affects performance, then little relevant information will be lost in the data analysis.

Pulse format. The perceived urgency of some formats (or types) of multitone and frequency-modulated signals had not been explored prior to this study. It was expected that at least one pulse format would be rated as more urgent than another. Both the magnitude estimation and paired comparison data indicate that sequential pure tones are perceived as significantly less urgent than simultaneous and frequency-modulated tones. The magnitude estimation data indicate that the perceived urgency of the simultaneous and frequency-modulated pulse formats differ depending upon time between pulses and pulse level. The paired comparison data indicate that there is no significant difference between the simultaneous and frequency-modulated tones. The paired comparison data indicate that pulse format is a relatively strong effect, accounting for approximately 24 percent of the total variance.

Because magnitude estimation and paired comparison results have a high, significant correlation ($\underline{r}$ = 0.95), future research could justifiably use only one of the two as a metric of perceived urgency. Free modulus magnitude estimation (Stevens, 1957) is the author's measure of choice. In this study, magnitude estimation data were obtained in only one-half the time of the paired comparison data. The magnitude estimation data transformation also retained a greater proportion of data points for the final data analysis, which permitted the use of response surface methodology.

The Effect of the Independent Variables on the Detection Time of Warning Signals

Signal detection time was influenced by level and spectral factors. Pulse level and pulse format were the only significant variables.

Pulse level. Prior to this study, there was little research describing the manner and extent of the relationship between signal level and detection time. Teichner (1954) found that the reaction time to an auditory stimulus becomes shorter, up to an asymptotic value, as the intensity of the stimulus is increased. The results of this study partially concur with this finding. Signals with a higher pulse level have shorter detection times, although the asymptote was never reached. However, the mean pulse level detection times (440 and 500 ms for 79 dBC and 65 dBC levels, respectively) are greater than the typical 140 ms reaction time to auditory stimuli predicted by Woodworth and Schlosberg (1954). The increased detection times observed in this experiment are in most part due to the distraction imposed by the loading task and by the presence of the 68 dBC ambient pink noise during the detection task.

The difference in mean detection times for high and low pulse levels is 60 ms, which in terms of human response is an extremely short interval. A response time difference of 60 ms may be important to a pilot performing nap-of-earth maneuvers in a high-speed aircraft where fast response is essential. Short response time differences may also be important to operators of automobiles, who may have to respond quickly to warnings of sudden changes in traffic or environmental conditions. Operators of high-speed trains may also benefit from short response-time differences, when they have to respond quickly when warned of sudden changes occurring on or around their path of travel. In applications where such short response is not essential, such as responding to a fire alarm, a difference of such small magnitude may not make a practical difference.

Pulse format. Previous research had not explored the relationship between pulse format and detection time. The results of this study indicate that there is a significant difference in the detection time for different pulse formats. The mean detection time for sequential pulses is significantly greater than for the frequency-modulated and the simultaneous pulses. The mean response times for pulse format, which range from 452 ms to 492 ms, are greater than the typical 140 ms reaction time to auditory stimuli predicted by Woodworth and Schlosberg (1954). As with the pulse level data, the greater response times observed in this experiment were in most part due to the distraction imposed by the loading task and by the presence of the 68 dBC ambient pink noise during the detection task.

The largest difference in mean detection times is 40 ms, which, in terms of human response, is an extremely short interval. The relevance of a difference of such small magnitude is application-dependent. As before, a response time difference of 40 ms may be important to persons performing tasks which require extremely fast response, but in applications where such short response is not essential, a difference of such small magnitude may not make a practical difference.

Listener attentional demands. Demands on listener attention were created by the presence of the probability monitoring task and by subject uncertainty as to the time of occurrence of both the visual bias condition and the auditory signal. As was previously stated, Criterion Task Set (CTS) probability monitoring task data were not included as dependent variables but were collected only to assure that subjects were attending to the loading task. Close monitoring of subject performance for the probability monitoring task (reaction time to correctly identify signals, number of false positives, number of misses) indicated that subjects were attending to the task throughout the experimental session.

Although the subjects' level of attention to the loading task was sufficient, the level of difficulty provided by the task was not high. In a sense, the probability monitoring task did not completely fulfill the objective, which was to present a workload task demanding a high level of operator mental effort in order to obtain the maximum difference between independent variables. CTS developers (Shingledecker, 1984) and validators (Schlegel and Gilliland, 1990) stated that the "high" loading level for this task did contain a high level of difficulty. But Modified Cooper-Harper workload ratings and post-task interviews obtained in this study indicated that subjects perceived the task to be only mildly difficult and not very demanding. This difference in ratings could be due to the difference in workload rating scales. CTS validation studies used the Subjective Workload Assessment Technique (SWAT) as the subjective measure of difficulty and reliable difference in task performance as the objective measure (Amell, Eggemeier, and Acton, 1987). In any case, future experiments should include a task with a more demanding workload level. Pilot testing involving the utilization of a workload scale and subject interviews should be performed to realistically determine the level(s) of difficulty of potential loading tasks.

Future research might also explore the effect of different levels of attentional demand on the detection time of auditory signals. After a loading task of sufficient difficulty has been defined, subjects could be exposed to different loading levels of that task. Alternatively, performance under various loading conditions could be compared to performance with no loading task.

The Relationship Between The Perceived Urgency and Detection Time of Warning Signals

The correlation for the paired comparison and magnitude estimation data indicate that the two measures are highly correlated (r = 0.95), and that the correlation is statistically significant. The correlational analysis for the magnitude estimation and detection time data indicate that there is a negative correlation of small magnitude (r = - 0.164) between these variables. Nonetheless, the correlation proved to be statistically significant. This indicates that there is probably a small but significant relationship between perceived urgency (as measured by both paired comparison and magnitude estimation) and detection time. As perceived urgency increases, detection time probably decreases.

As was hypothesized, pulse level is the one signal design characteristic which consistently affects all measures of perceived urgency and detection time. A significant difference between pulse levels was found in the magnitude estimation, paired comparison, and detection time measures. The higher the pulse level, the greater is the perceived urgency of the signal and the shorter is the detection time for that signal.

Pulse format significantly affects detection time and one measure (paired comparison) of perceived urgency. Subject detection time for sequential pure tone signals is significantly greater than for simultaneous and frequency-modulated signals. Subjects also rated this signal as sounding less urgent than the others. Time between pulses has no effect on detection time.

Unexpected Results

One unexpected outcome was the significant difference in pulse format detection times. Results indicated that the response latency for sequential pulses is significantly greater than for the frequency-modulated and the simultaneous pulses. It might be expected that pulse format would not be significantly different, because the subjects would respond only to the stimulus onset, not to differences in pulse format, which primarily occur after onset. However, other researchers (Adams and Trucks, 1967) also reported significantly different reaction times to different signal types which change over time. These results indicate that subject response latency may be dependent upon some spectral or temporal aspect of the signal pulse that occurs after the moment of onset. Therefore, simple reaction time may not be a pure measure of detectability for signal pulses which change over time. Future research should explore different measures of detectability for signal pulses which occur above masked threshold, and which change over time.

A second unexpected observation was that subjects did not perceive existing differences in the amplitude envelopes of different pulse formats. Figures 18a through 20b compare pulse format amplitude envelopes at the loudspeaker and at the listener's ear, without the broadband noise background. The envelopes for all pulses show a relatively constant amplitude over time at the loudspeaker (Figures 18a, 19a, and 20a) with a reduction in amplitude at the listener's ear (Figures 18b, 19b, and 20b). A greater amplitude fluctuation occurs at the listener's ear for both the sequential signals and frequency-modulated pulse signals, than for the simultaneous signals. None of these fluctuations were detectable at the listener's ear position. When asked to listen to the signals, subjects not only indicated that they perceived no between-format difference in amplitude fluctuation, but also indicated that they perceived no occurrence of amplitude fluctuation in any signal. At present, the experimenter can offer no theory regarding the failure of subjects to detect any difference in amplitude fluctuation for the different pulse formats.

