USING SIGNAL DETECTION THEORY TO MODEL THE DETECTION OF WARNING SIGNALS IN NORMAL AND HEARING-IMPAIRED LISTENERS WHILE WEARING HEARING PROTECTION.

by

Gary S. Robinson

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Industrial and Systems Engineering

APPROVED:

John G. Casali

J. G. Casali, Chairman

R. C. Williges

J. C. Woldstad

A. M. Prestrude

C. P. Koelling

October 29, 1993

Blacksburg, Virginia

USING SIGNAL DETECTION THEORY TO MODEL THE DETECTION OF WARNING SIGNALS IN NORMAL AND HEARING-IMPAIRED LISTENERS WHILE WEARING HEARING PROTECTION

by

Gary S. Robinson

Chairman: Dr. John G. Casali

Industrial and Systems Engineering

(ABSTRACT)

The question of whether or not an individual suffering from a hearing loss is capable of hearing an auditory alarm or warning is an extremely important industrial safety issue. International standard ISO 7731–1986(E), Danger Signals for Work Places – Auditory Danger Signals, requires that any auditory alarm or warning be audible to all individuals in the workplace, including those suffering from a hearing loss and/or wearing hearing protection devices (HPDs). Very little research has been conducted to determine how an individual's hearing level affects his/her ability to detect an auditory alarm or warning in a high-noise environment while wearing an HPD.

The research effort described herein was undertaken to determine how the ability to detect an alarm or warning signal changed for individuals with normal hearing and two levels of hearing loss as the levels of masking noise and alarm were manipulated. Pink noise was used as the masker since it is a generally-accepted, generic substitute for industrial noise. A heavy-equipment reverse alarm was used as the signal since it is a common alarm in industrial facilities and construction sites. The rating method paradigm of signal detection theory was used as the experimental procedure in order to separate the subjects' absolute sensitivities to the alarm from their individual criteria for deciding to respond in an affirmative manner.

Results indicated that even at a fairly low signal-to-noise ratio (0 dB), individuals with a substantial hearing loss [a pure-tone average (PTA) hearing level on the order of 45-50 dBHL in both ears] are capable of hearing the alarm while wearing a high-attenuation earmuff. Predictive models were developed using nonlinear regression techniques. These models may be used to predict whether or not individuals with known hearing levels will be capable of hearing the alarm under known conditions or to determine the level of alarm presentation in order to be heard reliably by individuals with a specified range of hearing for given noise levels

ACKNOWLEDGMENTS

The author would like to thank those individuals without whom this research effort would not have been possible. First, special thanks and gratitude are extended to Dr. John G. Casali, committee chairman, for his constant encouragement and advice, his belief in my ability, and his patience throughout the research effort. Thanks are also due Dr. Robert C. Williges and Dr. Jeffrey C. Woldstad for their helpful suggestions concerning the statistical analyses and to Dr. C. Patrick Koelling and Dr. Albert M. Prestrude for their invaluable advice and encouragement at each phase of the project.

This research effort was sponsored by the Corporate Health Department of the Aluminum Company of America (ALCOA), Pittsburgh, Pennsylvania. Christine Dixon-Ernst served as ALCOA technical monitor. Thanks are due Christine Dixon-Ernst and H. Dean Belk, M.D. for suggesting the research problem and for providing funding for its study. In addition to these two individuals, the author would like to thank Nancy Sussman, Jay Harper, M.D., and Stephen I. Roth of ALCOA for their beneficial comments, criticisms, and suggestions on the research effort. Gratitude is also extended to Joe Hazelwood at ALCOA's Southplant Truck Shop for his assistance in obtaining the alarm used in the experiment.

The author is also indebted to Randy Waldron of the Industrial and Systems Engineering Department for fabrication of the special test fixtures used in the experiment, to Dan Mauney for the use of his computer program for determining the physical attenuation of earmuffs, and to Joe Jenkins for his helpful advice in writing the C-language program used to control the experiment.

Finally, special thanks are extended to Rose, without whose love, support, understanding, and encouragement this achievement would not have been possible.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS
(Continued)

Page

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF FIGURES

LIST OF FIGURES
(continued)

LIST OF FIGURES
(continued)

LIST OF TABLES

LIST OF TABLES
(continued)

INTRODUCTION

Individuals who work in noisy environments often complain that wearing hearing protection devices (HPDs) interfere with their ability to hear auditory warning and indicator sounds (Kerivan, 1979; Suter, 1989; Wilkins and Acton, 1982; Wilkins and Martin, 1978). However, studies have shown that at high noise levels (greater than 80 to 85 dBA), individuals with normal hearing will be able detect auditory signals just as well as, if not better than, they would if they were not wearing HPDs (Abel, Kunov, Pichora-Fuller, and Alberti, 1983a; Forshaw, 1977; Wilkins and Martin, 1982). The reason for such an effect is that although the HPD attenuates both the signal and the noise equally, it reduces their levels to the point where the cochlear distortion present at high noise levels is reduced or eliminated, thus enabling the ear to better distinguish the signal from the noise (Suter, 1989). This effect, however, is likely HPD-specific in that the attenuation of the device must be sufficient to adequately reduce the distortion present in the cochlea. Wilkins and Acton (1982) suggest that one reason for the continued belief that HPDs interfere with auditory perception is the mistaken assumption on the part of the user that since HPDs reduce the audibility of signals when worn in quiet, they must do the same in noise.

Although HPDs may improve signal audibility in noise for persons with normal hearing, the same cannot be said for persons suffering from a hearing loss. The three studies which have investigated the effects of wearing HPDs on warning signal detection by hearing-impaired persons have found that HPDs reduce the audibility of such signals (Abel et al., 1983a; Coleman, Graves, Collier, Golding, Nicholl, Simpson, Sweetland, and Talbot, 1984; Forshaw, 1977). The explanation for such an effect is that the HPD attenuates the signal (and also the noise) to the point that the sound reaching the ear is below the auditory threshold (Abel et al., 1983a; Lazarus, 1980).

Both Lazarus (1980) and Coleman et al. (1984) suggest procedures for predicting an individual's masked threshold for warning signals while wearing an HPD. The method proposed by Lazarus (1980) requires knowledge of the third-octave spectrum of the noise, the spectral attenuation characteristics of the particular HPD used, the masked threshold of the signal if it were being listened for without wearing an HPD, and the pure-tone hearing threshold of the individual being considered. No empirical evidence of the accuracy of this method is reported. The method proposed by Coleman et al. (1984) uses critical band theory as modified by Patterson (1974, 1976) and Patterson, Nimmo-Smith, Weber, and Milroy (1982) and requires even more information than the method proposed by Lazarus (1980). Not only is it necessary to know the third-octave spectral characteristics of the background noise, the attenuation characteristics of the HPD, and the individual's hearing threshold, but it is also required to have an estimate of the shape of the individual's auditory filter at the frequency of interest [the frequency of the tone being detected, or for a complex signal, the frequency (or frequencies) at which most of the signal energy is centered]. Experimental results indicate that the procedure consistently overestimates the mean masked threshold by approximately 5 dB.

Although either of the methods mentioned above (both of which will be discussed in more detail in a later section) may provide suitably accurate predictions of the masked thresholds of individuals (with or without a hearing loss) wearing HPDs in noise, the amount and type of information necessary to implement either procedure is likely not available to the industrial hygienist or safety professional responsible for overseeing the health and safety of an employer's workforce. In addition, it may be prohibitively expensive to generate the data necessary to apply a procedure such as that described by Coleman et al. (1984). The need, therefore, is for a new predictive procedure which relies on information of the type which is available to the individuals who will be implementing the procedure while minimizing the need for additional data.

An experiment to develop such a model is described herein. Independent variables used included the broadband A-weighted sound pressure level (SPL) of the masking noise, the broadband A-weighted SPL of the warning signal, and the pure-tone average hearing level of the experimental subjects. Each of these measures are (or should be) readily available to industrial hearing conservationists, or are easily and economically obtainable.

HEARING AND HEARING LOSS

Anatomy and Physiology of the Ear

The ear, Figure 1, is divided into three major anatomical subdivisions: the outer ear, the middle ear, and the inner ear. Each of these subdivisions serves a different function in the process of converting the acoustical energy in the air into neural impulses for transmission to the brain.

Outer ear. The outer ear is composed of the pinna and the auditory canal. These structures serve the dual purposes of modification of the incoming sound energy and protection of the delicate structures of the middle and inner ear (Goldstein, 1989; Ward, 1986a). The pinna is the cartilaginous structure located on the side of the head. It collects and modifies the incoming sound waves, funneling them into the ear canal. Due to its shape and tissue characteristics, some frequencies of the incoming sounds are amplified, while others are attenuated (Ward, 1986a). The auditory canal is a tubular duct about 3 cm long leading to the tympanic membrane (eardrum). With a resonant frequency of about 3,400 Hz (varying slightly among individuals), the auditory canal also modifies incoming sound waves, causing the frequencies between approximately 2,000 and 5,000 Hz to be amplified by as much as 10 to 15 dB, thus making the ear more sensitive to sounds in this frequency range (Goldstein, 1989; Ward, 1986a). It is partially for this reason that noises in this frequency range are the most hazardous to hearing (Ward, 1986a), and that permanent hearing loss is often first discovered as an elevated threshold at 4000 Hz (Ward, 1986b). The length of the auditory canal helps maintain the tympanic membrane and the middle ear at a constant temperature, and the ear wax (cerumen) collects fine particulate matter before it reaches the eardrum (Goldstein, 1989).

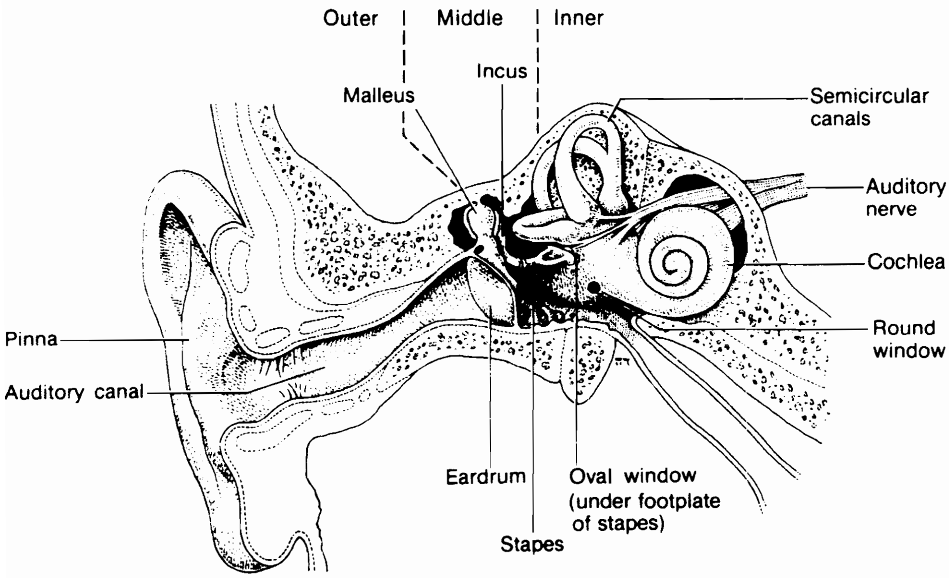


Figure 1. *Anatomy of the ear*. (from Goldstein, 1989)

Middle ear. The middle ear is a cavity with a volume of approximately 2 cc which contains three bones (ossicles) and is bounded on one side by the tympanic membrane and on the other side by the oval window of the cochlea. The tympanic membrane separates the outer ear from the middle ear and is the first step in the process of converting the pressure variations in the air to mechanical vibrations which are later converted to nerve impulses. Pressure waves traveling down the auditory canal strike the eardrum and cause it to vibrate at the same frequency as the incident pressure wave. As the tympanic membrane vibrates, the vibrations are passed on to the three bones inside the middle ear: the malleus (hammer), the incus (anvil), and the stapes (stirrup); and are then transmitted to the oval window. The malleus is attached to the tympanic membrane, the stapes is attached to the oval window, and the incus is situated between the two and acts as a lever to amplify the vibrations reaching it. Due to the shape and arrangement of the three bones and the size difference between the tympanic membrane and the oval window (the eardrum is 17 times larger than the oval window), the vibrations of the tympanic membrane are amplified by a factor of between 22 and 100 before they reach the oval window. This amplification is necessary due to the differences in the acoustic impedance of the air in the outer ear and the liquid (perilymph) in the inner ear (Goldstein, 1989; Ward, 1986a). If this amplification did not take place, only about 1/1000th of the acoustical energy in the original air pressure wave would be transmitted to the inner ear (Ward, 1986a).

Inner ear. The inner ear is composed of the cochlea and the auditory nerve. When the stapes vibrates the oval window, the vibrations are transmitted to the liquid (perilymph) inside the cochlea, which, in turn, causes the basilar membrane to vibrate. Situated on top of the basilar membrane is the organ of Corti, Figure 2. The two primary components of the organ of Corti are the hair cells and the tectorial membrane. As the

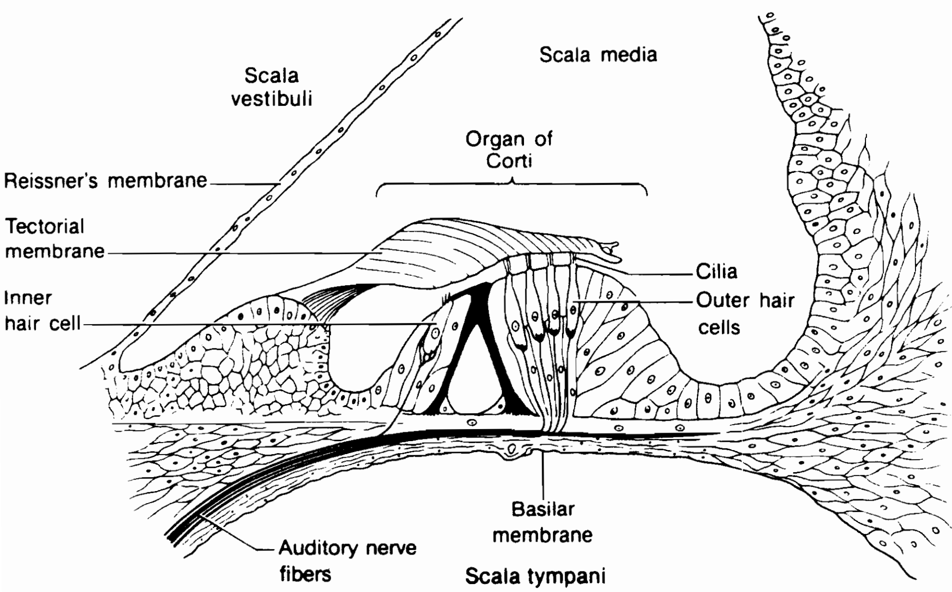


Figure 2. *Cross-section of the organ of Corti*. (from Goldstein, 1989)

basilar membrane vibrates, the cilia are deformed by the relative motion of the basilar membrane and the tectorial membrane. Deformation of the cilia causes the hair cells to fire a bio-electrical impulse to the auditory nerve fibers, and these signals are then transmitted to the brain via the auditory nerve. At high sound pressure levels, the ability to discriminate signals from noise is lost, due in part to nonlinear distortion within the cochlea [i.e., the generation of nerve impulses corresponding to harmonic overtones of the stimulating signals (Békésy, 1960)]. Detailed discussions of the mechanism for conversion of vibratory energy to nerve impulses and construction of the cochlea and its internal structure appear in Goldstein (1989) and Ward (1986a).

Auditory Pathways

Normally, the sensation of hearing is caused by airborne pressure waves which travel down the auditory canal and impinge upon the tympanic membrane causing it to vibrate. These vibrations are then conducted via the structures of the middle ear to the inner ear where they are converted to nerve impulses and transmitted to the brain. This is the air conduction pathway. However, another sound path is also available: that of bone conduction. In bone conduction, vibrations are conducted through the bones and tissues directly to the outer, middle and/or inner ear, causing the structures in these locations to vibrate. These vibrations are converted to nerve impulses just as the vibrations which arrive via air conduction. The stimuli for bone conduction may be internal or external to the body (Gales, 1979).

Auditory Dysfunctions

Conductive hearing loss. Conductive hearing loss is associated with physical damage to one or more of the structures of the middle or inner ear, or blockage of the sound conduction pathways of the outer ear. The most common causes of conductive hearing loss are otitis media, an inflammation or infection of the middle ear, and otosclerosis, a softening of the bones of the middle ear (Newby, 1979). However, a ruptured or severely scarred eardrum, excessive wax buildup in the auditory canal, or damaged or dislocated ossicles in the middle ear can also result in conductive hearing loss. Hearing loss of this type is usually associated with disease or a single traumatic incident such as a blow to the head and is often reversible with proper treatment and/or surgery. Hearing aids may also offer relief for the victims of permanent conductive hearing loss. Occupationally-related conductive hearing loss, while it does occur, is not common (Ward, 1986a).

Conductive hearing loss is characterized by a fairly flat audiogram in which the thresholds at all of the test frequencies are elevated by approximately equal amounts (Morrill, 1986). A typical audiogram characteristic of a conductive hearing loss is illustrated by the dashed line in Figure 3.

Neural hearing loss. Neural hearing loss is associated with permanent damage to the delicate structures of the inner ear (i.e., the basilar membrane, the cilia, etc.), to the auditory nerve, or to the higher neural auditory pathways up to and including the brain. Neural hearing loss is irreversible.

Noise-induced hearing loss may occur in either of two ways. The first of these is "acoustic trauma." Acoustic trauma refers to a single exposure to a high-intensity acoustic event which causes permanent physical damage to the delicate sensory structures of the inner ear, such as complete separation of the tectorial membrane from the cilia, or tearing of the inner structure of the cochlea. The second and more common means of inducing neural damage is prolonged exposure to moderate or loud noise. Exposures of this type cause the structures of the inner ear to fatigue. Anatomical characteristics of such fatigue might include swelling or twisting of the cilia or a reduction in the enzyme level in the cochlear fluid (Ward, 1986a). Perceptually, these changes might be characterized by tinnitus (a ringing in the ears) or a raised auditory threshold. If not

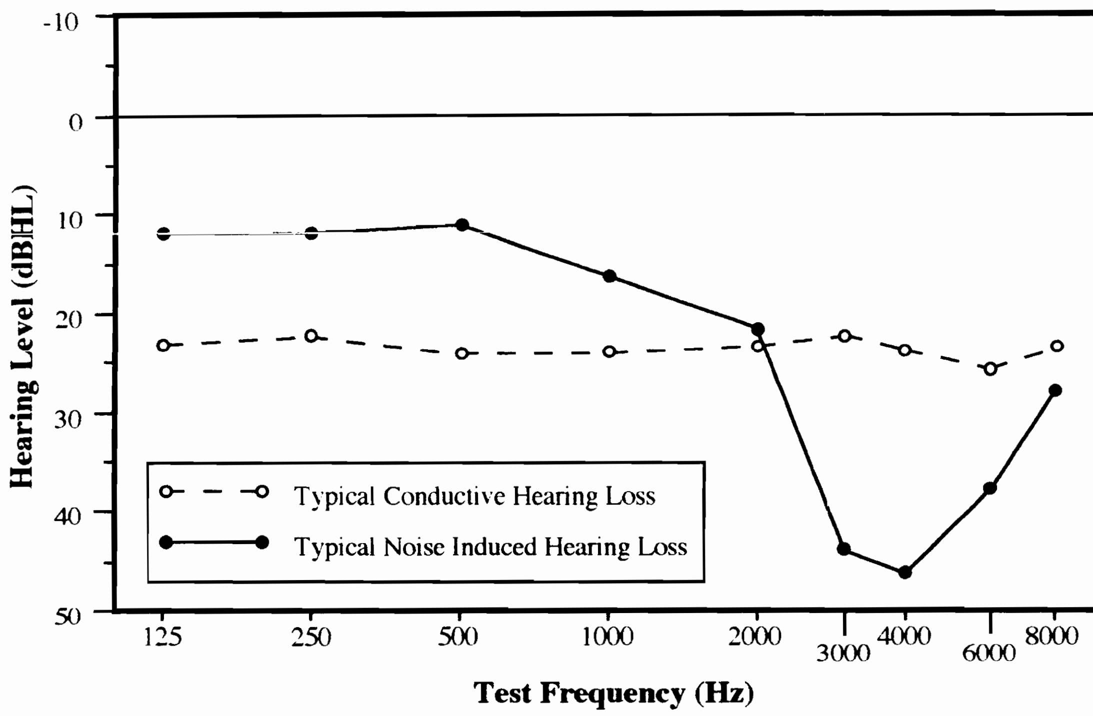


Figure 3. *Typical audiogram characteristics of conductive and noise-induced hearing loss.* (adapted in part from Melnick, 1979)

given sufficient time to recover, the structures of the inner ear continue to degenerate until permanent damage occurs. Neural hearing loss is the most common form of occupationally-related hearing loss (NIH, 1990).

Noise-induced hearing loss is often first noticed as an elevated threshold at 4000 Hz. As the condition worsens with continued exposure, this "notch" deepens and spreads to include the adjacent frequencies, eventually including the frequencies from 1000 Hz to 8000 Hz (Melnick, 1979). However, the characteristic notch shape is retained. A typical audiogram characteristic of noise-induced hearing loss is illustrated by the solid line in Figure 3.

Classifying hearing loss. In addition to classifying hearing loss as either conductive or neural, it is also necessary to specify the degree of loss. Hearing loss may be specified categorically by relating the amount of hearing loss to the degree of difficulty in understanding speech, or as a percentage loss in one or both ears (Kryter, 1985; Miller and Wilber, 1991). One categorical classification scheme is illustrated in Table 1 which bases the categories on the pure-tone average (PTA) hearing level at the frequencies of 500, 1000, and 2000 Hz. However, Kryter (1985) argues that the use of these three frequencies underestimates the difficulty encountered in understanding speech and suggests calculating the PTA hearing level using the frequencies of 1000, 2000, and 3000 Hz or possibly even 1000, 2000, 3000, and 4000 Hz.

The method of rating hearing impairment endorsed by the American Academy of Otolaryngology – Head and Neck Surgery (AAO–HNS) is to calculate a percent hearing impairment for one or both ears (Miller and Wilber, 1991). The *monaural* percent hearing impairment is calculated by subtracting 25 dB from the PTA hearing level at 500, 1000, 2000, and 3000 Hz and then multiplying the result by 1.5 percent. The *binaural* percent hearing impairment is calculated by multiplying the monaural percent hearing impairment for the better ear by 5, adding the result to the monaural percent hearing

TABLE 1

A Classification Scheme for Hearing Impairment (from Miller and Wilber, 1991)

Hearing (threshold) level, dB*	Class	Degree of handicap	Average hearing (threshold) level for 500, 1000, and 2000 Hz in the better ear: more than	Average hearing (threshold) level for 500, 1000, and 2000 Hz in the better ear: Not more than	Ability to understand speech
	A	Not Significant		25 dB	No significant difficulty with faint speech
25	B	Slight Handicap	25 dB	40 dB	Difficulty only with faint speech
40	C	Mild Handicap	40 dB	55 dB	Frequent difficulty with normal speech
55	D	Marked Handicap	55 dB	70 dB	Frequent difficulty with loud speech
70	E	Severe Handicap	70 dB	90 dB	Can understand only shouted or amplified speech
90	F	Extreme Handicap	90 dB		Usually cannot understand even amplified speech

* Hearing level (HL) is a weighted dB level as per ANSI S3.6-1989, Specification for Audiometers (ANSI, 1989).

impairment for the poorer ear, and dividing the total by 6. The National Academy of Sciences, National Research Council, Armed Forces Committee on Hearing, Bioacoustics and Biomechanics (CHABA) recommends an identical procedure to that endorsed by the AAO–HNS, except that the PTA hearing level is calculated for the three frequencies of 1000, 2000, and 3000 Hz . In addition, various states require the use of different frequencies in settling hearing loss claims (Miller and Harris, 1979).

Although not strictly a measure of hearing loss, OSHA (1989) uses a "Standard Threshold Shift" (STS) as a measure of the change in hearing due to excessive noise exposure. An STS is defined as a change, relative to a baseline audiogram, in the PTA hearing level (using frequencies of 2000, 3000, and 4000 Hz) of 10 dB or more in either ear. Melnick (1984) criticizes OSHA's reliance on thresholds measured only at high frequencies because hearing loss due to other etiologies will not be detected. In addition, he makes the point (applicable to all of the rating methods described above) that using an average hearing level at multiple frequencies could easily fail to detect a significant threshold shift at one frequency if other frequencies did not change or changed little.

HEARING PROTECTION

The Need: Prevalence of Industrial Noise

Occupational noise exposure is the most common cause of noise-induced hearing loss (NIH, 1990). In 1981, it was estimated that as many as 9 million workers were exposed to occupational noise levels exceeding a time-weighted average (TWA) of 85 dBA for an 8-hour day (EPA, 1981). Exposures of this magnitude are sufficient to cause permanent hearing loss if repeated over a period of years (NIH, 1990). It is estimated that in just the 50-59 age group, as many as 1.7 million workers already suffer from occupationally-related noise-induced hearing loss (Robinette, 1984). As alarming as these statistics are, they do not take into account those individuals who are noise-exposed in non-occupational pursuits. It is estimated that as many as 20 million Americans are regularly exposed to noise levels sufficient to cause permanent hearing loss (NIH, 1990). Clearly, proactive countermeasures must be taken to preserve the hearing of noise-exposed individuals. Because of their convenience and relatively low cost, HPDs are the most widely used solution to the problem of employee noise exposure in industry today.

Hearing Protection Devices

Although there are many different types of HPDs available, the most commonly used devices in industry are earplugs (including premolded, user-molded, and custom-molded designs), earmuffs (circumaural devices that completely surround the ear), and semi-aurals (devices that seal the opening of the ear canal, but do not cover the ear). Each of these categories will be discussed briefly.

Earplugs. Earplugs are of three general types: premolded, user-molded, or custom-molded. Regardless of type, earplugs are meant to be inserted to various depths in the ear canal to block the air conduction pathway. In general, earplugs are least affected by the wearing of eyeglasses or other items of safety equipment. However, proper sizing and fitting of the earplug to the user and proper insertion technique are critical to obtaining a proper seal with any earplug (Casali and Epps, 1986; Nixon, 1979).

Premolded earplugs are generally formed from soft rubber, vinyl, or silicone compounds. They usually have one or more flanges around their circumference to aid in sealing the ear canal, and some premolded plugs are available in multiple sizes to fit a wide range of ear canals. As mentioned earlier, proper fitting is essential to obtaining a correct seal with a premolded earplug. Some individuals may require a different size device for each ear, while others may not even be able to obtain an adequate seal due to the shape and/or size of their ear canals. Although the ear canal is generally elliptical in cross-section, some individuals have ear canal openings resembling elongated slits. This makes fitting premolded earplugs difficult, if not impossible. Premolded earplugs are also susceptible to modification by the wearer to make the device more comfortable. Modifications include trimming the flanges or puncturing internal air pockets, thus causing the device to collapse when inserted (Gasaway, 1984). Finally, premolded earplugs tend to loosen over time and with increased physical activity on the part of the wearer and must be reinserted periodically to maintain a tight seal (Casali and Park, 1990).

User-molded or user-formed earplugs are usually made from materials such as spun fiberglass (ear down), waxed cotton, or vinyl or polyurethane foams. Although these devices generally have a shorter useful lifetime than the premolded earplugs, they are usually cheaper, and are, with some exceptions, intended as a one-size-fits-all type of protector. These devices form a seal by assuming the shape of the ear canal. Therefore, variations in the size and shape of an individual's ear canal are less of a problem with user-molded plugs than with premolded plugs, although size extremes or slit-shaped canals may contraindicate the use of user-molded earplugs.

Custom-molded earplugs are, as the name implies, custom-made to fit a single individual. The general procedure for the manufacture of a custom-molded plug is to first make impressions of the ear canals of the person who will be using the device. These impressions are then used as a model for the manufacture of the earplugs. Although these devices do not have the same problems associated with sizing and fit as do premolded or user-molded plugs, the skill of the person making the impressions is of the utmost importance. If the initial impressions are not made properly, the resulting earplugs will not perform as they should and may be uncomfortable to wear.

Earmuffs. Earmuffs are circumaural devices that consist of earcups that completely enclose the ear and fit snugly against the side of the head, and a headband that is attached to each of the earcups. The function of the headband is to hold the earcups in place and to provide the necessary force to press the earcups' cushions against the side of the head. In lieu of a headband, some devices are attached to a hard hat via a spring-loaded mechanism. Earmuffs block the air conduction sound pathway by way of the cushioned seal of the earcups against the side of the head. Although earmuff fit is not affected by variations in ear canal shape and size, it is very much affected by head size, jaw shape and movement (Casali and Park, 1990), hair type and length, beards, eyeglasses, and items of safety equipment worn on or about the head. Earmuffs are often selected when an HPD is needed for intermittent use, but when worn for extended periods they can cause discomfort for the user. As with premolded earplugs, earmuffs are commonly modified by the user to enhance comfort but with a concomitant decrease in the protection afforded by the device. A common modification involves drilling holes in the earcups in an effort to improve communications, personalize the device, or to promote air circulation (Gasaway, 1984).

Semi-aurals. Semi-aural devices seal the opening of the ear canal, but are not inserted as deeply into the ear canal as an earplug. These devices are equipped with a small headband to hold them in place. This headband also allows the device to be stored around the neck when not in use. Attenuation performance of semi-aural devices is generally not as good as most earplugs or earmuffs, and their use is recommended only for intermittent exposures. They do, however, have the advantage over earmuffs in that they are generally unaffected by the wearing of eyeglasses and other pieces of safety-related headgear.

HPD Performance Characteristics

Each of the three types of HPDs described above possesses different spectral attenuation characteristics, as shown in Figure 4, where the manufacturer's advertised spectral attenuation data for four of the devices discussed above are plotted. As the data were taken directly from the HPD packaging, the plots represent the *best possible* performance attainable with the particular device. It has been shown (e.g., Park and Casali, 1991) that field performance of HPDs can fall far short of the performance that the manufacturer's advertised data would suggest. In addition, it should be noted that the data appearing in the figure represent the performance of four specific devices. The attenuation characteristics of the various products available on the market differ greatly across a single product type (i.e., not all premolded earplugs possess the same attenuation characteristics). However, the data are adequate to illustrate the general strengths and weaknesses of the various HPDs available in today's marketplace.

As can be seen, the largest differences in attenuation occur at frequencies below 2000 Hz. Earplugs attenuate low-frequency noise much better than either earmuffs or semi-aural devices. However, at the middle frequencies, earmuffs and earplugs have similar performance characteristics with earmuffs slightly outperforming the premolded earplug. At high frequencies, the devices again diverge with the earplugs and canal caps

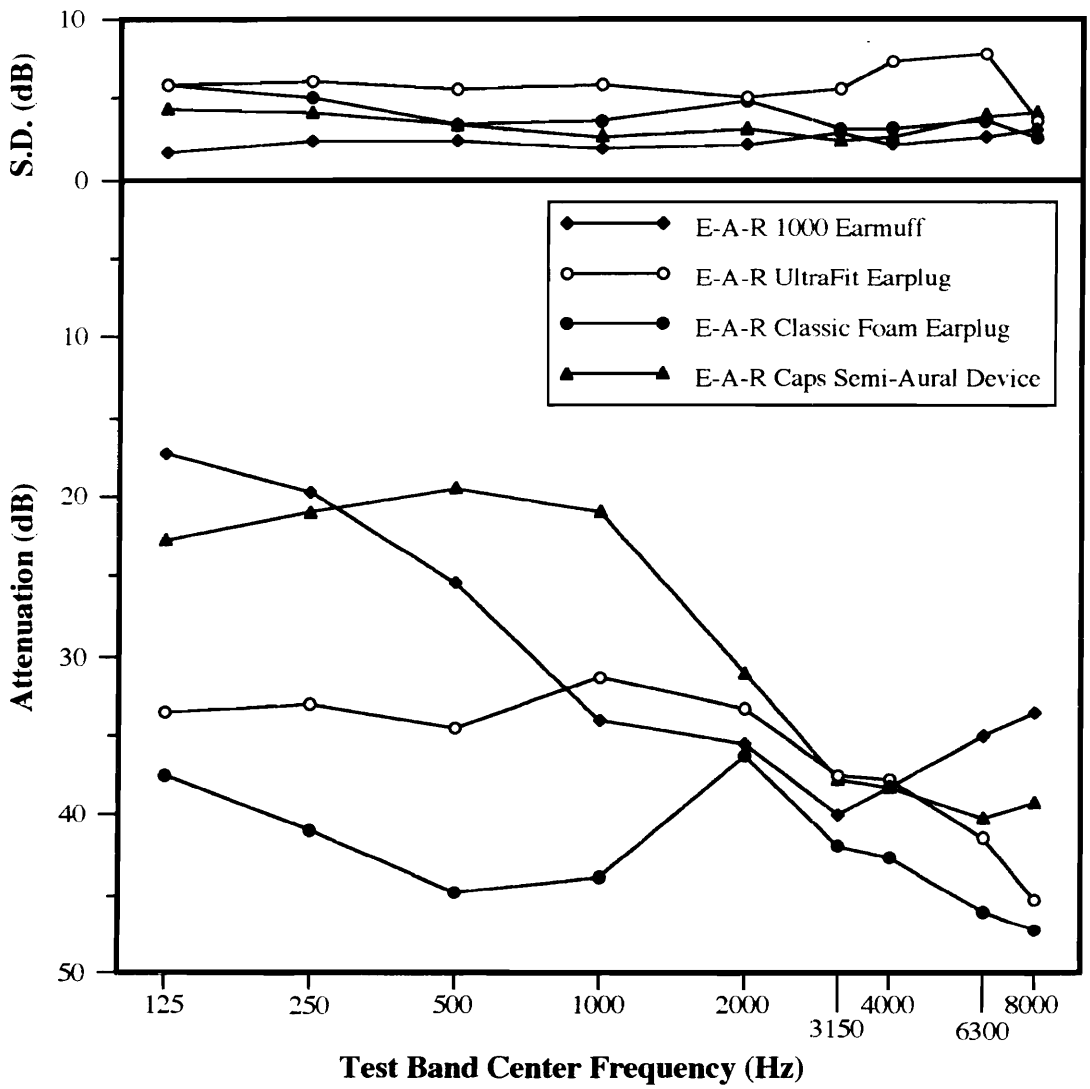


Figure 4. *Typical attenuation characteristics of various HPD types*. (adapted from manufacturer's data, circa 1991)

performing somewhat better than the earmuffs. In general, if an HPD is to be worn in noise that is predominated by low frequencies, an earplug would likely be preferred over an earmuff or semi-aural device. If, on the other hand, the noise is composed predominately of high frequencies, then an earmuff might be the better choice, all other things being equal. Semi-aural devices are most suitable in situations where the worker is only intermittently exposed to noise and must don and doff the device frequently (Casali, 1986).

AUDITORY ALARMS AND WARNINGS

Wilkins and Martin (1978) classify auditory alarm and warning devices into four broad categories depending on their design. These categories are: siren, horn, bell, and electronic device. Regardless of type, the sound produced by any given alarm or warning device can be varied in an almost infinite variety of ways by manipulating only three parameters [level, spectral (frequency) content, and temporal patterning (periodicity)], although several other parameters may also be varied. Level, and to a slightly less extent spectral content, are the primary determinants of signal detectability, while temporal patterning and spectral content aid discrimination and identification (Wilkins and Martin, 1978). Although discrimination and identification are extremely important issues when considering alarm and warning signals and will be discussed briefly, the experiment described herein was concerned only with detection.

Alarm and Warning Standards

Of the three parameters mentioned above, level and spectral content are the most likely to be specified in an alarm or warning standard. However, surprisingly few such standards exist. Perhaps the most comprehensive standard is International Standard ISO 7731–1986(E), "Danger Signals for Work Places–Auditory Danger Signals" (ISO, 1986). This standard not only specifies the spectral content and minimum signal-to-noise ratios of the signals, but also presents guidelines for calculation of the effective masked threshold of audibility using broadband, octave, or third-octave band analysis, and requires manufacturers of such devices to consider individuals suffering from hearing loss or those wearing HPDs (although it gives no quantifiable recommendations or procedures for doing so). The major ISO requirements are summarized in Table 2.

TABLE 2

Summary of Major Requirements of ISO 7731–1986(E), Danger Signals for Work Places–Auditory Danger Signals

Parameter	Requirement or Guideline
Level	Broadband estimate: A-weighted SPL of signal should exceed that of the background noise by 15 dB or more. Octave band estimate: Signal shall exceed the masked threshold (calculation method specified) by at least 10 dB in one or more octave bands as specified (see spectral content requirement below). Third-octave band estimate: Signal shall exceed the masked threshold by at least 13 dB in one or more third-octave bands as specified.
Spectral Content	The signal shall have its energy concentrated in the frequency range from 300 to 3000 Hz. Sufficient energy shall be present in the frequency range below 1500 Hz to satisfy the needs of individuals suffering from hearing loss or wearing HPDs.
Temporal Patterning	Pulsed signals are preferred. Pulses should be between 0.2 and 5 Hz. Pulse rate and duration are to be different from any periodically varying ambient noise in the work area.
Audibility and Discriminability	On-site listening tests are to be performed to ensure that the signal is both audible and discriminable. A minimum of ten subjects are to be used and the test is to be repeated five times. Subjects are to be representative of the workers who will be working in the area (in terms of age and hearing levels) and shall wear hearing protection if appropriate. 100% detection/discrimination is required.

From a review of the standards literature, it appears that only a few United States standards quantify requirements for audible alarms and warnings (DoD, 1981; NFPA, 1978, 1979; SAE, 1978; UL, 1978, 1981). For the most part, these standards specify only minimum levels and bandwidths, they do not consider individuals with hearing loss or wearing HPDs. The requirements of these standards are summarized in Table 3.

Although none of the standards are in complete agreement with each other, two summary observations can be made. There appears to be a consensus that the signal level should be about 15 dB above the noise level; and, while wider bandwidths are allowed, all of the standards which contain specific bandwidth information include the frequency range from 700 to 2800 Hz.

Other Alarm/Warning Guidelines

In addition to the various standards summarized in Tables 2 and 3, several authors have developed alarm and warning signal design guidelines which attempt to maximize signal detectability and/or discriminability.

Based on a lengthy series of experiments, Coleman et al. (1984) make the following recommendations concerning a warning signal's audibility, ability to gain attention, and discriminability. Signals should be at least 15 dB above the masked threshold (across its entire spectrum whenever possible), and no more than 25 dB above threshold (to avoid a startle response). Signals should have rise and fall times on the order of 20 ms to avoid startle. When signal levels reach 90 dB or higher, consideration should be given to the possibility of the signal contributing to the noise dose of the exposed individuals. Temporal patterning [inverse of the alarm's period -- how many periods (i.e., on/off cycles) per unit time] should be on the order of 1 to 4 Hz while modulation (amplitude and/or frequency fluctuations within a single period of the alarm)

TABLE 3

Summary of Requirements of U.S. Standards for Auditory Alarms and Warnings

Standard	Requirements
SAE J994b–1978; Performance, Test and Application Criteria for Electronically Operated Backup Alarm Devices – SAE J994b	Spectral content: Predominant frequencies are to be in the range from 700 to 2800 Hz. Level: Levels required vary by type. Five types specified: A – 112 dBA, B – 107 dBA, C – 97 dBA, D – 87 dBA, and E – 77 dBA. Measurement to be made at a distance of 4 ft from the alarm. Temporal Pattern: Recommends periods of 1 to 2 s with 50% duty cycles.
UL 464–1981; Audible Signal Appliances	Level: Minimum sound output of 75 dBA. Measurement of signal output specified to be in accordance with ANSI S1.21-1972 "Methods for the Determination of Sound Power Levels of Small Sources in Reverberation Rooms" which has been superseded by ANSI S1.31-1980 "Precision Methods for the Determination of Sound Power Levels of Broad-Band Noise Sources in Reverberation Rooms" and ANSI S1.32-1980 "Precision Methods for the Determination of Sound Power Levels of Discrete-Frequency and Narrow-Band Noise Sources in Reverberation Rooms." Temporal Pattern: Single stroke devices shall operate at a rate of 60 impulses/min with a 50% duty cycle.

TABLE 3 (continued)

Summary of Requirements of U.S. Standards for Auditory Alarms and Warnings

Standard	Requirements
UL 1023–1978; Household Burglar–Alarm System Units	Level: Minimum sound output of 85 dBA measured at a distance of 10 ft from the alarm. This minimum does not apply to units intended to be mounted in the same room with the users of the system, but rather to units intended to be place outside the building or centrally located within a building.
NFPA 74–1978; Household Fire Warning Equipment 1978	Level: Minimum sound output of 85 dBA measured at a distance of 10 ft from the alarm. The alarm "shall be clearly audible in all bedrooms over background noise levels with all intervening doors closed." Appendix A of the standard suggests that a signal 15 dB above the background noise (when measured in a bedroom) is adequate to awaken sleeping persons.
NFPA 72A–1979; Local Protective Signaling Systems 1979	Level: Recommends that the signal output should be 15 dB above the steady state background noise level. If the noise varies, recommendation made that the signal output be 5 dB higher than the *maximum* noise level. Temporal Pattern: Recommended on-time of 0.5 to 1 s and off-time of 0.5 s.

TABLE 3 (continued)

Summary of Requirements of U.S. Standards for Auditory Alarms and Warnings

Standard	Requirements
MIL–STD–1472C(1981); Human Engineering Design Criteria for Military Systems, Equipment and Facilities	Spectral content: Predominant frequencies are to be in the range from 200 to 5000 Hz, but preferably from 500 to 3000 Hz. When the distance to the alarm exceeds 300 m, only frequencies below 1000 Hz should be utilized. Level: Requires the signal to be 20 dB above the noise in at least one octave band in the operating frequency range.

should be 20 Hz or higher. Complex signals should be used which consist of harmonically-related components with a fundamental frequency below 1000 Hz.

Sorkin (1987) makes the following recommendations concerning auditory signals. Signals which are 6 to 10 dB above their masked threshold should ensure 100% detectability, while signals which are approximately 15 dB above their masked threshold will elicit rapid operator response. Warning signals should not exceed their masked threshold by more than 30 dB, and no signal should exceed 115 dB.

Wilkins and Martin (1978) also suggest that a signal should be at least 15 dB above its masked threshold to be detected reliably. In addition, they expressed a desire for improvements in the various methods for predicting such a threshold, citing that most prediction methods are only accurate to within ±5 dB.

A major shortcoming in most of the auditory warning and alarm standards and literature cited above is the general assumption that listeners will possess normal hearing and that HPDs will not be used. The single exception is ISO 7731-1986. However, even that standard fails to give quantitative guidelines concerning listeners with a hearing impairment or wearing HPDs, specifying only that the signal must be audible.

Identification and Urgency

The above discussions have been concerned primarily with signal detectability. However, there are other important aspects which must be considered when designing, specifying, selecting, or evaluating warning signals and alarms. Two such aspects are meaning and importance. Several attempts have been made to determine if there are certain invariant auditory qualities which cause some signals to be associated with dangerous situations. In one such experiment (Bock, Lazarus, and Hoege, 1983), 48 subjects listened to 36 signals (only 20 were of importance to the experimenters, the other 16 were artificial) presented in 4 levels of background noise. Each signal was rated on a

7-point bipolar rating scale from not dangerous to dangerous. It was found that of the 20 signals of interest, only 11 were rated consistently across all of the background noise levels. These signals were therefore deemed by the experimenters to be likely candidates for standardization.

In a second experiment (Höge, Schick, Kuwano, Namba, Bock, and Lazarus, 1988), cross-cultural differences in warning signal perception were investigated. In their experiment, 36 German and 74 Japanese subjects rated 41 actual or synthesized warning signals on 19 7-point bipolar rating scales. Results of the study indicate that cultural differences do exist which influence perception of warning signals and this may prevent the development of a standardized international danger signal.

Edworthy, Loxley, and Dennis (1991) conducted a series of experiments to determine what signal parameters influenced the perceived urgency of auditory warnings. Initially, a series of seven experiments were conducted in which various signal parameters (fundamental frequency, amplitude envelope, harmonic delay, rhythm, speed, pitch range, pitch contour, among others) were systematically manipulated. In each experiment, subjects rank-ordered signals according to their perceived urgency. These first seven experiments resulted in a set of rank-ordered signal parameters which could be manipulated to vary the perceived urgency of a warning or alarm signal. To verify the initial results, a set of 13 signals was constructed with a hypothesized order from most to least urgent. Subjects compared each signal to all the other signals and ordered each on a scale of perceived urgency. The subjective ordering agreed extremely well with the predicted order, with only one signal being rated out of its predicted sequence.

Haas and Casali (in press) also investigated how warning signal pulse parameters influence perceived urgency and detection time. The factors manipulated included pulse format (sequential, simultaneous, and frequency modulated pulses), pulse level, and inter-pulse interval (time between pulses). They found that perceived urgency increased as pulse level increased and as the inter-pulse interval decreased. The pulse format was also found to affect the perceived urgency of an auditory alarm in that pulses composed of sequentially-presented components were rated as less urgent than the other pulse formats investigated. In addition, it was found that detection time decreased as perceived urgency increased.

SIGNAL DETECTABILITY IN NOISE

The Masked Threshold

Masking can be defined as an increase in the absolute threshold of one (masked) sound caused by another (masking) sound (Gales, 1979). The masked threshold is therefore the level at which the masked signal is just audible. In the literature, the masked signal is usually a pure tone, while the masking noise may be one or more pure tones, a narrow band of noise, broadband noise, or complex sounds such as speech or music.

Masking by pure tones. Several plots of masking vs. frequency of the masked tone are illustrated in Figure 5. The number at the top of each plot is the frequency of the masking tone and the label on each solid curve is the level above threshold of the masking tone. The ordinate is the amount (in dB) by which the absolute threshold of the masked tone is raised by the masking tone while the abscissa is the frequency of the masked tone. As can be seen, masking is greatest at frequencies just to either side of the masking tone, the spread of masking is greater at frequencies above the masking tone than at frequencies below the masking tone, and these effects are magnified as the level of the masking tone increases. The notch at the frequency of the masking tone is due to the presence of beats (periodic fluctuations in amplitude due to the superposition of two simple harmonic waveforms of slightly different frequency — ANSI, 1973) in the region around the masking tone, making detection of the masked tone easier (Gales, 1979). This beats phenomenon is also apparent at frequencies above that of the masking tone which correspond to its harmonic frequencies (Deatherage, 1972). In practical situations, masking by pure tones would seldom be a problem, except in rare instances where the noise consists only of pure tones (or contains strong tonal components), or if two warnings with similar frequencies were activated simultaneously.

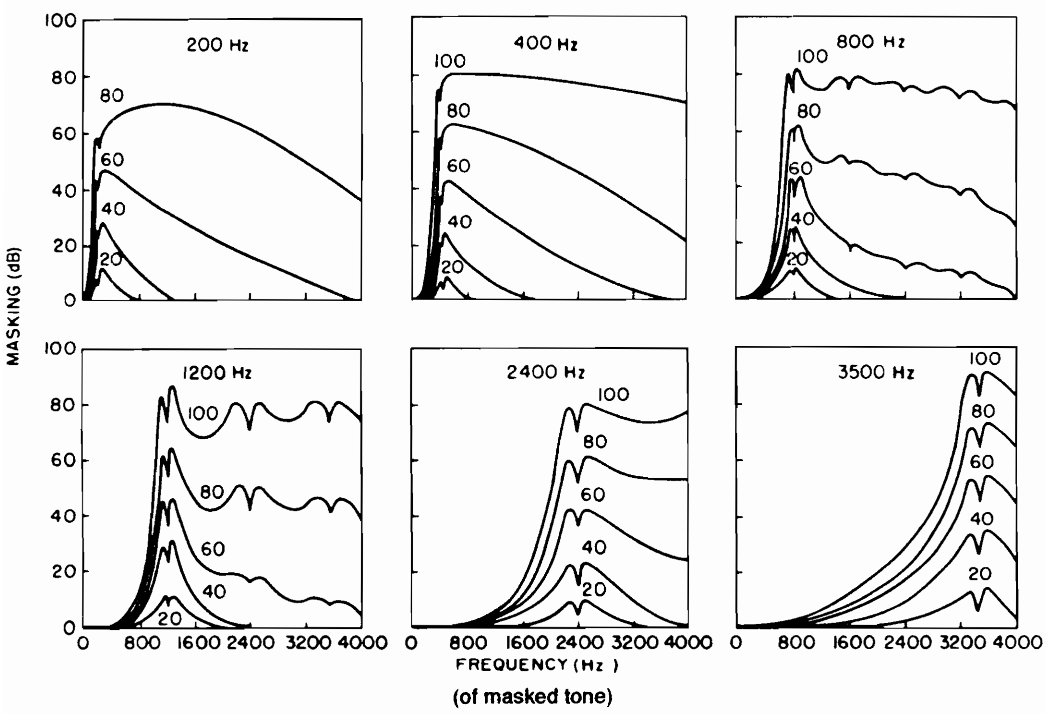


Figure 5. *Plots of masking vs. frequency of the masked tone for masking of pure tones by pure tones.* (from Deatherage, 1972)

Masking by narrow bands of noise. Egan and Hake (1950) investigated masking of pure tones by narrow bands of noise. They found that at the center frequency of the masking noise, the beats phenomenon so evident with pure tones was not apparent. They also found that narrow bands of noise were much more efficient maskers than were pure tones. Figure 6 is a plot of masking (increase in the threshold of the masked tone) vs. frequency of the masked tone for three levels (40, 60, and 80 dB) of a 90 Hz wide band of noise centered at 410 Hz. As is the case with pure tones, the spread of masking is much greater at frequencies above the band center frequency than at frequencies below the center frequency, and becomes more pronounced as the level of the masking noise increases.

Another phenomenon takes place when narrow bands of noise are used as maskers. This phenomenon is termed remote masking (Bilger and Hirsh, 1956; Spieth, 1957). In remote masking, narrow bands of noise centered at high frequencies cause considerable masking at low frequencies. It is suspected that such masking is due to low-frequency distortion in the cochlea (Kryter, 1985).

Masking by broadband noise. Perhaps the most common form of masking, especially in industrial workplaces, occurs when a signal is masked by a broadband noise. Hawkins and Stevens (1950) investigated how broadband noise masked pure-tone stimuli. In their experiment, they measured the masked threshold for 16 pure tones ranging in frequency from 100 to 9000 Hz when masked by uniform white noise at eight sensation levels (decibels above threshold) ranging from 20 to 90 dB. Their results appear in Figure 7. The perturbations in the plots above 6000 Hz were attributed to nonlinearities in the experimental apparatus. In the graph, the solid curves represent the masked thresholds for pure tones when masked by broadband noise of uniform spectral content. The labels on each curve represent the spectrum level of the masking noise with

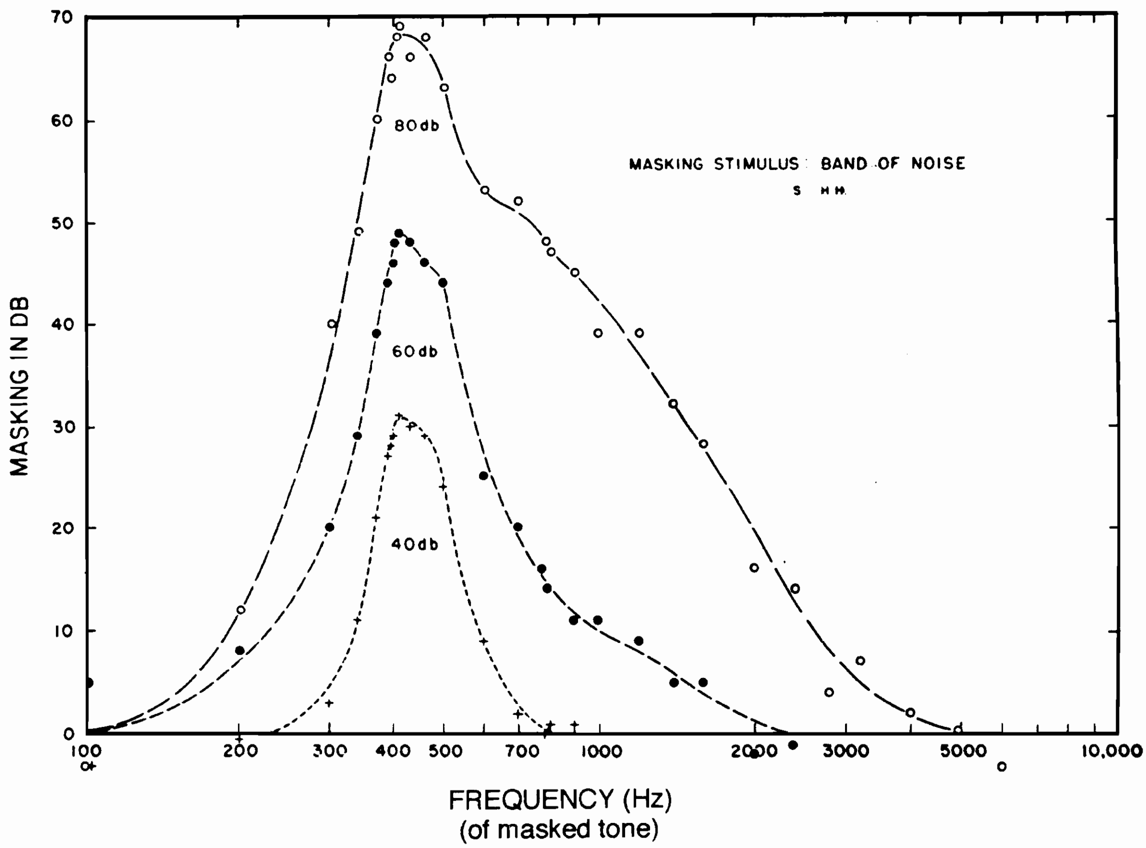


Figure 6. *Plot of masking vs. frequency of the masked tone for masking of pure tones by narrow bands of noise*. (from Egan and Hake, 1950)

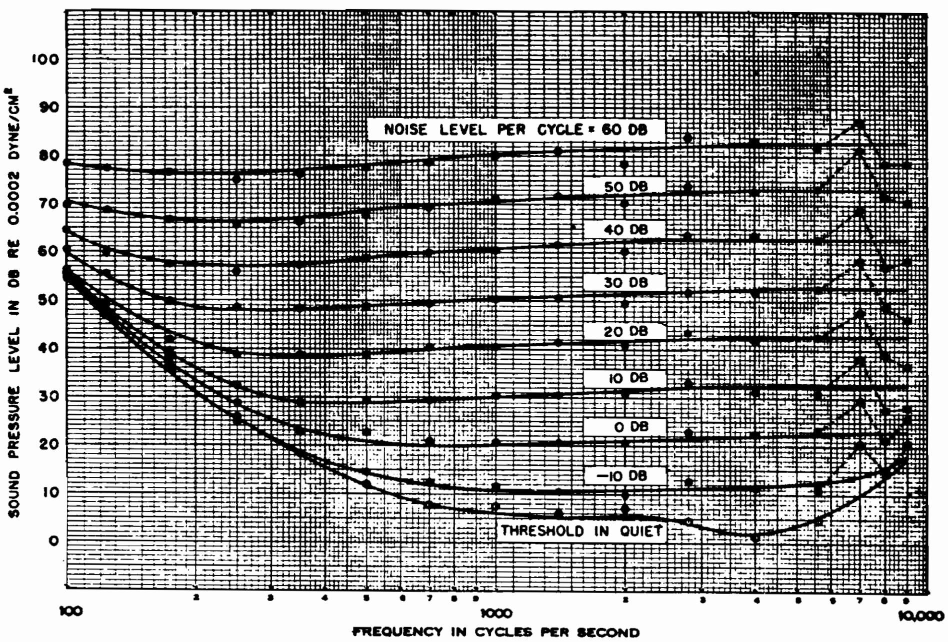


Figure 7. *Masked thresholds for pure tones when masked by broadband white noise.* (from Hawkins and Stevens, 1950)

the bottom curve representing the threshold of the experimental subjects in quiet. Due to the equal separation of the curves, the authors concluded that masking was directly proportional to the noise level. However, French and Steinberg (1947) found evidence of nonlinearities in the relationship when masking exceeded 50 dB, although the differences between the results of the two studies were 2 dB or less (Hawkins and Stevens, 1950).

