INTELLIGIBILITY OF SYNTHESIZED VOICE MESSAGES IN COMMERCIAL TRUCK CAB NOISE FOR NORMAL-HEARING AND HEARING-IMPAIRED LISTENERS

by

H. Boyd Morrison

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Industrial and Systems Engineering

APPROVED

Dr. John G. Casali, Chairman

Dr. Harry L. Snyder

Mrs. Beverly Williges

December 6, 1993

Blacksburg, Virginia

INTELLIGIBILITY OF SYNTHESIZED VOICE MESSAGES IN COMMERCIAL TRUCK CAB NOISE FOR NORMAL-HEARING AND HEARING-IMPAIRED LISTENERS

by

H. Boyd Morrison

Chairman: Dr. John G. Casali

Industrial and Systems Engineering

(ABSTRACT)

A human factors experiment was conducted to assess the intelligibility of synthesized speech under a variety of noise conditions for both hearing-impaired and normal-hearing subjects. Modified Rhyme Test stimuli were used to determine intelligibility in four speech-to-noise (S/N) ratios (0, 5, 10, and 15 dB), and three noise types, consisting of flat-by-octaves (pink) noise, interior noise of a currently produced heavy truck, and truck cab noise with added background speech. A quiet condition was also investigated. During recording of the truck noise for the experiment, in-cab noise measurements were obtained. According to OSHA standards, these data indicated that drivers of the sampled trucks have a minimal risk for noise-induced hearing loss due to in-cab noise exposure when driving at freeway speeds because noise levels were below 80 dBA. In the intelligibility experiment, subjects with hearing loss had significantly lower intelligibility than normal-hearing subjects, both in quiet and in noise, but no interaction with noise type or S/N ratio was found. Intelligibility was significantly lower for the noise with background speech than the other noises, but the truck noise produced intelligibility equal to the pink noise. An analytical prediction of intelligibility using Articulation Index calculations exhibited a high positive correlation with the empirically obtained intelligibility data for both groups of subjects.

ACKNOWLEDGMENTS

I would like to thank my graduate committee chairman, Dr. John Casali, for providing me with the opportunity to conduct interesting and, I hope, useful research, to expand my abilities both in the laboratory and in the classroom, and to learn. I also extend thanks to my other committee members: Dr. Harry Snyder, for his statistical expertise and overall guidance; and Mrs. Beverly Williges, for making available her wealth of knowledge concerning synthesized speech and for supplying the speech synthesis equipment.

Others provided much needed assistance during this research. I greatly appreciate the assistance of Bob Meyer and Jeff Dunn of the Volvo/GM Heavy Truck Corporation for providing trucks and drivers for noise measurements and recordings and Jeff Snyder for construction of equipment used in the truck noise measurement. I would also like to thank Gary Robinson, Bill Wright, Dan Mauney, and Jim Sayer for their help in preparation of the laboratory equipment and subsequent analysis of the data.

Finally, and most of all, I would like to thank my wife, Randi, for providing love, support, understanding, and encouragement during this, our first, year of marriage. Without her, this endeavor would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Background

In the 1990 Statement of National Transportation Policy, Transportation Secretary Samuel K. Skinner said that "... no industry in the Nation is more important to U.S. economic growth and international competitiveness than transportation" (Reed, 1990, p. 27). Unfortunately, this growth is hampered by two costly problems facing the road vehicle transportation industry: safety and congestion.

The greatest probability of fatality for persons under the age of 40 is from road traffic accidents (Hancock, Caird, Johnson, Shekhar, Yang , Coyle, and Pawlacyk, 1991). Worldwide, Rumar (1982) estimated that 250 000 deaths and 10 million injuries occur every year in traffic accidents. Even with the advent of safety devices such as passive restraint systems and anti-lock brakes, the ever-increasing number of vehicles could lead to 100 000 traffic fatalities per year in the U.S. alone by the year 2020 (Reed, 1990).

The growing number of vehicles also adds to the congestion facing urban commuters. Approximately 2 billion vehicle-hours of delay per year can be attributed to road congestion, with some estimates as high as 11 billion vehicle-hours per year by the year 2005 (Morans, Kamal, Okamoto, Hase, and Hellaker, 1991). Historically, government spending on road construction has not been sufficient to alleviate this problem. According to Federal Highway Administration data, total vehicle-miles of travel increased 78% between 1970 and 1988, while inflation-adjusted highway expenditures rose only 13.6% (Willis, 1990). It does not seem as if this situation will change in the near future. The Department of Transportation (Committee on Public Works and Transportation, U.S. House of Representatives, 1989) noted that the U.S. has neither the resources nor the funds to reduce traffic congestion solely through the construction of new roads.

The combined cost of these two problems is staggering. Voorst (1992) reported that traffic congestion and accidents could cost the U.S. up to $120 billion per year. Reed (1989, 1990) estimated that economic losses from accidents alone exceed $100 billion per year, and excess or wasted travel costs at least $46 billion per year.

To address these growing traffic safety and congestion concerns, plans for Intelligent Vehicle Highway Systems (IVHS) have been initiated by the U.S. Department of Transportation as flexible alternatives to major highway construction (Mast and Peters, 1992). IVHS offers the ability to help the driver detect possible hazards and roadway obstructions before he or she would otherwise be aware of them. This information might include the location of hazards such as icy bridges, upcoming accident scenes, dangerous curves, and sight-restricted intersections or bridge abutments. These systems might also include information about how to navigate around congested areas to reduce travel time. Accessing roadside sensors or Global Positioning System satellites, onboard maps can display vehicle location and desired destination in unfamiliar areas.

One of the most important applications of IVHS will be the warning of impending collisions and dangers to the driver and vehicle. Metzler (1988) estimated that by initiating a maneuver one-half second earlier, drivers could avoid approximately 60% of rear collisions, 50% of road junction accidents, and 30% of oncoming vehicle accidents. If the maneuver were initiated one second earlier, drivers could avoid over 90% of rear collisions and road junction accidents and over 60% of oncoming vehicle accidents. The immediate comprehension of a warning signal in these cases would be crucial to the driver's safety.

There are various methods for displaying the IVHS-generated information available to the driver. Roadside electronic signboards are used in some cities to display traffic information to drivers heading into congested areas. Moving map displays located on the vehicle dashboard give real-time navigational information; however, there is the danger that the map display would distract the driver's attention from the road. Because driving is largely a high-demand visual tracking task, many displays may need to be auditory so that this danger could be avoided.

Some efforts to implement these systems have already begun. A project called Pathfinder, developed jointly by the Federal Highway Administration, the California Department of Transportation, and General Motors, was designed to direct motorists driving specially equipped cars around accidents and areas of congestion (Beck, Meyer, Lewis, and Hager, 1988). TravTek, a demonstration project in Orlando, allows drivers using modified rental cars to access navigational information via visual display terminals and voice message systems (Fleischman, Carpenter, Dingus, Szczublewski, Krage, and Means, 1991). Erlichman (1992) described a pilot study involving the In-Vehicle Safety Advisory and Warning System (IVSAWS), a program proposed by the Federal Highway Administration which would use in-vehicle displays and portable radio beacons placed near road hazards to alert drivers of these situations.

Commercial vehicle operators, especially those driving heavy trucks, have unique concerns which can be addressed by IVHS, and it is to this application that this thesis is addressed. Because medium and heavy trucks have a higher fatality rate than the total vehicle population (3.5 fatalities per 100 million vehicle-miles for trucks vs. 2.1 per 100 million for all vehicles in 1988) (IVHS America, 1992), safety enhancements will be particularly important in IVHS design for commercial vehicles. Warning systems may give information about potentially dangerous truck conditions, such as load shift, possible rollover, low tire pressure, and brake condition, in addition to collision avoidance warnings and notification of hazardous roadway conditions (Mast, 1991). Because of their size and mass, trucks need much longer stopping and maneuvering distances than do passenger cars, and therefore advance warning of hazards is critical.

To be of use, navigation and route information must take into account the clearance and weight limitations and hazardous cargo restrictions specific to trucks. These navigation data could then help guide operators through the streets of unfamiliar cities and around unexpected congestion. Such devices could save the drivers time and money, which in turn would make the devices more acceptable to truck operators (Willis, 1990). Willis (1990) also reported that ambulance and cab companies have already begun using these systems to aid in vehicle location and guidance.

Another serious problem in long-haul trucking, as well as automobile driving, is driver impairment due to fatigue. Kearney (1966) reported that up to 35 percent of all highway fatalities could be the direct result of driver fatigue or drowsiness. Tilley, Erwin, and Gianturco (1973) found that almost two-thirds of all drivers had experienced drowsiness while driving automobiles and that almost one-third of those reporting drowsiness were not aware that drowsiness was impending. IVHS may be able to detect the onset of driver impairment and issue warnings to the operator and others in the immediate area (Mast, 1991).

The design of displays that are used in IVHS will be of paramount importance to their usefulness and acceptance by drivers. Auditory displays will be even more critical to the performance of these systems because they have unique advantages in situations where immediate detection and comprehension are needed (i.e., warnings and real-time navigation). Because drivers are not highly trained to listen for coded nonverbal warning sounds, voice message technology offers considerable potential benefit as a primary mode of auditory display.

Auditory Displays

Guidelines defined by Deatherage (1972) were established for deciding between the use of auditory and visual displays. Situations in which auditory displays should be used include (1) when warnings are given, because auditory warnings are omnidirectional, (2) when the visual system of a person is already overburdened, (3) when the message is short and simple, (4) when the message calls for immediate action, and (5) when the message indicates an event in time, not a location in space. It has already been noted that an important application of IVHS will be the use of warning signals. Most messages will be of short duration and will need to be acted upon immediately by the driver. Because a driver's visual attention is focused on the road, auditory displays are probably the least distracting mode of stimulus (Parkes and Coleman, 1990; Wickens, 1980), although if improperly designed, they may elicit a startle reaction which can be disturbing and time-consuming.

Deatherage (1972) also listed guidelines stating when speech should be used in place of nonspeech messages. These situations include (1) when flexibility is needed, (2) when listeners are without special training, (3) when the message deals with events in future time that require some preparation, and (4) when situations of stress might cause the listener to "forget" the meaning of the signal code. These guidelines address the need for flexibility required in different driving situations, including ones that can be highly stressful (e.g., hazardous road conditions, traffic, and inclement weather), as well as the fact that truck drivers are not specially trained to listen for warning signals, as are aircraft pilots. The need for the driver to prepare for route changes and the system to display the related directional information represents another situation in which the use of speech messages in truck cabs could be recommended.

Research Issues

The use of auditory displays that incorporate speech in the truck cab environment has many potential applications. However, various factors must be taken into account when designing these interfaces. Because of the high noise levels found in some truck cabs

(Reif, Moore, and Steevensz, 1980), exposure levels may be great enough to damage the hearing of operators over long periods of time. This is supported by a study by Nerbonne and Accardi (1975), who found a slow but steady increase in hearing loss with years of experience driving trucks. The possibility of having to accommodate an operator population which includes a significant number of hearing-impaired individuals cannot be ignored when installing auditory displays that could be important to driver safety. The speech levels necessary to overcome masking effects of the cab noise may contribute substantially to the exposure risk, especially for the already impaired operator who requires higher levels.

The high noise levels also present a concern with respect to the intelligibility of speech messages. Numerous studies (Abel, Alberti, Haythornthwaite, and Riko, 1982; Hawkins and Stevens, 1950; Pollack and Pickett, 1958) have shown that the speech level must be increased in noise to produce the desired intelligibility and have quantified the signal-to-noise (S/N) ratios needed for intelligibility. To achieve a 90% intelligibility score, normal-hearing individuals require an S/N ratio of at least 10 dB (Acton, 1970). As the S/N ratio is decreased, the intelligibility can be expected to fall as well, with approximately 30% intelligibility at an S/N ratio of -5 dB. For hearing-impaired individuals, the decrease in intelligibility has been found to be more pronounced as the S/N ratio decreases (Pekkarinen, Salmivalli, and Suonpää, 1990). However, when using synthesized speech, which is the speech technology most likely to be used with IVHS displays, the masking effects of high noise levels are not as well-defined.

The intent of the present study was to determine the effects that truck cab noise and speech levels have on the segmental (i.e., phonemic) intelligibility of synthesized speech and how these factors interact with the hearing threshold level of the listener.

HUMAN AUDITION AND NOISE

For the purposes of this document, sound is defined as a pressure oscillation of a medium having viscosity and elasticity, which is capable of being detected by a human as an auditory sensation. Noise is defined as unwanted sound (ANSI, 1960). Noise can also be defined as a sound which is random and unpredictable by nature, such as the hiss of a radio tuned between stations (Boff and Lincoln, 1988).

Human Audition

The normal, healthy ear receives sound pressure oscillations and transforms them into sensory information that the brain resolves through nerve impulses. Figure 1 shows that the ear is divided into three sections: the outer, middle, and inner ears. The outer ear consists of the pinna and the auditory canal, the open channel leading to the tympanic membrane, or eardrum. Because of its shape, the pinna provides information which aids in the recognition and localization of sounds. Midrange frequencies are greatly amplified due to the shape and dimension of the auditory canal (and to some extent the pinna), giving a net gain of 10 to 15 dB in the 2 to 4-kHz region (Ward, 1986).

The middle ear, consisting of the tympanic membrane, ossicles (comprised of three small bones called the malleus, incus, and stapes), and the oval window, converts the air pressure oscillations into mechanical vibrations. Because it provides a seal between the auditory canal and the middle ear, the tympanic membrane protects the middle and inner ear by keeping out foreign bodies. Two muscles attached to the malleus and stapes reduce the amount of high intensity noise (over 80 dB) reaching the inner ear by placing the ossicles in tension. Some protection is provided, particularly by the stapedius reflex; however, the muscular activity shows partial adaptation to the noise over time, and there

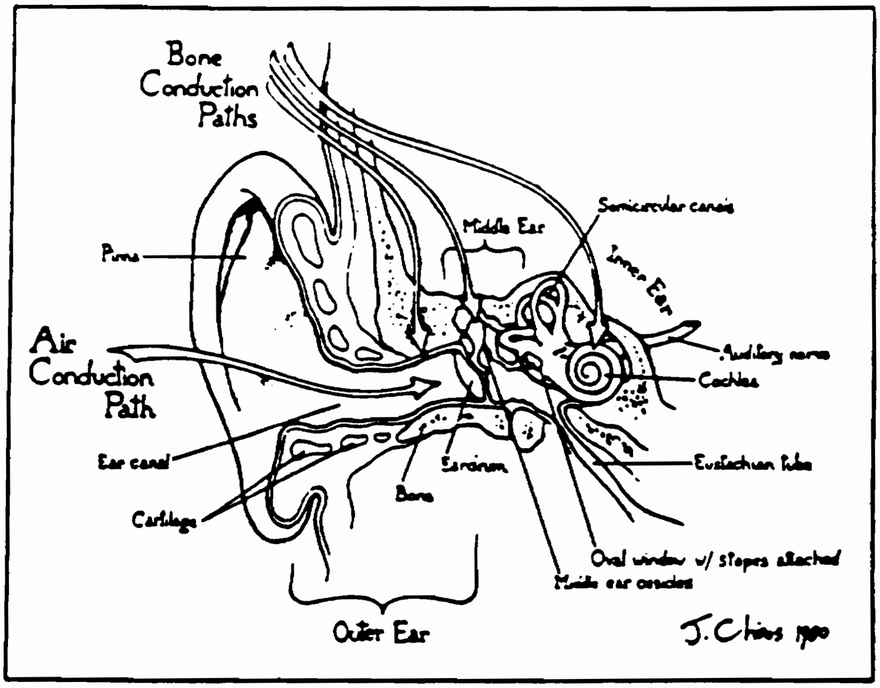

Figure 1. *Basic anatomy of the ear and sound paths (adapted from Berger, 1980).*

is a finite lag time in reflex response that limits protection against rapid onset noises (Ward, 1986).

The inner ear converts the mechanical vibration at the oval window into fluid motion in the cochlea. This fluid motion stimulates hair cells located on the basilar membrane, the deflection of which is then transformed into neural impulses which are processed by the brain. The hair cells are susceptible to damage from noise exposure, this damage becoming irreversible as exposure increases (Ward, 1986).

Hearing Loss

Conductive hearing loss. Conductive hearing loss is damage to the outer or middle ear which interferes with the conduction of sound to the inner ear. This condition is usually caused by a ruptured or scarred eardrum, disease (e.g., otitis media and otosclerosis), a blow to the head, or a sudden explosion near the head which results in immediate "acoustic trauma." This type of hearing loss is often reversible through medical or surgical means. While employment-related conductive hearing loss may occur, it is uncommon (Ward, 1986).

Conductive hearing loss is characterized by a fairly uniform amount of hearing loss at most frequencies (Morrill, 1986). An audiogram characteristic of persons with conductive hearing loss is shown in Figure 2 as a dashed line.

Sensorineural hearing loss. Damage to the inner ear is defined as sensory hearing loss, while damage to the auditory nerve is defined as neural hearing loss. Both of these types of hearing loss can be permanent and are irreversible by known medical means. Noise primarily produces sensory hearing loss (Ward, 1986). Although the structures damaged in sensory and neural hearing loss are different, they are normally grouped together because the hearing loss symptoms are similar.

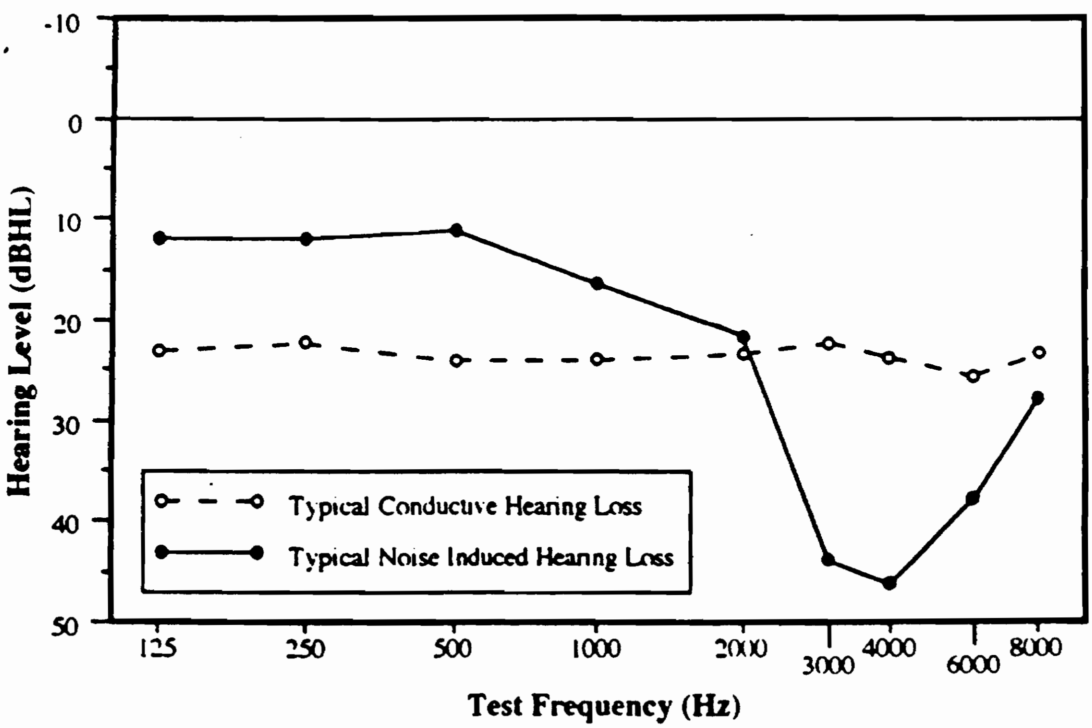

Figure 2. *Typical audiogram characteristics of conductive and noise-induced hearing loss (adapted in part from Melnick, 1979).*

Two types of noise can produce sensory hearing loss. Acoustic trauma, produced by a noise of very high intensity but short duration, results in instantaneous neural damage by rupturing the structures of the inner ear. As with conductive loss, an explosion or gunshot near the ear can cause acoustic trauma. The more common etiology of sensory hearing loss is long duration exposure to moderate or loud noise. After prolonged exposure to loud noise, the hair cells become damaged and insensitive to the fluid motion of the inner ear. As the severity of the exposure increases and the hair cells are not allowed time to recover, this damage becomes irreversible.

The Effects of Noise on Humans

Noise exposure. Sensory hearing loss is the most common type of hearing loss resulting from occupational noise exposure (NIH, 1990). This permanent hearing damage due to overexposure to noise is termed noise-induced hearing loss (NIHL). This problem has become so prevalent that Miller (1978) estimated that more people experience NIHL than all other occupational illnesses combined.

The Walsh-Healey Public Contracts Act (U.S. Department of Labor, 1969) was passed and stated that no employer contracting to the U.S. government could expose a worker to a daily 8-hour time weighted average (TWA) noise exposure of over 90 dBA. In 1970, the Occupational Safety and Health Act was passed to provide this protection to all of general industry. OSHA (1983) amended the original Occupational Health and Safety Act to add an "action level" that would be imposed at 85 dBA TWA. Noise exposure is a multiplicative function of sound level and duration. Integrated over an 8-hour day, any exposure which exceeds 90 dBA TWA (a 100% dose) is considered to be over the OSHA limit and hazardous, and engineering or administrative controls must be established (OSHA, 1983). For those exhibiting a pre-defined standard threshold shift, 85 dBA is the exposure limit.

Because the symptoms of NIHL are insidious and normally painless, many people are not aware of their condition until it has approached the advanced stages (Casali, 1986). The EPA (1981) estimated that over 9 million workers were exposed to levels exceeding 85 dBA daily. This number includes construction, mine, transportation, and agricultural workers that are not covered by the federal noise standard (Suter, 1986). If the current trend goes unchanged, Shipley (1985) contended that NIHL could become the most expensive legal and medical problem facing industry.

Characteristics of noise-induced hearing loss. Noise-induced hearing loss can first be seen in an audiogram as a "notch" in the hearing threshold level at 4000 Hz, which is not surprising considering the aforementioned amplification that occurs in the auditory canal (Ward, 1986). This notch becomes more pronounced and spreads to include frequencies in the 1000-Hz to 8000-Hz range as the condition worsens (Melnick, 1979). An audiogram characteristic of persons with noise-induced hearing loss is shown in Figure 2 as a solid line.

Degree of hearing loss. Specifying the degree of hearing loss is important when describing listener characteristics. The amount of hearing loss may be categorized by the degree of difficulty of understanding speech, or as a percentage pure-tone hearing loss in one or both ears (Kryter, 1985; Miller and Wilber, 1991). Most classification schemes use the pure-tone average (PTA) hearing level found at discrete frequencies which are referenced to audiometric zero. Decibel hearing level (dBHL) is the unit used to describe the hearing threshold level (i.e., dB weighted according to a normal audibility curve) at which a pure tone can be detected. Table 1 illustrates one of these classification schemes, which categorizes hearing loss using the PTA at 500, 1000, and 2000 Hz.

Other studies conclude that, if frequencies above 2000 Hz are ignored when classifying hearing loss, prediction of loss of speech perception is hindered. Smoorenburg, de Laat, and Plomp (1982) studied the speech reception threshold (SRT)

TABLE 1

One 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	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 normal 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

(the level at which 50% of spoken sentences could be reproduced correctly) in quiet and in noise with listeners of various levels of hearing loss. Their findings indicate that, while the threshold shift at 500 Hz correlates best with the SRT in quiet, the SRT in noise correlates best with shifts in the 2000–3000 Hz range. Kryter (1985) advocated the use of frequencies of 1000, 2000, and 3000 Hz and perhaps 4000 Hz to account for the frequencies normally found in everyday speech in quiet.

The American Academy of Otolaryngology-Head and Neck Surgery (AAO-HNS) endorses a method which calculates percent impairment for one or both ears (Miller and Wilber, 1991). "Impairment" denotes a medical condition that causes a deviation from one's normal or non-impaired characteristics (Cudworth, 1986). The most critical impairment would be that which decreases the ability to hear ordinary speech. A non-impaired individual would have a hearing threshold level of 25 dBHL or lower in each ear, while an individual with total impairment would have a hearing threshold level of 92 dBHL or higher in each ear. The percent impairment calculations therefore interpolate the difference between these two levels. The *monaural* percent hearing impairment is calculated by subtracting 25 dB from the PTA 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 impairment for the poorer ear, and dividing the total by 6. Each of these calculations would assign a 100% impairment to an individual whose PTA for each ear was at or above 92 dB and a 0% impairment to an individual whose PTA for each ear was at or below 25 dB.

Because there is no agreed upon definition of hearing loss or level of hearing loss, many researchers and government agencies use their own range of values. For example, Abel et al. (1982) defined bilateral noise-induced high frequency loss as a slope in loss of 35 to 65 dB between 500 and 4000 Hz. In addition, states vary in the method and frequencies to use when defining hearing loss for workers' compensation claims (Cudworth, 1986).