The cause of the amplitude envelope fluctuation is hypothesized to be the constructive interference (positive addition or phase reinforcement) and destructive interference (negative addition or phase cancellation) of pulse components reflected from the floor (the primary reflective surface) of the semi-reverberant chamber. The precise characteristics of the constructive and destructive interference are dependent upon the room acoustics and the position of the listener in the room. To test the hypothesis of constructive and destructive interference as a cause of amplitude fluctuation, future research could be conducted to determine the extent to which the interference of each pulse format occurs in reverberant environments, and whether the differences found in this study hold for free-field environments. If the hypotheses are true, all signals will show less amplitude fluctuation in a free-field or

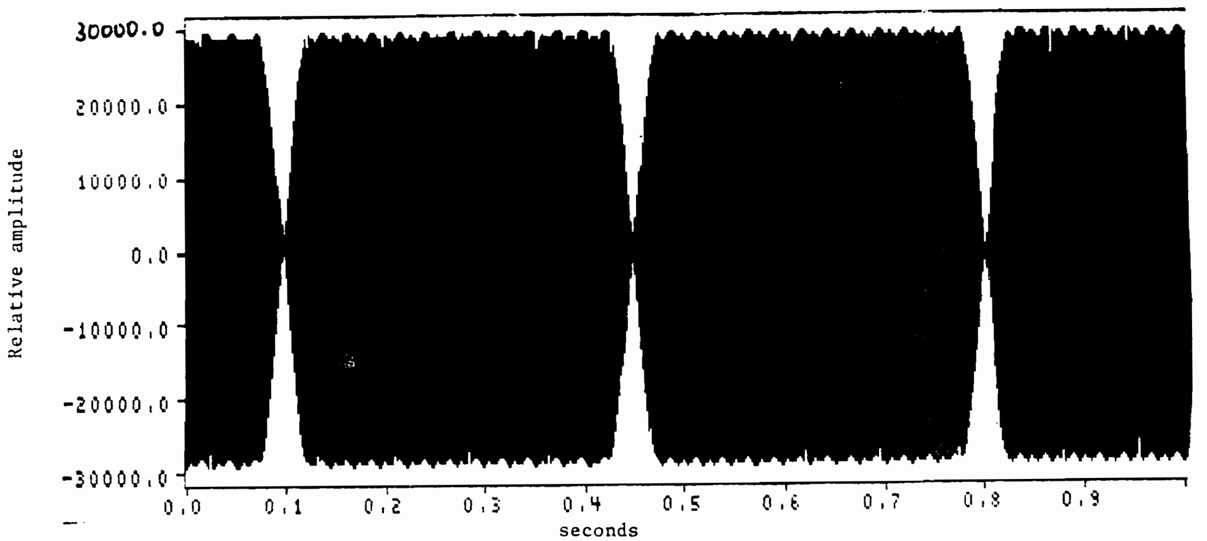


Figure 18(a). Amplitude envelope for the simulataneous **pulse format**, at the loudspeaker.

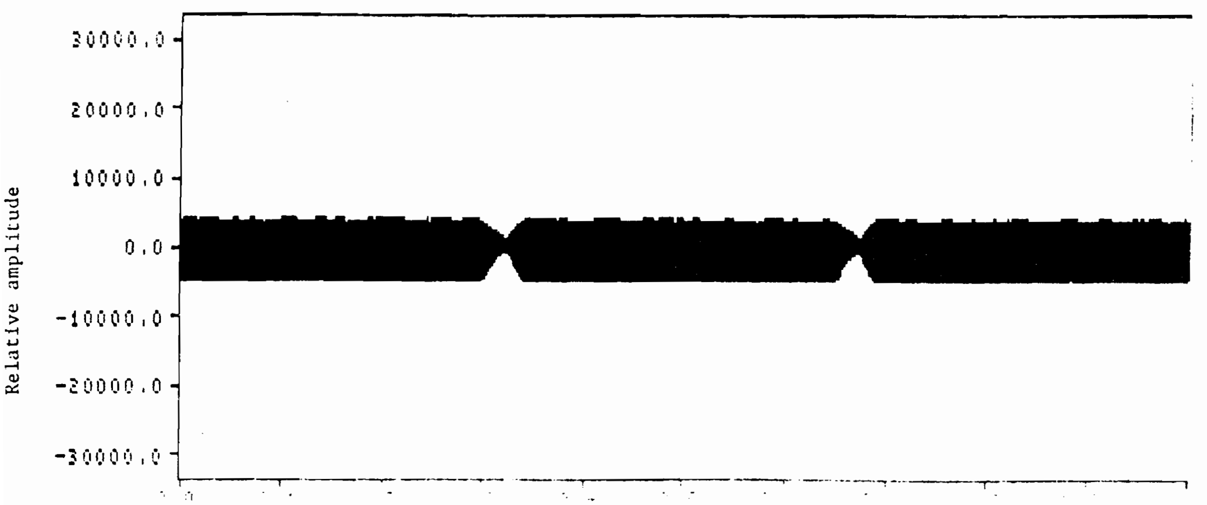


Figure 18(b). Amplitude envelope for the simultaneous pulse format, at the listener's ear.

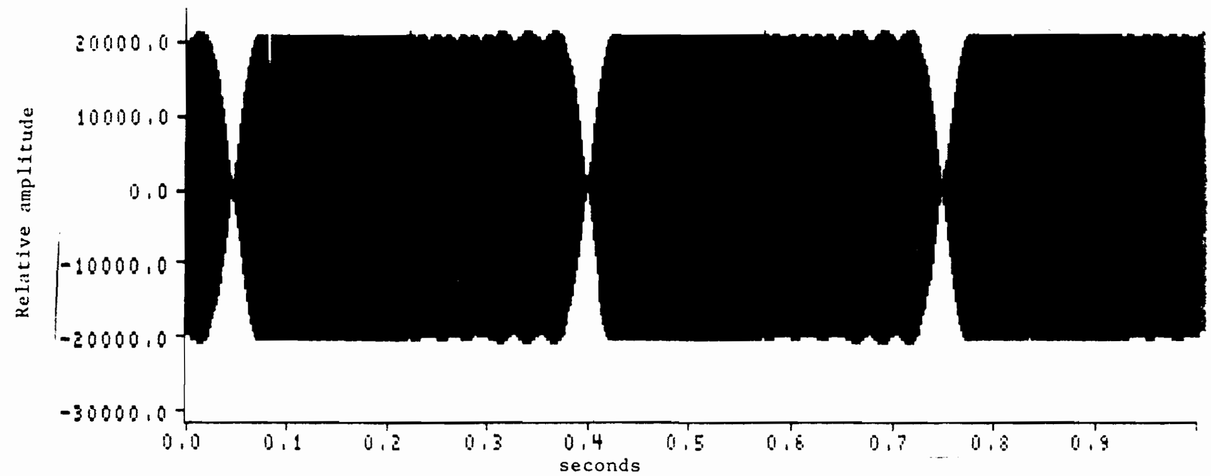


Figure 19(a). Amplitude envelope for the sequential pulse format, at the loudspeaker.

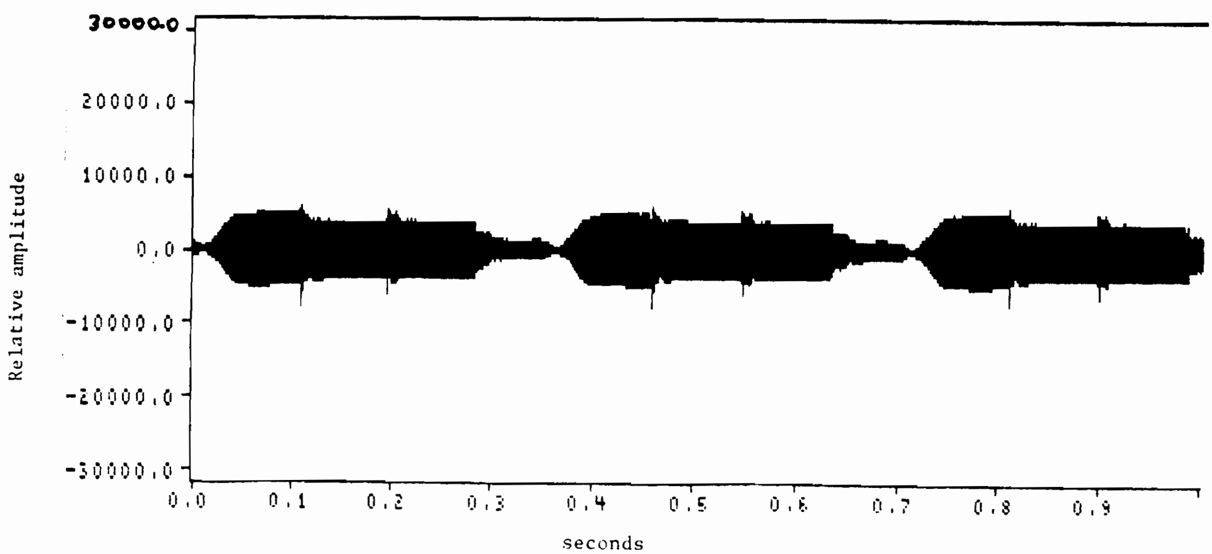


Figure 19(b). Amplitude envelope for the sequential pulse format, at the listener's ear.