To explain observed masking phenomena, Fletcher (1940) developed what would become critical band theory. According to this theory, the ear behaves as if it contained a series of overlapping filters, with each filter's bandwidth being proportional to its center frequency. As such, when masking of pure tones by broadband noise is considered, only a narrow "critical band" of the noise centered at the frequency of the tone is effective as a masker and that the width of the band is dependent only on the frequency of the tone being masked. In other words, the masked threshold of a pure tone could be predicted simply by knowing the frequency of the tone and the spectrum level (dB per Hz) of the masking noise (assuming that the noise spectrum is fairly flat in the region around the tone). Thus, the masked threshold of a tone in white noise would simply be:

$$L_{spectrum} + 10 \, \mathrm{Log}_{10}(BW)$$

where: $L_{spectrum}$ is the spectrum level of the masking noise, and
BW is the critical bandwidth centered around the tone.

Critical band theory as reported by Fletcher (1940) and Hawkins and Stevens (1950) apparently gained wide acceptance as evidenced by its citation in numerous textbooks and handbooks (Beranek, 1986; Deatherage, 1972; Gales, 1979). However, recent evidence (Patterson, Nimmo-Smith, Weber, and Milroy, 1982) indicates that the critical band is not merely a function of frequency, but also of age. Patterson et al. (1982) measured the width of the auditory filter by using a uniform white noise with a notch centered around the tone to be detected. As the notch width was varied in increments from 0.0 to 0.8 times the center frequency of the notch, the masked threshold of a tone whose frequency was centered in the notch was determined for 16 listeners ranging in age from 23 to 75 years. The results are illustrated in Figure 8. In the graph, the right ordinate gives, on an inverted scale, the effective bandwidth relative to the center frequency (f_c) of the auditory filter. Thus, the auditory filter for a 30 year old subject would be about 0.13 or 13% of the filter's center frequency. For a 60 year old subject, the filter width would be about 18% of the filter's center frequency. The wider the auditory filter, the higher the signal's masked threshold of detection.

To account for age differences, the masked threshold for a tone in white noise would be calculated in the usual manner using the equation given earlier. However, for noise that is not uniformly flat, it becomes necessary to integrate over the width of the auditory filter (Patterson et al., 1982; Sorkin, 1987).

Hearing Loss and the Masked Threshold

Only a few studies have investigated how masked thresholds vary with hearing level, and these generally address the problem as only a part of a larger issue. Berkowitz and Casali (1990) investigated the influence of age on the audibility of different telephone ringers. As part of their study, they measured the masked threshold of three telephone ringers in quiet and in 65 dBA pink noise for two groups of listeners: young normal-hearing listeners and elderly listeners who exhibited significant age related hearing loss (presbycusis). Their findings indicated that the masked thresholds of the elderly listeners were greater than those of the normal listeners even though the masked threshold of the normal listeners had been elevated to a point slightly above the thresholds measured in quiet for the elderly group.

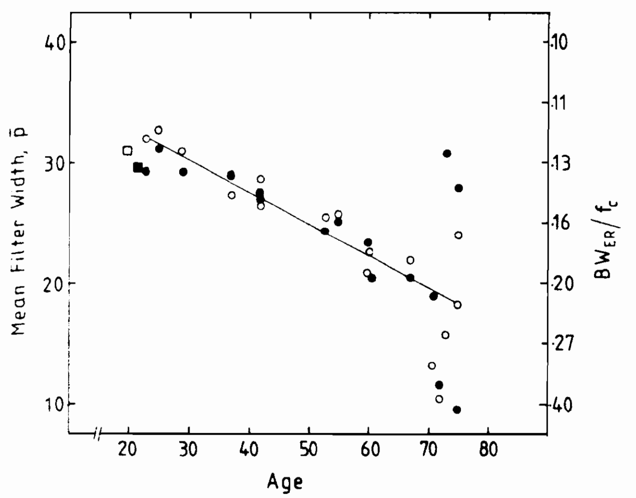


Figure 8. *Effective bandwidth* (BW_{ER}/f_c) *vs. age.* (from Patterson et al., 1982)

Researchers in Canada (Abel, Kunov, Pichora-Fuller, and Alberti, 1983a) obtained masked thresholds for both normal listeners and for two groups of hearing-impaired listeners as part of a study concerned with determining how wearing HPDs and hearing loss affects detection of auditory warnings. (This study will be discussed in greater detail in an upcoming section.) In the portion of the study pertinent to this discussion, third-octave wide signals centered at 1000 and 3000 Hz were presented via headphones in quiet and against two types of background noise (both of which were presented at 84 dBA) and the masked thresholds obtained. Their results indicated that the masked thresholds of the signals presented against an 84 dBA background noise for the normal and hearing-impaired listeners were comparable. However, statistical analysis was not performed due to unequal Ns and small sample size (Abel et al., 1983b).

In a field study intended to investigate the interaction of hearing loss and HPDs with warning signal detectability, Wilkins (1984) collected masked threshold data for normal and hearing-impaired listeners in noise without HPDs. In the experiment, two signals (the sound of clinking metal and a horn) were presented against a background noise whose level varied from 80 to 85 dBA. The signals were presented at five predetermined levels (the method of constant stimuli) and the response rates were measured. For the sound of clinking metal (composed of predominately high-frequency components), the response rates (collapsed across signal presentation level) were significantly less for the hearing-impaired listeners than for the normal listeners. This was the case both when the signals were listened for deliberately and when they were presented during normal working conditions. For the horn (composed predominately of low and middle frequencies), no such differences were found. Although it would have been possible to generate psychometric functions from the data and calculate a threshold based on a 50 or 75% detection criterion, no such calculations were reported.

At the 1948 San Diego County Fair, researchers (Webster, Himes, and Lichtenstein, 1950; Webster, Lichtenstein, and Gales, 1950) measured the masked thresholds of pure tones at 880 and 3520 Hz for 3666 individuals. The masking noise used was reported to be essentially flat from 256 to 9000 Hz with a spectrum level (dB per Hz) of 40 dB. The results indicate that masked thresholds for individuals judged to have neural hearing loss were not significantly different from those judged to have normal hearing, but that the normal and neural loss group had slightly lower masked thresholds than those individuals who were judged to have conductive or mixed conductive/neural hearing loss. Another interesting finding was that when considering only hearing-impaired subjects, those individuals who were aware of a hearing problem had masked thresholds significantly lower than those individuals who were unaware of a hearing problem.

The Effects of HPDs on Signal Detectability in Noise – Normal and Hearing-Impaired Listeners

Although it might be possible to separate the discussions of normal and hearing-impaired listeners in noise while wearing HPDs, it makes more sense to discuss them together and compare experimental results for the two groups. Few studies have addressed the detectability of auditory warning signals presented in noise while wearing hearing protection. Fewer still have included hearing-impaired listeners. Usually, only the occluded and unoccluded masked thresholds are obtained and compared. However, two researchers (Lazarus, 1980 and Coleman et al., 1984) do suggest methods for predicting the occluded masked threshold of an auditory warning or alarm.

Forshaw (1977) investigated the detectability of pure tones (ranging in frequency from 250 to 8000 Hz) against two types of background noise (broadband noise at 88 dBA which included all octave bands with center frequencies from 31.5 to 8000 Hz and tonal noise with line spectra at 460, 855, 1750, 1850, 2240, and 2560 Hz having levels ranging from 59 to 81 dB) found to exist in a ship's engine room. Detection thresholds of the pure tones in noise were determined both without HPDs and with earmuffs. Three normal-hearing subjects and one hearing-impaired subject participated in the first experiment using the broadband noise, while only one (normal-hearing) subject participated in the second experiment which utilized the tonal masking noise. Results indicated that although each of the normal subjects performed slightly better while wearing earmuffs in detecting the test tones against a background of broadband noise, the differences were not statistically significant. The hearing-impaired listener was said to have shown no adverse effect when wearing the HPD, but he was also said to have been completely unable to detect the 3000 Hz tone when in the occluded state. For the single subject exposed to the tonal masking noise, no difference in detectability with or without the earmuff was found.

In a study aimed at determining if HPD use impaired the detection of subtle machinery noises that often warn of impending failure, Kerivan (1979) conducted an experiment in which subjects were to detect a 10% upward shift in frequency of octave wide bands of noise centered at 500 and 2000 Hz. The background noise utilized was spectrally shaped to simulate the background noise found in a submarine engine room at a SPL of 70 dBA. Test signals were presented at signal-to-noise ratios of 0, -3, and -6 dB. All stimuli (noise and signal-plus-noise) were presented via earphones. Filtering was used to simulate the attenuation characteristics of an earplug and an earmuff. Though it would not have been practical to use a real earmuff with earphone presentation of test stimuli, use of a real earplug rather than a filtered simulation would have been possible. Although use of filter simulations of HPDs did produce equal attenuation across subjects, the attenuation values used were most likely those specified by the manufacturer and therefore unrealistically high, especially for the earplug (Park and Casali, 1991). The signals and noise would therefore have been attenuated more than they would have had real HPDs been used under realistic conditions. In addition, use of filter simulations of HPDs and earphone presentation of test stimuli does not allow for interaction between the subject and the sound field (i.e., bone and tissue conduction) or between the HPD, sound field, and the mechanisms responsible for distortion in the middle and inner ear. Results indicated that in all cases where HPDs were simulated, detection performance was significantly lower than in the unfiltered condition, with earmuffs degrading performance more than earplugs. These results are not surprising given the low noise level and signal-to-noise ratios used in the experiment.

Canadian researchers (Abel et al., 1983a) conducted an experiment which included not only normal listeners, but hearing-impaired listeners with two levels of hearing loss represented (high-frequency loss at 3000 Hz of between 35 and 85 dBHL, and high-frequency loss at both 1000 and 3000 Hz of at least 35 dBHL). In their experiment, listeners were asked to detect the presence of a third-octave band of noise centered at either 1000 or 3000 Hz in quiet and against two different background noises presented at 84 dBA. Although the test stimuli may not have been representative of most warning signals or alarms, the frequencies chosen did fall within the ranges specified by both ISO 7731–1986 and MIL–STD–1472 (1981). The background noises were taped samples of industrial noise (mining operations) and were of different spectral content. One was low-frequency biased, whereas the other was biased in the middle range of frequencies, but with strong low-frequency components. The detection tasks were carried out in both the unoccluded and occluded states. The HPD used was a high attenuation foam plug. Use of an earplug was necessary since all stimuli were presented via headphones. The normal listeners participated in 12 sessions and were exposed to all 12 conditions (3 levels of noise, 2 signals, and 2 protection states). However, the hearing-impaired listeners experienced only 6 conditions, with half of the subjects in each of the hearing-impaired groups listening for one of the signals and the other half of the subjects listening for the other signal.

Results indicated that for subjects with normal hearing, use of an earplug (with a mean attenuation of 28.1 dB at 1000 Hz and 40.6 dB at 3000 Hz) reduced the masked threshold of the signal centered at 3000 Hz by 3 to 6 dB, but did not significantly affect the detectability of the signal centered at 1000 Hz. Both hearing-impaired groups showed significant increases in the threshold of detectability for the signal centered at 3000 Hz when the earplug was worn (thresholds were greater than 100 dB for all three noise conditions). However, for the signal centered at 1000 Hz, no significant changes in masked thresholds were found when the earplugs were used. (For the listeners with a loss only at 3000 Hz, the mean attenuation achieved with the earplug was 28.8 dB at 1000 Hz and 40.0 dB at 3000 Hz. For the listeners with a loss at both 1000 and 3000 Hz, the mean attenuation achieved with the earplug was 34.4 dB at 1000 Hz and 41.0 dB at 3000 Hz.)

The authors concluded that for high-frequency signals, HPDs can improve the detectability of signals in high-noise environments for normal-hearing subjects. Also, for individuals with mild hearing loss (30 to 35 dBHL) in the frequency region of the signal being listened for, detectability will not be adversely affected by use of HPDs, but for individuals with more severe hearing loss (60 dBHL or more) detection thresholds will increase substantially. These results, however, may not be generalizable since only one high-attenuation foam earplug was used. In addition, as was the case with the study conducted by Kerivan (1979), experimental stimuli were presented to the subjects via earphones. The authors suggested further study to determine at what point hearing loss becomes sufficient to impair an individual's ability to detect signals in noise.

Wilkins (1984), and Wilkins and Martin (1977, 1981, 1982, 1984, 1985) performed a series of studies examining the problems associated with detecting signals in noise when wearing hearing protection. The first two experiments (Wilkins and Martin, 1977, 1982) consisted simply of determining the masked thresholds (with and without use of HPDs) of six warning sounds presented against various background noises. In the first experiment, two signals (a bell and a siren), two levels (75 and 95 dBC) of a white background noise, and three HPDs (an earplug, an earmuff, and a filter simulation of the earmuff) were used. Results showed that for the 95 dBC noise level, the masked thresholds for all of the occluded conditions were significantly lower than those measured in the unoccluded condition (however, the numerical differences were small).

In the second experiment (Wilkins and Martin, 1982), four signals (a siren, a wavering high/low signal, and the individual high and low components of the high/low signal) and two different noise spectra (a white noise and a taped sample of industrial noise) presented at 90 dBC were used. Only one HPD (an earmuff) was used in the second experiment. Again, masked thresholds were obtained for subjects in both the occluded and unoccluded condition. Statistically significant differences were found only between the occluded and unoccluded conditions for the siren and the high/low wavering signal (independent of noise spectra) with the occluded condition producing the lower masked thresholds.

In addition to the simple masked threshold experiments described above, Wilkins and Martin (1981, 1982, 1984, 1985) also performed a series of experiments aimed at determining how inattention affected detection of warning signals, if various noise and signal parameters interacted during periods of inattention, and if these interactions affected occluded masked thresholds.

In the first of these attentional demand experiments (Wilkins and Martin, 1981, 1982), a recorded siren was used as the warning sound and was presented against a background of random noise at a level of 75 dBC. The psychophysical method of constant stimuli was used with the five signal presentation levels separated by 5 dB. Presentation levels of the signals were not specified. Signals were presented in random order and the temporal spacing was varied randomly from 20 to 160 seconds. Detection tasks were performed in both the occluded (wearing an earmuff) and the unoccluded conditions. Finally, detection also took place either under a vigil condition in which the signal was intentionally listened for or in a loaded condition during which the subjects performed a secondary task, consisting of a single player video game. Subjects received a monetary reward based on their total performance in the detection task and in the video game.

Results of the experiment revealed that there was no difference in detection performance between the vigil condition and the loaded condition. This finding contradicts results of an experiment reported by Fidell (1978) in which detection thresholds were found to be higher than predicted when a loading task was performed. Wilkins and Martin (1981, 1982) suggest that the different results may have been do to use of the ascending method of limits in the earlier study, or to the fact that the subjects in the earlier study were not as well motivated as were the subjects in their own study. If the latter is the case, the increased detection thresholds measured in the Fidell (1978) study may have actually been due to shifts in the subjects' criteria rather than a change in their sensitivities.

In the second attentional demand experiment (Wilkins and Martin, 1982, 1984), the procedures were essentially the same as in the previous experiment with a few minor changes. First, a second warning signal (the sound of a grinder) was included. This new signal was added in an effort to determine how well an indistinct, incidental machinery sound would be perceived in noise while wearing earmuffs and performing a loading task. In addition, three distracting sounds (an engine, lathe, and drill) were also presented interspersed with the signal being detected to determine how well each of the target signals could be recognized while performing a loading task in noise and wearing earmuffs.

Results indicated that at signal levels above threshold, subjects were able to reliably recognize both warning sounds amid the distracting sounds. However, the sound of the grinder was detected significantly less often while the subjects were occupied with a loading task.

In their last attentional demand experiment, Wilkins and Martin (1982, 1985) tried to determine why the grinder sound used in the previous experiment proved to be a less effective warning signal than the siren as indicated by the significant difference in detection rate for the grinder between the vigil and loaded conditions. They hypothesized that signal effectiveness is a function of the contrast of the signal to irrelevant sounds (C_S) and with the noise (C_N). The experiment was conducted in a manner exactly like the previous experiment except that the two target signals to be detected were a pure tone of 2000 Hz, and a third-octave band of noise centered at 2000 Hz. Irrelevant stimuli were pure tones of 800 and 5000 Hz. Therefore, the 2000 Hz tone had a high contrast with the background noise (C_N) and low contrast with the irrelevant stimuli (C_S). The third-octave band of noise, on the other hand, had high C_S but low C_N.

Results indicated that both signals were perceived significantly less often in the loaded condition than in the vigil condition, and the 2000 Hz tone was detected significantly less often than the third-octave band of noise. Use of earmuffs significantly impaired detection of both signals. Wilkins and Martin (1982, 1985) interpreted these results to mean that both C_S and C_N are important parameters in detection tasks, with C_S probably being slightly more important than C_N.

Finally, Wilkins (1984) conducted a field study intended to investigate the interaction of hearing loss and HPDs with warning signal detectability. Details of the experimental procedure were presented earlier and are not repeated here. As discussed before, the unoccluded masked threshold for both signals (clinking metal and a horn) were significantly higher for the hearing-impaired subjects than for the normal-hearing subjects. In addition, it was found that hearing-impaired listeners detected significantly fewer clinking stimuli than did the normal-hearing subjects when HPDs were worn. Neither group showed any significant differences in detection of the horn between the occluded and unoccluded conditions. However, the author cautioned readers about taking the results too literally and cited several possible confounding variables of his study. Weaknesses cited include (Wilkins, 1984, p. 433): fluctuating noise levels in the factory, the presence of uncontrollable irrelevant sounds which might have elicited inappropriate responses from the subjects, some subjects might have made responses based on observations of other subjects, and finally, an inability to ethically require subjects to not wear their HPDs when determining unoccluded masked thresholds.

Coleman et al. (1984) took a different approach to the problem of examining occluded masked thresholds for normal and hearing-impaired listeners. Rather than simply measure masked thresholds for groups of subjects, these researchers developed a method by which the occluded masked thresholds could be predicted. Their experiments then attempted to validate the prediction method. The model was based on the auditory filter work of Patterson (1974, 1976) and Patterson et al. (1982) discussed briefly in a previous section. To apply their model and predict the occluded masked threshold of an individual (or population), it is necessary to know the spectral makeup of the masking noise, the spectral attenuation characteristics of the HPD used, the spectral characteristics of the signal being considered, the pure-tone threshold of the individual (or group) in question, and finally an estimate of the auditory filter width of the individual (or a suitable population estimate). Details of the procedure are contained in Coleman et al. (1984). When predicted occluded masked thresholds were compared to the measured thresholds for a representative sample of mine workers, the mean predicted values consistently overestimated the mean measured thresholds by about 5 dB (thus, the predicted thresholds were slightly conservative). Although shown to be fairly accurate, the procedure requires information that may not be generally available to individuals responsible for administering hearing conservation programs. Therefore, it is doubtful this prediction method will find general acceptance.

Lazarus (1980) also proposed a method for determining whether or not a signal would be audible to an individual wearing an HPD in noise. The proposed method requires knowledge of the third-octave spectral characteristics of the noise, the spectral attenuation characteristics of the HPD considered, and the masked threshold of the signal if it were being listened for without an HPD. To apply the method, it is first necessary to calculate the masked threshold of the signal in the background noise (L_M) and the masked threshold of the signal if the noise were to be reduced by an amount equal to the attenuation of the hearing protector (L'_M). Next, the attenuation of the HPD (R) is subtracted from the unoccluded masked threshold ($L_M - R = L$) and the result compared to L'_M. If $L \geq L'_M$, then the signal should be audible. If $L < L'_M$, the signal would likely be inaudible. [The notation used is that of Lazarus (1980).] No empirical evidence of the accuracy of this method is reported.

A second group of Canadian researchers (Laroche, Tran Quoc, Hétu, and McDuff, 1991) have developed a computer program capable of predicting masked thresholds in noise (both with and without hearing protectors) which takes into account hearing loss due to age. The model does not, however, consider the effects of noise-induced hearing loss or hearing loss due to injury, disease, or other etiology. The authors do acknowledge the model's shortcomings and discuss plans for expanding the model not only to include additional sources of hearing loss, but also to include the effects of sound propagation in sound fields with vastly different reverberation characteristics, and variation of the spectral and temporal characteristics of both the background noise and the warning signal (the current model assumes steady-state conditions).

Concern over an individual's ability to hear warning sounds while in a noisy environment and wearing a noise attenuation device is not limited to industrial interests. In an experiment aimed at assessing the noise exposure to motorcycle riders, Van Moorhem, Shepherd, Magleby, and Torian (1981) measured the insertion loss characteristics of two common types of motorcycle helmets (full-face and conventional). Also measured were typical noise levels inside the helmet and the spectral (octave band) characteristics of an emergency vehicle siren and an approaching automobile at a distance of 25 m. Calculation of the signal levels under the helmet for each of these "warning sounds" indicated that they would still be audible to a listener with normal hearing.

In summary, it appears that for normal-hearing individuals, the use of HPDs will not adversely affect their ability to detect warning or indicator sounds in high-noise environments. In fact, their use may actually improve signal audibility in some circumstances. For hearing-impaired listeners, particularly those with high-frequency neural losses, it appears that use of HPDs in noise may impair the ability to detect warning signals. Similar conclusions were reached by Wilkins and Martin (1987) in a review of several foreign language articles and unpublished British research papers. The effects, however, may be limited to frequencies above 1000 Hz (assuming noise-induced hearing loss). In addition, if the results reported by Abel et al. (1983a) are valid, impaired detection ability may not be apparent until hearing loss exceeds some level between 35 and 60 dBHL. Also, although there is some evidence to the contrary, it would appear that for well designed warning signals, inattention may not significantly affect the detection of the signals in noise when wearing HPDs. However, for this to be true, the signal should be distinct from both the background noise and any incidental sounds which may occur. Finally, although methods have been developed to predict the occluded masked threshold of auditory alarms and warnings, they are somewhat complex and do not readily lend themselves to general use in industry.

FACTORS AFFECTING SIGNAL DETECTABILITY IN NOISE WHILE WEARING HEARING PROTECTION

As can be seen, there are numerous factors which will (or may) affect the audibility of an alarm or warning signal when presented against a background of noise while the listeners are wearing HPDs. These factors include, but may not be limited to: the level and spectral content of the background noise, the level and spectral content of the signal, the absolute signal-to-noise ratio, the attenuation characteristic of the HPD, the hearing thresholds' of the listeners, and motivation of and/or attentional demand on the listeners. Each of these factors will be briefly discussed in the following section as to how they might affect warning signal audibility and how the problem might be circumvented or eliminated.

Characteristics of the Background Noise

Perhaps the most important characteristic of the background noise which affects signal audibility is its overall sound pressure level (SPL). Based on the research cited in the previous section, it would appear that in high noise levels, listeners with *normal hearing* will be better able to detect a warning signal while wearing an HPD than they would if they were not wearing the HPD. The research conducted by Forshaw (1977, using 88 dBA noise), Abel et al. (1983a, using 84 dBA noise), and Wilkins and Martin (1977, 1982, using 95 dBC noise) support this finding. However, when Kerivan (1979) used noise at only 70 dBA, he found that signals were less reliably detected when using HPDs than when HPDs were not used. Wilkins and Martin (1977, 1982) found no difference in detection thresholds with noise presented at 75 dBC. It would appear, therefore, that at some point between about 70 dBA and 85 dBA, the use of HPDs (by individuals with normal hearing) will begin to improve the likelihood of detecting an auditory alarm signal. Exactly where this point falls is not certain and it is likely to be dependent on a number of interacting factors. Although, if average noise levels are only about 70 to 80 dBA, it is doubtful that HPDs would be in general use unless the noise were of an intermittent nature. This brings up another point; the use of HPDs in intermittent noise may also reduce the likelihood of detecting an alarm or warning signal. If during a period of relative quiet a signal were audible to an individual not using an HPD, then it is quite possible that use of an HPD could lower the level of the signal to the point that it would be less audible, if not inaudible. However, in such a situation, the signal would probably not be audible during a period of high noise (Wilkins and Martin, 1978).

Another important characteristic of the background noise which may affect the detection of alarm and warning signals is its spectral content. Although not addressed directly in the body of research discussed above, it is quite possible that noise with extremely strong low-frequency bias may mask signals more efficiently than a spectrally flat noise due to upward spread of masking from the low frequencies. This point was mentioned by Lazarus (1980) when discussing his prediction method. Finally, as mentioned by Wilkins and Martin (1982, 1985), the presence of irrelevant sounds in the background noise or noise with strong tonal qualities (Forshaw, 1977) may reduce the likelihood of detecting an alarm or warning signal.

A solution to the above problems would be to implement controls that would eliminate, or at least reduce, the workplace noise causing the problem. If the overall noise level were to be reduced to a level of less than 80 to 85 dBA, HPD use would not be required. In a similar vein, if the noise output from irrelevant sound sources could be reduced or eliminated, the likelihood of a signal being detected would be increased. However, this is infeasible from both an economic and engineering standpoint in many industrial situations.

Characteristics of the Signal

The audibility of alarm and warning signals appears to be primarily dependent on two factors: the intensity of the signal relative to the level of the background noise (or signal-to-noise ratio) and the spectral content of the signal (including the contrast of the signal to the background noise). To some degree, both the intensity and the spectral content of such signals is specified in the various standards discussed earlier. However, those design recommendations seem to disregard individuals with hearing loss and/or individuals who will be using HPDs. The single exception is ISO 7731–1986(E), which does require that HPDs and hearing-impaired listeners be considered and that the alarm be audible to those individuals.

Considering the research cited earlier, it would appear that signals with most of their energy centered at frequencies near 1000 Hz are more likely to be detected than signals at higher frequencies (Abel et al., 1983a) and those signals which are in sharp contrast to the noise (Wilkins and Martin, 1982, 1985) are more likely to be detected when the listener is wearing an HPD. It is therefore the responsibility of the employer to carefully select warning and alarm signals used in the workplace.

HPD Attenuation Characteristics

Most of the research cited in the previous section used only one type of HPD. The two exceptions were Kerivan (1979) and Wilkins and Martin (1977, 1982) who both investigated an earplug and an earmuff (or filtered simulations of the two devices). Neither study showed a difference between the two devices. However, Kerivan (1979) used a low level of noise (70 dBA) which would likely not have shown a difference anyway. Although Wilkins and Martin (1977, 1982) did not show a difference between the two types of devices, this does not mean that a difference cannot exist, given the proper conditions. Properly-fit earplugs generally do a much better job of attenuating low-frequency noise than do earmuffs. Therefore, with a signal whose energy is centered in the range from 1000 to 1500 Hz and a noise having a low-frequency bias, it is possible that the use of an earmuff would still allow upward spread of masking to occur under the device and reduce the audibility of the signal. This problem was mentioned by Lazarus (1980) and by Wilkins and Martin (1978). In cases of this sort, the solution would be to use an earplug rather than an earmuff.

Hearing Level

Although the evidence is rather limited, it is probably safe to say that individuals with some degree of hearing loss will be less likely to hear an auditory alarm or warning than a normal-hearing individual when they are wearing HPDs. However, the degree of hearing loss necessary to make the above statement true is, at present, unknown. Abel et al. (1983a) estimate the point to be between 35 and 65 dBHL at 3000 Hz for detection of a signal at 3000 Hz. It so happens that the listener in the study performed by Forshaw (1977) also had a problem at 3000 Hz. What is needed is research aimed at determining the point in a hearing loss profile at which detection of signals begins to be degraded. Furthermore, more work is needed to determine the interactive effect of noise level on the hearing-impaired listener's ability to detect signals under HPDs.

Motivation and Attentional Demand

Not only is the research in this area as it pertains to the problem at hand sparse, but it is contradictory as well. As mentioned earlier, Wilkins and Martin (1981, 1982, 1984, 1985) conducted a series of attentional demand experiments and found no decrease in detection performance due directly to the existence of a loading task. What differences they did find (Wilkins and Martin, 1982, 1984, 1985) were actually interactions with signal and noise contrast factors. The differences found by Fidell (1978) could easily have been due to factors other than the presence of the loading task. In the first place, Fidell (1978) did not determine his subjects' detection thresholds prior to the experiment, but rather relied on predicted thresholds estimated by a method developed earlier (Fidell, Parsons, and Bennett, 1974). Although the prediction method was validated in earlier experiments using relatively low level noise (the spectrum level of the most low-frequency biased noise ranged from 40 dB at 125 Hz to 10 dB at 8000 Hz – broadband level measures were not specified), the background noise used in the 1978 experiment was extremely low-frequency biased and presented at a level several times that used to verify the prediction method (the spectrum level ranged from 84 dB at 125 Hz to 38 dB at 8000 Hz). Secondly, the ascending method of limits was used in the experiment and errors of habituation (Gescheider, 1985) may have elevated the thresholds. Finally, the subjects were told that the driving task they were performing as a loading task was actually the primary task and that the detection of the auditory stimuli was incidental. This could have affected the subjects' criteria for responding to the signal.

Therefore, based on the preceding discussion and on the results of the investigations of Wilkins and Martin (1981, 1982, 1984, 1985) discussed earlier, it is believed that the elevated thresholds found by Fidell (1978) are likely experimental artifacts and that if a signal is well designed and contrasts sufficiently with the environmental noise, it will likely be just as detectable during periods of inattention as during periods of attention. It is the responsibility of the employer to ensure that a properly designed signal is chosen.

THEORETICAL BASIS OF MASKING

Several phenomena have been observed in research in the area of auditory masking. These phenomena include the nonlinear growth of masking (Coleman et al., 1984), the widening of the auditory filter (Weber, 1977), and cochlear distortion (Lawrence and Yantis, 1956). Each of these phenomena will be discussed individually below.

Many authors (Beranek, 1986; Deatherage, 1972; Gales, 1979; Hawkins and Stevens, 1950) state that masking is a linear function of the level of the masking noise such that an increase in masker level of 10 dB will result in an increase of 10 dB in the masked threshold of the signal. However, results presented by other researchers (Coleman et al., 1984; French and Steinberg, 1947) showed a slight nonlinearity in the growth of masking with an increase in level of the masking noise. Referring back to Figure 7, it would seem that masking is indeed linear for broadband noise since the curves representing masked thresholds for various noise levels are equally spaced. The nonlinearities observed by French and Steinberg (1947) and by Coleman et al. (1984) were very small and insufficient information was presented to determine if their results were reasonable. Most of the research investigating masked thresholds has been conducted using pure tones or white noise (or narrow bands of noise with uniform spectral levels), and earphone presentation. Coleman et al. (1984), on the other hand, used taped samples of real mining noise and presented the pure-tone test stimuli via loudspeaker inside a reverberant room. It is therefore possible that the differences might have been due to experimental artifacts.

Weber (1977) reported measurable increases in the width of the auditory filter as the level of the masking stimulus increased. Kryter (1985) also makes the statement that the critical bandwidth increases with the level of the masker. But Scharf (1970), in an earlier review of the literature on critical bands, states that the critical band is independent of masker level, as has Deatherage (1972). If the auditory filter does widen as the intensity of the masking noise increases, then it follows that masking levels should also increase in a nonlinear fashion.

In has already been stated that Patterson et al. (1982) found that the width of the auditory filter increases with age. However, because of presbycusis, it is not possible to completely separate age from hearing loss if a wide age range is represented. In fact, if the data presented by Patterson et al. (1982) are examined closely, it is found that auditory filter width and hearing level (as measured using pure-tone audiometry) are highly correlated (r = -0.88) at 4000 Hz. This frequency was the highest frequency tested in the experiment, but it does represent a frequency at which moderate to substantial hearing loss could be expected due to either presbycusis or noise-induced hearing loss. In a related discussion, Scharf (1970) states that conflicting evidence exists as to whether or not the width of the auditory filter increases as a result of cochlear damage (including hearing loss).

The term *cochlear distortion* is often used to describe the overloading of the ear by intense stimulation. At high noise levels, it is this cochlear distortion which leads to the inability to discriminate signals from noise. It is supposed that the reason HPDs improve an individual's ability to discriminate speech and/or signals in high noise environments (at least for individuals with normal hearing) is that the HPD reduces both the speech (or signal) and noise such that the cochlear distortion no longer impairs the discrimination of the signal from the noise. Lawrence and Yantis (1956) conducted an experiment aimed at determining at what levels cochlear distortion first appeared. In their experiment, the presence of the second harmonic (as determined by beats of a probe tone) of a pure-tone stimulus was used to signal the presence of distortion components. [Probe tones were used to investigate harmonic distortion in the ear as follows: Two pure tones were simultaneously presented to the ear, the second (probe) tone differing in frequency from the second harmonic of the first (stimulus) tone by only a few Hz. The frequency of the probe tone was then adjusted until the beats produced by the interaction of the probe tone with the second harmonic of the stimulus tone (which was adjusted in amplitude) were audible (Lawrence and Yantis, 1956).] Their results indicate that distortion components occur at stimulus levels as low as 8 dB above threshold at 100 Hz increasing to 50 dB above threshold at 1000 to 5000 Hz. The presence of distortion components in the form of harmonics of the stimulus signal might also explain, at least partly, the phenomenon of upward spread of masking.

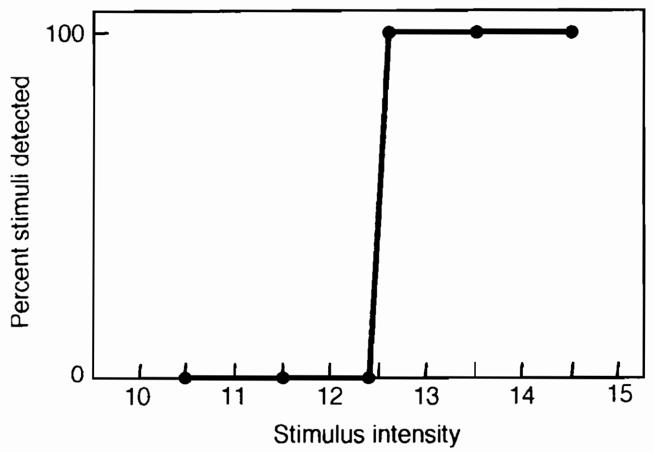
The phenomena discussed briefly in the preceding discussion cannot be considered in isolation, but rather must be considered together, as interrelated parts of the same problem. Theories concerning the existence of nonlinear growth of masking, critical bands, and cochlear distortion are empirical in nature and have been developed in an attempt to explain how the human ear functions when exposed to noise. Science has yet to determine the true nature of the mechanisms underlying the observed phenomena. Until these true mechanisms are discovered, these empirical models will continue to be refined, extended, or discarded and new models developed.

CLASSICAL HIGH THRESHOLD THEORY VERSUS SIGNAL DETECTION THEORY

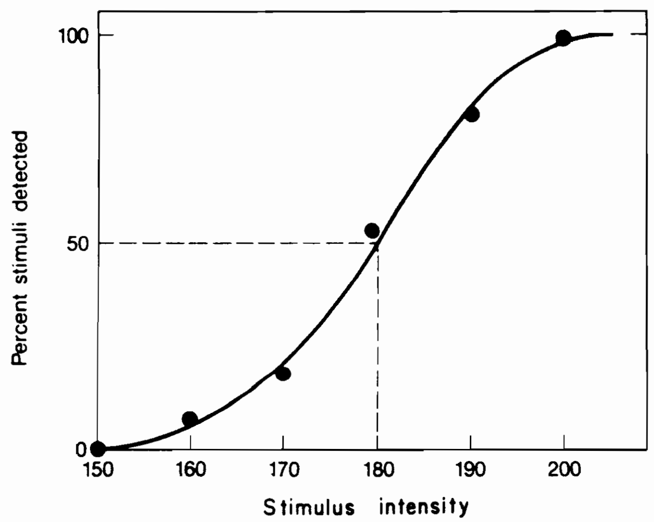
High Threshold Theory

Classical high threshold theory assumes that when the energy of a stimulus is below a certain threshold level, the stimulus is incapable of eliciting a response from an observer. (In the context of this discussion, only auditory stimuli will be considered; however, the theory applies to any sensory stimulus.) Any stimulus with energy less than this threshold will not be detected, while a stimulus with energy greater than this threshold will be detected. The concept of such a threshold pertains not only to the minimum energy necessary for *detection* of a stimulus (the absolute threshold), but also for the incremental increase in stimulus energy necessary for perceiving a change in the stimulus (difference threshold). The concept of the difference threshold is that an observer's sensory system requires a certain incremental increase in energy before a difference is detectable. If the change in the stimulus is less than this incremental amount, no change will be observable, but as soon as this increment is exceeded, a change will be observable (Gescheider, 1985). When considering masking phenomena described earlier, the masking noise serves to load an observer's auditory system so as to require a much larger initial sound pressure level before an auditory signal can be detected.

However, when experimental data for the percent of the correct detections are plotted against stimulus intensity (referred to as a psychometric function), the nearly step-shaped plot predicted by theory, Figure 9(a), does not result, but rather an ogive-shaped plot is obtained, Figure 9(b), (Gescheider, 1985; Green and Swets, 1988). Proponents of threshold theory reconcile this finding with theory by proposing that at any instant in time, a subject's threshold is indeed a step function as hypothesized by threshold theory,



(a) hypothesized



(b) observed

Figure 9. *Psychometric functions of classical threshold theory.* (from Goldstein, 1989)

but that this momentary threshold varies randomly over time with an underlying normal distribution resulting in the smooth ogive-shaped psychometric functions obtained in their experiments (Gescheider, 1985). The threshold is therefore taken as that signal level corresponding to some predetermined percentage of correct positive responses, usually the 50% point (Green and Swets, 1988).

In addition to the observed random variation of the sensory threshold itself, other factors may also affect the results of psychophysical experiments. Such factors include inattention on the part of the subject, practice effects, inexperience, anxiety, or criterion shifts. The developers of the classical psychophysical methods attempted to control these factors as much as possible in the design of the methodologies (Green and Swets, 1988) as well as in the selection and training of subjects. Implicit in these procedures is the assumption that only the use of carefully screened, well-trained, experienced, and motivated subjects would produce reliable results. In addition, it is assumed that any bias on the part of the subject could be either eliminated or controlled sufficiently to prevent it from affecting the results of the threshold determinations.

There are essentially three classical psychophysical methods for threshold determination: the method of constant stimuli, the method of adjustment, and the method of limits (Gescheider, 1985; Goldstein, 1989; and Green and Swets, 1988). Each method may be used to determine either absolute or difference thresholds. However, within the context of the discussions contained herein, only absolute threshold determinations will be considered.

Method of constant stimuli. In the method of constant stimuli, several (usually five to nine) levels of the stimulus are chosen such that about half of the levels will fall below the suspected threshold and about half of the levels will fall above the suspected threshold. The upper and lower levels are chosen such that the upper level will always be detected and the lower level will never be detected. Each of the stimulus levels are presented to the subject an equal number of times throughout the experiment in random order [and often at random times (Wilkins and Martin, 1985)]. Although Goldstein (1989) implies that as few as ten presentations of each stimulus level may be adequate, Gescheider (1985) states that as many as 100 presentations of each stimulus level may be necessary to generate reliable data. Data recorded during the experiment consist of the number of yes responses made for each stimulus presentation.

Once the data are obtained, the proportion of yes responses at each of the stimulus levels is plotted against stimulus intensity. The resulting psychometric function (illustrated in Figure 10) can then be used to determine the stimulus level which would produce a 50% positive response rate.

As originally envisioned, the method of constant stimuli required that a stimulus be presented during each trial interval. However, it was recognized that some of the positive responses at the lower stimulus levels were likely the result of guessing on the part of the subject. In an attempt to take this factor into account, a modified procedure was developed which used blank trials (where no stimulus was presented) in an effort to quantify the guessing rate of the subject (Green and Swets, 1988). The data resulting from such an experiment could then be corrected for guessing. The data were handled as follows: The positive response rates at each stimulus level obtained in the experiment could be decomposed into hit and false alarm rates. The hits would be the proportion of positive responses given a signal [P(Y|s)] and the false alarms would be the proportion of positive responses given no signal [P(Y|ns)]. However, the hit rate would include not only true positive responses [$P^*(Y|s)$], but also positive responses resulting from guesses. The false alarm rate would therefore be used to estimate the proportion of the hits due to

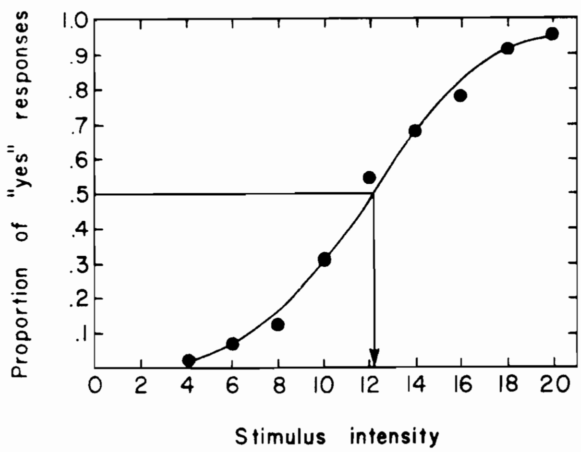


Figure 10. *Psychometric function obtained using method of constant stimuli*. (from Gescheider, 1985)

guessing. The true positive response rate would be calculated as (from Green and Swets, 1988, p. 129):

$$P^*(Y|s) = \frac{P(Y|s) - P(Y|ns)}{1 - P(Y|ns)}$$

Method of adjustment. The method of adjustment is probably the simplest of the classical psychophysical methods. In this procedure, the stimulus level is set initially either considerably above or below the suspected threshold. The stimulus intensity is then adjusted by the subject until it is just noticeable (if its initial level was below threshold) or until the sensation just disappears (if its initial level was above threshold) (Gescheider, 1985; Goldstein, 1989). The stimulus intensity is usually continuously variable. A large number of ascending *and* descending trials are usually performed, with the mean of the trial endpoints taken as the subject's threshold (Gescheider, 1985). The fact that the subject may adjust the stimulus intensity him/herself is considered by some as a motivational tool and a means of ensuring the subject pays attention to the task during the experiment (Goldstein, 1989).

Method of limits. The method of limits is similar to the method of adjustment in that the stimulus level initially presented to the subject may be either well above or well below the suspected threshold. The stimulus intensity is then adjusted by the experimenter in small discrete steps until the subject indicates that it is just perceptible or just imperceptible. At that point, the trial is terminated and a new (ascending or descending) trial initiated. The trial's "transition point" (Gescheider, 1985) is defined as the midpoint of the stimulus intensities for the last two responses of the trial. The subject's threshold is then taken as the mean of all the trial transition points (Gescheider, 1985; Goldstein, 1989).

Two types of constant error are possible with the method of limits. These are errors of habituation and errors of anticipation (Gescheider, 1985; Goldstein, 1989). An error of habituation occurs when the subject gets into the habit of making the same pattern of responses on successive trials. In this case the subject would continue to report that the stimulus was still present for several intervals after it passed the true transition point on a descending trial; or, for an ascending trial, the subject might report that the signal was not present for a few intervals after it did become noticeable. The result of such errors would be to artificially increase the threshold on ascending trials, while artificially decreasing the threshold on descending trials. When committing errors of anticipation, the subject would anticipate the appearance or disappearance of the signal and report that the signal was perceptible when in fact it was not, or that it had ceased to be perceptible when in fact it was still above threshold. In an effort to control for such errors, experiments are structured so that the initial stimulus levels vary from trial to trial, and excessively long trial sequences are avoided. Training and instruction are also used to reduce or eliminate these errors.

A modification of the method of limits often used in audiometry is Békésy tracking (the general procedure applicable to non-auditory stimuli is referred to simply as threshold tracking). This procedure combines features of both the method of limits and the method of adjustment. In Békésy tracking, the stimulus intensity is controlled by the subject. The initial level may be established either above or below the suspected threshold. As soon as the subject detects the stimulus, he/she presses a switch causing the stimulus intensity to decrease at a selected rate. When the stimulus is no longer perceptible, the subject releases the switch causing the stimulus intensity to increase. This procedure is repeated until the subject's responses become stable (Gescheider, 1985). This procedure also has its disadvantages in that the subject may develop a rhythm in his/her response, thus biasing the data.

***Comparing the classical psychophysical methods*.** The differences between the method of limits and the method of adjustment have to do with the manner in which the stimulus intensities are varied (in discrete steps or continuously) and with who (the experimenter or the subject) is responsible for the adjustment. In the method of adjustment, the stimulus is most often continuously variable and the adjustments are made by the subject. In the method of limits, on the other hand, the experimenter adjusts the stimulus in discrete steps (although the steps can be quite small). Both methods are susceptible to subject bias. In the method of adjustment, it is assumed that the fact that the subject is allowed to take an active part in the experiment helps control this bias to some degree and serves as a motivational tool. In both methods, consistent training and the use of only experienced, well-motivated subjects is another means for controlling for any possible bias on the part of the subject. However, unlike the method of constant stimuli, neither the method of limits nor the method of adjustment are amenable to the introduction of catch trials in order to estimate the subject's bias.

The method of constant stimuli differs from the other two methods in that the stimuli are presented at predetermined levels rather than being adjustable over some range of values. However, as in the method of limits, the subject's task is simply to state whether or not a stimulus is present in a given trial. Since a large number of trials are required to get an accurate estimate of the subject's threshold (Gescheider, 1985), this procedure is much more time consuming than either of the other two methodologies. It is also said to be the most accurate of the three classical psychophysical methods (Gescheider, 1985; Goldstein, 1989). As mentioned previously, the method of constant stimuli is the only method discussed which can provide any indication of the presence or absence of the subject's response bias.

Signal Detection Theory

The theory of signal detection (TSD), unlike the three classic psychophysical methods discussed above, makes no assumptions about the presence of a sensory threshold. Rather, signal detection theory assumes that all sensory events take place against a background of noise and that the ability of a subject to detect a signal depends not only on the relative strengths of the signal and the background noise, but also on the criterion established by the subject for indicating the presence or absence of a signal. This background noise may be internal to the observer (i.e., spontaneous neural activity) or external (as in masking noise) and is assumed to vary randomly. (In all of the discussions concerning signal detection theory, the terms *noise* and *signal-plus-noise* will be used repeatedly. The terms are not meant to refer to auditory noise, but may refer to any sensory input and the "noise" which masks it.)

In signal detection theory, the perceptual strength in response to some stimulus (be it noise or a signal imbedded in noise) is assumed to vary (usually normally) along a continuum. When only noise is present, the mean of the probability density function will be lower on the scale of response strength than when a signal is imbedded in the noise. This concept is illustrated by the three graphs in Figure 11 in which plots of the probability density functions for noise and signal-plus-noise are shown for three different signal strengths are shown. In the graphs, the abscissa represents the perceptual response strength internal to the observer and the ordinate represents the probability of occurrence of a given response strength given either noise alone (N) or a signal-plus-noise (SN) (Gescheider, 1985; Green and Swets, 1988). Since a signal is always added to the noise, the SN distribution will always be to the right of the N distribution (Gescheider, 1985). For very weak signals, Figure 11(a), the two distributions will exhibit a great deal of overlap and it will be extremely difficult for an observer to

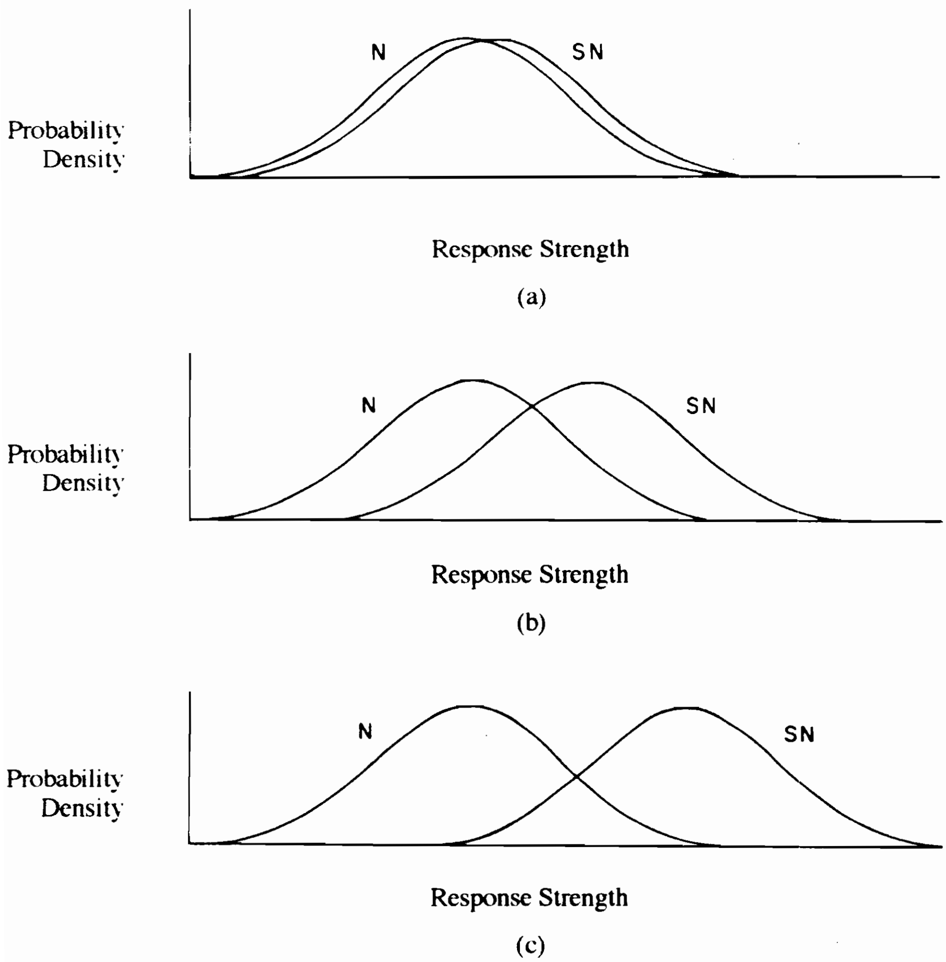


Figure 11. *Probability density functions for three signal strengths.*
(adapted from Gescheider, 1985)

correctly identify a signal. As the signal strength increases, Figure 11(b) and (c), the overlap decreases and the task of detecting the signal becomes much easier.

Unlike classical high-threshold theory, signal detection theory stipulates that the ability of a subject to detect a signal in noise depends not only on his/her sensitivity (the separation of the two probability distributions), but also on the criterion adopted by the subject for making a positive response. This criterion can be thought of as a response strength above which the subject will always respond affirmatively and below which the subject will always respond negatively, regardless of whether or not a signal was present. The concept of a criterion is illustrated in Figure 12. As in the previous figure, the plot in Figure 12 represents the probability density functions for conditions of noise alone (N) and also signal-plus-noise (SN). The difference between the means of the two distributions (d', in units of standard deviation) is a measure of the observer's sensitivity (Gescheider, 1985). Since the abscissa represents an internal response strength to a particular stimulus, the observer's sensitivity increases as the separation between the two distributions increases. The vertical line, C, represents one possible criterion which the subject may adopt in deciding how to respond. For occurrences of the stimulus which fall to the left of the criterion, the subject will respond that only noise is present, but for occurrences of the stimulus which fall to the right of the criterion, the subject will respond that a signal was also present.

