Speech Intelligibility in Adverse Conditions in Recorded Virtual Auditory Environments

LTC Nancy L. Vause, Ph.D.

U.S. Army Research Laboratory,
Human Research and Engineering Directorate
Aberdeen Proving Ground, Maryland 21005
410-278-5999
nvause@arl.mil

D. Wesley Grantham, Ph.D.

Hearing and Speech Sciences, Vanderbilt University
Nashville, Tennessee
615-340-8278
d.wesley.grantham@vanderbilt.edu

Abstract

The purpose of this investigation was to examine the effects of presentation mode on speech intelligibility in adverse listening conditions as signal-to-noise ratio was systematically varied in anechoic and reverberant environments. Speech intelligibility scores were obtained from 21 normally hearing listeners using a nonsense syllable test. The syllables were recorded in three environments (mono anechoic, spatial anechoic and spatial reverberant) at three SNR (0, 5, and 9dB) using two simultaneous interfering sound sources. The findings indicate (a) percent correct performance was about 40% lower with the traditional diotic presentation compared to a virtual presentation; (b) performance in the virtual reverberant was about 5% lower than in the virtual anechoic environment.

1 Introduction

Spatial hearing is undoubtedly one of the most complex of all biological abilities. Understanding the processes underlying spatial hearing presents a major challenge to psychoacoustic research. Interest is becoming widespread, because this ability is critical in the applied and basic research of various disciplines such as communications, entertainment, architectural design, rehabilitation, and safety. This study will focus on one important aspect of spatial hearing as it relates to safety and communication. Specifically, the experiment is concerned with speech recognition in clinical and virtual adverse conditions.

Understanding speech in adverse conditions is an extremely important and challenging task for the human auditory system. During daily conversations, most people possess the ability to “tune out” interfering noises that emanate from various directions, focusing instead on signals of interest. When adverse conditions disrupt speech perception, miscommunication is usually a temporary, albeit annoying, inconvenience because most conversations offer ample opportunity to repeat words or phrases not initially understood. However, the opportunity to clarify disappears in some occupations such as the military, surveillance, telecommunication, entertainment, and security where communication is conducted in unique and demanding circumstances.

In these occupational settings, the environment imposes severe challenges to the accurate understanding of speech. Civilian industrial or military personnel must routinely accomplish demanding missions relying heavily on concise clear communication in adverse conditions [1-3]. For example, military personnel must simultaneously monitor several radio channels or decipher speech in the presence of competing talkers and/or background noise. Soldiers often conduct these listening tasks in the reverberant surroundings of tanks, tactical operation environments, communication centers, intelligence support vehicles, and so forth. Unfortunately, research that examines communication in real-world complex environments is extremely limited. Furthermore, qualifying auditory examinations for personnel are usually conducted in artificial environments and may not evaluate real-world communication abilities important for safe and effective job performance.

Traditionally, many clinical and psychoacoustic experiments are conducted under headphones in sound-treated or anechoic chambers often using stimuli such as pure tones, bands of noise, or synthetic speech, which do not occur in nature. While these environments and stimuli are more easily controlled and provide a baseline for optimum performance, they can only approximate real-world performance. Auditory images produced by traditional recordings and headphone presentation do not contain the natural ear-specific auditory cues produced with a free-field presentation. Consequently, the former stimuli are perceived as originating from inside the head halfway between the two ears. Yet, headphone speech test results are often taken as an accurate portrayal of real-world performance.

The real-world or free-field acoustic characteristics of a sound measured at the eardrum depend on the source itself and the pathway the sound must travel to the ear. For example, the distance of the path will affect the phase of the signal in relation to the source. Additionally, the head position in the pathway will result in head-shadow and change the intensity of the signal relative to the intensity of the source. Furthermore, if a listener is wearing safety equipment or protective headgear the signal acoustics are further modified. Moreover, the external ear's filtering effects also influence the intensity and phase of the signal relative to the sound source in a very individualized manner. The auditory input, processing, and subsequent perceptual performance are further complicated by adverse environments such as background noise, multiple competing talkers, or reverberation. The outer ear and ear canal behave as a filter and thus alter the frequently composition of the signal. These complex spectral auditory cues are thought to be responsible for externalization of sound images.

Signal processing technology now allows reproduction of the complex spectral effects of the path from a sound source to the eardrum. In fact, digital signal processing can employ filters to reproduce the same spectral response that an individual would receive at his or her two eardrums. These techniques can achieve a three-dimensional (3-D) quality of sound under headphones and simulate realistic virtual spatial locations and listening environments [4–8]. When filtered appropriately, sounds are heard externalized in headphone presentation. In other words, listening through traditional headphones gives the listener the illusion that stimuli emanate from external locations as in a natural listening environment. Auditory images presented in this manner are commonly referred to 3-D audio or virtual auditory displays.

A myriad of areas could benefit from research on auditory spatial processing and speech communication in “real-world," often adverse environments. Specific areas include hearing aid design and use [9-12], aircraft or ship positioning and diver navigation [13,14], cockpit virtual target acquisition displays [2, 15, 16], navigational aids for the blind [17], entertainment [18, 19], military communication systems [20], military operations using headgear and hearing protection [21, 22], and clinical diagnosis [4, 23].

The purpose of the present research was to examine the effects of traditional and virtual adverse conditions (noise and reverberation) on speech intelligibility. This study examined the effect of presentation method on speech intelligibility in conditions of noise and reverberation as signal-to-noise ratios were systematically varied. This investigation also evaluated the influence of talker gender on speech intelligibility in anechoic and reverberant environments as a function of signal-to-noise ratio and presentation method. By employing carefully controlled stimuli and environmental factors, the experimenters attempted to bridge the gap between previous laboratory experiments and the practical challenges faced by personnel who must communicate in real-world settings. Additionally, the results of this study question whether existing clinical methods that evaluate speech communication in noise should be replaced by techniques that employ rather than eliminate natural binaural acoustic cues.

2 Review of the Literature

2.1 Monaural Factors Underlying Speech Perception

The need to understand and concentrate on one talker in the presence of competing talkers and/or noise is a familiar occurrence. Defined by Cherry [24] as the “cocktail party effect,” understanding speech in noisy environments depends upon several complex variables. To this date our knowledge about the processes required to accomplish this feat is incomplete [25]. Experiments have determined that speech intelligibility includes both monaural and binaural factors [26-31]. The absolute contribution of monaural cues appears to be greater than that of binaural cues [20]. However, the binaural advantage appears to be especially important in degraded listening conditions [20, 32, 33].

Acoustic parameters, environmental variables, and contextual variables contribute to speech intelligibility in both monaural and binaural listening conditions. These variables include, but are not limited to: the type and position of speech or noise (masker), signal-to-noise ratio (SNR), room reverberation characteristics, and the position of a listener with respect to the talker. Investigations have manipulated factors relevant to monaural listening such as signal characteristics [34-44], masker characteristics [20, 45-60], signal-to-noise ratio [28, 45, 47, 51, 61-66], and redundancy of the speech message [54, 62, 67-74].

2.2 Binaural Factors Underlying Speech Perception

Even though most people hear signals very well with headphones, free-field binaural hearing offers several advantages including localization of signals in space, signal separation, and signal enhancement in noise and reverberation. Binaural hearing allows the ability to selectively listen from one direction and tune out sounds coming from other directions that could interfere (cocktail party effect). Spatial hearing may also offer solutions to communication difficulties experienced by radio operators, tankers, and surveillance personnel. The simultaneous monitoring of several communication channels is usually accomplished using diotic or monotic headphones inside a small mobile communication center. Current military communication equipment designs do not employ stereophonic, dichotic, or virtual listening technologies, although future combat military operational plans include the visual "virtual battlefield."