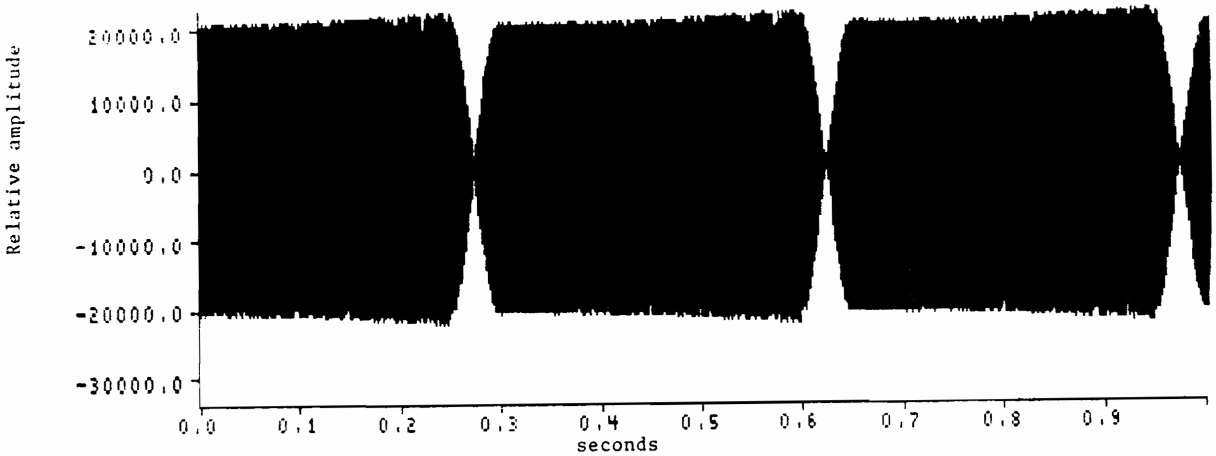


Figure 20(a). Amplitude envelope for the frequency modulated sawtooth pulse format, at the loudspeaker.

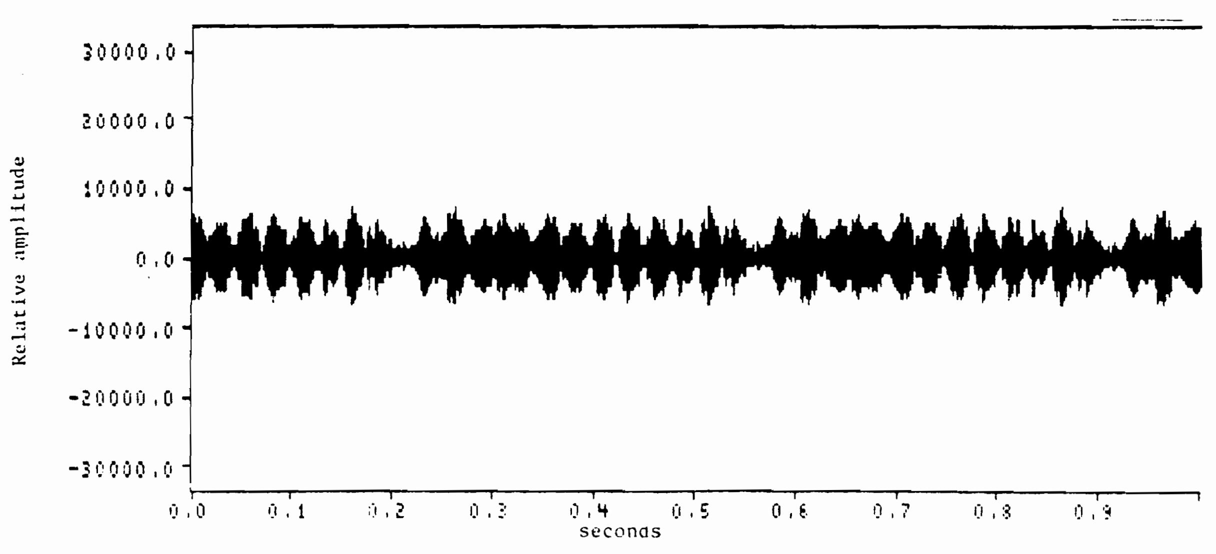


Figure 20(b). Amplitude envelope for the frequency modulated sawtooth pulse format, at the listener's ear.

anechoic chamber than in a reflective field or reverberant environment due to the absence of reflective surfaces which can cause standing waves and superposition effects. The reverberant environment's potential for standing waves, in which the sound would not be scattered but would occur in a succession of maxima and minima, offers a severe test for warning signal design. Standing waves, which result from both constructive and destructive interference, can be detected by moving a microphone along the principal wave propagation path and noting the presence of a succession of sound level maxima and minima.

RECAPITULATION

The overall purpose of this experiment was to investigate potential signal pulse parameters which provide a specific level of perceived urgency and detectability in environments with broadband ambient noise. Conditions included a loading task which presented additional attentional demands upon the subject during the signal detection task. Signal effects were measured using a multivariate approach.

The research fulfilled its overall purpose. A detailed, testable description of the relationship among auditory warnings, perceived urgency, and detection time was obtained for the parameters utilized in this experiment. Results indicate that both perceived urgency and detection time are influenced by pulse level and format. The higher the pulse level, the greater is the perceived urgency of the signal and the shorter is the detection time. Subjects took longer to detect sequential pure tone signals and rated those signals as less urgent than signals with simultaneous or frequency-modulated pure tones. Under most conditions, there is no significant difference in the perceived urgency or detection time of simultaneous and frequency-modulated pulses.

Time between pulses has a consistent effect only on the perceived urgency of warning signals. The shorter the time between pulses, the greater is the perceived urgency of the signal. Signals with no time between pulses (0 ms) were rated as most urgent of all.

There are several military and industrial environments where the perceived urgency of auditory warning signals should be of issue. Military and civilian aircraft cockpits utilize several acoustic signals, may of which can be activated simultaneously. The British Aircraft Corporation (BAC) Boeing 747 has several auditory warnings, including overspeed, landing-configuration, and takeoff-configuration warnings (Patterson, 1982). Hospital wards such as critical care, post-natal intensive care, and hemodialysis units often assemble many pieces of physiological monitoring equipment to observe patient status. Each piece of equipment may use several acoustic signals as well as several levels of signals to signify change in patient condition as well as status of equipment operation. In each application, urgency coding can aid the listener in distinguishing the level of response to each signal.

Signals designed to communicate a lower level of perceived urgency also serve an important purpose in military and industrial applications. Auditory icons (sometimes known as earcons) are caricatures of naturally occurring sounds, in which dimensions of a sound's sources are used to convey information. One example of an auditory icon is a computer generating an auditory signal which sounds like a ticking clock, to indicate to the user that the computer is processing information. From this perspective, sound provides information about materials interacting at a location in the environment. The strategy behind auditory icons is to use naturally occurring sounds as a source of sound to stand for a source of information. The advantage of this approach is that if a good mapping between a source of sound and a sound signal can be defined, the meaning of an auditory icon should be easily learned and remembered.

SUGGESTIONS FOR FUTURE RESEARCH

1) Pulse level affected both perceived urgency and signal detection time and should be investigated further. Future research should determine the relationship between signal level and perceived urgency, using a greater number and range of pulse levels. To examine the perceived urgency and detectability of signals which approach the recommended levels, future experiments should incorporate pulse levels which range from 15 to 30 dB above the level of ambient noise. However, care should be taken that these pulse levels provide no threat of subject hearing loss.

2) For generalization of results to a larger number of environments, future studies should present warning signals against other types of background noise spectra, such as white noise, voice babble, speech spectrum, vehicle, and aircraft noise. The background noise could also be presented at different levels. Suggested levels include a reference of 40 dB (level at a quiet residence) (Deatherage, 1972); 60 dB (level of conversational speech at 1.5 m) (Deatherage, 1972); 85 dB (level at an average factory); and 100 dB (level at a noisy factory) (Adams and Trucks, 1976). More severe environments, such as those inside helicopter cockpits, should also be included as special "worst-case" design situations.