Three possible criteria are illustrated in Figure 13. The criterion on the far left of the graph represents a lax or liberal criterion in that the subject will respond that a signal is present much more often than it is, resulting in many false alarms (responding yes when there is no signal). The criterion on the far right of the graph is considered to be a conservative criterion in that the subject will make few false alarms, but at the same time will miss many of the signals. The criterion in the middle is considered to be neutral in that about as many false alarms will be reported as signals are missed. Since the subject's

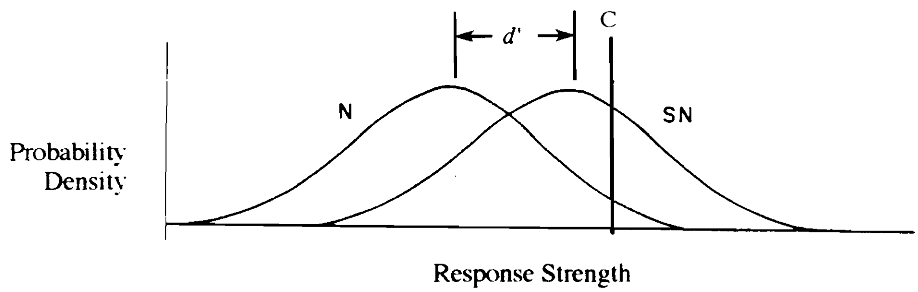


Figure 12. *Illustration of sensitivity and criterion.*
(adapted from Gescheider, 1985)

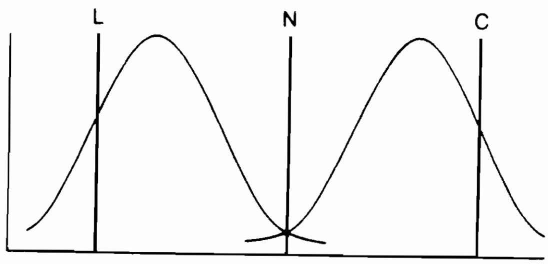


Figure 13. *Illustration of liberal (L), neutral (N), and conservative (C) criteria.*
(from Goldstein, 1989)

criterion (β) is taken to be the ratio of the ordinates of the signal-plus-noise and noise distributions (Gescheider, 1985), the neutral criterion illustrated in Figure 13 would have a β equal to 1.0.

Signal detection theory assumes that when a subject adopts a criterion for indicating that a signal is present, or only noise, that he/she does so based on the *a priori* probabilities of occurrence of noise [P(N)] and of the signal [P(SN)] as well as on the costs and benefits associated with wrong and right decisions. The assumed relationships between these factors is illustrated by the following equation.

$$\beta \propto \frac{P(N)}{P(SN)} \times \frac{V_{(\text{correct rejection})} - K_{(\text{false alarm})}}{V_{(\text{hit})} - K_{(\text{miss})}}$$

where: $V_{(\text{correct rejection})}$ is the value associated with a correct rejection,

$K_{(\text{false alarm})}$ is the cost associated with a false alarm,

$V_{(\text{hit})}$ is the value associated with a hit, and

$K_{(\text{miss})}$ is the cost associated with a miss.

If the subject were to behave as an ideal observer, the relationship would become an equality. Implicit in this relationship is the assumption that the subject has all of the information available to him/her (costs, values, and probabilities). If the costs and values are held neutral, then only the *a priori* probabilities are involved in establishing a criterion. The most common means of manipulating subjects' criteria in the yes/no paradigm is through the manipulation of the signal and noise probabilities.

In addition to manipulating the *a priori* probabilities of the signal and noise trials, it is also assumed to be possible to force the subject to adopt multiple criteria during the course of an experiment by simply asking that he/she do so. This is the principal assumption underlying the rating procedure. Finally, in the application of the forced-choice procedure, it is assumed to be possible to force the subject to adopt a neutral criterion. (The various experimental methodologies used in signal detection research will be described in a later section.) Implicit in all of the TSD experimental procedures is the assumption that subjects will be able to maintain constant criteria throughout the experiment.

In a signal detection experiment, there are four possible outcomes for a given stimulus event: 1) the subject responds that a signal was present when there was indeed a signal present (a hit), 2) the subject responds that a signal was present when there was not a signal present (a false alarm), 3) the subject responds that no signal was present when there was a signal present (a miss), or 4) the subject responds that there was no signal present when there was indeed no signal present (a correct rejection). With each of these four possible responses, there is an associated conditional probability, illustrated in Figure 14. For a given set of conditions and *a priori* probabilities of occurrence of noise and signal-plus-noise, only two of these conditional probabilities [$P(Y|SN)$ and $P(Y|N)$, or $P(N|SN)$ and $P(N|N)$] are necessary to completely describe a given situation. By convention, the probabilities used in signal detection theory are $P(Y|SN)$ and $P(Y|N)$, the hit rate and false alarm rate respectively (Gescheider, 1985).

Once the hit and false alarm rates are known, it is a simple matter to calculate the sensitivity measure, d', and the criterion, β, in the following manner. To calculate the subject's sensitivity, it is necessary to first convert $1-P(Y|N)$ (one minus the false alarm rate) and $1-P(Y|SN)$ (one minus the hit rate) to z scores (Z_N and Z_{SN}, respectively). Once this has been accomplished, the sensitivity measure can be calculated as:

$$d' = Z_N - Z_{SN}$$

The criterion may be estimated by taking the ratio of the ordinate values corresponding to Z_N and Z_{SN}.

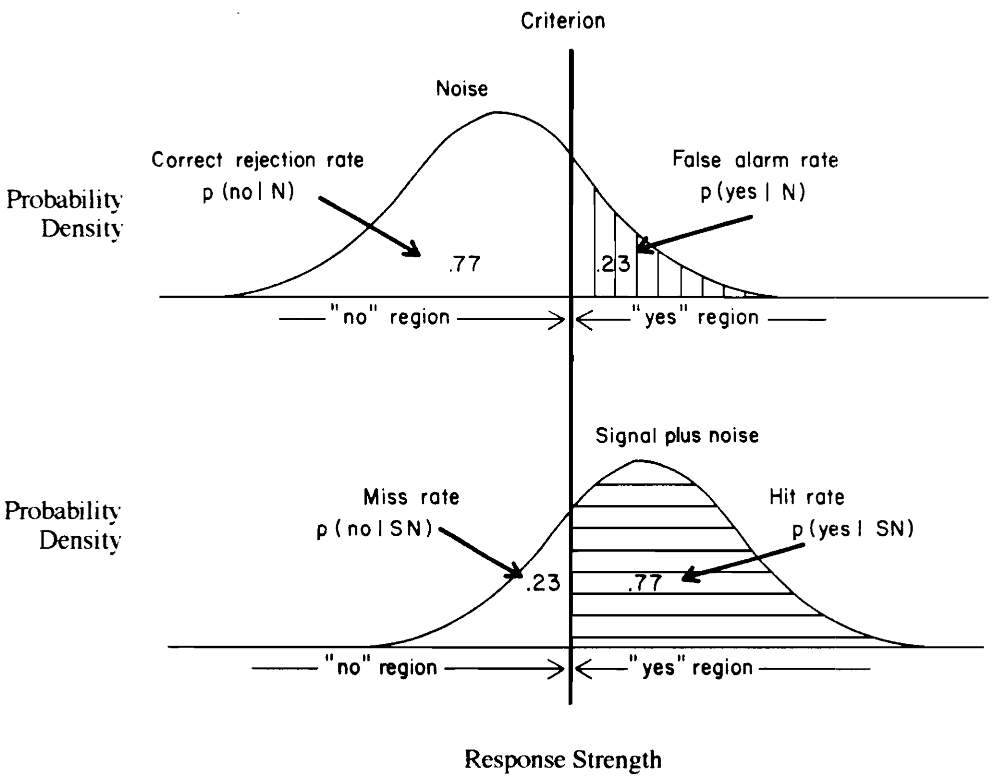


Figure 14. *Conditional probabilities*.
(from Gescheider, 1985)

The above discussion concerning the calculation of d' and β assumed data from only one experimental session for which hit and false alarm rates were calculated. This is legitimate as long as *none* of the basic underlying assumptions of signal detection theory are violated. Namely, that the noise and signal-plus-noise distributions are normally distributed and of equal variance. If either of these assumptions are violated, then the procedures outlined above for calculating d' and β are not valid. However, even when one or both assumptions are violated, procedures are available for calculating alternative, but comparable sensitivity and criterion measures. Usually, a signal detection experiment is planned such that the subjects' criteria are manipulated so as to generate several data points and the assumptions of normality and equal variance are then tested to determine which measures are most appropriate.

If the hits are plotted against the false alarms for all possible criteria for given distributions of noise and signal-plus-noise, a curve similar to that shown in Figure 15(a) would be obtained. This curve is called the Receiver Operating Characteristic (ROC) curve (Gescheider, 1985). Each point on the curve represents the hit and false alarm rates which would result from the adoption of different criteria for a given pair of noise and signal-plus-noise distributions. If the noise and signal-plus-noise distributions were close together (i.e., exhibited considerable overlap), the ROC curve would be flatter, approaching the positive diagonal shown in the figure. If the two distributions were relatively far apart (i.e., exhibited little overlap), the apex of the ROC curve would approach the upper left corner of the figure.

If the hit and false alarm rates were normalized by converting them to z scores, the ROC curve would become a straight line, parallel to the positive diagonal as shown in Figure 15(b). (The normalized ROC curve will be a straight line parallel to the positive diagonal only if the noise and signal-plus-noise distributions are both normal and of equal

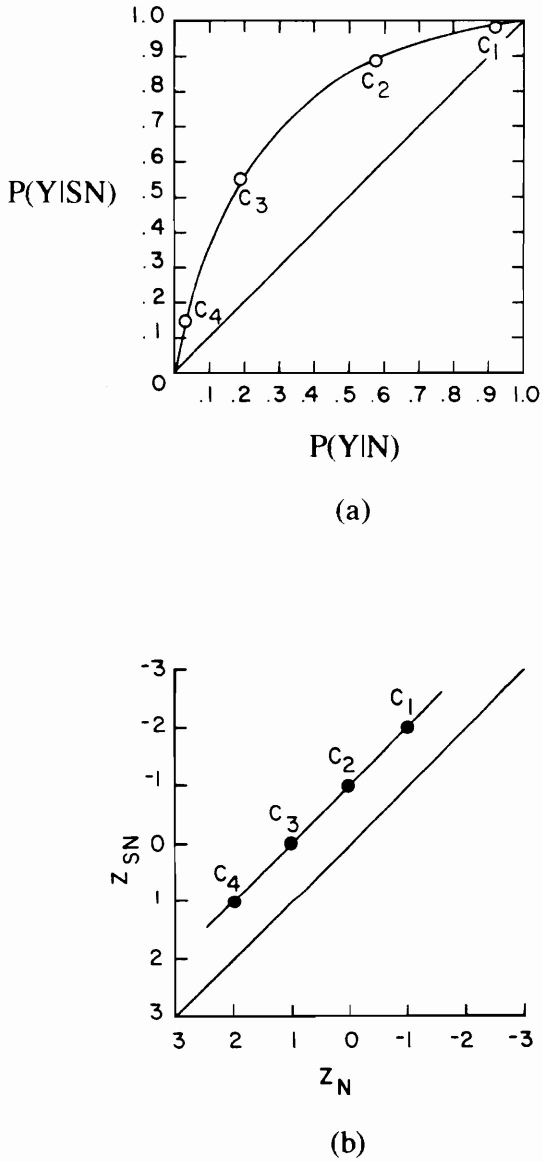


Figure 15. *Receiver operating characteristic (ROC) curves*.
(from Gescheider, 1985)

variance. In cases where the distributions are not normal or are of unequal variance, the normalized ROC curve will differ from that shown in the figure. These points will be addressed more fully later.) The sensitivity measure, d', is the difference between Z_N and Z_{SN} at any point along the normalized ROC curve [approximately 1.0 in the graph shown in Figure 15(b)].

Experimental procedures. Several methods are available to gather the data necessary to generate an ROC curve and/or determine the sensitivity measure of interest. The three more common procedures are described below. With two of the procedures (the yes/no and rating procedures), it is necessary to force the subject to adopt several criteria in order to generate multiple points along an ROC curve. The third procedure (forced-choice) does not require generation of an ROC curve but does not allow the examination of an observer's response criterion (Green and Swets, 1988).

Yes/no procedure. Perhaps the simplest experimental method to implement is the yes/no paradigm. In this procedure, the subject is presented with a series of stimulus intervals and asked after each presentation to indicate whether the interval contained a signal in addition to noise (yes) or only noise (no). To generate multiple points on the ROC curve, the subject's criterion is most often manipulated by varying the *a priori* probability of the signal. However, the subject's criterion may also be manipulated by varying the costs and payoffs associated with false alarms and hits respectively, while maintaining a constant signal probability. A single trial session may contain anywhere from a few hundred to more than a thousand individual stimulus intervals while an entire experiment may contain from three to nine sessions (for each subject in the experiment) (Green and Swets, 1988). Obviously, this procedure can be very time consuming.

Rating procedure. Since each experimental session in the yes/no paradigm described above generates only a single point on an ROC curve, the procedure requires a subject to attend multiple experimental sessions in order to generate a single ROC curve (and thus obtain a single sensitivity estimate). The rating procedure is a more economical alternative in that an entire ROC curve is generated in a single experimental session. In the rating procedure, the observer is asked not only to determine whether or not a signal was present during the observation interval, but also to state how confident he/she was of his/her response by choosing one of several alternative responses. For example, if five alternatives are used, they might be: 1) absolutely sure a signal was presented, 2) fairly sure a signal was presented, 3) not sure a signal was presented, 4) fairly sure a signal was not presented, and 5) absolutely sure a signal was not presented (Gescheider, 1985). Giving an observer n alternatives forces the observer to adopt a $n - 1$ criteria, illustrated in Figure 16 for the case of five alternatives. To construct an ROC curve, the resulting hit and false alarm rates for each of the criteria are calculated as if they were obtained using the yes/no paradigm described earlier. Therefore, there are a total of $n - 1$ points obtained on the ROC curve given n alternatives.

The *a priori* probability of a signal is usually set at 0.50 and remains constant throughout the experiment. Although this procedure requires much less data to generate multiple data points than does the yes/no procedure, it not only assumes that the subjects will be able to establish multiple distinct criteria, but that they will be able to hold these criteria constant throughout the experiment.

Forced-choice procedure. Another common paradigm is the forced-choice procedure . In this method, the subject is presented with multiple (usually two) test intervals and asked to identify which interval contained the signal. This procedure is sometimes extended to include more than two intervals or to require the observer to identify which of several signals was present (Green and Swets, 1988). Although it is possible to vary the *a priori* probabilities of occurrence of the signal between the two intervals and also vary the cost and benefits associated with a false alarm or correct identification and thus generate an ROC curve, the usual implementation requires only a

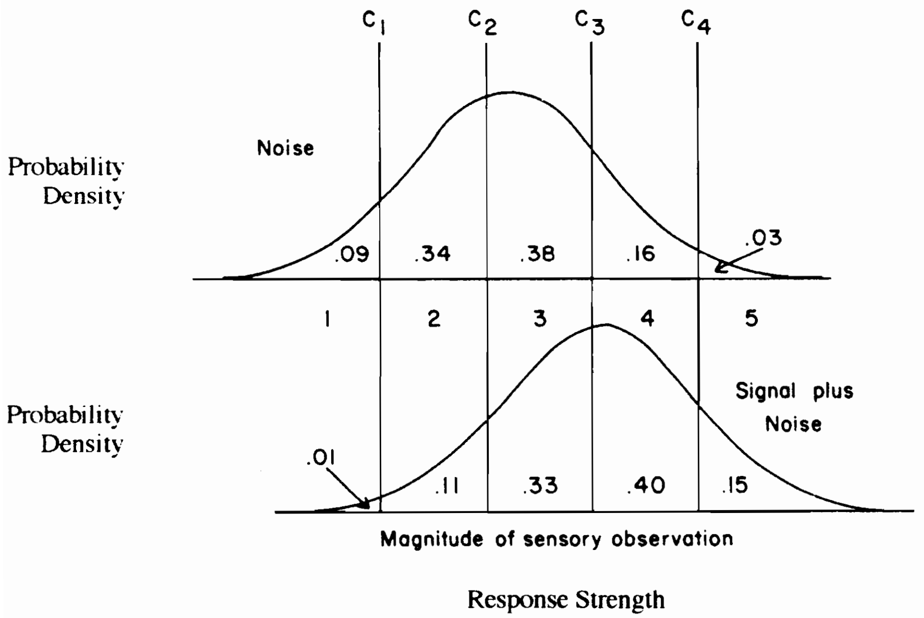


Figure 16. *Multiple criteria resulting from the use of the rating procedure.* (from Gescheider, 1985)

single experimental session to generate a measure of the observer's sensitivity (Green and Swets, 1988), making this procedure just as economical as the rating procedure. However, this economy does not come without cost, since it is not possible to determine a subject's criterion using this procedure. For this reason, Green and Swets (1988) recommend that this method be used only when purely sensory processes are being studied and the observer's motivation and/or response processes are of no concern.

Choice of method. Green and Swets (1988) state that for given signal and noise strengths and for a given observer, the three methods described above produce consistent results. Therefore, any of the three methodologies should be satisfactory. The major advantage of the rating procedure is one of economy. Its use can reduce the size of a signal detection experiment by a factor of five or six. The forced-choice method can be easily extended to test recognition as well as detection and is as economical as the rating procedure when questions of response bias are unimportant (Green and Swets, 1988). The choice of method should therefore be based on which method is most compatible with the intent of the experiment.

Data treatment. Once the data have been collected (regardless of the methodology used) it becomes necessary to determine the observers' sensitivities. If the yes/no or rating procedure were used, this involves the generation of ROC curves, testing the assumptions of normality and equal variance, and calculation of the appropriate sensitivity measure. In addition, it may also be desirable to determine the criteria adopted by the observers and use it as a dependent variable in addition to the sensitivity measure. (The process is streamlined considerably if the forced-choice procedure is used, as will be discussed shortly.)

The first step in data reduction after the hit and false alarm rates have been calculated is to plot the normalized data as shown in Figure 17. With the data plotted in this manner, it is easy to test the assumptions of normality and equal variance. The

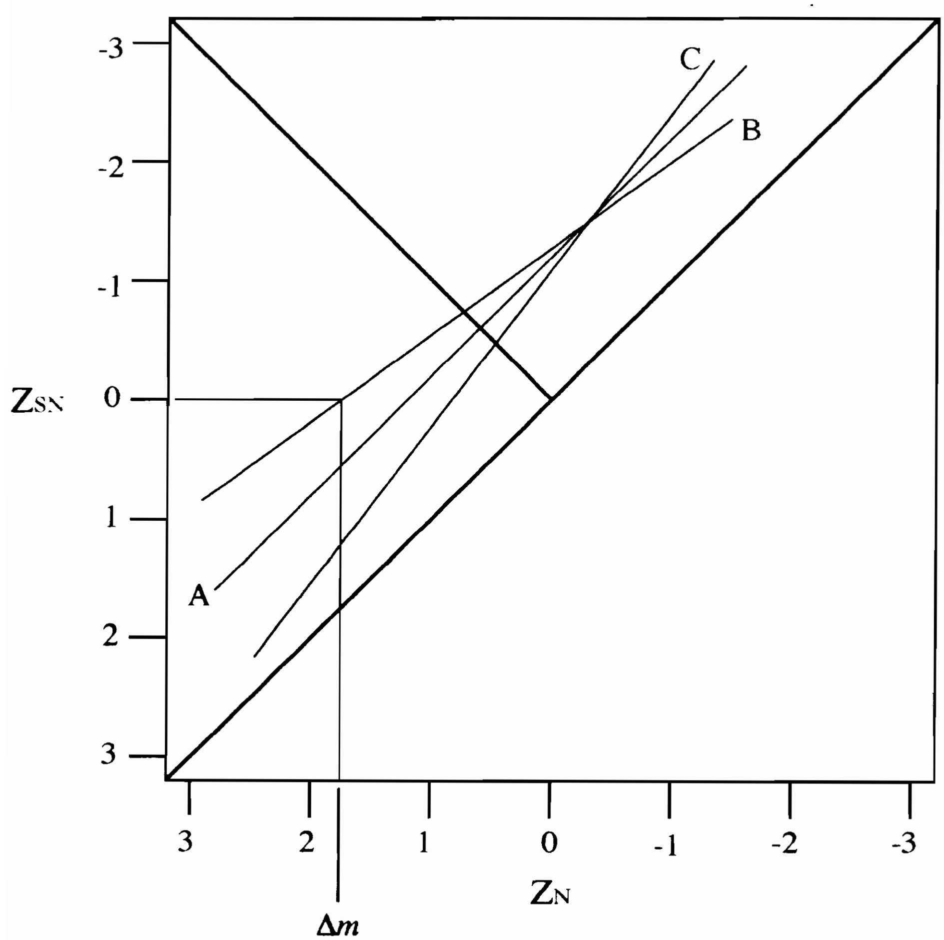


Figure 17. *Normalized ROC curve*. (Labels explained in text.)

equation of the best-fitting line through the normalized data is then determined. Although least squares estimates are often used when determining the equation of the line of the normalized ROC curve (Gescheider, 1985), several authors (Ogilvie and Creelman, 1968; Swets, 1986) recommend using maximum-likelihood estimates, for several reasons: both values (hits and false alarms) are dependent variables and are both subject to error so the least squares method is inappropriate since it minimizes error in only one direction, and the least squares method does not take into account the correlation between data points when the rating procedure is used.

If the data are linear, then the assumption of normality is valid. If the slope of the line is equal to 1.0, then the noise and signal-plus-noise distributions are of equal variance (line "A" in Figure 17). If, on the other hand, the slope of the line deviates from 1.0, the variance of the two distributions are not equal. If the variance of the signal-plus-noise distribution is greater than that of the noise distribution, the slope of the normalized ROC curve will be less than 1.0 (line "B" in Figure 17). If the slope of the normalized ROC curve is greater than 1.0 (line "C" in Figure 17), the noise distribution has greater variance than the signal-plus-noise distribution. For auditory stimuli, experimental data tend to support the assumption of normality, while at the same time indicating that the signal-plus-noise distribution tends to have slightly larger variance than the noise distribution (Egan, 1975; Green and Swets, 1988), resulting in normalized ROC curves with slopes less than 1.0 (line "B" in Figure 17).

Dependent measures. Once the normalized data have been plotted, the appropriate sensitivity measure can be calculated. Several such measures are in common use. If both the assumptions of normality and equal variance are shown to be correct, then perhaps the most common sensitivity measure is d'. This measure is obtained from the normalized plot by subtracting the Z_{SN} from Z_N at any point along the normalized ROC curve. For the example shown in Figure 17, $d' \approx 1.2$. It is also possible to calculate a value for d' when the noise and signal-plus-noise distributions are normal but of unequal variance. To do so, Z_{SN} is multiplied by the reciprocal of the slope of the normalized ROC curve and the product subtracted from Z_N. An alternative sensitivity measure for the case where the distributions are of unequal variance is Δm. This measure is obtained from the normal-normal plot by determining the value of Z_N when $Z_{SN} = 0$, as shown in Figure 17. Yet another alternative sensitivity measure (d_e') is the absolute difference between Z_N and Z_{SN} where the normalized ROC curve crosses the negative diagonal. Each of the measures discussed above relate directly to the differences between the means of the noise and signal-plus-noise distributions. Numerical differences between them are due to the manner in which each weights the variances of the two distributions when they are unequal. If the variances of the two distributions are equal, then each of the measures would be numerically equal to one another.

The measures described above suffer from two major weaknesses. First, each measure requires that the noise and signal-plus-noise distributions be normal. Although this assumption has often been shown to be valid for simple auditory stimuli (Egan, 1975; Green and Swets, 1988), it may not always be the case. Second, each of the measures are expressed in units of standard deviation and are thus difficult to relate to real world phenomena. If it is desired only to determine if two experimental treatments differ, then the above measures are quite adequate. However, if it is necessary to relate the experimental results to a real word situation, then a measure capable of being interpreted in the desired context must be found.

Two measures which do not rely on the normal distribution assumption and which may be interpreted in more meaningful terms than those described above are P(C) and P(A). The sensitivity measure P(C) is the proportion of correct responses when the two-interval forced-choice paradigm is used. When this procedure is implemented, it is not necessary to normalize the data, test the assumptions of normality and equal variance, or plot the data, but only to calculate the proportion of correct responses. As mentioned earlier, this makes data reduction considerably easier than when the yes/no or rating procedure is implemented. Not only does this measure have a specific and direct meaning, but it also does not require that the underlying distributions be normal or of equal variance. However, the maximum possible P(C) will be obtained when the underlying distributions are indeed normal and of equal variance (Robinson and Watson, 1972).

The other nonparametric measure, P(A) – often referred to as *accuracy* (Swets, 1988b), is the proportion of the area under the ROC curve [when plotted as in Figure 15 (a)] and is applicable when the yes/no and rating procedures are used. This measure is recommended by Swets (1986, 1988b) as a means of comparing results from many different studies which may or may not have used similar methodologies. Although this measure does not have a direct physical correspondence to a real world measure as does P(C), it may range in value from 0.5, representing chance performance when the ROC curve falls on the positive diagonal of the plot, to 1.0, representing perfect performance (Swets, 1988b). However, a value of P(A) obtained using the yes/no or rating procedure would be numerically equal to P(C) if the two-interval forced-choice procedure had been used (Egan, 1975; Robinson and Watson, 1972; Swets, 1986, 1988b; Swets, Pickett, Whitehead, Getty, Schnur, Swets, and Freeman, 1979). Therefore, Swets (1988b) suggests one interpretation of P(A) is that if a system (i.e., a person attempting to detect an alarm or warning signal in a background of noise) known to perform with an accuracy of 0.80 were presented with two stimulus intervals, one of which contained a signal, the correct interval would be chosen 80% of the time. In other words, P(A) appears to be a prediction of a system's performance in a two-interval forced-choice task obtained by way of the rating or yes/no procedure.

It is also possible to use the observed β as a dependent variable in a separate analysis and compare the results to those obtained using a sensitivity measure. The purpose for doing so would be to determine if observed differences between groups or experimental conditions were due to different sensitivities, the adoption of different criteria, or both. Just such an analytical technique was used by Williges (1969) when investigating the "vigilance decrement" (defined as a reduction in an operator's detection performance over time as measured by the hit rate – Wickens, 1984) associated with a visual monitoring task. His results indicated that the reduced detection performance over time was due to a shift in criterion rather than a change in sensitivity.

Comparison of High Threshold Theory and Signal Detection Theory

When comparing methodologies associated with high threshold theory to signal detection theory, it is necessary consider the differences in the assumptions underlying the two theories. The primary difference between the two theories is that threshold theory explicitly assumes that some clearly definable energy barrier exists such that a stimulus must contain enough energy to exceed this barrier before it can be detected by an observer. And, although this barrier does vary randomly over time, the presence of a stimulus will be reported infrequently in the absence of a signal (Green and Swets, 1988). Also, implicit in the controls built into the psychophysical procedures is the assumption that any bias on the part of the subject can be either eliminated or controlled to the point that it will have negligible effects on the results of the experiment.

Signal detection theory, on the other hand, makes no assumption of an energy threshold. Rather, it hypothesizes that the detection of a signal will depend on the strength of the signal relative to the background noise against which it is presented and upon the criterion established by the subject for deciding whether or not the internal sensory experience was sufficient to allow a positive response. In other words, signal detection theory accepts and makes allowances for a variable criterion on the part of the subject. Signal detection theory also places no limits on the possibility for false alarms as does high threshold theory.

In the forced-choice methodology, signal detection theory does attempt to force the subject to adopt a specific (neutral) criterion. This could be thought of as an effort similar to the attempts made by researchers using the classical methodologies to control (or alternatively limit or eliminate) their subjects' bias. The yes/no and rating procedures, on the other hand, although they do attempt to force the subject to adopt multiple criteria, make no effort to control how liberal or conservative the subject sets his/her criteria.

The methodologies associated with each of the two theories also produce vastly different metrics. Classical high threshold theory produces a measure of threshold which is readily interpretable. In audition, for example, the threshold of hearing at a given frequency is expressed in dB, which can be interpreted as the sound pressure level at that frequency which must be exceeded before a signal will be audible. Signal detection theory produces no such measure. Instead, the measure associated with signal detection theory is *sensitivity*, usually expressed in units of standard deviation (the difference between the means of two normal distributions). The criterion measure, likewise, is also difficult to interpret. It is a dimensionless ratio of the ordinates of two normal distributions. If greater than 1.0, the subject is said to be responding conservatively, if less than 1.0, the subject is said to be lax, and if β equals 1.0, the subject is said to be responding in a neutral fashion (Gescheider, 1985; Goldstein, 1989).

Only when the forced-choice procedure is used does signal detection theory produce a measure similar to that produced by any of the classical methodologies. In the classical method of constant stimuli, several stimulus intensities are presented and the proportion of correct positive responses at each of the stimulus levels is recorded. The threshold is then assumed to be that stimulus level which would produce a 50% correct response rate. A similar pattern of responses may be produced by the use of the forced-choice procedure of signal detection. In such a case, an experiment would be conducted using several experimental sessions where the signal level would be varied session to session. The resulting measures (proportion of correct responses) would then be plotted against the various signal levels to produce a psychometric function. The "threshold" could then be considered to be the signal level that produced a P(C) of 75% (Gescheider, 1985). [The 75% value of P(C) is used because 50% would represent chance performance, whereas 75% lies halfway between chance and perfect performance (Green and Swets, 1988).] This is indeed what is often done when signal detection experiments are conducted using auditory stimuli (Abel, Kunov, Pichora-Fuller, and Alberti, 1983; Patterson, 1974, 1976; Patterson, Nimmo-Smith, Weber, and Milroy, 1982; Watson, Franks, and Hood, 1972; Weber, 1977). Green and Swets (1988) suggest that when purely sensory phenomena are being considered, the forced-choice procedure is probably the best procedure to use, due primarily to its economy. However, an equally good argument could also be made based on the interpretability of the resulting data. This, in fact, may be why it is used so often.

Despite the fact that the measures resulting from the application of signal detection theory do not easily lend themselves to physical interpretation, they are ideally suited for investigating whether differences between two groups or treatments are due to differences in sensitivities or to differing criteria. An example of one such experiment was the "vigilance decrement" experiment conducted by Williges (1969) mentioned earlier. Another such experiment was conducted by Moskowitz and McGlothlin (1974) investigating the effects of marihuana on an auditory detection task. Their findings indicated that the reduced performance in the experimental task was due to a change in sensitivity rather than a change in criterion. Such research might not have been possible if a methodology capable of distinguishing between a subject's sensitivity and criterion, such as signal detection theory, were not available.

Testing the Assumptions

One of the most damaging arguments to classical high threshold theory is that different thresholds are obtained by manipulation of non-sensory factors in an experiment. Gescheider (1985) describes the re-examination of data from an earlier study (Gescheider, Wright, Weber, and Barton, 1971) which measured vibrotactile threshold obtained for two levels of the probability of signal occurrence [P(s) = 0.30 and 0.70] using the same subjects and holding all other factors in the experiment constant. The results showed the threshold determined with P(s) = 0.70 to be much lower than the threshold determined using P(s) = 0.30 (the difference being approximately 0.5 microns vibration amplitude). Even after correcting for bias using the relationship:

$$P^*(Y|s) = \frac{P(Y|s) - P(Y|ns)}{1 - P(Y|ns)}$$

where: $P^*(Y|s)$ is the corrected hit rate,

$P(Y|s)$ is the measured hit rate, and

$P(Y|ns)$ is the measured false alarm rate,

the two psychometric functions still did not fall on top of one another as predicted by classical high threshold theory (although the thresholds for the two levels of P(s) were much closer together after the correction, differing by only about 0.2 microns). Swets, Tanner, and Birdsall (1961) report similar findings using visual stimuli. After correcting experimentally obtained psychometric functions for false alarms, the resulting corrected psychometric functions did not overlap as predicted by high threshold theory.

Another argument may be made against classical high threshold theory by comparing experimental data for hits and false alarm rates to the receiver operating characteristic (ROC) curves (again, simply a plot of hit rate vs. false alarm rate) predicted by the two theories. Classical high threshold theory predicts linear ROC curves following the equation given below.

$$P(Y|s) = P^*(Y|s) + \{P(Y|ns)[1 - P^*(Y|s)]\}$$

This equation (a rearrangement of the earlier equation) is simply that of a straight line with a y-intercept equal to the true hit rate [$P^*(Y|s)$] and a slope equal to $1 - P^*(Y|s)$. Signal detection theory, on the other hand, predicts a curvilinear ROC curve. When experimental data are plotted, the data fall along a curvilinear path as predicted by signal detection theory. These points are illustrated in Figure 18, in which data from a yes/no experiment are plotted as are typical ROC curves predicted by threshold theory and signal detection theory (from Swets, 1988a).

It would appear therefore that classical high threshold theory fails to account fully for all aspects of sensory detection. Signal detection theory, on the other hand, does seem better able to explain the data. This does not mean, however, that TSD represents a *true* model of the human sensory process, only that it seems to work. Other threshold theories (low threshold theory and neural quantum theory) have been developed whose predictions seem to fit existing empirical data just as well as signal detection theory (Green and Swets, 1988). For example, the results of the vibrotactile experiment described earlier, when interpreted in terms of low threshold theory, produce estimates of vibrotactile threshold which are independent of the probability of signal occurrence (Gescheider, 1985).

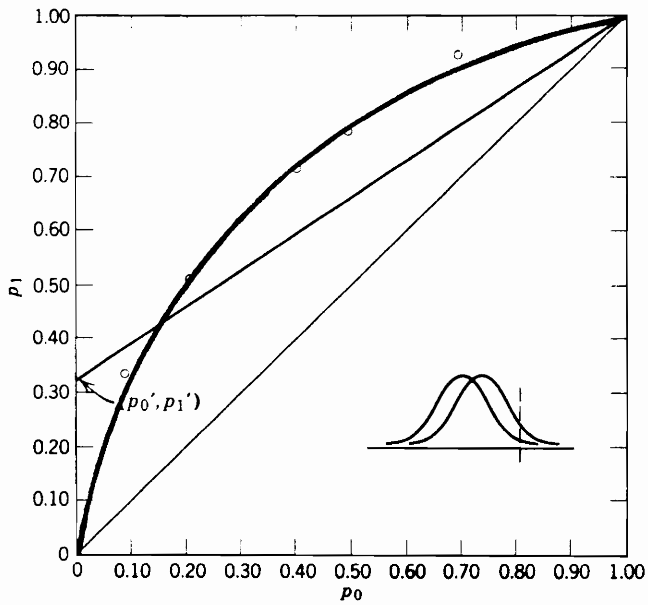


Figure 18. *Comparison of ROC curves for threshold theory and signal detection theory*. (adapted from Swets, 1988a)

Although the predictions of signal detection theory appear to closely match the results of empirical studies, the underlying assumptions (noise and signal-plus-noise distributions are normal and of equal variance) of signal detection theory can be tested rather easily simply by normalizing the ROC curve. If the normalized data are linear, then the assumption that the two distributions are normal should be valid. If the slope of the line through the normalized data is 1.0, then the two distributions have equal variance. Much of the existing data, although appearing to have underlying normal distributions, exhibit slopes of less than 1.0 when plotted as a normalized ROC curve, indicating that the signal-plus-noise distributions possess a greater variance than the noise distribution. Therefore, the theory has been modified somewhat to allow an increase in variance as the internal response strength increases [i.e., the variance of the distribution is proportional to its mean (Swets, Tanner, and Birdsall, 1961)] and alternative measures of sensitivity have been developed to take this factor into account.

Signal detection theory assumes that subjects will adopt criteria consistent with the proportionality relationship given earlier, that they can adopt multiple criteria in a single session, and that they can hold their criteria constant during the course of a single experimental session. In order to evaluate these assumptions, it is necessary to show that subjects can be forced to adopt different criteria while their sensitivity remains constant. Swets, Tanner, and Birdsall (1961) conducted a pair of visual detection experiments aimed, in part, at testing the assumptions underlying the criteria measures. In their first experiment, the yes/no procedure was used and the subjects' criteria were manipulated by varying the *a priori* probabilities of signal and noise. When the data were analyzed, it was found that the subjects did indeed adopt different criteria which varied in accordance with the proportionality equation presented earlier. Their second experiment utilized the rating procedure. They found that their subjects were able to adopt and maintain several distinct criteria during the course of the experiment.

Egan, Schulman, and Greenberg (1959) conducted a similar set of signal detection experiments using auditory stimuli. However, rather than manipulating the *a priori* probabilities to force their subjects to adopt different criteria when using the yes/no procedure, they asked them to adopt "strict," "medium," and "lax" (Egan, Schulman, and Greenberg, 1959, p. 770) criteria and then practiced them at each of these criterion levels. They also conducted the same experiment using the rating procedure, again asking the subjects to respond in the same manner. Their results indicated that not only did their subjects adopt multiple criteria in the rating procedure and change their criteria in the yes/no procedure, but the results of the two experiments were nearly identical in terms of the sensitivity measures obtained.

Williges (1969, 1971, 1973) has investigated how subjects establish criteria in vigilance tasks. His results indicate that if subjects know the *a priori* probabilities associated with the noise and signal-plus-noise trials, then they do adopt criteria in accordance with the equation presented earlier . However, if the subject does not have the necessary or correct information, he/she will not behave as predicted. Finally, at least in the experimental conditions tested, it was found that manipulation of the signal/noise probabilities had a much greater impact on the subjects' criteria than did manipulating the values and costs associated with correct and incorrect detections (Williges, 1971).

It appears as if classical high threshold theory does not adequately explain much of the sensory data currently available as far as the predicted independence of experimentally determined thresholds from non-sensory experimentally manipulable-conditions. It does appear, however, that the theory of signal detection adequately fits available experimental data, and the assumptions dealing with the underlying normal distributions of noise and signal-plus-noise are supported. In addition, it appears that signal detection paradigms do allow for the separation of the effects of sensitivity from the effects associated with the specific criterion adopted by the subject. However, it may be that the proper experiment has just not been conducted that would disprove these assumptions. After all, it is not possible to prove the null hypothesis, only to fail to reject it. Long and Waag (1981) caution against the wholesale acceptance of signal detection theory and cite examples of three small scale studies in which the theory did not hold up. However, in each case, one or more of the requirements for application of the theory was violated (i.e., there were too few trials, the subjects were not adequately practiced, the subjects did not know the *a priori* probabilities associated with the signal and noise, and the difference between the means of the noise and signal-plus-noise was excessively large).

It may also be that the theory only holds for certain types of stimuli. It has been shown that when applied to auditory tasks, the theory holds up remarkably well (Egan Schulman, and Greenberg, 1959; Green and Swets, 1988). However, some inconclusive results can be obtained when visual or vigilance tasks are considered (Swets, Tanner, and Birdsall, 1961; Long and Waag, 1981, Wickens, 1984) and its use in pain research is discouraged (Rollman, 1977). However, the theory has been successfully applied to such diverse situations as weather forecasting, eyewitness testimony, and medical diagnosis, to name just a few (Gescheider, 1984; Swets, 1988b; Wickens, 1984).

There seems to be two advantages to having a means of separating sensitivity from criterion. First, it allows an estimation of sensitivity which is uncontaminated by the subject's bias (conscious or unconscious). Secondly, in many situations, it allows researchers to determine if a difference between two groups and/or treatments is due to a difference in sensitivity, or to a difference in criteria. This was just the question asked by Williges (1969) when he investigated the vigilance decrement. His results indicated that the decrease in detections over time was due to the subjects adopting more conservative criteria (higher β), thus resulting in fewer signals being reported (fewer hits).

However, when the forced-choice procedure of signal detection theory is used in an experiment, it is not possible to obtain an estimate of β since the subject is assumed to have been using a neutral criterion. Data resulting from the application of this procedure are often handled in much the same fashion as data resulting from the classical method of constant stimuli. One study (Watson, Franks, and Hood, 1972) was found in which data obtained using a two-interval forced-choice procedure were compared to previous data obtained using classical psychophysical procedures. In this study, normal-hearing subjects were tested to determine their absolute "thresholds" for pure-tone stimuli in the absence of any intentionally generated background noise. The results were then compared to the existing ISO standard for audiometric zero. The researchers found that the stimulus levels required for 76% proportion of correct responses [P(C)] was in close agreement with the ISO levels for audiometric zero at frequencies below 4000 Hz. This finding leads to the question of just how big a difference exists between the two theories when purely sensory process are considered. What is needed is a series of experiments in which data obtained using both procedures are obtained for various sensory stimuli (i.e., auditory, visual, tactile, etc.) and subjected to statistical analysis.

However, when using TSD methodology in a psychophysical experiment, care must be taken to avoid violating the procedural requirements of the theory (i.e., sufficient trials, subjects possess knowledge of *a priori* probabilities, etc.). If these procedural requirements are violated, the results may not be reliable.

SIGNAL DETECTION LITERATURE RELEVANT TO THE RESEARCH EFFORT

Very little research relating directly to the study described herein has been conducted. However, several of the studies mentioned in the previous literature review did use TSD procedures. These studies include the detection experiment conducted by Abel et al. (1983a, 1983b); the masked threshold prediction technique developed by Fidell et al. (1974); and the critical band research conducted by Patterson (1974, 1976), Patterson et al. (1982), and Weber (1977). Each of these studies will be discussed in terms of the TSD principles and procedures used.

The only study performed relating directly to the current research topic which utilized TSD procedures was that of Abel et al. (1983a, 1983b) described earlier. In their study, the experimental task involved the detection of third-octave bands of noise against broadband background noises by both normal and hearing-impaired listeners in both the occluded and unoccluded states. The TSD procedure used was the two-interval forced-choice paradigm with the *a priori* probability of occurrence of the signal being 0.5 for each interval. The dependent measure used was P(C). The signal levels were varied such that the obtained values of P(C) would range from 0.5 (chance performance) to 1.0 (perfect performance). For each data point, a total of 150 trials were conducted (three blocks of 50 trials each). Psychometric functions were developed with P(C) = 0.8 defined as "threshold." [The authors did not indicate why a value of P(C) = 0.8 was used rather than P(C) = 0.75.]

Although the results of the study seem reasonable (discussed in detail earlier), there were several problems with the manner in which the study was implemented. First, only 150 trials were run with each subject in each experimental condition. Conventional wisdom dictates that several hundred trials [Green and Swets (1988) recommend a minimum of 500 trials, while Robinson and Watson (1972) recommend between 500 and 1000 trials] be conducted for each experimental condition. Although compromises are often necessary, no extenuating circumstances were mentioned by the study's authors. Secondly, no mention was made of any attempt to train the subjects in the experimental task prior to their first experimental session. In addition, no mention was made of any statistical tests used to determine if detection performance changed across the three trial blocks comprising each experimental session. Robinson and Watson (1972) state that the data from the first trial are routinely discarded by many researchers due to such learning or practice effects. Finally, no mention was made of any attempt on the part of the researchers to correct the data for any possible response bias favoring one interval over the other as suggested by Gescheider (1985).

The Abel et al. (1983a, 1983b) study was the only study found relating directly to any of the independent variables in the present research topic. The lack of research relating to the use of HPDs is not surprising since HPD test procedures are very rigidly standardized (ANSI, 1974; ANSI 1984; ISO, 1990). Most researchers in this field use these standardized test protocols as a matter of course, regardless of whether or not another procedure may actually be better in a given situation.

It is surprising, however, that no studies utilizing TSD procedures were found that dealt with hearing-impaired listeners. It would seem that TSD is particularly well suited to investigating hearing impairment, particularly with regard to how the level and type of impairment may affect an observer's criterion. The study discussed earlier, conducted at the San Diego County Fair (Webster, Himes, and Lichtenstein, 1950; Webster, Lichtenstein, and Gales, 1950), did find a slight difference in thresholds for individuals who were aware of a hearing loss when compared to those individuals suffering from a comparable hearing loss, but who were unaware of the loss. This would seem to indicate the existence of a criterion difference between the two groups. The only method available to investigate the possible existence of such a difference is signal detection theory.

Watson, Franks, and Hood (1972) did, however, conduct a signal detection experiment in which data obtained from a two-interval forced-choice task were compared to audiometric zero at six frequencies. In their study, normal-hearing listeners were required to detect a signal (at one of six signal frequencies: 125, 250, 5000, 1000, 2000, and 4000 Hz) presented to their right ear via audiometer earphones. Unlike standard TSD studies, the signals were *not* presented against a background of noise. Each experimental session was composed of seven or eight blocks of trials with each trial block containing 100 individual trials. Signal frequency was held constant within a session. Within each trial block, the signal level was constant, but was adjusted between blocks in order to vary the P(C) measure between 0.60 and 0.90. Each subject received a total 10 hours of practice (5 hours with the 1000 Hz signal, 1 hour with each of the other 5 signal frequencies). From the detection data, psychometric functions were calculated with a P(C) value of 0.76 taken as "threshold." Results of the study indicated good agreement between the signal detection data and the ISO standard for audiometric zero.

Patterson (1974, 1976), Patterson et al. (1982), and Weber (1977) also used the two-interval forced-choice procedure in their studies dealing with the critical band and the auditory filter. Patterson (1974, 1976) and Weber (1977) followed a standard procedure where, for any given background noise, the signal was varied over four or five levels (between trial blocks) and a psychometric function generated. Threshold was defined as a P(C) equal to 0.75. However, Patterson et al. (1982) used a different implementation of the two-interval forced-choice procedure. In their study, the signal level was varied within a block of trials based on the observer's performance in the preceding trials. This procedure, referred to as "forced choice tracking" by Gescheider (1985, p. 117), allows an experimenter to adjust the signal level so as to produce a constant P(C) rather than a range of P(C) values. The procedure used by Patterson et al. (1982) was to adjust the signal level down 2 dB after two correct responses and adjust the signal level up 2 dB after an incorrect response. Threshold was defined as the average of the levels at which the signal was adjusted.

Although Fidell et al. (1974) did use TSD and the concept of the ideal observer (Green and Swets, 1988; Tanner Jr. and Sorkin, 1972) to develop their threshold prediction technique, they used the ascending method of limits to verify the model. No other studies were found in the TSD literature which directly addressed any of the independent variables to be addressed in the research effort described herein, either individually or in combination.

RESEARCH OBJECTIVES

From the preceding discussion, it is obvious that there are many factors which affect how well individuals will be able to perceive warning and alarm signals in noisy environments while wearing HPDs. However, one of the most important issues and the one that has received the least attention deals with the point at which a hearing loss becomes too great to allow an individual to detect auditory alarms or warnings and how this "point of impairment" changes with noise and signal level. The research effort described herein attempted to answer that question.

The experiment described in the following sections utilized the theory of signal detection (TSD, Green and Swets, 1988) to develop a mathematical model which could be used to predict whether or not an individual with a given hearing level would be capable of detecting a signal of a given level in a given level of background noise while wearing an HPD. Due to the large size of the experiment and the large number of TSD trials needed per condition, it was infeasible to use multiple HPDs or background noises with different spectral shapes. It was decided, therefore, to develop the model for the "worst case" condition and use pink noise (low-frequency biased) and an earmuff (relatively poor attenuation at the low frequencies with substantial attenuation at high frequencies). In addition, only one signal (a standard reverse or "back-up" alarm) was used.

Inputs for the resulting model include the pure-tone average hearing level of the individual being considered as well as the broadband A-weighted sound pressure levels of the background noise and the signal. Model output is a measure of the accuracy with which individuals will be able to discriminate signals from noise (or, if the assumptions concerning the dependent measure used in the experiment are unacceptable, an indication of whether or not the signal level is above or below threshold). The model may be applied in several ways. Not only can the model be used to predict if an individual is capable of hearing a signal in a given situation, it might also be used to estimate by what degree a noise must be reduced or at what level a signal must be presented to allow the greatest number of people to hear it.

The experiment described herein investigated the detectability of signals in noise only for the occluded (wearing an HPD) condition. This was done for several reasons. First, the intent of the experiment (as requested by the research sponsor) was to determine when it becomes unsafe for an individual suffering from a hearing loss to work in a noisy environment in which the use of HPDs is required. The experiment was not intended to determine the difference in signal detectability between the occluded and unoccluded conditions. Secondly, the experimental scenario was practical in that OSHA (1989) requires HPDs be worn when exposures exceed a 90 dBA time-weighted average (TWA) per 8-hour day. Furthermore, employers are required to supply HPDs to all employees whose 8-hour TWA is 85 dBA or greater, although the employees are not required to wear their HPD unless they have experienced a standard threshold shift (defined earlier). Therefore, in noise levels of 85 dBA and higher, HPDs should be worn by most employees. However, whether or not this is in fact the case depends in large part on the emphasis placed on hearing conservation by the employer and OSHA's enforcement of the law.

EXPERIMENTAL METHOD AND DESIGN

HPD Selection

The decision to use an earmuff was made since, based on the literature review, it is believed that an earmuff represents a "worst case" scenario for two reasons. First, since earmuffs generally show less attenuation than earplugs at the low frequencies, there is greater opportunity for the upward spread of masking beneath the earcups to reduce the audibility of a signal when using an earmuff. Secondly, earmuffs typically exhibit slightly better attenuation at frequencies from 1000 to 4000 Hz than do many premolded earplugs, therefore an earmuff would likely attenuate warning signals (which are usually in this frequency range) slightly more than an earplug. Because of this choice, the resulting model should be slightly conservative in its predictions since, if a sound is audible while wearing an earmuff, it should also be audible when using an earplug, but the converse would not likely be true.

The earmuff chosen for use in the experiment was a Bilsom Viking earmuff, Figure 19, manufactured by Bilsom International, Inc. This large-volume, high-attenuation earmuff was chosen because it was identified by ALCOA representatives (C. Dixon-Ernst, personal communications, April 14, 1992) as an earmuff they would consider appropriate for use in the noise levels being investigated (85 to 95 dBA). In addition, since the experimenter had considerable previous experience with the device in testing, and having found that consistent fits across sessions with the same subject as well as across subjects were easily obtained with the device, it was believed that differences in signal detection due to HPD fitting problems would be minimized.

Figure 19. *Bilsom Viking earmuff.*

Background Noise Spectra

Pink noise was used as the background noise in the experiment. Use of pink noise provided the greatest opportunity for upward spread of masking to decrease the audibility of the signal, but less opportunity for direct masking as a midrange-biased noise would. Also, interaction of the pink noise with the relatively poor low-frequency attenuation characteristics inherent with earmuffs was expected to further reduce the audibility of the signal at all noise levels. Furthermore, pink noise is a popular "generic" noise used in psychoacoustic studies which have industrial workplace implications. It represents all bands with equal energy when measured using proportional-bandwidth filters and is the noise used for calculating HPD attenuation as per the ANSI S3.19-1974 HPD real-ear testing standard.

Warning Signal

The warning signal used in the study was a standard back-up alarm, Figure 20, (manufactured by Caterpillar, Inc. PN 3T-1815) commonly found on heavy equipment. This type of warning signal was identified as one of the most common alarm/warning signals across all ALCOA facilities (S. I. Roth, personal communications, April 14, 1992) which represent typical heavy industrial plants which rely heavily on diesel powered vehicles. The spectrum of the particular reverse alarm is illustrated in Figure 21 while the corresponding third-octave levels are given in Table 4. These spectral measurements were made in the Auditory Systems Laboratory's anechoic chamber using a Larson•Davis (L•D) 800B sound level meter, an ACO 7013 1/2 in measurement microphone and an L•D Model 825-10 preamplifier. When making the measurements, the microphone was located at a distance of 113 inches from the alarm and oriented frontally-incident to it.

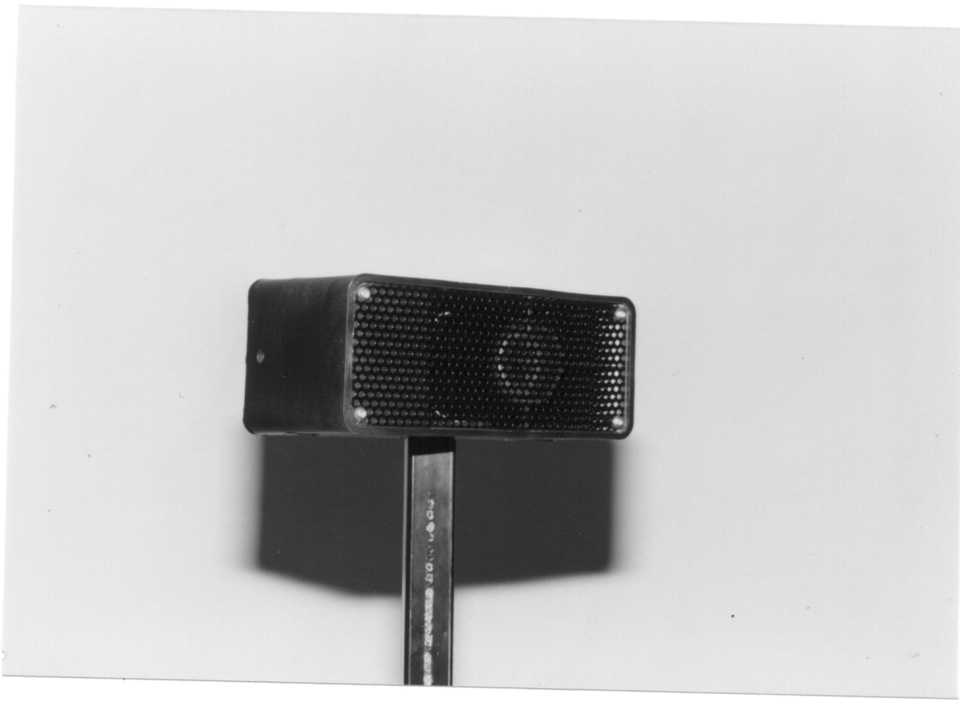


Figure 20. *Back-up alarm used in the experiment.*

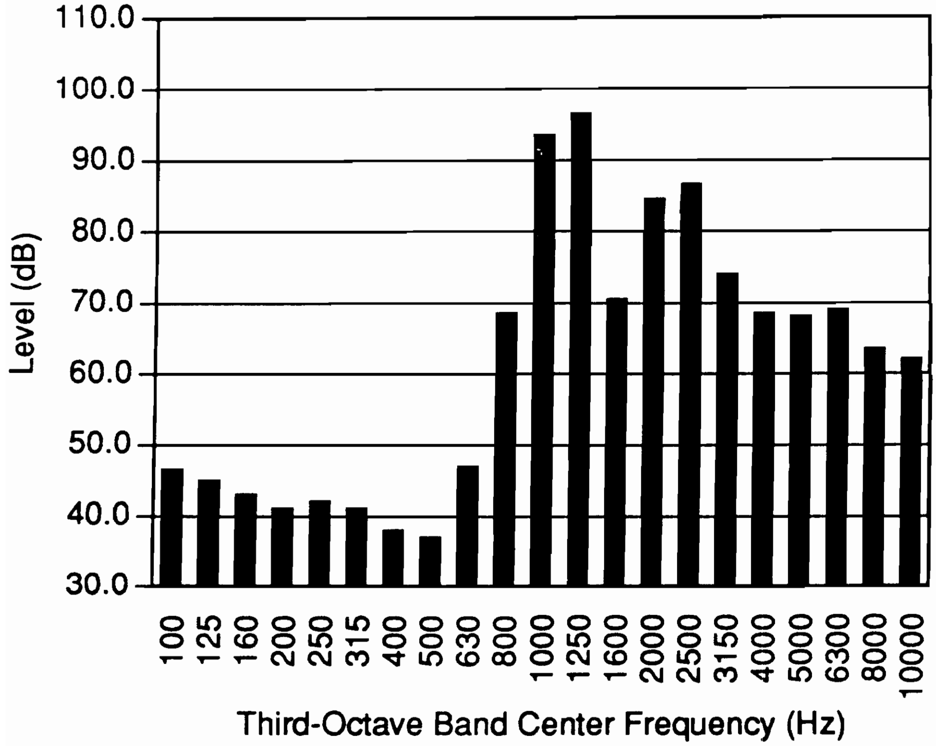


Figure 21. *Alarm spectrum.*

TABLE 4

1/3 Octave SPLs of Alarm

Third-Octave Center, Hz	SPL, dB
100	46.5
125	45.0
160	43.0
200	41.0
250	42.0
315	41.0
400	38.0
500	37.0
630	47.0
800	68.5
1000	93.5
1250	96.5
1600	70.5
2000	84.5
2500	86.5
3150	74.0
4000	68.5
5000	68.0
6300	69.0
8000	63.5
10000	62.0

As depicted in Figure 21, the alarm has most of its energy in the 1000-1250 Hz range with fairly strong harmonics present in the 2000-2500 Hz range. Very little energy is present below 800 Hz. These characteristics are in keeping with the warning signal design standards discussed earlier. The alarm operates with a 1 s period and a 50% duty cycle, also in line with the aforementioned standards. During the "on" portion of its duty cycle, the alarm output is constant between onset and offset. No amplitude- or frequency-modulation of the signal is used.