Truck Cab Noise

Historically, measurements of truck cab noise show that drivers often experience hazardous exposure levels. Truck-related sources of noise in truck cabs include the cooling fan, engine, transmission, exhaust, aerodynamic or wind noise, road or tire noise, and CB and radio chatter. Clarke (1978) reported that truck drivers were exposed to sound pressure levels at or above 90 dBA more than 40% of the driving time. During controlled SAE J336 (Society of Automotive Engineers, 1968) acceleration tests, Leneman (1977) found that while the exterior noise of a conventional short cab truck was 90 dBA, the interior noise was 94 dBA. Higher interior noise levels are consistent with the fact that most research into controlling truck noise has been focused on reducing exterior environmental noise, rather than interior cab noise. Using the same SAE J336 tests, Werner and Boyce (1976) found interior noise levels as high as 95 dBA, 5 dBA above the OSHA permissible limit if sustained over 8 hours. Although both truck cabs used in this study were cab-over-engine (COE) models, the findings should be similar for cab-behind-engine (CBE) models because no noise disparities are apparent between COE and CBE models (Fee and Schwendeman, 1976).

Reif and Moore (1983) measured noise levels in 58 relatively new and well-maintained 1970s vintage truck cabs and found that equivalent levels (L_{eq}) exceeded 90 dBA TWA in 40% of the trucks for a 10-hour time period (the average drivers spend behind the wheel in a day) and exceeded 85 dBA TWA in 90% of the trucks for a 10-hour time period. (L_{eq} is defined as the continuous sound level, usually A-weighted, that is equivalent in energy to a variable sound level when integrated over the same period of time (Earshen, 1986).) They measured the sound levels for three different types of driving: city, highway, and freeway. The highest noise exposure was found to exist during freeway hauls, while the lowest exposure occurred during city driving. Measurements were made from three locations: the center of the truck cab, the driver's right ear, and the driver's left ear. The highest noise levels were experienced at the left ear and decreased moving towards the right ear and center of the cab. All measurements were made with the radio off and conversation held to a minimum. Reif and Moore (1983) estimated that the use of the radio and CB radio would result in a 25% increase in noise exposure and included that figure in their evaluation of noise exposure.

In an earlier related study (Reif, Moore, and Steevensz, 1980), narrow band frequency spectra were obtained, and results showed that sound energy of truck cab noise was concentrated in the frequency ranges below 2000 Hz. A typical spectrum analysis of the noise at the right ear is shown in Figure 3. The effects of speech and CB radio on cab noise levels were also measured. Results showed that speech can increase the instantaneous sound level to 95 dBA (Figure 4), while CB radio output can increase the instantaneous sound level to 100 dBA (Figure 5). The highest increases due to the CB radio were found at the center of cab position during freeway driving. No information is given about the location of the CB radio or how the sound level of the CB was controlled.

A review of the literature reveals no studies conducted on the interior noise levels of trucks produced after 1980. It is not known whether improvements in technology have decreased the noise levels present in truck cabs and the corresponding noise exposure experienced by truck operators.

Because the noise levels reported above are well over the limits set by OSHA, they could pose a substantial exposure risk to drivers. Nerbonne and Accardi (1975) showed a slow but steady increase in hearing level (HL) with years of driving and that the mean HLs of truck drivers were significantly worse by 10 to 20 dB at 250 to 4000 Hz than non-noise-exposed men of similar age. This presents the possibility that a significant

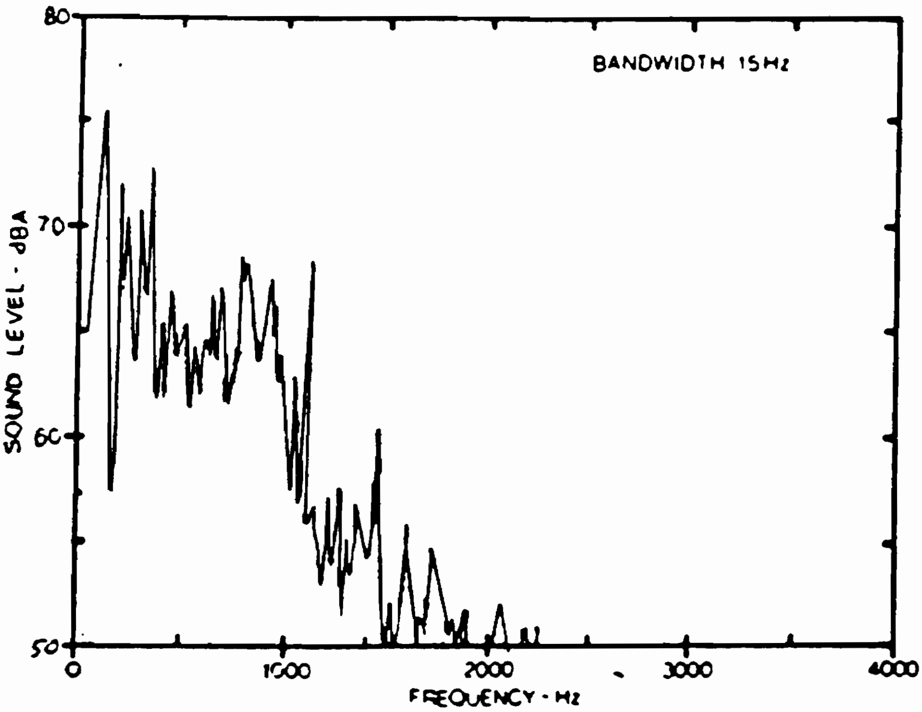

Figure 3. *Frequency analysis of noise during highway driving monitored at the driver's right ear (from Reif, Moore, and Steevensz, 1980).*

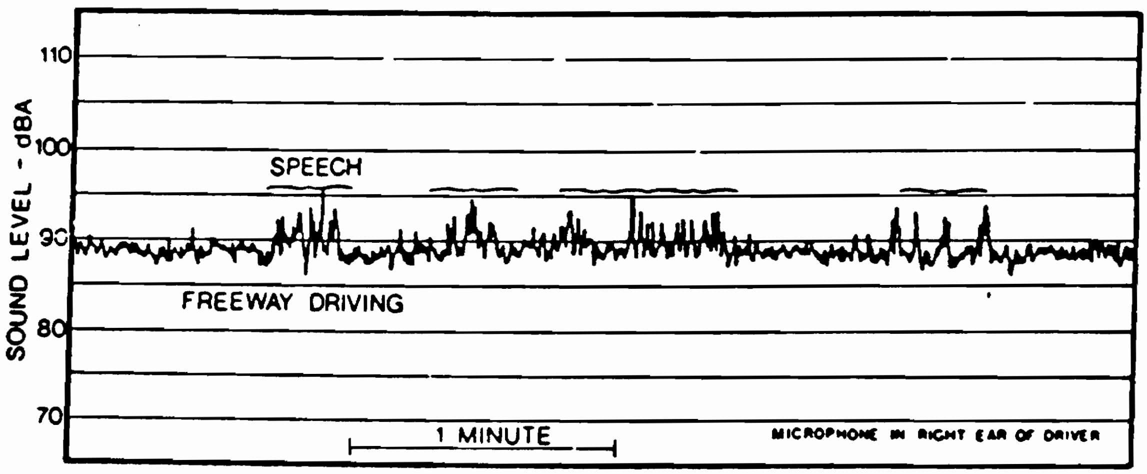

Figure 4. *The effect of speech on in-cab noise levels during freeway driving (from Reif, Moore, and Steevensz, 1980).*

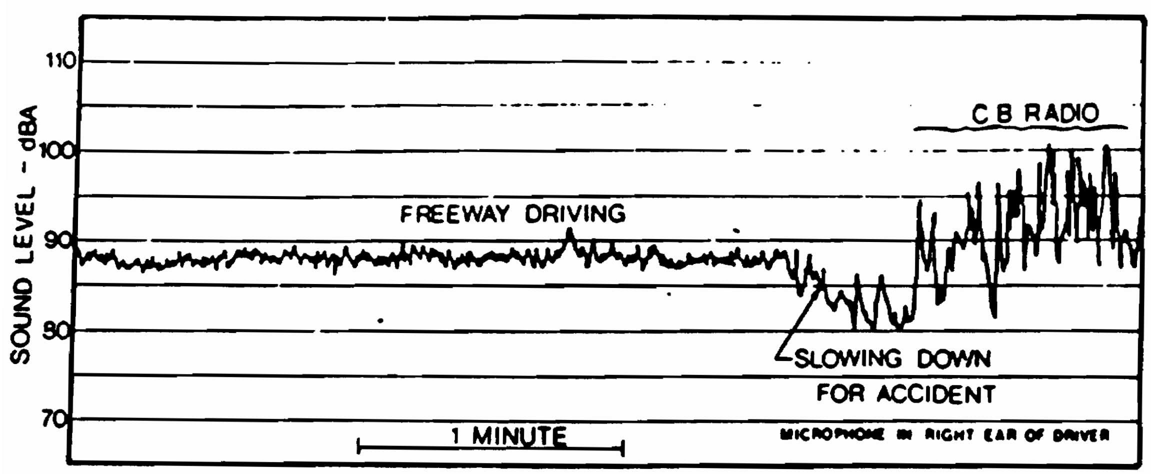

Figure 5. *The effect of CB radio on in-cab noise levels during freeway driving (from Reif, Moore, and Steevensz, 1980).*

proportion of the truck driver population has some hearing loss. Compared to the OSHA limit of 90 dBA per 8 hour day average, the 95 dBA truck noise would only be legally allowed for a 4-hour duration per day. Of course, trucks are driven at sustained highway speeds for much longer periods during a single day. In addition, the noise levels found inside truck cabs could mask speech and auditory signals if the signal-to-noise ratio is not high enough, and high speech levels from the CB radio and in-cab speech could interfere with speech warnings.

SPEECH INTELLIGIBILITY

Background

Spoken language can be divided into a number of different units. *Phonemes* are the basic building blocks of speech and are the smallest units which, if changed, will alter the meaning of the word. Phonemes generally correspond to letters or combinations of letters of the alphabet and comprise the consonant and vowel sounds. Syllables are formed by two or more phonemes and are the basic unit, or segment, of speech perception. The syllabic unit is usually described as a combination of consonant and vowel (a CV unit). *Words* are then formed by combinations of syllables.

Speech intelligibility is the ability to correctly recognize one of these units of speech. Test materials include nonsense syllables, single words, or entire sentences. To evaluate speech intelligibility, it is important to note which of the parts of speech contribute most to its comprehension.

The parts of the speech waveform that are usually associated with consonants have substantially less power than those parts associated with vowels. Figure 6 shows that, when using amplitude distortion to clip the peaks of the speech waveform, intelligibility is not substantially affected; clipping the center drastically reduces intelligibility. Therefore, consonants contribute much more to the intelligibility of speech than do vowels.

Using frequency distortion, which removes selected frequencies from a sound through filtering, French and Steinberg (1947) found that vowel sounds carry most of their power at lower frequencies than consonants. In the same study, results showed that intelligibility of speech was most degraded by high-pass filtering with a cutoff frequency above 1000 Hz and low-pass filtering with a cutoff frequency below 4000 Hz (see Figure 7).

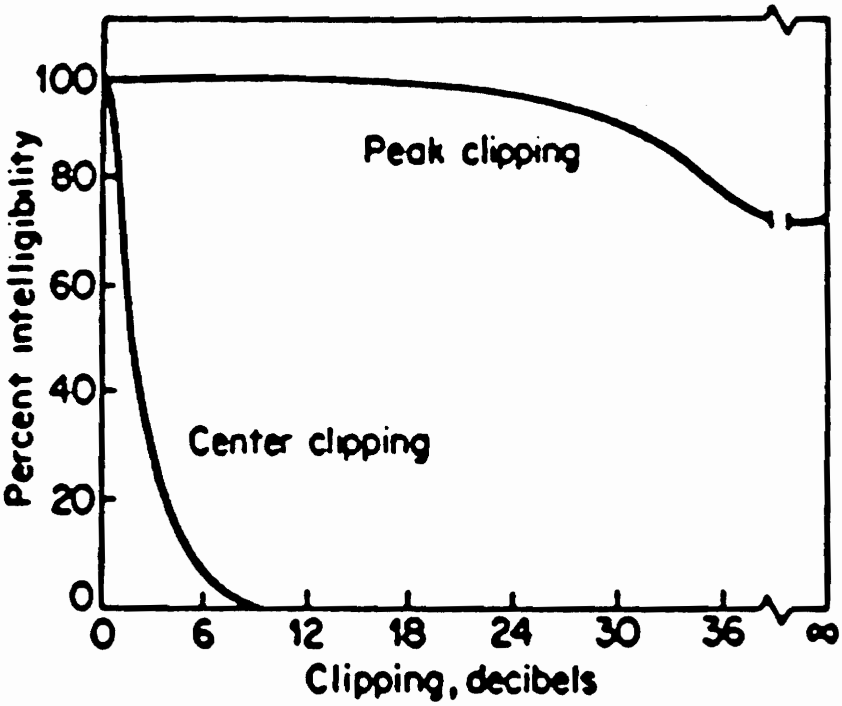

Figure 6. *The effects on speech intelligibility of various amounts of center clipping and peak clipping (from Licklider and Miller, 1951).*

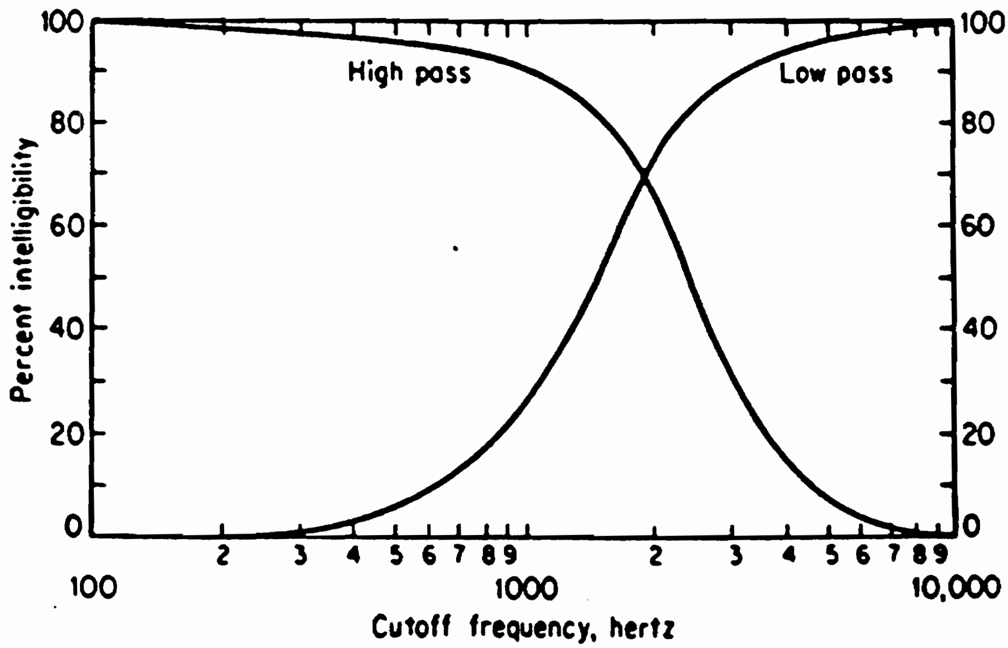

Figure 7. *The effect of low pass and high pass filtering on the intelligibility of speech (adapted from French and Steinberg, 1947).*

Types of Intelligibility Tests

A number of tests have been developed to evaluate the segmental intelligibility of speech. Segmental intelligibility refers to the ability to discriminate individual phonemic elements from each other. Sentences provide contextual information about the speech which aids in intelligibility; monosyllabic words must be recognized solely by the acoustic information provided and therefore offer a more objective measure of intelligibility. Because the consonants are so important to speech intelligibility, most tests emphasize the ability to distinguish them. ANSI S3.2-1989, a consensus standard for measuring the intelligibility of speech when evaluating communications systems, recommends three different methods for assessing speech intelligibility: the phonetically balanced (PB) monosyllabic word test, the Modified Rhyme Test (MRT), and the Diagnostic Rhyme Test (DRT). All three tests give results in percent intelligibility, which is the ratio of correct word identifications to total number of words presented. These tests each have advantages and disadvantages and are presented to listeners in different ways.

Phonetically balanced monosyllabic word test. Phonetically balanced word lists are sets of words in which the frequency of occurrence of various phonemes is proportional to their frequency in everyday English. Egan (1948) developed the "articulation test" of 20 PB lists, containing 50 monosyllabic English words each. The stimulus words are presented to the listener via a carrier sentence ("Would you write ________ now"). The listener responds in an open format by writing down the response.

Listeners require at least 10 hours of training to attain and maintain stable performance if PB words are used (ANSI, 1989). It then takes approximately 200 seconds to administer each PB list in an experiment. The open-response format allows an unlimited number of spelling errors and could bias the results. Despite these disadvantages, PB lists are useful because they are very sensitive to changes in signal-to-noise ratio. Changes in the signal-to-noise ratio in the range of -5 dB to +5 dB can produce large changes in intelligibility (ANSI, 1989).

Modified Rhyme Test. Developed by House, Williams, Hecker, and Kryter (1965), the MRT consists of 50 six-word lists of monosyllabic English words. Most of the words follow the consonant-vowel-consonant (CVC) sequence. The listener can respond in either a closed-response, multiple-choice format by circling one of the six choices, or by an open-response format in which the listener writes the word that he or she thought was presented. The closed-response format reduces the uncertainty of the listener and eliminates any problems due to spelling or illegible handwriting. A carrier phrase may or may not be used. Only the initial or final consonant differs in each set because consonants carry most of the information about the word.

When using carrier sentences, the MRT requires between 120 and 180 seconds to administer each 50 word set; when carrier sentences are not used, approximately 75 seconds are required (ANSI, 1989). In addition, the MRT requires much less training than PB lists do. The MRT also provides information about what types of consonant confusions are made.

Diagnostic Rhyme Test. The DRT (Voiers, 1977) is similar to the MRT; however, the DRT distinguishes differences between only the initial consonants and uses no carrier sentence. The listener has only two choices for each of the 96 pairs of words presented. This makes the test quick to administer; approximately 260 seconds are required for a list of 192 words (ANSI, 1989). The main disadvantage lies in the smaller word list.

The Effects of Noise on Speech Intelligibility

Noise presented simultaneously with a speech signal can cause loss of intelligibility and usually consists of complex sounds such as white noise or a babble of voices. This interference with the speech signal is called *masking*. The signal-to-noise (S/N) ratio (in this case the signal is speech) is a significant factor affecting speech intelligibility and is defined as a signed difference (as opposed to a fractional ratio) between the signal and noise levels (Casali, 1989). A negative S/N ratio indicates that the speech has a lower intensity level than the noise; at the same time, a positive S/N ratio indicates that the noise has a lower intensity level than the speech. For example, for speech at 85 dB in noise at 90 dB, the S/N ratio is -5 dB.

Other variables such as the characteristics of the noise, the characteristics of the listener, and the type of voice used can also affect the intelligibility of speech in noise. The following is a summary of the effects these variables have on speech intelligibility.

Signal-to-noise ratio. Hawkins and Stevens (1950) measured the monaural thresholds for speech over a range of S/N ratios. The criterion for determining the threshold of intelligibility was defined as the "level at which the listener is just able to obtain without perceptible effort the meaning of almost every sentence and phrase" (Hawkins and Stevens, 1950, p. 11) of a passage read from *The Wealth of Nations*. White noise was presented simultaneously with the recorded passage. Results showed that the S/N ratio needed to achieve the speech threshold was constant for noise levels between 35 and 110 dB (see Figure 8). That is, for every 10 dB increase in the noise level, the speech also needs to be raised 10 dB to maintain the threshold. These findings were corroborated in a later study which had a more definite criterion for the speech reception threshold, but only used noise levels up to 70 dB (Smoorenburg, de Laat, and Plomp, 1980).

At noise levels beyond the ones tested in the above studies, Pollack and Pickett (1958), using PB word lists, found a deterioration in word intelligibility between 100 and 130 dB that could not be overcome by a constant rise in the S/N ratio (see Figure 9). At an S/N ratio of 0 dB and noise levels between 80 and 110 dB, the results of this study showed a small but significant decrease in intelligibility, which contradict the findings of

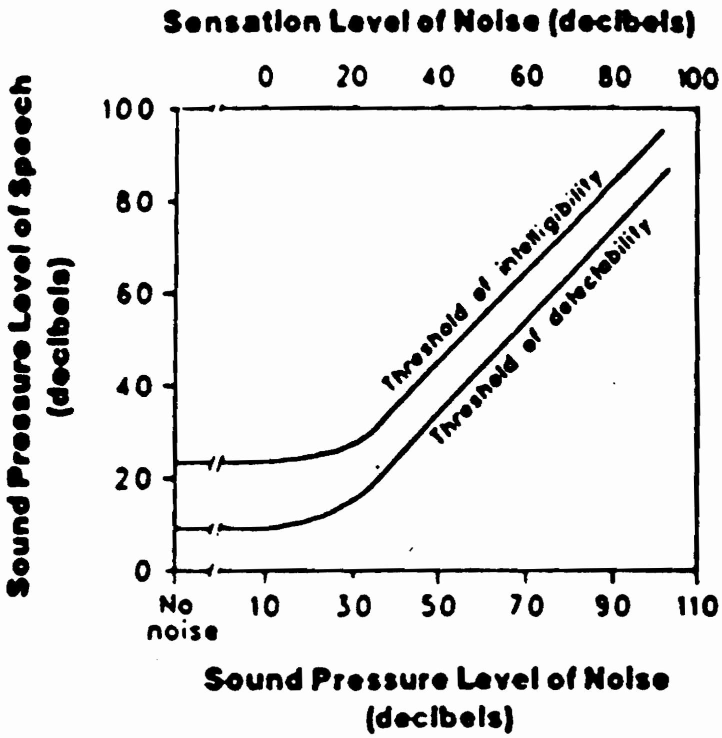

Figure 8. *Threshold level for intelligibility and detectability of speech as a function of noise level (adapted from Hawkins and Stevens, 1950).*

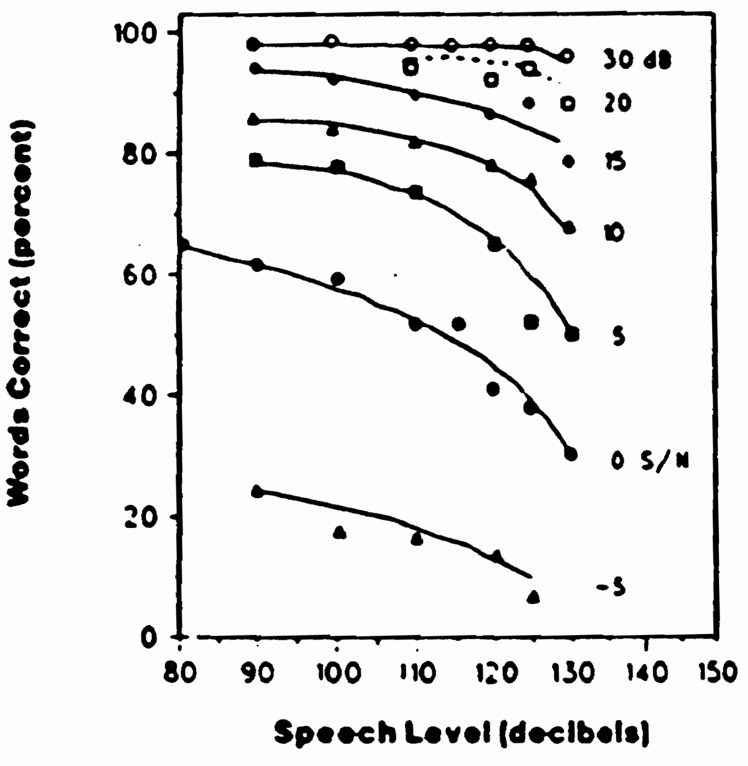

Figure 9. *Intelligibility of spoken phonetically balanced monosyllabic words presented in noise as a function of speech level and signal-to-noise ratio (from Pollack and Pickett, 1958).*

Hawkins and Stevens. However, this could be due to the redundancy of the connected discourse used by Hawkins and Stevens which is not present in PB word lists.

It is important to specify the method for measuring noise and speech levels. The long-term root mean square (rms) is the square root of the average of the squared pressure oscillations. This resolves the problem encountered when simply taking the arithmetic average of a sinusoidally varying function, which would give a value of zero. The peak-to-peak measurement gives the algebraic difference between the extreme values of pressure changes in the sound wave. The long-term rms value is used to compute slow, fast, and/or L_{eq} measurements of the noise levels. L_{eq}, being the constant dBA level that is equivalent to a time-varying level over a specific time window, is the conventional measure for the average sound level over a period of time. The Articulation Index (ANSI, 1969) uses the long-term rms plus 12 dB to represent the peak amplitudes of speech. Therefore, by measuring the rms levels of the speech, it is possible to estimate its L_{eq} and the resulting S/N ratio.