3) Subsequent studies should utilize a loading task with a more demanding workload level. Future research might also explore the effect of different levels of attentional demand on the detection time of auditory signals.

4) Based on the findings of this study, and for the sake of parsimony, it is reasonable to conclude that future studies could utilize only one measure of perceived urgency. Magnitude estimation would be preferred over paired comparison methodology, because magnitude estimation permits data collection in less time and retains a higher proportion of data points on a quantitative continuum, which permits response surface analysis.

5) Response surface methodology (Myers, 1976) was a useful tool in defining what values of significant continuous independent variables are optimum (greatest) for a dependent variable. For response surface procedures to be used to greater advantage, future experiments should employ three or more pulse levels and times between pulses in order to detect the presence of second-order effects. A fractional factorial central composite design (Myers, 1976) may be considered as a useful data collection technique in the response surface analysis of higher-order functions.

6) Future studies might utilize signal identification as a criterion variable in signal design. Identification of signals is important in situations where multiple signals may occur and the listener must identify and react to specific signals. The effect of multidimensional coding on signal identification could be studied. Listeners can identify a limited number of signals (approximately seven) when they are coded in one stimulus dimension (i.e., with differences in intensity, pitch, or interruption rates) (Miller, 1956). However, if auditory signals with multidimensional codes are utilized as signal pulses, signal identifiability may improve. In addition, the effect of auditory icons ("earcons"), which are caricatures of naturally occurring sounds, could be studied.

REFERENCES

Abel, S.M., Alberti, P.W., Haythornthwaite, C., and Riko, K. (1982). Speech intelligibility in noise: Effects of fluency and hearing protector type. Journal of the Acoustical Society of America, 71 (3), 708-715.

Abel, S.M., Kunov, H., Pichora-Fuller, M.K., and Alberti, P.W. (1985). Signal detection in industrial noise: Effects of noise exposure history, hearing loss, and the use of ear protection. Scandinavian Audiology, 14, 161-173.

Acton, W.H., and Crabtree, M.S. (1985). User's guide for the criterion task set (Tech. Report AAMRL-TR-85-034). Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory.

Adams, S.K., and Trucks, L.B. (1976). A procedure for evaluating auditory warning signals. In Proceedings of the Human Factors Society 20th Annual Meeting (pp. 166-172). Santa Monica, CA: Human Factors Society.

Amell, J.R., Eggemeier, F.T., and Acton, W.H. (1987). The criterion task set: An updated battery. In Proceedings of the Human Factors Society 31st Annual Meeting (pp.405-409). Santa Monica, CA: Human Factors Society.

American National Standards Institute (ANSI) (1973). American National Standard psychoacoustical terminology (ANSI S3.20). New York, New York: American National Standards Institute.

ANSI S3.21-1978. (1978). Methods for manual pure-tone audiometry. New York: American National Standards Institute, Inc.

ANSI S12.6-1984. (1984). Method for the measurement of real-ear attenuation of hearing protectors. New York: American National Standards Institute, Inc.

ANSI S3.6-1989. (1989). American National Standard specifications for audiometry. New York: American National Standards Institute, Inc.

Berger, E.H. (1981). Re-examination of the low-frequency (50-1000 Hz) normal threshold of hearing in free and diffuse sound fields. The Journal of the Acoustical Society of America, 70 (6), 1635-1645.

Berger, E.H. (1986). Hearing protection devices. In E.H. Berger, W.D. Ward, J.C. Morrill, and L.H. Royster (Eds.), Noise and hearing conservation manual (pp. 37-95). Akron, Ohio: American Industrial Hygiene Association.

Bock, M., Lazarus, H., and Hoege, H. (1982). The effects of noise on the efficiency of danger signals. Unveroffentlichter Forschungsbericht det Bundesanstalt fur Arbeitsschutz und Unfallforschung. Dortmund, Germany, 517-521.

Botsford, J.H. (1970). Current trends in hearing damage risk criteria. Sound and Vibration, 4 (4), 16-19.

Bray, J.H., and Maxwell, S.E. (1985). Sage university papers: Multivariate analysis of variance. London: Sage Publications.

Casali, J.G., and Horylev, M.J. (1987). Speech discrimination in noise: The influence of hearing protection. In Proceedings of the Human Factors Society 31st Annual Meeting (pp. 1246-1250). Santa Monica, CA: Human Factors Society.

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.

Chiles, W.D., and Alluisi, E.A. (1979). On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 21 (5), 515-528.

Cohn, J. (1894). Experimentelle Untersuchunger uber die Gefuhlsbetonung der Farben, Helligkeiten, und ihrer Combinationen. Philosophische Studien, 10, 562-603.

Coles, R.R.A. and Rice, C.G. (1965). Letter to the editor: Earplugs and impaired hearing. Journal of Sound and Vibration, 3 (3), 521-523.

Creelman, D.C. (1973). Simultaneous adaptive threshold estimation with frequency uncertainty. Journal of the Acoustical Society of America, 54 (1), 316.

Daniel, W.W. (1990). Applied nonparametric statistics (2nd ed.). Boston: PWS-Kent.

Deatherage, B.H. (1972). Auditory and other sensory forms of information presentation. In H.P. Van Cott and R.G. Kinkade (Eds.), Human engineering guide to equipment design (Rev. ed, pp. 123-160). Washington D.C.: U.S. Government Printing Office.

Doll, T.J., and Folds, D.J. (1986). Auditory signals in military aircraft: Ergonomics principals versus practice. Applied Ergonomics, 17 (4), 257-264.

Draper, N,. and Smith, H. (1966). Applied regression analysis. New York: Wiley.

Drazin, D. (1961). Effects of fore-period, fore-period variability and probability of stimulus occurrence on simple reaction time. Journal of Experimental Psychology, 62, 43-50.

Durlach, N.I., and Colburn, H.S. (1978). Binaural Phenomenon. In E.C. Carterette and M.P. Friedman (Eds.), Handbook of perception, Vol. IV Hearing. New York: Academic Press.

Durrant, J.D., Lovrinic, J.H. (1984). Bases of hearing science (2nd ed.). Baltimore: Williams and Wilkins.

Earshen, J. J. (1986). Instrumentation and noise descriptors. In E.H. Berger, W.D. Ward, J.C. Morrill, and L.H. Royster (Eds.), Noise and hearing conservation manual (pp. 37-95). Akron, Ohio: American Industrial Hygiene Association.

Edworthy, J. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. Human Factors, 33 (2), 205-232.

Edworthy, J., and Loxley, S. (1990). Auditory warning design: The ergonomics of perceived urgency. In E.J. Lonesey (Ed.), Contemporary Ergonomics 1990. London:Taylor and Francis.

Edworthy, J., Loxley, S., Geelhoed, E., and Dennis, I. (1989). The perceived urgency of auditory warnings. Proceedings of the Institute of Acoustics, 11 (5), 73-81.

Engen, T. (1971a). Psychophysics I: Discrimination and detection. In Kling, J., and Riggs, L. (Eds.), Woodworth and Schlosberg's Experimental Psychology (3rd ed., pp.11-34). New York: Holt, Rinehart and Winston.

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

Fechner, G.T. (1860). Elements der Psychophysik. Leipzig: Breikopf and Harterl. English translation of Vol.1 by H.E. Adler (D. H. Howes and E.G. Boring, Eds.) New York: Holt, Rinehart and Winston, 1966.

Fletcher, H. (1940) Auditory patterns. Review of Modern Physics, 12, 47-65.

Fletcher, H. and Munson, W.A. (1933). Loudness, its definition, measurement and calculation. Journal of the Acoustical Society of America, 5 (2), 82-108.

Forshaw, S.E. (1977). Listening for machinery malfunctions in noise while wearing ear muffs (Tech. Report DCIEM-TR-

77X43). Defence and Civil Institute of Environmental Medicine, Ontario, Canada: Behavioural Division.

Garner, W.R., and Miller, G.P. (1947). The masked threshold of pure tones. Journal of Experimental Psychology, 37, 293-303.

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

Green, D.M. (1961). Detection of auditory sinusoids of uncertain frequency. The Journal of the Acoustical Society of America, 33 (7), 897-903.