The alarm itself is switch-selectable for three sound output levels, HIGH [corresponding to a sound level rating of 112 dB(A)], MED. [107 dB(A)], and LOW [100 dB(A)]. The HIGH and MED. ratings correspond to Types A and B of SAE J994b, mentioned earlier. For experimental purposes, the alarm was tested and sampled with the switch set to HIGH since that is how it was received from ALCOA and conversations with ALCOA maintenance personnel (J. Hazelwood, personal communication, June 1992) indicated that they do not adjust the alarms when performing maintenance on the equipment.

Subjects

A total of 12 subjects, ranging in age from 18 to 73 years, participated in the experiment as paid volunteers. Each subject received compensation at a rate of $5 per hour for the time spent in the laboratory.

Screening criteria were based primarily on the subject's pure-tone hearing threshold. Prior to the screening procedures, each potential subject was asked to read and sign a written description of the experiment (Appendix A) as well as an informed consent form (Appendix B), the subject's rights were explained, and any questions answered. The screening session included asking the subject about his or her otological history, a brief otoscopic inspection of the outer ear, and a pure-tone audiometric examination. If the subject qualified and chose to participate in the study, he/she was scheduled for his/her first experimental session (a training session).

Facilities and Instrumentation

All experimentation was conducted in the Auditory Systems Laboratory on the Virginia Tech campus. This laboratory contains two audiometric test chambers, a reverberant room and an anechoic room, as well as a variety of support equipment and instrumentation. Instrumentation includes a Norwegian-Electronics type 828 integrated HPD signal presentation and measurement system controlled by an IBM PS/2 Model 70 microcomputer, an Apple Macintosh IIci microcomputer, a Beltone 114 clinical pure-tone audiometer, a Beltone 2000 clinical audiometer, and closed-circuit television (CCTV) system. In addition, a variety of audio signal generation and presentation equipment is available, as are several laboratory grade microphones and sound measurement instruments. The laboratory itself is an acoustically-isolated area so as to maintain a quiet environment for testing purposes.

Reverberant room. The reverberant room was used for all experimental sessions. This was be done to approximate the sound field conditions encountered inside large industrial plants with reflective wall and floor surfaces. The reverberant room is an extensively modified Industrial Acoustics Corporation (IAC) audiometric test booth. The chamber is of double wall steel construction, with approximately 4 in of fiberglass acoustic insulation sandwiched between the inner and outer skins. To achieve a reverberant sound field within the test space, the walls and ceiling are lined with one sheet of 0.5 inch thick gypsum board, on top of which is placed one sheet of 0.25 inch thick hard-tempered masonite. In addition, the carpet has been removed to expose the bare sheet-metal floor. The interior and exterior dimensions of the modified IAC test booth are given in Table 5, and ambient noise levels measured at the subject's head center position are shown in Table 6.

Anechoic chamber. The anechoic chamber was used for all audiometric tests, spectral analysis and digital sampling of the back-up alarm, and matching the spectral output of the digitized alarm with that of the original alarm. This chamber is a modified Eckel Corporation anechoic chamber and is of double wall steel construction with 3 inches of fiberglass acoustic insulation sandwiched between the inner and outer skins. Acoustic foam wedges line the six inner surfaces of the chamber providing a low-frequency cutoff of approximately 125 Hz, and an acoustically-transparent, expanded-metal grating suspended above the bottom wedges serves as a floor. The interior and exterior dimensions of the anechoic chamber are given in Table 7. The entire chamber is supported by 6 Hz vibration isolators to limit the structural-borne vibration reaching the test space. Ambient noise levels inside the test space, measured at the subject's head center position, are shown in Table 8.

Experimental apparatus. A schematic diagram of the experimental apparatus is illustrated in Figure 22. Presentation of all test stimuli (signals and noise) and recording of all subject response data were performed using a Macintosh IIci microcomputer. The Larson•Davis 3100D RTA served as a pink noise generator and was controlled via its RS232 serial port. The pink noise output of the L•D 3100D RTA was directed to a Scott Model 458A (65 w/ch) integrated audio amplifier and a Realistic Model 31-2000A octave band equalizer, used to shape the noise. The noise output of the Scott amplifier was directed to a pair of Infinity RS6b 3-way loudspeakers situated inside the reverberant room as shown in Figure 22 (speakers 1 and 2).

The warning signal was digitized and presented via the computer's digital audio output. The signal was shaped via an AudioControl octave band equalizer and a Ross

TABLE 5

Reverberant Test Chamber Dimensions
(all dimensions in inches)

	Interior Dimensions	Exterior Dimensions
Length	110	120
Width	74.25	83.25
Height	92.5	103.75

TABLE 6

Ambient Noise Levels in the Reverberant Test Chamber

Octave Band (OB) Center, Hz	Ambient OB Level, dB*
125	20
250	14
500	6.5
1000	4.5
2000	2.7
4000	5.1
8000	8.1

*From Casali and Robinson (1990).

TABLE 7

Anechoic Test Chamber Dimensions
(all dimensions in inches)

	Interior Dimensions*	Exterior Dimensions
Length	91	144
Width	114.5	168
Height	85	138

* Measured between foam wedge tips.

TABLE 8

Ambient Noise Levels in the Anechoic Test Chamber

Octave Band (OB) Center, Hz	Ambient OB Level, dB*
125	23.3
250	5.5
500	5.7
1000	7.5
2000	5.6
4000	7.3
8000	9.3

*From Casali (1988).

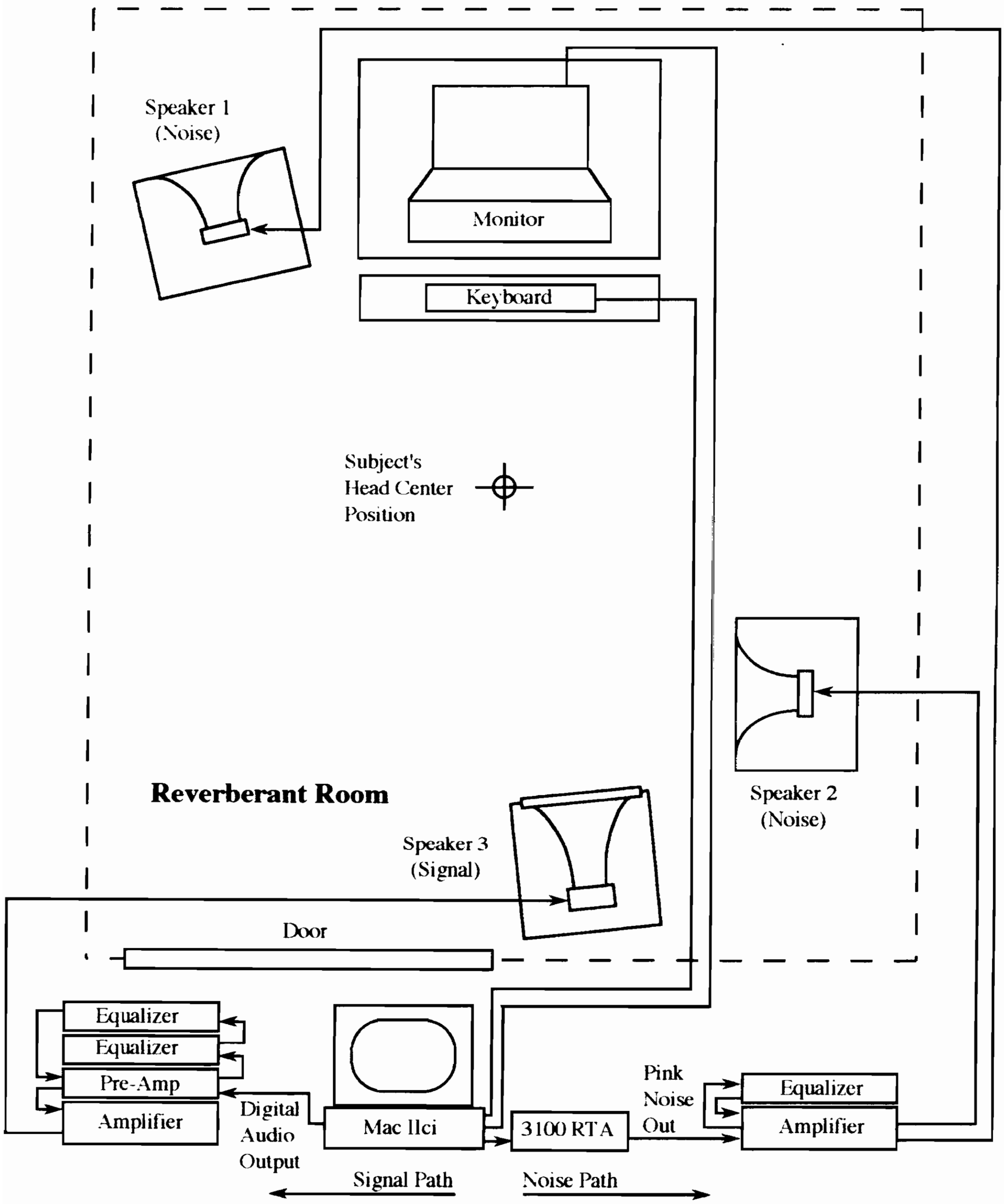


Figure 22. *Schematic diagram of the experimental apparatus.*

R31M third-octave band equalizer and amplified using an Adcom GFP-545II (100 w/ch) and GFP-555II amplifier/pre-amplifier combination. Output from the Adcom amplifier was directed to a single Klipsch K57K midrange horn driver located behind the subject, just to the right of the door (speaker 3). This bank of instrumentation provided a faithful reproduction of the acoustic characteristics of the original back-up alarm.

A Hewlett-Packard 98754A 20 inch Trinitron VDT display was used to present visual information, instructions, and feedback to the subject during the course of the experiment. Subject responses were made using a modified computer keyboard. The subject's responses were monitored on an Apple 12 inch monochrome monitor located at the experimenter's station. The monitor and keyboard as seen from the subject's view are illustrated in Figure 23, while the loudspeaker arrangement can be seen in Figure 24.

Several acoustic measurements were made in order to characterize the acoustical environment inside the test space as it was configured for the experiment. These measurements included reverberation time (RT_{60}) and diffusivity measurements as well as verifying that the pink noise as used in the experiment was indeed flat by octaves. Reverberation times at nine third-octave bands from 125 to 8000 Hz are give in Table 9. As stated earlier, the Realistic octave-band equalizer was used to shape the pink noise to ensure that it was flat by octaves as measured in the test space. An octave-band spectrum of the resulting pink noise is shown in Figure 25. The corresponding octave band SPLs are given in Table 10. As can be seen, the measurements are flat (within 3 dB) from 63 to 8000 Hz. Two additional tests were conducted to ensure that the sound field was as uniform as possible about the subject's head center position. These tests involved: 1) examining the differences in the measured SPL at six positions about the subject's head center position in each of nine third-octave bands and 2) examining the differences in the average SPL measured in each of the three principal planes of the test space at nine third-octave bands using a directional (cosine) microphone rotated in 15° increments about

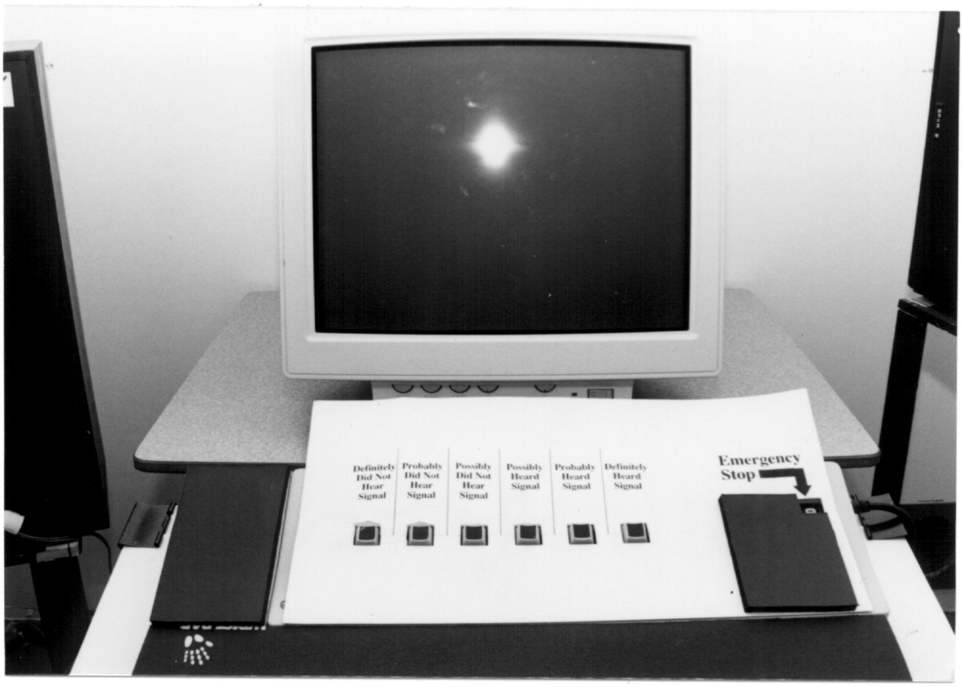


Figure 23. *Monitor and keyboard as seen by the subject.*

Figure 24. *Test chamber as seen from the door.*

TABLE 9

Reverberation Times in Test Space

Third-Octave Center, Hz	Reverberation Time RT_{60}, s
125	0.51
250	0.82
500	1.20
1000	1.13
2000	1.05
3150	1.02
4000	0.97
6300	0.80
8000	0.73

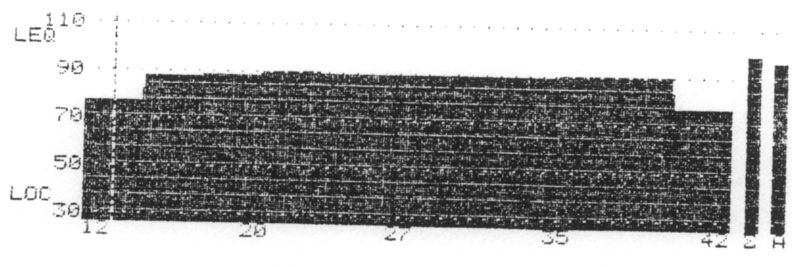


Figure 25. *Octave-band spectrum of the pink noise used in the experiment, as obtained using a L•D 3100D real-time spectrum analyzer.*

TABLE 10

Octave Band Spectral Measurements for Pink Noise

Octave Band (OB) Center, Hz	SPL, dB*
16	76.2
31.5	87.0
63	87.7
125	89.7
250	89.1
500	88.5
1000	89.0
2000	88.8
4000	89.4
8000	89.8
16000	76.2

each of the three principal axes of the room. Both tests were conducted using pink noise at an overall SPL of 95 dBA. [These tests are commonly used to characterize the sound field diffusivity for HPD test facilities operating under either ANSI S3.19-1974, "Method for the Measurement of Real-Ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs" (ANSI, 1974) and/or ANSI S12.6-1984, "Method for the Measurement of the Real-Ear Attenuation of Hearing Protectors" (ANSI, 1984). Complete details as to exactly how each of these tests are performed may be found in either standard.]

Results of these two tests are shown in Tables 11 and 12 respectively. As can be seen in Table 11, the maximum differences found in the six-position test (5.0 and 2.9 dB) occurred at 250 and 500 Hz respectively. The maximum left-right differences (3.1 and 2.2 dB) also occurred at these two frequencies. These differences were not considered serious since these frequencies were well below the frequencies at which most of the signal's energy was centered (1000-1250 Hz). All other differences (L-R or six-position) were less than 2 dB. As shown in Table 12, the difference between the average SPL in any of the three principal planes never exceeded 2.4 dB in any of the third-octave bands tested. It was therefore concluded that the sound field in the test space, as it was configured for the experiment, was reasonably diffuse and non-directional.

As stated earlier, the spectral analysis of the back-up alarm was carried out with the alarm located in the anechoic room. This was done to prevent room acoustics (i.e., reverberation) from affecting the measurements. Similar procedures were followed when the alarm was digitally sampled. For this purpose, an AKG C414B-ULS dual diaphragm microphone (set to its cardiod pickup pattern) was placed coaxially with the alarm at a distance of 116 in. Sampling was accomplished using MacRecorder Sound System Pro sound digitizing hardware and software in conjunction with the Macintosh IIci microcomputer at an 11 kHz sampling rate. The alarm was sampled for a total of five

TABLE 11

SPL Variation at Six Positions About Head Center Position

1/3 OB Center (Hz)	dB Right -15,0,0*	dB Left 15,0,0	R-L Δ**	dB UP 0,0,15	dB Down 0,0,-15	dB Front 0,-15,0	dB Back 0,15,0	6-Pos Δ***
125	88.0	87.6	0.4	87.7	87.7	87.4	87.3	0.7
250	82.7	85.8	3.1	80.8	85.0	81.9	83.6	5.0
500	83.0	85.2	2.2	85.9	84.8	84.0	85.3	2.9
1000	84.1	83.8	0.3	83.8	83.5	84.7	84.6	1.1
2000	85.1	83.5	1.6	84.3	83.8	84.4	84.1	1.6
3150	84.1	84.6	0.5	85.0	83.6	83.9	83.7	1.4
4000	84.1	84.7	0.6	85.5	84.1	85.9	85.0	1.8
6300	86.4	86.3	0.1	86.9	86.2	86.7	87.2	1.0
8000	86.2	86.1	0.1	86.1	85.7	85.9	86.1	0.4

* All dimensions are in cm.

** Absolute value dB difference between right and left microphone positions.

*** Maximum absolute value dB difference between all pairs of the 6 microphone positions.

TABLE 12

Sound Field Directionality in Test Space (Data are in dB)

	1/3 OB Center (Hz)								
Axis	125	250	500	1000	2000	3150	4000	6300	8000
X	74.9	84.3	82.6	80.7	79.6	81.5	83.3	85.1	85.5
Y	75.6	81.9	83.4	81.2	80.1	81.6	83.9	85.7	85.8
Z	74.5	84.0	82.7	81.6	80.7	82.2	84.1	87.5	87.6
Δ Max	**1.1**	**2.4**	**0.8**	**0.9**	**1.0**	**0.6**	**0.8**	**2.4**	**2.1**

periods. The sound editing software was then used to edit the sample to just two periods for use in the experiment. When the custom C-language software used to control the presentation of the signal for experimental purposes was complete, the Klipsch midrange driver was placed in the anechoic room and the digital signal output was shaped to match that of the original signal. This was done to eliminate any coloration of the signal due to nonlinearities in the hardware or software used in the sampling procedure. In this way, it was assured that the spectrum of the signal emanating from the horn driver used in the experiment would be as close as possible to that of the original back-up alarm.

Support instrumentation. A Beltone Model 114 clinical pure-tone audiometer, used in conjunction with TDH 50 earphones was used to determine the pure-tone hearing threshold level of all subjects during the screening process as well as for the before- and after-session thresholds. All audiograms were obtained in the anechoic room. An intercom system was used to allow the experimenter and the test subject to communicate during the course of an experimental session whenever necessary. The intercom system utilizes a "hot" subject microphone so that the subject is not required to depress a push-to-talk switch to communicate with the experimenter. A closed-circuit television (CCTV) system was used to visually monitor the subject during each experimental session.

Sound measurement instrumentation used during the course of the research effort included an L•D Model 800B precision sound level meter and an L•D Model 3100D real-time spectrum analyzer. These devices were used in conjunction with one or more of the following preamplifiers and/or microphones: an L•D Model 900B preamplifier, an L•D Model 825-10 preamplifier, an L•D Model 2540 half-inch microphone, an ACO Model 7013 half-inch microphone, an ACO Model 7023 one-inch microphone, and an AKG C414B-ULS dual diaphragm microphone. Many of these instruments were used to calibrate the system at the beginning of the day and to monitor the SPLs within the test space during an experiment.

Experimental Design

The experimental design used in the research effort described herein was the mixed three-factor design illustrated in Figure 26. Data analysis procedures included nonlinear regression and repeated measures analysis of variance (ANOVA).

Independent variables. The three independent variables represented in the design were hearing level (HL), noise level (NL) and signal-to-noise ratio (SN).

Hearing level was the single between-subjects variable and had three levels (normal hearing, slight loss, and mild/marked loss). The descriptive terms used are in general agreement with those used by Miller and Wilber (1991) when related to the hearing levels represented by each category as described below.

Subjects were screened based on their pure-tone average (PTA) hearing levels over the frequency range from 500 to 2000 Hz. This was done since the back-up alarm used in the experiment had most of its energy contained in the 1000 to 2000 Hz range and it was believed, based on the results of the literature review, that the frequencies above 2000 Hz would have little impact on the detectability of the signal. Normal-hearing subjects were those individuals whose PTA hearing levels in both ears in the frequency range of interest was between 0 and 20 dBHL. [Hearing level (HL) is a weighted dB level for each audiometric frequency of interest as defined in ANSI S3.6-1989, "Specification for Audiometers" (ANSI, 1989). Therefore, pure-tone hearing thresholds are given in units of dBHL.] Subjects falling into the second group (slight loss) were required to have PTA hearing levels in both ears between 20 and 40 dBHL. Subjects whose PTA hearing levels in both ears were above 40 dBHL but below approximately 60 dBHL qualified for the third group. An additional requirement placed on all subjects was

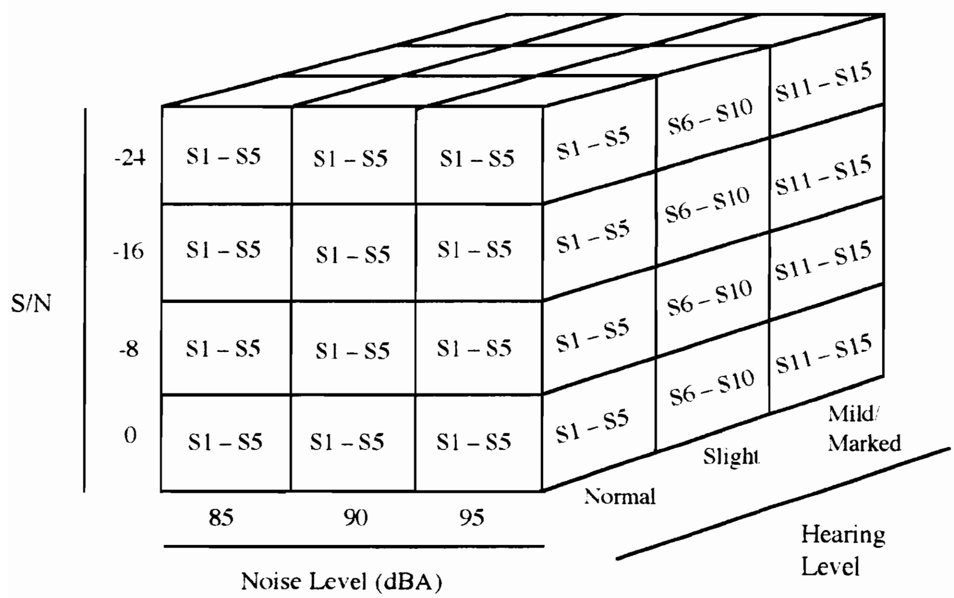


Figure 26. *Experimental design.*

that their range of hearing in either ear across the frequencies of interest (500 to 2000 Hz) could not exceed 30 dBHL. It was intended that only five subjects would be recruited for each of the HL categories because availability of subjects with the necessary hearing loss was an unknown factor. (As it turned out, just enough subjects were found who met the criteria for the third group (the group with the greatest loss) but only two subjects meeting the criteria for the middle group could be recruited.)

Each subject's pure-tone hearing threshold at the frequencies of interest as well as the monaural PTA hearing levels for the frequencies from 500 to 2000 Hz (used for categorizing purposes) are given in Table 13. The monaural PTA hearing levels using two other frequency ranges are also given. These additional classification schemes were utilized in the regression analysis, and will be discussed later.

Noise level was a within-subject variable with three levels representing the broadband A-weighted sound pressure level (SPL) of the background noise against which the signal was presented. The three levels used in the experiment were 85, 90, and 95 dBA. This range of noise levels encompasses over 90% of the levels commonly encountered in industry (EPA, 1981).

Signal-to-noise ratio was a within-subject variable with four levels (0, -8, -16, and -24 dB). This variable represents the broadband A-weighted SPL of the signal relative to that of the noise. Thus, a signal-to-noise ratio of -8 dB indicates that the broadband SPL of the signal is 8 dB less than the broadband SPL of the background noise. The levels chosen for the signal were based on pilot tests (described in Appendix C) which indicated that the masked threshold for the signal in pink noise for normal listeners was in the vicinity of -20 to -25 dB.

Counterbalancing. Presentation order of conditions for each subject was accomplished by way of random assignment. In addition, due to the size of the experiment, it was decided that use of a single earmuff would be unwise since the

TABLE 13

Experimental Subjects' Pure-Tone Hearing Thresholds

Subject		Audiometric Frequency (Hz)						Range of PTA Hearing Level		
Number	Ear	500	1000	1500	2000	3000	4000	.5-2 kHz	.5-4 kHz	1-2 kHz
1	R	0	5	0	-5	-5	0	0.0	-0.8	0.0
	L	10	5	10	10	10	0	8.8	7.5	8.3
2	R	5	5	15	10	5	5	8.8	7.5	10.0
	L	5	5	5	5	0	5	5.0	4.2	5.0
31	R	5	0	5	5	0	15	3.8	5.0	3.3
	L	10	15	10	10	20	20	11.3	14.2	11.7
4	R	5	5	5	5	10	15	5.0	7.5	5.0
	L	10	15	10	10	20	20	11.3	14.2	11.7
5	R	15	5	10	15	10	10	11.3	10.8	10.0
	L	15	0	5	10	15	5	7.5	8.3	5.0
6	R	30	20	15	20	35	55	21.3	29.2	18.3
	L	30	25	20	20	35	60	23.8	31.7	21.7
7	R	10	20	35	40	50	55	26.3	35.0	31.7
	L	15	15	30	40	75	80	25.0	42.5	28.3
11	R	40	65	50	40	55	50	48.8	50.0	51.7
	L	35	55	55	60	60	60	51.3	54.2	56.7
12	R	50	55	50	40	60	75	48.8	55.0	48.3
	L	55	55	55	40	50	65	51.3	53.3	50.0
13	R	65	60	60	55	55	60	60.0	59.2	58.3
	L	65	70	70	60	60	60	66.3	64.2	66.7
14	R	40	45	50	45	55	60	45.0	49.2	46.7
	L	40	55	60	55	65	65	52.5	56.7	56.7
15	R	45	55	45	45	40	40	47.5	45.0	48.3
	L	55	60	60	60	55	65	58.8	59.2	60.0

headband force might decrease over the course of the experiment resulting in poorer fits for those individuals finishing the experiment last. It was therefore decided to use four earmuffs and randomly assign them to conditions for each subject such that each subject would use each earmuff three times during the course of the experiment.

Dependent measure. The dependent measure used in the experiment was the proportion of the area under the receiver operating characteristic (ROC) curve, P(A), as recommended by Swets (1986, 1988b). One advantage to using P(A) is that it is independent of the underlying distributions governing the subjects' responses (Robinson and Watson, 1972). In addition, since values of P(A) range from 0.50 for chance performance to 1.0 for perfect detection (Swets, 1986), the measure may be thought of as the *accuracy* with which a subject can discriminate a signal from noise (Swets, 1988b; Swets, Pickett, Whitehead, Getty, Schnur, Swets, and Freeman, 1979). If this assumption is unacceptable, then at the very least, a value of P(A) of 0.75 may be thought of as an indication of "threshold" since it represents a level of performance halfway between chance and perfect performance.

An example of the dependent measure is illustrated in Figure 27, in which three representative ROC curves are plotted. The ordinate and abscissa of the graph are the probability of a correct detection [probability of a hit, P(H)] and the probability of a false alarm [P(FA)] respectively. The dependent measure, P(A), is the proportion of the unit area of the graph below and to the right of the ROC curve. Thus, the accuracy corresponding to each of the three ROC curves shown in the figure are 0.95, 0.85, and 0.75. The diagonal represents chance performance, or an accuracy of only 0.50.

Use of a signal detection theory protocol in an experiment such as the one described herein has the advantage of allowing the separation of a subject's criterion for making a positive response to a stimulus from the subject's sensitivity to the particular stimulus being used to elicit the response.

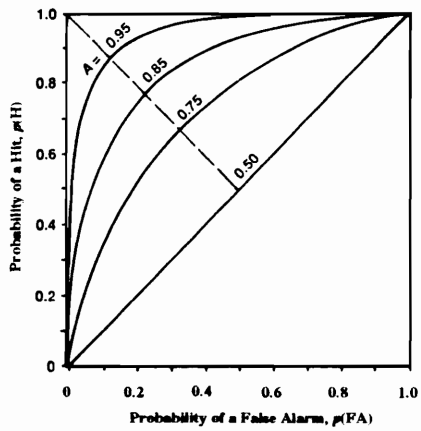


Figure 27. *Illustration of the dependent measure*, P(A). (from Swets, 1988)

Experimental Procedure

The experiment was conducted using the rating procedure methodology of signal detection theory. Signal detection theory was chosen since it was felt that if one of the classical psychophysical methodologies had been used, the data might have been biased by one or more of the non-experimental factors mentioned earlier (i.e., inattention, motivation, anxiety, experience, etc.) which have been shown to influence a subject's criterion. The rating procedure was selected over the other methodologies not only for its economy, but also because it was hoped that it might allow examination of criteria as a second dependent variable. As it turned out, this was not possible. A discussion of the reasons why this is so is contained in the Results section.

Each subject was required to attend 12 experimental sessions in addition to the screening session and one practice session. Each session (including the practice session) was broken down into four parts; a pre-test audiogram, fitting and fit-testing of the earmuff, the signal detection task, and a post-test audiogram. An outline of the overall session structure is illustrated in Table 14. The practice sessions were structured exactly like the experimental sessions (with the exception that the noise and signal levels were slightly different from, but within the range of, those used in the experiment) so as not to confuse the subject with a change in procedure. Details of the experimental procedures are given below.

Pre- and post-test audiograms (at the frequencies of 500, 2000, 3000, and 4000 Hz in each ear) were performed not only to ensure that the subject's hearing had not changed drastically since the previous session (due to tinnitus, head cold, TTS, etc.), but also to determine if the subject experienced a temporary threshold shift as a result of his/her participation in the experiment. As stated earlier, all audiometric tests were performed in

TABLE 14

Outline of the Procedures in a Typical Experimental Session

- Set/check the signal and noise levels before the subject arrives.
- Conduct pre-test audiogram in anechoic room (only at 500, 2k, 3k, & 4k Hz).
- Review instructions with the subject (Table 15).
- Fit the miniature microphones in each ear.
- Fit the earmuff and attach the miniature microphones to the earcups' exterior.
- Ask subject to enter the reverberant room.
- Perform the noise reduction (NR) test of the earmuff fit and compare NR value at 1000 Hz with previous values obtained.
- If necessary, adjust earmuff and repeat NR measurement.
- Ask subject to exit the reverberant room.
- Set/check the signal and noise levels.
- Ask subject to re-enter the reverberant room.
- Preview the signal and noise for the subject.
    - Six intervals of the signal alone,
    - Six intervals of the noise alone,
    - Six intervals of the signal and noise together.
- Perform first three blocks of trials.
- Ask subject to exit the reverberant room.
- Check the signal level, adjust if drift detected.
- Ask subject to re-enter the reverberant room.
- Perform last three blocks of the experiment.
- Ask subject to exit the reverberant room.
- Check the signal level.
- Remove the earmuff and miniature microphones from the subject.
- Conduct the post-test audiogram in anechoic room (only at 500, 2k, 3k, & 4k Hz).
- Schedule the subject for his/her next session.

the anechoic room. The pre-test audiogram was performed immediately after the subject entered the laboratory.

Once the pre-test audiogram was completed, the instructions for the remainder of the session, Table 15, were reviewed with the subject. (A much more detailed set of instructions were read to the subject during the training session, these appear in Appendix D.) If there were no questions, miniature microphones (Knowles Model BT-1759) were placed in the concha of each ear, Figure 28. Next, the earmuff was fit on the subject and a second pair of microphones were attached to the exterior of the earcup, Figure 29. The purpose of the miniature microphones was to obtain a noise reduction measurement for determining the adequacy of the earmuff's fit. After seating the subject in the reverberant room, pink noise was played at a level of 95 dBA for approximately 2 minutes while a 30 s Leq measurement was obtained from each microphone. Once data were obtained from each microphone, the pink noise was terminated and the noise reduction measured in the third-octave band centered at 1000 Hz was calculated. [Casali and Park (1992) have shown that attenuation measured in the third-octave band centered at 1000 Hz is an acceptable predictor of the broadband attenuation of earmuffs. The physical measurements themselves were accomplished by means of a computer program (Mauney, 1992) implemented on an IBM/PS2 Model 70.] The noise reduction values for each earcup were examined and compared to the values obtained for that subject in all previous sessions. If the obtained measures were such that the highest measured value was no more than 20% greater than the minimum measures obtained, the fit was considered acceptable. (With earmuff attenuation typically ranging from 30 to 35 dB at 1000 Hz, allowing 6 to 7 dB of variability in the fit of the earmuff was considered to be practical compromise between the need for a consistent fit and the reality of achieving that consistency.) If, however, the measures were not in this range (either too high or too

TABLE 15

Instructions Given the Subject During Each Session

Remember, the first thing I will do is to fit the miniature microphones in the concha of each ear. Then I will fit the earmuff and attach another set of miniature microphones to the exterior of each earcup. Once I get what I think is a good fit, I will ask you to enter the room.
Once I get the leads for the mini-mics connected, I will perform the physical test of the earmuff's fit. Remember, during this part of the test, you do not make any response, just sit as still as you can and hold your head as you do during the experiment.
After the physical test is complete, I will ask you to exit the room so I can check the signal level. Once the signal level is OK, I will ask you to re-enter the room.
Once you are seated and comfortable, I will preview the signal and noise at the levels you will be listening to them today. I will present six intervals of the signal by itself, followed by six intervals of the noise by itself, followed by six intervals of the signal and noise together.
If you do not have any questions, we will then proceed with the experiment. Remember, during the experiment, concentrate on listening for the signal. Try to do the best you can and try to use as many of the response categories as you can.
After three blocks of trials, I will again ask you exit the room so I can recheck the signal level and you can stretch your legs.
Once that is complete, I will then ask you to re-enter the room and we will finish the last three blocks of trials.
Do you have any questions?

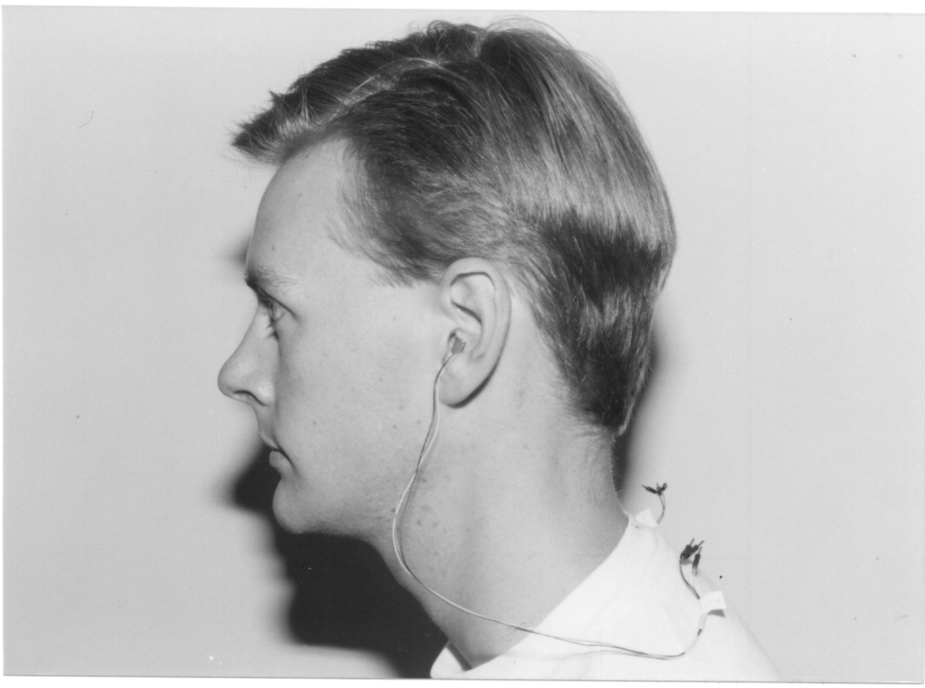


Figure 28. *Knowles miniature microphone in the concha of the ear.*

Figure 29. *Knowles miniature microphone on the earmuff's exterior.*

low), the earmuff was refit and the noise reduction procedure repeated until an acceptable fit was achieved.

Once an adequate fit had been obtained, the signal and noise levels for the session were set. The noise level was set using the L•D 3100 RTA using exponential averaging and a 1 s (SLOW) averaging time. The signal level was also set using the L•D 3100 RTA and exponential averaging, but using an averaging time of 1/32 s and reading the maximum SPL (L_{max}). The faster time constant and L_{max} were necessary because the alarm operated on a very fast cycle with an "on" time of only 1/2 s. This was roughly equivalent to using the "impulse" response on a standard sound level meter. (The subject was not in the room when the noise and signal levels were being set).

After setting the signal and noise levels to be used in the session, the actual data collection portion of the experiment began. Prior to actually starting the session, however, the noise and signal were previewed for the subject (with the subject in the room) at the levels to be used in the session. For this preview, six intervals of the signal alone were presented, followed by six intervals of the noise alone, followed by six intervals of the signal and the noise together.

The experiment itself consisted of 6 blocks of 84 individual trials (a total of 504 trials) during which a brief period of noise was presented to the subject. Exactly half of the trials in each block contained a signal (consisting of two "on" cycles of the back-up alarm) in addition to the noise. The subject's task was to indicate whether or not a signal was heard and how sure he or she was of his or her response. Responses were made by pressing one of six switches on the keyboard located in front of them. Each switch represented a response ranging from "Definitely Did Not Hear Signal" to "Definitely Heard Signal." Intermediate responses were: "Probably Did Not Hear Signal," "Possibly Did Not Hear Signal," "Possibly Heard Signal," and "Probably Heard Signal." (Using an even number of responses was an attempt to prevent subjects from responding in an ambivalent manner.)

Each presentation was preceded by a warning message appearing on the monitor to alert the subject that a trial was imminent. Immediately following the warning, the test stimulus was presented. The length of this presentation interval was approximately 2 s. Coincident with the initiation of the noise, a large question mark appeared on the monitor to indicate that the subject could respond. Two seconds after the noise was terminated, the question mark disappeared (indicating the end of the response completion interval) and one of two feedback messages appeared to indicate whether there had or had not been a signal present in the previous trial. These messages were displayed for approximately 1 s. The feedback messages were then removed and the warning message appeared once again. This cycle was repeated until all 84 trials had been completed. (Visual stimuli were used for warning and feedback so as not to confuse the subject about the auditory signal being listened for.)

An illustration of the trial structure is shown by the timeline in Figure 30. [The structure of the trial blocks (pre-session signal preview, number of blocks, trials per block, total number of trials, warning, and feedback) are based on the recommendations of Green and Swets (1988).] Before starting each block of trials, the subject was encouraged to do his/her best and to try to use as many of the response categories as possible. Upon completion of each block of trials, the subject was given the opportunity to take a short break, if desired. After the third block of trials, the subject was asked to step out of the room and the signal level was checked (and adjusted if any drift was detected). After the break, the subject returned to the reverberant room and started the next block of trials.

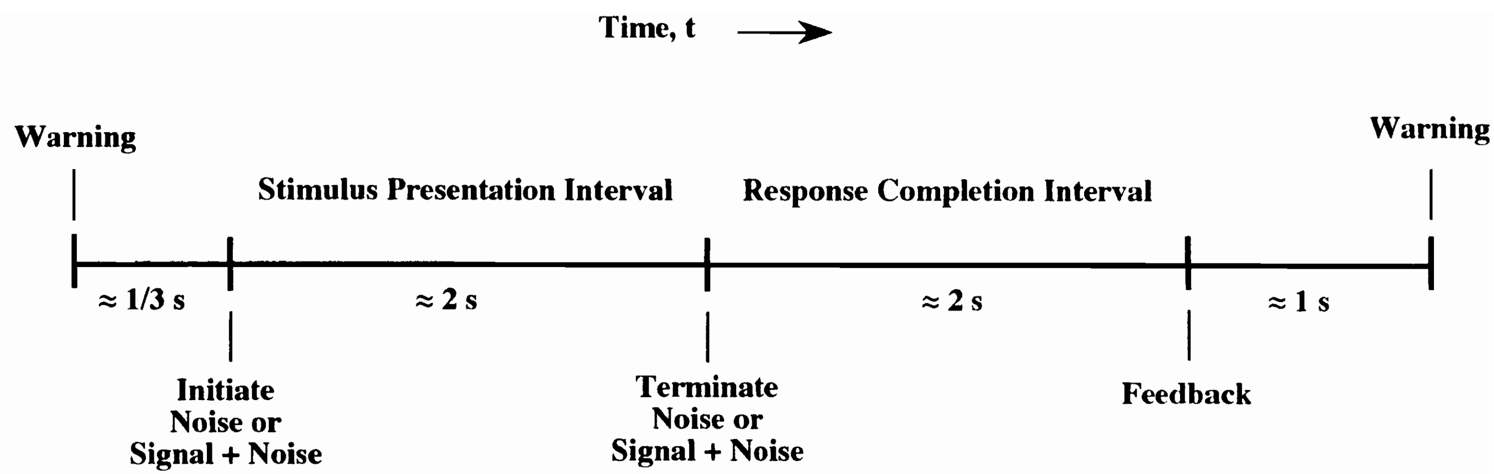


Figure 30. *Trial timeline.*

Once the sixth block of trials was completed, the subject was asked to step out of the room, the earmuff and miniature microphones were removed, and the signal level was checked once again. Following this, the post-session audiogram was performed. Each experimental session lasted between 1-1/2 and 2 hours.

RESULTS

Data Reduction and Analysis

Before proceeding with the discussion of how the data were analyzed, it is first necessary to state that only the sensitivity data were analyzed, criteria data were not. The reason for this is simple. Upon examination of the raw data, it was evident that in the conditions in which the normal listeners were responding in a manner that would allow generation of criteria measures, the subjects in the group with the greatest hearing loss (group 3) could not hear the signal at all and were responding accordingly. Likewise, in the conditions where the group 3 subjects were responding so as to allow calculation of criteria measures, the normal listeners had no problem hearing the signal and responded accordingly. With few exceptions, the listeners in the middle group responded in a manner similar to the better hearing subjects. It was therefore decided to make no attempt to generate or analyze criteria data since so little data would have resulted and the analysis would have been reduced to comparing individuals in specific conditions rather than groups across multiple conditions. In other words, there just was not enough data to analyze. In addition, there was a considerable age difference between the normal-hearing subjects (ranging in age from 18 to 26 years with a mean age of 20.2 years) and the most hearing-impaired subjects (ranging in age from 34 to 66 years with a mean age of 54.6 years). So even if there had been enough criteria data to analyze, the results would have been hopelessly confounded with age.

The raw data, as it existed at the end of an experimental session, consisted of a record of what response choice was made by the subject for each of the 504 trials and whether or not a signal had been presented during each trial. Before any analysis could be performed, however, it was necessary to get the data into a useful form. Once that had been accomplished, it was necessary to generate receiver operating characteristic (ROC) curves for each subject and calculate the area beneath each curve and thus generate the dependent measure used in the subsequent analyses.

As mentioned briefly in an earlier section, the data were analyzed using both nonlinear regression and repeated measures analysis of variance (ANOVA) techniques. In the discussions which follow, the procedures used to generate the dependent measure will be described first, followed by discussions of the procedures used in the regression analysis and analysis of variance.

Generation of the Dependent Measure

As stated above, the raw data obtained during each experimental session existed simply as an indication of what rating response was made by the subject in each trial and whether or not a signal had been presented during that trial. From these data, cumulative response frequencies were determined which indicated how many times the subject used each response for trials when just the noise was presented (noise) and when a signal was presented in addition to the noise (signal-plus-noise). [Missed trials (i.e., trials in which the subject made no response for whatever reason) were not counted.] These cumulative response frequencies were then converted to proportions and normalized. Each step of this procedure is illustrated in Table 16. In the table, response categories R1 through R6 correspond to the six response options available to the subject in the experiment such that R1 would correspond to "Definitely Did Not Hear Signal" and R6 to "Definitely Heard Signal."

For example, the subject used response category R4 (corresponding to a response of "Possibly Heard Signal") 95 times when a signal was presented. Adding this number to the 79 additional responses made using response categories R5 and R6 (corresponding to responses of "Probably Heard Signal" and "Definitely Heard Signal") and dividing by 247 (the total number of responses made when both signal and noise were presented)

TABLE 16

Response Proportions for One Experimental Session (Subject 1, NL = 95 dBA, SN = -24 dB)

Raw Responses

	R1	R2	R3	R4	R5	R6
Signal-plus-noise	0	43	30	95	78	1
Noise	23	95	52	55	22	1

Raw Cumulative Response Proportions

	R1	R2	R3	R4	R5	R6
Signal-plus-noise	1.0	1.0	0.826	0.704	0.320	0.004
Noise	1.0	0.907	0.524	0.315	0.093	0.004

Normalized Response Proportions

	R1	R2	R3	R4	R5	R6
Signal-plus-noise	—	—*	0.938	0.537	-0.47	-2.65
Noise	—	1.324	0.061	-0.48	-1.32	-2.65

* z-scores for response proportions are at $\pm \infty$ and are not considered.

results in a cumulative response proportion for response R4 of 0.704. The normalized response proportion is then found by simply referring to a table of the cumulative normal distribution and finding the z-score corresponding to a cumulative proportion of 0.704.

Once the normalized response proportions had been obtained, the z-scores were plotted (Figure 31) such that for each response category (R1 through R6), responses made during signal-plus-noise trials were considered hits [Z(H)] and responses made during noise trials were considered false alarms [Z(FA)]. The least-squares solution for the best straight line through the normalized data was then obtained. [Although a computer program had been obtained (Dorfman, 1983) that would provide maximum likelihood estimates of a straight line, its use would have required collapsing the data in cases where a subject used a response category infrequently or did not use a category for both noise and signal-plus-noise conditions. It was felt that collapsing the response data in this manner would not be representative of how the subject actually performed and could bias the results more so than using the least squares solution.] Two special cases existed, when the normalized data provided only one or two points which could be plotted on the normalized ROC curve. In the case of a single data point, the line passing through the point with a slope = 1 was used since this is the value predicted by signal detection theory. When only two points were available, the equation of the line passing through both points was used.

The equation for the normalized ROC curve was then used to generate 400 points for the ROC curve as illustrated in Figure 32. Also shown in Figure 32 are the raw response proportions from Table 16. The area under this ROC curve, P(A), was then calculated using the multiple application of Simpson's 1/3 rule extension of the Newton-Cotes numerical integration formulae (Chapra and Canale, 1985).

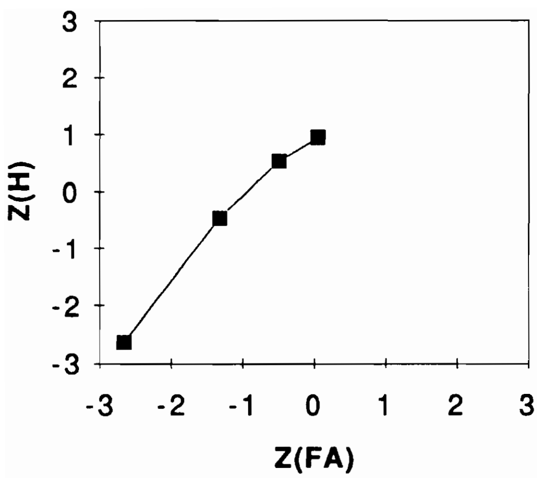


Figure 31. *Normalized ROC curve.*

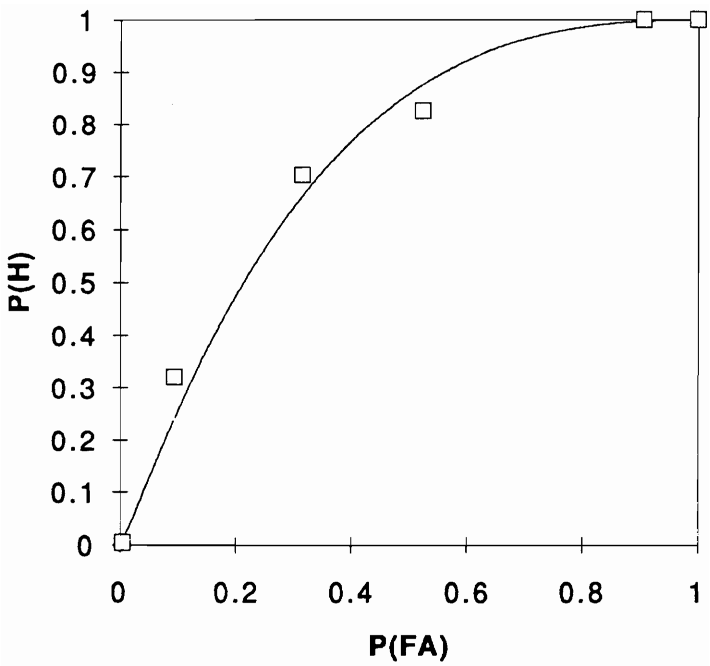


Figure 32. *ROC curve*.

Testing the Significance of Blocks

The procedures briefly outlined in the preceding discussion were carried out using the data from all six blocks of trials as well as for the data from just the last five blocks of trials. This was done since some statistical reservations had been expressed concerning the effect that inclusion of data from the first block of trials would have on the data (N. Sussman, personal communications, April 14, 1992) and because Robinson and Watson (1972) make the statement that discarding the first block of data is common in TSD research. The data thus obtained were subjected to a paired *t*-test to determine if use of the data obtained in the first block of trials made any difference in the dependent measure. Results indicate that there was not a significant difference between the P(A) measures when the data from the first block of trials were included or excluded (t_{141} = -0.825, $p = 0.4110$). For this reason, the remaining analyses were conducted using data from all six blocks of trials.

Regression Models

Model selection. Although multiple regression techniques were considered, it was eventually decided to use a logistic regression model which included a natural response frequency of the form:

$$P(A) = C + (1 - C)*(\frac{e^{f(x)}}{1 + e^{f(x)}})$$

where: P(A) = the area under the ROC curve,

C = the natural response frequency of the model,

$f(x) = b_0 + b_1x_1 + \cdots + b_nx_n$,

and the $x_1, x_2, \cdots x_n$ terms represent the hearing level (HL), noise level (NL), and signal level (SL) terms as well as the two- and three-way interaction terms. (For the regression model, *signal level* rather than *signal-to-noise ratio* was used because it was believed that use of signal level would be more intuitive for the end user of the model.)

A logistic model of the form shown above was chosen for several reasons. First, psychometric functions obtained in auditory experiments often exhibit a sigmoid shape such as that produced by logistic regression models and, upon inspection, the raw data did appear to exhibit such a shape in several conditions. Secondly, although threshold theory attributes the sigmoid shape of auditory psychometric functions to the underlying normal distribution of the random variation in the instantaneous threshold (Gescheider, 1985), the logistic distribution produces a very similar curve and is much easier to handle mathematically than is the normal distribution. The inclusion of the natural response term would allow the model to approach an asymptote at a value for P(A) of approximately 0.50, which is the minimum possible value of the P(A) term predicted by theory.

The regression analysis was performed using the SAS PROC NLIN (SAS, 1990) procedure. Both PROC NLIN and PROC PROBIT (SAS, 1990) could have been used in the analysis, but prediction limits could only be obtained through the use of PROC NLIN. However, PROC PROBIT was used to provide the initial estimates needed by PROC NLIN.