Another possible contributor to increased speech intelligibility in free-field listening as opposed to monaural headphone listening is the central auditory system's ability to suppress noise internally in some binaural listening conditions based on interaural differences. This ability, called binaural release from masking or masking level difference (MLD), is predominately a laboratory phenomenon, but it indicates that the auditory system can internally improve the SNR if the interaural differences for the signal and masker are different [73, 74]. It is reasonable to consider that even though the MLD is a threshold measurement and listening to conversations at a party is a supra-threshold task, the same underlying binaural mechanisms may operate in both situations.

Experiments that have examined speech intelligibility in free-field environments have manipulated and examined the importance of the following variables: reverberation [36, 58, 72, 75-82], SNR [20, 25, 26, 28, 33, 78, 80, 83-88], noise correlation [20, 89-91], angular separation of the signal and masker [20, 24, 28, 29, 32, 50, 80, 92, 93], and presentation methods [20, 36, 94-97].

2.3 Speech Perception in Virtual Environments

Listeners generally perceive real-world sounds as externalized. On the other hand, traditional binaural headphone presentation of sounds generally produces images that are perceived *inside* the head. Externalization of sound with free-field presentation (and their generally accurate localization) occurs because of the complex filtering of the acoustic signal produced by a listener's head, torso and pinnae. Because such filtering effects are not present in the traditional headphone presentation of signals, the latter mode of presentation does not result in externalization of sound images.

Several studies in recent years have attempted to simulated under headphones the actual acoustic signals that would arrive at the eardrums of the listener if the sound source were in a specific location relative to the listener [4-8]. This is done by filtering the inputs to the two ears by the so-called "head-related transfer functions," or HRTFs, that are the source-to-eardrum transfer functions (one for each ear) specific for the particular source position being simulated. The result, where this technique is applied correctly, is that the listener wearing headphones perceives the sound as externalized to the appropriate location in space. In other words, listening through headphones gives the listener the illusion that stimuli emanate from external locations as in a natural listening environment. The three-dimensional spatial effect thus achieved under headphones is called "virtual" auditory space, and the presentation mode is often referred to as "virtual listening" or "virtual or 3-D auditory displays." Investigators are now employing 3-D or virtual technology to study speech communication and intelligibility in a variety of adverse conditions [2, 4, 20, 36, 94].

The task of measuring numerous individual HRTFs (i.e., one for each ear and for each spatial location that must be simulated) is expensive and time-consuming, as well as technologically challenging. An alternative method to achieve externalization involves recording experimental signals and noise through microphones placed in the ear canals of KEMAR (Knowles Electronics Manikin for Acoustic Research), thus reproducing the signals arriving at KEMAR's eardrums. The two-channel recordings are subsequently presented to subjects through insert earphones to the two ears (Fig. 1). Recordings of this nature incorporate the effects of the head and pinnae on the incoming signals and reproduce the interaural differences as they exist in free-field listening conditions [32]. However, any method that employs a generic manikin will not always produce realistically externalized sounds. Thus, a listener presented with KEMAR-recorded signals may not externalize them in the same way as if the signals were filtered by his *own* HRTFs (i.e., due to differences between KEMAR's pinnae shapes and the listener's individual pinnae shapes).

In summary, these studies have demonstrated that the ability to understand speech in the presence of noise and reverberant environments is complex and depends upon numerous variables. The cues provided by monaural listening appear to contribute to speech intelligibility to a greater degree than do binaural cues. Chief among these cues is the relative energy of the signal and noise (SNR). However, the advantages of spatial hearing provided by binaural listening emerge in adverse conditions such as low SNRs, reverberation, or a combination of these conditions. The literature indicates that angular separation in real or virtual listening conditions improves speech intelligibility, but no study has compared the effect of adverse conditions of noise and reverberation on talker gender between traditional clinical evaluation methods and performance using recordings which employ generic binaural cues. Additionally, many clinical instruments examining speech intelligibility and communication headsets present speech in noise in a diotic presentation, ignoring the advantages of spatial hearing. The present study attempts to compare performance with traditional diotic presentation to that which incorporates the binaural advantages afforded from listening with two ears in more of a real-world environment.

3 METHODS

3.1 Subjects

Twenty-one adult native American English speakers with bilateral normal hearing were paid volunteers in this study. Subjects' age ranged from 21-38 years. Nineteen subjects were female and two were male. Subjects initially passed a pure-tone audiometric and immittance screening test confirming normal middle ear function and auditory thresholds. In an effort to avoid Idiopathic Discriminatory Dysfunction (IDD), no subject who reported (during initial interview) disturbed speech intelligibility in background noise was included [98]. Frequent breaks were allowed within a session to avoid fatigue.

3.2 Testing Environment and Stimuli

Subjects were tested individually in a lighted sound-insulated listening station. Subjects were grouped and tested simultaneously at three different listening stations (Fig. 1). During all sessions, each subject was seated in the same individual listening station wearing the same ER4 insert earphones.