Noise exposure. Pollack (1958) showed that speech intelligibility decreases after short term exposure to high noise and speech levels. The cumulative effects of the exposure were most pronounced at noise and speech levels above 115 dB. Above 120 dB, the intelligibility scores were halved 13 minutes into testing. The function of these noise-induced temporary threshold shifts (NITTS) in decibels is approximately proportional to the logarithm of the exposure duration (Ward, 1973).

Noise frequency. The effect of introducing a masking signal depends greatly upon its frequency spectrum. When masking a pure tone with another pure tone, the masking effect is greatest when the masker frequency is near that of the signal frequency (Deatherage, 1972). However, as the masker sound level increases, the masking effect spreads to higher frequencies (see Figure 10). The masking effect is asymmetrical; that is, frequencies above the masker frequency are masked more than those below it. This

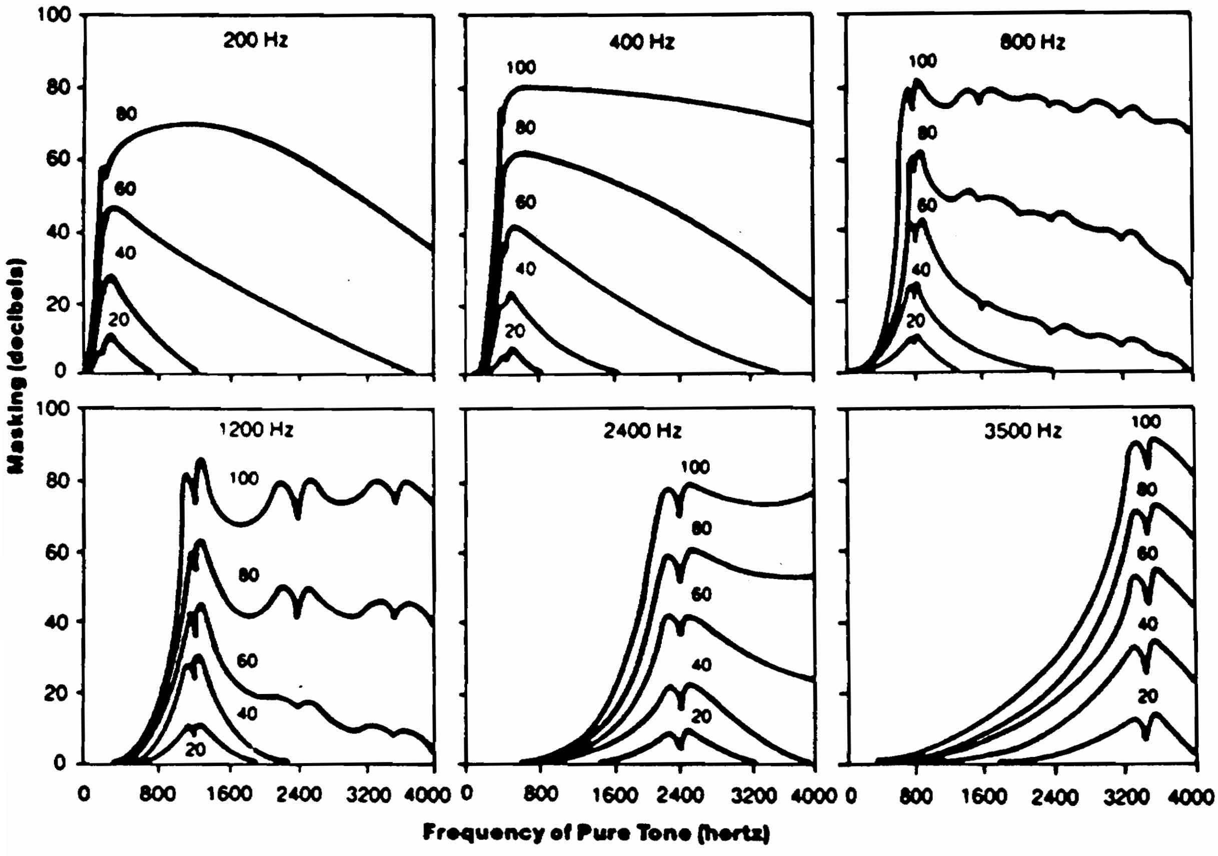

Figure 10. *Masking of pure tones by pure tones as a function of signal frequency (from Deatherage, 1972).*

upward masking effect is most pronounced with masking tones of low frequency. Narrow-band noise also produces an upward spread of masking. Figure 11 shows that as the level of a 90-Hz wide band of noise centered at 410 Hz increases, the upward spread of masking also increases.

Upward spread of masking also occurs when the signal consists of speech. Using monosyllabic words as the speech signal, Miller (1947) determined that wide-band noise is a more effective masker of speech presented at 95 dB than any low-frequency, narrow-band noise. Frequency bands below 1100 Hz (135–400 Hz, 350–700 Hz, and 600–1100 Hz) were more effective maskers than higher frequency maskers when the intensity of the speech was below that of the noise. At an S/N ratio of -18 dB, these low frequency bands rendered speech almost totally unintelligible. The lower frequency bands produced a performance decrement function similar in shape to the wide-band noise, but no narrow-band frequencies produced lower performance than the wide-band noise (see Figure 12).

Klein, Mills, and Adkins (1990) conducted a study to determine the effect of upward masking of pure tones on normal-hearing and hearing-impaired young and elderly listeners. Their results indicate that the upward-masked thresholds of hearing-impaired listeners were generally higher than those of listeners with normal hearing at a masker level of 90 dB. These findings are consistent with those of earlier studies (Hannley and Dorman, 1983; Jerger, Tillman, and Peterson, 1960). The data also showed that upward-masked thresholds were similar for both age groups.

Noise type. The type of noise background has been shown to have an effect on speech intelligibility. Miller (1947) reported that as the number of masking voices increases, up to four, the ability to discriminate the signal speech decreases. Above four, little further decrement was experienced (see Figure 13). He found no difference in masking effectiveness whether the masking voices spoke in English or in a language the listener did not understand, suggesting that the frequency spectrum of the masker, not a

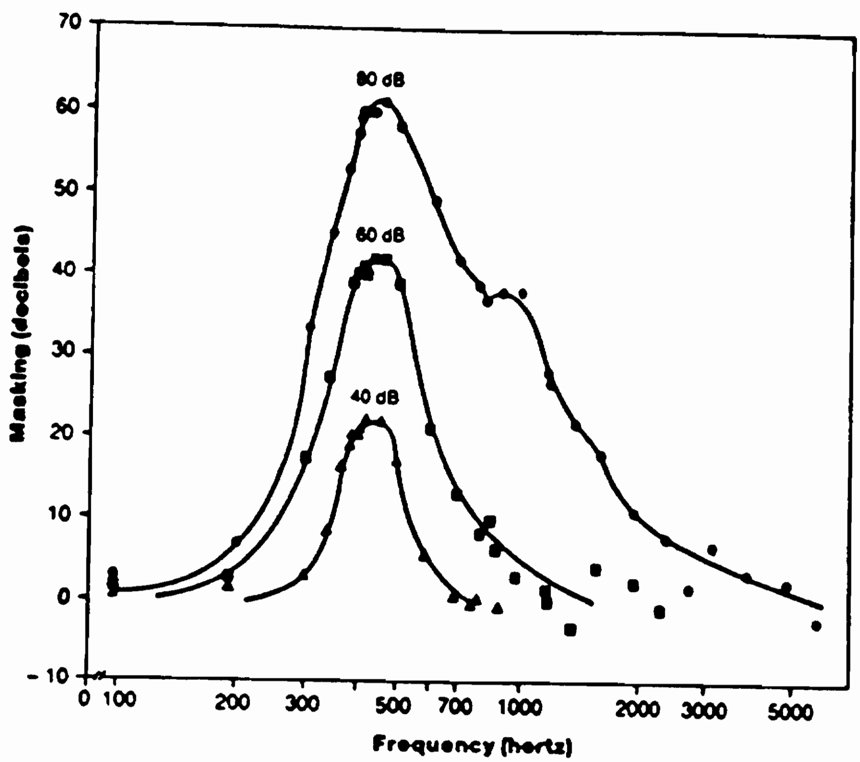

Figure 11. *Masking of pure tones by narrow band noise (90 Hz wide, centered at 410 Hz), as a function of signal frequency and overall sound pressure level of the noise (from Egan and Hake, 1950).*

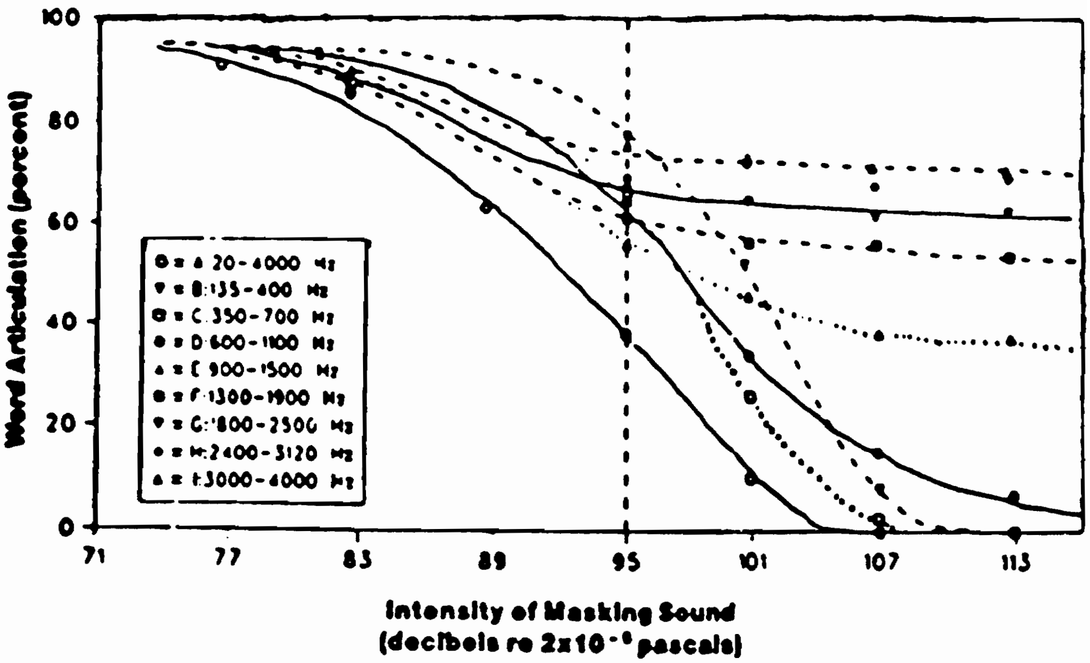

Figure 12. *Percent intelligibility as a function of noise intensity and noise frequency (from Miller, 1947).*

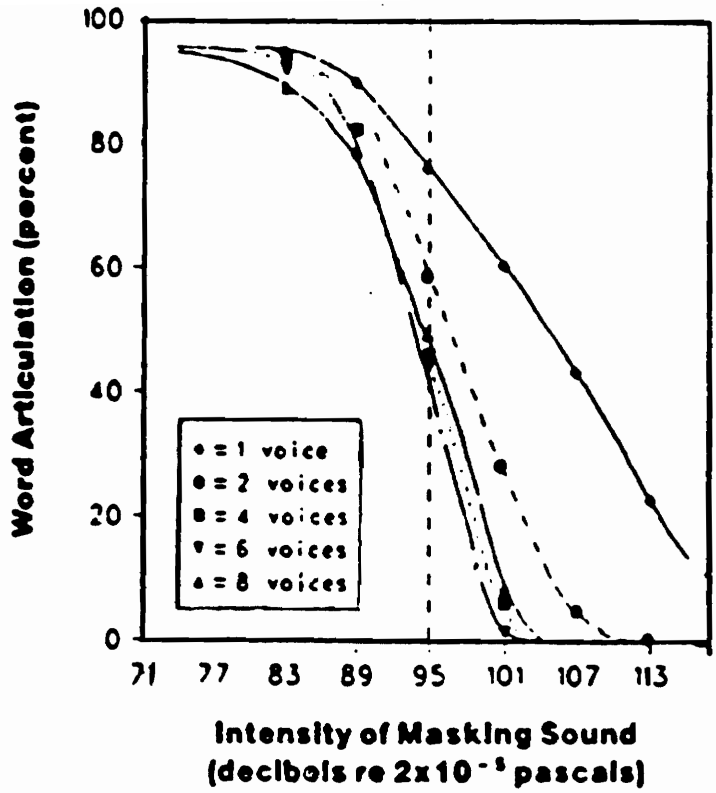

Figure 13. *Percent intelligibility as a function of noise intensity and number of masking voices (from Miller, 1947).*

distraction of attention, was the influential factor.

Egan, Carterette, and Thwing (1954) presented speech and noise in both monaural and dichotic conditions. The speech signal was adjusted until the subject could just understand it with effort. Noise consisted of either a white noise (100–4000 Hz) or a spoken prose passage read by the same voice as the signal. Articulation tests showed that the voice masker was more effective in reducing intelligibility than was the white noise in both the monaural and dichotic conditions.

Abel et al. (1982) compared speech presented in a backgrounds of quiet, white noise or taped crowd noise and determined that taped crowd noise was a more effective masker of speech than white noise. In addition, intelligibility dropped significantly in both types of noise with a 10 dB reduction in the speech level. No information was given on how the noise measurements were made, nor on how speech levels were quantified.

The Effect of Hearing Loss on Speech Intelligibility

Sensorineural hearing loss can have a significant effect on speech intelligibility in a noisy background. In a noise background of 60 dBA, subjects with more than mild noise-induced hearing loss levels performed worse on intelligibility tests than a control group with normal hearing (< 15 dBHL) (Acton, 1970). Mean pure-tone audiograms of all hearing-impaired groups are shown in Figure 14. Surprisingly, a group with mild hearing loss performed better than the normal-hearing group (see Figure 15). This was attributed to a conditioning process on initial noise exposure that the normal-hearing group did not undergo. Figure 15 also indicates that a drop-off in intelligibility occurs at an S/N ratio of 10 dB for all groups and gets progressively worse. Although no information was given concerning the interaction between hearing loss and S/N ratio, it appears from inspection of the data that there was none below 10 dB.

Pekkarinen, Salmivalli, and Suonpää (1990) evaluated four groups with sensorineural

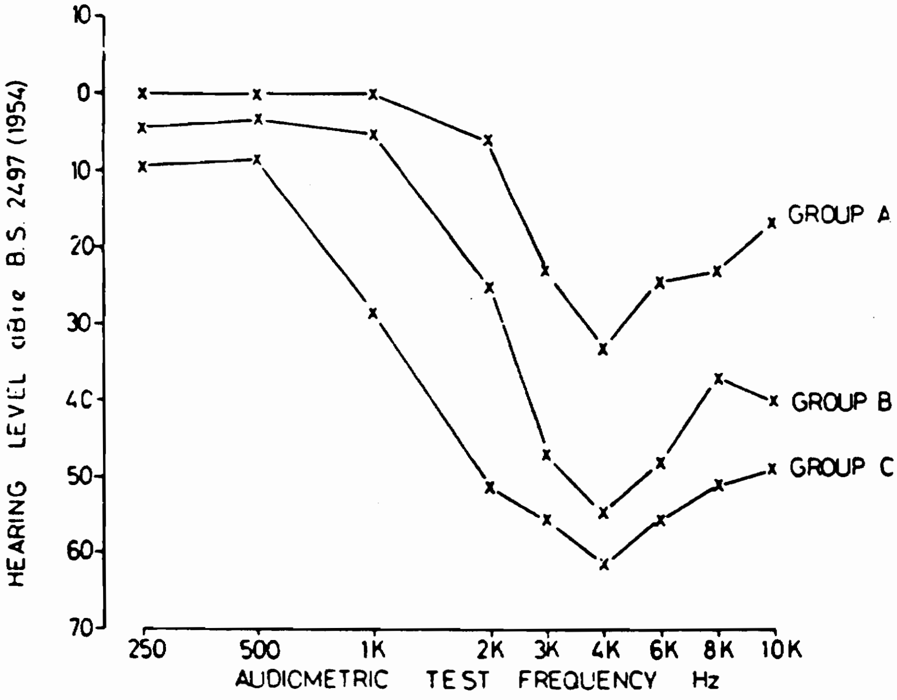

Figure 14. *Mean pure-tone audiograms for all hearing-impaired groups (from Acton 1970).*

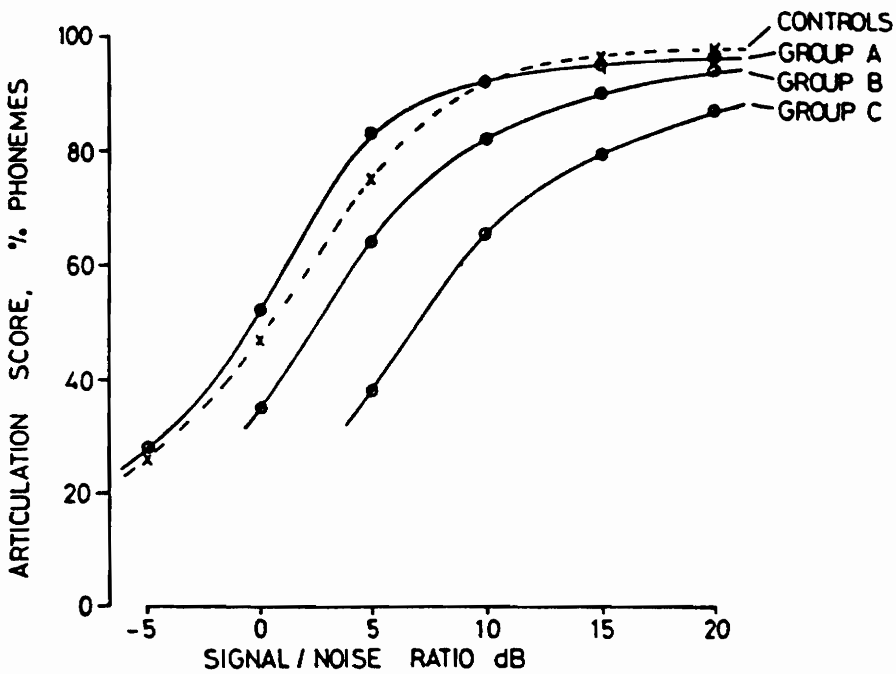

Figure 15. *Word articulation as a function of signal-to-noise ratio. "Controls" denotes the normal-hearing group, while hearing-impairment worsens from group A to group C (from Acton, 1970).*

hearing loss, one group with conductive hearing loss, and one normal-hearing group in pink noise at S/N ratios of -3, -8, and -13 dB. Pure-tone audiogram means and standard deviations for all groups are shown in Figure 16. Subjects with sensorineural hearing loss had significantly lower speech discrimination scores than the subjects with normal hearing or conductive losses and were more adversely affected as the S/N ratio was reduced (see Figure 17). Similar results were reported by Lyregaard (1982). The performance of subjects with frequency loss above 1000 Hz suffered similarly to those with flat or sloping loss (Pekkarinen et al., 1990).

In a previously mentioned study, Abel et al. (1982) compared intelligibility scores of subjects with normal hearing ("according to conventional audiometric tests" (Abel et al., 1982, p. 372)), those with noise-induced high-frequency loss (5 to 25 dBHL at 500 Hz with a slope in loss of 35 to 65 dB between 500 and 4000 Hz), and those with bilateral flat loss (30 to 50 dBHL at 500 Hz and 45 to 65 dBHL at 4000 Hz) when speech was presented with an 85 dBA masking noise. When the S/N ratio was lowered from +5 dB to -5 dB, the intelligibility scores of all groups dropped. The subjects with hearing loss had substantially lower speech discrimination than the normal-hearing group. In addition, there were significant interactions between hearing loss configuration and noise background, hearing loss configuration and S/N ratio, and noise background and S/N ratio. However, no specific information about the sources of the interactions was given.

Other Factors Influencing Speech Intelligibility

Visual cues. When speech intelligibility is substantially degraded due to noise, lip-reading cues may provide a higher level of discrimination. Erber (1969) reported that as the S/N ratio falls, intelligibility of speech without visual cues decreases more rapidly than when the speech is presented with accompanying visual lip-reading cues (see Figure 18).

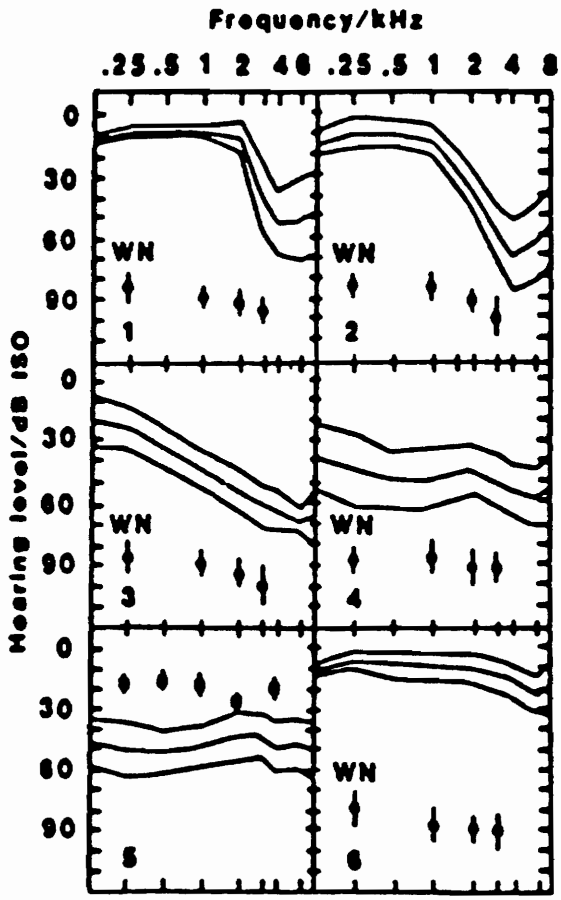

Figure 16. *Pure-tone audiogram means and standard deviations for all groups. Panel 5 denotes the group with conductive loss, panel 6 denotes the normal-hearing group, and the other panels denote the groups with sensorineural hearing loss (from Pekkarinen, Salmivalli, and Suonpää, 1990).*

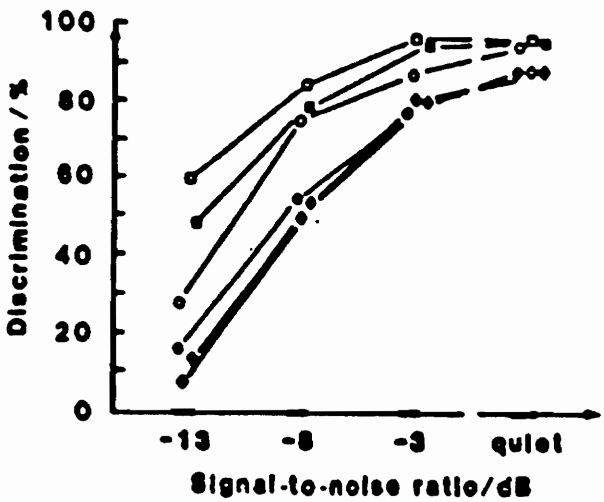

Figure 17. *Word discrimination as a function of signal-to-noise ratio and hearing loss level. The solid square represents the normal-hearing group, the hollow square represents the group with conductive hearing loss, and the other symbols represent groups with sensorineural hearing loss (from Pekkarinen, Salmivalli, and Suonpää, 1990).*

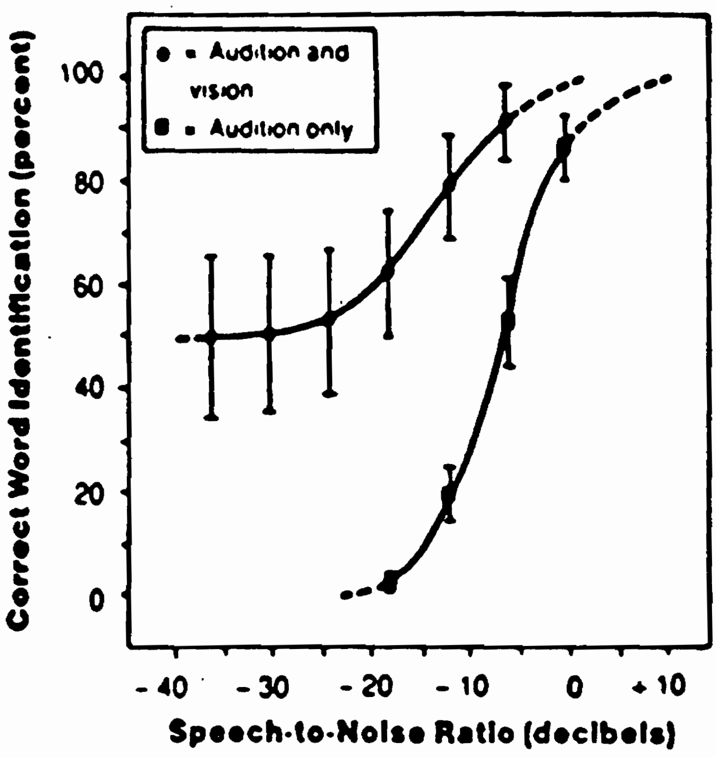

Figure 18. *Percent intelligibility as a function of signal-to-noise ratio and the use of visual cues (from Erber, 1969).*

Martin, Howell, and Lower (1976) found that visual cues improved intelligibility for subjects with normal hearing in noise levels at or below 80 dBA. Above 80 dBA, visual cues increased intelligibility by 30%. The important implication is that useful information is lost when the talker is not visible, as in communication systems, or when there is no talker, as with speech synthesis devices.