Alternate forms of the model. Although subjects were screened based on their pure-tone average hearing levels in the range from 500 to 2000 Hz, it did not mean that this range would be optimal for use in the regression analysis. It was therefore decided to investigate several methods for quantifying hearing level in the analysis and determine which method provided the best model. Six schemes for quantifying hearing level were investigated: 1) binaural PTA hearing level over the frequency range from 500 to 2000 Hz, 2) minimum (left or right) PTA hearing level over the frequency range from 500 to 2000 Hz, 3) binaural PTA hearing level over the frequency range from 500 to 4000 Hz, 4) minimum PTA hearing level over the frequency range from 500 to 4000 Hz, 5) binaural PTA hearing level over the frequency range from 1000 to 2000 Hz, 6) minimum PTA hearing level over the frequency range from 1000 to 2000 Hz. In all cases, the binaural PTA was calculated by multiplying the PTA in the better ear by five, adding this to the PTA of the poorer ear and dividing the result by six. This method of calculating binaural PTA hearing level was used because it emphasizes that both ears are important in an auditory detection task, but that the better hearing ear is *more* important when a large binaural difference exists. This is similar to the methodology used for determining binaural hearing impairment discussed by Miller and Wilber (1991). Finally, in calculating PTA hearing levels, audiometric data for 1500 Hz were included since the alarm used in the experiment had most of its energy in that vicinity.

The first step used in the regression analysis was to fit the full model (which included all of the main effect and interaction terms as well as the intercept term) for each of the six hearing level schemes mentioned earlier, and examine the mean square error (MSE). The results are shown in Table 17. As can be seen, the MSE for the models using the PTA over the frequency range from 500 to 4000 Hz are nearly double the MSE for the other four models. It was therefore decided that these two models would be dropped and only the four models with the lowest MSE would be pursued further.

Model refinement. The next step was to refine the models by eliminating any main effect and/or interaction terms which could be dropped with a minimum impact on the models' usability. Elimination of unnecessary terms makes the resulting models easier to understand and explain, simpler to implement, and also reduces the variance of the predictions (Montgomery and Peck, 1982). [The natural response term, C, was never considered a candidate for exclusion. To do so would have forced the models to approach an asymptote at zero, rather than at a level near P(A) = 0.5.] A backward elimination procedure (Neter, Wasserman, and Kutner, 1989) was used in the model

TABLE 17

MSE Fitting Full Model for each Method of Quantifying Hearing Level

Scheme used to Quantify Hearing Level	Mean Square Error, Full Model
Binaural Pure-Tone Average, 500 – 4000 Hz	0.008156362
Minimum Pure-Tone Average, 500 – 4000 Hz	0.008109418
Binaural Pure-Tone Average, 500 – 2000 Hz	0.004314772
Minimum Pure-Tone Average, 500 – 2000 Hz	0.004808894
Binaural Pure-Tone Average, 1000 – 2000 Hz	0.004221177
Minimum Pure-Tone Average, 1000 – 2000 Hz	0.00444377

refinement process. Exactly the same procedure was used simultaneously for each of the models based on each alternative method of quantifying hearing level.

The first step in the process was to drop each of the main effect and interaction terms (i.e., HL, SL, NL, HL*SL, etc.) individually, examine the resulting MSEs and determine which term(s) could be dropped with minimal change in the MSE. The MSEs (sorted in ascending order) for each of the models with a single term eliminated appear in Table 18. (In Tables 18 through 22, B0 refers to the intercept term of the regression equation. Main effect and interaction terms are given explicitly.)

The versions of the models with the minimum MSE (those appearing Table 18 above the double horizontal line) were carried to the next level in the analysis. This involved taking each of these versions of the models and dropping each of the remaining terms, resulting in a set of models with two terms eliminated, and examining the resulting MSEs. The results (again, sorted in ascending order by MSE) appear in Table 19.

This procedure was repeated until a total five terms had been dropped from the models for each alternative method of quantifying hearing level. The sorted MSEs for each of these subsequent stages in the analysis appear in Tables 20 though 22. (As in Tables 18 and 19, the versions of the models carried to the next level in the analysis appear above the double horizontal line.) At the fourth level in the analysis, the MSE had increased by as much as 16%, which was considered to be excessive, and the process could have been terminated at this point. However, it was still desired to examine the MSEs at the next level, so the single version of the models with the minimum MSE after dropping four terms was carried over to the next level in which five terms were dropped.

Plots of minimum MSE at each stage of the analysis versus the number of terms dropped from the model were then examined to determine at which point dropping additional terms caused an adverse increase in the MSE (appearing as a knee in the plot).

TABLE 18

Sorted MSEs (Ascending Order) for each Alternative Model Dropping One Term from the Model

Binaural PTA 500-2000 Hz		Minimum Monaural PTA 500-2000 Hz		Binaural PTA 1000-2000 Hz		Minimum Monaural PTA 1000-2000 Hz	
Terms Dropped	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**
B0	0.0042882000	NL	0.0047737620	NL	0.0041948540	NL	0.0044157620
HL*NL	0.0042950230	HL*NL	0.0047866720	B0	0.0042098960	B0	0.0044301070
NL*SL	0.0043196240	B0	0.0047961970	NL*SL	0.0042365740	HL*NL	0.0044482860
NL	0.0043198120	NL*SL	0.0048033170	HL*NL	0.0042370310	NL*SL	0.0044537140
HL	0.0046647700	HL	0.0048062610	HL	0.0043237410	HL	0.0044630180
HL*NL*SL	0.0046818790	HL*NL*SL	0.0048218000	HL*NL*SL	0.0043368060	HL*NL*SL	0.0044740270
SL	0.0047223280	HL*SL	0.0048622360	SL	0.0043737350	SL	0.0045104140
HL*SL	0.0047277320	SL	0.0048624030	HL*SL	0.0043832190	HL*SL	0.0045160160

TABLE 19

Sorted MSEs (Ascending Order) for each Alternative Model Dropping Two Terms from the Model

Binaural PTA 500-2000 Hz		Minimum Monaural PTA 500-2000 Hz		Binaural PTA 1000-2000 Hz		Minimum Monaural PTA 1000-2000 Hz	
Terms Dropped	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**
B0, HL*NL	0.0043261160	B0, HL*NL	0.0047606950	B0, HL*NL	0.0042064690	B0, HL*NL	0.0044156649
HL*NL, NL*SL	0.0043980050	NL, HL*NL	0.0047626320	HL*NL, NL*SL	0.0042138710	HL*NL, NL*SL	0.0044207740
NL, HL*NL	0.0046229100	HL*NL, NL*SL	0.0047686640	NL*SL, HL	0.0042918760	NL*SL, HL	0.0044299900
B0, HL	0.0046303560	B0, HL	0.0047708430	B0, HL	0.0042964430	B0, HL	0.0044334550
NL*SL, HL	0.0046312600	NL*SL, HL	0.0047731840	NL*SL HL*NL*SL	0.0043124870	NL*SL HL*NL*SL	0.0044471950
NL*SL HL*NL*SL	0.0046497140	NL*SL HL*NL*SL	0.0047868960	NL, HL*NL	0.0043198600	NL, HL*NL	0.0044598130
B0, HL*NL*SL	0.0046573070	B0, HL*NL*SL	0.0047913660	B0, HL*NL*SL	0.0043285150	B0, HL*NL*SL	0.0044610830
NL, HL	0.0046811180	NL, HL	0.0048065380	HL*NL, SL	0.0043677890	NL, HL	0.0045201770
HL*NL, SL	0.0047169990	NL*SL, HL*SL	0.0048502790	NL, HL	0.0043928200	HL*NL, SL	0.0045378900
NL*SL, HL*SL	0.0047257500	HL*NL, SL	0.0048607060	NL*SL, HL*SL	0.0044142820	NL*SL, HL*SL	0.0045398110
B0, HL*SL	0.0047456050	B0, HL*SL	0.0048664550	B0, HL*SL	0.0044538310	B0, HL*SL	0.0045761130
B0, NL*SL	0.0047644280	NL, HL*NL*SL	0.0048842640	NL, HL*NL*SL	0.0045100090	NL, HL*NL*SL	0.0046302700
NL, HL*NL*SL	0.0047681590	B0, NL*SL	0.0049042790	B0, NL*SL	0.0045543600	B0, NL*SL	0.0046992510
NL, HL*SL	0.0049043040	NL, HL*SL	0.0050119850	NL, HL*SL	0.0047356930	NL, HL*SL	0.0048518510
HL*NL, HL*NL*SL	0.0050829670	HL*NL, HL*NL*SL	0.0052072170	HL*NL, HL*NL*SL	0.0050178860	HL*NL, HL*NL*SL	0.0051576630
HL*NL, HL*SL	0.0057585030	HL*NL, HL*SL	0.0058798810	HL*NL, HL*SL	0.0061024870	HL*NL, HL*SL	0.0062428970
HL*NL, HL	0.0072453590	HL*NL, HL	0.0074212160	HL*NL, HL	0.0079257570	HL*NL, HL	0.0080673730
NL, NL*SL	0.0077561960	NL, NL*SL	0.0079506380	NL, NL*SL	0.0081450200	NL, NL*SL	0.0081101770
B0, SL	0.0077972620	B0, SL	0.0079675250	B0, SL	0.0082437070	B0, SL	0.0082293690
NL, SL	0.0114986450	NL, SL	0.0116665730	NL, SL	0.0126115970	NL, SL	0.0128173130
B0, NL	0.0178507700	NL*SL, SL	0.0199063630	B0, NL	0.0182923500	B0, NL	0.0186101030
NL*SL, SL	0.0195969110	B0, NL	0.1814841900	NL*SL, SL	0.0200328110	NL*SL, SL	0.0203684550

TABLE 20

Sorted MSEs (Ascending Order) for each Alternative Model Dropping Three Terms from the Model

Binaural PTA 500-2000 Hz		Minimum Monaural PTA 500-2000 Hz		Binaural PTA 1000-2000 Hz		Minimum Monaural PTA 1000-2000 Hz	
Terms Dropped	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**
B0, HL*NL, NL*SL	0.0047373790	B0, HL*NL, NL*SL	0.0048734450	B0, HL*NL, NL*SL	0.0045215900	B0, HL*NL, NL*SL	0.0046648830
B0, HL, NL*SL	0.0047727930	B0, HL, NL*SL	0.0049035700	B0, HL, NL*SL	0.0045553740	B0, HL, NL*SL	0.0046937710
B0, NL*SL, HL*NL*SL	0.0047860810	B0, NL*SL, HL*NL*SL	0.0049160650	B0, NL*SL, HL*NL*SL	0.0045672950	B0, NL*SL, HL*NL*SL	0.0047048670
B0, HL, HL*NL*SL	0.0049293610	B0, HL, HL*NL*SL	0.0050683670	B0, HL, HL*NL*SL	0.0047021040	B0, HL, HL*NL*SL	0.0048444370
NL*SL, HL, HL*NL*SL	0.0049559870	NL*SL, HL, HL*NL*SL	0.0050953000	NL*SL, HL, HL*NL*SL	0.0047290750	NL*SL, HL, HL*NL*SL	0.0048717560
B0, HL*NL, HL*NL*SL	0.0059689270	B0, HL*NL, HL*NL*SL	0.0061072390	B0, HL*NL, HL*NL*SL	0.0054306500	B0, HL*NL, HL*NL*SL	0.0056061140
HL*NL, NL*SL, HL*NL*SL	0.0062407920	HL*NL, NL*SL, HL*NL*SL	0.0063671610	HL*NL, NL*SL, HL*NL*SL	0.0060580740	HL*NL, NL*SL, HL*NL*SL	0.0061266974
NL, HL*NL, HL*NL*SL	0.0067114290	NL, HL*NL, HL*NL*SL	0.0068340590	NL, HL*NL, HL*NL*SL	0.0067474770	B0, HL, HL*SL	0.0068975680
B0, HL, HL*SL	0.0067681620	B0, HL, HL*SL	0.0068845880	B0, HL, HL*SL	0.0068898270	NL, HL*NL, HL*NL*SL	0.0069710090
NL*SL, HL, HL*SL	0.0071815230	NL*SL, HL, HL*SL	0.0072881960	NL*SL, HL, HL*SL	0.0076911390	NL*SL, HL, HL*SL	0.0078067770
NL, HL*NL, HL	0.0071921170	NL, HL*NL, HL	0.0073666510	NL, HL*NL, HL	0.0078683220	NL, HL*NL, HL	0.0080088180
B0, HL*NL, HL	0.0072078600	B0, HL*NL, HL	0.0073809080	B0, HL*NL, HL	0.0078873540	B0, HL*NL, HL	0.0080273990
HL, HL*NL NL*SL	0.0072130910	HL, HL*NL, NL*SL	0.0073853850	HL, HL*NL, NL*SL	0.0078919850	HL, HL*NL, NL*SL	0.0080317060
B0, HL*NL, HL*SL	0.0075105170	B0, HL*NL, HL*SL	0.0076295240	B0, HL*NL, HL*SL	0.0081321720	B0, HL*NL, HL*SL	0.0082541880
HL*NL, NL*SL, HL*SL	0.0077611420	HL*NL, NL*SL, HL*SL	0.0078849160	HL*NL, NL*SL, HL*SL	0.0083994500	HL*NL, NL*SL, HL*SL	0.0085175810
NL, HL*NL, HL*SL	0.0081006950	NL, HL*NL, HL*SL	0.0082189910	NL, HL*NL, HL*SL	0.0087120550	NL, HL*NL, HL*SL	0.0088284580
NL*SL, HL*NL*SL, NL	0.0087111290	NL*SL, HL*NL*SL, NL	0.0088128080	NL*SL, HL*NL*SL, NL	0.0089568150	B0, HL*NL, SL	0.0090799000

TABLE 20 (continued)

Sorted MSEs (Ascending Order) for each Alternative Model Dropping Three Terms from the Model

Binaural PTA 500-2000 Hz		Minimum Monaural PTA 500-2000 Hz		Binaural PTA 1000-2000 Hz		Minimum Monaural PTA 1000-2000 Hz	
Terms Dropped	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**
NL*SL, HL*SL, HL*NL*SL	0.0087278600	B0, HL*NL, SL	0.0088855430	B0, HL*NL, SL	0.0089643290	NL*SL, HL*NL*SL, NL	0.0090809150
B0, HL*NL, SL	0.0087897860	NL*SL, HL*SL, HL*NL*SL,	0.0088904070	NL, HL*NL, NL*SL	0.0091019430	NL, HL*NL, NL*SL	0.0092252060
NL, HL*NL, NL*SL	0.0088717660	NL, HL*NL, NL*SL	0.0089675330	B0, HL, SL	0.0091660810	B0, HL, SL	0.0092786210
B0, HL, SL	0.0089394980	B0, HL, SL	0.0090282170	NL*SL, HL*SL, HL*NL*SL	0.0091943890	NL*SL, HL*SL, HL*NL*SL,	0.0093229510
NL*SL, HL, NL	0.0090231950	NL*SL, HL, NL	0.0091124660	NL*SL, HL, NL	0.0093142140	NL*SL, HL, NL	0.0094299610
NL, HL*NL, SL	0.0129848250	NL, HL*NL, SL	0.0130592880	NL, HL*NL, SL	0.0134588990	NL, HL*NL, SL	0.0135739560
B0, HL, NL	0.0177232580	B0, HL, NL	0.0180189400	B0, HL, NL	0.0181604820	B0, HL, NL	0.0184762830
B0, HL*NL, NL	0.0177486670	B0, HL*NL, NL	0.0180450400	B0, HL*NL, NL	0.0181828030	B0, HL*NL, NL	0.0184996840
NL*SL, HL, SL	0.0194591310	NL*SL, HL, SL	0.0197655480	NL*SL, HL, SL	0.0198893590	NL*SL, HL, SL	0.0202221540
NL*SL, HL*NL*SL, SL	0.0194683570	NL*SL, HL*NL*SL, SL	0.0197742620	NL*SL, HL*NL*SL, SL	0.0198961290	NL*SL, HL*NL*SL, SL	0.0202286600
SL, HL*NL, NL*SL	0.0194908880	SL, HL*NL, NL*SL	0.0197966260	SL, HL*NL, NL*SL	0.0199157320	SL, HL*NL, NL*SL	0.0202483930

TABLE 21

Sorted MSEs (Ascending Order) for each Alternative Model Dropping Four Terms from the Model

Binaural PTA 500-2000 Hz		Minimum Monaural PTA 500-2000 Hz		Binaural PTA 1000-2000 Hz		Minimum Monaural PTA 1000-2000 Hz	
Terms Dropped	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**
B0, HL, NL*SL, HL*NL*SL	0.0050110670	B0, HL, NL*SL, HL*NL*SL	0.0051394790	B0, HL, NL*SL, HL*NL*SL	0.0047955330	B0, HL, NL*SL, HL*NL*SL	0.0049273180
B0, HL*NL*SL, HL*NL, NL*SL	0.0068639510	B0, HL*NL*SL, HL*NL, NL*SL	0.0069984410	B0, HL*NL*SL, HL*NL, NL*SL	0.0076121070	B0, HL*NL*SL, HL*NL, NL*SL	0.0077272130
B0, HL, NL*SL, HL*NL	0.0072340610	B0, HL, NL*SL, HL*NL	0.0073927120	B0, HL, NL*SL, HL*NL	0.0078830870	B0, HL, NL*SL, HL*NL	0.0080174130
B0, HL, HL*NL, HL*NL*SL	0.0073764180	B0, HL, HL*NL, HL*NL*SL	0.0075325670	B0, HL, HL*NL, HL*NL*SL,	0.0080527890	B0, HL, HL*NL, HL*NL*SL	0.0081853830
NL*SL, HL*NL, HL, HL*NL*SL	0.0073955730	NL*SL, HL*NL HL, HL*NL*SL	0.0075524680	NL*SL, HL*NL, HL, HL*NL*SL	0.0080774750	NL*SL, HL*NL, HL, HL*NL*SL	0.0082096420
B0, HL*NL, NL*SL, HL*SL	0.0081244230	B0, HL, NL*SL, HL*SL	0.0082651110	B0, HL*NL, NL*SL, HL*SL	0.0087554200	B0, HL, NL*SL, HL*SL	0.0088728140
B0, HL, NL*SL, HL*SL	0.0081431740	B0, HL*NL, NL*SL, HL*SL	0.0082739750	B0, HL, NL*SL, HL*SL	0.0087751360	B0, HL*NL, NL*SL, HL*SL	0.0088740170
NL*SL, HL*SL, HL*NL*SL, B0	0.0086995560	NL*SL, HL*SL, HL*NL*SL, B0	0.0088542910	NL*SL, HL*SL, HL*NL*SL, B0	0.0091640020	NL*SL, HL*SL, HL*NL*SL, B0	0.0092870400
B0, HL, HL*SL, HL*NL*SL	0.0088391620	B0, HL, HL*SL, HL*NL*SL,	0.0089893820	B0, HL, HL*SL, HL*NL*SL,	0.0093256800	B0, HL, HL*SL, HL*NL*SL,	0.0094458660
NL*SL, HL*SL, HL, HL*NL*SL	0.0088577570	NL*SL, HL*SL, HL, HL*NL*SL	0.0090076210	NL*SL, HL*SL, HL, HL*NL*SL	0.0093455820	NL*SL, HL*SL, HL, HL*NL*SL	0.0094651530
NL*SL, HL, HL*NL*SL, NL	0.0111929800	NL*SL, HL, HL*NL*SL, NL	0.0113071240	NL*SL, HL, HL*NL*SL, NL	0.0116486770	NL*SL, HL, HL*NL*SL, NL	0.0117679230
B0, HL, SL, HL*NL*SL	0.0121847210	B0, HL, SL, HL*NL*SL	0.0122975550	B0, HL, SL, HL*NL*SL	0.0125750900	B0, HL, SL, HL*NL*SL	0.0126842670
B0, HL, NL, NL*SL	0.0176286550	B0, HL, NL, NL*SL	0.0179185660	B0, HL, NL, NL*SL	0.0180620040	B0, HL, NL, NL*SL	0.0183723920
NL*SL, B0, NL, HL*NL*SL	0.0176355250	NL*SL, B0, NL, HL*NL*SL	0.0179252030	NL*SL, B0, NL, HL*NL*SL	0.0180674790	NL*SL, B0, NL, HL*NL*SL	0.0183778250
B0, HL*NL, NL*SL, NL	0.0176645600	B0, HL*NL, NL*SL, NL	0.0179543340	B0, HL*NL, NL*SL, NL	0.0180918840	B0, HL*NL, NL*SL, NL	0.0184026470
B0, HL, NL, HL*NL*SL	0.0177251940	B0, HL, NL, HL*NL*SL	0.0180065590	B0, HL, NL, HL*NL*SL	0.0181633070	B0, HL, NL, HL*NL*SL	0.0184662740

TABLE 21 (continued)

Sorted MSEs (Ascending Order) for each Alternative Model Dropping Four Terms from the Model

Binaural PTA 500-2000 Hz		Minimum Monaural PTA 500-2000 Hz		Binaural PTA 1000-2000 Hz		Minimum Monaural PTA 1000-2000 Hz	
Terms Dropped	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**
B0, HL, SL NL*SL	0.0193442790	B0, HL, SL NL*SL	0.0196460330	B0, HL, SL NL*SL	0.0197682650	B0, HL, SL NL*SL	0.0200964880
NL*SL, B0, SL, HL*NL*SL	0.0193522380	NL*SL, B0, SL, HL*NL*SL	0.0196535680	NL*SL, B0, SL, HL*NL*SL	0.0197742650	NL*SL, B0, SL, HL*NL*SL	0.0201022540
B0, HL*NL, NL*SL, SL	0.0193898730	B0, HL*NL, NL*SL, SL	0.0196908410	B0, HL*NL, NL*SL, SL	0.0198043040	B0, HL*NL, NL*SL, SL	0.0201322460
NL*SL, HL, HL*NL*SL, SL	0.0194448310	NL*SL, HL, HL*NL*SL, SL	0.0197383550	NL*SL, HL, HL*NL*SL, SL	0.0198720720	NL*SL, HL, HL*NL*SL, SL	0.2019314400

TABLE 22

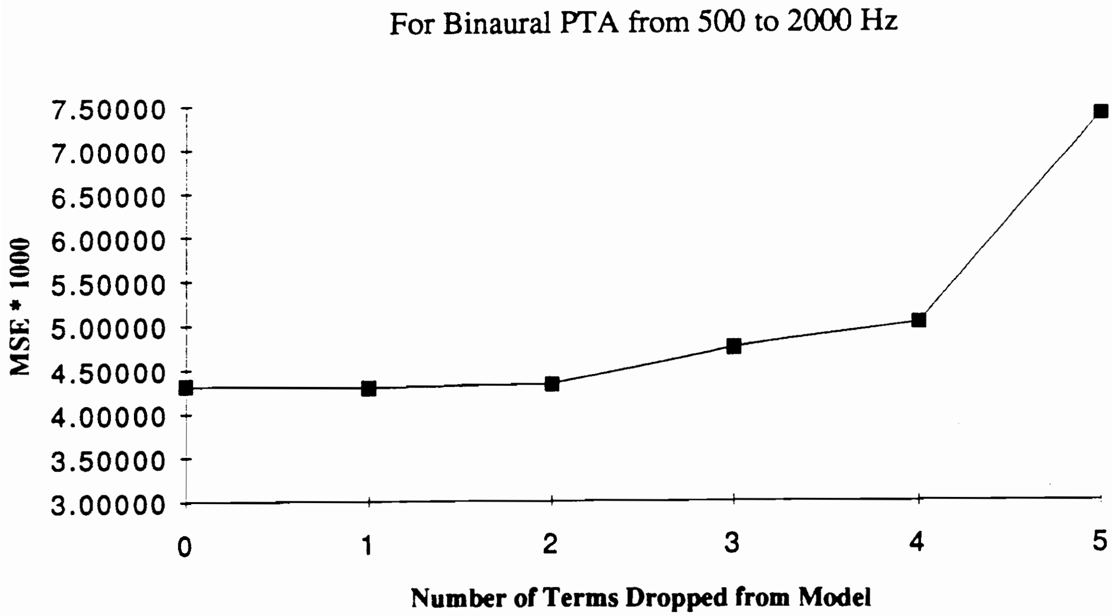
Sorted MSEs (Ascending Order) for each Alternative Model Dropping Five Terms from the Model

Binaural PTA 500-2000 Hz		Minimum Monaural PTA 500-2000 Hz		Binaural PTA 1000-2000 Hz		Minimum Monaural PTA 1000-2000 Hz	
Terms Dropped	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**	**Terms Dropped**	**MSE**
B0, HL, NL*SL, HL*NL*SL, HL*NL	0.00737532	B0, HL, NL*SL, HL*NL*SL, HL*NL	0.007537919	B0, HL, NL*SL, HL*NL*SL, HL*NL	0.008085813	B0, HL, NL*SL, HL*NL*SL, HL*NL	0.008217047
B0, HL, HL*NL*SL, HL*SL, NL*SL,	0.008857577	B0, HL, NL*SL, HL*NL*SL, HL*SL	0.009009012	B0, HL, NL*SL, HL*NL*SL, HL*SL	0.009344454	B0, HL, NL*SL, HL*NL*SL, HL*SL	0.009461989
B0, HL, NL*SL, HL*NL*SL, NL	0.017664728	B0, HL, NL*SL, HL*NL*SL, NL	0.017941442	B0, HL, NL*SL, HL*NL*SL, NL	0.018090041	B0, HL, NL*SL, HL*NL*SL, NL	0.018388557
B0, HL, NL*SL, HL*NL*SL, SL	0.019370881	B0, HL, NL*SL, HL*NL*SL, SL	0.019660789	B0, HL, NL*SL, HL*NL*SL, SL	0.01978631	B0, HL, NL*SL, HL*NL*SL, SL	0.020104122

These plots are shown in Figure 33 for each of the four candidate schemes of quantifying hearing level.

Upon examination, the plots seem to indicate that four parameters could be dropped without an excessive increase in the MSE. However, due to the large increase in the MSE between cases in which four terms were dropped and those in which five terms were dropped, it was decided to drop the last data point (representing the cases in which five terms were dropped) and plot the data using an expanded ordinal scale to determine if the compressed scale in the previous plots concealed another knee in the graph. These plots are shown in Figure 34. As can be seen, there does indeed appear to be a knee in the graph at the point where two terms were dropped from the model.

At this point, it was decided to develop plots for the models in which two terms were dropped and see how they fit the experimental data. Upon examination, these plots appeared to fit the data "too well" as can be seen from the example plots in Figure 35. It is believed that the direct search procedure used in the regression analysis was producing a situation similar to that which occurs in polynomial regression in which a "perfect" fit may be obtained by including one less term in the model than there are data points (Montgomery and Peck, 1982). It was therefore decided to examine the graphs for the models in which three terms were dropped and see if the fit were more reasonable. Upon inspection, the plots did appear to provide a more reasonable fit to the experimental data as can be seen in Figure 36, in which the same conditions appearing in Figure 35 are plotted using the newer models. By "reasonable," it is meant that the quick decrease in P(A) with a small increase in hearing level appearing in Figure 35 is not supported by the data and that the curves represented in Figure 36 appear to fit the experimental data better. However, this is somewhat speculative because of the gap in the data due to the missing subjects.



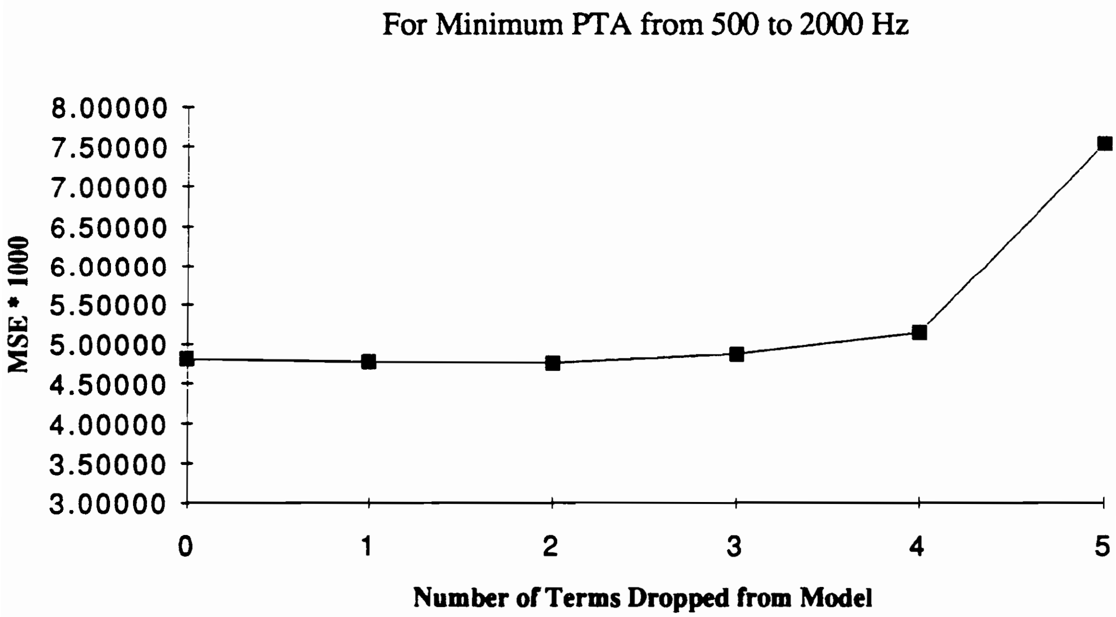
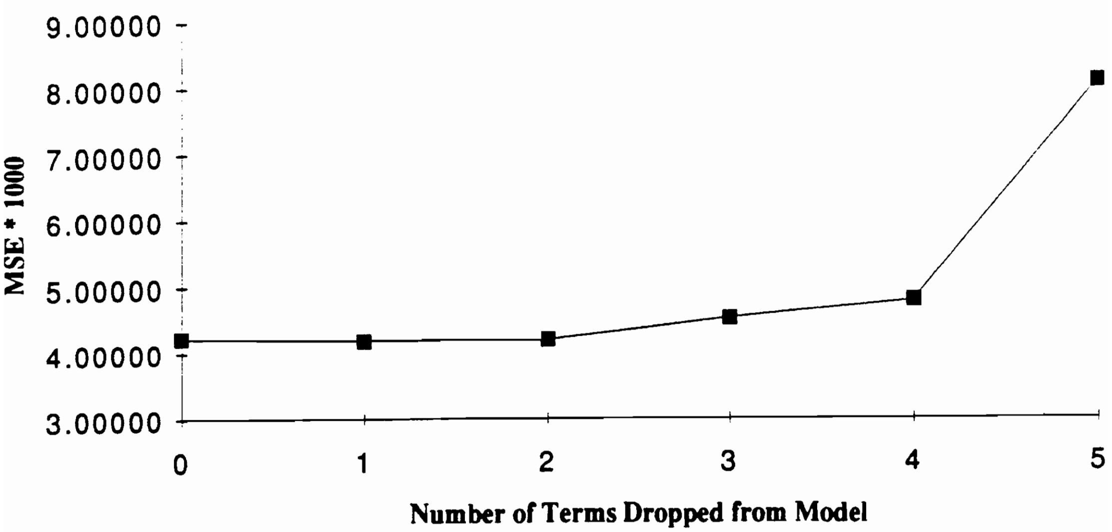


Figure 33. *Plots of minimum MSE versus number of terms dropped from the model.*

For Binaural PTA from 1000 to 2000 Hz



For Minimum PTA from 1000 to 2000 Hz

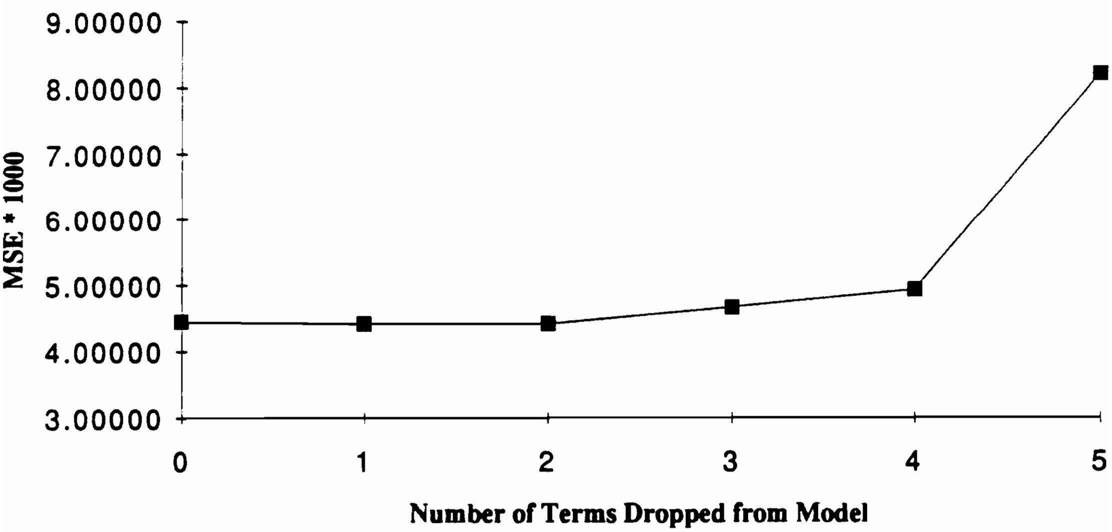
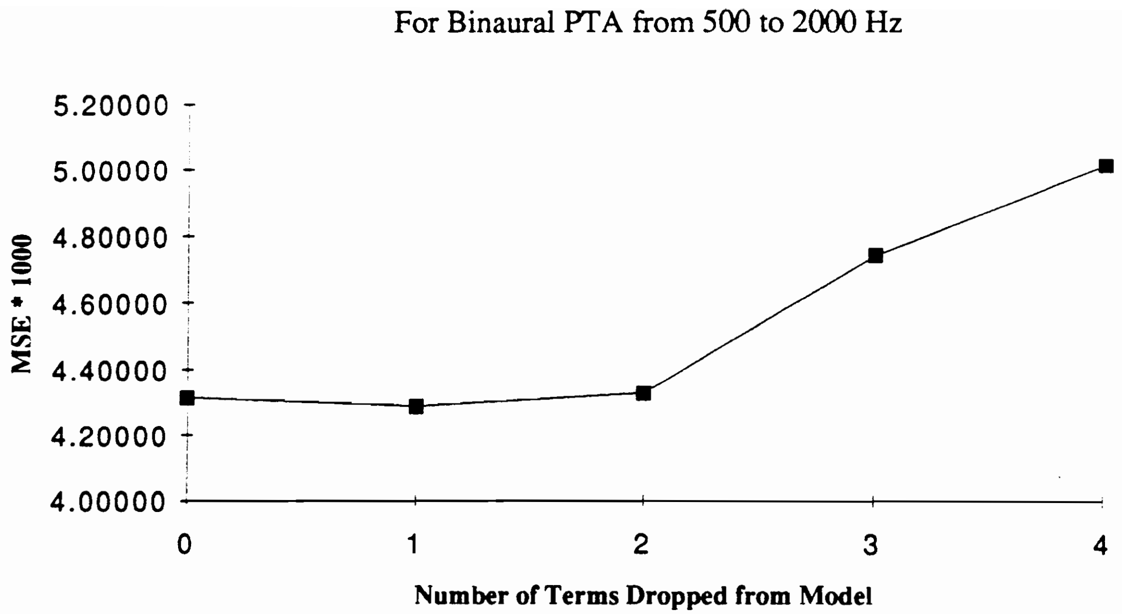


Figure 33 (continued). *Plots of minimum MSE versus number of terms dropped from the model.*



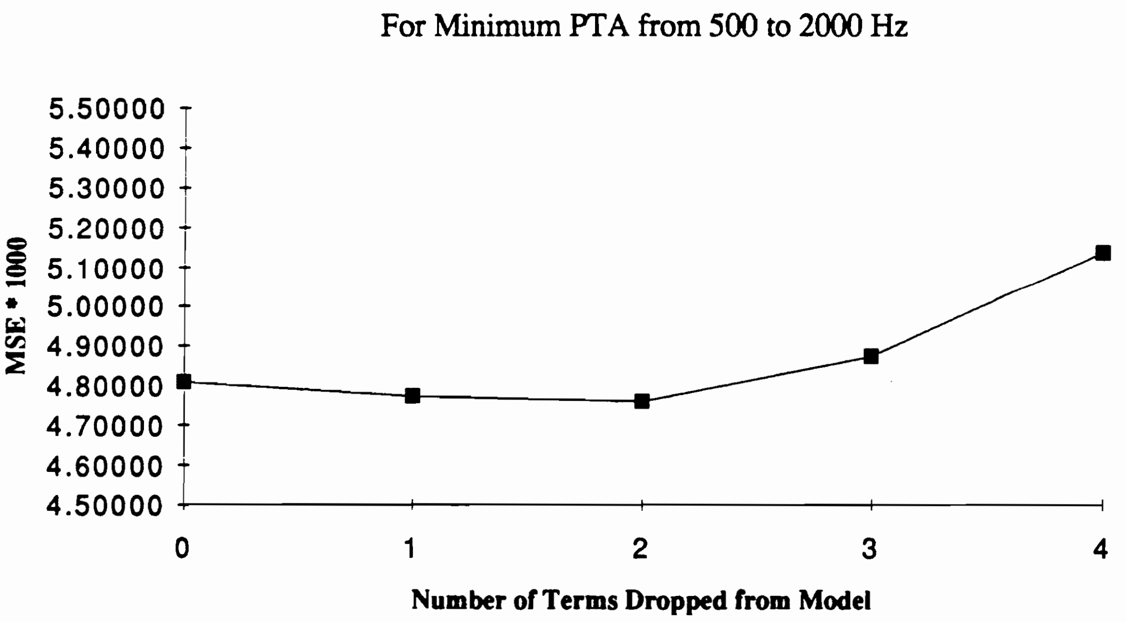
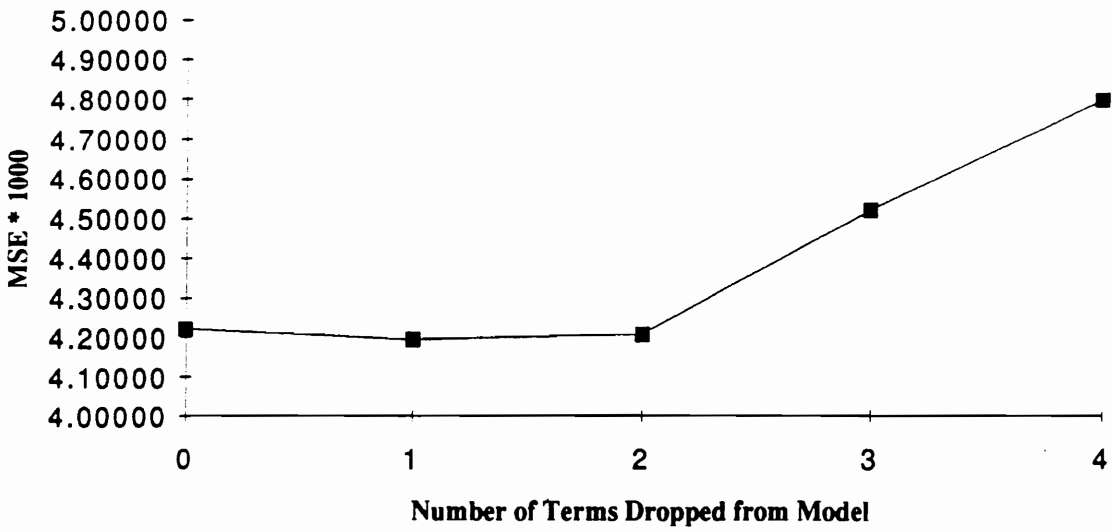


Figure 34. *Plots of minimum MSE versus number of terms dropped from model using an expanded ordinal scale and dropping the last data point.*

For Binaural PTA from 1000 to 2000 Hz



For Minimum PTA from 1000 to 2000 Hz

Figure 34 (continued). *Plots of minimum MSE versus number of terms dropped from model using an expanded ordinal scale and dropping the last data point.*

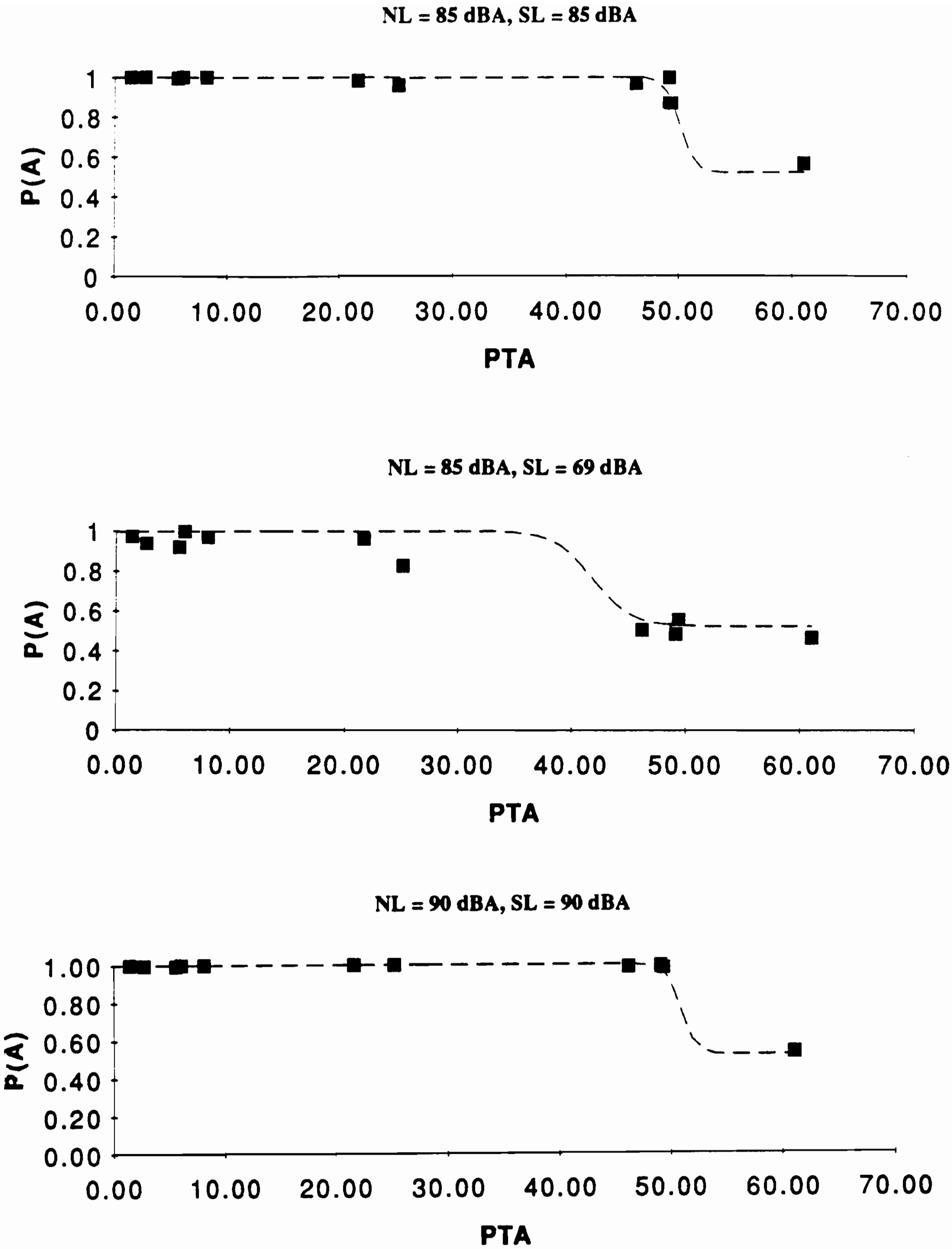
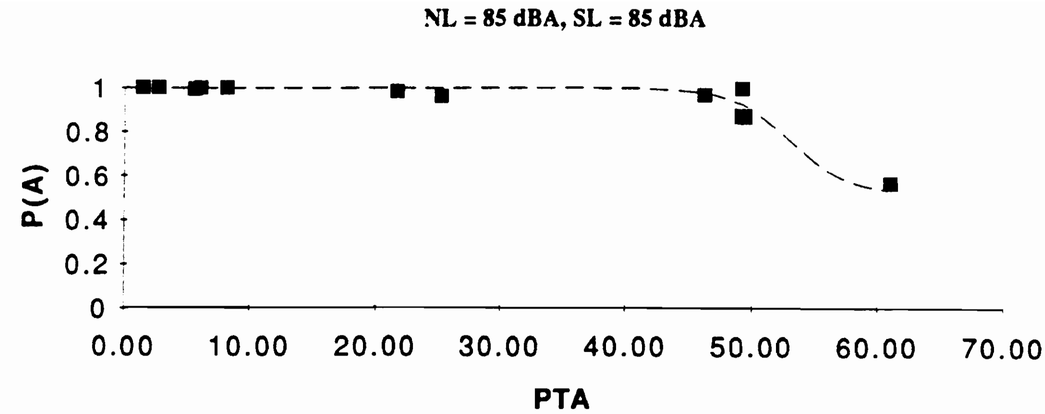
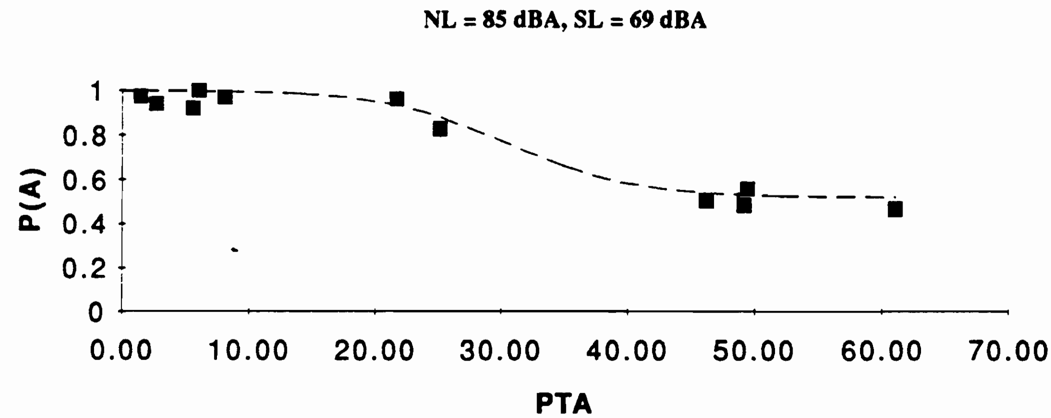


Figure 35. *Example plots of the regression equation (dotted line) and the experimental data with two terms dropped from the model (i.e., the nearly-"perfect" fit explained in the text).*





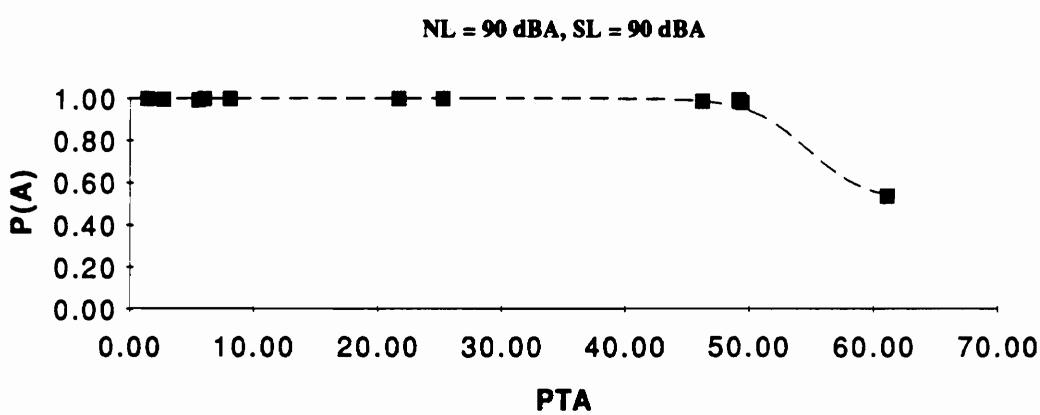


Figure 36. *Corresponding plots of the regression equation (dotted line) and the experimental data with three terms dropped from the model (i.e., the reasonable fit explained in the text).*

Final regression models. It was therefore decided to accept the models in which three terms were dropped as the final models. Fortunately, the models for each of the four schemes for quantifying hearing level included the same terms. This fact strengthens the argument that the procedures used in the model refinement process were correct since the same point was reached via four different routes. The terms included in the final models are: Hearing Level (HL), Noise Level (NL), Signal Level (SL), the HL × SL two-way interaction term, and the HL × NL × SL three-way interaction term. (The terms dropped from the models included the intercept term as well as the two two-way interactions of HL × NL and NL × SL.) The equation of the final model is given below. Parameter estimates and MSE for each of the four models corresponding to each of the four schemes of quantifying hearing level are given in Table 23.

$$P(A) = C + (1 - C)*(\frac{e^{(b_1*HL+b_2*NL+b_3*SL+b_4*HL*SL+b_5*HL*NL*SL)}}{1+e^{(b_1*HL+b_2*NL+b_3*SL+b_4*HL*SL+b_5*HL*NL*SL)}})$$

As can be seen in the table, the differences in magnitude of the MSE for the four candidate models are quite small. However, note that the MSE for the models using a binaural PTA are less than the corresponding models using the minimum PTA hearing levels. In addition, the MSE for the models considering only the frequencies from 1000 to 2000 Hz are consistently less than the corresponding models utilizing the broader frequency range of 500 to 2000 Hz. Plots of each model against the experimental data appear in Appendices A through D.

Confidence and prediction intervals may be obtained for the given models, but to do so requires rather intensive calculations using the partial derivatives of the model with respect to each estimated parameter (b_n) evaluated at the point of interest. However, the formulae (Myers, 1990) necessary for determining these intervals are given below so they may be calculated by interested readers.

TABLE 23

Parameter Estimates and MSE for the Final Models

Method used to Quantify Hearing Level	Parameter	Parameter Estimate
Binaural PTA,	C	0.51561123
500 – 2000 Hz	b_1	0.778290188
(MSE = 0.004737379)	b_2	-0.762423071
	b_3	1.026309075
	b_4	-0.025340819
	b_5	0.000131875
Minimum	C	0.516859065
Monaural PTA,	b_1	0.754151889
500 – 2000 Hz	b_2	-0.749209661
(MSE = 0.004873445)	b_3	1.007393843
	b_4	-0.024797419
	b_5	0.000129168
Binaural PTA,	C	0.520382132
1000 – 2000 Hz	b_1	1.067938668
(MSE = 0.004521590)	b_2	-0.98546813
	b_3	1.325599931
	b_4	-0.033250017
	b_5	0.00016938
Minimum	C	0.523284916
Monaural PTA,	b_1	1.085218025
1000 – 2000 Hz	b_2	-1.012811967
(MSE = 0.004664883)	b_3	1.360911799
	b_4	-0.033964061
	b_5	0.000172247

For the 100(1 - α)% confidence interval on the mean response at the point of interest:

$$\pm t_{\alpha/2, n-p-1} MSE \sqrt{w_0' (W' W)^{-1} w_0}$$

and, for the 100(1 - α)% prediction limits for a new observation at the point of interest:

$$\pm t_{\alpha/2, n-p-1} MSE \sqrt{1 + w_0' (W' W)^{-1} w_0}$$

where: n = number of observations in the data (142),

p = number of parameters estimated in the model (6),

w_0 = is a column vector of the partial derivatives of the model with respect to each of the parameters evaluated at the point of interest,

w_0' = the transpose of w_0,

W = the matrix of the partial derivatives of the model evaluated at each of the experimental data points, and

W' = the transpose of W.

The partial derivatives used in constructing the w_0 column vector are given below, the W'W matrices for each of the candidate models are given in Table 24.

$$\frac{\partial}{\partial C} = 1 - \left(\frac{1}{1 + e^{f(x)}}\right)$$

$$\frac{\partial}{\partial HL} = (1 - C) * \left[\frac{HL * e^{f(x)}}{(1 + e^{f(x)})^2}\right]$$

$$\frac{\partial}{\partial NL} = (1 - C) * \left[\frac{NL * e^{f(x)}}{(1 + e^{f(x)})^2}\right]$$

$$\frac{\partial}{\partial SL} = (1 - C)*[\frac{SL*e^{f(x)}}{(1 + e^{f(x)})^2}]$$

$$\frac{\partial}{\partial HL*SL} = (1 - C)*[\frac{HL*SL*e^{f(x)}}{(1 + e^{f(x)})^2}]$$

$$\frac{\partial}{\partial HL*NL*SL} = (1 - C)*[\frac{HL*NL*SL*e^{f(x)}}{(1 + e^{f(x)})^2}]$$

where: $f(x)$ = b1*HL+b2*NL+b3*SL+b4*HL*SL+b5*HL*NL*SL, and

C = the natural response for the model used.

Analysis of Variance

Although multivariate analysis of variance (MANOVA) was considered for the analysis as a means of controlling for an inflated Type I error in the repeated measures analysis, it was decided to perform the univariate analysis of variance (ANOVA) and adjust the numerator and denominator degrees of freedom using the Greenhouse-Geisser epsilon (ê). This was done since the multivariate test is discouraged when the number of subjects in a repeated measures analysis does not exceed the number of repeated measures (Vasey and Thayer, 1987). The Greenhouse-Geisser correction was chosen over the Huynh-Feldt correction since it has been shown to be the more conservative approach (Vasey and Thayer, 1987). Type III sums of squares were used since they are the appropriate sums of squares for use in an unbalanced design (Speed, Hocking, and Hackney, 1978) such as the one described herein (the middle hearing level group contains only two subjects, whereas the other two hearing level groups contain five subjects each).