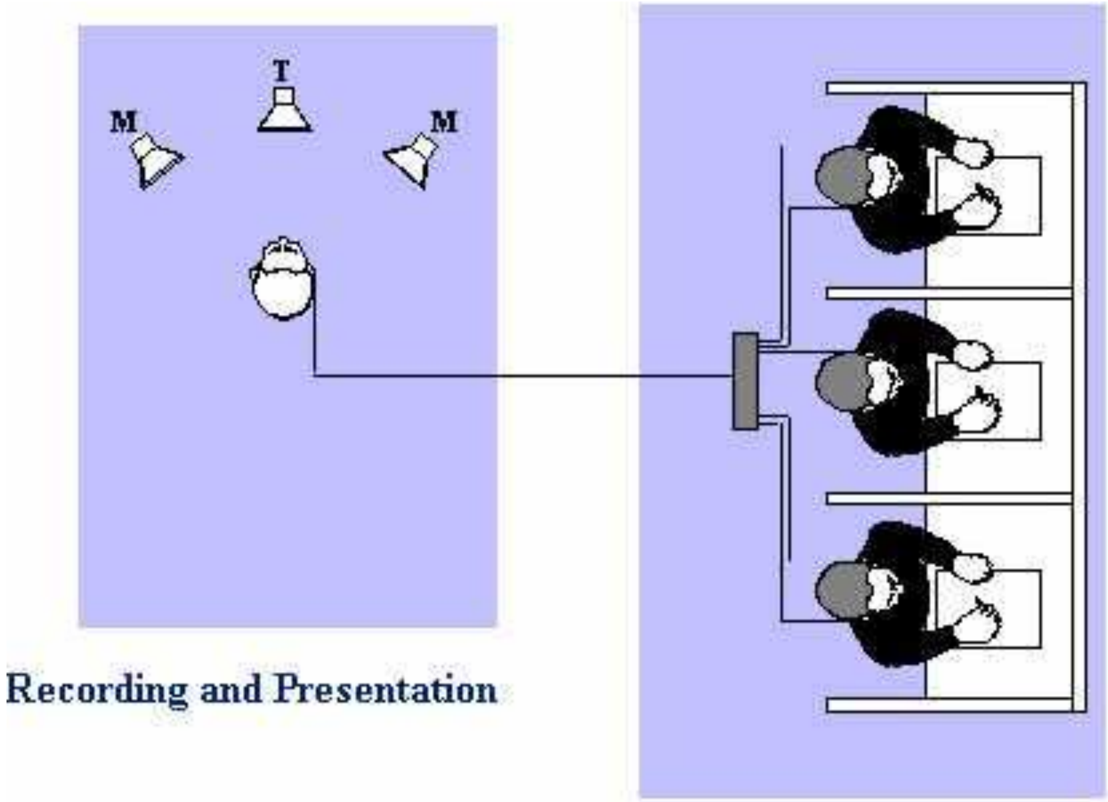

Figure 1: Recording and Presentation Environment

Nonsense syllables were chosen as the speech tokens to be identified. This choice was made in an attempt to control the monaural factors underlying speech perception in an auditory-only presentation. For example, it is well documented that clearly spoken speech is easier to understand than conversational speech for even normal hearing adults particularly in noise and reverberation [44]. The nonsense syllable ensured that articulation and speaking style did not introduce unwanted sources of variance. Additionally, this measure was selected because of the high correlation between nonsense syllables recognition and audiometric configuration and because of its sensitivity to the adverse influence of reverberation [72]. Additionally, the brief nature of the nonsense syllables closely duplicates brief communication found in a secure military environment.

The experimental speech stimulus tokens employed were the Nonsense Syllable Test (NST) [99]. The syllables used were taken from recordings of the UCLA Nonsense Syllable Test (NST) [100]. They consisted of the combination of 23 consonants, three vowels (/a/, /u/, /i/), with consonant position either initial (CV) or final (VC). Some consonants do not naturally occur in the English language in the initial and final position; thus, these combinations were not included in the syllable combination. The syllables were digitally recorded, calibrated and standardized for both a male and female talker. For a given run the 129 syllables were randomly presented without replacement.

The competing noise was digitally recorded multi-talker babble obtained from Auditect of St. Louis. To ensure an uncorrelated constant SNR with minimal envelope fluctuation, two digital 20-talker babble separate recordings were used. The multi-talker babble was the most effective masker for speech tokens, as it provided a worst-case listening scenario (i.e., a masker with a long-term speech spectrum and with little or no temporal variation in the envelope). This spectrum mirrors the long-term rms 1/3 octave-band idealized speech spectrum derived from ANSI S3.5-1969.

Speech recordings were made in an anechoic chamber (6 m x 6 m x 6 m). The syllable lists consisted of computer-generated CV and VC nonsense syllables randomly combined from all initial/final consonant tokens in all 3-vowel environments for a male and female talker. The CV and VC stimulus were presented from a single loudspeaker in the center of the loudspeaker array, and two independent babble sources were presented continuously through two different loudspeakers positioned at ~ 45 azimuth, 1.8 m from KEMAR. A 1000 Hz calibration tone was also recorded and subsequently used to set stimulus levels appropriately. For the virtual conditions, all stimuli (nonsense syllables plus babble) were recorded though the KEMAR manikin's two ears while it was positioned in the center of the anechoic chamber facing 0 azimuth. Etymotic ER-11 microphones were placed at the position of KEMAR's eardrums. Equalization filters were used to compensate for the frequency response of the ER4 insert earphones that subjects used. For the diotic control condition, KEMAR was removed from the chamber, and a single-channel recording was made with a single microphone at the position of KEMAR's head.

Each of the 129 syllables, spoken by the two talkers, was digitally stored in 9 different experimental conditions: 6 virtual presentation environments (2 virtual environments [virtual anechoic vs. virtual reverberant] x 3 SNRs [0, +5, +9]), plus three diotic conditions (3 SNRs) for both talkers. Additionally, 2 practice recordings (diotic and virtual) were made in an anechoic quiet environment for each talker, yielding a total of 22 recordings, each of which consisted of a different random ordering of only the 129 syllables. This resulted in 2838 recorded syllables that were later presented to subjects according to a random schedule described below. When the stimuli were played back to the subjects, the level of the target syllables was set such that the average level was 65 dB SPL. Babble presentation level was set (via the recording calibration tone) to 56, 60, or 65 dB SPL, to achieve SNRs of +9, +5 or 0 dB, respectively.

Subjects were presented the stimuli in one of two listening conditions: (a) in the 3-D or virtual condition, KEMAR's left and right ear recording was presented to the subjects' left and right ears, respectively, thus reproducing the interaural information and spectral cues provided by KEMAR's torso, head, and pinnae and, (b) in the diotic or mono condition, the single-channel (microphone) recording was presented identically to the subject's left and right ears. Diotic presentation was employed only for the anechoic condition to replicate typical recordings used in clinical evaluations. Reverberant room conditions were created by positioning acoustically reflective panels within the anechoic chamber around KEMAR.

3.3 Procedure

The subjects were seated comfortably in the auditory lab wearing ER 4 insert earphones. All testing presented the speech targets at 65 dB SPL. This level was well above threshold, while at the same time the multi-talker babble at all SNR ratios did not exceed uncomfortably loud or noise-hazardous loudness levels. The subjects were given an open-response version of the test with answer sheets containing a numbered set of blank lines on which to write down each syllable they heard. This open-response task was used to gather the broadest possible information about speech perception in the absence of any response constraints. The subject's task was to write each nonsense syllable on a response sheet provided by the experimenter as it was heard.

Subjects were tested in seven groups of three; the three subjects within a group were always presented the same stimuli in the same conditions; the order and gender of the conditions was counterbalanced across groups. Syllables were presented in lists of 129 syllables, where each syllable was randomly chosen (without replacement). The condition and speaker were held constant for each 129-syllable list. After each syllable presentation, there was a 5000-ms pause, during which time the subject recorded his or her response. A list included two blocks of syllables from either the male or female recording. The first block of 64 syllables was completed before the next block of 65 syllables began. The talker gender remained constant during the entire 129 syllable list. One half of the subject groups began with the male talker recordings while the other half began with the syllables spoken by a female talker. Short breaks were given between each syllable block with longer breaks between lists. Each 129-syllable list required approximately 20 minutes. All subjects practiced listening to the recorded stimuli in all conditions for a minimum of 3 hours. Each subject was required to pass a "qualification" test with 100% correct selected from stimuli presented in the diotic anechoic quiet condition. No subject could proceed until this high level of proficiency was achieved.

Thus, the nine experimental conditions can be represented as follows:

1. Virtual--anechoic (V-A) presented at +9, +5, and 0 SNR.
2. Virtual--reverberant (V-R) presented at +9, +5, and 0 SNR.
3. Diotic--anechoic (D-A) presented at +9, +5, and 0 SNR.

3.4 Practice

Three practice sessions consisting of at least two 64-syllable blocks for all 22 conditions were administered prior to data collection. Subjects were allowed to learn the task and reach asymptotic performance prior to data collection. The 1408 practice tokens exceeds the recommended minimum of 250 to minimize procedural learning effects [35]. Additionally, a 10-syllable trial warm-up was provided at the beginning of each condition.

4 RESULTS

4.1 Overall Analysis

All response scoring was blind. The number of nonsense syllables correctly identified for each listening condition was tabulated for each subject. The mean raw scores and their standard deviations for each condition are shown in Table 1. However, raw score means do not adequately describe stimulus intelligibility. Traditional speech testing normally evaluates performance in a percentage correct scale. Therefore, these raw scores were converted to percentage correct in the figures to better illustrate listener ability in a more meaningful manner.

