INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

EVALUATION OF A BINAURAL LISTENING SYSTEM
USING A MONOPHONIC INPUT

By

MARYROSE HANNON MCINERNEY

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2002

UMI Number: 3047243

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

4-16-02
Date

Harry Levitt, Ph.D.
Chair of Examining Committee

4/16/02
Date

Robert Goldfarb, Ph.D.
Executive Officer

Supervisory Committee:

Arlene Neuman, Ph.D.
Associate Professor
Graduate Center of CUNY

Patrick Nye, Ph.D.
Haskins Laboratories
New Haven, CT

External Examiner:

Mark Ross, Ph.D.
Storrs, CT

THE CITY UNIVERSITY OF NEW YORK

Abstract

Evaluation of a Binaural Listening System Using A Monophonic Input

By

Maryrose Hannon McInerney

Advisor: Distinguished Professor Harry Levitt

This dissertation evaluated a simulated binaural signal processing system for separating speech from noise using a monophonic input. Sinewave modeling was used to separate the peaks of the speech-plus-noise spectrum from the residual spectral components. Interaural time and/or intensity differences were used to alter the perceived locations of the spectral peaks and the residual spectral components. It was hypothesized that the sound image associated with the spectral peaks, which is heard as speech, would be perceived towards one side of the head while the sound image associated with the residual spectral components, which is heard as noise, would be heard towards the opposite side of the head. It was also hypothesized that the combination of interaural time and intensity differences would produce even greater differences in perceived location than either interaural time or interaural intensity differences alone. An additional hypothesis was that the differences in perceived location would also improve ease of listening.

Ten normal hearing adults were tested under four experimental conditions: Interaural time difference (IT), Interaural intensity difference (II), Interaural time and intensity difference (ITI) and no processing (NP). There were 4 replications for each of 4 speech-to-noise ratios (2dB, 0dB, -2dB, -4dB) for a total of 64 trials per subject. Judgments were made of the perceived locations of the speech and noise images, spatial diffuseness of the speech image and ease of listening of the speech image.

Results supported the hypothesis regarding the effect of signal processing on the perceived location of the speech and noise. The magnitude of the effect was greater for the ITI condition than for either the IT or II conditions. An unexpected finding was that the processing technique had a greater effect on the perceived location of the noise than on the perceived location of the speech. Speech-to-noise ratio did not have a significant effect on the changes in perceived location produced by the processing technique. The processing technique also did not have a significant effect on the ratings of spatial diffuseness of the speech image or on the ratings of ease of listening.

ACKNOWLEDGMENTS

"Two roads diverged in a wood, ... I took the one less traveled by, and that has made all the difference." Robert Frost

I have met many individuals along the way. So many have contributed with their kind words and supportive thoughts. For those that admire my persistence, I can only say that I would not have completed this task if I did not have my family and friends. I humbly thank you all.

Most of all I would like to thank my husband, Bill, for without his encouragement and unconditional love the completion of this dissertation would not have been possible. As well I would like to thank my children, Maggie, Caroline and Michaela who provided me constant support and inspiration in good times and bad.

I thank my parents, Ann and Gerard Hannon, and my surrogate parents, Rose and Tom McInerney, who gave me faith and family.

Without my chair, my mentor, my friend, Harry Levitt, this journey would not have come to an end. I wholeheartedly thank him for sharing his abundant knowledge, patience and good humor with me. As he ended so many conversations, "Looking forward to hearing from you, … all the best".

I thank Arlene Neuman and Patrick Nye who served on my committee. Arlene Neuman, who taught me the skill of research and the art of believing in myself. Patrick Nye, for his dedication and patience to the 'cause'.

I graciously thank Mark Ross who accepted the role of external reader. Robert Goldfarb who was there for the touchdown. In memory of Irv Hochberg, who was instrumental in my completion of the program.

To all my friends, especially Pat, Debbie and Barbara Kurman, who were always whispering in my ear "you can do it, you can do it", I give a special squeeze.

Thank you also to all my friends at the City University of New York, especially Linda Ashour and Loretta Walker who provided laughter, making those gray days a little brighter.

Finally I would like to thank all my colleagues, especially the senior Audiology staff, Audiology support staff and Pediatric Administration at Hackensack University Medical Center for their unselfish support. I especially would like to thank Dr. Marvin Gottlieb, Chairman of Pediatrics, who first inspired me to pursue my doctorate.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION and PURPOSE

Binaural listening offers many advantages over monaural listening. These include the ability to localize sound as well as improved intelligibility and ease of listening for speech-in-noise when the speech and the noise are spatially separated.

The underlying mechanism a normal hearing listener uses to identify the apparent location of a sound source is complex involving many variables. Binaural cues for localization include all of the position-dependent differences between the signals received by the two ears from a single source: interaural time differences, interaural intensity differences, and interaural spectral differences (the differential effect of frequency on interaural intensity differences, particularly for frequencies above 4.0 Hz) (Hirsh, 1950). There are also other related variables such as head movement (Wightman and Kistler, 1997), source familiarity, memory, listener expectations, cue plausibility (Rakerd and Hartmann, 1985) and vision (Shelton and Searle, 1980).

Specifically, in the horizontal plane, localization depends mainly on information of interaural differences of time, intensity and spectrum. A sound on the right side of the head will reach the right ear (near ear) a bit sooner than it will reach the left ear (far ear) resulting in an interaural time difference (IT). Interaural time differences occur because sound arrives fractionally sooner at the ear nearer the sound source.

Interaural intensity differences (II) occur because the sound intensity decreases with distance. As a result the sound that is farther from the sound source is also less intense. The spectrum of the signal will also be a slightly different spectrum because the head-shadow will selectively attenuate the higher frequencies at the further ear (Middlebrooks, 1992).

Interaural time differences (IT's) are readily predictable in a free-field environment and relatively frequency independent over large ranges of frequency (Kuhn 1977, 1987). Interaural intensity differences (II's) are frequency-dependent because of the differential effect that the head and body have on low versus high frequency sounds. Low frequency sounds have long wavelengths and bend around the body while high frequency sounds are attenuated by the head. Differences in spectrum also have some role in horizontal localization, especially in discriminating front from rear sounds. For example if a sound is on one side in the front of a listener, the high frequency components of the signal received at the near ear will have a relatively greater intensity than they would if the same sound were at the same distance from, but to the rear of the listener (Musicant and Butler, 1984).

The auditory system is highly sensitive to IT at frequencies below 1000 kHz; above 1500 Hz the auditory system is relatively insensitive to IT. Sensitivity to II is roughly constant with frequencies. In terms of interaural cues that reach the ears in normal soundfield listening, IT dominates over

II at low frequencies (Wightman, 1992). At high frequencies, II is the dominant cue with interaural spectral differences as a secondary cue.

There is a direct connection between the ability to localize sounds and the ability to hear speech in noise (Hirsh, 1950). A difference in spatial location that usually occurs between a desired speaker and competing noise usually results in improved intelligibility. Dirks and Wilson (1969), for example, showed higher intelligibility scores for both spondaic and phonetically balanced (PB) words in noise when the sources of the speech and the noise were spatially separated. This is the perception of normal hearing people and many hearing-impaired listeners as the ability to single out a specific talker from the noisy background babble at a party.