In addition to being unbalanced, the data also contained two missing values since one subject had to withdraw from the experiment two sessions short of finishing for reasons unrelated to the experiment. It was decided to use the model developed earlier and described in the previous section to predict the performance of this subject in the two

TABLE 24

W'W Matrices for the Candidate Models

W'W when using Binaural PTA from 500 to 2000 Hz

74.01615497	54.76555151	144.5811128	126.7242662	4554.91024	410429.134
54.75555151	425.8414333	841.3037804	762.5550454	35208.43822	3177113.477
144.5811128	841.3037804	2447.757943	2074.939889	68821.25867	6223945.175
126.7242662	762.5550454	2074.939889	1778.41074	62607.451	56660096.596
4554.91024	35208.43822	68821.25867	62607.451	2922654.195	264234670.6
410429.134	3177113.477	6223945.175	5660096.596	26234670.6	23937816550

W'W when using Minimum PTA from 500 to 2000 Hz

73.83693668	52.94989817	144.8566527	126.7522671	4403.371467	396745.4472
52.94989817	402.7211333	809.1753653	734.5806846	33338.5969	3009254.964
144.8566527	809.1753653	2440.635959	2067.717842	66285.16388	5993622.471
126.7522671	734.5806846	2067.717842	1771.272767	60392.85459	5459532.015
4403.371467	33338.5969	66285.16388	60392.85459	2771401.406	250646998.8
396745.4472	3009254.964	5993622.471	5459532.015	250646998.8	22714917265

TABLE 24 (continued)

W'W Matrices for the Candidate Models

W'W when using Binaural PTA from 1000 to 2000 Hz

74.61351873	50.23316377	125.7863423	110.5174981	4189.893811	377792.0694
50.23316377	418.2329566	789.8205269	721.4116617	34760.17527	3133050.829
125.7863423	789.8205269	2188.569364	1868.335786	65036.53906	5875326.429
110.5174981	721.4116617	1868.335786	1613.059421	59617.61982	5384714.957
4189.893811	34760.17527	65036.53906	59617.61982	2901718.891	262090166.2
377792.0694	3133050.829	5875326.429	5384714.957	262090166.2	23721315731

W'W when using Minimum PTA from 1000 to 2000 Hz

74.56720939	46.44258351	120.8790473	105.9546131	3873.449196	349470.4814
46.44258351	381.8253648	734.7253609	672.5281908	31781.41896	2865155.484
120.8790473	734.7253609	2108.592231	1800.628618	60618.26459	5475119.536
105.9546131	672.5281908	1800.628618	1555.07758	55682.75387	5029079.477
3873.449196	31781.41896	60618.26459	55682.75387	2657529.778	240104420.7
349470.4814	2865155.484	5475119.536	5029079.477	240104420.7	21737562335

conditions in which he did not participate as suggested by Winer, Brown, and Michels (1991). This was not considered overly risky since the two conditions missed were the two highest (most likely detectable) signal-to-noise ratio conditions (S/N = 0 and -8 dB) at the 95 dBA noise level. In addition, the subject's performance in the next lowest signal-to-noise ratio condition (SN = -16 dB) at the same noise level was nearly perfect [P(A) = 0.9388]. The version of the model used quantified hearing level based on the binaural PTA in the range from 500 to 2000 Hz. This particular model was utilized simply because the method used in quantifying hearing level matched the original screening criteria.

The ANOVA summary table is presented in Table 25. As can be seen, significant (at an adjusted G-G $p < 0.05$) main effects and interactions included HL ($F = 18.81$, $p = 0.0006$), and SN ($F = 49.44$, $p = 0.0001$), and SN × HL ($F = 5.73$, $p = 0.0049$). Each of these effects will be discussed separately below with the discussion of the interaction presented first, followed by a discussion of the main effects.

SN × HL interaction. Post-hoc tests of the SN × HL interaction were conducted using simple-effect F-tests followed up with Student-Newman-Keuls tests to determine the locus of each of the simple main effects. The Student-Newman-Keuls procedure was chosen because it apportions α depending on how far apart the means being compared are in relation to the ordering of all means considered.

The simple-effect F-tests indicated that simple main effects of HL existed at signal-to-noise ratios of -8 ($F = 23.47$, $p < 0.0001$), -16 ($F = 39.57$, $p < 0.0001$), and -24 dB ($F = 4.65$, $p = 0.0184$); but not at 0 dB ($F = 2.72$, $p = 0.0840$). The Student-Newman-Keuls tests indicated that at signal-to-noise ratios of -8 and -16 dB, the individuals with the greatest hearing loss (group 3) performed significantly poorer than individuals in the other two groups. At a signal-to-noise ratio of -24 dB, individuals in both groups 2 and 3 performed significantly poorer than normal listeners, but their performance did not differ

TABLE 25

ANOVA Summary Table for the Experiment

Source	df	MS	*F*	*p*	G-G *p*
<u>Between-Subjects</u>					
Hearing Level (HL)	2	0.962635	18.81	0.0006	
Subjects S(HL)	9	0.0511898			
<u>Within-Subject</u>					
Noise Level (NL)	2	0.0102952	2.81	0.0869	0.0929
NL × HL	4	0.00329884	0.90	0.4849	0.4794
NL × S(HL)	18	0.00366835			
Greenhouse-Geisser Epsilon = 0.9155					
Signal-to-Noise Ratio (SN)	3	0.89453701	49.44	0.0001	0.0001
SN × HL	6	0.10374625	5.73	0.0006	0.0049
SN × S(HL)	27	0.01809457			
Greenhouse-Geisser Epsilon = 0.6156					
NL × SN	6	0.00235756	0.66	0.6828	0.6201
NL × SN × HL	12	0.00545753	1.53	0.1437	0.1858
NL × SN × S(HL)	54	0.00357752			
Greenhouse-Geisser Epsilon = 0.6461					
Total	143				

significantly from one another. These differences are illustrated in Figure 37. In the figure, levels of the HL variable with different letters are significantly different at a given level of SN. Levels labeled with the same letter at a given SN level do not differ significantly from one another.

Simple-effect F-tests were also conducted to determine if simple main effects of SN existed at any of the three levels of HL. Results revealed simple main effects of SN at all three levels of the HL variable: ($F = 23.01, p < 0.0001$) for normal listeners, ($F = 16.14, p < 0.0001$) for individuals with a slight hearing loss, and ($F = 25.34, p < 0.0001$) for listeners suffering mild to marked hearing loss. Post-hoc comparisons using the Student-Newman-Keuls procedure revealed that normal listeners and listeners with only a slight hearing loss exhibited significantly degraded performance only at a signal-to-noise ratio of -24 dB. However, listeners with mild to marked hearing loss showed significantly degraded performance at each level of signal-to-noise ratio with the exception that their performance at SN = -16 dB and SN = -24 dB did not differ significantly. These differences are illustrated in Figure 38. As before, levels of SN labeled with different letters at a given level of HL indicates a significant difference exists between the levels of SN.

Main effects. As stated earlier, significant main effects present in the analysis included hearing level ($F = 18.81$, $p = 0.0006$) and signal-to-noise ratio ($F = 49.44$, $p = 0.0001$). The main effect of SN is illustrated in Figure 39. As can be seen, post-hoc tests revealed that the subjects' performance dropped significantly as signal-to-noise ratio decreased. The main effect of hearing level is illustrated in Figure 40. Here, post-hoc tests indicated that the performance of the subjects suffering the greatest hearing loss was significantly poorer than the subjects in the other two groups, but the performance of the subjects with only a slight hearing loss did not differ significantly from that of the normal-hearing subjects. These main effects should be viewed with caution, however,

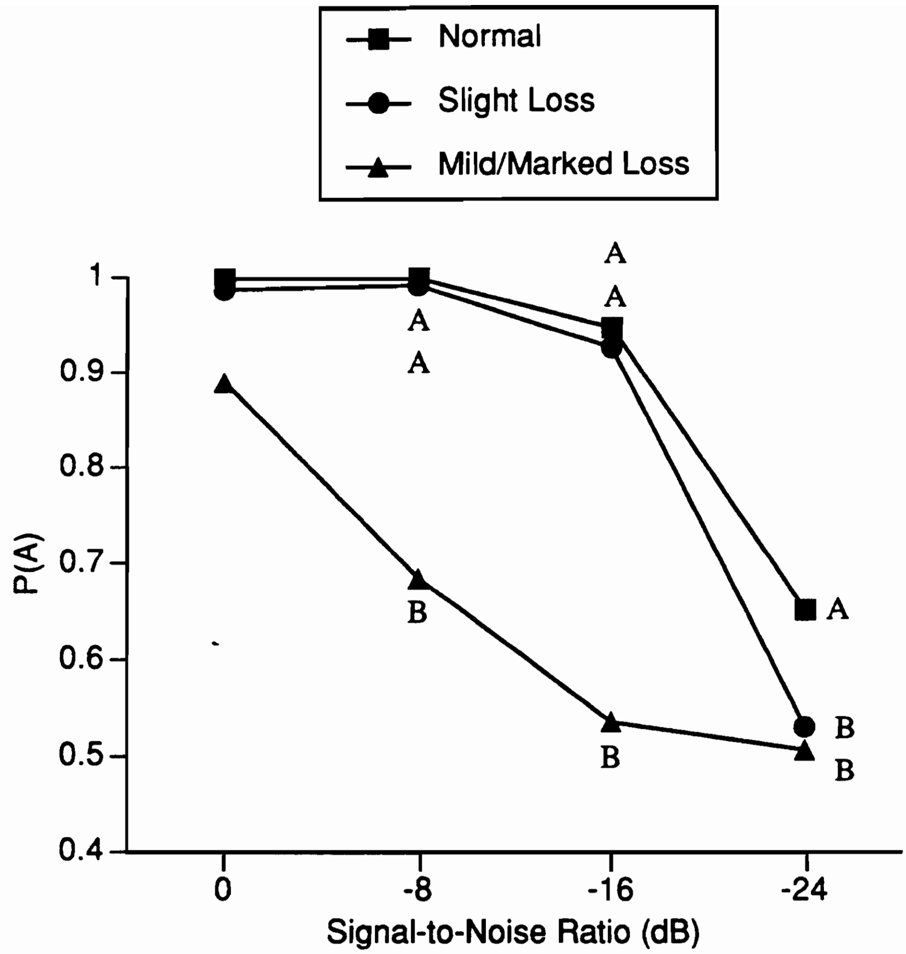


Figure 37. *Simple effects of HL at each level of SN. (Levels of the HL variable with different letters are significantly different at a given level of SN. Levels labeled with the same letter at a given SN level do not differ significantly from one another.)*

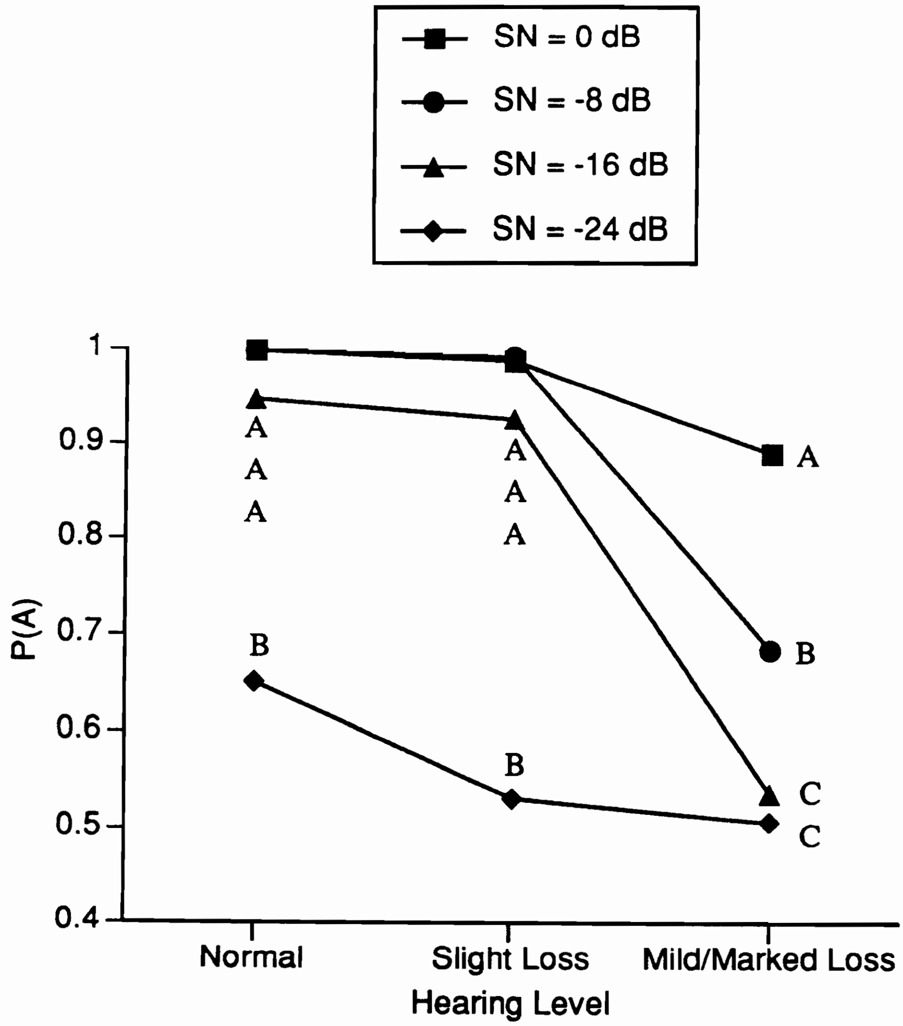


Figure 38. *Simple effects of SN at each level of HL. (Levels of the SN variable with different letters are significantly different at a given level of HL. Levels labeled with the same letter at a given HL level do not differ significantly from one another.)*

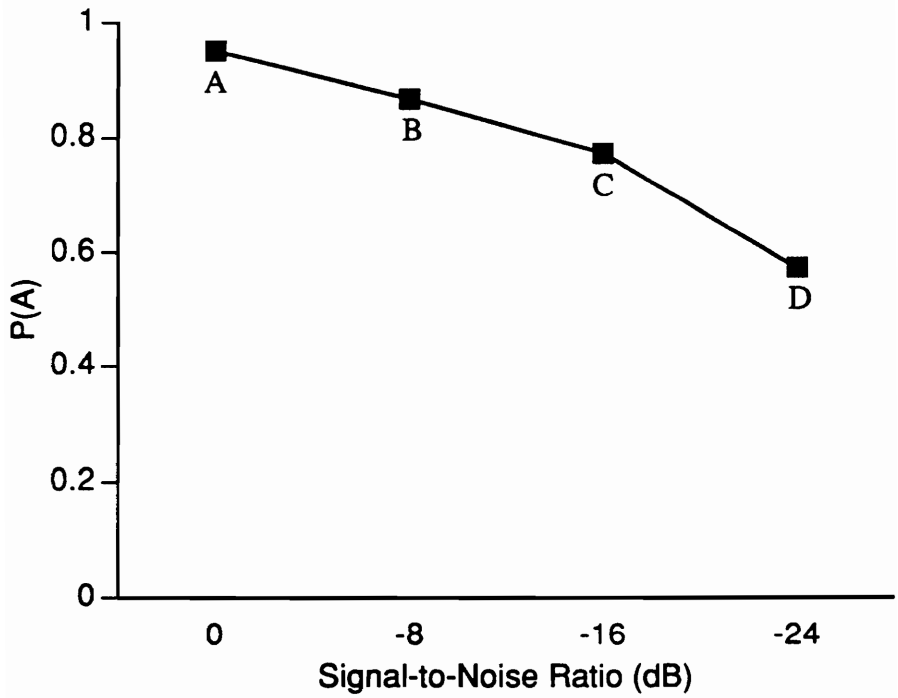


Figure 39. *Main effect of signal-to-noise ratio.*

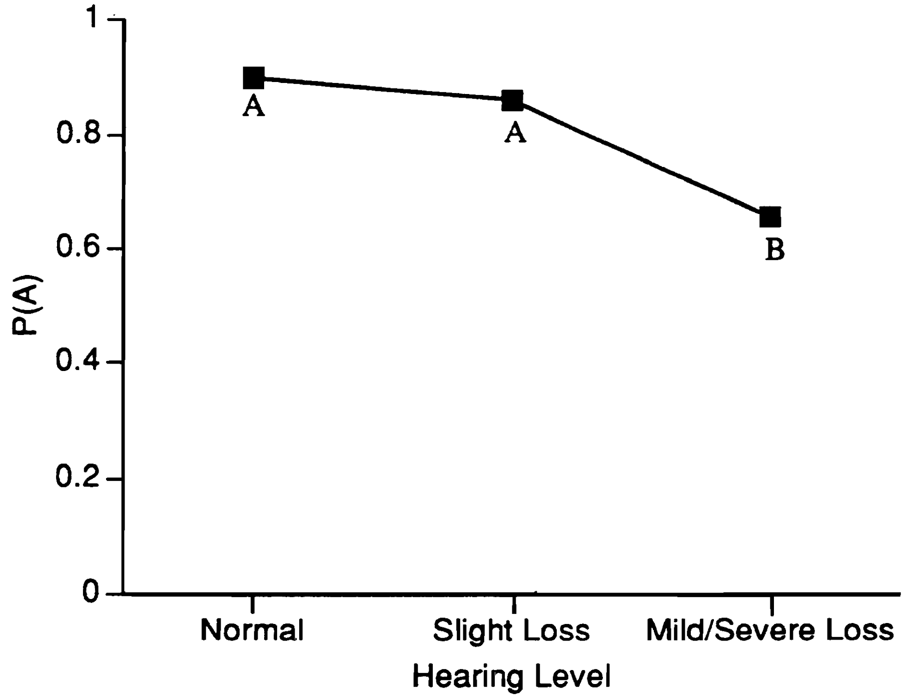


Figure 40. *Main effect of hearing level.*

since both the hearing level and the signal-to-noise ratio variables were involved in the SN × HL interaction discussed above. More specific information can be gleaned from a careful examination of the interaction than from the main effects.

One additional issue must be addressed before the discussion of the ANOVA results can be concluded. One of the subjects (Subject 13) in the group with the greatest hearing loss had thresholds in both ears of approximately 60 dBHL (depending of course on the frequency range considered). In the experiment, this subject's performance indicated that she never really heard the signal. Her data were always clustered close to the negative diagonal of the ROC curve and calculated P(A) values hovered around 0.5, indicating chance performance. There was no reason to think that her data were inappropriate for inclusion in the experiment, but there was some concern as to whether or not she belonged in the third group or if she represented yet another, more hearing-impaired group. For this reason, it was decided to perform the analysis a second time, without this subject, and see how the two analyses differed.

The pattern of significance of the second analysis (excluding Subject 13's data) was identical to that obtained in the first analysis (including Subject 13's data), including the significant differences found in the post-hoc analyses, with the exception that the noise level (NL) main effect became significant. It was therefore decided to present the results of the post-hoc test of this NL effect from the second analysis along with the data from the first analysis. Plots of the data from both analyses are presented in Figure 41. Removal of subject 13's data caused a slight increase in the mean performance measure at all three noise levels. Post-hoc tests on the data from the second analysis (dashed line) revealed that the performance at the 85 dB noise level was significantly poorer than at the 90 or 95 dB noise levels. Implications and interpretations of this and all other results will be discussed in the next section.

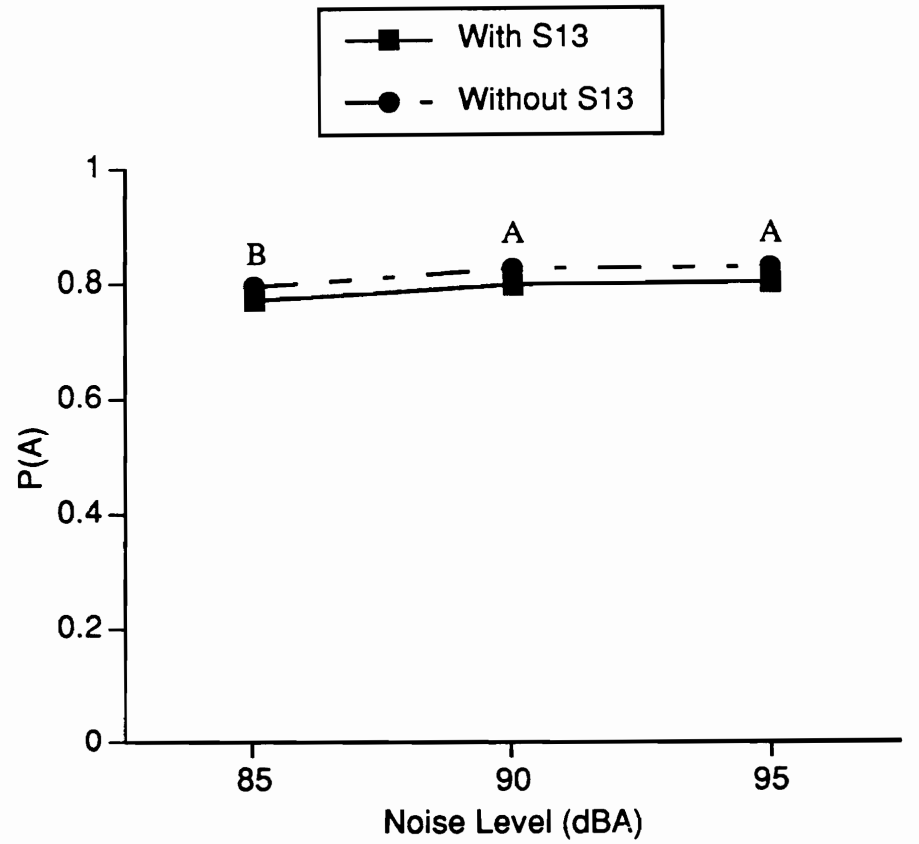


Figure 41. *Noise level effect with (solid line) and without (dashed line) Subject 13's data.*

DISCUSSION

Limitations of the Experiment

Before the significance of the results presented in the previous section can be discussed, it is necessary to point out the limitations of the research effort and how they may impact interpretation of the results. Most obvious is the fact that out of the thousands of potential subjects contacted through classified advertisements in three local newspapers and flyers posted across the Virginia Tech campus, only two subjects with a slight hearing loss that fit the requirements for the middle group could be recruited for the experiment. The effect this had on the analysis was to reduce the overall power ($1-\beta$) of the ANOVA and thus increase the potential for making a type II error (β - failing to reject the null hypothesis when in fact it is false). Also, the fact that both of these subjects' hearing levels placed them at the lower (better hearing) end of the range defining the middle group may have resulted in the mean performance for the middle group being higher than would have been the case if individuals with more varied hearing levels had been found. In the regression analysis, the fact that these two subjects had very similar hearing levels and were both at one end of the hearing level range for the middle group left a gap in the data in terms of hearing levels represented by the data. This can be seen by examination of the plots presented in the appendix. For example, when quantifying hearing level using the binaural PTA from 500 to 2000 Hz (Appendix E), there are no data between approximately 23 and 45 dBHL. In some of the experimental conditions, this range of hearing is where performance should have begun to degrade.

It would also be desirable, in a follow-on study, to add to the overall number of subjects represented in the data. If more subjects had been available, the power ($1-\beta$) of the subsequent statistical tests would have been greater. However, including more subjects in the experiment was not possible because all recruitment avenues were exhausted within the time and budget constraints. *All* of the hearing-impaired subjects meeting the *a priori* qualifications for the experiment who were willing to participate in the study were used as subjects; there were no *extra* subjects who met the hearing loss requirements.

In a similar vein, due to the small number of qualified, hearing-impaired subjects, it was not possible to select subjects based on the origin or cause of the loss. Therefore, the results could possibly be confounded by etiology. For example, of the two subjects in the middle group, one (Subject 6) exhibited a fairly flat audiogram indicative of conductive hearing loss. The other subject's (Subject 7) audiogram indicated that his hearing loss was primarily due to presbycusis. Also, all of the subjects in the mild/marked group (group 3) experienced tinnitus to some degree, which is quite common as a concomitant symptom with hearing loss. However, this should not be viewed as a serious problem since all types of hearing loss are represented in an industrial workforce.

Constraints on Generalization

Several factors existed which serve to limit the ability to generalize the results of the research described below. The first of these is the choice of alarm. As mentioned earlier, the specific alarm used in the study was chosen because it had been identified as an extremely common alarm in many industrial facilities and construction sites. The results of the experiment should apply equally well to alarms with the same or very similar characteristics (i.e., a 1 s period, 50% duty cycle, energy contained primarily in the 1000 to 1500/2000 Hz range). However, the results cannot be generalized to different alarms such as a siren or bell, or even to a horn, with characteristics which differ considerably from those of the horn used in the study. This would be the case for any *real* alarm tested. But, it is felt that the results of this experiment are more generalizable than would have been the case if a pure tone or a third-octave band of noise had been used, both of which have typically been used in other laboratory signal detection studies.

Also worth mentioning is the choice of HPD. An earmuff was chosen for use in the experiment because it was believed an earmuff would have a greater detrimental effect on the detection of an auditory alarm of predominately high-frequency output, the model described earlier applies specifically to the particular earmuff used (Bilsom Viking). Although the trends described by the data may apply when other conventional, passive HPDs are used, the specific results should not be applied to other HPDs, nor even to other earmuffs which exhibit attenuation characteristics much different from those of the earmuff used in the study.

Regression Model

The original intent of the research effort was to develop a predictive model which could be used to predict if an individual with a known hearing loss would be able to hear an auditory alarm in noise while wearing an HPD. Additionally, the model was seen as having some utility in determining what the level of an alarm must be if it were to be heard in a known noise by individuals suffering from a specified hearing loss while wearing HPDs. Such a model was successfully developed. However, due to the limitations discussed in the previous paragraphs, the model should be employed conservatively. For instance, if using the model to predict whether or not an individual can hear an alarm in a given situation, it may be prudent to base any decision on the lower prediction limit (remembering that the width of the prediction interval increases as α decreases). Likewise, in predicting what the sound pressure level of an alarm should be for a given set of conditions, a safety factor of 10 to 15 dB might be added to the model's output [which is consistent with the recommendations made by Sorkin (1987) and Wilkins and Martin (1978)].

Dependent measure. Although the dependent measure [P(A)] used may be interpreted as an indication of the *accuracy* with which an individual would be able to correctly discern a signal in a noisy environment, it is prudent to simply consider a value of $P(A) = 0.75$ as a "threshold" value (since this value represents a level of performance halfway between chance and perfect performance) and values less than 0.75 as being below threshold and values greater than 0.75 being above threshold. In this way, the model may be considered similar to what might have been developed if a more traditional psychophysical procedure had been used, such as the method of constant stimuli.

Response function. The choice of response function (logistic with a natural response) is believed to be appropriate for the application. Use of such a response function allows the model to approach an asymptote at the extremes of the data, as would be expected. (e.g., If an individual cannot hear a signal presented at a given level, there is absolutely no reason to believe that they would be able to hear the signal better if it were presented at an even lower level.) Also, response functions with such a sigmoid shape are often obtained in psychophysical studies utilizing both standard psychophysical techniques and the two-interval forced-choice procedure of signal detection theory. This characteristic shape is due to the underlying normally-distributed random variation in the subjects' thresholds over time. However, the logistic distribution is easier than the cumulative normal distribution to manipulate mathematically (Grey and Morgan, 1972; Ogilvie and Creelman, 1968) and provides a response surface of almost the same shape as would be obtained using the cumulative normal distribution (Neter, Wasserman, and Kutner, 1989).

Model selection. Insofar as which of the four models is "best," it is believed that the model quantifying hearing level based on the binaural PTA from 1000 to 2000 Hz (including the threshold at 1500 Hz) would be the best overall model for the alarm used in the experiment. The reasons this model is preferred over the other three is that it emphasizes the point that *both* ears are important in such a detection task, and that detection performance depends primarily on the hearing level in the frequency range containing most of the alarm's energy. If an individual has a substantial difference in thresholds between ears, then the model using the minimum monaural PTA in the frequency range from 1000 to 2000 Hz may be more appropriate. However, since there is so little difference in the MSE for each of the four models, any of them could probably be used with equal success.

Model Usage

As given in the Results section, all of the models developed are of the form:

$$P(A) = C + (1 - C)*(\frac{e^{(b_1*HL + b_2*NL + b_3*SL + b_4*HL*SL + b_5*HL*NL*SL)}}{1 + e^{(b_1*HL + b_2*NL + b_3*SL + b_4*HL*SL + b_5*HL*NL*SL)}})$$

where: C = the natural response frequency of the model,

HL = the PTA hearing level (calculated in one of four ways),

SL = the A-weighted SPL of the signal,

NL = the A-weighted SPL of the noise,

HL*SL = the product of the HL and SL terms,

HL*NL*SL = the product of HL, NL, and SL terms, and

b_n = estimates of the regression parameters.

The b_n terms will differ depending on which of the four possible methods is used to quantify hearing level; their values can be found in Table 23.

With the model in this form, it is possible to predict if an individual would be capable of hearing an alarm under known conditions simply by substituting the hearing level, signal level, and noise level terms into the model. For example, given a binaural PTA hearing level of 47 dBHL (in the frequency range of 1000 to 2000 Hz), 90 dBA noise, and a signal level of 82 dBA, the P(A) predicted by the model would be 0.85. This value is above that considered to be a "threshold" value [P(A) = 0.75], but not substantially greater. It would therefore be assumed that the individual was capable of hearing the alarm, but not with absolute certainty.

It is also possible to use the model to predict the masked "threshold" of an individual with a known hearing level in a given noise. However, to do so, it is first necessary to rearrange the above equation so that the SL term may be solved for when assuming P(A) = 0.75. The equation then becomes:

$$SL = \frac{\ln\left(\frac{P(A)-C}{1-P(A)}\right) - b_1 * HL - b_2 * NL}{b_3 + b_4 * HL + b_5 * HL * NL}$$

where the terms are as specified earlier.

Using this equation and assuming P(A) = 0.75, NL = 90, HL = 47 (again using the binaural PTA hearing level from 1000 to 2000 Hz), the predicted masked threshold for the signal is 80 dBA. As expected, the masked "threshold" predicted using this equation is slightly less (2 dB) than the signal level used in the previous calculation of detectability. As per the suggestions made earlier, the conservative approach would be to add 10 to 15 dB to this estimate to ensure that the signal is detectable under normal circumstances.

In a similar fashion, the equation can be rearranged and solved for either the HL or the NL terms. Doing so would allow the end user to determine what minimum hearing level should be required of individuals expected to be exposed to known noise conditions, or to determine the level that the ambient noise must be reduced to so that existing alarms/warnings can be heard by all of the workers in the area.

However the model is used, it is important to remember that the signal and noise levels be within the ranges used in the experiment. That is noise levels from 85 to 95 dBA, and signal levels within the range represented by signal-to-noise ratios from 0 to -24 dB.

It is also important to remember that the noise and signal levels used in the model are those present at the location of the individual(s) being considered, not the levels measured at the alarm's location or in some central location. This point cannot be overemphasized. Noise level information may be available in the form of sound level contour depicting the SPL at regularly spaced intervals throughout the work area such as that illustrated in Figure 42 (Royster, Berger, and Royster, 1986). Data of this type may be gathered as part of an industrial noise survey in support of a hearing conservation program. Signal level information of the type necessary would not likely be available, however. It would therefore be necessary to generate such data. If the alarm being considered is fixed, then perhaps a contour map of the signal levels similar to the noise level contour in Figure 42 can be generated. If the alarm is not fixed (i.e., a back-up alarm on a piece of construction equipment or an alarm on an overhead crane), then perhaps a signal level contour for the specific piece of equipment could be generated. These data, plus the hearing levels of the workforce, which should be readily available if a hearing conservation program is in place, are all the information necessary for utilization of the regression model described earlier.

ANOVA

Since the hearing level (HL) and signal-to-noise ratio (SN) factors are both involved in the HL × SN interaction, the main effects will not be discussed independently. Rather, only the interaction will be discussed. Perhaps the best single plot of the HL × SN interaction is that which appears in Figure 37 and is repeated here in Figure 43

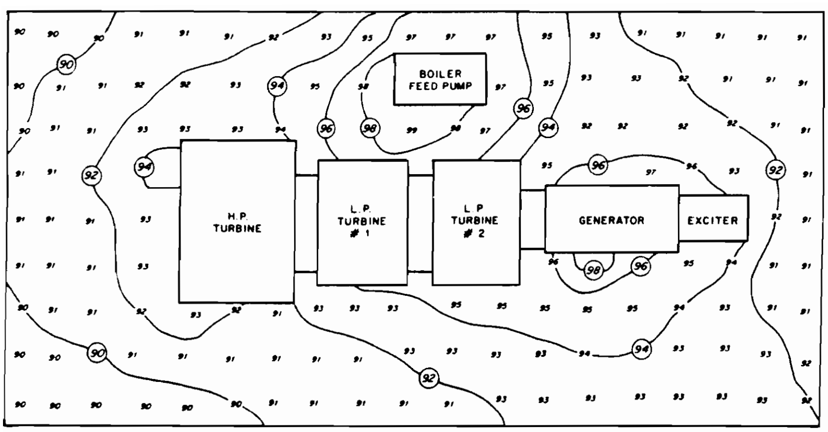


Figure 42. *Illustration of a sound level contour*. (from Royster, Berger, and Royster, 1986)

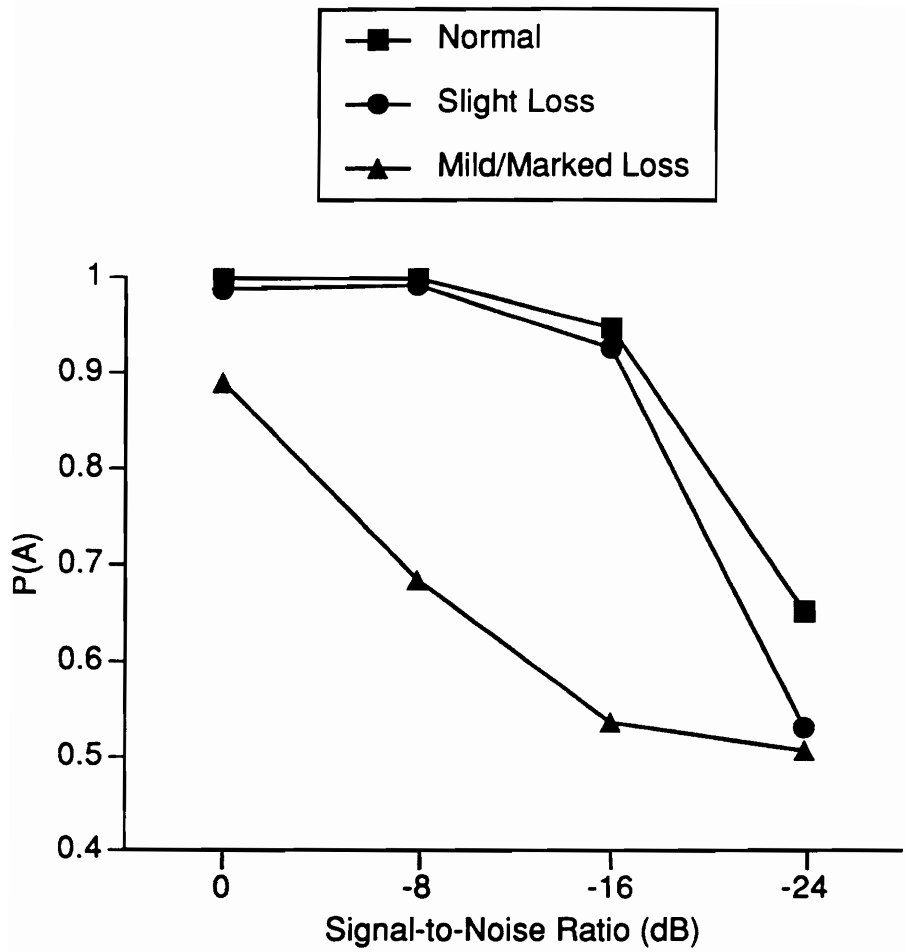


Figure 43. *SN × HL two-way interaction*.

(without the alphabetic labels). As can be seen, the performance of the subjects in each hearing level group decreases as signal-to-noise ratio decreases (the simple main effect of SN for each HL group). What is different is at what point performance drops below what would be an acceptable level in an industrial situation in which the alarm were actually being listened for. For the better hearing subjects (groups 1 and 2), performance does not drop off drastically until the signal-to-noise ratio drops to -24 dB. For the more hearing-impaired subjects (group 3), performance is unsatisfactory at all signal-to-noise ratios other than 0 dB. [Although the figure shows the mean performance of the subjects in group 3 (mild/marked loss) at the 0 dB signal-to-noise ratio to be at a level of P(A) = 0.89, when Subject 13's data are eliminated, the performance of this group improves to P(A) = 0.97. Although the mean performance level of this group also improves slightly at the three lower signal-to-noise ratios when Subject 13's data are eliminated, the performance remains below P(A) = 0.75.] The nature of this simple main effect of SN for the subjects exhibiting only a slight hearing loss (group 2) might have been slightly different had more subjects that fit into this category been found and had their hearing levels varied over the entire range allowed for this group.

It is important to note that even subjects with a fairly severe hearing loss (having hearing levels on the order of 50 dBHL in the frequency range of the signal) are still capable of detecting the signal used in the study when presented at a fairly low signal-to-noise ratio (0 dB). The threshold [assuming P(A) = 0.75 to represent "threshold"] for this signal in the conditions examined in the experiment appears to be at a signal-to-noise ratio of about -8 dB. "Threshold" for the normal-hearing subjects, on the other hand, appears to be at a signal-to-noise ratio between -16 and -24 dB.

As expected, there were significant differences found in the performance of the three hearing level groups at three of the four levels of signal-to-noise ratio (-8, -16, and -24 dB). However, the nature of the differences involving the group of subjects showing only a slight hearing loss (the middle group) might have been different if more subjects belonging to this group had been present in the study and if these subjects had better represented the range of hearing allowed for this group. It is believed that as hearing level increases (hearing becomes worse) performance would drop off first at the lowest signal-to-noise ratios (as evidenced by the significant difference between groups 1 and 2 at the -24 dB signal-to-noise ratio) and then to progress to higher signal-to-noise ratios with the degree of the performance decrement dependent on hearing level at any given signal-to-noise ratio. However, the study described herein can only allude to this effect due to the small number of subjects and the gap in the data caused by the missing subjects. If more subjects had been available, more and narrower hearing level categories utilized, and more levels of signal-to-noise ratio used, the data would have been more persuasive in this regard.

The noise level effect (illustrated in Figure 41), although not significant in the original analysis, was significant when the data were analyzed after eliminating the data obtained from Subject 13. The effect was such that the performance of the subjects in the 85 dBA noise level was slightly poorer than their performance at the other noise levels (90 and 95 dBA). This may have been due to the subjects in the most hearing-impaired group performing slightly poorer at the 0 dB signal-to-noise ratio in 85 dBA noise than at the 0 dB signal-to-noise ratio in the other noise levels. If this is the case, it is likely caused by the fact that the signal in that condition was at such a low level, that the earmuff attenuated it to such an extent that it approached threshold for the listeners in this group.

This leads to the suspicion that there may be additional factors important in determining whether or not an individual can hear a signal in noise while wearing an HPD, perhaps in the form of the three-way NL × SN × HL interaction. Although this interaction was not significant in the ANOVA ($F = 1.53$, $p = 0.1858$), it was found to be a necessary term in the regression model. As mentioned several times previously, it may be that with more subjects (resulting in a more powerful test), this interaction may indeed be shown to be significant.

FINAL CONCLUSIONS

Perhaps the most important conclusion reached based on the results of the research effort described herein is the fact that even individuals with a substantial hearing loss (on the order of 45 to 50 dBHL) are capable of hearing a back-up alarm when presented at a reasonably low signal-to-noise ratio (0 dB) in all of the noise levels investigated while wearing a high-attenuation earmuff. However, when considering individuals with greater hearing loss or noise levels other than those investigated in the experiment, the picture is less clear. The marginal noise level effect seems to indicate that at noise levels less than 85 dBA, individuals with a mild to marked hearing loss may begin to experience difficulty in detecting such an alarm at such a low signal-to-noise ratio. Also, detectability of such an alarm in noise while wearing an HPD depends, at the very least, on the signal-to-noise ratio and the hearing level of the individual and probably also on the level of the masking noise.

At reasonable signal-to-noise ratios (0 dB), the hearing level at which a performance decrement begins to become apparent appears to be in the range between approximately 50 to 60 dBHL. This range is considerably narrower than the 35 to 65 dBHL range mentioned by Abel, et al. (1983a). Also, it would appear that only the hearing levels in the range of the alarm being listened for is important in the detection task and that both ears are important in such a task, at least when the between-ear difference is less than 15 to 20 dB.

RECOMMENDATIONS FOR FURTHER RESEARCH

Although the experiment described herein was successful, the scope of the research was limited in that only one alarm, one noise, and one HPD were used and that it was not possible to screen subjects based on etiology of hearing loss. Additional experiments should be conducted which investigate the issue of audibility of auditory alarms in noise, but they should include more subjects and narrower hearing level categories (on the order of 10 to 15 dB). Also, additional independent variables should be investigated, including etiology, protector type (plug versus muff) with attenuation as a variable, alarm type (i.e., bell, siren, horn) and spectra, and attentional demand. Although it would not be possible to include all of the above dimensions into a single study, it may be possible to use sequential experimentation techniques and answer the research questions as part of a multi-year effort. Since criteria would not be investigated (a separate experiment to investigate criteria differences is described below), the two-interval forced-choice procedure of signal detection theory would probably be the best and most efficient method of data collection.

The advantage of sequential experimentation is that a series of small experiments can be conducted in which only a few (one or two) of the potential independent variables would need to be manipulated while other independent variables are manipulated between experiments. In this way, the data from multiple experiments may be analyzed independently as well as if they had been obtained from a single large experiment. Doing so has the added benefit of breaking down what could be a large, complicated, time consuming effort into smaller, more economical, and more easily-solved parts.

However, care must be taken in planning such an undertaking. Statistically, each portion of the study must be designed so it can stand alone, and the parts must form a logical *and* analyzable whole. In addition, logistical concerns must be addressed, such as whether each of the smaller experiments will be conducted in the same laboratory using the same equipment, or if they will be conducted in different laboratories. If different laboratories are to be used, consensus and consistency must be achieved with respect to how the physical environment is to be quantified, what equipment is to be used, and what procedures are to be followed. (Though not impossible, as evidenced by the various ANSI acoustical test standards, this process could be a major impediment to the research effort.) If the experiments are to be conducted in the same laboratory, the concerns about the physical environment and equipment become less important, but the question of procedures must still be addressed, especially if different experimenters will be involved.

However, since so few hearing-impaired subjects could be found that both met the requirements of the experiment and were willing to participate in the study, perhaps the biggest obstacle standing in the way of using sequential experimentation in a research effort such as that described herein is the requirement that the between-experiment variables also be between-subject variables.

A separate experiment designed specifically for investigating the question of criterion differences should also be conducted. Signal levels should be set such that at each noise level used, the sensitivity for all subjects would be approximately equal regardless of hearing level. In such an experiment, it would probably be possible to use a single experimental session utilizing the yes/no procedure since there would be no interest in generating a complete ROC curve, but rather just a criterion measure. Use of the yes/no procedure would have the added benefit of allowing the subject to establish his/her own natural criterion, whereas use of the rating procedure's requirement that the subject establish multiple criteria could cloud the issue by confusing the subject's natural criterion with artificial criteria adopted due to the experimental protocol.

REFERENCES

Abel, S. M., Kunov, H., Pichora-Fuller, M. K., and Alberti, P. W. (1983a). *Hearing protectors: Signal detection in industrial noise and the effects of noise exposure history*. Mount Sinai Hospital, Toronto: Silverman Hearing Research Laboratory.

Abel, S. M., Kunov, H., Pichora-Fuller, M. K., and Alberti, P. W. (1983b). The effect of hearing protection on narrowband signal detection in industrial noise. *The Journal of Otolaryngology, 12*(2), 83–88.

ANSI (1973). *Psychoacoustical terminology*, ANSI S3.20-1973. New York: American National Standards Institute.

ANSI (1974). *Method for the measurement of real-ear protection of hearing protectors and physical attenuation of earmuffs*, ANSI S3.19-1974. New York: American National Standards Institute.

ANSI (1980a). *Precision methods for the determination of sound power levels of broad-band noise sources in reverberation rooms*, ANSI S1.31-1980. New York: American National Standards Institute.

ANSI (1980b). *Precision methods for the determination of sound power levels of discrete-frequency and narrow-band noise sources in reverberation rooms*, ANSI S1.32-1980. New York: American National Standards Institute.

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

ANSI (1989). *Specification for audiometers*, ANSI S3.6-1989. New York: American National Standards Institute.

Békésy, G. von (1960). *Experiments in hearing*. New York: McGraw-Hill.

Beranek, L. L. (1986). *Acoustics*. New York: Acoustical Society of America.

Berkowitz, J. P. and Casali. S. P. (1990). Influence of age on the ability to hear telephone ringers of different spectral content. In *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 132–136). Santa Monica, CA: Human Factors Society.

Bilger, R. C. and Hirsh, I. J. (1956). Masking of tones by bands of noise. *Journal of the Acoustical Society of America*, *28*(4), 623–630.

Bock, M., Lazarus, H., and Hoege, H. (1983). Effects of noise on the efficiency of danger signals. In *Proceedings of the Fourth International Congress on Noise as a Public Health Problem* (pp. 517–521). Milano, Italy: Centro Ricerche E Studi Amplifon.

Casali, J. G. (1986). In defense of industrial workers' hearing. *IE News*, *20*(4), 1–6.

Casali, J. G. (1988). *A computer-controlled facility for hearing protection research and attenuation testing verification re ANSI S12.6-1984* (IEOR Technical Report No. 8801). Blacksburg, VA: Virginia Polytechnic Institute and State University, Industrial Engineering and Operations Research Department.

Casali, J. G. and Epps, B. W. (1986). Effects of user insertion/donning instructions on noise attenuation of aural insert hearing protectors. *Human Factors*, *28*(2), 195-210.

Casali, J. G. and Park, M. Y. (1990). Attenuation performance of four hearing protectors under dynamic movement and different fitting conditions. *Human Factors*, *32*(1), 9-25.

Casali, J. G. and Park, M. Y. (1992). A regression-based methodology for efficient prediction of broadband attenuation of hearing protectors. *Noise Control Engineering Journal*, *38*(3), 97–108.

Casali, J. G. and Robinson, G. S. (1990). *A reverberant computer-controlled facility for hearing protection research and attenuation testing: verification re ANSI S3.19-1974* (ISE Technical Report No. 9007; Audio Lab No. 12/1/90/5-HP). Blacksburg, VA: Virginia Polytechnic Institute and State University, Industrial and Systems Engineering Department.

Chapra, S. C. and Canale, R. P. (1985). *Numerical methods for engineers* (2nd ed.), New York: McGraw-Hill.

Coleman, G. J., Graves, R. J., Collier, S. G., Golding, D., Nicholl, A. G. McK., Simpson, G. C., Sweetland, K. F., and Talbot, C. F. (1984). *Communications in Noisy Environments* (Report Number TM/84/1). Edinburgh, England: Institute of Occupational Medicine, Ergonomics Branch.

Deatherage, B. H. (1972). Auditory and other sensory forms of information presentation. In H. P. Van Cott and R. G. Kincade (Eds.), *Human Engineering Guide to Equipment Design* (pp. 123–160). New York: Wiley.

DoD (1981). *Human engineering design criteria for military systems, equipment and facilities*, MIL–STD–1472C. Washington, D. C.: Department of Defense.

Dorfman, D. D. (1983). *RSCORE III* [Computer program]. University of Iowa, Department of Psychology.

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

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.

Egan, J. P. and Hake, H. W. (1950). On the masking pattern of a simple auditory stimulus. *Journal of the Acoustical Society of America*, *22*(5), 622–630.

Egan, J. P., Schulman, A. I., and Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America*, *31*(6), 768–773.

EPA (1981). *Noise in America: the extent of noise problem* (EPA Report 550/9-81-101). Washington, D.C.: Environmental Protection Agency.

Fidell, S., Parsons, K. S., and Bennett, R. (1974). Prediction of aural detectability of noise signals. *Human Factors, 16*(4), 373–383.

Fidell, S. (1978). Effectiveness of audible warning signals for emergency vehicles. *Human Factors, 20*(1), 19–26.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics, 12*, 47–65.

Forshaw, S. E. (1977). *Listening for machinery malfunctions in noise while wearing ear muffs* (DCIEM Technical Report No. 77x43). Ontario, Canada: Department of National Defense – Canada, Defense and Civil Institute of Environmental Medicine, Behavioral Division.

French, N. R. and Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 19*(1), 90–119.

Gales, R. S. (1979). Hearing characteristics. In C. M. Harris (Ed.), *Handbook of noise control* (2nd ed.) (pp. 8-1 to 8-18). New York: McGraw-Hill.

Gasaway, D. C. (1984). 'Sabotage' can wreck hearing conservation programs. *National Safety News, 129*(5), 56-60 and 66.

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

Gescheider, G. A., Wright, J. H., Weber, B. J., and Barton, W. G. (1971). Absolute thresholds in vibrotactile signal detection. *Perception and Psychophysics, 10*(6), 413-417.

Goldstein, E. B. (1989). *Sensation and perception* (3rd ed.). Belmont, CA: Wadsworth.

Green, D. M., and Swets, J. A. (1988). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula.

Grey, D. R. and Morgan, B. J. T. (1972). Some aspects of ROC curve-fitting: normal and logistic models. *Journal of Mathematical Psychology*, *9*, 128–139.

Haas, E. C. and Casali, J. G. (in press). The perceived urgency and detection time of multi-tone and frequency-modulated warning signals in broadband noise. *Human Factors*.

Hawkins, J. E. and Stevens, S. S. (1950). The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America*, *22*(1), 6–13.

Höge, H., Schick, A., Kuwano, S., Namba, S., Bock, M., and Lazarus, H. (1988). Are there invariants of sound interpretation? The case of danger signals. In *Noise as a Public Health Problem, Hearing, Communication, Sleep and Nonauditory Physiological Effects* (pp. 253–258). Stockholm, Sweden: Swedish Council for Building Research.

ISO (1990). *Acoustics – measurement of sound attenuation of hearing protectors – subjective method*, ISO 4869–1. Geneva, Switzerland: International Standards Organization.

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

Kerivan, J. E. (1979). *The effects of earplugs and earmuffs an pitch discrimination in noise* (NSMRL Report Number 888). New London, CN: Naval Submarine Medical Research Laboratory.

Kryter, K. D. (1985). *The effects of noise on man*. Orlando: Academic Press.

Laroche, C., Tran Quoc, H., Hétu, R., and McDuff, S. (1991). 'Detectsound': A computerized model for predicting the detectability of warning signals in noisy workplaces. *Applied Acoustics*, *32*, 193–214.

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*(5), 852–858.

Lazarus, H. (1980). The effects of hearing protectors on the perception of acoustic signals. *Zentrablatt für Arbeitsmedizin*, *30*(6), 204–212.

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

Mauney, D. W. (1992). *PHYTEST* [computer program]. Virginia Polytechnic Institute and State University, Industrial and Systems Engineering Department.

Melnick, W. (1979). Hearing loss from noise exposure. In C. M. Harris (Ed.), *Handbook of noise control* (2nd ed.) (pp. 9-1 to 9-16). New York: McGraw-Hill.

Melnick, W. (1984). Evaluation of industrial hearing conservation programs: A review and analysis. *American Industrial Hygiene Association Journal, 45*(7), 459–467.

Miller, M. H. and Harris, J. D. (1979). Hearing testing in industry. In C. M. Harris (Ed.), *Handbook of noise control* (2nd ed.) (pp. 10-1 to 10-12). New York: McGraw-Hill.

Miller, M. H. and Wilber, L. A. (1991). Hearing evaluation. In C. M. Harris (Ed.), *Handbook of Acoustical Measurement and Noise Control* (3rd ed.) (pp. 19-1 to 19-17). New York: McGraw-Hill.

Montgomery, D. C. and Peck, E. A. (1982). *Introduction to linear regression analysis*. New York: Wiley.

Morrill, J. C. (1986). Hearing measurement. In E. H. Berger, W. D. Ward, J. C. Morrill, and L. H. Royster (Eds.), *Noise and hearing conservation manual* (4th ed.) (pp. 1-18). Akron, OH: American Industrial Hygiene Association.

Moskowitz, H. and McGlothlin, W. (1974). Effects of marihuana on auditory signal detection. *Psychopharmacologia*, *40*(2), 137–145.

Neter, J., Wasserman, W., and Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Homewood, IL: Irwin.

Newby, H. A. (1979). *Audiology* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

NFPA (1978). *Household fire warning equipment 1978*, ANSI/NFPA 74–1978. Boston: National Fire Protection Association.

NFPA (1979). *Local protective signaling systems*, ANSI/NFPA 72A–1979, Boston: National Fire Protection Association.

NIH (1990). *Noise and hearing loss*. NIH consensus development conference statement 1990 Jan 22-24, 8(1), Bethesda, MD: National Institutes of Health.

Nixon, C. W. (1979). Hearing protective devices: ear protectors. In C. M. Harris (Ed.), *Handbook of noise control* (2nd ed.) (pp. 12-1 to 12-13). New York: McGraw-Hill.

Ogilvie, J. C. and Creelman, D. (1968). Maximum-likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology, 5*, 377–391.

OSHA (1989). *Occupational noise exposure limits*. Code of Federal Regulations, 29 Parts 1900-1910, 179–195.

Park, M.–Y. and Casali, J. G. (1991). A controlled investigation of the in-field attenuation performance of selected insert, earmuff, and canal cap hearing protectors. *Human Factors*, *33*(6), 693–714.

Patterson, R. D. (1974). Auditory filter shape. *Journal of the Acoustical Society of America*, *55*(4), 802–809.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America*, *59*(3), 640–654.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *Journal of the Acoustical Society of America*, *72*(6), 1788–1803.

Robinette, M. S. (1984). Audiometric programs' value rising with hearing loss claims. *Occupational Health and Safety*, February, 23–24.

Robinson, D. E. and Watson, C. S. (1972). Psychophysical methods in modern psychoacoustics. In J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory, Volume II* (pp. 101 –131). New York: Academic Press.

Rollman, G. B. (1977). Signal detection theory measurement of pain: A review and critique. *Pain*, *3*, 187-211.

Royster, L. H., Berger, E. H., and Royster, J. D. (1986). Noise surveys and data analysis. In E. H. Berger, W. D. Ward, J. C. Morrill, and L. H. Royster (Eds.), *Noise and hearing conservation manual* (4th ed.) (pp. 97-176). Akron, OH: American Industrial Hygiene Association.

SAE (1978). *Performance, test, and application criteria for electronically operated backup alarm devices*, ANSI/SAE J994b–1978. Warrendale, PA: Society of Automotive Engineers, Inc.

SAS (1990). *PROC NLIN* [Computer program]. Cary NC, SAS Institute Inc.

SAS (1990). *PROC PROBIT* [Computer program]. Cary NC, SAS Institute Inc.

Scharf, B. (1970). Critical bands. In J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory, Volume I* (pp. 159 –202). New York: Academic Press.