	Signal-to-Noise Ratio		
Condition	**0dB**	**+5 dB**	**+9 dB**
Female Stimuli			
Diotic Anechoic	23.86(5.18)	34.95(4.75)	40.05(6.10)
Virtual Anechoic	74.38(11.57)	99.67(8.43)	102.43(7.19)
Virtual Reverberant	65.48(11.70)	93.67(9.42)	103.91(5.85)
Male Stimuli			
Diotic Anechoic	17.29(3.00)	32.38(5.92)	42.05(5.84)
Virtual Anechoic	56.62(8.93)	81.43(6.43)	91.48(5.33)
Virtual Reverberant	51.57(9.16)	76.19(8.47)	89.71(6.87)

Note. A raw score of 129 denotes perfect performance.

Table 1. Raw Score Means and Standard Deviations for All Conditions

An overall three-factor [3 presentation methods (diotic, virtual anechoic and virtual reverberant) X 3 SNRs (0dB, +5dB and +9dB) X 2 talker genders (male vs. female)] repeated measures analysis of variance (ANOVA) was conducted. The experimental score consisted of all correctly identified nonsense syllables (out of possible 129 total tokens) and was the dependent variable for statistical treatments. All statements addressing statistical significance were based on = 0.05.

All three main effects were significant: presentation method [F (2,40) = 1635.98, $p < .001$], SNR [F (2,40) = 916.12, $p < .001$], and talker gender [F (1,20) = 334.93, $p < .001$]. First order interactions occurred for presentation method x SNR [F (4,80) = 38.06, $p < .001$], SNR x talker gender [F (2,40) = 8.2, $p < .001$], and presentation method x talker gender [F (2,40) = 80.55, $p < .001$]. Additionally, a second order interaction emerged for environment x SNR x talker gender [F (4,80) = 2.8, $p < .031$].

4.1.1 The Effects of Presentation Method and SNR

The main effects of presentation method and SNR are apparent in Figure 2 where the data are collapsed across talker gender. Here it can be seen that intelligibility was significantly less in the diotic presentation method than in either virtual (anechoic or reverberant) condition. Clearly, subjects had greater difficulty identifying the nonsense syllables when recorded and presented in a diotic manner. It can also be seen that performance improved with increasing SNR. The predictable response pattern to SNR found in diotic presentations appears to also be present in the virtual presentations, as scores behaved in a similar manner.

Considering performance in the two virtual conditions, performance (collapsed across genders) in the reverberant environment was degraded compared to in the anechoic condition except in the most favorable SNR (+9dB). To test the significance of this effect, another three-factor [2 environment (anechoic vs. reverberant) X 3 SNR (0, +5 dB, +9 dB) X 2 talker gender] repeated measures analysis of variance on the raw scores between the virtual conditions was conducted. Significant main effects of environment [F (1,20) = 39.95, $p < .001$], SNR [F (2,40) = 662.03, $p < .001$], and talker gender [F (1,20) = 378.183, $p < .001$] were obtained. The only significant interaction was between SNR and environment [F (2,40) = 9.5, $p < .001$]. Thus, the small performance difference (5%) at 0 and 5 dB SNR was statistically significant, but performance in the anechoic and reverberant conditions at a +9 dB SNR was essentially the same.

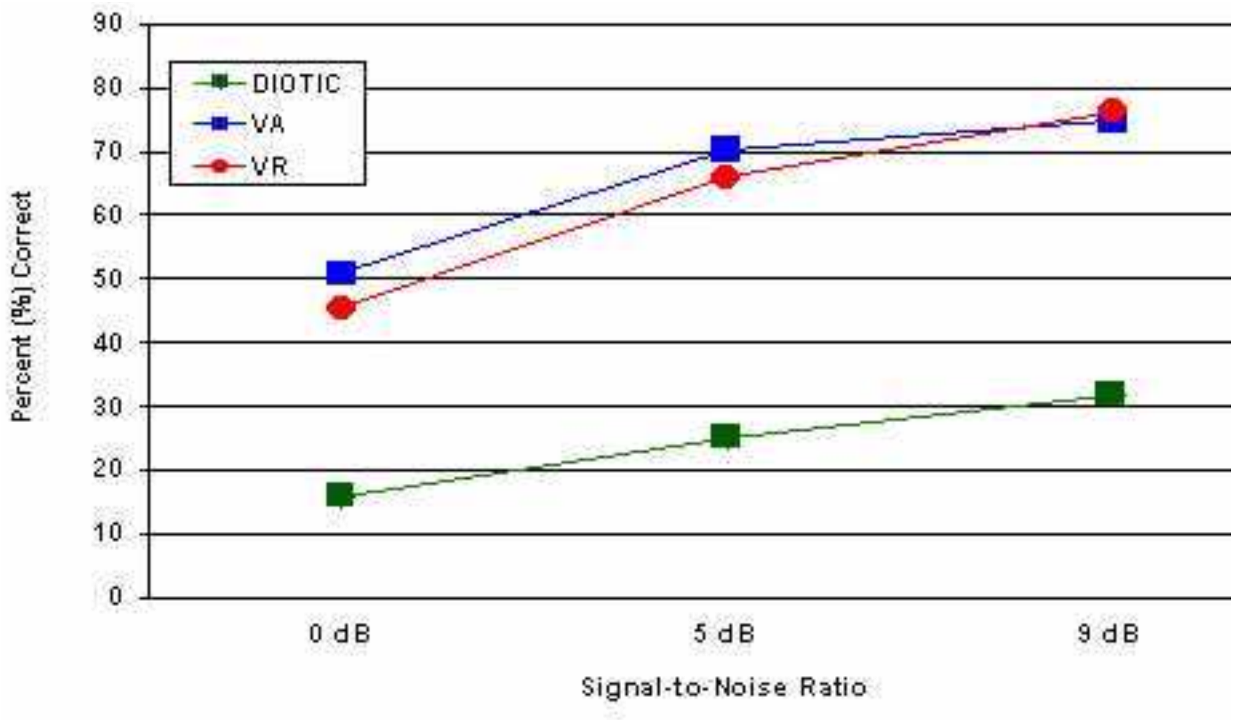

Figure 2. Effect of Signal-to-Noise Ratio on Speech Intelligibility by Condition

4.1.2 The Effects of Talker Gender

The effect of talker gender was also significant. Subjects identified female nonsense syllables with greater accuracy than male tokens for both virtual recordings at all three levels of SNR. This was not the case for the stimuli recorded in a diotic mode. Intelligibility in the most adverse listening condition (0 dB SNR) exhibited the same trend as in the virtual recordings, with higher identification accuracy for female tokens. However, in the more favorable diotic listening conditions (SNR = +5 and +9 dB) there was little if any difference in the identification of male versus female tokens.

5 Discussion

5.1 The Effect of Virtual Presentation Mode on Identification Performance

The results of this study definitely indicate that in the presence of a multi-talker noise (positioned at two locations 45), virtual nonsense syllables recorded through a manikin (in either anechoic or reverberant environments) produced higher identification scores than identical nonsense syllables recorded in diotic conditions. The amount of benefit from the virtual presentation, collapsed across talkers and SNR, was approximately 41%. This improvement is slightly higher than the benefit (15%-28%) noted by previous investigators [20, 36].