Many speech communication systems and amplification systems are limited to a single microphone input. The separation of speech from noise is more difficult under monotic or diotic listening conditions. It is possible to process speech and noise picked up by a single microphone so as to create an apparent spatial separation between the speech and noise. Signal processing of this type has the potential for making speech in noise more intelligible. The purpose of this study is to evaluate a newly developed processing system of this type.

TERMINOLOGY

Antiphasic

When the signals at the two ears are in opposite phase.

Binaural Fusion

The phenomenon when a single sound is heard for two separate inputs, one to each ear.

Binaural Interaction Advantage

The ability of the binaural auditory system to enhance the separation of target from interference by making use of interaural differences in the received sounds.

Binaural listening

Listening with two ears.

Binaural Masking Level Difference (BMLD or MLD)

This is a measure of the improvement in detectability of a signal that can occur under binaural listening conditions. It is the difference in masked threshold (in dB) for the case where the signal and masker have the same phase and level relationships at the two ears and the case where the interaural phase and/or level relationships of the signal and masker are different.

Diotic Listening

Acoustic signal delivered identically to both ears.

Dichotic Listening

Acoustic signal delivered to both ears with one or more interaural differences.

Homophasic

When the signals at the two ears are identical in phase.

Interaural Intensity Difference (II)

Interaural Intensity Difference is the difference in sound intensity between the two ears.

Interaural Time Difference (IT)

Interaural Time Difference is the difference in time between the two ears.

Interaural Intensity and Time Difference (ITI)

Interaural Time Difference is the difference in time between the two ears.

Lateralization

The process when the apparent lateral position of a sound in the head is identified.

Localization

The process when the apparent source location of a sound is identified in the soundfield.

Monaural Listening

Listening with one ear.

Monotic Mode of Presentation

Acoustic signal delivered to one ear only.

Monophonic System

Audio system that has only one acoustic channel (e.g. only one loud speaker).

Stereophonic System

Audio system that has two acoustic channels driving two transducers (loudspeakers, earphones) with between channel differences that create apparent differences in spatial location among the components of the acoustic signal.

Suffixes o and π:

The suffix o indicates the homophasic condition; the suffix π indicates the antiphasic condition. For example, SoNπ indicates that the signal is binaurally in-phase (homophasic condition) and that the noise is π radians out-of-phase between ears (antiphasic condition).

CHAPTER TWO

REVIEW OF THE LITERATURE

The first half of this review is concerned with the effect of interaural differences on the localization and lateralization of sound, on release from masking, and on subjective advantages resulting from interaural differences, binaural fusion and localization. The second half of this review deals with signal processing for noise reduction.

Under normal listening conditions the signals reaching our ears are similar but not identical; yet we typically hear one sound. This phenomenon in which a single sound is heard for two separate inputs is known as binaural fusion. In the laboratory a binaurally fused image can be created by means of headphone listening. In the most basic case, two identical stimuli are applied, one to each ear, (i.e. diotically). Under these conditions a normally hearing listener almost always claims to hear a single, fused sound image within or near his/her head and close to or on the median plane (i.e., roughly in line with his/her nose). (Licklider, 1948; Hirsh, 1948)

Stereophonic listening via loudspeakers also produces a binaurally fused sound image. In this case, two or more loudspeakers in spatially separate locations are used. The binaural sound image produced in this way is not perceived between the ears or near the head, but is located in space, usually at some location between the loudspeakers. Stereophonic sound images are perceived as being externalized while those produced by headphone listening are usually internalized (i.e., perceived either between the ears or near the head). Recent research has shown that it is possible to create the perception of an external sound image (e.g. as in surrounding speakers). In order to do this, it is necessary to simulate, by electronic means, the sound transmission paths from each loudspeaker to each ear (Wightman and Kistler, 1989). This aspect of binaural listening is beyond the scope of the studies reported here.

Since the location of an internalized sound image is usually perceived on the left-right axis it is convenient to specify the position of an internalized sound image in terms of its position on that axis. Even in the case of externalized sounds, lateral position is usually the most salient cue, height is usually less salient and depth is difficult to perceive reliably. Jeffress (1961) has shown that judgment of lateral position of external images and lateralization of internal images are highly correlated for similar internal differences. This study supported the use of headphone listening in studying the effect of internal differences on localization.

If an interaural time difference is introduced, the apparent location of the binaural sound image is shifted away from the ear receiving the delayed signal. For sinusoidal stimuli, an interaural time difference (IT) is equivalent to an interaural phase difference (IP).

i.e., the interaural phase difference (θ) is equal to (Δt / T) • 360

Where Δt is the interaural time difference and T is the period of sinusoidal signal.
Since the frequency f is equal to 1/T,

$$\theta = \Delta t \cdot 360 \cdot f$$

NOTE: Phase is cyclical and more than one Δt corresponds to θ
e.g. if $\theta = 90°$ for a 1000 Hz tone, Δt could be 0.25 msec or 1.25 msec or 2.25 msec, etc.

Cherry (1953) demonstrated that systematically increasing interaural phase difference produces cyclic changes in apparent lateral position from the center of the head to one or the other extremity and back again. The interaural phase difference for the extreme condition is equal to one-half the period of the signal (i.e., an IPD of 180 degrees or π). An interaural phase difference (or time delay) equal to the period brings the tone back to its median plane position. For aperiodic stimuli such as pulses, increasing the interaural time delay beyond 2 or 3 msec causes the binaural sound image to split into two parts, a separate sound being heard in each ear (Zwislocki, 1956). For speech, a larger interaural time delay of up to 15 msec can be tolerated before the binaural sound image is split (Cherry and Taylor, 1954).

LOCALIZATION

Sensitivity has also been measured in terms of the minimum audible angle. Tobias and Curtis (1959) showed an improvement of speech intelligibility in the out-of-phase condition over the in-phase condition by an amount equivalent to a 5dB improvement in signal-to-noise ratio. Sayers (1964) indicated that the perceived location of a tone varied over frequencies was relatively unaffected by frequency (up to about 1200 Hz). Sayers (1964) indicated that the midline perceived location of the sound image at 0° azimuth to be diffuse (appearing on both sides of the head) as the interaural phase shift approaches 180°.

Watson and Mittler (1965) showed that the movement of the sound image is nearly a linear function of dichotic intensity-difference from 0 to about 4.5 dB, while large differences in intensity have a decreasing incremental effect on the sound image. Yost (1981) measuring interaural intensity differences and perceived location for a variety of overall levels and stimulus durations, and a wide range of frequency showed similar results as Watson and Mittler (1965). As the tonal frequency increases beyond 1,200 Hz, interaural phase or interaural time did not appear to influence the lateral position of sinusoids. The smallest detectable interaural time difference (the just noticeable difference or jnd) varies with the type of signal employed. Jnds's as low as 15 to 20 microseconds have been reported for pulses. For tones, the jnd for an interaural time difference rises as a function of frequency. Zwislocki and Feldman (1956) observed a minimum jnd on the order of 30 to 50 microseconds at a frequency of 300 Hz, rising to about twice this value at higher frequencies.