Sorkin, R. D. (1987). Design of auditory and tactile displays. In G. Salvendy (Ed.), *Handbook of Human Factors*. New York: Wiley.

Speed, F. M., Hocking, R. R., and Hackney, O. P. (1978). Methods of analysis of linear models with unbalanced data. *Journal of the American Statistical Society*, *73*(361), 105–112.

Spieth, W. (1957). Downward spread of masking. *Journal of the Acoustical Society of America*, *29*(4), 502–505.

Suter, A. H. (1989). *The effects of hearing protectors on speech communication and the perception of warning signals* (Technical Memorandum 2–89). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, *99*(1), 100–117.

Swets, J. A. (1988a). *Signal detection and recognition by human observers*. Los Altos, CA: Peninsula.

Swets, J. A. (1988b). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293.

Swets, J. A., Pickett, R. M., Whitehead, S. F., Getty, D. J., Schnur, J. A., Swets, J. B., and Freeman, B. A. (1979). Assessment of diagnostic technologies. *Science*, *205*(4407), 753–759.

Swets, J. A., Tanner, W. P., and Birdsall, T. G. (1961). Decision process in perception. *Psychological Review*, *68*(5), 310–340.

Tanner Jr., W. P. and Sorkin, R. D. (1972). The theory of signal detectability. In J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory, Volume II* (pp. 63 – 98). New York: Academic Press.

UL (1978). *Household burglar–alarm system units*, ANSI/UL 1023–1978. Melville, NY: Underwriters Laboratories Inc.

UL (1981). *Audible signal appliances*, ANSI/UL 464–1981. Melville, NY: Underwriters Laboratories Inc.

Van Moorhem, W. K., Shepherd, K. P., Magleby, T. D., and Torian, G. E. (1981). The effects of motorcycle helmets on hearing and the detection of warning signals. *Journal of Sound and Vibration*, *77*(1), 39–49.

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. (1986a). 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, OH: American Industrial Hygiene Association.

Ward, W. D. (1986b). Auditory effects of noise. In E. H. Berger, W. D. Ward, J. C. Morrill, and L. H. Royster (Eds.), *Noise and hearing conservation manual* (4th ed.) (pp. 197-216). Akron, OH: American Industrial Hygiene Association.

Watson, C. S., Franks, J. R., and Hood, D. C. (1972). Detection of tones in the absence of external masking noise. I. Effects of signal intensity and signal frequency. *Journal of the Acoustical Society of America*, *52*(2, Part 2), 633 – 643.

Weber, D. L. (1977). Growth of masking and the auditory filter. *Journal of the Acoustical Society of America*, *62*(2), 424–429.

Webster, J. C., Himes, H. P., and Lichtenstein, M.,. (1950). San Diego county fair hearing survey. *Journal of the Acoustical Society of America*, *22*(4), 473–483.

Webster, J. C., Lichtenstein, M., and Gales, R. S. (1950). Individual differences in noise masked thresholds. *Journal of the Acoustical Society of America*, *22*(4), 483–490.

Wickens, C. D. (1984). *Engineering psychology and human performance*. Glenview, IL: Scott, Foresman.

Wilkins, P. A. (1984). A field study to asses the effectiveness 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*, *25*(3), 249–260.

Wilkins, P.A. and Martin, A. M. (1977). *The effect of hearing protection on the masked thresholds of acoustic warning signals.* Paper presented at the 9th International Congress on Acoustics, Madrid, Spain.

Wilkins, P.A. and Martin, A. M. (1978). *The effect of hearing protectors on the perception of warning and indicator sounds – A general review* (Technical Report No. 98). University of Southampton, England: Institute of Sound and Vibration Research.

Wilkins, P.A. and Martin, A. M. (1981). The effect of hearing protectors on the attention demand of warning sounds. *Scandinavian Audiology, 10,* 37–43.

Wilkins, P.A. and Martin, A. M. (1982). 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 Press.

Wilkins, P.A. and Martin, A. M. (1984). Attention demand and recognition in the perception of warning sounds and the effects of wearing hearing protection. *Journal of Sound and Vibration, 94*(4), 483–494.

Wilkins, P.A. and Martin, A. M. (1985). The role of acoustical characteristics in the perception of warning sounds and the effects of wearing hearing protection. *Journal of Sound and Vibration, 100*(2), 181–190.

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. (1969). Within-session criterion changes compared to an ideal observer criterion in a visual monitoring task. *Journal of Experimental Psychology, 81*(1), 61 – 66.

Williges, R. C. (1971). The role of payoffs and signal ratios in criterion changes during a monitoring task. *Human Factors, 13*(3), 261–267.

Williges, R. C. (1973). Manipulating the response criterion in visual monitoring. *Human Factors, 15*(2), 179–185.

Winer, B. J., Brown, D. R., and Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

APPENDIX A

Description of Experiment

Description of the Signal Detection Experiment

Written Instructions to Subject Participant

This experiment is intended to determine how well people can hear an alarm or warning sound when it is played in a noisy area while they are wearing earmuffs. If you become a subject in the experiment, you will be asked to participate in a total of 12 experimental sessions in addition to this screening session and 1 training session.

The screening session will consist of your being presented with an informed consent form, filling out a questionnaire concerning your hearing and past experience with hearing protection devices, having your outer ears visually examined, and undergoing the administration of a hearing test. If you qualify as a subject, the experimenter will then show you the experimental apparatus, demonstrate the signals to be used in the experiment, and explain the experimental procedures. Feel free to ask questions at any time. The entire screening session will last approximately one hour.

The training and experimental sessions will be structured exactly alike and will begin with a short version of the hearing test conducted during the screening session. The experimental procedures will then be explained. Next, a miniature microphone will be placed securely at the opening of your ear canal, you will be fitted with an earmuff, and a second miniature microphone will be affixed to the exterior surface of the earmuff. Once the earmuff is properly fit, you will be escorted into the test room. The experimenter will then check the fit of the earmuff by playing a noise for about 30 seconds. If the experimenter is not satisfied with the fit of the earmuff, he will adjust it and recheck its fit. Once the experimenter is satisfied with the fit of the earmuff, the signal you are to listen for will again be demonstrated, after which the training procedure will begin. The training procedure will be arranged into 6 blocks with 84 trials in each block, for a total of 504 trials. During a given trial, a noise will be presented for about 2 seconds. While the noise is on, the signal which you are to listen for may or may not also be presented. Your task is to indicate, by pressing the appropriate switch, whether or not you heard the signal and how sure you are of your response. You may respond at any time while the noise is on or anytime before the next noise presentation. During each of the blocks of trials, exactly half of the trials will contain the signal as well as the noise, and exactly half the trials will contain only the noise. At the end of each block of trials, you will be given the opportunity to take a break. However, to ensure that the fit of the earmuff does not change, you are asked to refrain from touching the device and to avoid talking or otherwise moving your jaw either during the experiment or during the breaks. These procedures will continue until all 500 trials have been completed. Once the last block of trials has been completed, the earmuff and miniature microphones will be removed, and a hearing test performed exactly like the one done at the start of the session will be conducted. It is expected that the length of each of these sessions will between 1-1/2 and 2 hours.

No known risk is posed by the experiment except possibly of fatigue due to the length of the experimental sessions and perhaps some discomfort because of the snug fit of the protectors, but the devices will not harm you in any permanent way.

Please sign below to indicate that you have read and understood these instructions.

____________________________ ____________________________

Subject's Printed Name Subject's Signature

APPENDIX B

Subject's Informed Consent

SUBJECT'S INFORMED CONSENT

AUDITORY SYSTEMS LABORATORY-VA TECH
(AUDIOMETRY AND SIGNAL DETECTION EXPERIMENT)

First, your right and left ear hearing 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 your ability to hear alarm and warning sounds in noise while wearing a hearing protector. In the hearing test, your hearing will be tested with very quiet pulsating tones played through a set of earphones. You will have to be very attentive and listen carefully for these tones. **Depress the button on the hand switch and hold it down whenever you can hear the tone and release it when you do not hear a tone.** The tones will be very faint and you will have to listen very carefully to hear them. During the experimental session, you will be asked to listen for an alarm signal presented during a short period of noise. Again, you will have to be very attentive and listen carefully for the signal. During this period of noise, you will indicate whether or not you hear the alarm signal and how sure you are of the presence or absence of the signal by pressing one of six switches. Some of the signals may be loud enough that you will have no trouble hearing them while others may be so quiet you may not hear them at all.

During the experimental sessions, you will always be wearing a hearing protector when the noises and signals are played. The test will be conducted in a sound-proof booth with the experimenter sitting outside. The door to the booth will be shut but not 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 these hearing tests. Also, realize that they are not designed to assess or diagnose any physiological or anatomical hearing disorders. The tests will only be used to determine your hearing ability today.

The purpose of the study is to determine how well people can hear auditory alarms and warnings when in a noisy area while they are wearing hearing protectors. The experimenter will always fit the protector on you, but your feedback is important so that the best possible fit can be obtained. The protectors are intended to provide a snug fit so that noise will be blocked. Therefore, they may seem tight around your ears. Some minor discomfort may result from the tight fit, but the protectors will not harm you in any permanent way.

Several physical measurements may also be obtained as a part of the study. These will include dimensional measurements of the ear and head width, obtained with simple rulers, calipers, and an ear gauge. Your middle ear pressure may be checked using a clinical tympanogram, a simple, non-invasive procedure to determine if the pressure is within a normal range. None of these screening procedures pose any risk to your well-being or cause any pain. If you desire, the experimenter will show you the measurement instruments at this time.

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 a member of the research team.

2) You have the right to inspect your data and to withdraw it from the experiment if you feel that you should. In general, 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 else 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 (six 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 pre-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, being paid only for the portion of time that I spend in the study.

Signature ______________________________________

Printed Name ___________________________________

Date ____

Phone ____

REMEMBER, you are supposed to press the button (and keep it pressed) whenever you hear the tone and release it when you do not hear the tone.

The research team for this experiment consists of Gary Robinson, a Ph.D. student in ISE, and Dr. John G. Casali, Director of the Auditory Systems Laboratory. They 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

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

Dr. Ernie Stout
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

Pilot Study to Determine the Signal Levels to be Used in the Experiment

PILOT STUDY TO DETERMINE THE SIGNAL LEVELS TO BE USED IN THE EXPERIMENT

A short of pilot study was performed to determine what signal levels should be used in the experiment. An additional question which also had to be decided was whether the signal level independent variable should be represented by four discrete levels of the signal across all three noise levels, or if it should be in terms of the signal-to-noise ratio, allowing different absolute signal levels in each level of noise while maintaining the relative difference between the noise and signal levels constant.

With only a few minor exceptions, the pilot tests were performed using the same TSD rating procedure that was to be used in the actual experiment. Doing so allowed the experimental set-up (both hardware and software) and procedures to be tested. The procedures used in the pilot study differed from those to be used in the experiment in that only 150 to 250 trials per condition were used and the HPD's attenuation was not checked.

Four subjects participated in the pilot study as unpaid volunteers. Although one of the subjects did have a hearing loss which was severe enough to require the use of hearing aids in both ears during normal daily activity, his PTA hearing levels over the 500 to 2000 Hz range were only slightly higher than the other three subjects (by about 10 dB) and his individual thresholds at the frequencies through 1500 Hz were normal ($\leq$ 15 dBHL) . This subject had never participated in any of the experiments conducted in the laboratory and was therefore considered to be a novice listener. The other three subjects were experienced listeners and had participated in numerous experiments in the preceding months/years.

Only two noise levels (85 and 90 dBA) were used in the pilot study, and signals were presented at signal-to-noise ratios of -15, -20, -25, and -30 dBA for a total of eight test conditions. No subject participated in more than four of the eight conditions.

The data obtained were not analyzed quantitatively. Instead, ROC curves were generated and a simple visual examination was performed to determine whether or not the subject could hear the signal. This was done since the intention of these tests was not to quantify these subjects' performance, but rather to estimate masked thresholds in these two noise levels.

Results indicated that at a signal-to-noise ratio of -15 dB, none of the subjects had any difficulty distinguishing the signal from the noise. At a signal-to-noise ratio of -30 dB, however, the subjects were incapable of hearing the signal at all. At a signal-to-noise ratio of -20 dB, most of the subjects could hear the signal quite well as evidenced by the example ROC curves shown in Figures C1 through C4. Why Subject DM's performance differed so drastically from that of Subject RR at a signal-to-noise ratio of -20 dB in the 90 dBA noise is not known. However, since the subject with a known hearing loss performed at a level only slightly poorer than Subject RR at the -20 dB signal-to-noise ratio conditions, Figures C5 and C6, DM's poor performance was considered to be anomalous. The slight difference between the performance of Subject BT and that of Subject RR was not considered to be substantial, and may have been due to Subject BT's inexperience as a subject rather than to his hearing loss. The only subject run at a signal-to-noise ratio of -25 dB produced an ROC curve in the 85 dB noise level (a signal level of 60 dB) indicating that the signal was near threshold, Figure 7, whereas he was unable to hear the signal at all in the 90 dB noise level (a signal level of 65 dB), Figure 8.

These data seemed to indicate that the masked threshold for the signal at the two noise levels investigated was in the vicinity of 20 to 25 dB below the noise level for

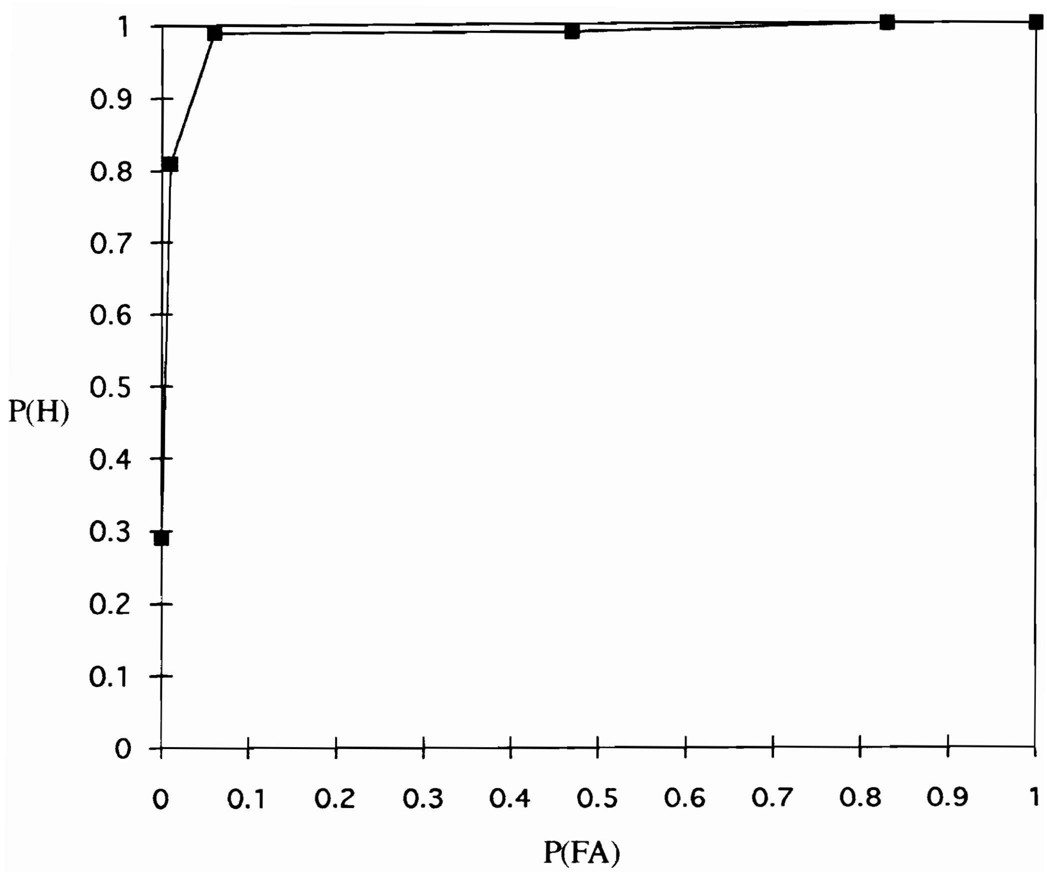


Figure C1. *ROC curve for subject RR in 85 dBA noise with the signal level set to 65 dB (S/N = -20 dB).*

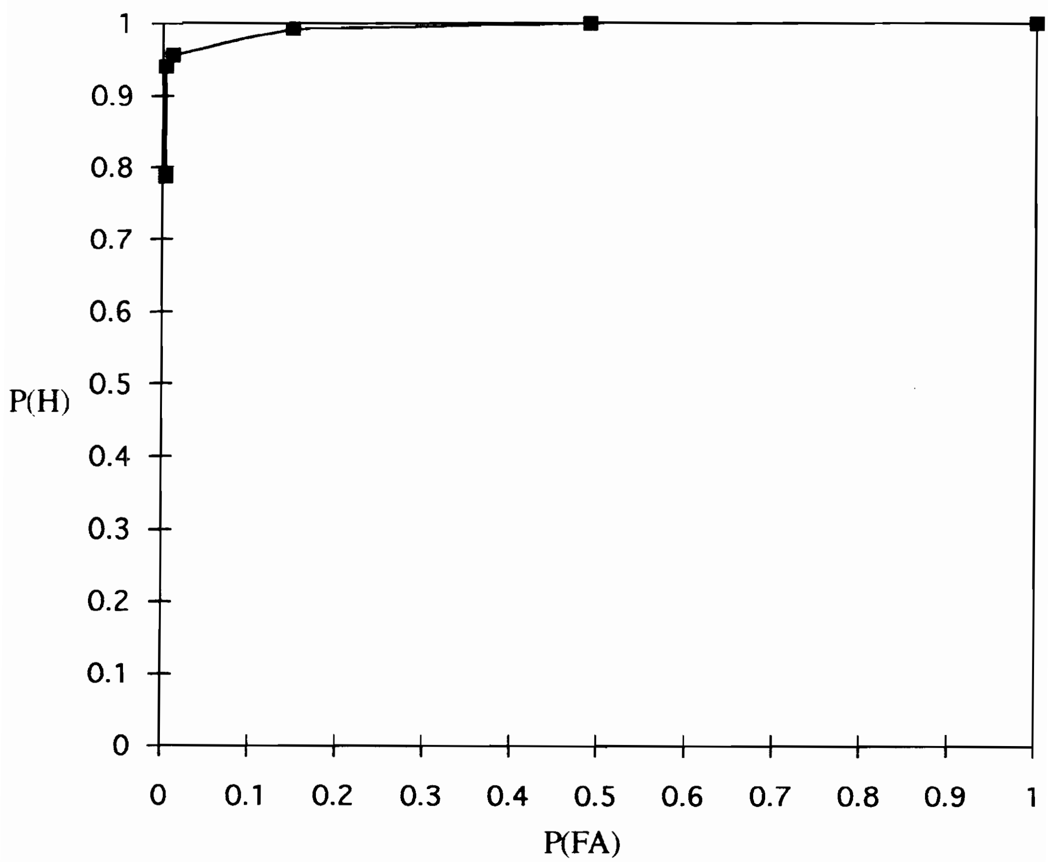


Figure C2. *ROC curve for subject DM in 85 dBA noise with the signal level set to 65 dB (S/N = -20 dB).*

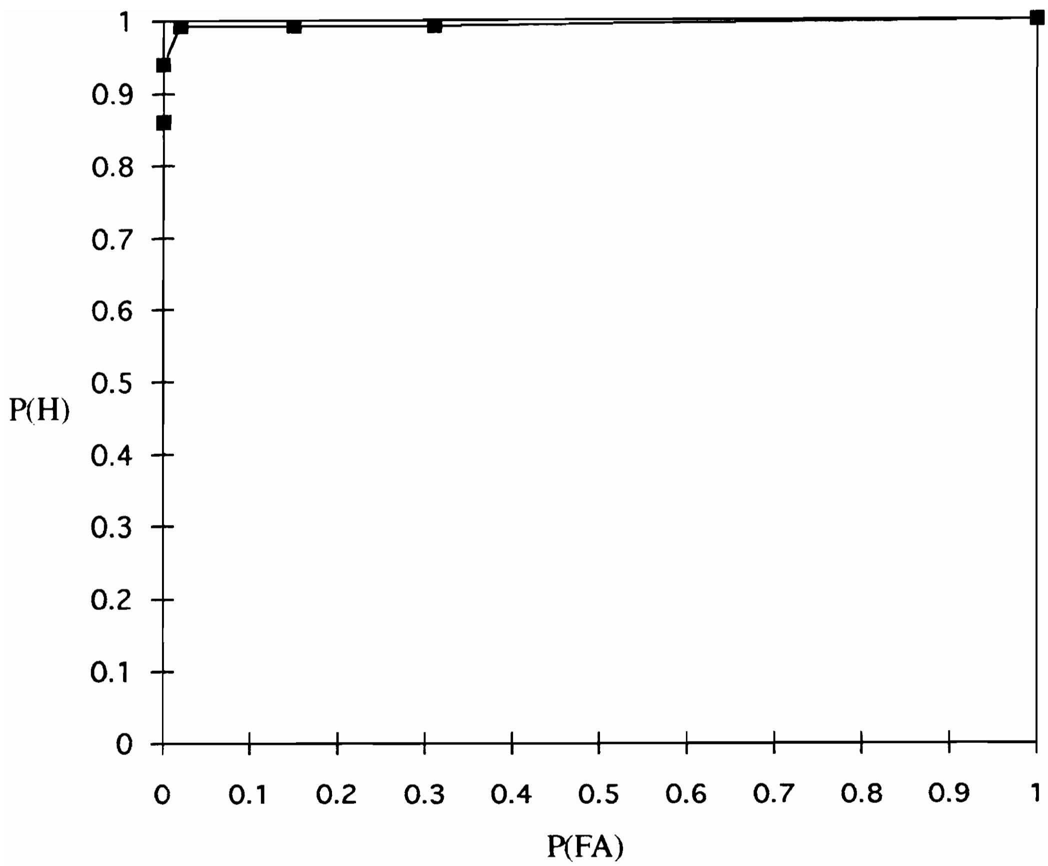


Figure C3. *ROC curve for subject RR in 90 dBA noise with the signal level set to 70 dB (S/N = -20 dB).*

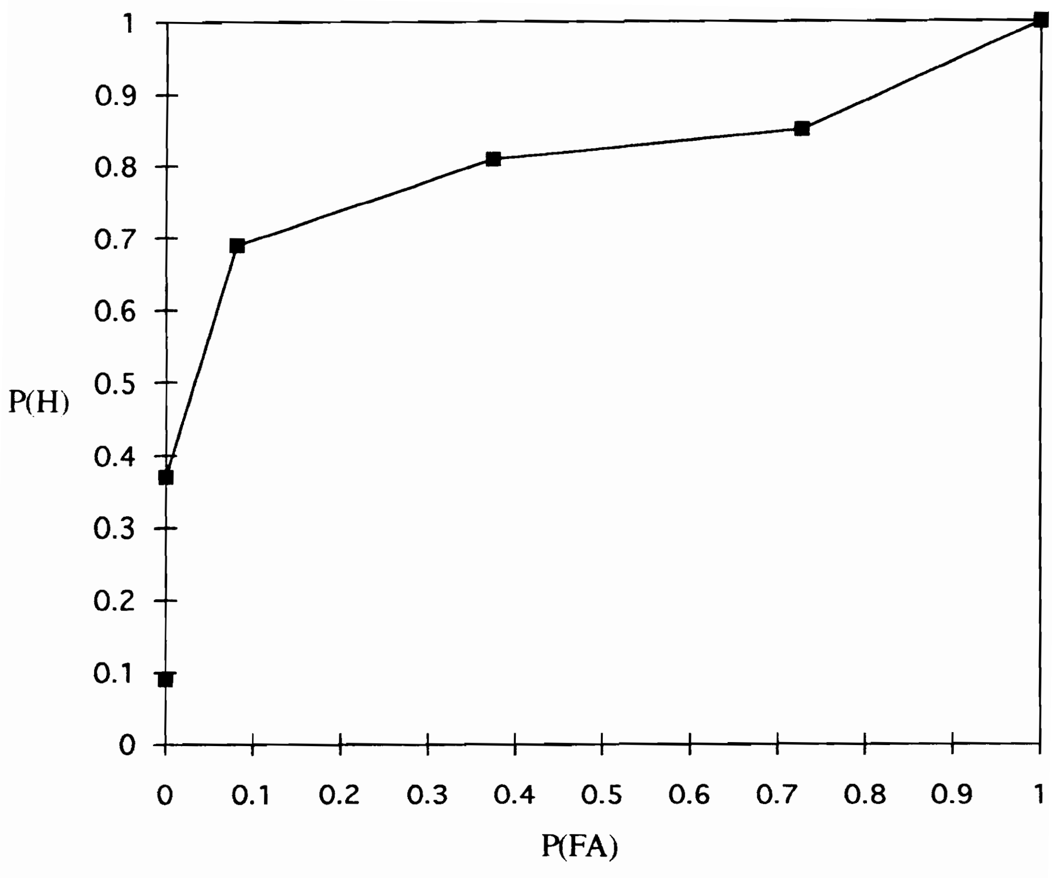


Figure C4. *ROC curve for subject DM in 90 dBA noise with the signal level set to 70 dB (S/N = -20 dB).*

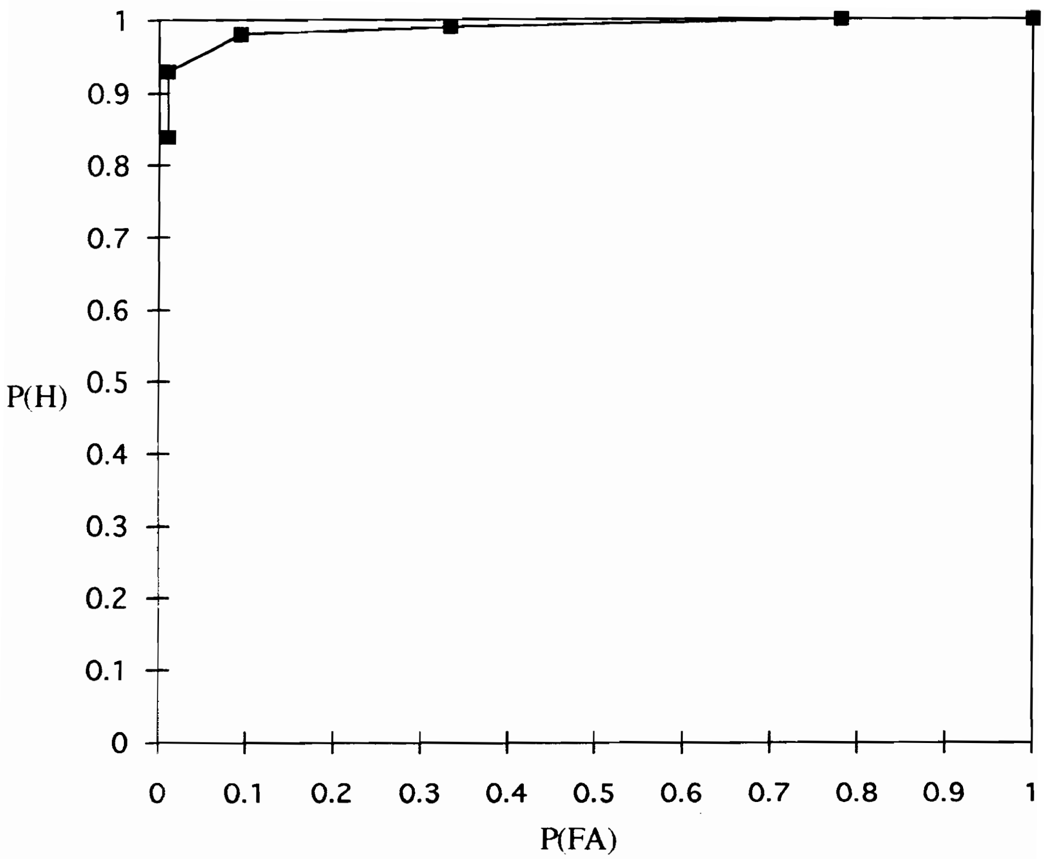


Figure C5. *ROC curve for subject BT in 85 dBA noise with the signal level set to 65 dB (S/N = -20 dB).*

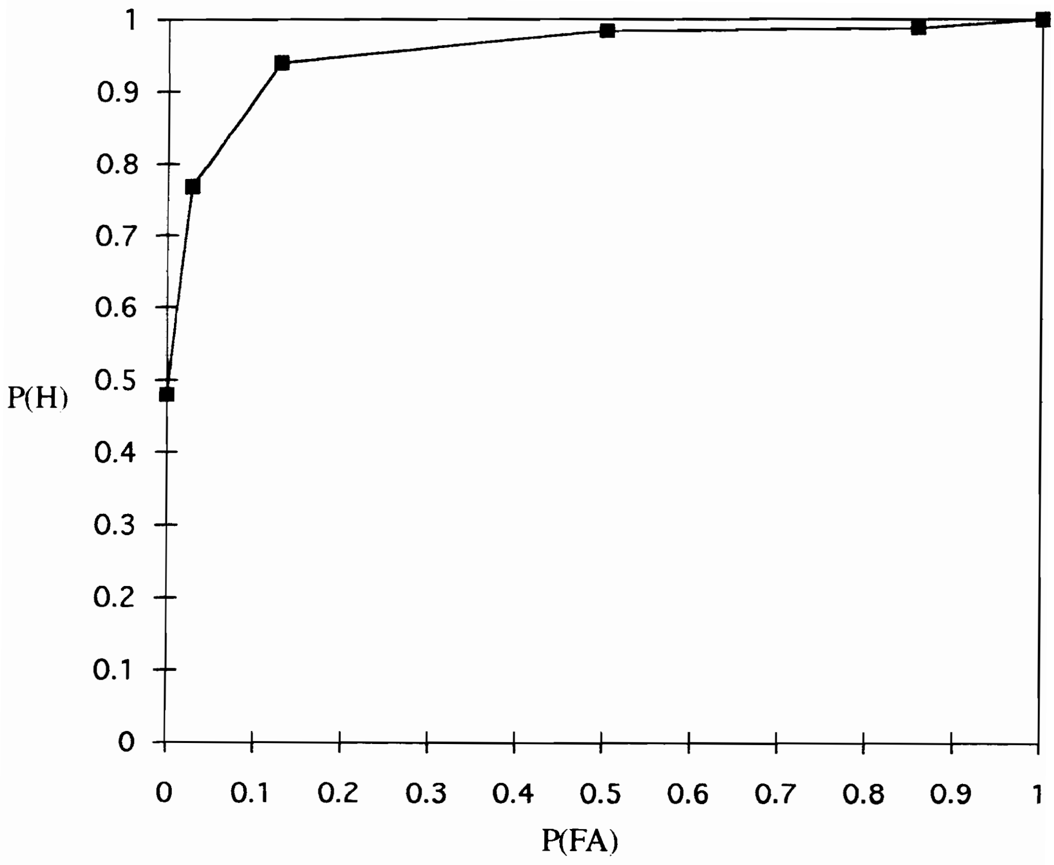


Figure C6. *ROC curve for subject BT in 90 dBA noise with the signal level set to 70 dB (S/N = -20 dB).*

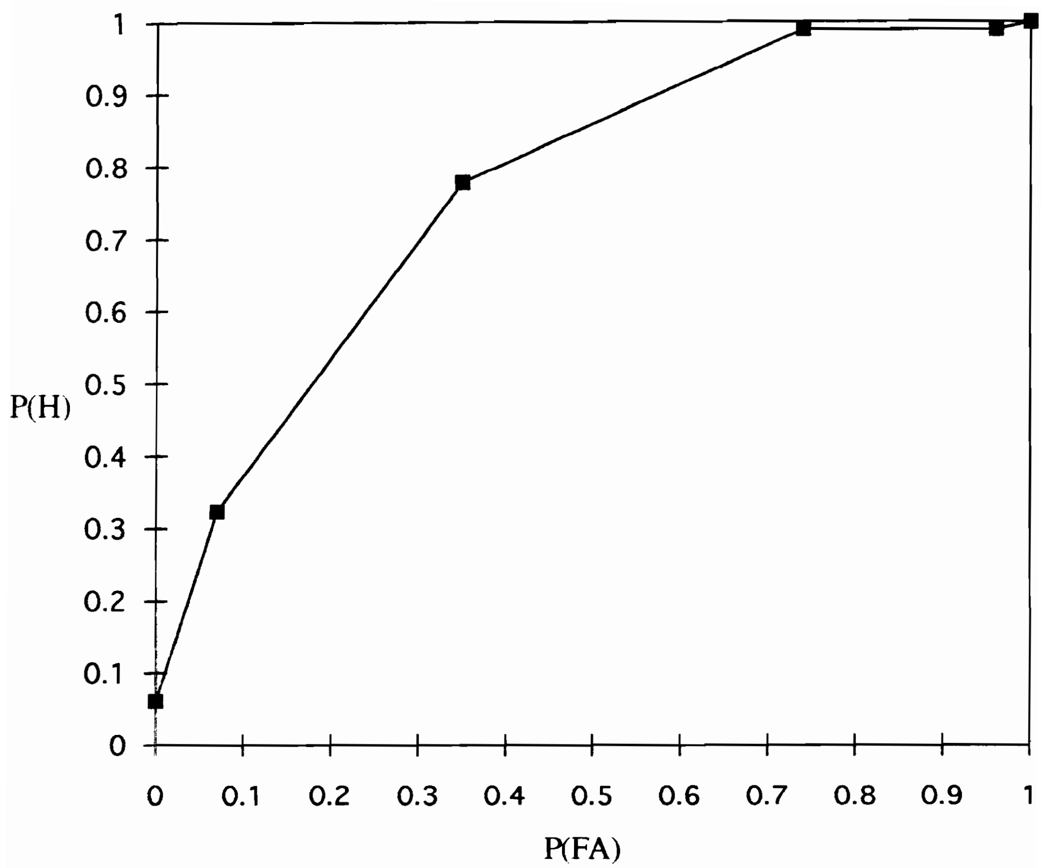


Figure C7. *ROC curve for subject BM in 85 dBA noise with the signal level set to 60 dB (S/N = -25 dB).*

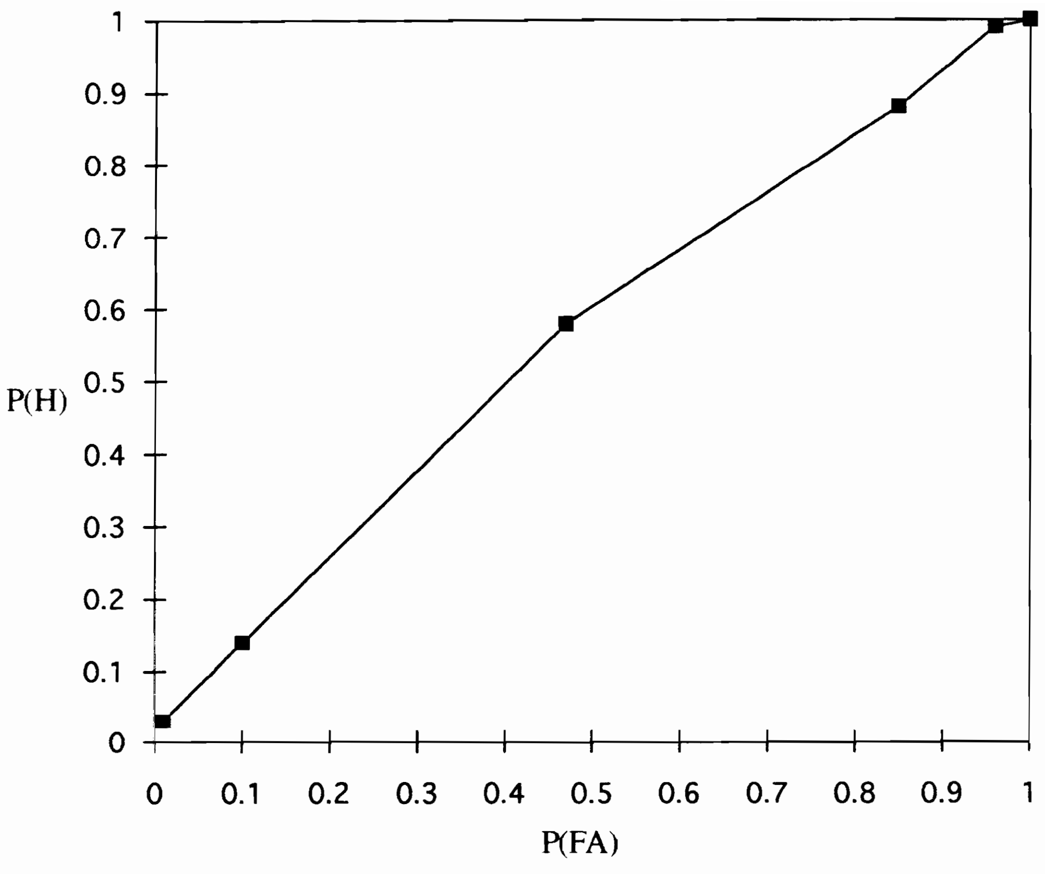


Figure C8. *ROC curve for subject BM in 90 dBA noise with the signal level set to 65 dB (S/N = -25 dB).*

normal listeners. This answered the first question dealing with the masked threshold for the signal. However, it was still necessary to determine what scheme to use for the signal level independent variable.

It was desired to set the signal levels such that the masked threshold for all of the subjects would be contained in the range of levels used. Otherwise, there would be no need for including a group if they would never be able to hear the signal or would never have any trouble hearing the signal. To include the masked threshold for normal listeners in this range, it would be necessary to have the lowest signal level used be in the vicinity of 60 dBA (a signal-to-noise ratio of -25 dB in 85 dBA noise). Although the masked threshold for the most hearing-impaired listeners was not known, the maximum practical signal level was considered to be 95 dBA (providing a signal-to-noise ratio of 0 dB in the 95 dBA noise condition). Using these values as upper and lower limits, the signal level would have to be varied over a range of 35 dB if discrete signal levels were used across all three noise levels. With only four levels to be used in the experiment, it would have been necessary to use levels of 95, 83, 71, and 59 dBA to completely include this range. The 12 dB difference between adjacent signal levels was considered to be too large to be practical. In addition, it was feared that the large difference between adjacent levels could lead to a situation wherein listeners would always be able to hear the signal in one condition and then never hear the signal in the next condition, completely skipping the region where they would have some uncertainty as to whether or not they heard the signal and thus generate less than ideal data. If signal-to-noise ratio were used instead of absolute signal level, however, the difference between adjacent signal levels could be cut to 8 dB if the levels used were 0, -8, -16, and -24 dB. This was considered to be a better solution and one that was believed to be likely to produce more reliable data.

It was therefore decided to use signal-to-noise ratio as the independent variable and to present the signal at signal-to-noise ratios of 0, -8, -16, and -24 dB.

APPENDIX D

Detailed Instructions Given Each Subject During the Training Session.

Experimental Instructions

(Signal Detection portion of the experiment)

During this part of the experiment, short periods of noise will be presented at equally spaced intervals. Within each noise interval, a signal may or may not also be presented. Your task will be to indicate whether or not you hear(d) the signal and how sure you are of your response using the keyboard in front of you.

There will be a total of 504 trials/session, separated into 6 blocks of 84 trials each. A signal will be presented in exactly half of the trials in each block (42 times/block or 252 times/session). The levels at which the noise and signals are played will remain constant throughout the entire session. In other words, I **will not** be adjusting the noise or signal levels trial-to-trial or block-to-block.

Each of the 504 trials will be structured in the following way.

- The word "WARNING" (in red) will be displayed at the top of the monitor directly in front of you to indicate that a trial interval is starting and the noise is about to be presented.
- The noise (and possibly a signal) will then be presented.
- Almost simultaneously with the initiation of the noise (and possibly signal) a large question mark will be displayed on the monitor and will continue to be displayed for a couple of seconds after the noise in terminated. You may respond at any time while the question mark is displayed.
- At the end of the response interval (when the question mark disappears), one of two windows will be displayed to indicate whether or not a signal was presented with the noise in the preceding trial. If a signal was presented, a window containing the words "Yes -- Signal" (in green) will appear in the center of the screen. If a signal was not presented, a window containing the words "No -- Signal" (in blue) will appear at the bottom of the screen.
- After a few seconds, the Yes or No windows will disappear and the warning message will reappear indicating the start of a new trial.
- **Do you have any questions before I explain how you should try to respond?**

As you can see, there are 6 possible responses available to you which run the gamut from "Definitely Did Not Hear Signal" to "Definitely Heard signal." I'll go through each response in order from left to right.

- •• If you are absolutely sure that you did not hear a signal during the trial interval, then you would press the key on the far left.
- •• If you don't think you heard a signal, but your not absolutely sure, you might press the second key.
- •• If you don't think you heard a signal, but your not really sure, then you might press the third key.
- •• Likewise, if your think you did hear a signal, but your not really sure, you might press the fourth key.
- •• If you think you did hear a signal, but your not absolutely sure, you might press the fifth key.
- •• And finally, if you are absolutely sure you did hear a signal, then you would press the sixth key (on the far right).

Please respond in a manner which reflects how you hear the signals, do not try to guess. Also, try to use as many categories as you can consistently use. Although there may be some conditions in which you may always be able to distinguish the signal from the noise and others in which you may not be able to hear the signal at all, it is hoped that most of the conditions will be such that there may be some uncertainty on your part. Remember, you may respond at any time while the question mark is being displayed.

If, at any time during a block of trials you feel it necessary to stop the experiment, just press the button labeled "Emergency Stop" a couple of times. This will interrupt the experiment.

You may relax between blocks of trials. However, it is important that you not remove or adjust the earmuff at any time during a session. If it becomes absolutely necessary to adjust the earmuff, please tell me and I will adjust it.

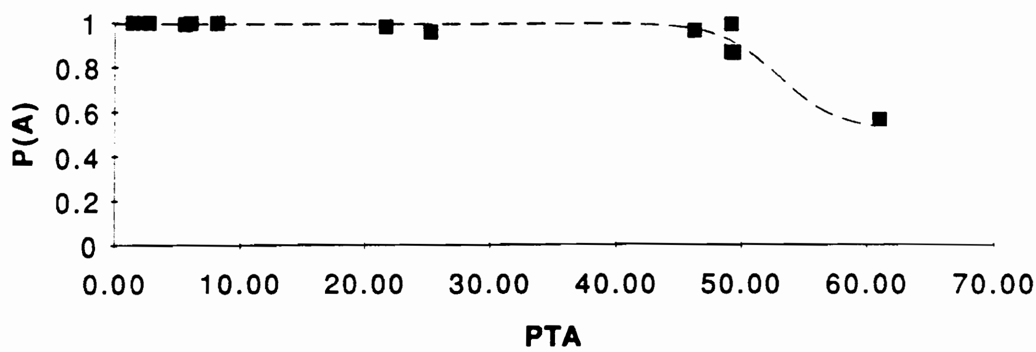
Before we start with the experiment, the signal and noise levels to be used in the session will be previewed. First, six presentations of only the signal will occur, second, six presentations of only the noise will occur, and finally, six presentations the signal and noise will occur. There is no need for you to respond during this preview.

Do you have any questions before we begin?

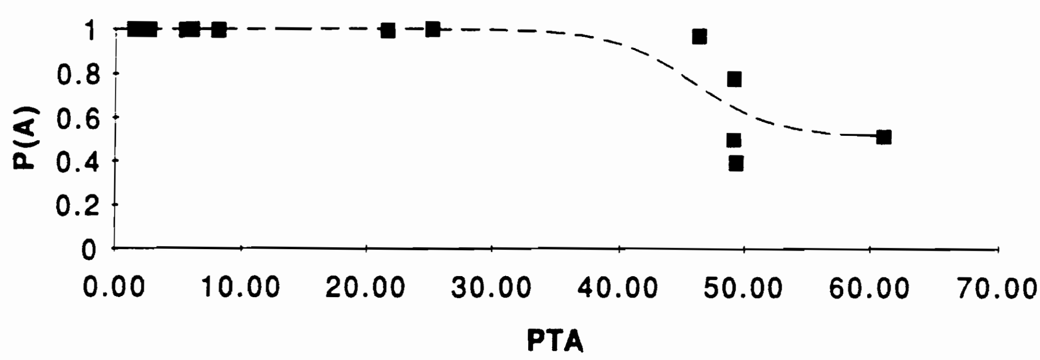
APPENDIX E

Plots of the Fitted Regression Model Against Experimental Data When Using the Binaural PTA from 500 to 2000 Hz to Quantify Hearing Level.

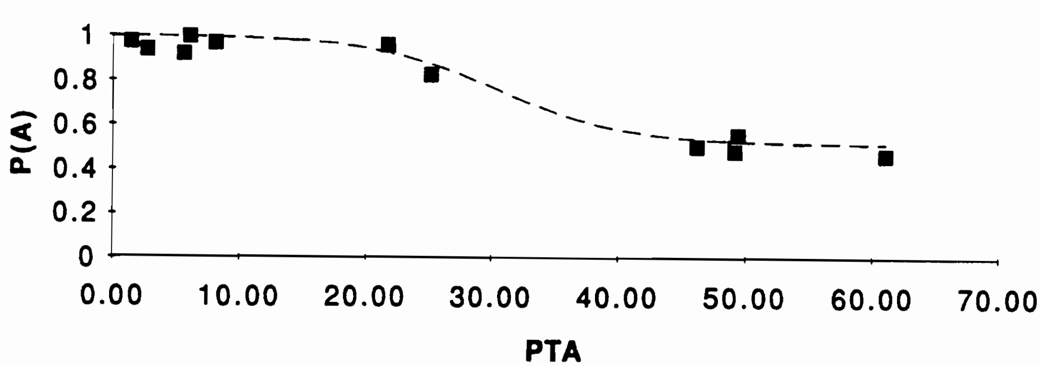
NL = 85 dBA, SL = 85 dBA

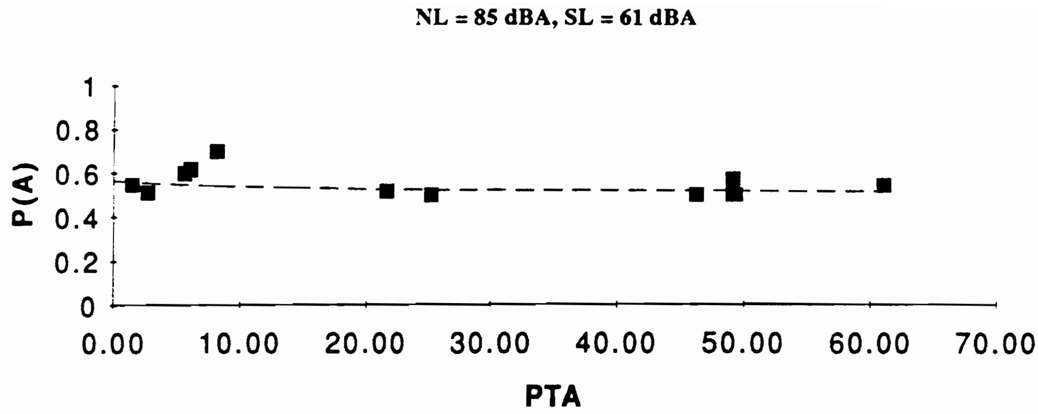


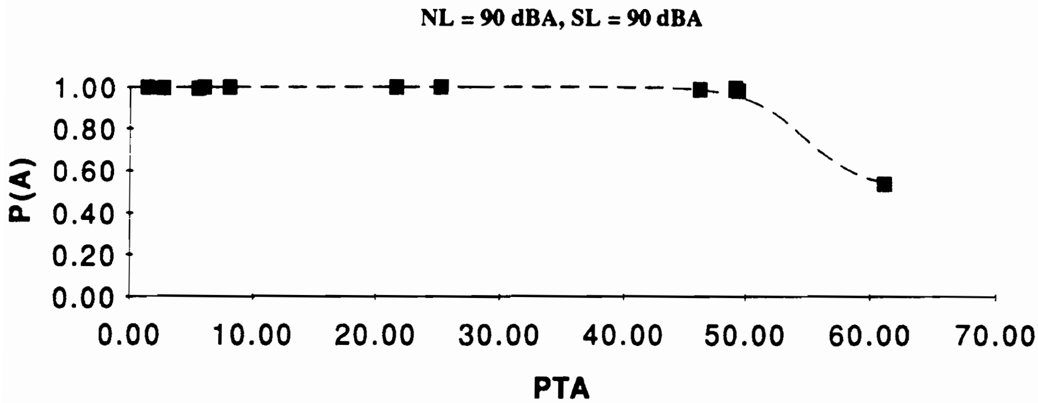
NL = 85 dBA, SL = 77 dBA

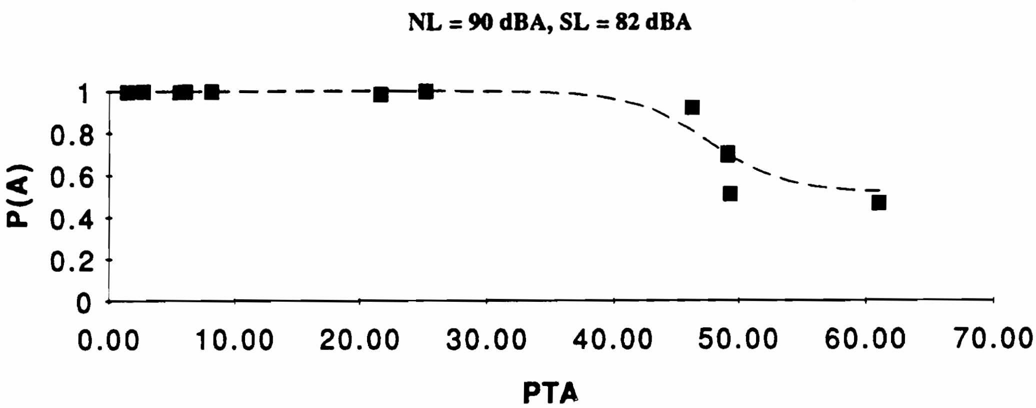


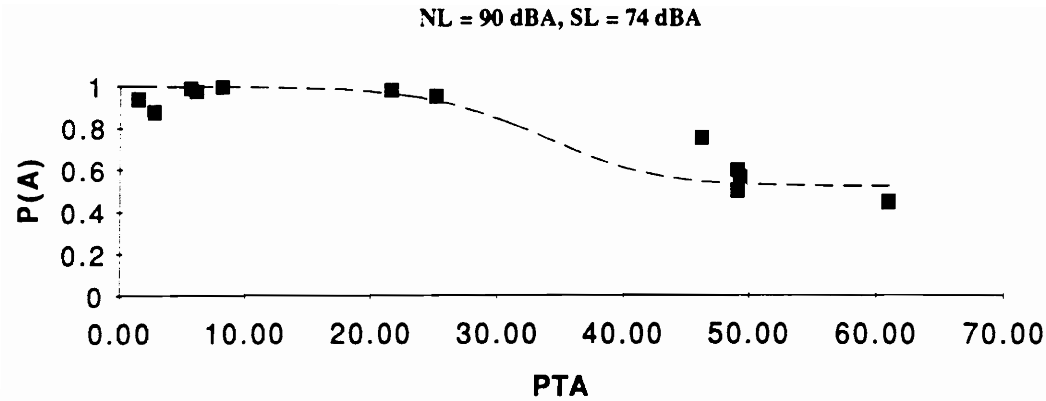
NL = 85 dBA, SL = 69 dBA

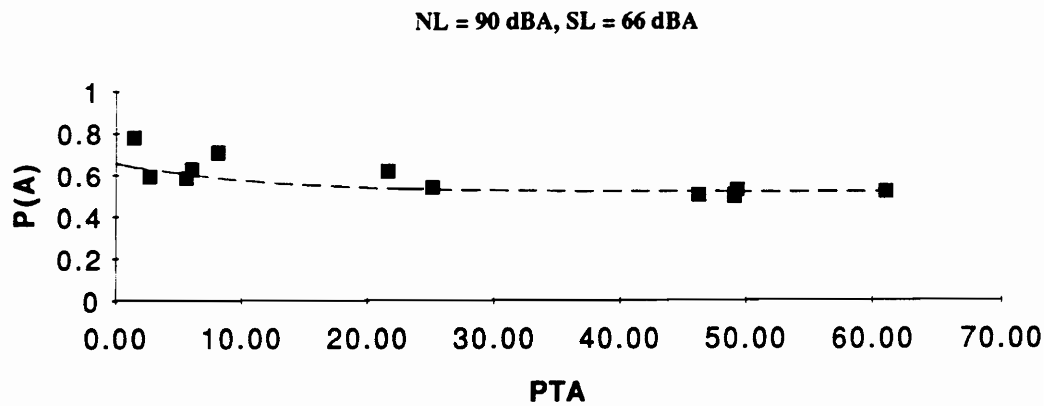


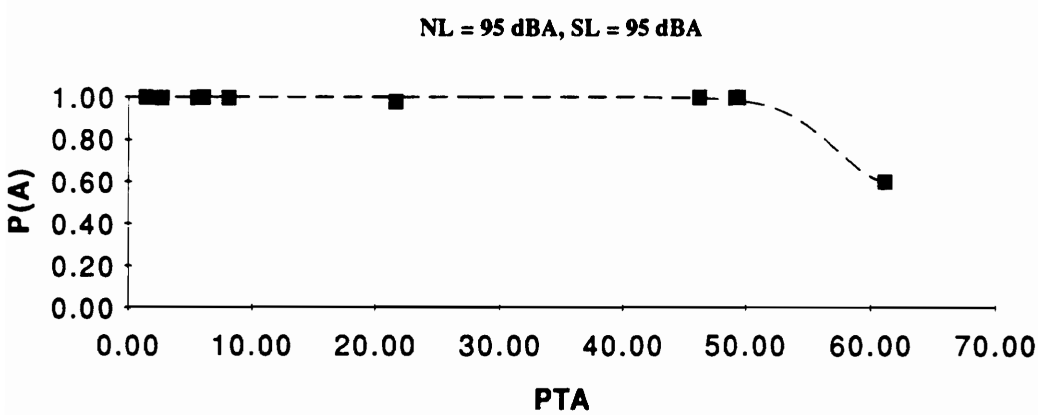

NL = 85 dBA, SL = 61 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00 10.00 20.00 30.00 40.00 50.00 60.00 70.00
PTA