In a similar manner, Bronkhorst and Plomp [32] also pre-recorded speech in an anechoic room through a manikin in an attempt to measure the effect of spatial separation of a single competing noise on speech intelligibility. In this study, the level required for subjects to achieve an overall 50% intelligibility threshold was measured by varying the signal while maintaining the same noise level. Separating the noise and speech in auditory space by 90 enabled subjects to obtain 50% correct intelligibility when the signal was 10.1 dB softer than when the noise and signal were located in the same position. In a more recent study that employed virtual signals created with individualized HRTFs, Koehnke and Besing [4] employed a similar adaptive threshold technique, but varied the noise while keeping the speech signal constant. They reported subjects gained an even larger advantage (13.7 dB) when 90 separated the single competing noise and the stimuli.

It is difficult to directly compare the performance improvement in the present experiment (where SNR was held constant) to that of the studies of Bronkhorst and Plomp [32] and Koehnke and Besing [4] who varied SNR to find the 50% performance level. However, the results are in qualitative agreement. In all three studies performance improved when signal and noise were spatially separated (as in the virtual condition) compared to the case when signals and noise came from the same location (as in the diotic condition).

It should be noted that the aforementioned studies employed only one noise source. Perhaps improvement differs with multiple noise sources. Subjects in the present study were able to identify very difficult stimuli sets even though competing noise sources simultaneously originated from not one, but two spatially separated locations. Additionally, the competing noise used in this study was a 20-talker babble. Multiple talkers tend to degrade speech intelligibility more than random noise, due to the similarity of speech signal spectra modulations. The improvement in intelligibility was corroborated by subject reports of their ability to better understand the syllables in virtual listening conditions.

5.2 The Effects of Reverberation on Performance in Virtual Environments

It is generally accepted that speech perception in noise is degraded by reverberation. In the present study, performance in the virtual reverberant environment (collapsed across SNRs) significantly decreased speech intelligibility compared to the virtual anechoic recordings (average difference: 3.4%). The decreased intelligibility in normal hearing listeners less than expected although generally consistent with previous investigations in reverberation and noise in real and virtual environments [4, 36].

One possible reason that reverberation had a relatively small effect on performance may be a function of methodology. In the present study, the nonsense syllables were presented in the traditional manner for the UCLA nonsense syllable test (without a carrier phrase). Houtgast & Steeneken [101] showed that the effect of reverberation on speech intelligibility was smaller for tests without carrier phrases than for tests with carrier phrases. This difference may be due to the fact that a carrier phrase precedes the speech target, increasing the reflected energy that could serve as a masker. On the other hand, speech tests without carrier phrases present the speech target immediately into the reverberant environment. Listeners hear the target speech token first without persistent reflected speech signals introduced by a carrier phrase. To the extent this difference holds, the present results may actually underestimate the influence of reverberation for meaningful material.

Even with spatial separation and absence of a carrier phrase, these results indicate that some persistent reverberant energy did significantly degrade performance. This delayed or reflected energy can influence some important aspects of a speech signal and interfere with intelligibility by distorting the original signal [102]. Low frequencies are typically absorbed less efficiently by reflective surfaces. Because vowel (low-frequency) sounds are usually more intense than consonant sounds (high-frequency), the reflected vowel phonemes tend to mask consonant manner and place information, particularly in the VC combinations. Noise and reverberation as found in this study combine synergistically to degrade speech recognition.

The effect of reverberation in this study has important military relevancy. For example, soldiers prosecute military operations in urban terrain by conducting house-to-house and room-to-room searches of areas of interest. Likewise, communication in tactical operations centers (TOC) occurs in rooms, tents, or enclosures with similar or higher reverberation times. The reflections from a reverberant environment may be thought of as a type of masking noise, although the effects of noise and reverberation are perceptually different. Soldiers can easily detect the presence of noise, whereas the presence of a highly reverberant field is often not apparent. The subtle effects of reverberation may not be evident to an untrained soldier listener, yet may result in miscommunication as the ability to understand speech is degraded. If the influence of reverberation can adversely affect the ability to communicate in these enclosed environments, soldiers must recognize the possibility of decreased communication ability in reverberant environments.

5.3 The Effects of Signal-to-Noise Ratio (SNR) on Performance

The SNR appeared to influence intelligibility in a similar manner for all presentation methods (virtual or diotic -- see Fig 3). Subject performance increased as SNR increased, although absolute identification remained poor in the diotic conditions for all SNRs. For example, identification of nonsense syllables doubled (15.95% to 31.82%) in diotic presentation as SNR improved from 0 dB to +9 dB. Likewise, increasing the SNR from 0 dB to +9 dB also significantly improved intelligibility, increasing scores an average of 24.2% and 29.7% in the virtual and reverberant environments, respectively. Apparently, the auditory system behaves in a similar manner with respect to signal-to-noise ratio in diotic and virtual presentation modes, although absolute intelligibility is very different.

5.4 Effects of Talker Gender on Identification of Performance

A major finding in the present study was that syllables spoken by this female talker were identified at a higher rate than when spoken by this male talker, at least in the virtual environments. This difference was not as evident for diotic presentation. Specifically, female tokens (collapsed across all SNRs) were more easily identified than male tokens in the virtual anechoic (difference: 12.2%) and reverberant (difference: 11.6%) presentations. However, in the diotic condition, the advantage for female tokens occurred only at the most difficult SNR (0 dB), and here the difference was only 5.1%. As SNR increased, the female talker advantage decreased to (2.0%). Finally, at the highest SNR (+9 dB), subjects identified male tokens with a slightly higher accuracy (1.6%) than female tokens. Hence, it appears for these stimuli, the advantage for female spoken tokens was evident when the signal was separated from the competing talkers in auditory space (i.e., under virtual listening conditions). Under diotic conditions there was generally little or no advantage with the female talkers. These results are similar to those reported by Ericson and McKinley [20]. In their investigations, intelligibility improved more dramatically for female competing targets when spatially separated than when presented diotically. These results may indicate that female speech may be improved by incorporating spatial separation in communication systems.