Interaural intensity differences also produce a change of the apparent lateral position of a binaurally fused sound image. There is a simple, monotonic relationship for all stimuli (including pure tones) between apparent lateral position and interaural intensity difference. The larger the interaural intensity difference is, the greater the degree of laterality. Mills (1958) showed that the jnd for interaural intensity differences is about .5 to 1.0 dB for tones. For extremely large interaural intensity differences the signal will be below threshold at one ear and heard only in the opposite ear.

BINAURAL RELEASE FROM MASKING

Hirsh (1948) demonstrated that the threshold of detection for a signal in a background of noise was lower if the signal was presented with different interaural phase characteristics than that of the noise masker. For instance, a signal presented with an interaural phase difference of π radians and masked by a noise that is binaurally in phase ($S_\pi N_0$) has a masked threshold that can be as much as 15 to 18 dB lower than one obtained when both the signal and masker are binaurally in phase (S_0N_0). Other interaural differences such as S_0N_π also yield an improved detectability. Reversing the interaural conditions, i.e., S_0N_π, yields a masked effect that is less than S_0N_0 but higher than $S_\pi N_0$. Other binaural conditions yielding smaller improvements in detectability are S_mN_0 and S_0N_m where the Speech and Noise, respectively, are presented to one ear only and S_0N_u and $S_\pi N_u$ where uncorrelated noise is delivered to each ear. Durlach (1963) has developed a general theory for predicting the amount of release from masking for each of these conditions. Binaural release from masking, as this effect is known, also occurs for the case of speech in noise although the magnitude of improvement is smaller. Licklider (1948), in a companion study to Hirsh's (1948) showed binaural improvements in speech intelligibility. The binaural release from masking for speech was less than for tonal signals (on the order of 3 to 6 dB, depending on the signal-to-noise ratio). He also discussed the introspective reports of the listener concerning the apparent locations of the speech and noise. Both the speech and the noise appeared to be at the center of the head for the S_0N_0 condition, but that for the $S_0N\pi$ condition, only the speech was located near the center of the head while the noise was heard at the two ears. The case in which speech and noise occupied different spatial locations also corresponded to improved speech intelligibility.

Hirsh (1950) found that similar improvements in speech intelligibility would occur if signals were sent to two loudspeakers differing in location. Carhart et al (1967) also investigated the relative importance of spatial factors. They measured thresholds for identifying one- and two-syllable utterances in the presence of four simultaneous maskers. Two of the maskers were modulated white noise and two of the maskers were whole sentences. They used several different listening conditions, including homophasic, antiphasic and conditions where the signal or the maskers were delayed at one ear relative to the other. In these latter conditions, the different maskers were sometimes given opposing delays, so that some would be located towards one ear and some towards the other. They found that when the speech and noise were spatially separate, intelligibility improved by 13 dB. The improvement was greater for bisyllabic words.

In addition to experimental studies showing improved intelligibility when speech and noise differ in terms of their interaural time or intensity differences, there are many subjective reports of the advantages of binaural listening. Koenig (1950) and Haas (1972) have reported, informally, that speech is easier to listen to binaurally than monaurally, Libby (1980) describes these advantages as: increased ease of listening, increased spatial balance, improved sound quality and increased listening success in difficult listening situations. Levitt and Rabiner (1967a) investigated the factors affecting release from masking for speech. They showed that the amount of release from masking depended on the speech-to-noise ratio as a function of frequency. They also developed a theoretical model (Levitt and Rabiner, 1967b) for predicting release from masking for speech from psychophysical data on binaural tonal masking. These predictions accounted for the diversity of experimental findings on release from masking for speech.

EASE OF LISTENING

Several researchers have identified listening effort or ease of listening as a dimension of interest in hearing aid research (Hafter and Schlauch (1992); Saunders and Levitt (1991); Feurerstein, 1992; Preminger, 1995; Humes, 1997). Listening effort has been described as the amount of effort a subject needs to give to the listening task in order to understand as much of the speech as possible.

Feuerstein (1992) conducted a study in which normal hearing listeners rated their perceived ease of listening and attention effort in monaural versus binaural listening conditions. Word recognition was also evaluated. This study showed no significant correlation between ease of listening and attention effort, however a significant correlation was found between perceived ease of listening and word recognition. Forty-eight subjects listened to list of sentences from the Revised Speech In Noise Test (SPIN) (Bilger, et al 1984) in a sound treated room. The sentences were presented from one loudspeaker and the multitalker background noise for each list was presented from a second loudspeaker located at an equal distance from the center of the subject's head. Perceived ease of listening was assessed using a modified direct estimation technique, as described by Geller and Margolis (1984). Subjects were required to rate ease of listening for each sentence list using a scale of 0 to 100, with '0' being defined as very, very difficult and '100' being defined as very, very easy. Results indicated that binaural listening was judged on average to be the easiest, with monaural-near ear listening to be the next easiest, and monaural-far ear to be the most difficult. An analysis of variance for perceived ease of listening indicated a significant main effect (F=149.95; p<0.01) for listening condition. Product- moment correlation coefficients indicated a significant and positive correlation (0.75) between ease of listening and word recognition, and significant negative correlations between ease and effort (-0.35) and between discrimination and effort (-.38).

Humes, et al (1997,1999) substantiated the use of ease of listening as an outcome measurement. Humes et al (1997) used ease of listening as one of several objective and subjective measures to evaluate the benefits of different circuitry provided by a hearing aid in a prospective, double blind, crossover controlled study. In the first study, Humes, et al (1997), 110 subjects evaluated the experimental Bill-processing option in ITC hearing aids. There were four listening conditions consisting of two different presentation levels (60 and 75 dB SPL) and two signal-to-noise ratios (+5 and +10 dB). After listening to passages from the Connected Speech Test and the NU-6, the listener rated 'ease of listening' on a 0-100 scale, with 100 representing extremely easy and 0 being extremely difficult. Percent-correct performance scores for both the NU-6 and CST tests were also obtained based on 50 items. An analysis of variance of the NU-6, CST scores and a magnitude estimation of listening effort test showed no statistically significant (p<. 05) effect of hearing loss group and no difference in performance between the linear and BILL settings.

In 1999, Humes et al, evaluated the binaural performance and benefit of a two channel wide dynamic range compression hearing aid and a linear hearing aid. Again, ease of listening and word recognition testing were two of the outcome measures that were used. Listeners rated the difficulty of listening to speech in babble background noise and in quiet using similar criteria as described above. Ease-of-listening results were found to be significantly enhanced with both hearing aids as opposed to unaided listening.

Preminger and Van Tassell (1995) used ease of listening and intelligibility as outcome measures in two experiments. The subjects rated five speech quality dimensions (intelligibility, pleasantness of tone, loudness, listening effort and total impression) as a function of changes to the frequency response of a listening system.

Four normal hearing subjects evaluated 18 different frequency response conditions in which ratings of intelligibility varied over a range, from 25 to 100. For 9 conditions, the low band was held constant at the output level, where intelligibility was rated 25; the high band was presented at the output level, where intelligibility was rated 0, and then increased in 4-dB steps. For the other 9 conditions, the low-and high frequency bands were adjusted in reverse. Ratings of Listening effort and intelligibility were highly correlated for both the low frequency band (.95) and high frequency band (.91).