NL = 90 dBA, SL = 90 dBA
P(A)
1.00
0.80
0.60
0.40
0.20
0.00
0.00 10.00 20.00 30.00 40.00 50.00 60.00 70.00
PTA

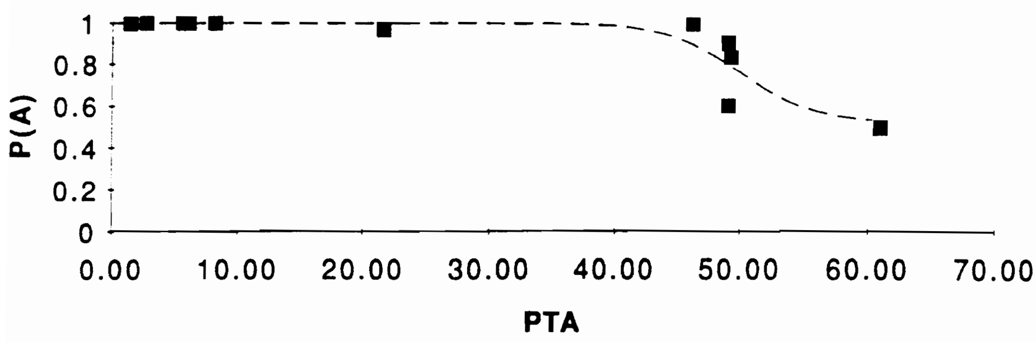

NL = 90 dBA, SL = 82 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00 10.00 20.00 30.00 40.00 50.00 60.00 70.00
PTA


NL = 90 dBA, SL = 74 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

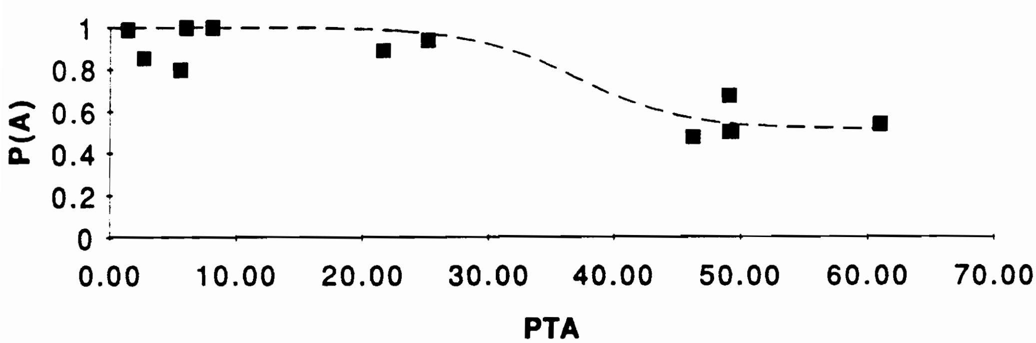

NL = 90 dBA, SL = 66 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA


NL = 95 dBA, SL = 95 dBA
P(A)
1.00
0.80
0.60
0.40
0.20
0.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

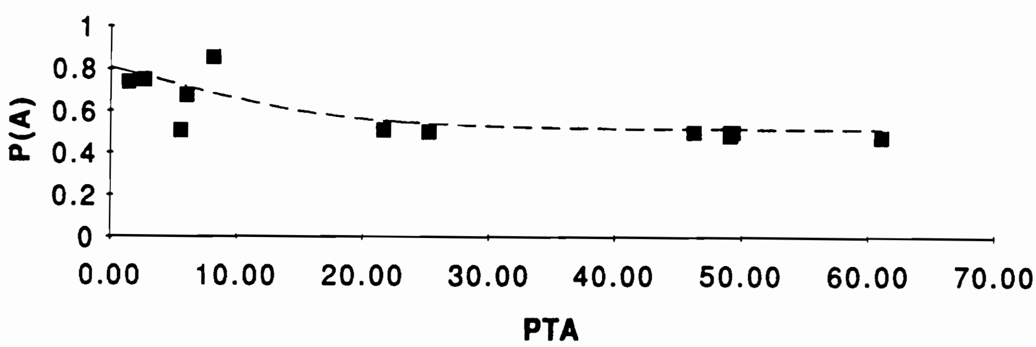
NL = 95 dBA, SL = 87 dBA



NL = 95 dBA, SL = 79 dBA

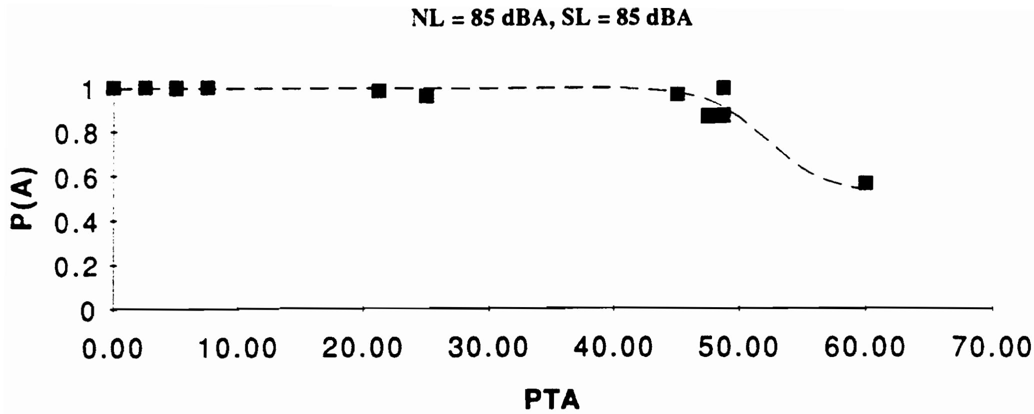


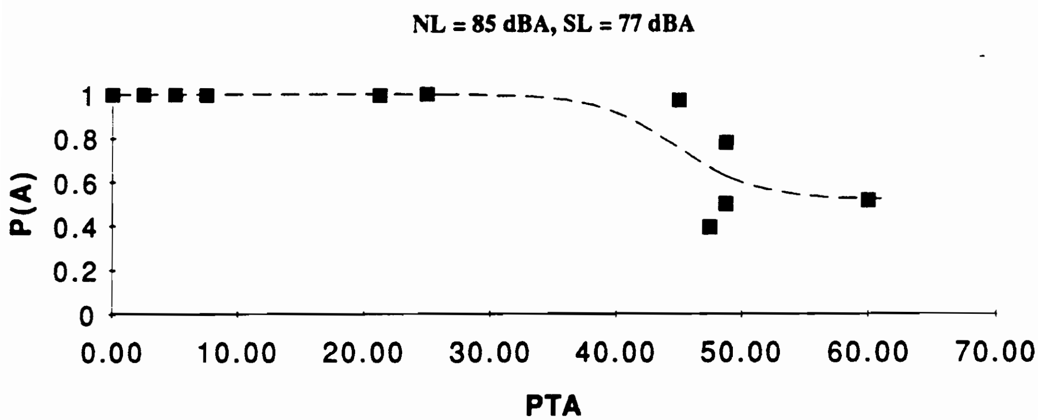
NL = 95 dBA, SL = 71 dBA

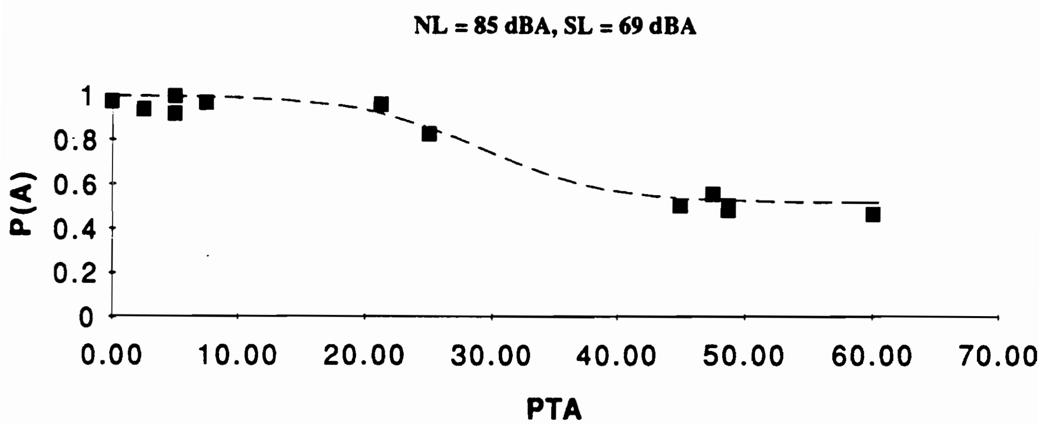


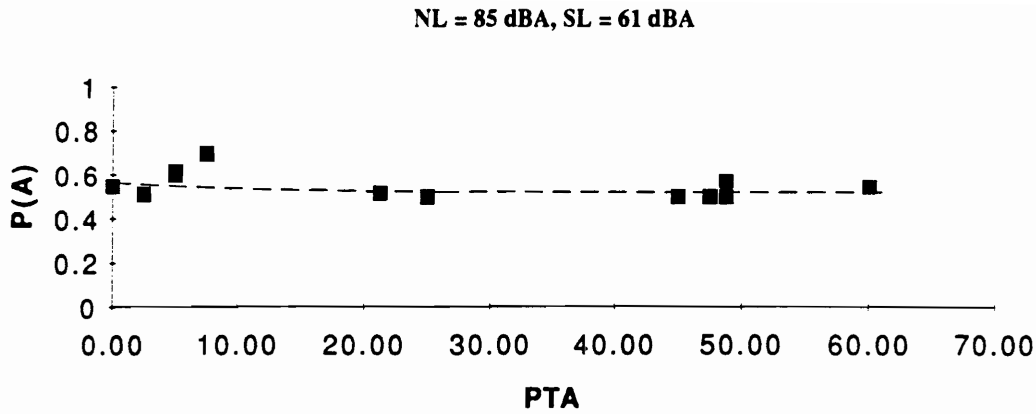
APPENDIX F

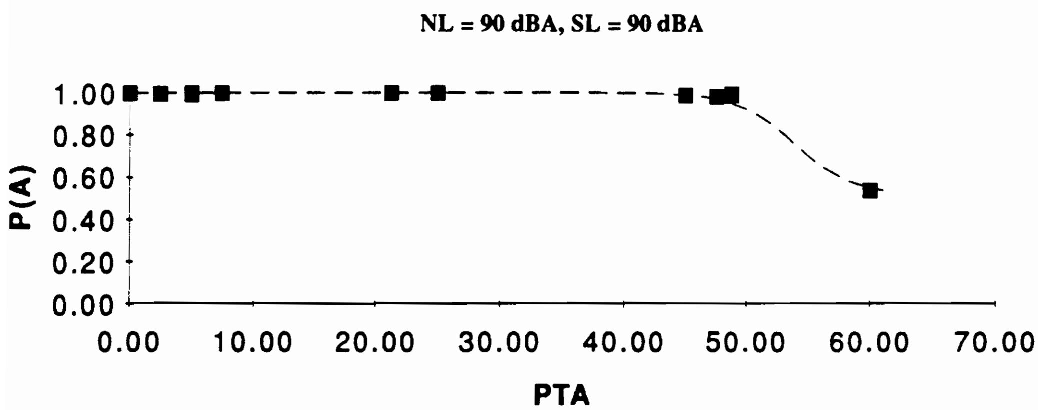
Plots of the Fitted Regression Model Against Experimental Data When Using the Minimum Monaural PTA from 500 to 2000 Hz to Quantify Hearing Level.

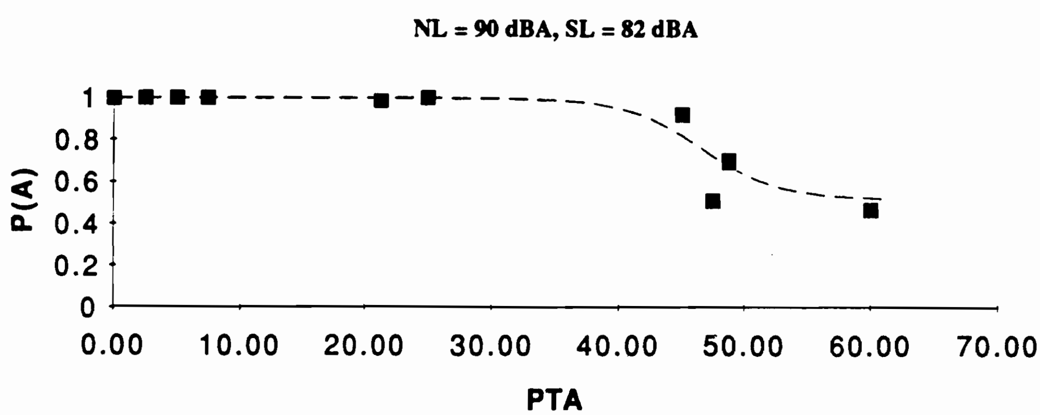

NL = 85 dBA, SL = 85 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

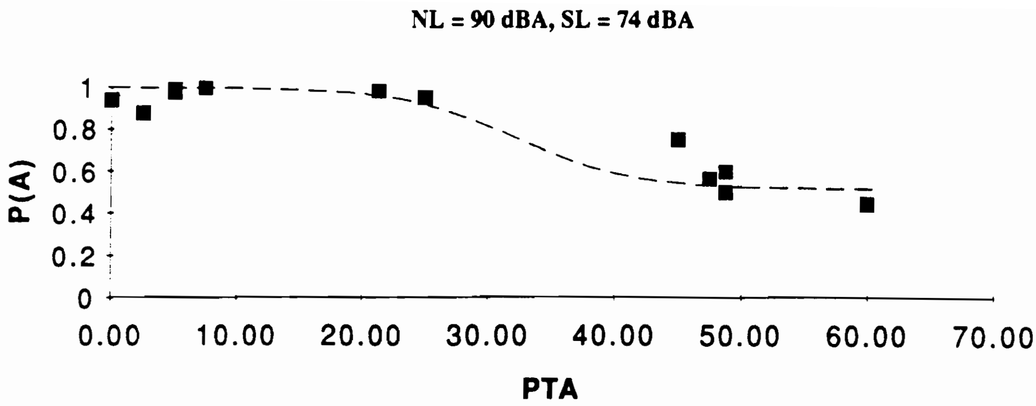

NL = 85 dBA, SL = 77 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

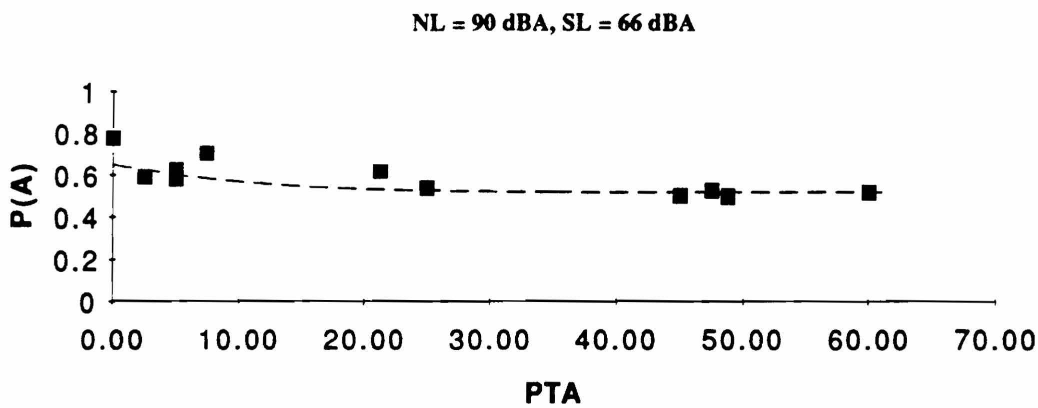

NL = 85 dBA, SL = 69 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

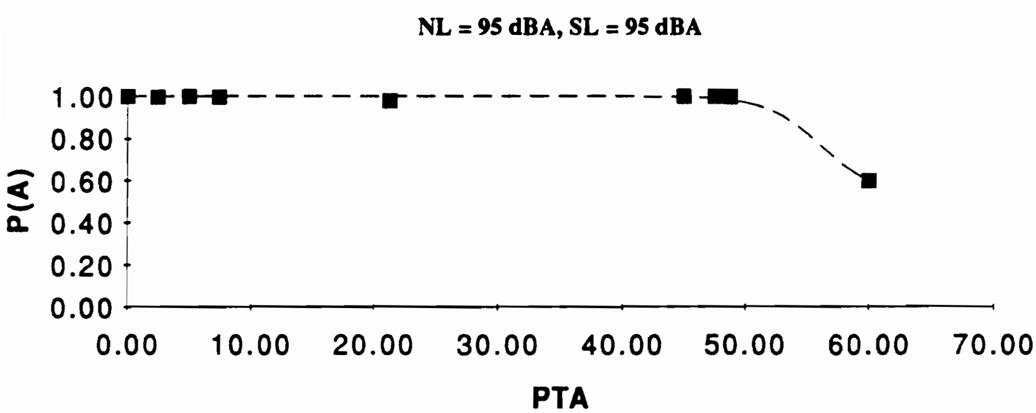

NL = 85 dBA, SL = 61 dBA
1
0.8
0.6
0.4
0.2
0
P(A)
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA


NL = 90 dBA, SL = 90 dBA
1.00
0.80
0.60
0.40
0.20
0.00
P(A)
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

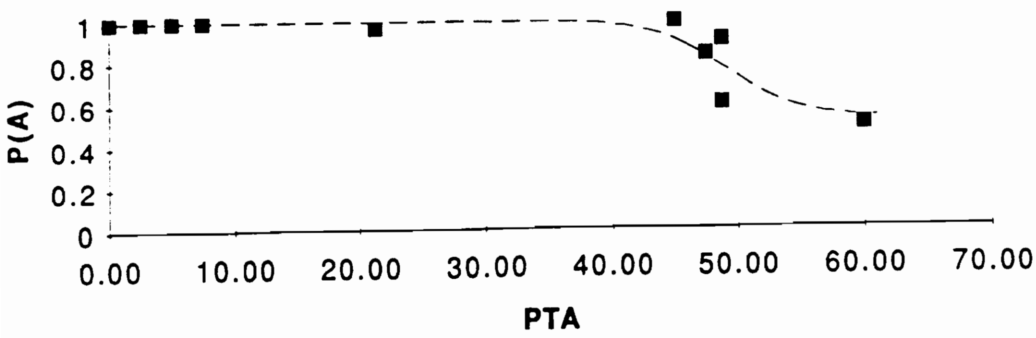

NL = 90 dBA, SL = 82 dBA
1
0.8
0.6
0.4
0.2
0
P(A)
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA


NL = 90 dBA, SL = 74 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

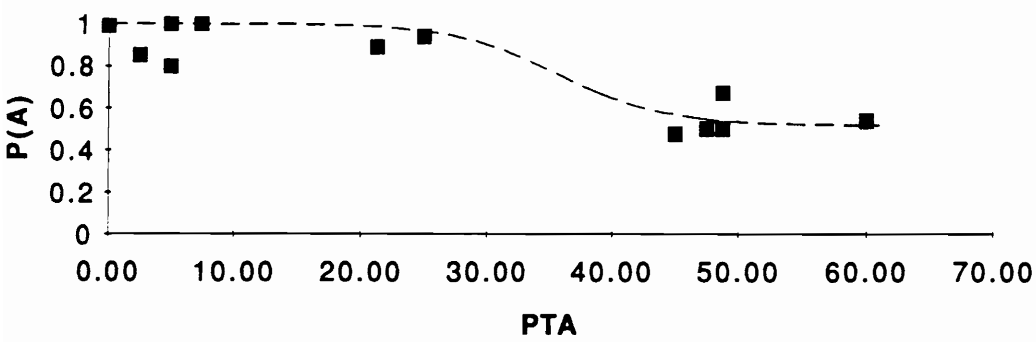

NL = 90 dBA, SL = 66 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA


NL = 95 dBA, SL = 95 dBA
P(A)
1.00
0.80
0.60
0.40
0.20
0.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

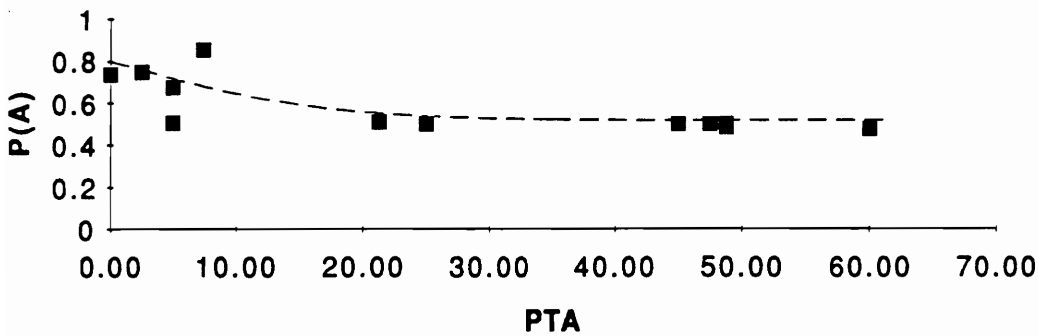
NL = 95 dBA, SL = 87 dBA



NL = 95 dBA, SL = 79 dBA



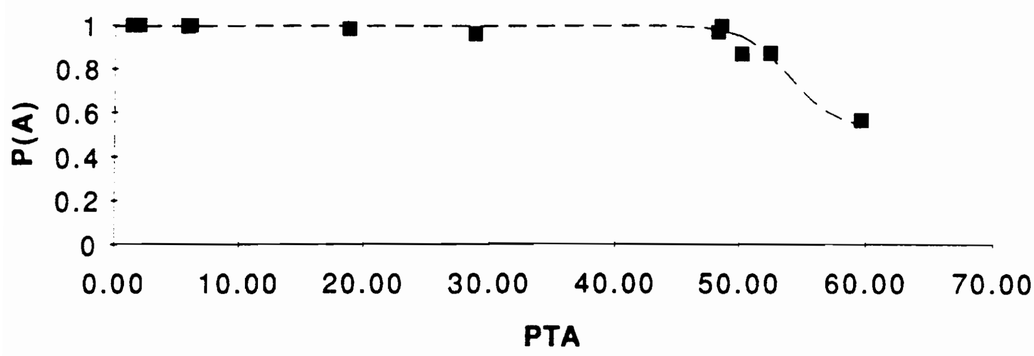
NL = 95 dBA, SL = 71 dBA



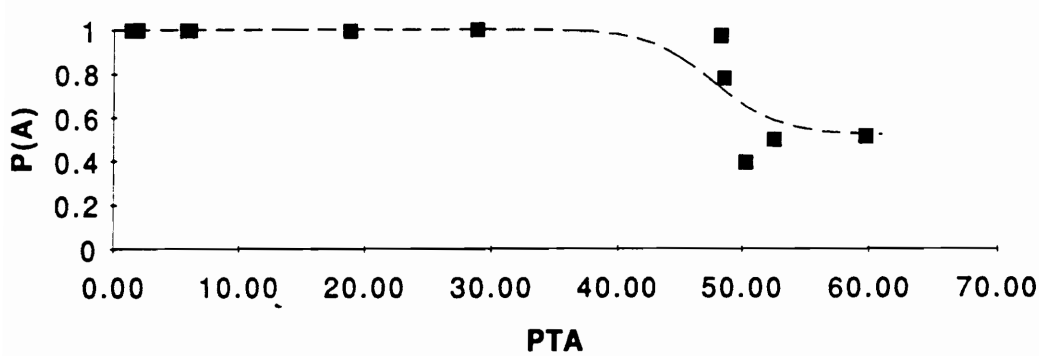
APPENDIX G

Plots of the Fitted Regression Model Against Experimental Data When Using the Binaural PTA from 1000 to 2000 Hz to Quantify Hearing Level.

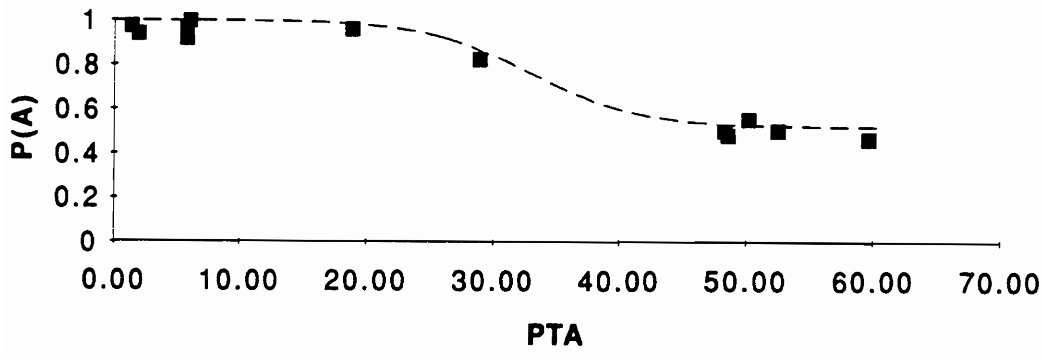
NL = 85 dBA, SL = 85 dBA



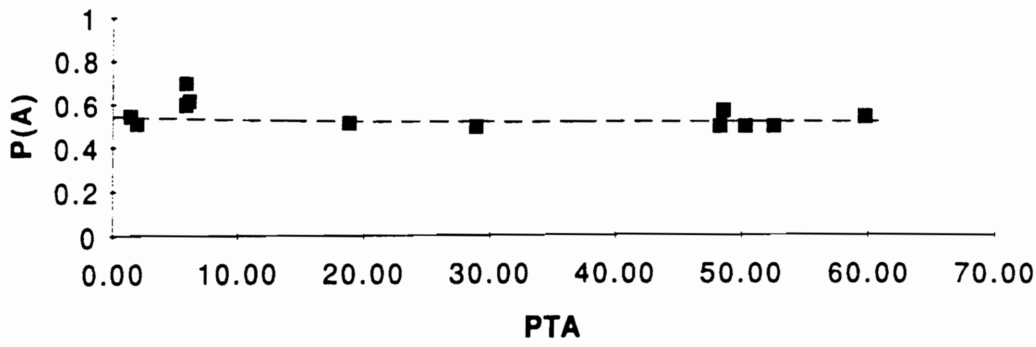
NL = 85 dBA, SL = 77 dBA



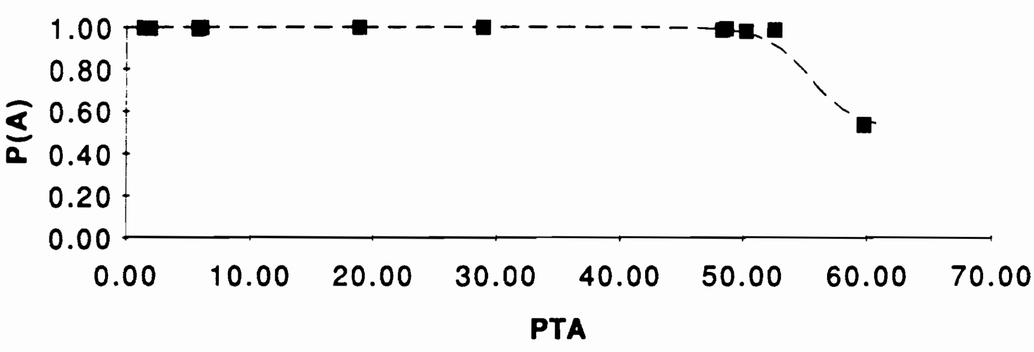
NL = 85 dBA, SL = 69 dBA



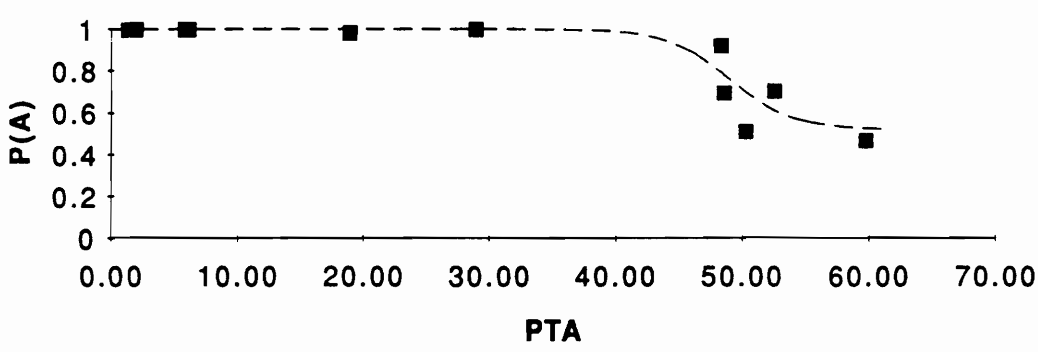
NL = 85 dBA, SL = 61 dBA



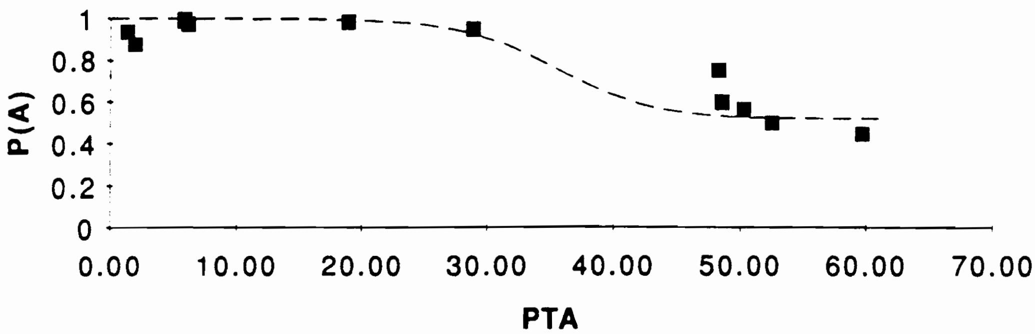
NL = 90 dBA, SL = 90 dBA



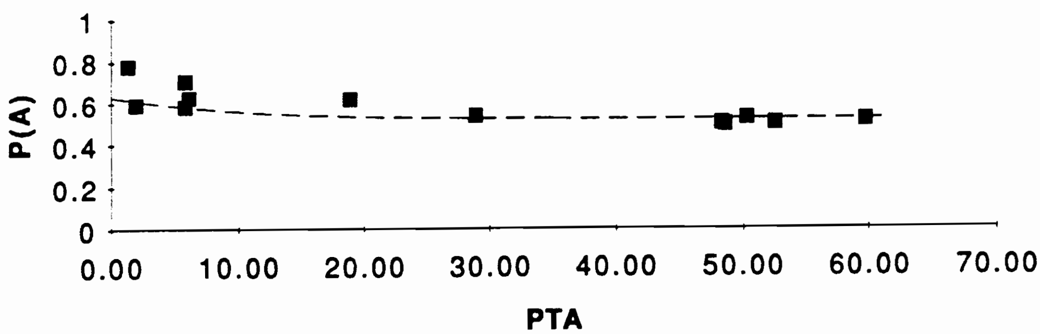
NL = 90 dBA, SL = 82 dBA



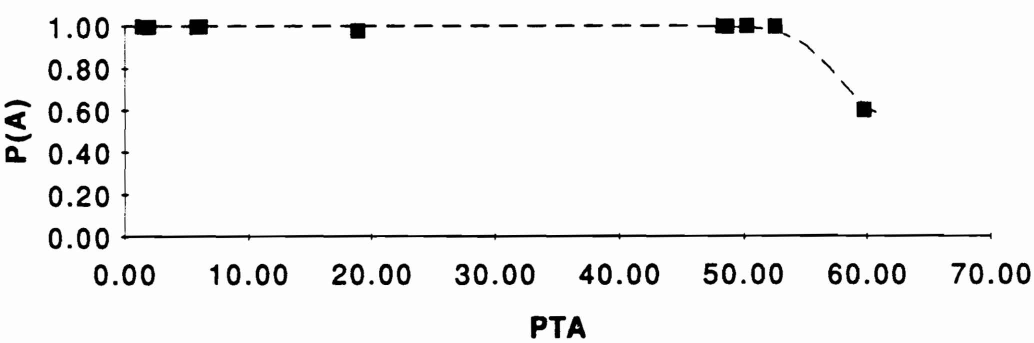
NL = 90 dBA, SL = 74 dBA

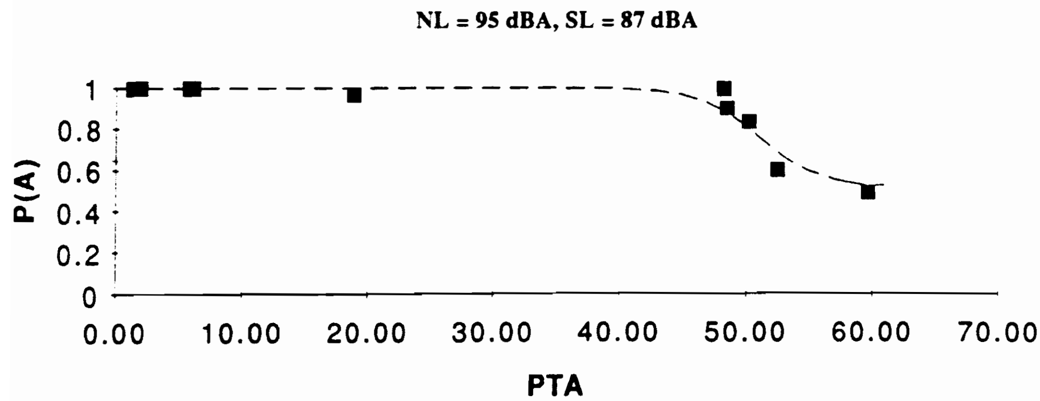


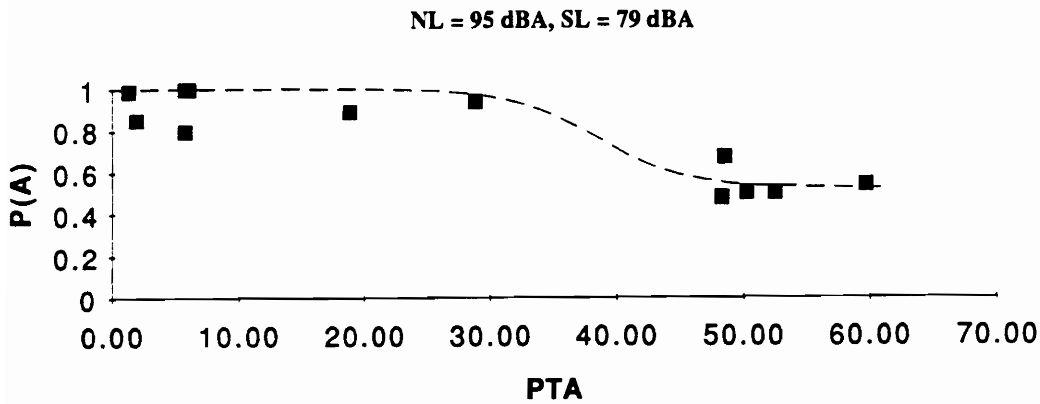
NL = 90 dBA, SL = 66 dBA

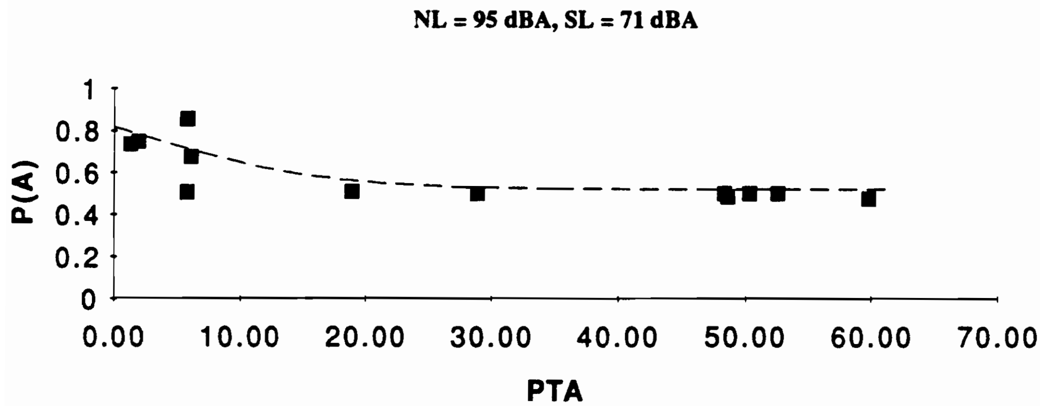


NL = 95 dBA, SL = 95 dBA



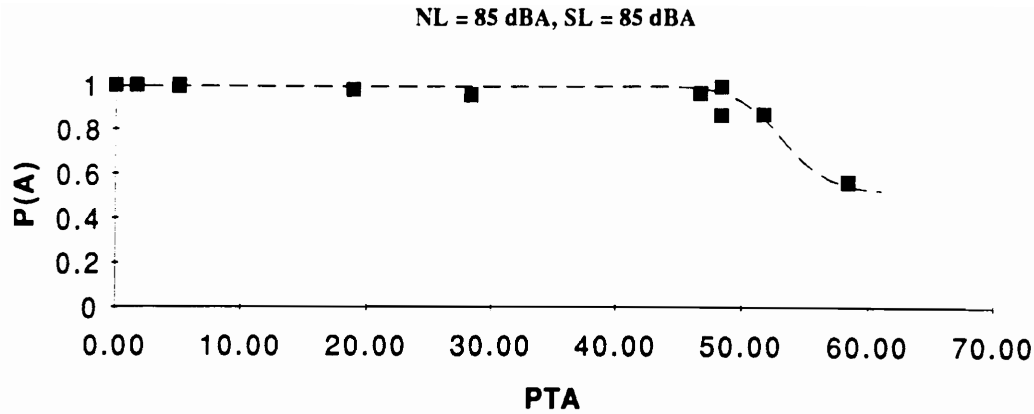

NL = 95 dBA, SL = 87 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

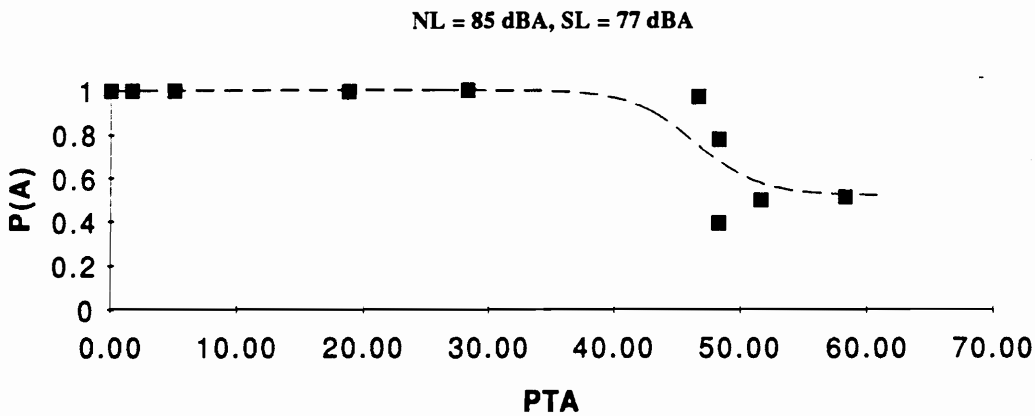

NL = 95 dBA, SL = 79 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

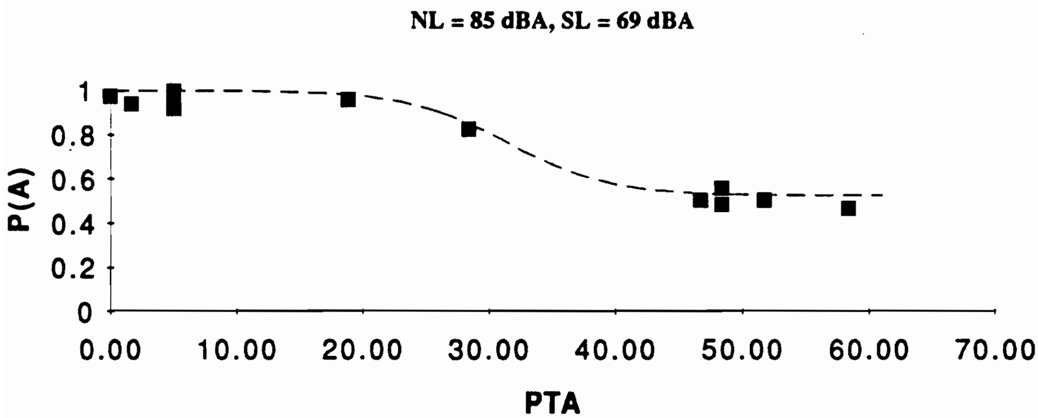

NL = 95 dBA, SL = 71 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

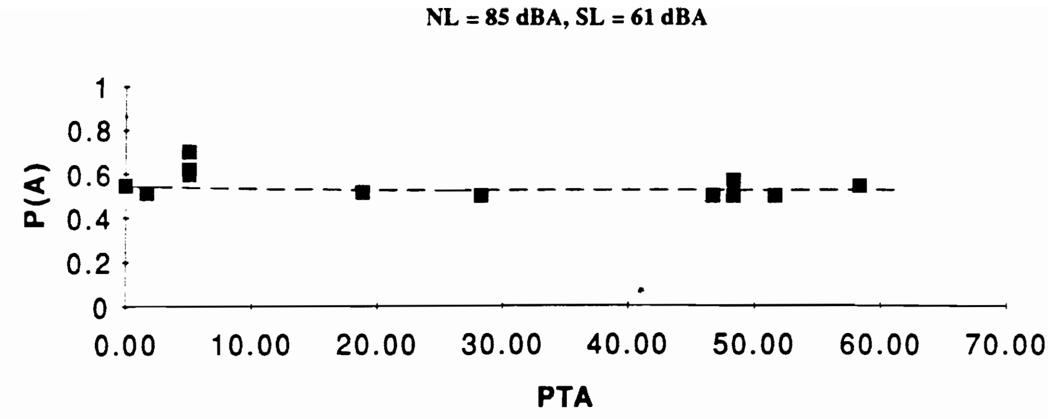
APPENDIX H

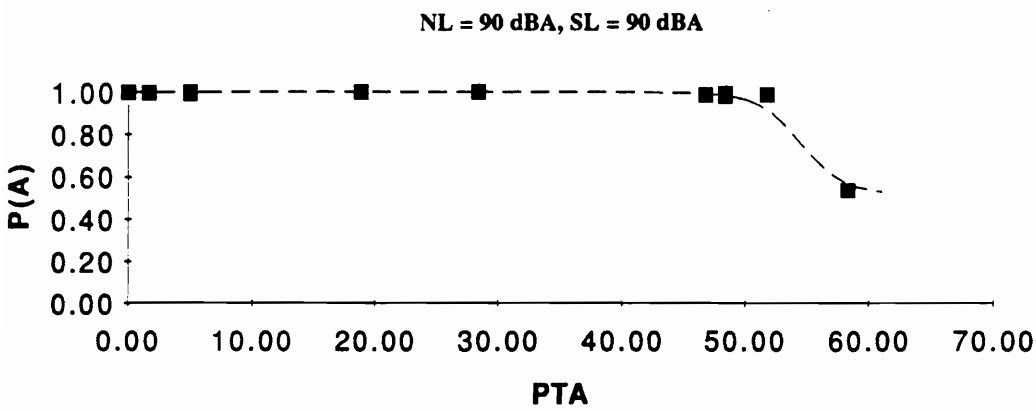
Plots of the Fitted Regression Model Against Experimental Data When Using the Minimum Monaural PTA from 1000 to 2000 Hz to Quantify Hearing Level.

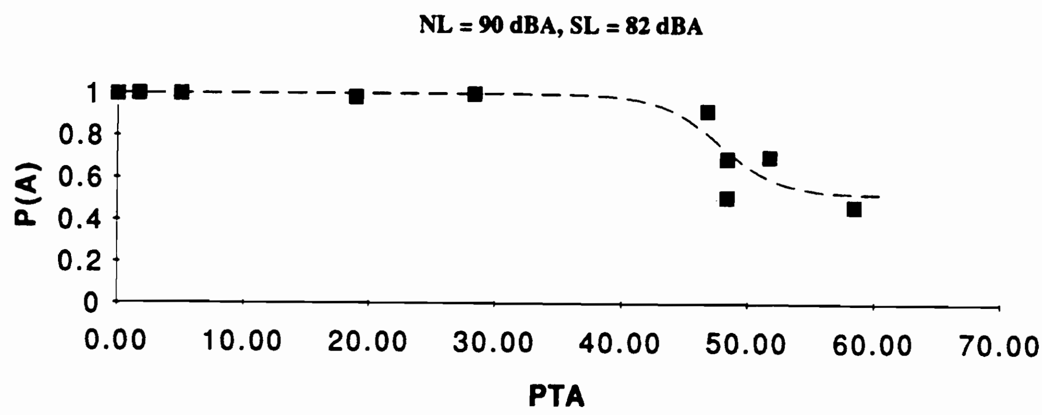

NL = 85 dBA, SL = 85 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

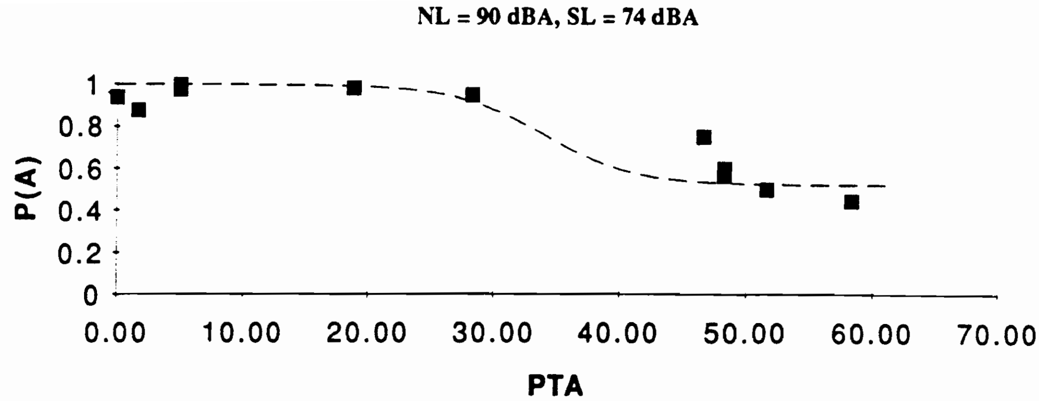

NL = 85 dBA, SL = 77 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

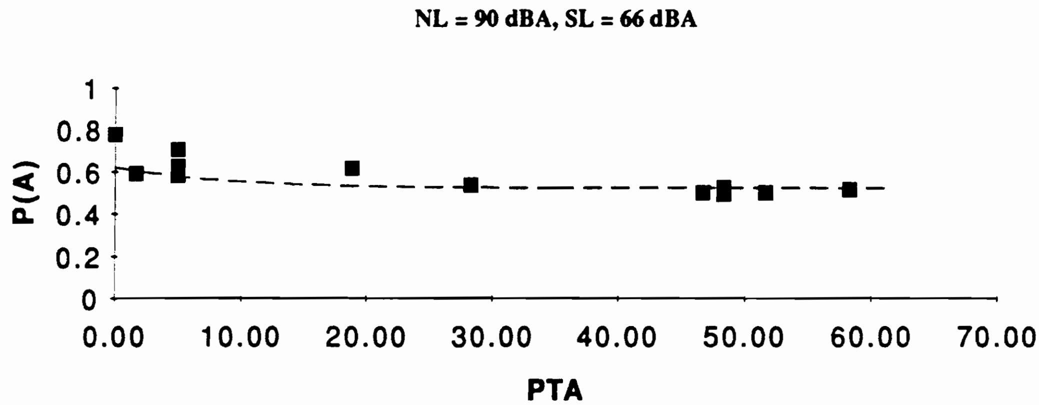

NL = 85 dBA, SL = 69 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

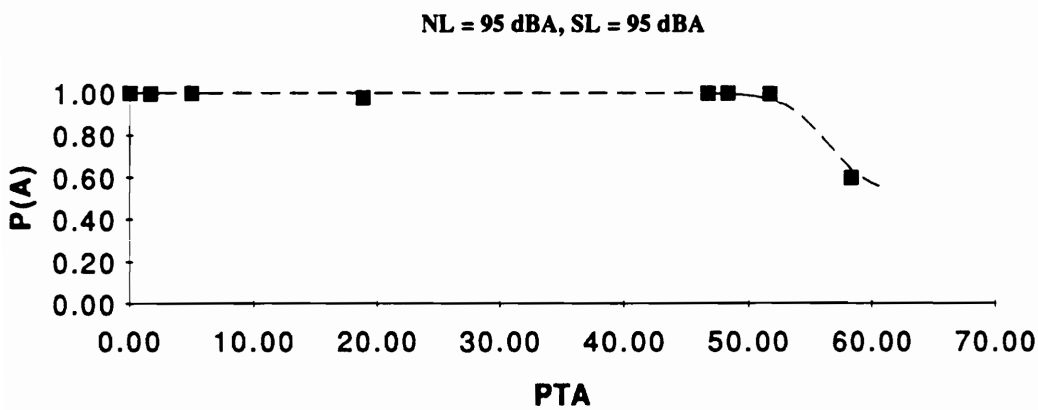

NL = 85 dBA, SL = 61 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA


NL = 90 dBA, SL = 90 dBA
P(A)
1.00
0.80
0.60
0.40
0.20
0.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

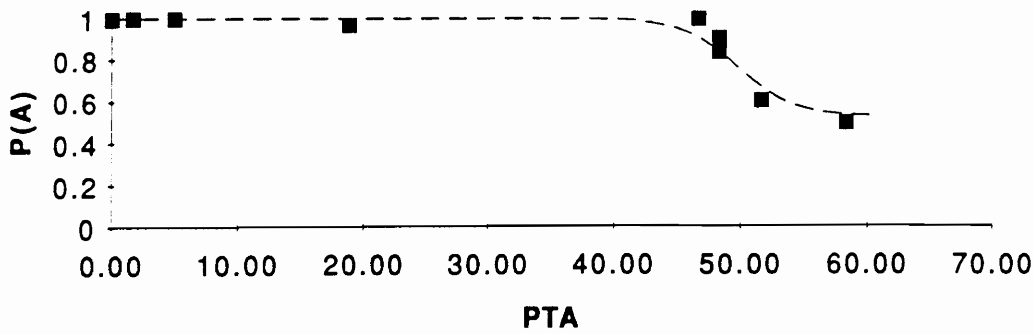

NL = 90 dBA, SL = 82 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA


NL = 90 dBA, SL = 74 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

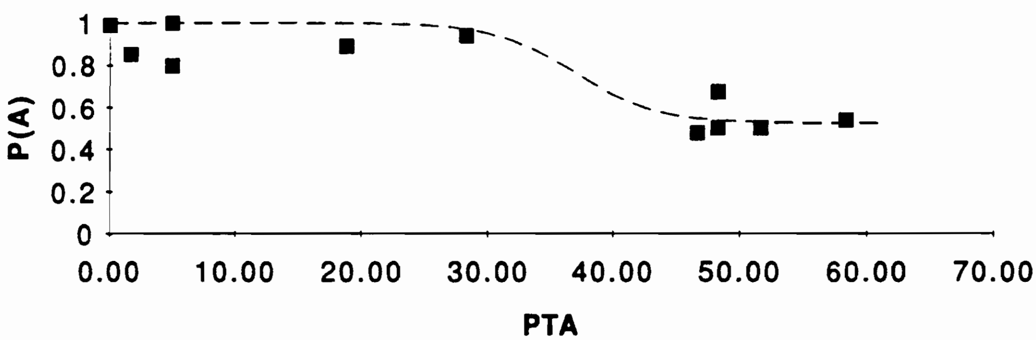

NL = 90 dBA, SL = 66 dBA
P(A)
1
0.8
0.6
0.4
0.2
0
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA


NL = 95 dBA, SL = 95 dBA
P(A)
1.00
0.80
0.60
0.40
0.20
0.00
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
PTA

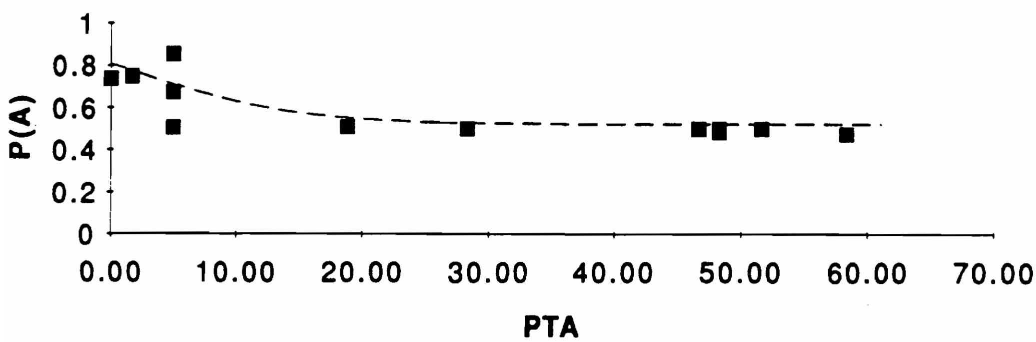
NL = 95 dBA, SL = 87 dBA



NL = 95 dBA, SL = 79 dBA



NL = 95 dBA, SL = 71 dBA



VITA

of

Gary S. Robinson

Mr. Robinson (born on April 1, 1957) received his B.S. in Mechanical Engineering from Mississippi State University in 1982. From 1982 to 1989, he worked as a mechanical engineer at the Naval Surface Warfare Center at Dahlgren, Virginia in the field of chemical warfare defense. Since 1989, he has been a full-time graduate student at Virginia Polytechnic Institute and State University. He received an M. S. degree in Industrial and Systems Engineering, with a concentration in human factors in December, 1991 and will receive his Ph.D. in the same field in December, 1993. As a graduate student, he has worked as a graduate research assistant in the Auditory Systems Laboratory and as a graduate teaching assistant for an Engineering Methods and Measurement laboratory.

Mr. Robinson is a member of the American Society of Mechanical Engineers, the Human Factors and Ergonomics Society, the Institute of Industrial Engineers, and the Institute of Noise Control Engineering, as well as the Alpha Pi Mu, Tau Beta Pi, and Phi Kappa Phi honor societies. His extracurricular interests include sailing, photography, and classic automobiles.