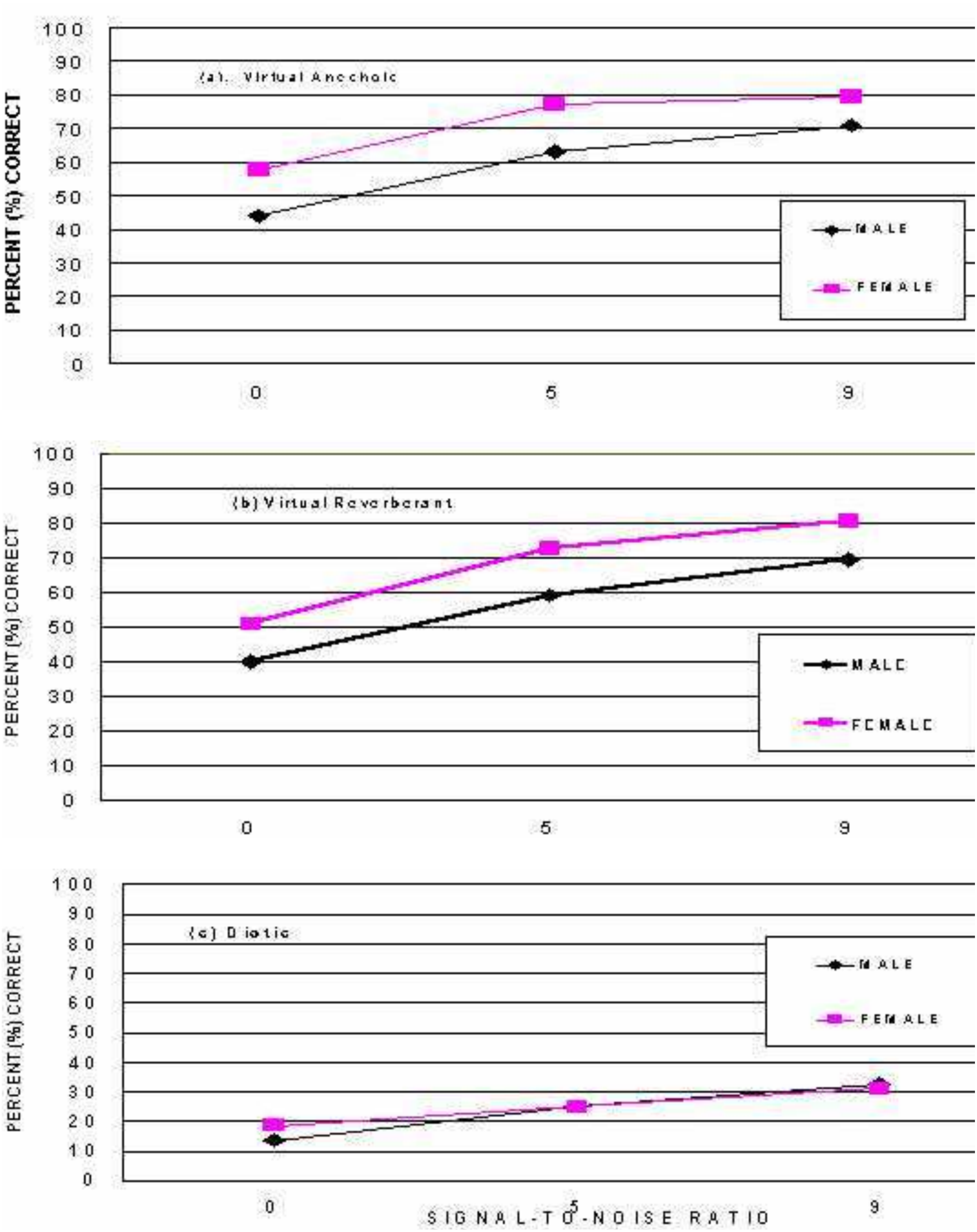

Figure 3. Effect of Gender by Signal-to-Noise Ratio and Condition

An attempt was made to explain the gender effect by computing the rms energy of each token. Although the rms energy was slightly higher for the female tokens in the // vowel context than for the male tokens (which could possibly account for some of the performance differences in the virtual conditions), this explanation is not entirely satisfactory, because it would predict a female advantage in the diotic conditions as well; as described earlier, there was little if any gender effect in the diotic condition for the more favorable SNRs. Several investigators have examined the spectral and aerodynamic vocal characteristics of male and female talkers and confirmed parameter differences between genders across speech conditions [103 -105]. Generally, female speakers are perceived to be breathier than male speakers. The degree of glottal closure appears to be related to the perception of breathiness. Sodersten et al. [105] reported a significant correlation between perceived breathiness and the relative amplitude of the first harmonic. Further study is required to determine which of these acoustic measures (if any) might underlie the consistent female token advantage in virtual listening conditions.

6 Conclusions

The findings of this study may be summarized as follows:

1. Traditional diotic presentation of UCLA NST stimuli degrades subject performance compared to virtual presentation and may not accurately represent real-world performance.

2. Signal-to-noise ratios did not influence performance trends between presentation methods, as intelligibility improved with increasing SNR for all three presentation modes. However, performance differences between the virtual anechoic and virtual reverberant presentations disappeared at the most favorable SNR (+9), as the adverse effect of reverberation was minimized for these normal-hearing listeners.

3. Reverberation decreased speech intelligibility, although the effects were not as dramatic as expected. These results may indicate that spatial separation reduces the adverse effects of reverberation.

4. Female talker tokens used in this study were more easily identified than male talkers in a virtual presentation, but not generally for diotic presentation. Talker gender differences may not be evident in a diotic presentation, but may affect real-world intelligibility.

In summary, in situations where communication is critical and adverse environments result in low SNRs, virtual auditory headphone presentation may provide improved speech communication. This study suggests that traditional diotically recorded and presented speech recognition test materials may not accurately reflect real-world performance. The results lend credence to the use of virtual testing as a clinical tool. Potential application in clinical speech intelligibility testing may provide a more accurate measure of real world performance than traditionally presented diotic testing, but only a real world test will validate this premise. Additionally, this study suggests that conventional presentation may not adequately differentiate gender advantages or disadvantages in speech recognition. Further study is indicated to investigate this finding.

7 References

1. Matschke, R. G. (1994). Communication in noise: Speech intelligibility of airplane pilots with and without active noise compensation. *HNO, 42*(8), 499-504.
2. McKinley, R. A., Ericson, M. A., & D'Angelo, W. R. (1994). 3-dimensional auditory displays: Development, application, and performance. *Aviat. Spa. Environ. Med*, *65*(5), Suppl., A31-8.
3. Ribera, J. E., Mozo, B. T., Mason, K. T., & Murphy, B. A, (1996). Communication survey of CH-47D crew members. *Military Medicine, 161*(7), 387-391.
4. Koehnke, J., & Besing, J. M. (1996). A procedure for testing speech intelligibility in a virtual listening environment. *Ear and Hearing, 17*(3), 211-217.
5. McKinley, R. A., & Ericson, M. A. (1989). A real-time digital auditory localization cue synthesizer. *Sound Localization by Human Observers Symposium Proceedings.* Washington, DC: National Academy Press.
6. Wenzel, E. M., Stone, P. K., Fisher, S. S., & Foster, S. H. (1990). A system for three-dimensional acoustic "visualization" in a virtual environment workstation. P*roceedings of the IEEE Visualization-90 Conference* (pp. 329-337). Los Alamitos, CA.
7. Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening: I. Stimulus synthesis. *J. Acoust. Soc. Am., 85,* 858-867.
8. Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening: II. Psychophysical validation. *J. Acoust. Soc. Am , 85*, 868-878.
9. Besing J., & Koehnke, J. (1995). A test of virtual auditory localization. *Ear and Hearing, 16*(2)*,* 220-229.
10. Byrne, D. (1992). Key issues in hearing aid selection and evaluation. J. Am. Acad. Of Aud., 3(2), 67-80.
11. Koehnke, J. Besing, J. M. & Abouchacra, K. (1995). Virtual reality: 3-D audio and clinical testing.
    [Presentation]. Dallas, Texas: American Academy of Audiology.
12. Koehnke, J., Besing, J. Goulet, C., Allard, M., & Zurek, P. M. (1992). Speech intelligibility, localization, and binaural detection with monaural and binaural amplification. *J. Acoust. Soc. Am., 94*, 2434.
13. Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *J. Acoust. Soc. Am., 98*(5), Pt. 1*,* 2542-2553.
14. Hollien, H., Hicks, J. W., Jr., & Klepper B. (1986). An acoustic approach to diver navigation. *Undersea Biomedical Research*, *13*(1), 111-128.