Message set size. In a study of the influence of message set uncertainty on speech intelligibility, Miller, Heise, and Lichten (1951) varied the message set size and the S/N ratio in a noise background of 90 dB. After the listeners were trained on the message content, they found that as the number of possible words increased, the 50% speech reception threshold also increased (see Figure 19). For a message set size of 256 words, the 50% threshold was attained at an S/N ratio of -4 dB.

The same results occurred when the predictability of the message set was evaluated. Miller et al. (1951) used digits, words in sentences, and nonsense syllables at S/N ratios of -18 to +18 dB in a background noise of 90 dB. The rate of increase in recognition accuracy was greatest for the digits, then words in sentences, and was the least for nonsense syllables.

Prediction of Speech Intelligibility

Because intelligibility testing is time consuming, several analytical methods have been developed to predict intelligibility from knowledge of the noise and signal spectra alone. The Preferred-Octave Speech Interference Level (PSIL) (Peterson and Gross, 1978) roughly estimates the type of speech possible in a noise environment by simply calculating the arithmetic average of the noise levels at octave bands centered at 500, 1000, and 2000 Hz. However, this method is of limited use because it does not depend on levels of speech and important components of the noise might lie outside the range of the measured octave bands.

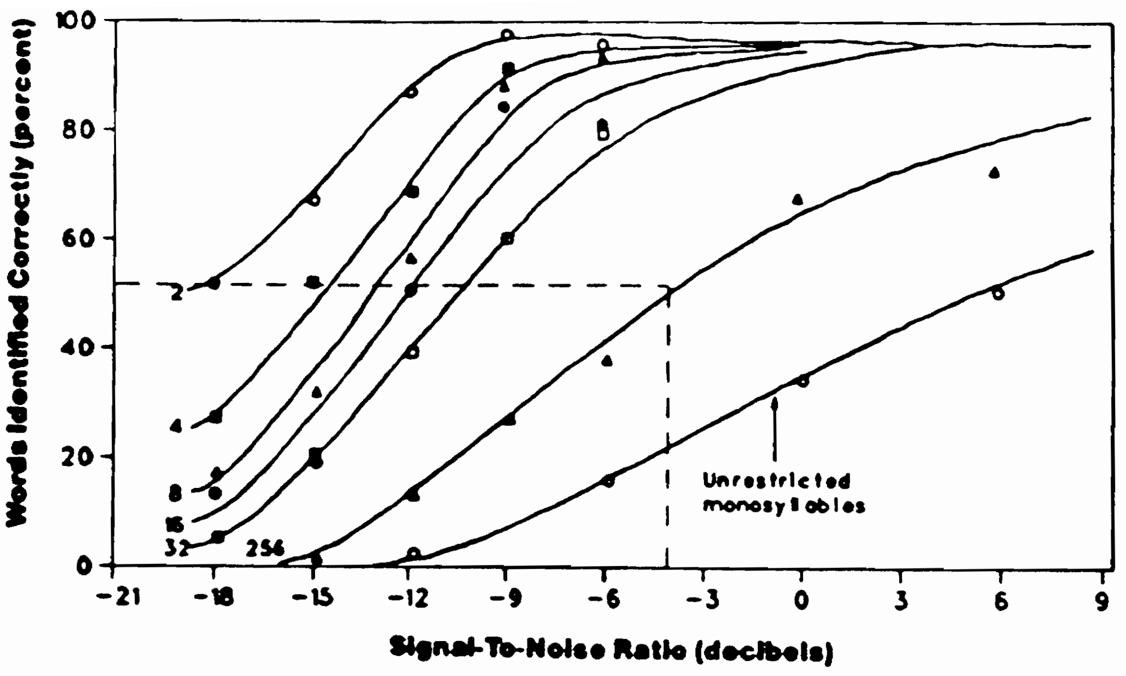

Figure 19. *Intelligibility of monosyllabic words presented in noise as a function of signal-to-noise ratio for seven vocabulary sizes (from Miller, Heise, and Lichten, 1951).*

When using the octave or one-third octave band method, the Articulation Index (AI) (French and Steinberg, 1947) measures both the speech and the noise levels and weights the frequency bands according to their contribution to speech intelligibility. AI values range from 0.0 to 1.0, with increasing values corresponding to higher predicted intelligibility. Although a more complex method than the PSIL, the AI is useful because it can predict speech intelligibility in more varied settings. Figure 20 shows the relationship between AI and intelligibility of various PB words, rhyme tests, and sentences. A limitation of the AI when used as per ANSI S3.5-1969 r 1986 is that it is intended to be used only when the speech is presented by male talkers. Because this standard was published before use of synthesized speech displays were common, it is not known whether the AI is applicable to such displays.

Recently, Parizet (1992) attempted to determine the appropriate weighting function to use in computing the AI in automobiles. In comparing different cars and driving conditions, he found that, because of the particularity of a car's frequency spectrum, the weighting functions advocated by different authors are not significantly different from each other or the ANSI S3.5-1969 r 1986 standard; therefore, the ANSI standard for AI estimation is sufficient. He also notes that the various weighting functions can give disparate results in the case of hearing impairment, although no evidence for this conclusion is given in the paper.

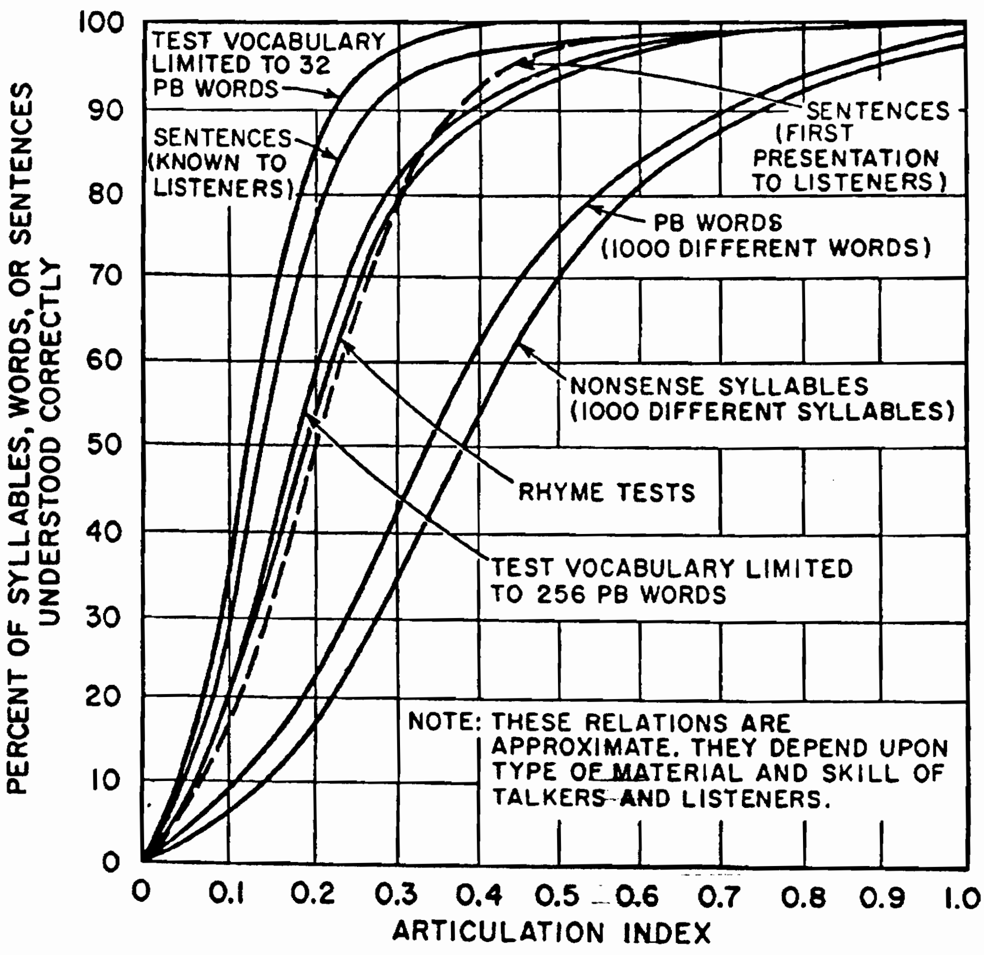

Figure 20. *Relationship between the Articulation Index and the intelligibility of various types of speech test materials (from ANSI S3.5-1969 r 1986, Figure 15).*

SYNTHESIZED SPEECH

Background

There are three methods for presenting speech in auditory displays: analog recordings of actual voice, digitized actual speech, and synthesized (machine-produced) speech. Analog recordings require relatively little storage space per sample and, when of high fidelity, sound exactly like live, natural speech, but they take considerable time to access specific words or phrases which makes them impractical for warning displays. Digitized recordings and speech synthesized by analysis can be accessed much faster than analog recordings, but they require a large amount of memory and sound unnatural when the connected prose must be concatenated together using words stored separately (i.e., when there is a lack of coarticulation). This occurs when a word must be retrieved from the vocabulary and then encoded for speech generation, which causes unnatural pauses to occur. In addition, both of these methods' messages are restricted to words that are previously recorded and stored.

Synthesis-by-rule does not encode previously recorded speech. Rather, it uses a dictionary of phonemes and linguistic rules to construct speech that has the coarticulation and prosody, or rhythm, of natural speech. This technique allows a flexibility not found with the other methods because words and messages do not have to be prerecorded, which makes the possible vocabulary set virtually unlimited. Synthesis-by-rule systems have also become relatively inexpensive, and some are approaching the quality of natural speech. Lack of natural quality has been their main disadvantage in the past.

Intelligibility of Synthesized Speech

Simpson and Navarro (1984) claimed that although every variable evaluated to date had been shown to influence the intelligibility of synthesized speech in the same manner that it influenced the intelligibility of human speech, the effect was often stronger with synthesized speech. This claim is consistent with findings that have shown synthesized speech to be less intelligible than human speech.

In a series of studies (Logan, Greene, and Pisoni, 1989; Pisoni, Greene, and Nusbaum, 1985; Pisoni and Nusbaum, 1986), the intelligibility of nine commercially available speech-to-text systems (MITalk-79, Telesensory Systems TSI Prototype-1, Digital Equipment DECtalk 1.8, Infovox SA 101, Speech Plus Prose 3.0, Votrax Type'n'Talk, Street Electronics Echo, Commodore Amiga 500, and First Byte Smoothtalker) was compared to that of human speech. Using the closed-response MRT and a subject pool of undergraduate students, the authors found that the DECtalk Version 1.8 produced the best overall intelligibility, with an error rate of 3.3% for the Perfect Paul voice (different voices can be selected with the DECtalk model), thus giving 96.7% intelligibility. Natural speech produced 0.5% errors, or 99.5% intelligibility. Error rates were calculated by combining the error rates found in both the initial and final consonants, and all tests were run with an S/N ratio of +35 dB. For the initial consonant, the Paul voice was not significantly different from natural speech. The worst system, Street Electronics Echo, produced a total error rate of 35.56%. The results of this study yielded somewhat elevated intelligibility scores because they were not corrected for chance or guessing.

A study comparing the later version 2.0 of DECtalk with human speech and other speech synthesizers found 88.2% intelligibility for the DECtalk Paul voice when using the Diagnostic Rhyme Test and 95.6% intelligibility for human speech (Pratt, 1987). The disparity in results between this study and the previous study can be accounted for not only by the difference in DECtalk versions, but also by the test used (DRT vs. MRT), by the correction for chance or guessing in the Pratt study, and by the fact that the Pratt study was conducted in England. The DECtalk speaks in American English, and therefore some differences can be expected between English and American listeners. In spite of these differences, the DECtalk was still the most intelligible speech synthesizer tested.

Kangas and Allen (1990) conducted a further comparison of the DECtalk Perfect Paul voice with recorded human speech and sought differences which might be attributable to the hearing loss level of the listener. With the stimuli presented at "a comfortable listening level" (Kangas and Allen, 1990, p. 752), 12 normal-hearing subjects and 12 subjects with sensorineural hearing loss performed an open-response MRT with both the Paul voice and human speech. Subjects with hearing loss had pure-tone average hearing levels at 500, 1000, and 2000 Hz ranging from 25 to 70 dB in the better ear. Significant differences in intelligibility were found between the two voices and the two hearing levels, yet there was no significant interaction between the two variables. Furthermore, the mean intelligibility score for the normal-hearing group listening to the DECtalk was only 73% correct, leading the authors to speculate that age is a factor in intelligibility of synthesized speech (ages of the subjects ranged from 49 to 69 years). A major confounding in this study was the authors' decision to let hearing-impaired subjects wear hearing aids if that was their preferred means of listening to speech. Only some of the hearing-impaired subjects wore hearing aids, and no method of controlling hearing aid volume was indicated.

In a study comparing the performance of normal-hearing, young subjects to that of hearing-impaired, elderly subjects listening to both DECtalk v1.8 Perfect Paul and human speech, Humes, Nelson, and Pisoni (1991) found that the hearing-impaired listeners performed equally well with either the synthesized or human speech, while the normal-hearing listeners performed significantly better listening to the human speech. In addition, the normal-hearing group performed significantly better than the hearing-impaired group for both types of speech. This study was conducted with both the open-and closed-response MRT format, but no detailed data about the intelligibility scores were given. Graphs are provided which suggest a 95% intelligibility for the normal-hearing group in the closed-response MRT and 80% intelligibility for the hearing-impaired group.

Another study which made use of the open-response format MRT in the comparison of three speech synthesizers was conducted by Mirenda and Beukelman (1987). Again, DECtalk was found to be the most intelligible, with 81.7% intelligibility for adult normal-hearing listeners.

There are indications that the difference in intelligibility between synthetic and human speech is increased in the presence of noise. Pisoni and Koen (1982) assessed the two types of speech at several S/N ratios and found that the intelligibility of synthetic speech was more affected by noise than human speech. This decrement was attributed to the lack of acoustic cues in the phonetic structure that is present in human speech. No information was given about the type of speech synthesizer used or the S/N ratios that were evaluated. Results from Pratt (1987) showed that human speech did not produce a decrease in intelligibility as severe as the DECtalk Perfect Paul voice when conditions were changed from no noise to a 0 dB S/N ratio. Human speech was 79.8% intelligible at the 0 dB S/N ratio while the Paul voice was 62.1% intelligible.

The results of a study by Clark (1983) showed that, although the vocalic segments of synthetic speech were equivalent in intelligibility to human speech, some types of consonants were less intelligible. Synthesized consonants were more vulnerable to masking noise than human speech consonants, suggesting that synthesized phonetic cues are less robust perceptually. The reader who is interested in the types of phoneme confusions that are made with synthesized speech is referred to Clark (1983) and Pratt (1987).

Numerous studies have evaluated the various features of synthesized speech in warning messages to pilots in aircraft cockpits. Simpson (1976) investigated the effect of linguistic redundancy on synthetic speech comprehension and found that comprehension is greater with short sentences than with simple two-word warnings. This result held true for monosyllabic words, but the effect was not significant for the polysyllabic words. The redundancy of the polysyllables was apparently good enough to be intelligible to pilots familiar with the vocabulary. These experiments were performed with well-trained, healthy pilots listening to simulated weather broadcasts at 60 dB. The S/N ratio ranged from -8 to +8 dB, and the articulation scores at the -8 dB S/N ratio varied from 15% to 30% intelligibility.

In wide-body jet cockpit noise, altitude callouts were heard by airline pilots at an S/N ratio of -10 dB (Simpson and Navarro, 1984). Although the pilots were unfamiliar with the accent of the speech synthesizer, intelligibility of 99.7% was produced. Voice messages were hand-edited at the phonetic level to restore phonetic errors, which could be partly responsible for the high levels of intelligibility.

Simpson and Marchionda-Frost (1984) evaluated the effects of pitch and speech rate on intelligibility in simulated helicopter noise of 85 dB. Again, the subjects were well-trained pilots with normal hearing. In addition, it is not clear whether the background noise included competing voice messages, but in a separate study (Simpson and Navarro, 1984) it was reported that the S/N ratio was -23 dB. Results showed no change in intelligibility for either independent variable, although a change in reaction time to the message was significant. The range of intelligibility was 98.7 to 99.8%, and speech rate ranged from 123 to 178 words per minute (WPM).

Somewhat contradictory are the findings of Marics and Williges (1988). Reduction of the message rate was found to significantly increase intelligibility when subjects were asked to transcribe messages from a speech synthesizer, but the range of the speech was

150 to 250 WPM as compared to 123 to 178 WPM in the Simpson and Marchionda-Frost (1984) study. Repetition of a message also increased intelligibility. This experiment was conducted in quiet conditions with untrained listeners.

An experiment by Slowiaczek and Nusbaum (1985) resulted in similar findings, which indicated that a speaking rate of 150 WPM produced significantly higher intelligibility of synthetic speech did than a rate of 250 WPM. Furthermore, they concluded that segmental information may be more important to the intelligibility of synthesized speech than suprasegmental factors such as speaking rate and pitch.

The effects of alerting cues have been investigated extensively. Simpson and Williams (1980) investigated whether the recommendation from some design guides that a tone precede a voice warning message to attract attention was warranted. The study showed that, while responses with an alerting tone were slower than without, lengthening the warning with another word rather than a tone did not increase system response time. System response time was defined as the time from onset of the warning signal to the pressing of the appropriate response button. The conclusion was that the alerting tone, which increased the time to obtain the message, was not needed by the pilot because the message served as its own alerting cue. It was suggested that signal words such as "Danger!" or "Warning!" be used in case other types of messages were also delivered to the listener via synthesized speech. This recommendation was supported in a later study by Hakkinen and Williges (1984).

Bucher, Karl, Voorhees, and Werner (1984) presented 85-dB synthesized speech to experienced helicopter pilots over headphones with a background chatter at 75 dB. Simulated helicopter noise was played through loudspeakers in the sound field at a level of 80 dB, and no information was given about the attenuation afforded by the headphones. Reaction time (i.e., the interval from message onset to the finger leaving the resting key) and movement time (i.e., the interval from the finger leaving the resting key to the pushing of the appropriate button) were evaluated. Statistical analysis revealed no significant differences between the use of alerting tone or word prefixes. System response time, as defined in the previous paragraph, was significantly shorter when no alerting prefix of any kind was used. Byblow and Corlett (1989), however, found that, even though system response time was shortest with the key word only, movement time was significantly longer, suggesting that more cognitive demand is placed on the subjects in this condition.

There are indications that synthetic speech may be more taxing on cognitive processes than is human speech. In a series of recall experiments using synthetic and natural speech as stimuli, Luce, Feustel, and Pisoni (1983) found significant decrements in recall when messages were presented by synthetic speech. Conclusions were that some of the acoustic-phonetic information from the synthesized speech was not properly encoded and that synthetic speech places more demands on short-term memory.

Using a time estimation task, Hart and Simpson (1976) found that key word messages required significantly more attentional demand than when the key words were embedded in sentences. This difference held true only for the monosyllabic key words. When listening to messages containing polysyllabic key words, the pilots performed equally well with sentences and key words only. Again, this result indicates that monosyllabic words are best presented in a sentence format.

Several studies have shown that the amount of training can have an impact on speech intelligibility. Simpson and Navarro (1984) reported that a significant improvement in intelligibility occurred as the subject went through the experiment. This improvement was attributed only to the familiarity with the synthesizer because the subjects were given no training prior to the experiment and no feedback on performance was given. Oshrin and Siders (1984), with a similar experiment, concluded that discrimination of synthetic speech monosyllables improves when errors are corrected and other appropriate training is applied.

When the quality of the speech synthesizer is poor, training shows an even greater effect. In a battery of tests, subjects were trained by synthetic speech, human speech, or received no training (Schwab, Nusbaum, and Pisoni, 1985). Analysis indicated that the subjects trained by synthetic speech had the same MRT scores as the other subjects on the first day of testing, but by the tenth day of testing scored significantly better. These and data from the other tests imply that familiarity with the speech synthesizer can be more important to intelligibility than familiarity with the experimental procedures.

RESEARCH OBJECTIVES

From the preceding discussion, it is obvious that there are many factors which affect the intelligibility of synthesized speech in noisy environments. However, there is little known about how these factors interact with hearing threshold levels of the user population. Nearly all previous studies have used normal-hearing listeners, although over 10% of the general population has some level of hearing impairment (NIH, 1990). In addition, relatively little research has been conducted on the intelligibility of synthesized speech in noise.

It is apparent that truck cabs may have significant potential to damage the hearing of operators over a period of exposure. The first objective of this research was to quantify the sound and exposure levels in a typical modern truck cab under several driving conditions. Four different truck cab/engine type combinations were analyzed for overall sound levels and frequency spectra. These measures allowed an estimation of truck cab noise exposure hazards and provided the necessary noise background recordings for the laboratory experiment.

The laboratory experiment had several objectives. Because truck cab noise is so prevalent, a significant portion of the driver population may already have some noise-induced hearing loss. With the advent of IVHS systems that use synthetic speech for auditory displays, it is necessary to understand the impact that hearing loss could have on the intelligibility of speech warning displays for critical situations. Therefore, the first objective of the laboratory experiment was to investigate the differences in synthesized speech intelligibility for hearing-impaired and normal-hearing listeners.

The second objective of the experiment was to evaluate the effect of noise type. Comparing the truck noise background with a laboratory-produced pink noise background allows the results of the experiment to be more readily repeatable. In addition, the effect of the addition of competing background speech, which may be present in the truck cab environment, was examined. A quiet condition was also evaluated to determine intelligibility differences between hearing-impaired and normal-hearing listeners in high S/N ratio environments and to directly compare the results to those of other studies.

It is important to relate these effects to the speech level at which the message will be presented because that is the factor most easily changed by design engineers or adjusted by end users. The third objective of this experiment, therefore, was to evaluate intelligibility at various speech levels and to examine how the resulting S/N ratios interact with hearing loss level and noise type.

Finally, the ability to measure the truck cab noise spectrum and analytically estimate the intelligibility that would be attained would be a helpful design tool. The Articulation Index (AI), as described in ANSI standard S3.5-1969 r 1986, is a physical measure that has been shown to be highly correlated with speech intelligibility and is used to predict the relative performance of communication systems. Although the standard is intended to be used with speech presented by male talkers only, the ANSI S3.5-1969 r 1986 articulation weighting factors were assessed to determine if they are reasonable to use with a synthesized speech display and with hearing-impaired listeners. The analytical AI results were compared with the empirical results for each noise condition and hearing level.

TRUCK CAB INTERIOR NOISE MEASUREMENTS

Measurement Convention

In quantifying the noise and speech levels for the measurements and experiment, all octave-band sound pressure levels (SPL) are referenced to 20 μPa and based on root-mean-square (rms) measurements in unweighted dB or dB(linear), hereafter simply termed dB, unless specified otherwise. Where appropriate, the overall A-weighted (dBA) SPL is provided.

Truck Noise Measurements and Recordings

Vehicles and instrumentation. Four different truck cab/engine combinations were supplied by Volvo/GM Heavy Truck Corporation of Dublin, Virginia for testing. The four combinations were a conventional cab with a Cummins model N14 engine, a conventional cab with a Detroit Diesel Series 60 engine, a cab-over-engine with a Cummins model N14 engine, and a cab-over-engine with a Detroit Diesel Series 60 engine. The cab-over-engine had a sleeping compartment which was closed for the measurements; the conventional cab had no sleeping compartment. The Detroit Diesel conventional cab was tested pulling an unloaded flatbed trailer. The other cab/engine combinations were tested without a trailer. Measurements were made using a Rion UC-53 omnidirectional microphone placed 15.24 cm to the right of the driver's right ear and a Rion NA-29E octave-band analyzer. Recordings were made using a Realistic super cardioid dynamic microphone and a Teac model 124 audio tape recorder.

Method. All recordings and measurements were made on days without precipitation, with the windows closed, and with all accessories turned off. First, reverberation time measurements of both cab types were made using a taped pink noise excitation stimulus to allow faithful laboratory reproduction of the truck cab reverberation characteristics.

Reverberation times for both cab types appear in Table 2. The reverberation time (T_{60}) refers to the time, in seconds, required for the sound level in an environment to fall 60 dB after a noise abruptly ceases. Because a noise decay of 60 dB was not possible with the equipment available, a 15 dB decay was measured. The resultant time was then multiplied by four to estimate the T_{60}.