Green, D.M. (1976). An introduction to hearing. Hillsdale, NJ: Lawrence Erlbaum Associates.

Grice, G. (1968). Stimulus intensity and response evocation. Psychological Review, 75, 359-373.

Hawkins, J.E. Jr., and Stevens, S.S. (1950). The masking of pure tones and speech by white noise. Journal of the Acoustical Society of America, 22, 6-13.

Hellier, E., and Edworthy, J. (1989). Quantifying the perceived urgency of auditory warnings. Canadian Acoustics, 17 (4), 3-11.

Hellier, E., and Edworthy, J. (1990). Altering the urgency of auditory warnings: An experimental study. In E.J. Lonesey (Ed.), Contemporary Ergonomics 1990. London:Taylor and Francis.

ISO R226-1961 (E). (1961). Normal equal-loudness contours for pure tones and normal thresholds of hearing under free-field listening conditions. Switzerland: International Organization for Standardization.

Johnson, D.L., and Nixon, C.W. (1974). Simplified methods for estimating hearing protection performance. Sound and Vibration, 8 (6), 20-27.

Kalikow, D. (1967). Psychofit. Unpublished computer program (FORTRAN) for the analysis of magnitude estimates, Brown University, Providence, Rhode Island. (Expansion of a program by D.W. Panek and J.C. Stevens).

Karplus, H., and Bonvallet, G. (1953). A noise survey of manufacturing industries. American Industrial Hygiene Association, 14, 235-263.

Kemmerling, P., Geiselhart, R., Thornburn, D.E., and Cronburg, J.G. (1969). A comparison of voice and tone warning systems as a function of task loading (Tech. Report ASD-TR-69-104). Wright-Patterson Air Force Base, OH: Aeronautical Systems Division.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, N.J.: Prentice-Hall.

Kohfeld, D.L. (1971). Simple reaction time as a function of stimulus intensity in decibels of light and sound. Journal of Experimental Psychology, 88 (2), 251-257.

Kreezer, G.L. (1959). Attention value of audio and visual signals (WADC Tech. Report AD 58-521). Wright-Patterson Air Force Base, OH: Aeromedical Laboratory.

Lane, H., Catania, A., and Stevens, S. (1961). Voice level: Autophonic scale, perceived loudness, and the effects of sidetone. Journal of the Acoustical Society of America, 33, 160-167.

Lambert, D.R. (1980). Effect of hearing protectors on thresholds of machinery cues in noise (Experiment 1:

Naive subjects). Unpublished paper, Naval Ocean Systems Center, San Diego, CA.

Laroche, C., Tran Quoc, H., Heut, R., and McDuff, S. (1989). "Detectsound": A computerized model for the prevention of accidents in noisy workplaces. In Proceedings of the Annual International Industrial Ergonomics and Safety Conference. Montreal, Canada.

Lawrence, M. and Yantis, P.A. (1956). Onset and growth of aural harmonics in the overloaded ear. Journal of the Acoustical Society of America, 28, 852-858.

Lazarus, H. (1983). Literature review 1978-1983 in German language on noise and communication. In G. Rossi (Ed.) Proceedings of the Fourth International Congress on Noise as a Public Health Problem (pp. 425-451). Milan, Italy: Centro Richerche e Studi Amplifon.

Michael, P. (1982). Single-number performance factors for hearing protectors. In P. Alberti (ed.), Personal Hearing Protection in Industry (pp. 221-247). New York:Raven Press.

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

Momtahan, K.L., and Tansley, B.W. (1989). An ergonomic analysis of the auditory alarm signals in the operating room and recovery room. Paper presented at the Annual Conference of the Canadian Acoustical Association, Halifax, Canada.

Morrill, J.C. (1984). Audiometric instrumentation: equipment choices, techniques. Occupational Health and Safety, 53 (10), 78-84.

Myers, R.H. (1976). Response surface methodology. Ann Arbor, MI: Edwards Brothers.

Nicolosi, L., Harryman, E., and Kreshek, J. Terminology of communication disorders speech-language-hearing. Baltimore:Williams and Wilkins

Ogden, G.D., Levine, J.M., and Eisner, E.J. (1979). Measurement of workload by secondary tasks. Human Factors, 21 (5), 529-548.

Patterson, R.D. (1982). Guidelines for auditory warning systems on civil aircraft (Civil Aviation Authority Paper 82017). Cheltanham, England: Civil Aviation Authority, Airworthiness Division.

Patterson, R.D., Lower, M.C., Cosgrove, P., and Milroy, R. (1989). Sounds for the inductive loop warning system (ISVR Report No. AC672\3). Southampton, England: Institute of Sound and Vibration Research.

Patterson, R.D., and Milroy, R. (1980). Auditory warnings on civil aircraft: The learning and retention of warnings (Civil Aviation Authority Contract Number 7D\S\0142). Cambridge, England: MRC Applied Psychology Unit.

Pickles, J.O. (1988). An introduction to the physiology of hearing (2nd ed.). New York: Academic Press.

Potter, R.C., Fidell, S.A., Myles, M.M., and Keast, D.N. (1977). Effectiveness of audible warning devices on emergency vehicles (Tech. Report DOT-TSC-OST-77-38). U.S. Department of Transportation, Washington, D.C.: Office of Noise Abatement.

Randle, R.J., Larsen, W.E., and Williams, D.H. (1980). Some human factors issues in the development and evaluation of cockpit alerting and warning systems (NASA Reference Publication 1055). National Aeronautics and Space Administration, Moffett Field, California: Scientific and Technical Information Office.

Richards, A.M. (1976). Basic experimentation in psychoacoustics. Baltimore: University Park Press.

Riley, M.W., Cochran, D.J., and Ballard, J.L. (1983). An investigation of preferred shapes for warning labels. Human Factors, 24 (6), 737-742.

Scharf, B. (1959). Critical bands and the loudness of complex sounds near threshold. Journal of the Acoustical Society of America, 31, 365-370.

Scharf, B. (1961). Loudness summation under masking. Journal of the Acoustical Society of America, 33 (4), 503-511.

Scharf, B. (1970). Critical bands. In J.V. Tobias (Ed.), Foundations of modern auditory theory, Vol. 1. New York: Academic Press

Scharf, B., and Buus, S. (1986). Audition I: Stimulus, physiology, thresholds. In K.R. Boff, L. Kaufman, J.P. Thomas (Eds.), Handbook of perception and human performance, (pp.14-1 - 14-71). New York: John Wiley and Sons.

Scharf, B., and Meiselman, C.H. (1977). Critical bandwidth at high intensities. In E.V. Evans and J.P. Wilson (Eds.), Psychophysics and physiology of hearing. London: Academic Press.

Shingledecker, C.A. (1984). A task battery for applied human performance assessment research (Tech. Report AFAMRL-TR-84-071). Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Sternberg, S. (1969). Memory scanning: mental processes revealed by reaction time experiments. American Scientist, 57, 421-457.

Schlegel, R.E., and Gilliland, D. (1990, January). Evaluation of the criterion task set - Part I: CTS performance and SWAT data, baseline conditions (Tech. Report AAMRL-TR-90-007). Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory.

Schroeder, M.R. (1963). New method of measuring reverberation time. Journal of the Acoustical Society of America, 37 (3), 409-412

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

Stevens, S.S. (1957). On the psychophysical law. Psychological Review, 64 (3), 153-181.

Talamo, J.D. (1982). The perception of machinery indicator sounds. Ergonomics, 25, 41-51.

Tatsuoka, M.M. (1971). Multivariate analysis: Techniques for educational and psychological research. New York: John Wiley and Sons.

Teichner, W.H. (1954). Recent studies of simple reaction time. Psychological Bulletin, 51 (2), 128-149.

Teichner, W., and Krebs, M. (1972). Laws of the simple visual reaction time. Psychological Review, 79, 344-358.

Technical Committee ISO/TC 159 (Ergonomics). (1986). International Standard (ISO 7731): Danger signals for work places - Auditory danger signals. Switzerland: International Organization for Standardization.

Thurstone, L.L. (1927). A law of comparative judgement. Psychological Review, 34, 273-286.

Torgerson, W.S. (1958). Theory and methods of scaling. London: John Wiley and Sons.

Trucks, L.B., and Adams, S.K. (1975). Predicting the effectiveness of auditory warning signals in industrial environments. In Proceedings of the Human Factors Society 19th Annual Meeting (pp. 368-374). Santa Monica, CA: Human Factors Society.