15. Gillingham, K. K. (1992). The spatial disorientation problem in the United States Air Force. *J. Vestibular. Res*., *2*, 297-306.
16. Rosenberg, M. E., Barson, J. V, Pollard, A. J. (1994). Auditory localizing acuity during exposure to altitude. *Aviation, Space and Environmental Medicine*, *65*, 649-653.
17. Loomis, J. M., Chick, H., Cicinelli, J. G. (1990). Active localization of virtual sound sources.*J.Acoust.SocAm.,88*,1757-1767.
18. McGee, K. (1994). Exploring the third dimension. *Roland Users Group: The magazine for the electronic musician*, *12*(1), 64-66.
19. Trahiotis, C., & Bernstein, L. R. (1987). Some modern techniques and devices used to preserve and enhance the spatial qualities of sound. In W. A. Yost & G. Gourevitch (Eds.), *Directional Hearing* (pp. 279-290). New York: Springer-Verlag.
20. Ericson, M. A., & McKinley, R. L. (1997). The intelligibility of multiple talkers separated spatially in noise. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural And Spatial Hearing In Real And Virtual Environments* (pp. 701-724). Mahwah, NJ: Lawrence Erlbaum.
21. Abouchacra, K. S. (1994). Directional masking: Detection, localization and confidence in responses to a target signal presented in directional noise. *J. Acoust. Soc. Am.,* 95, 2897.
22. Vause, N. L., & Grantham, D. W. (1997). Effects of earplugs and protective headgear on auditory localization ability in the horizontal plane, *Human Factors,* submitted.
23. Newton, V. E., & Hickson, F. S. (1981). Sound localization: Part II. A clinical procedure. *J. of Laryngology and Otology*, *95*(1), 41-48.
24. Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am., 25*, 975-979.
25. Yost, W. A. (1997). In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 329-347). Mahwah, NJ: Lawrence Erlbaum.
26. Hirsh, I. J. (1950). The relation between localization and intelligibility. *J. Acoust. Soc.Am., 22*(2), 196-200.
27. Kock, W. E. (1950). Binaural localization and masking. *J. Acoust. Soc. Am., 22*(6),801-804.
28. Plomp, R. (1976). Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acustica, 34*, 200-211.
29. Plomp, R., & Mimpen, A. M. (1981). Effect of orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences. *Acustica, 48,* 325-328.
30. Santon, F. (1976). Numerical prediction of echograms and of the intelligibility of speech in rooms. *J. Acoust. Soc. Am., 59*, 1399-1405.
31. Thompson, P. O., & Webster, J. C. (1964). The effect of talker-listener angle on word intelligibility: II. In an open field. *Acustica, 14,* 44-49.
32. Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J. Acoust. Soc. Am., 83*, 1508-1516.
33. Dirks, D. D., & Wilson, R. H. (1969). The effect of spatially separated sound sources on speech intelligibility. *J. Speech Hear. Res., 12*, 5-38.
34. Epstein, A., Giolas, T. G., & Owen, E. (1968). Familiarity and intelligibility of monosyllabic word lists. *J. Speech Hear. Res., 11*, 435-438.
35. Theodoridis, G. C., & Schoeny, Z. G. (1990). Procedure learning effects in speech perception tests. *Audiology, 29*, 228-239.
36. Abouchacra, K. S., Tran, T. V., Besing, J., & Koehnke, J. (1997). Performance on a selective attention task as a function of stimuli presentation mode. *ARO*, No. 210, p. 53.
37. Hübner, R., & Hafter, E. R. (1995). Cueing mechanisms in auditory signal detection. *Perception&Psychophysics*,*58*(2), 197-202.
38. Grant, R. W., & Braida, L. D. (1991). Evaluating the articulation index for auditory-visual input. *J. Acoust. Soc. Am., 89*, 2952-2960.
39. Warren, R. M., Riener, K. R., Bashford, J. A., Jr., & Brubaker, B. S. (1995). Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Perception & Psychophysics*, *57*(2), 175-182.
40. Bashford, J. A., Jr., Reiner, K. R., & Warren, R. M. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. *Perception & Psychophysics*, *51*, 211-217.
41. Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Intelligibility of average talkers in typical listening environments. *J. Acoust. Soc. Am., 81*, 1598-1608.
42. Hood, J. D., & Poole, J. P. (1980). Influence of the speaker and other factors affecting speech intelligibility. *Audiology, 19*, 434-455.
43. Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing: I. Intelligibility differences between clear and conversational speech. *J.Speech Hear. Res., 28*, 96-103.
44. Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J. Acoust. Soc. Am., 95*(3), 1581-1592.
45. Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *J. Acoust. Soc. Am., 46*, 694-703.
46. Pollack, I., & Pickett, J. M. (1958a). Masking of speech by noise at high sound levels. *J. Acoust Soc.Am., 30*(2), 127-130.
47. Festen, J. M., & Plomp, R. (1990). Effects of fluctuation noise and interfering speech on the speech-reception threshold for impaired and normal hearing, *J. Acoust. Soc. Am., 88*(4), 1725-1736.

48. Stuart, A., Phillips, D. P., & Green, W. B. (1995). Word recognition performance in continuous and interrupted broad-band noise by normal-hearing and simulated hearing-impaired listeners. *American J. of Otology*, *16*(5), 658-663.

49. Ainsworth, W. A., & Meyer, G. F. (1994). Recognition of plosive syllables in noise: Comparison of an auditory model with human performance. *J. Acoust. Soc. Am., 6*(2), Pt. 1, 687-694.

50. Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speech like maskers on binaural speech recognition in normal and impaired hearing. *J. Acoust. Soc. Am., 92*, 3132-3139.

51. Miller, G. A. (1947). The masking of speech. *Psychol. Bull., 44*, 105-129.

52. Pollack, I., & Pickett, J. M. (1958b). Stereophonic listening and speech intelligibility against voice babble. *J. Acoust. Soc. Am., 30*(2), 131-133.

53. Danhauer, J. L., & Leppler, J. G. (1979). Effects of four noise competitors on the California consonant test. *J. Speech Hear. Dis., 44*, 354-361.

54. Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am., 61,* 1337-1351.

55. Carhart, R., Johnson, C., & Goodman, J. (1975). Perceptual masking of spondees by combinations of talkers. *J. Acoust. Soc. Am., 58*(Suppl. 1), 535.

56. Lewis, H. D., Benignus, V. A., Muller, K. E., Malott, C. M., & Barton, C. N. (1988). Babble and random-noise masking of speech in high and low context cue conditions. *J. Speech Hear. Res., 31*, 108-114.

57. Needleman, A. R., & Crandell, C. C. (1995). Speech recognition in noise by hearing-impaired and noise-masked normal-hearing listeners. *J. Am. Acad. Audiol., 6*, 414-424.

58. Nábelek, A. K., Letowski, T. R., & Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *J. Acoust. Soc. Am., 86*(4), 1259-1265.

59. Gower, D. W., & Casali, J. G. (1994). Speech intelligibility and protective effectiveness of selected active noise reduction and conventional communication headsets. *Human Factors, 36*(2), 350-367.

60. Abel, S. M., & Spencer, D. L. (1997). Consonant discrimination in noise with earplugs and muffs. *J. Acoust. Soc. Am., 101*(5), Pt. 2, 3127A.

61. Plomp, R. (1986). A signal-to-noise ratio model for the speech reception threshold of the hearing impaired. *J. Speech Hear. Res., 29*, 146-154.