Subjective judgements are being used increasingly in evaluating acoustic amplification systems. These judgments involve category ratings, magnitude estimation and paired comparisons. An important advantage of rating methods is the relative ease and high efficiency with which evaluation data can be obtained. Subjective ratings, however, have their limitations; for example, criterion changes, bias and high test-retest variability. An alternative approach is the use of paired comparisons, which has a lower test-retest variability. This technique has been extensively used by Neuman (1987), Neuman et al (1995), Kuk (1995) and others. Paired comparison judgements provide a viable alternative to simple subjective measurements or rating scales (Eisenberg and Dirks, 1997).

The psychophysical scaling procedures of paired comparisons and category ratings have been directly compared by Purdy and Pavlovic (1992) and Eisenberg and Dirks (1995 and 1997). Purdy and Pavlovic (1992) compared the methods of paired comparison, category ratings and magnitude estimation in elderly adults who had hearing essentially within normal limits up through 2000 Hz. Judgements of perceived intelligibility were obtained for band-pass filtered speech using stimulus items that produced a monotonic increase in predicted intelligibility scores based on Articulation Index (AI) theory (French and Steinberg, 1947). The intelligibility judgements were shown to be highly correlated with the AI estimates of intelligibility for the three procedures. Although test retest reliability was shown to be marginally poorer for paired comparisons when compared to magnitude estimation, category ratings or with an objective measure of word recognition, sensitivity was equivalent among the three scaling procedures and the word recognition measure in differentiating between the band-pass filtered conditions.

Eisenberg and Dirks (1997) compared category ratings and paired comparison judgements. In their first experiment twelve normal hearing subjects made subjective judgements of the speech clarity of sentences from the Pediatric Speech Intelligibility Test (Jerger, 1984) processed by six band pass filters that increased monotonically in Articulation Index (AI) estimates. Three of the filters were widely spaced (Filters 1,3 and 6) and three filters were closely spaced (Filters 4,5, and 6). Results indicated that speech clarity judgements obtained by category ratings and paired comparisons were highly related to the AI estimate both for the normal and hearing-impaired individuals. The two scaling procedures were compared directly by converting paired comparison judgements into z scores, as described by Thurstone (1927), and then correlated with the category ratings. The six-filter and three-filter widely spaced tests were shown to have a high correlation (r>0.90). The three filter closely spaced tests, however, showed a reduced correlation (r=0.568) presumably reflecting the divergence in judgements observed between the two psychological scaling procedures for that particular condition. Specifically, the category ratings displayed a narrow range of high values across the three filters in contrast to the paired comparisons, which displayed a broader range. These results indicated that paired comparisons were more sensitive than category ratings for differentiating among more subtle acoustic differences.

SIGNAL PROCESSING FOR NOISE REDUCTION

There have been many attempts at reducing background noise in audio systems. The general problem of noise reduction is not new and has been addressed in great depth by statisticians, physicists, engineers, and others (Lim and Oppenheim, 1979; CHABA, 1989). The problem is central to the fields of Information Theory and Coding Theory. As a consequence, there is a substantial body of theory and methods of practical implementation that address the problem.

Noise is defined as any unwanted signal that interferes with a desired signal. There are three types of noise that are particularly damaging to speech intelligibility:

1. Random noise with an intensity-frequency spectrum similar to that of speech.
2. A second interfering voice. Note that the interference produced by many other voices of roughly equal intensity (known as speech babble) has physical characteristics similar to that of random noise with a speech-shaped intensity-frequency spectrum.
3. Substantial room reverberation. Reverberation is produced by sound being reflected off walls, floors, ceilings, tabletops, and other hard surfaces. Some reverberation is

helpful in reinforcing the speech signal, but too much reverberation will reduce speech intelligibility (and overall sound quality), particularly in the presence of other types of noise.

Advanced signal-processing techniques for noise reduction open up substantial new possibilities. There are two general principles that are helpful when identifying possible ways of addressing the problem of speech and noise. These are:

1. The more we know about the acoustic characteristics of the speech and noise, the more we can do to reduce the effects of the noise on the speech.
2. The larger the differences (i.e. spatial, spectral and temporal) between the speech and the noise, the more we can do to reduce the effects of the noise on the speech.

There are several different approaches to the problem of noise reduction. One approach is that of adaptive filtering (or frequency-dependent amplitude compression). In this approach estimates of the speech and noise spectra are obtained and those frequency bands in which the noise level exceeds the speech level are attenuated (Graupe, 1987). This approach can also be used to reduce reverberation by identifying the frequency bands with excessive reverberation and then attenuating those bands (Neuman and Eisenberg, 1991).

A practical problem in implementing the above approach is that of obtaining reasonably accurate estimates of the speech and noise spectra as they vary over time. One approach to this problem is to measure the noise spectrum during pauses or other short breaks in the speech signal and to obtain the speech-plus-noise spectrum when speech is present (Allen, 1977). Since these spectra are obtained over short interval of time, they are known as short-term noise spectra. If the short-term noise spectrum does not vary rapidly with time, the frequency bands in which noise levels exceed speech levels can be attenuated. (Gagne, 1988).

The mathematical theory of filtering provides a formula for an optimum filter that will maximize the signal-to-noise ratio (Wiener, 1949). This filter, known as a Wiener filter, requires that the spectra of both the signal and the noise do not vary with time- a requirement that clearly does not apply to speech, or to many noises. Many speech sounds, however, have spectra that are approximately constant over short intervals of time. It is thus possible to use a short-term Wiener filter in which the spectra of the speech and the noise are assumed not to vary significantly over short intervals of time. The potential gain in the speech-to-noise ratio, assuming the validity of this assumption, is relatively small, however. Thus far, Short-term Wiener filtering for speech in random noise has not proven successful for people with normal hearing (Levitt, et al, 1993).

A variation of the above approach is to take the short-term noise spectrum obtained during a pause in the speech and subtract it from the short-term speech-plus-noise spectrum when speech is once again present (Weiss and Aschkensay, 1975; Boll, 1975). This technique takes into account time-varying changes in the short-term speech spectrum but still assumes that the short-term noise spectrum does not vary significantly with time. This technique, known as spectrum subtraction, can improve speech-to noise ratios for many commonly encountered ambient noises by as much as 10 to 12 dB, but without a concomitant improvement in speech intelligibility. The signal processing involved produces audible distortions, referred to as processing noise, which counteract the potential improvements in intelligibility resulting from the reduction of background noise.

Another approach to the problem that produces a much-improved speech-to-noise ratio and improved sound quality but no significant change in intelligibility is that of sine wave modeling. Sinusoidal modeling (McAulay and Quatieri, 1986, 1989) is a speech analysis/ synthesis procedure in which the signal is reproduced using a limited number of sinusoidal components. In the signal analysis portion of the processing, the incoming signal is divided into overlapping segments. Each segment is multiplied by a Hamming (raised cosine) window, and the spectrum is then computed using the fast Fourier transform (FFT) algorithm. The indicated number of peaks is selected from the magnitude spectrum, and the frequency, amplitude, and phase of each peak are saved. The output signal is then synthesized by generating one sinusoid for each selected peak using the measured frequency, amplitude and phase values. The formulation of sinusoidal modeling proposed by McAulay and Quatieri (1986) used a large number of sinusoids to represent the speech, with approximations made in encoding the information about the peaks in order to reduce the data rate in a digital speech-coding system. McAulay and Quatieri (1986) used for much of their work a 10 kHz sampling rate with new segments acquired every 5 msec. They concluded that 40 to 60 sinusoids would result in a reconstructed speech signal virtually indistinguishable from the original.