U.S. Department of Defense. (1979). Military Standard: Noise limits for military materiel (MIL-STD-1474C). Washington, D.C.: Department of Defense.

U.S. Department of Defense. (1989). Military Standard: Human engineering design criteria for military systems, equipment and facilities (MIL-STD-1742D). Washington, D.C.: Department of Defense.

Vasey, M.W., and Thayer, J.F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. Psychophysiology, 24 (4), 479-486.

Ward, W.D. (1986). Anatomy and physiology of the ear: normal and damaged hearing. In E.H. Berger, W.D. Ward, J.C. Morrill and L.H. Royster (Eds.), Noise and hearing conservation manual (4th ed.), (pp. 177-196). Akron: American Industrial Hygiene Association.

Watson, C.S., Kelley, W.J., and Wroton, H.W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. The Journal of the Acoustical Society of America, 60 (5), 1176-1184.

Wickens, C.D. (1984). Engineering psychology and human performance. Columbus, Ohio: Charles E. Merrill Publishing Co.

Wierwille, W.W., and Casali, J.G. (1983). A validated rating scale for global mental workload measurement applications. In Proceedings of the Human Factors Society 27th Annual Meeting (pp. 129-132). Santa Monica, CA: Human Factors Society.

Wilkins, P.A. (1984). A field study to assess the effects of wearing hearing protectors on the perception of warning sounds in an industrial environment. Applied Acoustics, 17, 413-437.

Wilkins, P.A., and Acton, W.I. (1982). Noise and accidents - A review. Annals of Occupational Hygiene, 2, 249-260.

Wilkins, P.A., and Martin, A.M. (1980). The effects of hearing protection on the perception of warning sounds. In P.W. Alberti (Ed.), Personal hearing protection in industry (pp. 339-369). New York: Raven.

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

Williges, R.C., and Wierwille, W.W. (1979). Behavioral measures of aircrew mental workload. Human Factors, 21 (5), 549-574.

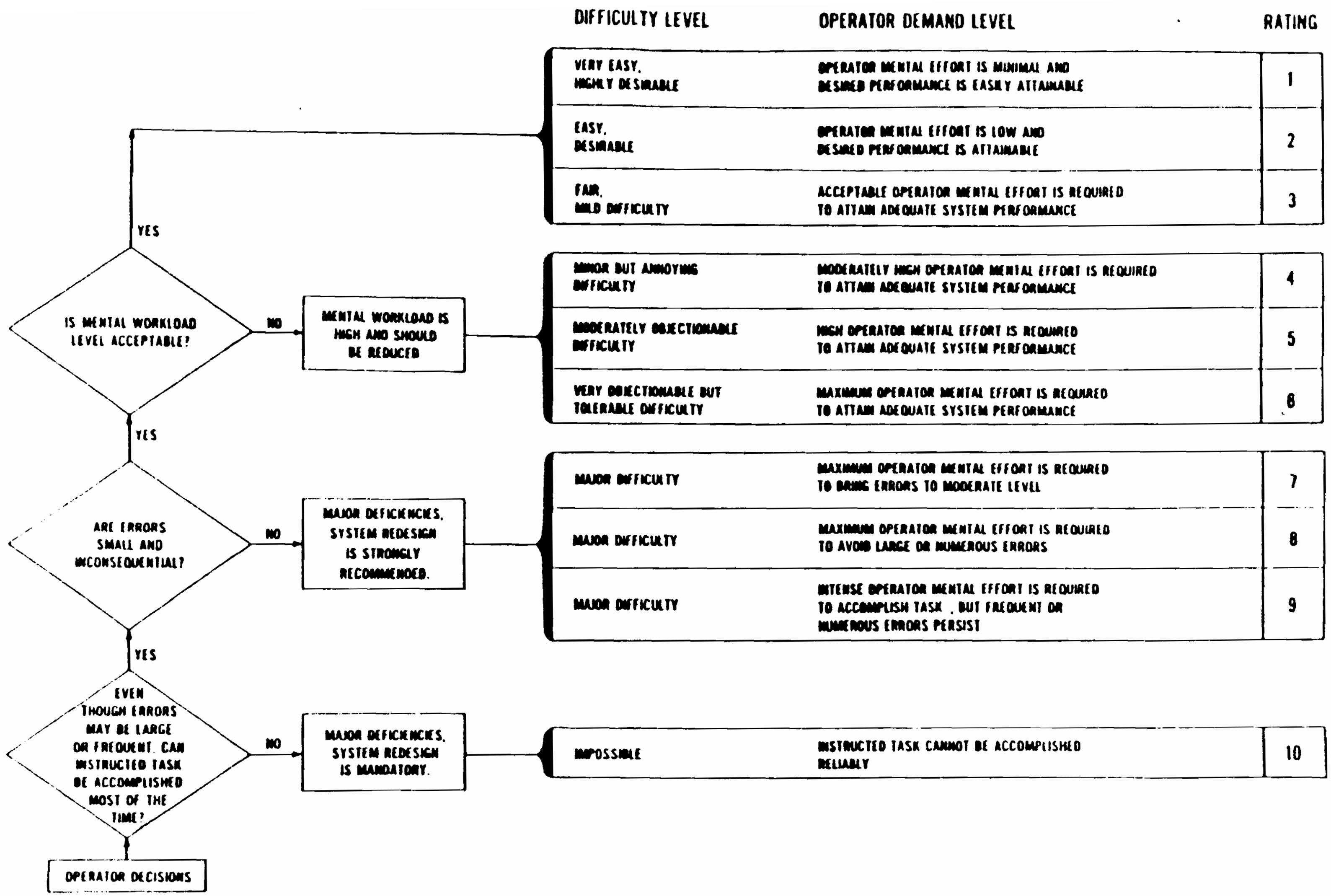
Winer, B.J. (1962). Statistical principles in experimental design. New York: Mcgraw-Hill.

Woodworth, R.S., and Schlosberg, H. (1954). Reaction time. In Experimental psychology (rev. ed., pp. 16-42). New York: Holt, Rinehart and Winston.

APPENDICES

APPENDIX A

MODIFIED COOPER-HARPER SCALE


DIFFICULTY LEVEL
OPERATOR DEMAND LEVEL
RATING
VERY EASY, HIGHLY DESIRABLE
OPERATOR MENTAL EFFORT IS MINIMAL AND DESIRED PERFORMANCE IS EASILY ATTAINABLE
1
EASY, DESIRABLE
OPERATOR MENTAL EFFORT IS LOW AND DESIRED PERFORMANCE IS ATTAINABLE
2
FAIR, MILD DIFFICULTY
ACCEPTABLE OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
3
MINOR BUT ANNOYING DIFFICULTY
MODERATELY HIGH OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
4
MODERATELY OBJECTIONABLE DIFFICULTY
HIGH OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
5
VERY OBJECTIONABLE BUT TOLERABLE DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
6
MAJOR DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO BRING ERRORS TO MODERATE LEVEL
7
MAJOR DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO AVOID LARGE OR NUMEROUS ERRORS
8
MAJOR DIFFICULTY
INTENSE OPERATOR MENTAL EFFORT IS REQUIRED TO ACCOMPLISH TASK, BUT FREQUENT OR NUMEROUS ERRORS PERSIST
9
IMPOSSIBLE
INSTRUCTED TASK CANNOT BE ACCOMPLISHED RELIABLY
10
YES
IS MENTAL WORKLOAD LEVEL ACCEPTABLE?
NO
MENTAL WORKLOAD IS HIGH AND SHOULD BE REDUCED
YES
ARE ERRORS SMALL AND INCONSEQUENTIAL?
NO
MAJOR DEFICIENCIES, SYSTEM REDESIGN IS STRONGLY RECOMMENDED.
YES
EVEN THOUGH ERRORS MAY BE LARGE OR FREQUENT, CAN INSTRUCTED TASK BE ACCOMPLISHED MOST OF THE TIME?
NO
MAJOR DEFICIENCIES, SYSTEM REDESIGN IS MANDATORY.
OPERATOR DECISIONS

APPENDIX B

INFORMED CONSENT DOCUMENT

APPENDIX B

SUBJECT'S INFORMED CONSENT: AUDITORY SYSTEMS LABORATORY - VA TECH

(THE PERCEIVED URGENCY AND DETECTABILITY OF MULTITONE AND FREQUENCY-MODULATED WARNING SIGNALS IN BROADBAND NOISE)

First, the hearing in your right and left ears will be tested with very quiet tones played through a set of headphones. Then, if qualified, you may also participate in a research experiment designed to investigate parameters which may affect the perceived urgency and detectability of auditory warning signals.