Following the reverberation measurements, one set of noise measurements for each cab/engine combination was made according to SAE standard J336 (1988) to determine the truck cab interior noise over the upper range of engine speeds. During each run, the vehicle was accelerated at wide-open throttle from one-half rated engine speed up to rated engine speed. Runs continued until four measurements of the maximum A-weighted SPLs were within a total range of 2 dBA. The reported SPL (Table 3) is the average of these four measurements. Measurements contaminated due to noises unrelated to normal vehicle operation (e.g., grinding gears) were discarded.

The second set of noise measurements was taken on an interstate freeway with the cruise control set at 62 mph (100 kph). A-weighted overall L_{eq} measurements, as well as octave-band spectral measurements, were taken over two five-minute runs on the same stretch of freeway traveling in opposite directions to minimize measurement differences due to wind direction and hills. These two measurements were then averaged to assess the driver's exposure to truck cab noise in a typical freeway setting. Recordings of the truck cab noise were also made at this time for playback during the experiment.

Results. Results of the truck cab noise measurements are shown in Table 3. The higher SPLs for the Detroit Diesel engine compared to the Cummins during both sets of tests are most likely due to variations in construction because both engines are four-stroke diesel designs. The conventional cab type had generally lower SPLs in the SAE J336 test, although the differences between the two types were not as great during the freeway measurements. In comparison of these noise measurement data with those collected in

TABLE 2

Truck Cab Reverberation Times

	Conventional Cab	Cab-over-engine
Octave Band (Hz)	T_{60} (s)	T_{60} (s)
63	0.29	0.39
125	0.26	0.23
250	0.16	0.16
500	0.16	0.16
1000	0.15	0.16
2000	0.16	0.16
4000	0.16	0.16
8000	0.15	0.16

TABLE 3

SAE J336 and Freeway Interior Truck Cab Noise Measurements

SAE J336 Measurements (Under Open-Throttle Acceleration):

Truck Type	Maximum (dBA)
Cab-Over-Engine with Cummins	80.1
Cab-Over-Engine with Detroit Diesel	84.5
Conventional cab with Cummins	77.9
Conventional cab with Detroit Diesel	81.9

Freeway Measurements (Spectral Measurements are in L_{eq} (dB)):

	Octave Band Center Frequency (Hz)								
Truck Type	63	125	250	500	1000	2000	4000	8000	Overall (dBA)
Cab-Over-Engine									
with Cummins	92.1	86.5	80.2	73.1	67.7	63.5	58.9	52.6	76.9
with Detroit Diesel	92.7	88.7	79.1	71.7	70.2	68.6	60.8	53.0	77.9
Conventional cab									
with Cummins	85.9	80.8	75.1	72.5	68.0	62.5	57.2	52.5	74.3
with Detroit Diesel	98.2	88.1	76.5	72.7	72.7	68.3	60.0	52.5	78.8

earlier studies, it can be seen that the modern trucks have a substantially lower overall noise level than the 1970s vintage trucks to the point that the trucks measured for this study pose a minimal noise exposure risk to the drivers. The noise levels for all four trucks are well below the OSHA action level, which is 85 dBA time-weighted average (TWA) for an eight-hour day (OSHA, 1983). Because the freeway L_{eq}s are below 80 dBA for all cab/engine configurations, the OSHA noise dose is 0% for each truck and therefore does not depend on the exposure time at freeway cruising speeds (This dose results from the fact that the OSHA threshold level for including a noise in the exposure measurement is 80 dBA). However, the addition of accessory noise and wind noise from an open window could add substantially to the driver's noise exposure. This reduction in interior noise from prior measurements on earlier model trucks can most likely be attributed to the improvement in cab design, drivetrain design, and passive noise control in the intervening years.

EXPERIMENTAL METHOD AND DESIGN

Experimental Design

The experimental design was a three-way, mixed-factors layout with one between-subjects variable and two within-subjects variables. A complete factorial design was used to allow the resolution of all main effects and interactions in an analysis of variance (ANOVA). A graphic layout of the experimental design is illustrated in Figure 21.

Hearing level variable. Two levels of this between-subjects variable were used; these were determined by assessing the subjects' pure-tone average (PTA) hearing level at 1000, 2000, 3000, and 4000 Hz during audiometric exams. These frequencies were chosen because they encompass the speech bandwidth critical to intelligibility. Subjects were grouped by normal-hearing (PTA below 25 dBHL) and hearing-impaired (PTA between 25 and 50 dBHL, inclusive). Under the classification scheme devised by Miller and Wilber (1991) as shown in Table 1, this level of impairment would result in a slight to mild handicap. The criterion was based on the better of the two ears, and a bilateral difference of no more than 20 dB was allowed. Hearing level data for all subjects can be found in Appendix A, and mean audiogram profiles of the subjects' better ears for each group are shown in Figure 22. The mean PTA for the normal-hearing group was 1.1 dBHL, and the mean PTA for the hearing-impaired group was 35.9 dBHL.

Noise type variable. Because pink noise is common as a representative broadband masking noise in speech intelligibility testing and can be accurately replicated across trial presentations, it was selected as one of the three noise types. The second noise type was the freeway interior truck cab noise recorded from the conventional cab with the Detroit Diesel engine because it had the highest measured overall L_{eq}. For the third noise type, the truck noise was mixed with selected speech from an audio cassette abridgment of the novel, *The Firm.* The selection was chosen to present a relatively constant level of

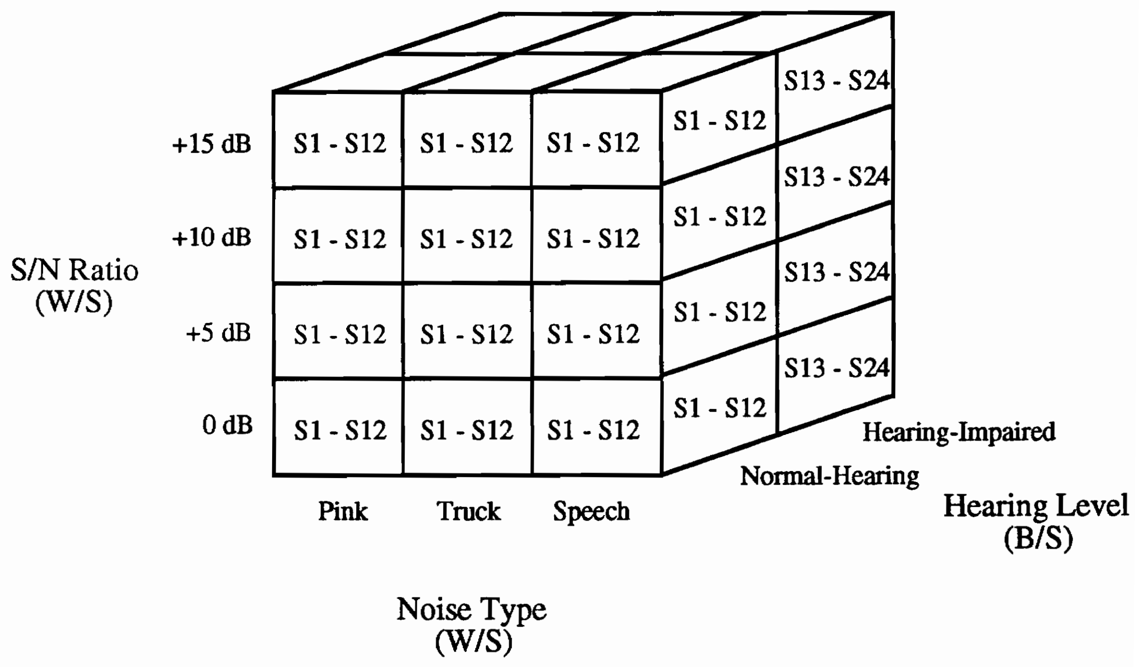

Figure 21. *Experimental design. (Specifics on variable levels are discussed in text.)*

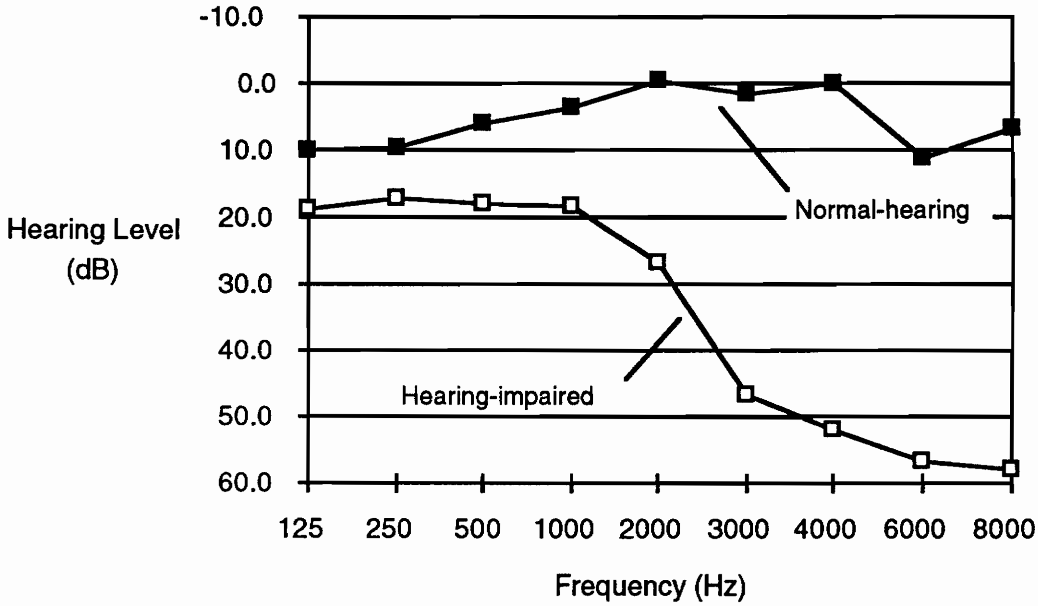

Figure 22. *Mean pure-tone hearing levels for subjects' better ears (12 per group).*

background speech without music, and the combination of truck noise and background speech (hereafter referred to as the speech noise) was intended to provide a simulation of in-cab radio chatter which might interfere with the synthesized speech message. The speech noise type was not intended to be used to compare noises equivalent in level but different in amount of speech content; instead, it was intended to show what differences would occur with the *addition* of background speech to the truck noise already present.

Speech-to-noise ratio variable. Four levels of speech-to-noise (S/N) ratio were manipulated by varying the level at which the speech was presented in the noise. The S/N ratios were 0, 5, 10, and 15 dB for the pink noise and truck noise. These S/N ratios were chosen because they encompass a large range of intelligibility and because attenuation of the speech signal was most easily accomplished in 5 dB increments using a detented audiometer attenuator. Because the speech noise had a higher overall noise level than the other noise types (due to the addition of the speech to the noise already present on the recordings), the S/N ratios described are higher than the actual S/N ratios for the speech noise type. For the sake of simplicity in the experimental design, the terminology used for the levels of S/N ratio are the same for all noise types; in other words, the synthesized speech levels were the same across noise type. The S/N ratio was used instead of speech level because it allows easier interpretation of the intelligibility data. Details of how the speech levels were measured and calibrated appear in the experimental procedures section.

Dependent variable. The dependent variable was the percent intelligibility for each treatment condition. The raw data were collected in the form of number of words correctly identified out of the total of number of words presented. The percent intelligibility measure allows comparison to the results of past research and allowed efficient collection of data.

Subjects. A total of 24 paid volunteers were equally blocked into normal-hearing and hearing-impaired groups. In each block of 12 subjects, there were 7 males and 5 females. The ages of the normal-hearing subjects ranged from 21 to 59 years (mean age 32 years), while the ages of the hearing-impaired subjects ranged from 20 to 63 years (mean age 38 years). Detailed characteristics of the subjects' dBHL at each pure-tone frequency, age, and gender can be found in Appendix A. Native English speakers were used to avoid confounding due to the effect of language fluency (Abel et al., 1982). Because subjects who are trained in the use of synthetic speech devices have been shown to perform better on intelligibility tests than those without such training (Schwab, Nusbaum, and Pisoni, 1985), only subjects who were naive in the use of synthesized speech were allowed to participate. Subjects were paid $5 per hour for participation.

Counterbalancing. To avoid the confounding of treatment effects and practice effects for the within-subjects factors, the experimental conditions were presented according to a balanced Latin square, which orders within-subjects treatment combinations so that each condition follows and precedes each of the other conditions once throughout the experiment. Subjects from each block were randomly assigned to treatment orders available from the Latin square design. Because there were 12 subjects in each between-subjects category and there were 12 treatment combinations, the Latin square was totally filled for both hearing levels.

Facilities and Apparatus

The experimental facility used is located within the Auditory Systems Laboratory at Virginia Tech. It includes a reverberant chamber, an anechoic chamber, and various audiometric testing equipment. The laboratory itself is acoustically isolated to maintain a quiet environment for testing.

Anechoic chamber. The anechoic chamber was used to audiometrically screen subjects. This Eckel Corporation chamber is of double-wall steel construction with 7.62 cm of fiberglass acoustic insulation sandwiched between the inner and outer skins. The six surfaces of the chamber interior are lined with acoustic foam wedges, and an acoustically transparent, expanded-metal grating located above the bottom wedges serves as the floor. Although the reverberant chamber can also be used for screening, the experimental set-up favored the use of the anechoic chamber. The interior and exterior dimensions of the anechoic chamber are given in Table 4, and ambient noise levels at the subject's head center position, which conform to the requirements of ANSI S12.6-1984 for threshold testing, are given in Table 5.

Reverberant chamber. The reverberant chamber was used for the experiment to more closely approximate the truck cab acoustics. It is an extensively modified Industrial Acoustics Corporation (IAC) test booth that has double-wall steel construction with 10.16 cm of fiberglass acoustic insulation sandwiched between the inner and outer skins. The reverberant environment results from walls and ceiling composed of 0.64-cm thick hard-tempered masonite layered over 1.27-cm thick gypsum board and a bare metal floor. The interior and exterior dimensions of the IAC test booth are given in Table 6, and ambient noise levels at the subject's head center position, which conform to the requirements of ANSI S3.19-1974 for threshold testing, are given in Table 7.

The sound field for establishing the noise environment was achieved with two two-way loudspeakers, one placed in the front left corner of the room facing the front left corner and the other placed in the right rear corner facing the left wall. This arrangement provided a uniform, diffuse field about the subject's head. An additional loudspeaker was located directly behind the subject for speech presentation. Figure 23 depicts a schematic of the interior layout of the reverberant chamber for the experiment, and a photograph of the room as it was set up for the experiment is shown in Figure 24.

TABLE 4

Interior and Exterior Dimensions of Anechoic Chamber (adapted from Casali, 1988)

	Interior Dimensions (cm)*	Exterior Dimensions (cm)
Length	231.14	365.76
Width	290.84	426.72
Height	215.90	350.52

*Measured from wedge tips.

TABLE 5

Anechoic Chamber Ambient Noise Levels (adapted from Casali, 1988)

Octave Band Center (Hz)	Measured Ambient Noise Levels (dB)
125	23.3
250	5.5
500	5.7
1000	7.5
2000	5.6
4000	7.3
8000	9.3

TABLE 6

Interior and Exterior Dimensions of Reverberant Chamber (adapted from Casali and Robinson, 1990)

	Interior Dimensions (cm)	Exterior Dimensions (cm)
Length	279.40	304.80
Width	188.60	211.46
Height	234.95	263.53

TABLE 7

Reverberant Chamber Ambient Noise Levels (adapted from Casali and Robinson, 1990)

Octave Band Center (Hz)	Measured Ambient Noise Levels (dB)
125	20
250	14
500	6.5
1000	4.5
2000	2.7
4000	5.1
8000	8.1

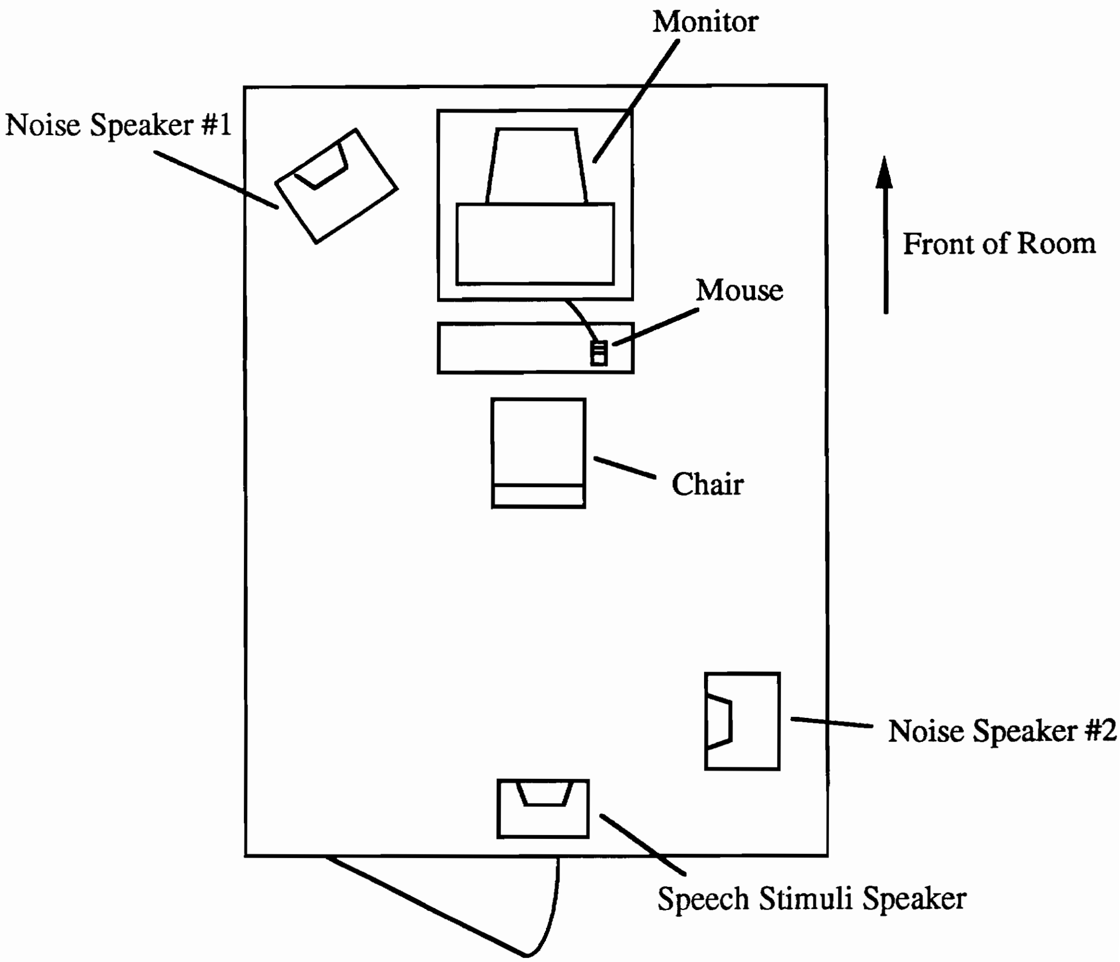

Figure 23. *Schematic of the interior layout of the reverberant chamber as it was set up for the experiment.*

Figure 24. *Photograph of the interior of the reverberant chamber as it was set up for the experiment.*

Five pieces of 5.08-cm thick Sonex™ foam lined the walls of the reverberant chamber to reduce the reverberation times to approximate the truck cab reverberation times. The reverberation times for the reverberant chamber as it was configured for the experiment are shown in Table 8. Although the reverberation times in the chamber were slightly higher than those in the truck, the difference provided a more conservative simulation than the anechoic chamber would have.

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. A Larson-Davis (L-D) Model AE100 artificial ear and an ACO Model 7023 2.54-cm microphone were used for audiometer calibration. An intercom system allowed the experimenter to hear the test subject at all times and to talk to the subject through the earphones or a separate loudspeaker. A closed-circuit television camera was used to allow the experimenter to manipulate the cursor shown on the monitor located in the chamber.

Noise levels in the reverberant chamber were measured using a Larson-Davis model 3100D real-time spectrum analyzer and were made at the head center position using a Larson-Davis model 2540 1.27-cm omnidirectional microphone. These instruments were used to calibrate the system at the beginning of the day and to measure the chamber reverberation times.

Speech stimuli were presented using the Hypercard 2.1 application running on an Apple Macintosh IIci. Word choices were presented visually on a Real Tech 48.26-cm monitor located in the reverberant chamber. Speech output was attenuated in 5-dB steps using the Beltone 114 audiometer and was presented through an Infinity RS 9b loudspeaker located directly behind the subject's head to minimize binaural differences. Subjects were instructed to look straight ahead and keep their head still during the testing. The loudspeaker was driven by an Adcom GFP-555II preamplifier and an Adcom GFA-

TABLE 8

Reverberant Chamber Reverberation Times

Octave Band (Hz)	T_{60} (s)
63	0.28
125	0.46
250	0.37
500	0.27
1000	0.21
2000	0.22
4000	0.21
8000	0.20

545II amplifier. The speech signal was attenuated by 12 dB above 8 kHz using a Ross R31M equalizer to reduce extraneous hiss outside the speech bandwidth. Truck and speech noises were presented from tape on a Teac model 124 tape deck, while pink noise was generated and shaped to be flat by octaves using a Yamaha GE-60 graphic equalizer. Two Infinity RS 6b loudspeakers driven by an NAD 1020B preamplifier and an NAD 2200 amplifier were used to produce a diffuse sound field about the subject's head for all noise environments. Figure 25 is a photograph of the sound generation equipment as it was set up for the experiment.

Speech stimuli. In determining which type of word list to present during this experiment, ANSI S3.2-1989 was consulted. Although it is stated in the standard that it is not intended for use with synthesized speech devices, it is the only relevant standard available and the three recommended sets of test materials have been successfully used in past studies on synthetic speech. The standard's suggested test materials were therefore deemed appropriate for consideration. Of the alternatives, Modified Rhyme Test (MRT) word list stimuli were selected for use because the MRT is better than the phonetically balanced word lists at identifying the sources of poor intelligibility, is easier and quicker to administer in its closed-response nature, requires less training time, is indicative of errors in both the initial and final consonant sounds (the Diagnostic Rhyme Test (DRT) uses only the initial consonant), and it provides a larger response set than the DRT (ANSI S3.2-1989). Furthermore, the MRT has been used in a number of past experiments with the DECtalk and therefore allows the results of this experiment to be compared with those of previous studies. The standard includes 50 six-word lists (see Appendix B); the order of presentation of each word from each list was randomized in the experiment. Each word was presented twice throughout the 12 experimental conditions so that a confounding of word position during the response was avoided.

Figure 25. *Photograph of the noise and speech generation equipment as it was used in the experiment.*

Synthesized speech presentation. Because of previously mentioned studies citing its superior performance relative to other commercial speech synthesizers, the DECtalk version 2.0 text-to-speech system was used to synthesize the speech stimuli. For the present experiment, the "Perfect Paul" voice and the DECtalk default settings were used. The default speaking rate was 180 words per minute. MRT words and the phrase "The word is" were entered into the DECtalk with periods after every word to avoid inflections, with standard orthographic spellings when possible, and phonetically when not (e.g., "fig." was interpreted by the DECtalk as an abbreviation and was pronounced "figure"). The spoken words were then digitized at a sampling rate of 22 kHz via direct line connection into SoundEdit Pro 1.0, a commercially available sound digitizing and editing program for the Macintosh which allows visualization of the sampled frequency spectrum to minimize signal clipping. The MRT words were equalized for presentation time and then saved to a Hypercard stack for playback at 22 kHz during the experiment. Because every sample period gives information about at least one-half of the wavelength, the 22 kHz sampling rate passes sound frequencies at least up to 11 kHz. Therefore, the frequency range was wider than that needed for the critical speech bandwidth.

Experimental Procedures

Setting noise levels. Both the pink noise and the truck noise were played back in the laboratory experiment at the A-weighted L_{eq} equivalent to that measured in the truck with the highest freeway L_{eq}, namely the conventional cab with the Detroit Diesel engine. The L_{eq} for the pink and truck noises in the experiment was 78.8 dBA. A representative 30-second section of the truck noise over which the noise level was relatively constant was selected and redubbed on to another tape so that a full 10-minute tape of uniform truck noise was available. It was then shaped using the Yamaha equalizer and was played back at spectrum levels within 5 dB of the actual measured levels in the critical speech bandwidth. Measurements at the left and right ear using the omnidirectional microphone showed less than a 1-dB difference for either noise type, and directional measurements using an AKG model C414 B-ULS cardioid microphone facing towards the front, back, right, and left of the head center position showed less than a 3-dB difference for both noise types. Results from both sets of measurements (Table 9) are within strict diffusivity standards set by ANSI S3.19-1974 for hearing protector testing and show that no advantage was afforded either ear in this experiment.