In the case of sinewave modeling, the major peaks in the speech-plus-noise spectrum are obtained. These peaks, which are frequently located at the harmonics of voiced speech sounds, consist mostly of speech and relatively little noise. The spectral components between these peaks, which consist mostly of noise, are discarded. The spectral peaks are then converted back to a time waveform with a much-improved speech-to-noise ratio (approaching 12 dB for speech in white noise), but with some processing noise.

Noise suppression using sinusoidal modeling is related to other techniques that have been proposed for improving speech intelligibility in noise. The common feature in these techniques is the preservation in the short-time spectrum of isolated frequencies or narrow frequency regions identified as containing speech information and the suppression of those frequency regions identified as containing predominantly noise.

It should be noted that the peaks in the long- term average speech spectrum convey important cues for speech perception. These peaks reflect either the harmonic structure of the speech signal at low frequencies (particularly for female and children voices) and the formant structure at higher frequencies. It has been shown that the information contained in the formants is sufficient for speech intelligibility. The harmonic structure conveys important prosodic cues as well as the gender, age, emotional state and related information regarding the speaker. The sinewave model of speech captures both harmonic and formant information efficiently. It has been suggested by Kates (1994) that sinusoidal modeling may not be limited to the same degree as the above-mentioned processing techniques because the undesired frequencies can be suppressed with a minimum of processing artifacts, resulting in a high quality signal.

Kates (1991) evaluated the potential benefits of sinusoidal modeling. The modeling he used differed from the McAulay and Quatieri (1996) model by reducing the number of spectral peaks used in the speech analysis while using exact information about the sinusoids for the signal reconstruction. Speech materials were low-pass filtered at 7.5kHz and sampled at 20 kHz using a 12-bit analog-to-digital converter. The windowed data segments were 25.6 msec in duration (512 samples) and a new segment was acquired every 9.6 msec (192 samples) for an overlap of 62.5%. A study was performed on a group of subjects with normal hearing and impaired hearing. Nine subjects compared unprocessed sentences with sentences modeled using 16, 8 and 4 sinusoids. Kates found that the ratings of perceived intelligibility by subjects with normal and impaired hearing were very similar both in quiet and at a 5 dB SNR. The competing signal added before processing was multi-talker speech babble. Both groups of subjects judged the original unprocessed material to be the most intelligible in quiet, but judged the original 16-sinusoid version of the materials to be equally intelligible at the 5 dB SNR.

Kates (1994) evaluated the effectiveness of using sine wave modeling as a noise reduction technique in two additional experiments. Five adult subjects with normal hearing were included in the study. The first experiment was consonant recognition using isolated speech tokens and the second experiment was a comparative measure of perceived intelligibility for the same test material. The speech stimuli were isolated consonant-vowel (CV) and vowel-consonant (VC) tokens read by a male talker. Thirteen consonants were used (p, t, k, b, d, g, h, th, s, sh, v, voiced th, and z). The vowel contexts were consonant +/a/, /a/+ consonant. The tokens were presented at a +25 dB or a +5 dB speech to babble ratio. Consonant- recognition performance was measured as a function of the presentation level. Perceived intelligibility for the same set of CV or VC tokens was also obtained using a round robin paired comparison technique. Analysis of the data revealed that a reduction in the number of sinusoids resulted in a reduced speech intelligibility both in quiet and in noise. Although the speech-to-noise ratio was increased, by the sine wave model processing, the process of discarding the low-level spectral components resulted in an audible-processing distortion that counteracted the effect of an improvement in speech-to-noise ratio.

Balachandran and Levitt (1998) have shown that the sinewave approximation is barely distinguishable from the original speech using only 16 sinewave components. However, the method of reconstructing the signal was different in this experiment from that of Kates (1994). Balachandran and Levitt used an overlap-add procedure rather than a linear interpolation between adjacent segment peaks. In this study the waveform was analyzed in a series of Hamming weighted windows at regular overlapping time intervals. The analysis and waveform reconstruction were performed independently for each time window. The reconstructed windows were then added together with appropriate time interval weighting to ensure no perceptual discontinuity at the boundary between adjacent windows.

The focus of their study was to determine the number of spectral peaks needed for a sinewave model of speech to be perceptually indistinguishable from the unprocessed speech signal. A three-interval oddity task was used to determine if the processed speech could be distinguished from the original speech signal. Three stimuli were presented on each trial. In each case, one of the three stimuli differed from the other two. For example, the odd stimulus could be the sinewave model of the speech signal with the other two stimuli being the unprocessed version, or vice versa. The order of the stimuli was randomized for each trial. An adaptive test procedure was used in which the number of sinewave components was varied until the subject could just distinguish between the processed speech 70.7 percent of the time. The test stimuli consisted of VCV nonsense syllables for all permutations of the fifteen consonants /p/, /t/, /k/, /f/, /th/, /s/, /sh/, /m/, /b/, /d/, /g/, /v/, /(v)th/, /z/, /n/, /ng/ with the vowels /a, i, u/.

The results revealed that two to four sine waves were adequate to recognize individual consonants and that the sinewave approximation was barely distinguished from the original speech when using 16 components. For some subjects the 16-component sine-wave model contained a subtle, low level processing distortion, which enabled them to distinguish between the original and processed speech signals; therefore, as many as 60 peaks were needed for the sine-wave approximation to be indistinguishable from the original speech signal. The use of 16 sinewave components was nevertheless considered to provide a good practical approximation to the original speech signal since it sounded quite natural and was barely distinguishable from the original.

The method of sinewave modeling provides a means for identify perceptually important components of speech signal. In the case of speech in noise the spectral peaks consist mostly of speech provided the speech to noise ratio is greater than -5dB

(Levitt, personal communication). The remaining spectral components (the Residual) consist mostly of noise. Levitt (personal communication) has developed a technique using sinewave modeling in which major spectral peaks are presented binaurally with interaural time and intensity differences such that the sound image associated with the spectral peaks are heard towards one side of the head. The residual is also presented binaurally, but with the opposite interaural time/intensity differences so that the sound image associated with the residual is heard towards the opposite side of the head. The peaks sound image is heard as speech and the residual sound image is heard as noise. Since the lower level spectral components are not discarded there are no audible signal-processing distortions, as was the case in the studies by Kates (1991, 1994).

The purpose of the present investigation is to evaluate the effect of the above method of signal processing on the apparent spatial locations of speech and noise and on ease of listening. Since the above method of signal processing is designed to create the illusion of a spatial separation from a monotic signal, it is referred to as a Simulated Binaural Listening System, as opposed to a true binaural listening system in which the interaural characteristics of the speech and noise are controlled independently before being added.

CHAPTER THREE

METHODS

This chapter will describe the stimuli, the equipment, the subjects and the procedure used in the experiment. Computer programs were written to process the speech and noise stimuli, to present the stimuli and to record and analyze the results. The processing technique engaged two principal operations at the outset. The first operation introduced an interaural intensity difference of selected components of the signal. The interaural intensity difference was defined as the difference (in dB) of the root mean square (RMS) levels between the signal reaching the right ear and the signal reaching the left ear. The second operation introduced a difference in the time of arrival of selected components of the signal between the left and right ears. This was achieved by adjustment to the phase spectrum.