62. Studebaker, G. A., Pavlovic, C. V, & Sherbecoe, R. L. (1987). A frequency importance function for continuous discourse. *J. Acoust. Soc. Am., 81*, 1130-1138.

63. Markides, A. (1986). Speech levels and speech-to-noise ratios. *Br. J. Audiol., 20*, 115-120.

64. Houtgast, T. (1981). The effect of ambient noise on speech intelligibility in classrooms. *Applied Acoust., 14*, 15-25.

65. Blauert, J. (1997). An introduction to binaural technology. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 593-609). Mahwah, NJ: Lawrence Erlbaum Associates.

66. Stone, M. A., & Moore, B. C. (1992). Spectral feature enhancement for people with sensorineural hearing impairment: effects on speech intelligibility and quality. *J. Of Rehabilitation Research and Development, 29*(2), 39-56.

67. Benoit, C. (1990). An intelligibility test using semantically unpredictable sentences: Towards the quantification of linguistic complexity. *Speech Communication, 9*, 293-304.

68. Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *J. Acoust. Soc. Am., 84*, 101-114.

69. Bell, T. S., Dirks, D. D., & Trine, T. D. (1992). Frequency-importance functions for words in high- and low-context sentences. *J. Speech Hear. Res., 35*, 950-959.

70. Studebaker, G. A., & Sherbeco, R. L. (1991). Frequency-importance and transfer functions for recorded CID W-22 word lists. *J. Speech Hear. Res., 34*, 427-438.

71. Mendel, L. L., & Danhauer, J. L. (1997). *Audiologic evaluation and management and speech perception assessment*. San Diego, CA: Singular.

72. Pekkarinen, E., & Viljanen, V. (1990). Effect of sound-absorbing treatment on speech discrimination in rooms. *Audiology, 29*, 210-227.

73. Levitt, H., & Rabiner, L. R. (1967a). Binaural release from masking for speech and gain intelligibility. *J. Acoust. Soc. Am., 42*, 601-608.

74. Levitt, H., & Rabiner, L. R. (1967b). Predicting binaural gain in intelligibility and release from masking for speech. *J. Acoust. Soc. Am., 42*, 820-829.

75. American Speech Language and Hearing Association, (1995). Position Statement and Guidelines for Acoustics in Education Settings, American Speech and Hearing Association, 37 (Suppl. 14, pp. 15).

76. Haas, H. (1951). On the influence of a single echo on the intelligibility of speech. *Acustica, 1*, 49-58.

77. Lochner, J. P. A., & Burger, J. F. (1964). The influence of reflections on auditorium acoustics. *J. Sound and Vibration, 1*, 426-454.

78. Moncur, J. P., & Dirks, D. (1967). Binaural and monaural speech intelligibility in reverberation. *J. Speech Hear. Res., 10*, 186-195.

79. Gelfand, S. A., & Hochberg, I. (1976). Binaural and monaural speech discrimination under reverberation. *Audiology, 15*, 72-84.

80. MacKeith, N. W., & Coles, R. R. A. (1971). Binaural advantages in hearing of speech. *J.Laryngol. Otol., 85,* 213-232.

81. Crum, M. A., & Tillman, T. W. (1973). *Effects of speaker-to-listener distance upon speech intelligibility in reverberation and noise*. Paper presented at Am. Sp. and Hrg Ass., Detroit. Michigan, USA.

82. Greenberg, J. E., & Zurek, P. M. (1992). Evaluation of an adaptive beamforming method for hearing aids. *J. Acoust. Soc. Am., 91*(3), 1662-1672.

83. Nábelek, A. K., & Pickett, J. M. (1974). Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing impaired listeners. *J. Speech Hearing Research, 17*, 724-739.

84. Asahina, N. (1994). A study on speech intelligibility in traffic noise. *J. of Oto-Rhino-Laryngology of Japan, 97*(3), 401-413.

85. Good, M. D., Gilkey, R. H., & Ball, J. M. (1997). The relation between detection in noise and localization in noise in the free field. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 349-376). Mahwah, NJ: Lawrence Erlbaum.

86. Kollmeier, B. (1997). Signal processing for hearing aids employing binaural cues. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 753-775). Mahwah, NJ: Lawrence Erlbaum.

87. Saberi, K., Dostal, L, Sadralodabai, T., Bull, V., & Perrott, D. R. (1991). Free-field release from masking. *J. Acoust. Soc. Am., 90*(3), 1355-1370.

88. Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level., *J. Acoust. Soc. Am.,* 66, 1333-1342.

89. Robinson, D. E., & Jeffress, L. A. (1963). Effect of varying the interaural noise correlation on the detectability of tonal signals. *J. Acoust. Soc. Am., 35,* 1947-1952.

90. Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *J. Acoust. Soc. Am., 20*, 150-159.

91. Doll, T. J., Hanna, T. E., & Russotti, J. S. (1992). Masking the three-dimensional auditory displays. *Human Factors, 34*(3), 255-265.

92. Cherry, E. C., & Taylor, W. K. (1954). Some further experiments upon the recognition of speech with one and with two ears. *J. Acoust. Soc. Am., 26*, 554-559.

93. Burkhard, M. D., & Sachs, R. M. (1975). Anthropometric manikin for acoustic research.*J. Acoust. Soc. Am., 58*, 214-222.

94. Ricard, G. L., & Meirs, S. L. (1994). Intelligibility and localization of speech from virtual directions. *Human Factors, 36*, 120-128.

95. Moore, T. J. (1981). Voice communication jamming research. *Advisory Group for Aerospace Research and Development Conference Proceedings*, No. 311, Aural Communication in Aviation, CP311, 2-1-2-6.

96. McKinley, R. L., & Moore, T. J. (1989). An information theory based model and measure of speech communication effectiveness in jamming. *Proceeding of Speech Tech. '89.* New York: Media Dimensions.

97. Yost, W. A., Sheft, S., & Dye, R.H. (1994). Divided auditory attention with up to three sound sources: A cocktail party. *J. Acoust. Soc. Am., 95*, 2916.

98. Rappaport, J. M., Phillips, D. P., & Gulliver, J. M. (1993). Disturbed speech intelligibility in noise despite a normal audiogram: A defect in temporal resolution? *Journal of Otolaryngology*, *22*(6), 447-453.

99. Resnick, S.B., Dubno, J.R., Hoffnung, S. & Levitt, H. (1975). Phoneme errors on a nonsense syllable test, *J. Acoust. Soc. Am.,* 58, S114.

100. Dubno, J. R., & Dirks, D. D. (1982). Evaluation of hearing-impaired listeners using a nonsense-syllable test: I. Test reliability. *J. Speech Hear. Res., 25*, 135-141.

101. Houtgast, T., & Steeneken, H. J. M., (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria., *J. Acoust. Soc. Am.,* 77(3), 1069-1077.

102. Houtgast, T., & Steeneken, H. J. M., (1973). The modulation Transfer Function in Room Acoustics as a predictor of speech intelligibility. *Acustica, 28,* 66-73.

103. Holmberg, E. B., Perkell, J. S. Hillman, R. E., & Gress, C. (1994), Individual variations in measures of voice. *Phonetica*, 51, 30-37.

104. Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am., 87*, 820-857.

105. Sodersten, M., & Lindestad, P. A. (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. *J. Speech Hear. Res., 33*(3), 601-611.