To set the levels for the speech noise type, five normal-hearing subjects listened to the background speech while the truck noise was played and were asked to set the level of the speech so that it would be comfortable to listen to for an extended period of time. The overall mean preferred level for the five listeners was calculated to have an L_{max} of 91.1 dBA. The L_{max} was used for this calculation because speech carries its power in the peaks. The background speech at this mean level and the truck noise at its original L_{eq} were then mixed together using a Fostex model 260 mixing board onto a single tape, and the overall L_{eq} for the speech noise type was 81.8 dBA. The truck noise itself had exactly the same spectrum and level for both the truck and the speech noise types, and the only difference was the addition of the speech for the latter type. Figure 26 shows the octave band spectral plots of the three noise types.

Calibration of speech levels. The speech level was calibrated using speech-weighted noise generated by a Beltone 2000 clinical audiometer as per ANSI S3.6-1989. The speech-weighted noise was routed through the Beltone 114 audiometer and then out through the speech sound system as described above. The system gain controls were adjusted so that the long-term rms output of the signal was equivalent to the A-weighted SPL of the pink noise and truck noise and so that the VU meter of the Beltone 114 was reading 0 dB. The Beltone 2000 was then disconnected, and the output line from the Macintosh was connected in its place. A representative set of digitized words was then

TABLE 9

Orthogonal Directional Noise Measurements in the Horizontal Plane and Noise Measurements at the Left and Right Ear (Head Center Level = 78.8 dBA)

Directional Measurements

Truck Noise		Pink Noise	
Direction	Noise Level (dBA)	Direction	Noise Level (dBA)
Front	71.2	Front	73.3
Back	73.5	Back	76.1
Left	72.6	Left	73.3
Right	73.0	Right	74.6

At-Ear Measurements

Truck Noise		Pink Noise	
Ear*	Noise Level (dBA)	Ear*	Noise Level (dBA)
Right	76.4	Right	77.3
Left	76.9	Left	77.6

*Measurements made 5.6 cm from head center position

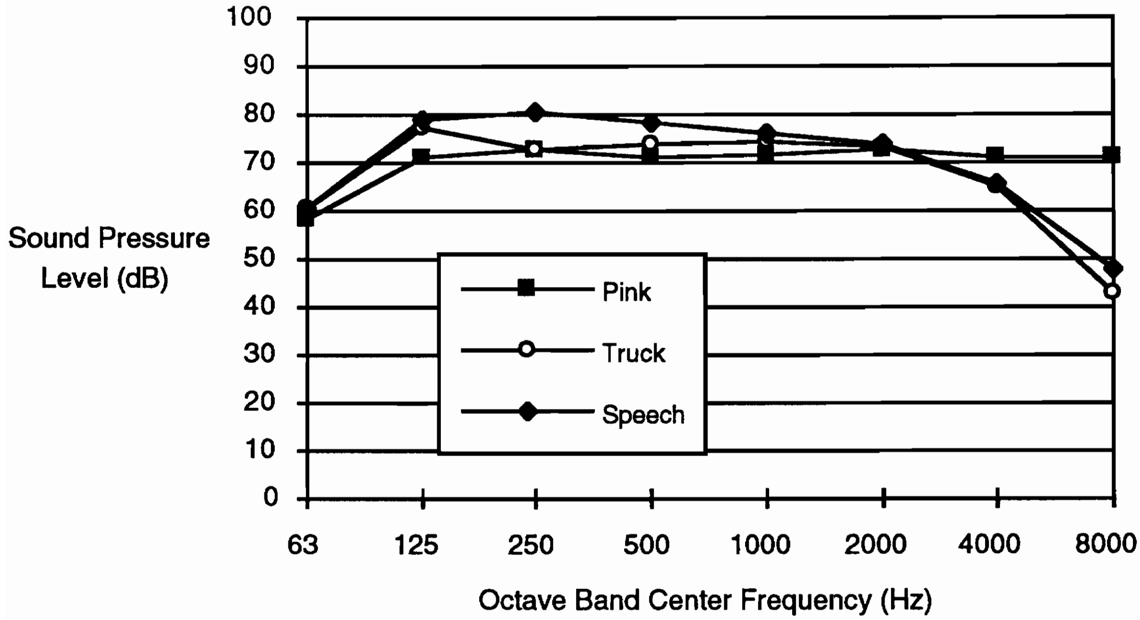

Figure 26. *Octave band spectral plots of pink, truck, and speech noise types, as presented in the reverberant test environment. (Overall dBA levels were 78.8 for pink and truck and 81.8 for speech noise.)*

played, and the gain on the Beltone 114 was adjusted to achieve speech peaks of 0 dB on the VU meter. At this point, the speech level was equivalent to the pink noise level and resulted in a 0 dB S/N ratio. The speech level could then be manipulated in 5 dB increments using the linear step attenuator on the Beltone 114, which in turn would change the S/N ratio in 5 dB increments given a constant noise level.

Subject screening. Subjects were screened prior to participation to determine whether they matched the experimental criteria. After entering the laboratory and being welcomed, subjects were required to read a description of the experiment and provide their informed consent in writing prior to participation. The experiment description given to subjects is shown in Appendix C, and the informed consent form used is in Appendix D. After completing the forms, subjects were asked questions about their age and otological history using the questionnaire in Appendix E. Following the questionnaire, a visual otoscopic examination of the outer ear was conducted using a Welch-Allyn 21700 otoscope to check for excessive ear wax. Subjects with ear wax obstructing the ear canal were excused from the experiment. Subjects were then screened with the Beltone model 114 clinical pure-tone audiometer using a Hughson-Westlake manual audiometry procedure (Carhart and Jerger, 1959). If the subject fit the audiometric criteria described above, the experimental session was scheduled. Otherwise, the subject was excused.

Intelligibility trials. Prior to the intelligibility task, subjects underwent an audiometric test similar to the screening procedure but only at 1000, 2000, 3000, and 4000 Hz in order to test for temporary threshold shift (TTS) during the experiment. This was not only for the safety of the subject, but also to account for any change in hearing level during the experiment that might cause a decrease in intelligibility.

For the MRT task, subjects sat inside the reverberant chamber, in a fixed chair facing the monitor, and were given a computer mouse with which to select their responses. Subjects were instructed to maintain approximately the same body position throughout the experiment and to face forward at all times. This constraint was to ensure that subjects would not turn to hear the synthesized speech with their better ear.

After the experimenter turned on the noise for the trial, the subject was instructed to begin by using the mouse to position the cursor within a box on the screen containing the word "begin" and pressing the mouse button. At that point, the 50-word trial began, and the first word was presented.

The Hypercard program was used to randomize and then present the words in a closed-response format in the experiment. For each treatment condition, a word from each of the MRT's 50 six-word lists was presented so that each condition contained exactly 50 words. Each word was presented following the carrier phrase "The word is" and a pause of approximately 0.5 second. As each word was auditorially presented, subjects saw each of the six word choices appear within separate boxes on the monitor. Because accuracy, not speed, was the emphasis of this experiment, the test was designed to be self-paced. Subjects would use the mouse to select the box which contained the word that they thought they had heard, and only then would the computer present the next word. This design also did not allow word presentations to go unanswered. Subjects were told to guess if they were unsure of their answer, and a correction factor for guessing was applied in score computation, as discussed later. The location of each word choice was the same throughout the experiment, and since each word was presented twice, spatial bias due to visual location of the correct response was avoided. Data about the word presented, the subject's response, and whether the response was correct were collected automatically by the computer.

Initially, the subjects ran through a practice trial which had no background noise (i.e., without the pink, truck, or speech noise). The speech level during the practice trial was 78.8 dBA, the speech level that gave a 0 dB S/N ratio during the experiment. Although the primary purpose of the practice trial was to familiarize the subjects with the task and the synthesized speech, data were collected for this trial. After subjects had shown understanding of the task in the practice trial, they proceeded with the 12 experimental conditions. Prior to each 50-word set, noise type and S/N ratio were adjusted according to the order determined through the counterbalancing procedure, and subjects were allowed to rest. After the 12 treatment conditions had been completed, a final trial with no noise (i.e., a "quiet" condition), which provided the same environment as for the practice trial, was administered.

After the experimental trials, a post-experiment pure-tone audiogram was performed at 1000, 2000, 3000, and 4000 Hz for comparison to the screening before the trials. If no TTS was found, the subject was paid for participation, thanked, and dismissed. The procedure in the case of a TTS, which was not used in the experiment because none was found for any subject, would have to been to measure the subject's hearing at one-hour intervals until hearing had returned to normal.

EMPIRICAL RESULTS

Noise Conditions

Data reduction. To compute the dependent measure of percent intelligibility, it was necessary to convert the scores from the raw observations according to ANSI S3.2-1989. Using the formula, $R_a = R - [W/(n-1)]$, where R_a is the number of items correct adjusted for chance/guessing, R is the number of items correct, W is the number of items incorrect, and n is the number of alternative choices per item (six in this case), the raw score was adjusted for the probability that some of the correct answers would result from chance or guessing due to the closed-choice format. Each of the 50-word raw scores was then divided by 50 and changed to a single percent intelligibility score.

Assumptions of the ANOVA. An analysis of the distributions for each treatment condition using the Kolmogorov-Smirnov goodness-of-fit test (Siegel and Castellan, 1988), which measures the maximum difference of the cumulative relative frequency between the observed and the normal distribution for each condition, showed that none of the distributions deviated significantly from the normal distribution at $p < 0.05$ (see Table 10). In addition, using Cochran's test for homogeneity of variance (Winer, 1971), which divides the largest distribution variance by the sum of variances for all conditions, variances of the treatment conditions were not significantly different at $p < 0.05$ (see Table 11). These results indicate that the underlying assumptions of the ANOVA requiring a normal sampling distribution and homogeneity of variance were not violated.

Because the experimental design contained within-subjects variables, the ANOVA incurred the additional assumption of sphericity or homogeneity of covariance among the repeated measures. Therefore, Greenhouse-Geisser ε corrections (Vasey and Thayer, 1987) were used to determine the adjusted degrees of freedom to protect against violations of the sphericity assumption, which can induce a positive bias in the ANOVA.

TABLE 10

Deviations from the Normal Distribution for the Kolmogorov-Smirnov Goodness-of-Fit

	Hearing-Impaired		Normal-Hearing	
S/N Ratio	Noise Trial	Difference	Noise Trial	Difference
15 dB	Pink	0.1458	Pink	0.1976
	Truck	0.1667	Truck	0.1993
	Speech	0.1471	Speech	0.1592
10 dB	Pink	0.2460	Pink	0.2426
	Truck	0.2499	Truck	0.1407
	Speech	0.1606	Speech	0.1219
5 dB	Pink	0.1702	Pink	0.1964
	Truck	0.1985	Truck	0.1609
	Speech	0.1298	Speech	0.2657
0 dB	Pink	0.2094	Pink	0.1065
	Truck	0.1781	Truck	0.1734
	Speech	0.2083	Speech	0.2227
Quiet	Practice	0.1141	Practice	0.1669
	Final	0.1095	Final	0.1657

*Critical difference at $p < 0.05$ is 0.375

TABLE 11

Treatment Condition Variances for Cochran's Test for Homogeneity of Variance

	Hearing-Impaired		Normal-Hearing	
S/N Ratio	Noise Trial	Variance	Noise Trial	Variance
15 dB	Pink	82.56	Pink	57.60
	Truck	89.50	Truck	149.06
	Speech	89.89	Speech	104.73
10 dB	Pink	118.17	Pink	38.23
	Truck	109.05	Truck	101.72
	Speech	123.40	Speech	143.48
5 dB	Pink	145.40	Pink	110.31
	Truck	158.49	Truck	72.74
	Speech	66.85	Speech	22.87
0 dB	Pink	48.52	Pink	107.30
	Truck	48.65	Truck	71.56
	Speech	46.35	Speech	87.31
Sum of Variances		2193.74	C = 158.49/2193.74 = 0.072	

*Critical value at $p < 0.05$ is 0.1108 for 20 variances and a df of 16

The effect of this correction is an increase in the p-values of the within-subjects sources in the ANOVA.

Overall analysis. Mean percent intelligibility scores for each treatment combination are shown in Table 12. Intelligibility scores were first subjected to a three-way, mixed-factor ANOVA. All main effects and interactions of hearing level, noise type, and S/N ratio were included as sources of variance. In the F-ratio for noise type and S/N ratio and for interactions containing these main effects, the interaction of source with subjects served as the denominator to determine significant effects; in the F-ratio for hearing level, subjects within hearing level served as the denominator. Analysis was accomplished using the SuperANOVA Macintosh statistical analysis application and confirmed by hand calculation. Using the Greenhouse-Geisser corrections where appropriate, the main effects of Hearing Level (HL) $\{F(1,22) = 20.07, p = 0.0002\}$, Noise Type (NT) $\{F(2,44) = 44.66, p < 0.0001\}$, and S/N Ratio (SN) $\{F(3,66) = 449.97, p < 0.0001\}$, and the interaction of Noise Type-by-S/N Ratio $\{F(6,132) = 2.421, p = 0.0469\}$ were significant using a $p < 0.05$ criterion. No other interaction was significant at $p < 0.05$. A complete ANOVA summary table appears in Table 13.

Main effects. The significant main effect of S/N Ratio was evaluated using a Newman-Keuls test, and the results are shown in Table 14. Analysis shows that all levels of S/N Ratio were significantly different from each other, with a steady increase in intelligibility as the S/N ratio increased. Figure 27 shows the mean intelligibility for each S/N ratio. A linear growth in intelligibility ($r = 0.999$) as S/N ratio increases is exhibited, indicating that a ceiling effect did not occur, at least within the range of the levels tested.

Since Hearing Level had but two levels, no post-hoc analysis was necessary. The mean intelligibility for the normal hearing group across all experimental trials was 42.7%, while the corresponding intelligibility score for the hearing-impaired group was 32.1%.

TABLE 12

Mean Percent Intelligibility Scores for the 12 Experimental Treatment Combinations

	Hearing-Impaired		Normal-Hearing	
S/N Ratio	Noise Trial	% Intelligibility	Noise Trial	% Intelligibility
15 dB	Pink	56.4	Pink	64.0
	Truck	59.0	Truck	72.0
	Speech	39.6	Speech	59.2
10 dB	Pink	44.4	Pink	50.8
	Truck	44.2	Truck	55.0
	Speech	23.6	Speech	42.4
5 dB	Pink	30.0	Pink	38.4
	Truck	28.8	Truck	38.6
	Speech	20.4	Speech	30.8
0 dB	Pink	15.2	Pink	22.2
	Truck	16.2	Truck	24.8
	Speech	7.7	Speech	14.7

TABLE 13

ANOVA Summary Table for the 12 Experimental Treatment Combinations

Source	df	MS	*F*	*p*	*G-G ε*	*G-G p*
Between-Subjects						
Hearing Level (HL)	1	8106.89	20.071	0.0002		
Subjects (S/HL)	22	403.91				
Within-Subjects						
S/N Ratio (SN)	3	22531.47	449.971	< 0.0001	0.902	< 0.0001
SN x HL	3	122.87	2.454	0.0709		0.0779
SN x S/HL	66	50.07				
Noise Type (NT)	2	4306.22	44.659	< 0.0001	0.883	< 0.0001
NT x HL	2	260.11	2.698	0.0785		0.0861
NT x S/HL	44	96.43				
SN x NT	6	141.21	2.421	0.0298	0.746	0.0469
SN x NT x HL	6	68.47	1.174	0.3240		0.3274
SN x NT x S/HL	132	58.31				
Total	287					

TABLE 14

Newman-Keuls Test Results for the S/N Ratio Main Effect

	0 dB	5 dB	10 dB	15 dB	CD	CD
Means	16.80	31.17	43.40	58.37	at $p < 0.05$	at $p < 0.01$
0 dB	———	14.37*	26.60*	41.57*	3.84	3.12
5 dB		———	12.23*	27.20*	3.57	2.84
10 dB			———	14.97*	3.14	2.36

*Significant at $p < 0.01$

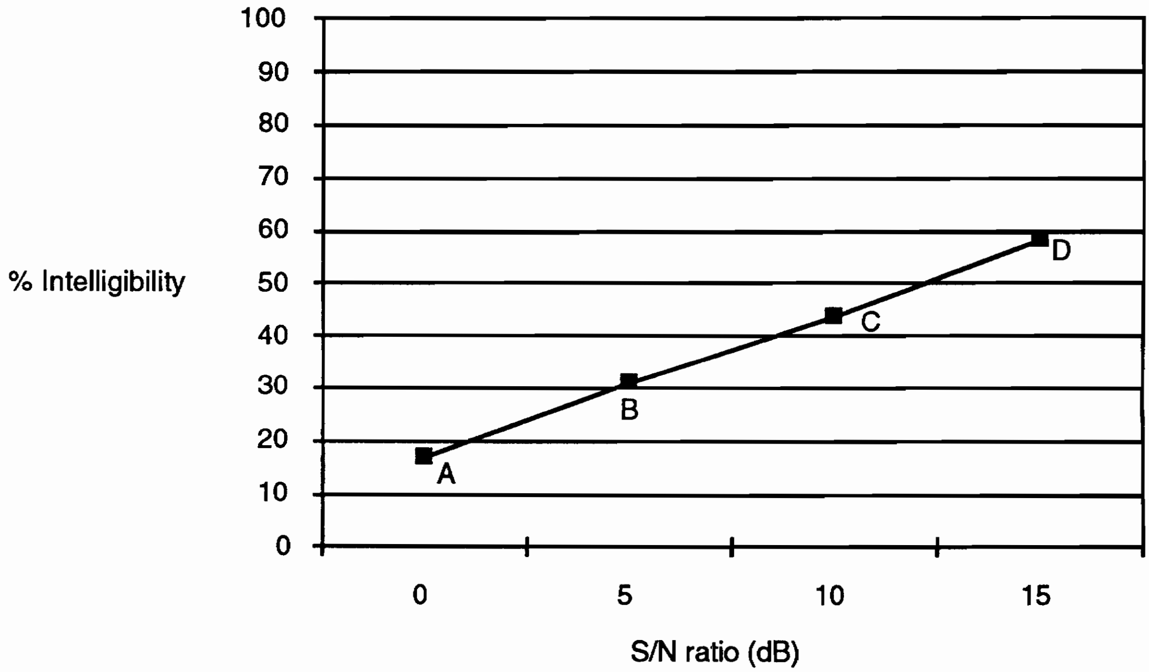

Figure 27. *Percent intelligibility means plotted against S/N ratio (Means denoted by different letters are significantly different at p < 0.01).*

A Newman-Keuls test of Noise Type at $p < 0.05$ showed that the speech noise type accorded significantly poorer intelligibility than either of the other two noise types (see Table 15). Figure 28 depicts the mean intelligibility scores across the three noise types. This result shows that qualitative differences between the natural speech already present and the synthesized speech display systems to be used in truck cab environments may not be enough to make these displays intelligible to the driver without a substantial increase in S/N ratio. The pink noise and the truck noise did not significantly differ from each other, which can be explained by the lack of a substantial intensity difference between the noises in the critical speech bandwidth, as they were reproduced in the experiment. Had a higher fidelity playback of the truck recording in the low frequencies below 125 Hz been possible, larger differences may have occurred.

Noise Type-by-S/N Ratio interaction. Because this was the only significant interaction, it was the only one subjected to subsequent analyses. Post hoc, a simple-effects F-test was first conducted on this interaction. The simple-effects F-test was used because it distributes the alpha error across only one level of a particular variable, while evaluating the other variable at that level to determine significance. All levels of both variables were significantly different at all levels of the other variable. ANOVA summary tables for the simple-effects tests appear in Tables 16 and 17. A Newman-Keuls analysis was then conducted on each simple-effects F-test to identify the source of the interaction, with the results shown in Table 18. From the analysis, the only source identified was a change in significance between the speech noise and the pink and truck noises from 0.01 to 0.05 at the two lower S/N ratios (see Figure 29) and a change from a 0.01 to a 0.05 significance level for the speech noise between the 5 dB and the 10 dB S/N ratios. This is a rather weak interaction, and general data trends are not easily identifiable from it.

TABLE 15

Newman-Keuls Test Results for the Noise Type Main Effect

	Means	Speech 29.80	Pink 40.18	Truck 42.33	CD at $p < 0.05$	CD at $p < 0.01$
Speech		———	10.38*	12.53*	3.45	4.38
Pink			———	2.15	2.87	3.83

*Significant at $p < 0.01$

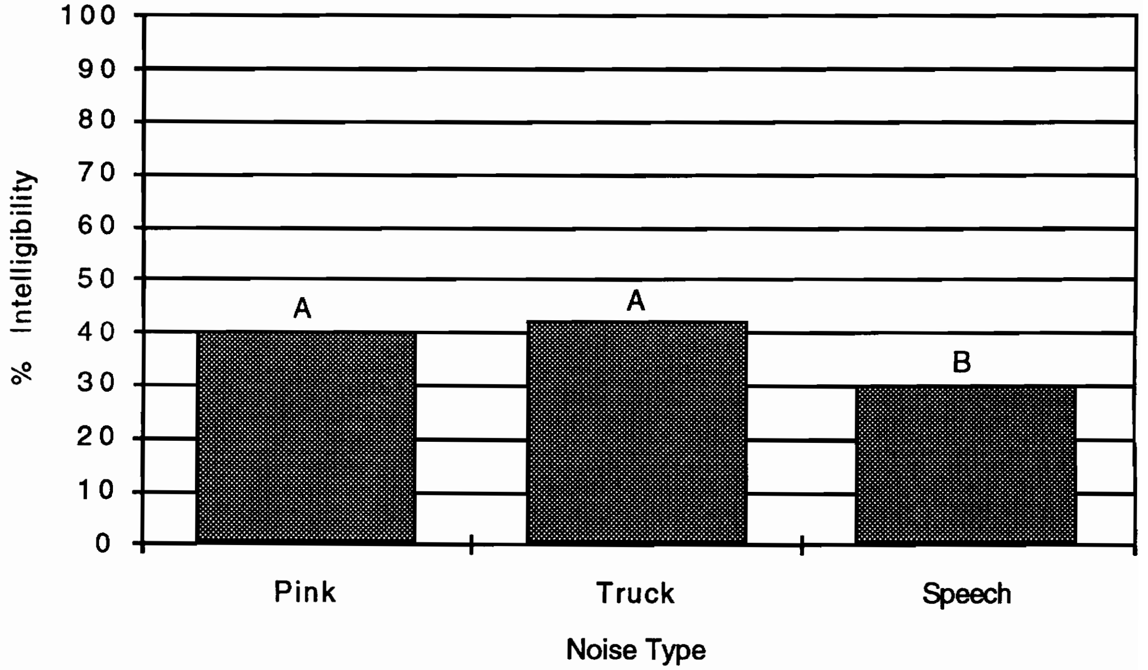

Figure 28. *Percent intelligibility means plotted against noise type (Means denoted by different letters are significantly different at $p < 0.01$).*

TABLE 16

Simple-Effects *F*-Test Results for Noise Type at Each S/N Ratio

S/N Ratio (dB)	MS	*F*	*p*	*G-G p*
15	1615.76	27.71	< 0.0001	< 0.0001
10	1970.88	33.80	< 0.0001	< 0.0001
5	559.28	9.59	0.0001	0.0025
0	583.92	10.01	< 0.0001	0.0021

[1]All *F*-ratios were calculated with MSE = 58.31.

[2]All *p*-values were calculated with numerator df = 2 and denominator df = 132.

[3]All *G-G p*-values were calculated with G-G $\varepsilon = 0.746$.

TABLE 17

Simple-Effects *F*-Test Results for S/N Ratio at Each Noise Type

Noise Type	MS	*F*	*p*	*G-G p*
Pink	7624.06	130.75	< 0.0001	< 0.0001
Truck	9125.82	156.51	< 0.0001	< 0.0001
Speech	6064.00	104.00	< 0.0001	< 0.0001

[1]All *F*-ratios were calculated with MSE = 58.31.

[2]All *p*-values were calculated with numerator df = 3 and denominator df = 132.

[3]All *G-G p*-values were calculated with G-G $\varepsilon = 0.746$.