3.1 THE STIMULI

The stimuli were derived from selected subtests of the Hearing in Noise Test (Nilsson, Soli, and Sullivan, 1994) available in a stereo CD format (Starkey, 1996). The lists of sentences are recorded on one track, and noise filtered to match the long-term average spectra of the sentences is recorded on a second track. The full Hearing in Noise Test (HINT) consists of

250 digitally recorded sentences organized into 25 phonetically balanced lists of 10 sentences. The sentences are derived from the Bamford-Kowal-Bench sentences developed in the United Kingdom (Bench and Bamford, 1979) and are at approximately a first grade reading level. The sentences are recorded by a male talker and are equated for naturalness, length and 50 percent intelligibility in the presence of their associated noise channels. For this experiment, sentence Lists 1,11,15,19,22 and 24 were used.

CONSTRUCTION OF THE EXPERIMENTAL SIGNAL SOURCE

STEP 1: ACQUISITION OF THE HINT SENTENCES

The instrumentation used to access the two-track recordings was an IBM PC and a series of MATLAB programs. As noted above, the initial stimulus materials were stereo files (speech on track 1 and noise on track 2) on a CD disk. Since software enabling the direct digital transfer of CD recordings to files stored in a computer was unavailable, it was necessary to first convert the CD recordings to analog form (i.e., play them) and then digitize them to be stored in two-track disk files in a PC computer. The signal levels of the sentences and their spectrally matched noise tracks were calibrated by means of the calibration tone provided for that purpose on the HINT CD. After the 1500 Hz calibration tone had been digitized, the RMS levels of both the speech and noise tracks were adjusted to equal the RMS level of the calibration tone.

STEP 2: CREATION OF MONAURAL FILES AT FOUR SIGNAL-TO-NOISE RATIOS

The computer-stored two-track speech plus noise files from the HINT CD were combined to make a set of single-track files at different speech-to-noise (S/N) ratios. Four single-track files were generated in which the speech was added to the noise at S/N ratios of -4dB, -2dB, 0 dB, and + 2dB. Each file was stored at a resolution of 16 bits. Consequently, to achieve maximum storage efficiency, the mean RMS level of each single-track file was adjusted to one sixth of the 16 bit dynamic range. Extreme peaks in the time waveform were then clipped so as not to exceed a range of ±3 times the RMS level. This was done to avoid peak clipping. Peak clipping at 3 times the mean RMS level occurred in frequently and was not audible.

STEP 3: PROCESSING of the STIMULI

The single-track speech-plus-noise stimuli represent the noisy speech signals to be used in the evaluating the simulated binaural listening system. The first stage of processing required that the spectrum of the speech plus noise be obtained. Each single track was subdivided into a sequence of time windows. Each window contained 1024 samples, which corresponds to 46.4 msec of sound at a sampling rate of 22,050 Hz. A FFT, carried out on each window, generated a short-term spectrum consisting of 512 spectral lines per window covering the range from 0 to 11,025 Hz. Each window was multiplied by a Hamming weighting function to improve spectral resolution. Overlapping windows were used to eliminate audible between-window discontinuities. Adjacent windows overlapped by 960 samples (i.e., a 94% overlap).

A fixed number of spectral peaks (N) were selected for each window. They were selected after the high frequencies of the speech had been boosted to compensate for the characteristic high frequency roll off that occurs in the average speech spectrum. Thus, the high frequencies were boosted at the rate of 9-dB/octave beginning at 500 Hz. This pre-emphasis function was selected so as to whiten the idealized speech spectrum as specified in ANSI S3.5-1997. Without pre-whitening, most of the highest peaks would normally be found in the low frequency region of the spectrum. The use of pre-whitening thus increases the high frequency content of sine-wave models of speech when employing a small number of peaks thereby capturing both low and high frequency information in the speech signal with a limited number of components. Note that after the required number of peaks had been selected, the inverse of pre-whitening (un-whitening) was used so as to ensure that the sine wave model of the speech had the same spectral shape as the unprocessed signal. The time waveform was then reconstructed by performing an inverse FFT on each processed window and then adding the overlapping time windows.

The spectral peaks were selected from the amplitude spectrum in the following manner. It was determined from previous studies (Balanchanrian and Levitt, 1998) and pilot studies performed by the author that 16 peaks provided an approximation to the speech signal in quiet that was virtually indistinguishable from the unprocessed speech signal. In order to improve the reduction of background noise, only those spectral peaks lying above a specified threshold level were selected for processing. This threshold level was 20 dB below the level of the highest peak. The 16 largest peaks in a frame were then selected starting with the highest spectral peak working downward in order of magnitude until the required number of peaks was obtained. In some frames, fewer than 16 peaks met the 20 dB criterion; it was thus not always possible to select 16 peaks per frame consistently throughout a whole sentence.

A two-track processed speech and noise file was then generated for each unprocessed single-track speech plus noise file. The sine waves corresponding to the 16 largest peaks were delayed by ΔT microseconds and attenuated by ΔI dB in one track and advanced by ΔT microseconds and increased in level by ΔI dB in the second track. The remaining spectral components were then added to the two tracks but with the opposite values of ΔT and ΔI.

When the processed signals on the two tracks are delivered binaurally to a listener the audio signal corresponding to the 16 highest-peaks (which consists mostly of speech) is heard on one side of the head. The audio signal corresponding to the remaining spectral components (which consists mostly of noise) is heard on the opposite side of the head.

Three combinations of ΔT and ΔI were used in the experiment. These were:

The Interaural Time (IT) condition:

ΔT=300 microseconds, ΔI=0dB

The Interaural Intensity (II) condition:

ΔT=0 microseconds, ΔI=3.0 dB

The Interaural Time plus Intensity (ITI) condition:

ΔT=300 microseconds, ΔI=3.0 dB

A MATLAB computer program was used to generate interaural time differences of 300 microseconds and the interaural intensity differences of 3 dB by manipulating the phase vector and the relative gain by applying an appropriate multiplier. Thus, the signal composed of the selected peaks that exceeded the 20dB threshold in the whitened FFT was given a positive phase increase in the right ear and a phase decrease of the same size in the left ear. The formula is given below,

Phase increase or decrease = $(2.\pi\ .f.\ \Delta T)/2$
(at frequency fHz)

Where ΔT is the interaural time delay (in seconds) and f is the frequency (in Hz) of the spectral component being delayed. Note that the phase shift is divided by 2 since the total phase shift is the sum of the two phase shifts one to each ear.

The "noise" signal consisting of all the non-peak FFT terms (in addition to peak terms that failed to meet the 20 dB threshold criterion) was given the same interaural phase shift in magnitude, but of opposite sign; i.e., the noise signal was delayed to the right ear and advanced in phase to the left ear.

Interaural intensity differences were introduced in the same way except that intensity rather than phase was manipulated. For an interaural intensity difference of 3 dB, the spectral peaks to the right ear were increased by 1.5 dB and decreased by 1.5 dB to the left ear. The remaining spectral components were decreased by 1.5 dB to the right ear and increased by 1.5 dB to the left ear.