The research experiment will consist of three sessions. In the first session, you will listen to auditory warning signals played through a loudspeaker. You will be asked to tell me how urgent the signals seem by assigning numbers to them. In the second session, you will be asked to monitor displays on a computer screen. You will be asked to press a button on a keypad if the displays show any abnormal movement. During this session, you will also be asked to press a button on a different keypad whenever you hear an auditory warning signal. In the third session, you will listen to pairs of auditory warning signals. You will be asked to tell me which signal sounds more urgent to you.

After the research experiment, you will again have your hearing tested with very quiet tones played through a set of headphones.

No loud or harmful sounds will ever occur during the study. The test will be conducted in a sound-proof booth with the experimenter sitting outside. The door to the booth will be shut but <u>not</u> locked; either you may open it from the inside or the experimenter may open it from the outside. There is also an intercom system through which you may communicate with the experimenter by simply talking. (There are no buttons to push).

There is no risk to your well-being posed by either the hearing tests or the experiment. Also, realize that the hearing tests and the experiment are not designed to assess or diagnose any physiological or anatomical hearing disorders.

As a participant in this experiment, you have certain rights, as stated below. The purpose of this sheet is to describe these rights to you and to obtain your written consent to participate.

1) You have the right to discontinue participating in the study at any time for any reason by simply informing the experimenter.

2) You have the right to inspect your data and to withdraw it from the experiment if you feel that you should. The confidentiality of your data will be protected. All data that will be collected will be coded; your name will appear only on this consent form. All data are processed and analyzed after all subjects have completed the experiment. Subsequently, your data will be kept confidential by the research team. No one will see your individual data with your name.

3) You have the right to be informed as to the general results of the experiment. If you wish to receive a summary of the results, include your address (three months hence) with your signature on the last page of this form. If, after receiving the summary, you would then like further information, please contact the Auditory Systems Laboratory and a more detailed report will be made available to you. To avoid biasing other potential subjects, you are requested not to discuss the study with anyone until six months from now.

4) You may ask questions of the research team at any time prior to data collection. All questions will be answered to your satisfaction subject only to the constraint that an answer will not bias the outcome of the study. If bias would occur, with your permission an answer will be delayed until after data collection, at which time a full answer will be given.

Before you sign this form, please make sure that you understand, to your complete satisfaction, the nature of the study and your rights as a participant. If you have any questions, please ask them of the experimenter at this time. Then if you decide to participate, please sign your name below and provide your phone number so that you may be contacted for scheduling.

I have read a description of this study and understand the nature of the research and my rights as a participant. I hereby consent to participate, with the understanding that I may discontinue participation at any time if I choose to do so.

Signature_______________________________________

Printed Name____________________________________

Date____

Phone____

The research team for this experiment consists of a graduate student in ISE, Ms. Ellen Haas, and Dr. John G. Casali, Director of the Auditory Systems Laboratory. Dr. Casali may be reached at the following address and phone number:

Auditory Systems Laboratory
Room 538 Whittemore Hall
VPI&SU
Blacksburg, VA 24061
(703) 231-9086

Ms. Haas may be reached at the following address and phone number:

U.S. Army Human Engineering Laboratory
Behavioral Research Directorate
Bldg. 520
Aberdeen Proving Ground, MD 21005-5001
(301) 278-2946

In addition, if you have detailed questions regarding your rights as a participant in University research, you may contact the following Individual:

Chairman, University Human Subjects Committee
301 Burruss Hall
VPI&SU
Blacksburg, VA 24061
(703) 231-5283

(PLEASE TEAR OFF AND KEEP THIS PAGE FOR FUTURE REFERENCE)

APPENDIX C

INSTRUCTIONS AND EXPLANATIONS OF EXPERIMENTAL PROCEDURES

INSTRUCTIONS FOR THE AUDIOMETRY SESSIONS

This is a hearing check. You will be listening for some tones. Each time you hear a tone, push down on the response switch. When the tone goes away, release the response switch.

No matter how faint the tone, push down on the response switch when you hear the tone and release the response switch when the tone goes away.

Upon completion of your hearing check, please remain seated and quiet until the experimenter gives you further instructions.

INSTRUCTIONS FOR THE MAGNITUDE ESTIMATION DATA COLLECTION SESSION

"You will hear a series of sounds. Your task is to tell me how urgent they are by assigning numbers to them.

When you have heard the first sound, give its urgency a number - any number greater than or equal to zero, which you think is appropriate. Then, tell me that number by speaking it aloud. You will then hear the next sound. Do the same thing - give the urgency of the second sound a number, then tell me that number by speaking it aloud. You will do the same thing will all of the sounds that you hear.

Try to make the number that you have assigned to the sound, proportional to the urgency of that sound. For instance, if the urgency of the sound is twice as high as the one before it, give it a number twice as high. Remember, you may assign any number greater than or equal to zero, and there is no limit to the number you assign. There is no right or wrong answer. I want to know how you judge the urgency of the sounds. Any questions?

I will talk to you over this intercom. I'll be using it to describe to you what you will be hearing. However, when you listen to the sounds, and when you talk to me (such as when you tell me the number that you have assigned to the sound), please look at the blue dot on the door. I will have no trouble hearing what you say. I just want you to keep your head oriented in one position; that is why I want you to look at the blue dot on the door at all times.

Do you have any questions?

The sounds will vary by time between pulses. Here are some examples. The experimenter will present sample stimulus 1. Three seconds later, the experimenter will present sample stimulus 2.

"The sounds will vary by type. Here are some examples". The experimenter will present sample stimulus 3. Three seconds later, the experimenter will present sample stimulus 4. Three seconds later, the experimenter will present sample stimulus 5.

"Some sounds will be louder than others. I will turn on the pink noise when I present these. Here are some examples". The experimenter will present sample stimulus 6. Three seconds later, the experimenter will present sample stimulus 7.

"Are there any questions? I will now present three more signals. Just say "yes" when you hear each signal". The experimenter presents stimuli 8, 9 and 10, with three-second pauses between each signal.

"You did very well. I will now turn the pink noise on, and will present the test sounds. Remember to keep looking at the blue dot on the door".

INSTRUCTIONS FOR THE REACTION TIME DATA COLLECTION SESSION

Workload Task

"In this session, you will be monitoring a screen with three displays which are intended to have the appearance of dials like those on a machine". The experimenter turns on the display. "The dials have six pointer positions and a pointer which appears below the positions and moves from one to another. Under normal conditions, the pattern of pointer movement is random. The pointer is equally likely to move to any position. Periodically, the pointer will tend to stay on one side of the dial more than the other. This will be called a "biased" movement. Your task is to watch the dials carefully for biased patterns of pointer movement. If you think you see a bias, press the button on the keypad that corresponds to the dial. Press the button with your (non-dominant) hand. When you correctly respond to a bias, it is eliminated, and the pointer goes back to moving randomly again. A given dial may show a biased movement more than once during the test period, but two dials will never be biased at the same time. A bias condition will last longer than a few seconds. Try to avoid responding unless you are confident that a dial is biased. Responses to nonexistent biases are scored against you. The screen will automatically go blank at the end of the monitoring period."

"To make this task similar to what is experienced in real-life conditions, I am going to play a continuous noise while you are performing this task. This type of noise is known as "pink noise", and the level at which it is presented will pose no harm to you".

"Rest your hands like this during the trial. Remember to press the button on the four-button keypad with your (non-dominant) hand. We will now have four practice trials. The screen will go blank between trials. When you see the picture on the screen come on with three boxes, just press any key of your 4-button keypad to begin the next trial. I will remind you to do this over the intercom. Are there any questions"?

The experimenter turns on the pink noise and presents four practice trials to the subject. When the subject has completed the four practice trials, the experimenter will stop the pink noise, and will present the instructions for the Reaction Time Task:

Reaction Time Task

"While you are monitoring the screen for bias conditions, you will hear some auditory signals which are like the sounds that you heard previously. Whenever you hear an auditory signal, please press the button on the single-button keypad with your`(dominant) hand as quickly as you can. Are there any questions? Let's try five practice trials, in which you will press the button when you hear the sound. I will turn on the pink noise while you listen for the sound".