TABLE 18

Newman-Keuls Test Results for the Simple Effects *F*-Tests

Noise Type at 15 dB S/N Ratio

	Speech	Pink	Truck	CD at $p < 0.05$	CD at $p < 0.01$
Means	49.4	60.2	65.5		
Speech	———	10.8**	16.1**	7.88	9.85
Pink		———	5.3	6.57	8.68

*Significant at $p < 0.05$
**Significant at $p < 0.01$

Noise Type at 10 dB S/N Ratio

	Speech	Pink	Truck	CD at $p < 0.05$	CD at $p < 0.01$
Means	33	47.6	49.6		
Speech	———	14.6**	16.6**	7.88	9.85
Pink		———	2.0	6.57	8.68

*Significant at $p < 0.05$
**Significant at $p < 0.01$

Noise Type at 5 dB S/N Ratio

	Speech	Truck	Pink	CD at $p < 0.05$	CD at $p < 0.01$
Means	25.6	33.7	34.2		
Speech	———	8.1*	8.6*	7.88	9.85
Pink		———	0.5	6.57	8.68

*Significant at $p < 0.05$
**Significant at $p < 0.01$

Noise Type at 0 dB S/N Ratio

	Speech	Pink	Truck	CD at $p < 0.05$	CD at $p < 0.01$
Means	11.2	18.7	20.5		
Speech	———	7.5*	9.3*	7.88	9.85
Pink		———	1.8	6.57	8.68

*Significant at $p < 0.05$
**Significant at $p < 0.01$

TABLE 18 (Continued)

S/N Ratio at Pink Noise

	0 dB	5 dB	10 dB	15 dB	CD	CD
Means	18.7	34.2	47.6	60.2	at $p < 0.05$	at $p < 0.01$
0 dB	———	15.5**	28.9**	41.5**	8.63	10.55
5 dB		———	13.4**	26.0**	7.88	9.85
10 dB			———	12.6**	6.57	8.68

*Significant at $p < 0.05$
**Significant at $p < 0.01$

S/N Ratio at Truck Noise

	0 dB	5 dB	10 dB	15 dB	CD	CD
Means	20.5	33.7	49.6	65.5	at $p < 0.05$	at $p < 0.01$
0 dB	———	13.2**	29.1**	45.0**	8.63	10.55
5 dB		———	15.9**	31.8**	7.88	9.85
10 dB			———	15.9**	6.57	8.68

*Significant at $p < 0.05$
**Significant at $p < 0.01$

S/N Ratio at Speech Noise

	0 dB	5 dB	10 dB	15 dB	CD	CD
Means	11.2	25.6	33.0	49.4	at $p < 0.05$	at $p < 0.01$
0 dB	———	14.4**	21.8**	38.2**	8.63	10.55
5 dB		———	7.4*	23.8**	7.88	9.85
10 dB			———	16.4**	6.57	8.68

*Significant at $p < 0.05$
**Significant at $p < 0.01$

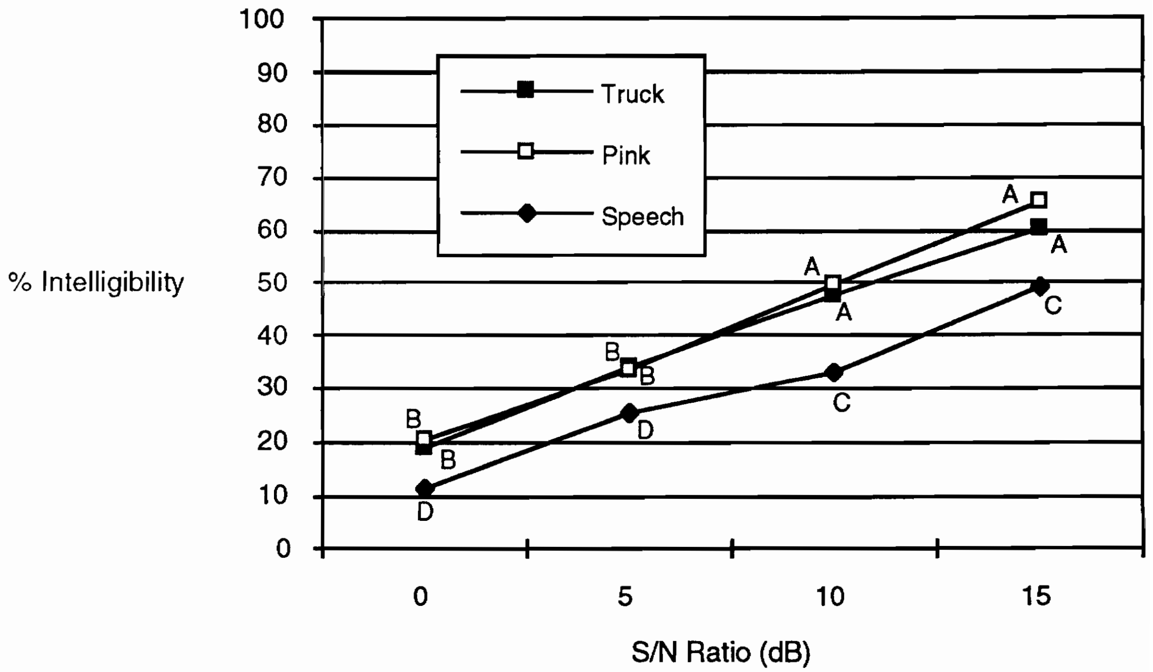

Figure 29. *Noise Type-by-S/N Ratio interaction for mean percent intelligibility (Means denoted by C are significantly different from means denoted by A at $p < 0.01$; means denoted by D are significantly different from means denoted by B at $p < 0.05$).*

Treatment Order

Treatment order and hearing level were analyzed using a two-way ANOVA to determine whether intelligibility increased during the 12 experimental conditions. The ANOVA summary table can be found in Table 19. No effect of order was found $\{F(11,242) = 0.139, p = 0.9866\}$, nor was there an interaction between order and hearing level. The low value of the Greenhouse-Geisser ε indicates that the ordering of the treatment conditions as dictated by the balanced Latin square design caused nonsphericity of the order effect. Because the order of the trials was not significant, it is unlikely that the order of the treatment combinations affected the measured intelligibility scores. In addition, this result seems to indicate that training to listen to synthesized speech in noise during a period as short as the one present in the experiment would provide no measurable improvement in the intelligibility scores.

Trials in Quiet Conditions

Using the data from the practice trial and the final trial, a two-way ANOVA of trials and hearing level was analyzed, with the ANOVA summary table in Table 20. The difference between the practice trial and the final trial was significant $\{F(1,22) = 5.489, p = 0.0286\}$, as was the difference between the hearing-impaired group and the normal-hearing group $\{F(1,22) = 25.010, p < 0.0001\}$, but there was no interaction between trials and hearing level. The Greenhouse-Geisser correction is not necessary in this case because nonsphericity is not possible with only two trials. Mean intelligibility for the hearing-impaired group practice trial was 74.6% compared with 79.4% for the final trial. For the normal-hearing group, the practice trial mean intelligibility was 88.4%, and the final trial mean intelligibility was 93.0%.

TABLE 19

ANOVA Summary Table for Effect of Treatment Order

Source	df	MS	*F*	*p*	*G-G ε*	*G-G p*
<u>Between-Subjects</u>						
Hearing Level (HL)	1	8106.89	20.071	0.0002		
Subjects (S/HL)	22	403.91				
<u>Within-Subjects</u>						
Order (O)	11	52.83	0.139	0.9995	0.460	0.9866
O x HL	11	108.80	0.287	0.9879		0.9295
O x S/HL	242	379.42				
Total	287					

TABLE 20

ANOVA Summary Table for Quiet Conditions

Source	df	MS	*F*	*p*
Between-Subjects				
Hearing Level (HL)	1	2252.28	25.010	< 0.0001
Subjects (S/HL)	22	90.05		
Within-Subjects				
Trial (T)	1	265.08	5.489	0.0286
T x HL	1	0.12	0.002	0.9607
T x S/HL	22	48.29		
Total	47			

Since no effect of treatment order was found, an explanation for these findings could be that subjects were inexperienced with the task during the practice trial, but once it was completed they could perform at a much higher level even with noise and thereafter showed little or no incremental improvement across the experimental trials. This explanation is supported by Kangas and Allen (1990), who found no order effect for the two word lists presented to subjects after the practice trials had been completed. These differences in mean intelligibilities for the two groups also provide more evidence that, even in quiet, hearing-impaired subjects have more trouble identifying synthesized speech words than normal-hearing subjects.

ANALYTICAL RESULTS

The Articulation Index (AI), as described in ANSI standard S3.5-1969 r 1986, is a physical measure that has been shown to be highly correlated with speech intelligibility and is used to predict the relative performance of communication systems. As with ANSI S3.2-1989, the AI cannot be assumed to apply to situations involving synthesized speech. However, because it is accepted as the most accurate analytic predictor of speech intelligibility in noise, the AI was applied to the speech and noise conditions in this experiment. The overall long-term rms for speech was approximated according to ANSI S3.5-1969 r 1986 by measuring the peak dBC level of 20 words, taking the arithmetic average, and then subtracting 3 dB. By plotting the measured octave-band noise levels for each of the noise types and the idealized speech spectrum raised by the difference between the overall long-term rms for speech and 65 dB, the octave-band method was used to calculate the AI for each of the 12 treatment combinations. This was done to assess whether the AI could be applied to intelligibility scores measured for synthesized speech.

The results of the AI calculations, the empirical intelligibility scores for both the hearing-impaired and normal-hearing groups, and the predicted intelligibility scores for rhyme tests (ANSI S3.5-1969 r 1986, Figure 15) are shown in Table 21. The AI calculations were highly correlated with the empirical results for both groups across all noise types ($r = 0.980$ for the normal-hearing group and $r = 0.941$ for the hearing-impaired group), but the empirical results showed substantial differences from intelligibility results for rhyme tests that would be predicted using ANSI S3.5-1969 r 1986, Figure 15. At higher AIs, the empirical results were much lower than predicted; while at lower AIs, the empirical results were higher than predicted. The curves also show a much more linear growth in intelligibility versus AI for the empirical results than

TABLE 21

Articulation Index Values, Intelligibility Scores Predicted from ANSI S3.5-1969 r 1986, and Empirical Intelligibility Scores

Noise Type	Articulation Index	Predicted % Intelligibility	Normal-Hearing % Intelligibility (Empirical)	Hearing-Impaired % Intelligibility (Empirical)
		15 dB S/N Ratio		
Pink	0.48	94.0	64.0	56.4
Truck	0.48	94.0	72.0	59.0
Speech	0.43	91.0	59.2	39.6
		10 dB S/N Ratio		
Pink	0.31	81.0	50.8	44.4
Truck	0.32	82.0	55.0	44.2
Speech	0.26	72.0	42.4	23.6
		5 dB S/N Ratio		
Pink	0.16	40.0	38.4	30.0
Truck	0.15	38.0	38.6	28.8
Speech	0.09	18.0	30.8	20.4
		0 dB S/N Ratio		
Pink	0.07	11.0	22.2	15.2
Truck	0.04	5.0	24.8	16.2
Speech	0.01	1.0	14.7	7.7

the typical ogive curve as represented in ANSI S3.5-1969 r 1986, Figure 15 (the curves for the normal-hearing group are shown in Figure 30; the curves for the hearing-impaired group are shown in Figure 31). Because of this mild slope and linearity, and given the strong correlation between the AI and the empirically-measured scores, the AI seems to be appropriate to use for predicting relative intelligibility of synthesized speech within the range of AIs tested in this experiment. However, the accuracy of the AI in predicting absolute intelligibility with synthesized speech cannot be inferred from this data. The data do not supply sufficient evidence to draw conclusions as to the cause of the differences between the empirical and predicted results.

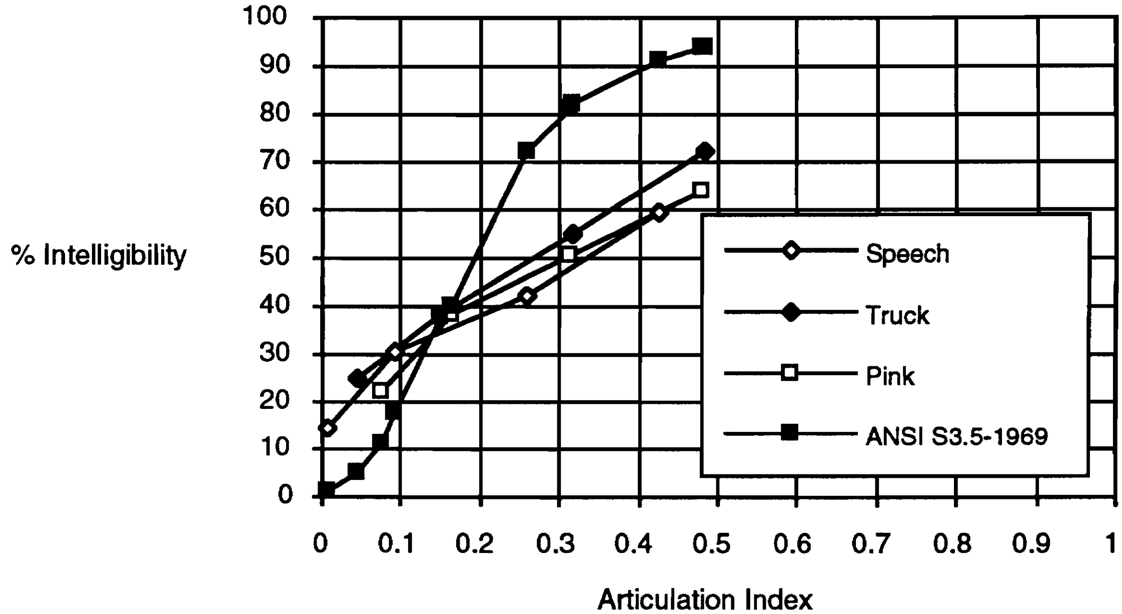

Figure 30. *Functional relationship between percent intelligibility and the Articulation Index for the normal-hearing group plotted with the function for rhyme tests as per ANSI S3.5-1969 r 1986.*

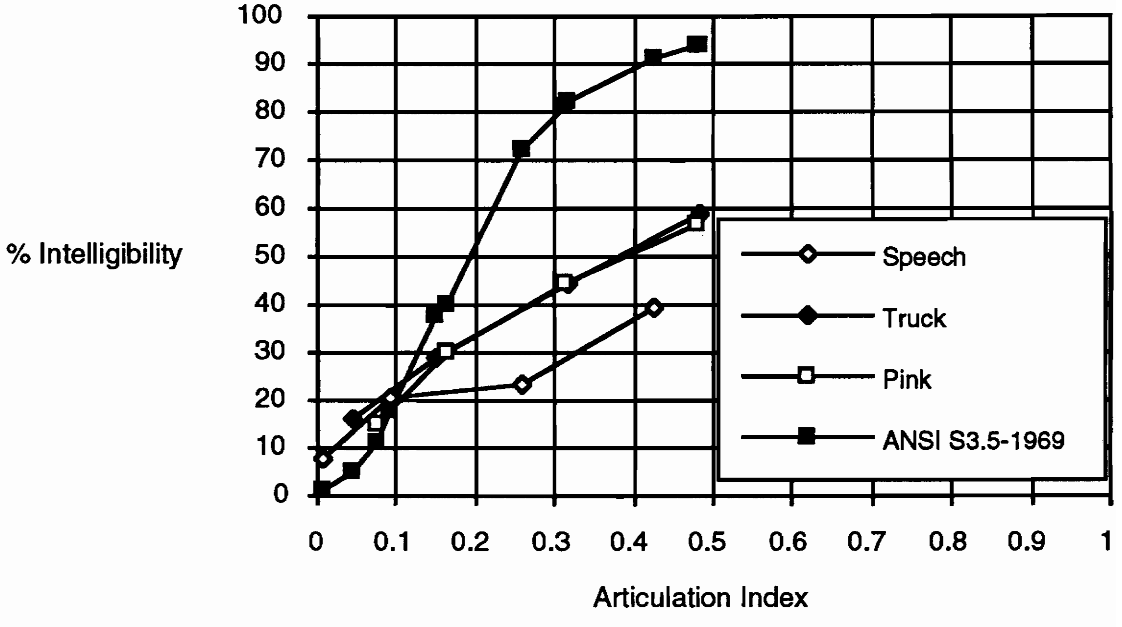

Figure 31. *Functional relationship between percent intelligibility and the Articulation Index for the hearing-impaired group plotted with the function for rhyme tests as per ANSI S3.5-1969 r 1986.*

DISCUSSION

Truck Noise

Several interesting results are apparent from the truck measurements taken. Firstly, the noise levels in the measured truck interiors fell considerably below those of trucks measured in the 1970s, indicating a substantial improvement in noise control and/or truck drivetrain design. In addition, these noise levels, projected out to eight hours, are below exposure levels deemed hazardous by OSHA (1983), and would be less than a 0% dose according to OSHA calculations when the trucks are driven with the windows up and all accessories turned off. Even though the noise levels measured during the acceleration tests were higher, they are still below OSHA action levels and would not be present for more than a small period of the time the driver would be behind the wheel.

Secondly, the octave-band measurements show that a large portion of the sound energy is located in the lower frequencies, as would be expected with slow-revolution diesel power plants. The sound levels in the critical speech bandwidth are much lower than the overall levels and so should not interfere as much with speech as a noise which is flat-by-octaves but equivalent in overall level. However, the reproduction of the truck noise in this experiment did not allow demonstration of this difference. But because of the relatively low overall sound intensities, upward spread of masking should not be a factor in intelligibility.

Synthesized Speech Intelligibility

Synthesized speech was not compared to natural speech because the differences between them are relatively well-established. The fact that all main effects were significant was not an unexpected result considering the literature on the effects of hearing level and S/N ratio on the perception of speech. However, the lack of interaction between hearing level and the other two variables is interesting for a number of reasons. Significant interactions between hearing level and S/N ratio had been found in several experiments involving natural speech, and intelligibility differences between hearing levels were found to be greater at lower S/N ratios (Pekkarinen et al., 1990). The lack of interaction when using synthetic speech implies that some of the auditory cues which help hearing-impaired listeners distinguish natural speech in high S/N ratios are not present. Furthermore, the lack of interaction between hearing level and noise type indicates that normal-hearing subjects have no advantage over listeners with hearing impairment when listening in noise which has background speech. The significant Noise Type-by-S/N Ratio interaction is weak, considering the strong significance of the main effects, and the vague results of the subsequent analysis of the interaction limits the conclusions that can be drawn from it to one of speculation of its direction.

The 93.0% intelligibility for normal-hearing subjects in the final quiet trial was comparable to that found by Logan et al. (1989), who found 96.7% intelligibility using a DECtalk version 1.8. Inspection of graphs in Humes, Nelson, and Pisoni (1991) yields approximately 95% intelligibility for the normal-hearing group and approximately 80% intelligibility for the hearing-impaired group for a closed-response MRT, which are similar to the results of this study (79.4% for the hearing-impaired group in the present study). The intelligibility scores for both groups, even during the practice trial, were much higher than those found by Kangas and Allen (1990), which were 73.0% intelligibility for the normal-hearing group and 43.6% intelligibility for the hearing-impaired group. This may be partially attributable to their open-response format and older subject population (49 to 69 years), who may not be as familiar with new technology such as speech synthesizers. Furthermore, the difference in intelligibility between the groups was much greater in the study by Kangas and Allen (29.4% versus 13.6% for the present study). Again, this may be due to their open-response format, but age should not be a factor in this case because the differences are for groups of comparable age. In addition, the fact that the hearing-impaired in their study were allowed to wear hearing aids should have reduced the differences between the groups. Another study utilizing the open-response MRT found 81.7% intelligibility (Mirenda and Beukelman, 1987), while a study using the DRT (Pratt, 1987) found 88.2% intelligibility, both for normal-hearing groups. These studies tend to suggest that the Kangas and Allen study may have underestimated the intelligibility that is possible using the DECtalk.

Articulation Index

The use of the AI as a tool to predict synthesized speech intelligibility in noise can be valuable to designers of auditory display systems. The high correlations in this study between the AI and the empirical speech intelligibility data suggest that it is appropriate to use the AI for synthesized speech for predicting relative intelligibility in the range of S/N ratios tested. Because the high correlations applied to all noise types, the AI could be used not only for steady state noise such as pink or truck noise, but also for the speech noise which had a steady overall level and unpredictable instantaneous changes in SPL. Furthermore, the AI showed high correlations for listeners with hearing impairment in all noise types, which indicates that the AI could be used relative prediction for these listeners as well. However, the results indicate that the AI may bear a different relationship to synthesized speech intelligibility than to natural human speech, as demonstrated by the data shown in Figures 30 and 31. In addition, the ability to predict absolute intelligibility using the AI cannot be inferred.

CONCLUSIONS

Experimental Conclusions

Based on the results of the truck cab noise measurements, the trucks used in this study do not pose an exposure risk to drivers, as least as long as their drivetrains and other noise-emitting components are maintained as new. While the same may be true for other currently produced trucks, a noise level analysis of each truck cab would be needed to confirm it. This would be very important to producers and operators of long-haul trucks because the noise levels would then be below those allowable by OSHA for employee exposure. Although transportation companies are not now regulated by OSHA, if such measures were implemented in the future, truck manufacturers and trucking companies may already be within allowable levels.

This study provides more evidence that knowledge of the environment, the user, and the delivery system is necessary when designing or deciding to implement the use of an auditory display. Specifically, synthesized speech displays may introduce unique problems when used with populations encompassing a wide variety of hearing levels. A measure of adjustability by the user may overcome some of these problems, but the addition of background speech to the noise environment may change the intelligibility situation. This study also gives some information as to how synthesized speech intelligibility varies with a change in S/N ratio. Although ceiling effects were not found, constant improvement in intelligibility is possible up to at least a 15 dB S/N ratio. This could be an important factor when trading off between intelligibility and noise level.

IVHS Implications

The results of this study yield several implications to be considered when using synthesized speech to present information in noisy environments such as commercial vehicle cabs. Because hearing-impaired listeners have significantly reduced ability to understand synthesized speech and may constitute a substantial portion of the driver population, designers of in-vehicle synthesized speech displays will need to account for the special needs of these operators in the development of IVHS systems. This will be true not only for drivers with noise-induced hearing loss, but also for the elderly population segment which is afflicted with presbycusis and may not be as accustomed as younger drivers to new technologies such as synthesized speech. Furthermore, the addition of chatter from radios or occupants to the noise already present in the car will need to be considered when establishing the appropriate S/N ratio for in-vehicle displays. The significant decrease in intelligibility due to the addition of chatter in this study shows that the auditory display system used may not only have to account for vehicle noise, but may also have to account for instantaneous noise originating from other sources which add to the overall noise level in the vehicle. Because of the lack of differences in performance between the pink noise and truck noise types in this study, noises similar in spectra and level may yield similar intelligibility data. However, the degree of similarity needed and the particular spectra and levels for which this holds is not known. Finally, the Articulation Index may be a useful tool in providing an estimate of the intelligibility that will be afforded the operator before the system is in place, given knowledge about the noise environment and the speech synthesizer.

SUGGESTIONS FOR FUTURE RESEARCH

Due to the relatively recent development of synthesized speech, many areas of research into its use remain. Changes wrought by the advancement of the technology and the introduction of more sophisticated speech synthesizers will have to be investigated. Therefore, the following is a list of topics for possible future study.

- Because this study was limited to the effects of one synthesizer, future research will be needed to analyze the intelligibility differences among newly developed synthesizers and human speech to help pinpoint the phonemes most affected in noisy environments and to determine if there are phoneme differences between hearing-impaired and normal-hearing populations.
- The effects of age on the ability to correctly understand synthesized speech will have to be studied if a substantial portion of the user population will be expected to use it properly.
- Asymptotic ceiling and floor effects in intelligibility were not present in this study, and a wider range of S/N ratios will need to be analyzed to determine them.
- Research into synthesized speech itself will allow exploration into how pitch and speaking rate interact with noise.
- Experimental comparison of MRT results to contextual sentence tasks is needed to more realistically predict the intelligibility that can be expected in some actual IVHS systems that utilize extended speech.
- Systematic comparison of the application of the AI to both human and synthesized speech and to normal-hearing and hearing-impaired listeners will allow a better quantification of these relationships, which could be useful in determining proper utilization of synthesized speech in noisy environments.

REFERENCES

Abel, S. M., Alberti, P. W., Haythornthwaite, C., and Riko, K. (1982). Speech intelligibility in noise with and without hearing protectors. In P. W. Alberti (Ed.), *Personal hearing protection in industry* (pp. 371–385). New York: Raven.

Acton, W. I. (1970). Speech intelligibility in a background noise and noise-induced hearing loss. *Ergonomics, 13* (5), 546–554.

ANSI S3.5-1969 r 1986 (1986). *Method for calculation of the articulation index.* New York: American National Standards Institute.

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

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

ANSI S3.2-1989 (1989). *Method for measuring the intelligibility of speech over communication systems.* New York: American National Standards Institute.

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

Beck, M., Meyer, M., Lewis, S. D., and Hager, M. (1988). Smart cars, smart streets. *Newsweek,* December 5, 86–87.

Berger, E. H. (1980). EARlog #5 - Hearing protector performance: How they work - and - what goes wrong in the real world. *Sound and Vibration, 14* (10), 14–17.

Boff, K. R., and Lincoln, J. E. (1988). *Engineering data compendium: Human perception and performance.* Wright Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Bucher, N. M., Karl, R. L., Voorhees, J. W., and Werner, E. (1984). Alerting prefixes for speech warning messages. In *Proceedings of the National Aerospace & Electronics Conference (NAECON)* (pp. 924–931). New York: IEEE.