A prominent artifact of the processing method is a distortion product that stems from an unavoidable lack of frame-to-frame coherence in the frequency locations of the selected peaks. The artifact is most noticeable when a speech signal composed exclusively of 16 peak FFT terms per analysis frame is heard. However, if the non-peak FFT terms are also included at a fraction of their original amplitude, a reduction in the perceived artifact can be achieved at the expense of a small loss in separation between the signal due to peaks and the rest of the FFT terms or "noise". The method that was adopted to control the fraction of non-peak FFT amplitude that would be included employed a vector of frequency values and a fractional parameter. In each analysis frame, this vector contained the frequency locations of all the non-peak FFT terms in that frame. The vector was called a mask because, when the parameter (a multiplier) was set to zero, the amplitudes of all the non-peak terms heard in the right-ear channel could be masked out (i.e., rendered inaudible).

Two pilot studies were conducted to determine the value of the masking parameter. The first pilot study indicated that a masking factor of 0.7 provided a favorable reduction of artifact for some sentences for three of four subjects. One subject, however, noted a different artifact, a warbling at the end of words.

A second pilot study was conducted to determine which processing technique subjects found the least annoying. Three subjects used a paired-comparison technique to compare five processed sentences in three different conditions using a 0.7 masking factor to five processed sentences using a 1.0 masking factor. Subjects reported that there was a clear difference between speech and noise when the 0.7 masking factor was used. However in two of three sentences in all three conditions a different artifact (warble echo sound) was introduced. All three subjects chose the sentences with 1.0 as the masking factor as less annoying. Thus it was decided to use a masking factor of 1. When the masking factor is 1, it permits the full amount of non-peak energy to appear in the right channel. As a consequence, the RMS levels in the two channels differed by a small amount.

A more serious problem than the small difference in overall level between the ears was the perception of a significant loudness imbalance between the ears. This was because the sum of 16 spectral peaks plus the attenuated residual in the right ear was less loud than the sum of the residual plus the attenuated spectral peaks reaching the left ear. A final adjustment, boosting the signal to the right ear, was used to balance between ear loudness levels.

VERIFICATION OF SIGNAL PROCESSIING

Verification of the signal processing was conducted using an IBM THINKPAD Laptop computer containing the processed two-track speech plus noise files, and an oscilloscope (Tektronix

TDS460A). The oscilloscope was connected to the output receptacle of the IBM THINKPAD. A two-track wave file was played by the laptop computer and monitored on the oscilloscope with the output of the track for the right ear being displayed on the upper trace of the oscilloscope and the output of the track for the left ear being displayed on the lower trace of the oscilloscope.

STEP 1A: VERIFICATION OF INTERAURAL TIME DIFFERENCES (ΔT)

The waveform corresponding to the largest spectral peaks were easily identified on the two oscilloscope traces. The magnitude of the between trace time difference was then read off of the time scale of the oscilloscope and was found to be 300 microseconds, as expected.

STEP 1B: VERIFICATION OF INTERAURAL INTENSITY DIFFERENCE (ΔI)

Verification of interaural intensity difference was obtained by measuring the RMS levels at the output of each channel of the laptop computer. This was done first for the 16 spectral peaks and then for the remaining spectral components.

Further verification of the interaural time and intensity differences was obtained using KEMAR. Oscilloscopic traces and RMS levels were examined as before, but this time at KEMAR's ears using the headphones to be used in the experiment. The processed sentence "The neighbor's boy has brown hair," was used. Recordings with a noted $\Delta I = 12$ dB difference in amplitude were obtained. Similar waveforms except for amplitude were noted. A voltage ratio of 4 or a 12 dB between channel differences was recorded between the two recordings. A similar check was performed for interaural time differences.

CALIBRATION AND VERIFICATION OF THE TRANSDUCERS

Sennheiser HD 414 earphones were calibrated using a Knowles Electronics Manikin for Acoustic Research (KEMAR) (Burkhard and Sachs, 1975) in a soundproof room. Zwislocki couplers were mounted in KEMAR (Db4005 in the left ear and Db1000 in the right ear). The couplers were fitted with B&K ½-inch pressure microphone (type 4134). The output of each coupler was fed into a B&K power supply (type 2801) through a B&K Preamplifier (type 2619). The calibration signal was generated by a Pistonphone (type 4220). The sound pressure level at each coupler was measured using a B&K Type 1613 Sound Level Meter. The frequency response of the Sennheiser earphones mounted on KEMAR was recorded and compared to the manufacture specifications.

SUBJECTS

Ten normal hearing subjects with no significant history of ear disease were recruited for this study. There were seven females and three males. All subjects were native speakers of English with no reported hearing problems. All subjects were under 55 years. Nine of the subjects were right handed and one subject was left-handed. Subjects received an audiological evaluation. Each had pure tone thresholds better than 20 dB HTL at all frequencies between 250 and 8000 Hz.

PROCEDURE

Each subject was seated in a quiet room. The investigator explained the nature of the study to the subject. The subject then was asked to read and sign the informed consent form. The subject then was instructed about test procedures (see appendix for consent form and instructions). Test stimuli were presented binaurally through Sennheiser HD 414 earphones. Prior to earphone placement, the output level was set to a level previously determined as comfortably loud for a group of normal hearing subjects. The subjects were asked if the signal level was comfortably loud while listening to a set of unprocessed sentences at different signal to noise ratios. If the signal level was not comfortably loud, the subject was instructed to adjust the volume level on the computer to obtain a comfortably loud level. Three practice trials prior to the first experiment were given to each subject to assure that he or she understood the task.

TABLE 3.1: EXPERIMENTAL CONDITIONS

	CONDITION	SPEECH TO NOISE RATIO	INTERAURAL DIFFERENCES TIME / INTENSITY Microseconds/ dB
1	**NP**	+2 dB S/N Ratio	0 microsecs/0dB
2	**II**	+2 dB S/N Ratio	0 microsecs/3dB
3	**IT**	+2 dB S/N Ratio	300 microsecs/0dB
4	**ITI**	+2 dB S/N Ratio	300 microsecs/3dB
5	**NP**	0 dB S/N Ratio	0 microsec/0dB
6	**II**	0 dB S/N Ratio	0 microsecs/3dB
7	**IT**	0 dB S/N Ratio	300 microsec/0dB
8	**ITI**	0 dB S/N Ratio	300 microsecs/3dB
9	**NP**	-2 dB S/N Ratio	0 microsecs/0dB
10	**II**	-2 dB S/N Ratio	0 microsecs/3dB
11	**IT**	-2 dB S/N Ratio	300 microsecs/0dB
12	**ITI**	-2 dB S/N Ratio	300 microsecs/3dB
13	**NP**	-4dB S/N Ratio	0 microsecs/0dB
14	**II**	-4 dB S/N Ratio	0 microsecs/3dB
15	**IT**	-4 dB S/N Ratio	300 microsecs/0dB
16	**ITI**	-4dB S/N Ratio	300 microsecs/3dB

EXPERIMENT ONE (SPATIAL LATERALIZATION)

In the first experiment the 16 experimental conditions listed in Table 1 were repeated four times each in a randomized sequence. A stimulus display and response-recording mechanism for spatial location judgments of the speech and noise was generated using MATLAB. The subject could listen to each trial up to three times by selecting the 'repeat sentence' button. The result was recorded in a computerized database.