The experimenter turns on the pink noise, and presents 10 practice trials to the subject. When the subject has completed the 10 practice trials, the experimenter will stop the pink noise, and will say:

"Let's do one more practice in which we'll perform both tasks together. Be sure to hold your hands like this. While you are monitoring the screen for bias conditions, you will hear the auditory signals. Remember to press a button on the four-button keypad with your (non-dominant) hand when you see a bias on the screen. When you hear an auditory signal, be sure to press the button on the one-button keypad with your (dominant) hand. This is important: the auditory signal is the most important signal. Whenever you hear the sound, respond to it as quickly as you can. If you hear a signal and see a bias condition on the screen at the same time, respond to the auditory signal first. Responding to the sound as quickly as you can is the most important task. Remember to position your hands like this". The experimenter demonstrates proper hand position on each keypad.

"Are there any questions? Let's begin this last practice task". The experimenter turns on the pink noise, and presents one practice task.

"Very good. Now we'll do three trials, in which you will again perform both tasks together. Be sure to hold your hands like this. As before, while you are monitoring the screen for bias conditions, you will hear the auditory signals. Remember to press a button on the four-button keypad with your (non-dominant) hand when you see a bias on the screen. When you hear an auditory signal, be sure to press the button on the one-button keypad with your (dominant) hand. Remember that the auditory signal is the most important signal. Whenever you hear the sound, respond to it as quickly as you can. If you hear a signal and see a bias condition on the screen at the same time, respond to the auditory signal first. Responding to the sound as quickly as you can is the most important task".

"Remember to position your hands like this". The experimenter demonstrates proper position on the keypad. "Are there any questions"?

The experimenter presents the stimuli.

INSTRUCTIONS FOR THE PAIRED COMPARISON DATA COLLECTION SESSION

"You will hear a two signals. Your task is to tell me which of the two sounds more urgent to you. Indicate which signal you think is the most urgent by speaking it aloud; if the first signal seems to be more urgent to you, say "one". If the second sound seems to be more urgent to you, say, "two". In the event that both signals sound equally urgent to you, you still have to choose one or the other. Say either "one" or "two".

"You will be presented with many pairs of signals. For each pair, do the same thing; after you hear the second signal of the pair, tell me which of the two sounds more urgent to you, by saying either "one" or "two". There is no right or wrong answer. I want to know which signal you think is more urgent.

"Remember, when you listen to the sounds, and when you talk to me, (such as when you tell me the number that you have assigned to the sound), please look at the blue dot on the door. I will have no trouble hearing what you say. I just want you to keep your head oriented in one position;

that is why I want you to look at the blue dot on the door at all times".

"Any questions? You will now perform two practice trials. After you have heard these, I will go ahead and present the signals for you to compare. I will turn on the pink noise when you hear these signals".

The experimenter turns on the pink noise, and presents the practice and the paired comparison signals.

APPENDIX D

SUBJECT CASE HISTORY FORM

CASE HISTORY FORM

NAME______________________________ DATE__________________

I. MEDICAL HISTORY

A. Pain____ Ear____
B. Discharge____ Ear____
C. Fullness or pressure in ears ____
D. Dizziness ____
E. Tinnitus____ Ear____
F. Medications (including aspiring and other nonperscription drugs)______________________________
____
____

G. Ear, nose and throat surgery
____
____

H. Noise exposure (prior and current)
____
____

I. Family history of hearing loss (limit to members of immediate family who sustained hearing loss prior to age 50)

J. History of ear infections____ frequency ____
severity____ treatment____

K. Head injuries, e.g. concussions or severe blows
____
____

L. High fevers____

M. Do you have or have you had:

High blood pressure	____
Diabetes	____
Tuberculosis	____
Meningitis	____
Venereal disease	____
Scarlet fever	____
Mumps	____
Measles	____
Seizures	____
Multiple sclerosis	____

N. General current medical condition:
Excellent___ Good___ Fair___ Poor___

II. AUDIOLOGIC HISTORY

A. Previous hearing tests?____ If yes, date__________

B. Do you feel you have a hearing loss?____

C. Discrimination ability:

1. Telephone ____
2. Conversation ____
3. In noise ____
4. In quiet ____
5. Radio, movies, TV ____
6. Localization ____
7. Female voice ____
8. Male voice ____
9. Conferences ____

III. SPEECH/LANGUAGE HISTORY

A. Is English your first language?____
If not, what is?_____________________

APPENDIX E

ANOVA STATISTICS

MAGNITUDE ESTIMATION AND DETECTION TIME

TABLE 1

Magnitude Estimation Data ANOVA Summary Table

Source of Variance	df	MS	*F**	*p*	H-F *p***
<u>Between-Subjects</u>					
Subjects	35	0.00			
<u>Within-Subjects</u>					
Pulse format (F)	2	282.609	10.06	<0.001	<0.001
S X F	70	28.099			
Time Between Pulses (T)	2	115.594	44.57	<0.001	<0.001
S X T	70	2.593			
Pulse Level (L)	1	3605.639	114.89	<0.001	***
S X L	35	31.383			
F X T	4	4.209	2.39	0.053	0.055
S X F X T	140	1.759			
F X L	2	76.861	24.92	<0.001	<0.001
S X F X L	70	3.085			
T X L	2	15.172	11.10	<0.001	<0.001
S X T X L	70	1.367			
F X T X L	4	0.503	0.35	0.841	
S X F X T X L	140	1.425			
Total	647				

* Denominators used for each source of variance in the *F* tests appear as the second term in each grouping in the table.

**All H-F *p* values appear for those effects in which the original uncorrected *p* value was significant because the correction is not useful for nonsignificant effects. For pulse format, the H-F epsilon was 0.968. For time between pulses, the H-F epsilon was 0.694. For the format x time between pulses interaction, the H-F epsilon was 0.834. For the format by level interaction, the H-F epsilon was 0.945. For the time between pulses x pulse level interaction, the H-F epsilon was 0.906.

*** H-F corrections not required for effects in which treatment df = 1 because the assumption of sphericity is always fulfilled.

TABLE 2

Detection Time Data ANOVA Summary Table

Source of Variance	df	MS	*F**	*p*	H-F *p***
<u>Between-Subjects</u>					
Subjects	35	0.11			
<u>Within-Subjects</u>					
Pulse format (F)	2	0.09	6.88	0.002	<0.01
S X F	70	0.01			
Time Between Pulses (T)	2	0.03	1.98	0.146	
S X T	70	0.02			
Pulse Level (L)	1	0.59	34.24	<0.001	***
S X L	35	0.02			
F X T	4	0.01	1.49	0.207	
S X F X T	140	0.01			
F X L	2	0.01	0.38	0.688	
S X F X L	70	0.01			
T X L	2	0.01	0.57	0.571	
S X T X L	70	0.01			
F X T X L	4	0.01	0.54	0.707	
S X F X T X L	140	0.01			
Total	647				

* Denominators used for each source of variance in the *F* tests appear as the second term in each grouping in the table.

**All H-F *p* values appear for those effects in which the original uncorrected *p* value was significant because the correction is not useful for nonsignificant effects. For pulse format, the H-F epsilon was 0.780.

*** H-F corrections are not required for effects in which treatment df = 1 because the assumption of sphericity is always fulfilled.

VITA

Ellen Haas

PERSONAL DATA:

Date of Birth: April 10, 1954
Place of Birth: Denver, Colorado
Married: Gary Haas

EDUCATION:

Dissertation in progress, Virginia Polytechnic Institute and State University (VPI&SU), Blacksburg, Virginia. Major research area (dissertation): Perceived urgency and detection time of auditory warning signals

Master of Science in Industrial and Systems Engineering (Human Factors Psychology), California State University at Long Beach, 1982. Major research area (thesis): Human factors design of a hemodialysis dialysate pressure monitor.

Bachelor of Arts in Psychology, 1974, University of Colorado at Boulder, Colorado.

PROFESSIONAL AFFILIATIONS:

Human Factors Society
Acoustical Society of America

EMPLOYMENT EXPERIENCE:

1984-Present: Human Factors Engineer, Army Research Laboratory.

1980-1983: Human Factors Research Assistant, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

1978-1980: Human Factors Engineering Consultant, NATEL Corporation, Orange, CA.

RESEARCH INTERESTS:

Human factors engineering, psychoacoustics, auditory displays, auditory icons, workload assessment.