Byblow, W. D., and Corlett, J. T. (1989). Effects of linguistic redundancy and coded voice warnings on system response time. *Applied Ergonomics, 20* (2), 105–108.

Carhart, R., and Jerger, J. F. (1959). Preferred methods for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders, 24*, 330345.

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

Casali, J. G. (1988). *A computer-controlled facility for hearing protection research and attenuation testing: verification re ANSI S12.6-1974* (IEOR Technical Report No. 8801). Blacksburg, VA: Virginia Polytechnic Institute and State University.

Casali, J. G. (1989). Multiple factors affect speech communication in the work place. *Occupational Health & Safety, 58* (7), 32–42.

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. 12/1/0-/5-HP). Blacksburg, VA: Virginia Polytechnic Institute and State University.

Clark, J. E. (1983). Intelligibility comparisons for two synthetic and one natural speech source. *Journal of Phonetics, 11* (1), 37–49).

Clarke, R. M. (1978). Stationary locomotive in-cab noise measurements. *Sound and Vibration, 12* (5), 18–23.

Committee on Public Works and Transportation, U.S. House of Representatives (1989). *The status of the nation's highways and bridges: Conditions and performance* (Committee Print 101-2). Washington, DC: U.S. Government Printing Office.

Cudworth, A. L. (1986). Workers' compensation. In E. H. Berger, W. D. Ward, J. C. Morrill, and L. H. Royster (Eds.), *Noise and hearing conservation manual* (4th ed.) (pp. 523–536). Akron, OH: American Industrial Hygiene Association.

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

Department of Transportation (1977). Medical regulatory criteria for evaluation under Section 391.41 (b) (11). Washington, DC: U.S. Department of Transportation.

Earshen, J. J. (1986). Sound Measurement: Instrumentation and Noise Descriptors. In E. H. Berger, W. D. Ward, J. C. Morrill, and L. H. Royster (Eds.), *Noise and hearing conservation manual* (4th ed.) (pp. 37–96). Akron, OH: American Industrial Hygiene Association.

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope, 58* (9), 955-991.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (1954). Some factors affecting multi-channel listening. *Journal of the Acoustical Society of America, 26*, 774-782.

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*, 622–630.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*, 423–425.

Erlichman, J. (1992). A pilot study of the In-Vehicle Safety Advisory and Warning System (IVSAWS) driver-alert warning system design. In *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 480–484). Santa Monica, CA: Human Factors Society.

Fee, J. A., and Schwendeman, M. (1976). *Preliminary analysis of the cab-over-engine vs. the cab-behind-engine designs* (Report No. HS-019 136). Washington, DC: Federal Highway Administration.

Fleischman, R. N., Carpenter, J. T., Dingus, T. A., Szczublewski, F. E., Krage, M. K., and Means, L. G. (1991). Human factors in the TravTek demonstration IVHS project: Getting information to the driver. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 1115–1119). Santa Monica, CA: Human Factors Society.

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.

Hakkinen, M. T., and Williges, B. H. (1984). Synthesized warning messages: Effects of an alerting cue in single- and multiple-function voice synthesis systems. *Human Factors, 26* (2), 185–195.

Hancock, P. A., Caird, J. K., Johnson, S. B., Shekhar, S., Yang, T. A., Coyle, M., and Pawlacyk, L. S. (1991). Human factors safety issues in intelligent vehicle highway systems. *HFS Safety News, 19* (1), 1–9.

Hannley, M., and Dorman, M. F. (1983). Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss. *Journal of the Acoustical Society of America, 74* (1), 40–51.

Hart, S. G., and Simpson, C. A. (1976). Effects of linguistic redundancy on synthesized cockpit warning message comprehension and concurrent time estimation. In *Proceedings of the 12th Annual Conference on Manual Control* (pp. 309–321). Washington, DC: U.S. Government Printing Office.

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.

House, A. S., Williams, C. E., Hecker, M. H., and Kryter, K. D. (1965). Articulation testing methods: Consonantal differences with a closed-response set. *Journal of the Acoustical Society of America, 37* (1), 158–166.

Humes, L. E., Nelson, K. J., and Pisoni, D. B. (1991). Recognition of synthetic speech by hearing-impaired elderly listeners. *Journal of Speech and Hearing Research, 34,* 1180–1184.

Intelligent Vehicle Highway Society of America (1992). *Strategic plan for intelligent vehicle-highway systems in the United States.* Washington, DC: Author.

ISO (1990). *Acoustics–Determination of occupational noise exposure and estimation of noise-induced hearing impairment, ISO 1999.* Geneva, Switzerland: International Organization for Standardization.

Jerger, J., Tillman, T. W., and Peterson, J. L. (1960). Masking by octave bands of noise in normal and impaired ears. *Journal of the Acoustical Society of America, 32* (2), 385–390.

Kangas, K. A., and Allen, G. D. (1990). Intelligibility of synthetic speech for normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Disorders, 55,* 751–755.

Kearney, P. W. (1966). *Highway homicide.* New York: Cromwell.

Klein, A. J., Mills, J. H., Adkins, W. Y. (1990). Upward spread of masking, hearing loss, and speech recognition in young and elderly listeners. *Journal of the Acoustical Society of America, 87* (3), 1266–1271.

Kryter, K. D. (1972). Auditory and other sensory forms of information presentation. In H. P. Van Cott and R. G. Kinkade (Eds.), *Human engineering guide to equipment design* (pp. 161–226). Washington, DC: U.S. Government Printing Office.

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

Leneman, G. (1977). *Truck noise 9: Noise reduction study of an in-service diesel-powered truck* (Report No. DOT-TSC-OST-76-3). Washington, DC: Department of Transportation.

Licklider, J. C. R., and Miller, G. A. (1951). The perception of speech. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1040-1074). New York: Wiley.

Logan, J. S., Greene, B. G., and Pisoni, D. B. (1989). Segmental intelligibility of synthetic speech produced by rule. *Journal of the Acoustical Society of America, 86* (2), 566–581.

Luce, P. A., Feustel, T. C., and Pisoni, D. B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors, 25* (1), 17–32.

Lyregaard, P. E. (1982). Frequency selectivity and speech intelligibility in noise. In O. J. Pedersen and T. Poulson (Eds.), *Binaural Effects in Normal and Impaired Hearing* (Scandinavian Audiology, Supplement 15) (pp. 113–122).

Marics, M. A., and Williges, B. H. (1988). The intelligibility of synthesized speech in data inquiry systems. *Human Factors, 30* (6), 719–732.

Martin, A. M., Howell, K., and Lower, M. C. (1976). Hearing protection and communication in noise. In S. D. G. Stephens (Ed.), *Disorders of auditory function* (pp. 47–62). New York: Academic Press.

Mast, T. (1991). Human factors in intelligent vehicle-highway systems: A look at the future. In *Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 1125–1129). Santa Monica, CA: Human Factors Society.

Mast, T. M., and Peters, J. I. (1992). Intelligent Vehicle Highway Systems (IVHS): Human factors issues in subsystem development. In *Proceedings of the Human Factors Society 36th Annual Meeting* (p. 1063). Santa Monica, CA: Human Factors Society.

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.

Metzler, H. G. (1988). *Computer vision applied to vehicle operations* (Technical Paper 881167). Warrendale, PA: Society of Automotive Engineers.

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin, 44*, 105–129.

Miller, G. A., Heise, G. A., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*, 329–335.

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.

Miller, R. (1978). Hearing protection: The state of the art. *National Safety News, 117* (3), 91–92.

Mirenda, P., and Beukelman, D. R. (1987). A comparison of speech synthesis intelligibility with listeners from three age groups. *Augmentative and Alternative Communication*, *3*, 120–128.

Morans, R. W., Kamal, M. M., Okamoto, H., Hase, M., and Hellaker, J. (1991). IVHS. *Automotive Engineering*, *99* (3), 13–20.

Morrill, (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. 233–292). Akron, OH: American Industrial Hygiene Association.

Nerbonne, M. A., and Accardi, A. E. (1975). Noise-induced hearing loss in a truck driver population. *The Journal of Auditory Research*, *15*, 119–122.

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

OSHA (1970). *Occupational safety and health act of 1970: Noise provisions.* Philadelphia, PA: U.S. Department of Labor.

OSHA (1983). Occupational noise exposure; Hearing conservation amendment; Final rule. In *Federal Register 48,* (pp. 9738–9785). Washington, DC: U.S. Department of Labor.

Oshrin, S. E., and Siders, J. A. (1984). Discrimination of computer synthesized speech. *Perceptual and Motor Skills, 59* (2), 619–622.

Parkes, A. M., and Coleman, N. (1990). Route guidance systems: A comparison of methods of presenting directional information to the driver. In E. J. Lovesey (Ed.), *Contemporary ergonomics 1990* (pp.480–485). London: Taylor and Francis.

Parizet, E. (1992). The influence of speech importance function upon articulation index computation in cars. *Noise Control Engineering Journal, 38* (2), 73–79.

Pekkarinen, E., Salmivalli, A., and Suonpää, J. (1990). Effect of noise on word discrimination by subjects with impaired hearing, compared with those of normal hearing. *Scandinavian Audiology, 19*, 31–36.

Peterson, A., and Gross, E., Jr. (1978). *Handbook of noise measurement* (8th ed.). New Concord, MA: General Radio.

Pisoni, D. B., Greene, B. G., and Nusbaum, H. C. (1985). Some human factors issues in the perception of synthetic speech. *Speech Tech '85*, 57–61.

Pisoni, D. B., and Koen, E. (1982). Intelligibility of natural and synthetic speech at several different signal-to-noise ratios. *Journal of the Acoustical Society of America, 71*, UU1.

Pisoni, D. B., and Nusbaum, H. C. (1986). *Developing methods for assessing the performance of speech synthesis and recognition systems* (Research on Speech Perception, Progress Report No. 12, pp. 379–388). Bloomington, IN: Indiana University.

Pollack, I. (1958). Speech intelligibility at high noise levels: Effect of short-term exposure. *Journal of the Acoustical Society of America, 30* (4), 282–285.

Pollack, I., and Pickett, J. M. (1958). Masking of speech by noise at high sound levels. *Journal of the Acoustical Society of America, 30* (2), 127–130.

Pratt, R. L. (1987). Quantifying the performance of text-to-speech synthesizers. *Speech Technology, 5*, 54–63.

Reed, D. (1989). Intelligent vehicles. *Automotive Engineering, 97* (3), 19.

Reed, D. (1990). National transportation policy. *Automotive Engineering, 98* (5), 27.

Reif, Z. F. and Moore, T. N. (1983). *Prediction of in-cab noise exposure of drivers* (Technical Paper 831028). Warrendale, PA: Society of Automotive Engineers.

Reif, Z. F., Moore, T. N., and Steevensz, A. E. (1980). *Noise exposure of truck drivers* (Technical Paper 800278). Warrendale, PA: Society of Automotive Engineers.

Rumar, K. (1982). The human factor in road safety. *Australian Road Research Board Proceedings, 2*, 66–80.

Schwab, E. C., Nusbaum, H. C., and Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27* (4), 395–408.

Siegel, S., and Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.

Shipley, L. B. (1985). Hearing loss comp becoming today's major industry issue. *National Safety and Health News*, 35–38.

Simpson, C. A. (1976). Effects of linguistic redundancy on pilot's comprehension of synthesized speech. In *Proceedings of the 12th Annual Conference on Manual Control* (pp. 294–308). Washington, DC: U.S. Government Printing Office.

Simpson, C. A., and Marchionda-Frost, K. (1984). Synthesized speech rate and pitch effects on intelligibility of warning messages for pilots. *Human Factors, 26* (5), 509–517.

Simpson, C. A., and Navarro, T. N. (1984). Intelligibility of computer generated speech as a function of multiple factors. In *Proceedings of the National Aerospace and Electronics Conference* (84CH1984 NAECON) (pp. 932–940). New York: IEEE.

Simpson, C. A., and Williams, D. H. (1980). Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings. *Human Factors, 22* (3), 319–330.

Slowiaczek, L. M., and Nusbaum, H. C. (1985). Effects of speech rate and pitch contour on the perception of synthetic speech. *Human Factors, 27* (6), 701–712.

Smoorenburg, G. F., de Laat, J. A. P. M., and Plomp, R. (1982). The effect of noise-induced hearing loss on the intelligibility of speech in noise. In H. M. Borchgrevink (Ed.), *Hearing and Hearing Prophylaxis* (Scandinavian Audiology, Supplement 16) (pp. 123–133).

Society of Automotive Engineers (1968). *Sound level for truck cab interior* (Report No. SAE J336). New York: Society of Automotive Engineers.

Suter, A. H. (1986). Hearing conservation. 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.

Tilley, D. H., Erwin, C. W., and Gianturco, D. T. (1973). *Drowsiness and driving: Preliminary report of a population survey* (Report No. 730121). Detroit, MI: Society of Automotive Engineers.

U.S. Department of Labor (1969). Occupational noise exposure. In *Federal Register 34,* (pp. 7946–7949). Washington, DC: U.S. Department of Labor.

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.

Voiers, W. D. (1977). Diagnostic evaluation of speech intelligibility. In M. E. Hawley (Ed.), *Speech intelligibility and speaker recognition* (pp. 374–387). Stroudsburg, PA: Dowden, Hutchinson and Ross.

Voorst, B. V. (1992). Why America has so many potholes. *Time, 139* (18), 64–65.

Ward, W. D. (1973). Adaptation and fatigue. In J. Jerger (Ed.), *Modern developments in audiology* (pp. 301–344). New York: Academic Press.

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

Werner, V. A., and Boyce, W. (1976). *Truck noise 11: Evaluation and reduction of heavy-duty truck noise* (Report No. DOT-TSC-OST-76-21). Washington, DC: Department of Transportation.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp.239–257). Hillsdale, NJ: Erlbaum.

Willis, D. K. (1990). IVHS technologies: Promising palliatives or popular poppycock? *Transportation Quarterly, 44* (1), 73–84.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York, NY: McGraw-Hill.

APPENDIX A

SUBJECT DATA

TABLE A-1

Subject Data for Normal-Hearing Group Concerning Pure-Tone Decibel Hearing Level (dBHL), Age, and Gender

		Pure-Tone Frequency (Hz)										
Subject	Ear	125	250	500	1000	2000	3000	4000	6000	8000	Age	Gender
1	R	5	5	15	20	-5	5	0	10	0	21	M
	L	0	5	20	30	5	15	15	10	15		
2	R	5	0	-5	0	-5	-5	-5	-5	0	22	M
	L	5	5	-5	-5	-10	-5	-5	0	5		
3	R	50	40	40	20	5	5	0	5	10	26	F
	L	25	25	20	15	5	5	-5	20	10		
4	R	5	5	0	0	-5	0	-5	10	10	25	M
	L	10	10	0	0	-5	0	-5	15	10		
5	R	10	5	0	0	0	-5	0	10	-5	26	M
	L	5	5	0	0	-5	-5	10	10	-5		
6	R	5	5	5	5	0	0	-5	0	0	26	M
	L	10	5	5	5	0	-5	-5	10	0		
7	R	10	20	10	5	5	0	0	5	5	27	F
	L	20	15	15	5	10	0	5	20	0		
8	R	0	5	0	0	-5	0	0	-5	-10	24	M
	L	5	5	0	5	0	0	5	5	-10		
9	R	15	20	15	10	5	10	5	10	15	48	F
	L	15	10	10	5	5	5	5	10	10		
10	R	15	10	5	0	5	0	5	10	5	35	F
	L	5	5	0	0	5	0	-5	0	15		
11	R	5	5	0	10	5	5	10	10	20	41	F
	L	10	10	5	0	0	5	5	15	5		
12	R	5	5	5	-5	5	15	15	25	15	59	M
	L	20	15	10	-5	0	15	15	50	35		

TABLE A-2

Subject Data for Hearing-Impaired Group Concerning Pure-Tone Decibel Hearing Level (dBHL), Age, and Gender

		Pure-Tone Frequency (Hz)										
Subject	Ear	125	250	500	1000	2000	3000	4000	6000	8000	Age	Gender
13	R	35	30	30	20	20	35	55	65	75	20	F
	L	40	30	30	25	20	35	60	65	80		
14	R	15	10	10	0	15	45	40	35	20	21	F
	L	10	15	5	-5	20	50	55	35	15		
15	R	20	25	30	35	45	40	45	45	40	21	F
	L	10	20	30	35	45	40	45	40	40		
16	R	15	10	5	0	55	75	90	85	85	59	M
	L	15	15	10	5	25	75	80	95	90		
17	R	35	35	45	55	45	40	40	50	55	62	F
	L	60	65	55	60	60	55	65	75	80		
18	R	20	20	15	25	45	50	55	45	60	20	M
	L	15	10	5	20	5	40	45	50	60		
19	R	10	10	10	10	55	55	60	45	45	37	M
	L	5	15	10	10	65	70	70	65	65		
20	R	10	10	15	10	20	45	60	55	60	31	M
	L	10	10	25	25	45	50	55	60	65		
21	R	30	30	30	30	30	40	25	25	30	23	F
	L	30	40	45	50	50	50	45	50	50		
22	R	15	10	15	15	10	40	80	75	65	41	M
	L	5	5	5	5	5	30	60	70	60		
23	R	20	10	15	20	40	50	70	55	70	63	M
	L	15	10	10	15	40	40	35	55	70		
24	R	30	25	30	25	25	65	75	80	75	52	M
	L	20	15	15	15	15	75	80	90	90		

APPENDIX B

MODIFIED RHYME TEST WORD LIST

went sent bent dent tent rent	not tot got pot hot lot	peel reel feel eel keel heel	mass math map mat man mad
hold cold told fold sold gold	vest test rest best west nest	hark dark mark bark park lark	ray raze rate rave rake race
pat pad pan path pack pass	pig pill pin pip pit pick	heave hear heat heal heap heath	save same sale sane sake safe
lane lay late lake lace lame	back bath bad bass bat ban	cup cut cud cuff cuss cub	fill kill will hill till bill
kit bit fit hit wit sit	way may say pay day gay	thaw law raw paw jaw saw	sill sick sip sing sit sin
must bust gust rust dust just	pig big dig wig rig fig	pen hen men then den ten	bale gale sale tale pale male
teak team teal teach tear tease	pale pace page pane pay pave	puff puck pub pus pup pun	wick sick kick lick pick tick
din dill dim dig dip did	cane case cape cake came cave	bean beach beat beak bead beam	peace peas peak peach peat peal
bed led fed red wed shed	shop mop cop top hop pop	heat neat feat seat meat beat	bun bus but bug buck buff
pin sin tin fin din win	coil oil soil toil boil foil	dip sip hip tip lip rip	sag sat sass sack sad sap
dug dung duck dud dub dun	tan tang tap tack tam tab	kill kin kit kick king kid	fun sun bun gun run nun
sum sun sung sup sub sud	fit fib fizz fill fig fin	hang sang bang rang fang gang	
seep seen seethe seek seem seed	same name game tame came fame	took cook look hook shook book	

APPENDIX C

DESCRIPTION OF EXPERIMENT

Description of the Synthesized Speech Experiment
Written Instructions to the Subject Participant

This experiment is intended to determine how well people understand synthesized speech in different noise environments. If you become a subject in the experiment, you will be asked to participate in one experimental session to be scheduled immediately following the screening session. The entire screening session will last approximately one-half hour. The experimental session will last approximately one and one-half hours. Please feel free to ask questions at any time.

The screening session will consist of your being presented with an informed consent form, filling out a questionnaire concerning your hearing, having your outer ears visually examined, and undergoing the administration of a hearing test. If you qualify as a subject, an experimental session will be scheduled. At the beginning of the experimental session, the experimenter will show you the experimental apparatus, demonstrate the noises and speech to be used in the experiment, and explain the experimental procedures.

The experimental session will consist of a short hearing test, a training session and then the actual experimental session. The hearing test will be conducted in the same room which was used for the screening. After the hearing test, you will be escorted to the test room, and the experimental procedures will be explained. You will listen to words spoken by a speech synthesis device in a noisy background. The three types of noise backgrounds that you will be experiencing will be demonstrated for 20 seconds each. The training session will then begin. You will hear 50 words, each preceded by the phrase "The word is", spoken in a noisy background. Your task is to indicate, by pressing the appropriate Macintosh screen button, which word you heard. You may respond any time after the word is spoken. After all 50 words have been spoken, you will be allowed to rest while your responses are graded. If you demonstrate understanding of the test procedure, you will then listen to 12 sets of 50 word blocks. The test procedure will be exactly the same for all sets and will be exactly like the training session. After each set of 50 words, you will be given the opportunity to take a break. Once the last set of words has been completed, a hearing test performed exactly like the one at the start of the experiment will be conducted.

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

____________________________ ____________________________

Subject's Printed Name Subject's Signature

APPENDIX D

SUBJECT'S INFORMED CONSENT

Subject's Informed Consent
Auditory Systems Laboratory - Virginia Tech
(Synthesized Speech Experiment)

This experiment involves research that is intended to determine how well people can understand synthesized speech in different noise environments. If you become a subject in the experiment, you will be asked to participate in 1 experimental session to be scheduled immediately follow this screening session. The screening session will last approximately one-half hour. The experimental session will last approximately one and one-half hours.

Your hearing ability is an important factor in this experiment. The purpose of the screening session is to determine whether your hearing ability today qualifies you to participate in the experimental session. If you qualify, you will be asked to schedule a time to participate in the experimental session. You will not be paid for the screening session.

First, your right and left ear 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 the 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 to words spoken by a speech synthesis device in noisy backgrounds. Again, you will have to be very attentive and listen carefully to the words. When a word is spoken, you will indicate the word you heard by pressing 1 of 6 buttons. Some words may be very easy to hear while others may be difficult to hear.

The noise and speech levels may seem loud and may possibly cause some discomfort. However, the noise exposure is at least 25% less than the duration allowed by the Occupational Safety and Health Administration (OSHA) over an 8 hour day for workers in industry. You will only undergo this noise for a short time during one day; the OSHA regulations apply to a 5 day per week, 8 hour per day exposure. Therefore, there is minimal risk to your hearing. In addition, an audiogram will be administered before and after the experimental session to assure that no temporary threshold shift has occurred. If a temporary threshold shift has occurred, you will be asked to stay until the experimenter can verify that your hearing has returned to normal. You may experience some fatigue due to the length of the experimental session.

There is no risk to your well-being posed by these audiometric screening (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 test will be conducted in a soundproof 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).

The scientific benefit of this research will be to allow the estimation of the intelligibility of synthesized speech in truck cab noise. This will be important in the design of speech warning and message displays to be used in truck cabs. You will be paid $5.00 per hour to participate in the experimental session.

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 six months hence, include your address as it will be six months from now, with your signature on the last page of this form. If, after receiving the summary, you would 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 the two signature pages of 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 on the next two pages.

Participant's Signature Page
(Participant Copy)

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 the time that I spend in the study.

Signature ______________________________________

Printed Name ______________________________________

Date ______________________________________

Witness ______________________________________

Printed Name ______________________________________

Date ______________________________________

The research team for this experiment consists of Boyd Morrison, a Master's 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 E

SCREENING QUESTIONNAIRE

SCREENING QUESTIONNAIRE

Name: ______________________ Age: _____ Phone: __________

Native English speaker? ______________________________________

Experience with speech synthesizers? ____________________________

Tinnitus or head noises? ______________________________________

Ear wax history? ____

Otopathological history? ______________________________________

Occupation? ____

Noisy hobbies? ____

HPD experience? ____

OTOSCOPIC DATA:

Occluding ear wax? ____

Ear canal irritation? ____

Unusual canal characteristics? __________________________________

Eardrum perforations? ____

Eardrum scar tissue? ____

Foreign matter? ____

COMMENTS:

____

____

____

____

VITA

Mr. H. Boyd Morrison, born February 17, 1967, received a B.S. in Mechanical Engineering in 1989 from Rice University. Upon completion of the M.S. in Industrial and Systems Engineering (Human Factors option) at Virginia Polytechnic Institute and State University, he intends to pursue a Ph.D. in the same program.

Before entering Virginia Tech, he was a human factors engineer for two years for Lockheed Engineering and Sciences Company working on NASA's Space Station Freedom project, concentrating on the human factors aspects of maintenance, assembly operations, and window viewing. As a graduate student, he was a teaching assistant for Introduction to Human Factors, worked as a consultant for Hay Systems on several projects for the U.S. Army, was awarded a graduate fellowship for Intelligent Vehicle Highway System research by General Motors, and developed a simulation of a shipboard radar detection task for the U.S. Navy. His research interests include transportation human factors, auditory systems, and visual displays.

He is a student member of the American Society of Mechanical Engineers and the Human Factors and Ergonomics Society, and he is a full member of Alpha Pi Mu, Phi Kappa Phi, and Toastmasters. His outside interests include alpine skiing, scuba diving, softball, Star Trek, and his dog and cat.

H. Boyd Morrison