Each subject was asked to make three responses to each stimulus. The first task was to identify the perceived location of the speech on a schematic drawing. Once the subject had identified the perceived location of the speech and clicked on the mouse button, an 'O' appeared on the screen to represent where the subject heard the *speech*. The subject was then required to identify the perceived spatial location of the *noise* on the same schematic drawing. In the third task, the subject judged the spatial diffuseness of the speech using the scale: very concentrated, concentrated, diffuse and very diffuse. Figure 3.1 is a copy of the display screen of the computer generated response form of a subject.

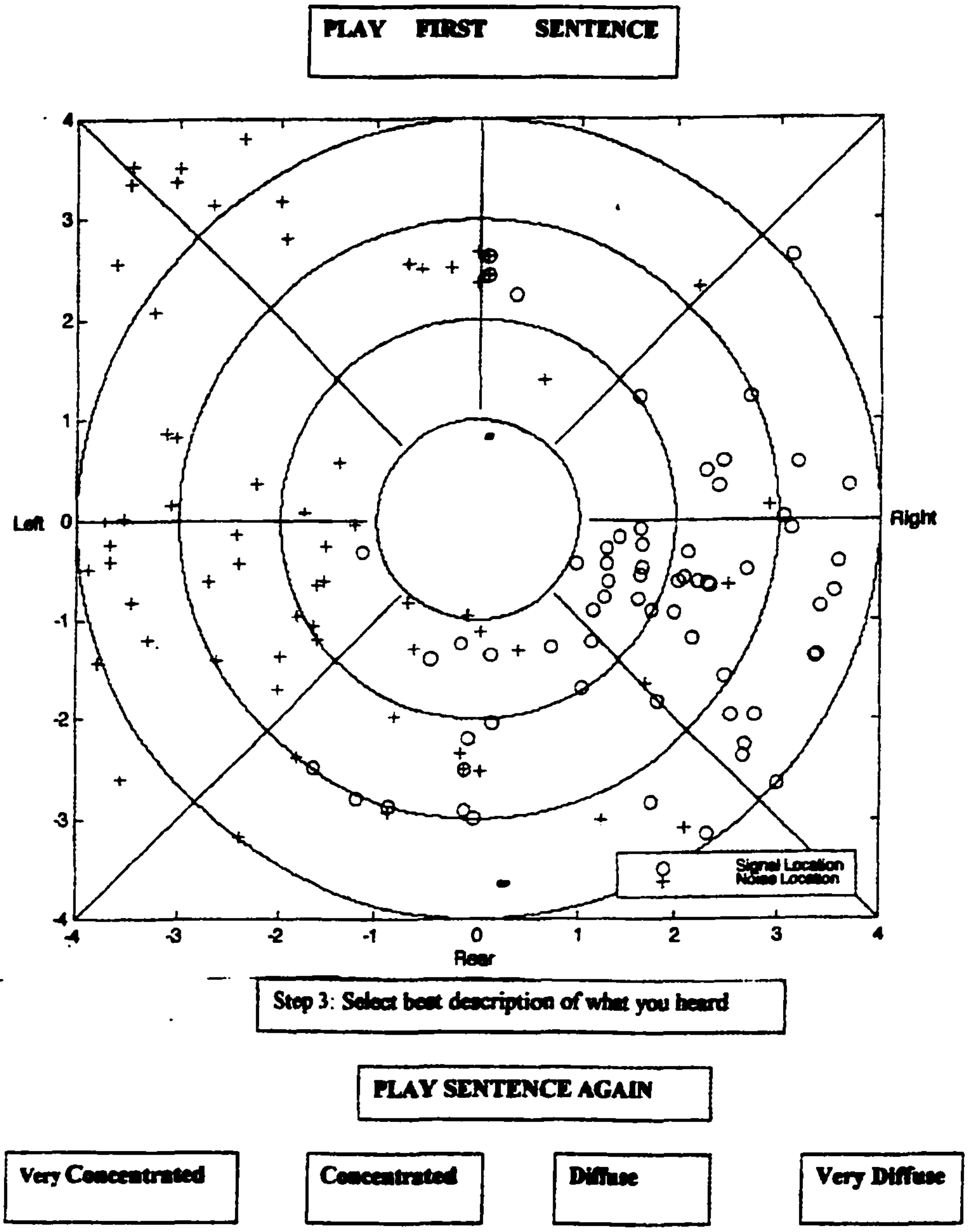


FIGURE 3.1: Computer generated response form for a given subject. Spatial location of the speech and noise (marked with 'o' and 'x') and diffuse ratings obtained with touch pad strokes.

EXPERIMENT TWO: EASE OF LISTENING (PAIRED-COMPARISONS)

In the second experiment, a paired comparison technique was used. Each Processing Technique was compared with every other Processing Technique forming a 4x4 matrix of paired comparisons. Each 4x4 matrix was repeated 4 times at each of the 4 signal to noise ratios (-4dB, -2dB, 0 dB, +2dB). Table 3.2 shows how paired comparisons were carried out for each 4x4 matrix. For a given cell in the matrix, the experimental condition corresponding to the row was presented first and the condition corresponding to the column was presented second i.e., cell CB compared condition C (presented first) with condition B (presented second). If condition C was judged to be easier to listen to than condition B, then a '1' was inserted in the cell CB. If condition B was judged to be easier to listen to than condition C, then a '0' was entered into this cell. The score for condition A was the sum of the entries in Row A. The score for condition B was the sum of the entries in row B. The score for condition C was the sum of entries in Row C. The score for condition D was the sum of entries in Row D.

TABLE 3.2: An example of the 4x4 response matrix produced by one subject's responses for 4 comparisons of 4 conditions at one signal to noise ratio.

SN RATIO		II	IT	ITI	NP	SUM
2dB	II		1	0	0	1
	IT	0		0	1	1
	ITI	0	1		1	2
	NP	0	1	0		1
0 dB	II		0	1	1	2
	IT	0		0	1	1
	ITI	1	1		1	3
	NP	0	0	0		0
-2 dB	II		1	0	1	2
	IT	0		1	1	2
	ITI	0	0		1	1
	NP	0	0	0		0
-4 dB	II		1	0	0	1
	IT	0		1	1	2
	ITI	0	1		1	2
	NP	0	0	0		0

CHAPTER IV RESULTS

The primary focus of this investigation was to evaluate the effect of simulated binaural separation on the perception of speech against a background of noise. A Simulated Binaural Listening system was used for this purpose. Sinewave modeling was used to increase the perceptual differences between the spectral peaks and the residual. The major spectral peaks, which consist mostly of speech, are presented with inter-aural time and intensity differences such that the sound image associated with these spectral peaks (speech image) is heard towards one side of the head. The residual spectral components, which consist mostly of noise, are presented with the opposite interaural time/intensity differences so that the residual image, is heard towards the opposite side of the head as noise.

The hypotheses that were addressed in this study were: 1) Interaural intensity differences applied to the spectral components of the speech plus noise, as described above, will produce differences in the perceived spatial locations of the speech and noise. 2) The interaural time differences applied to the spectral components of the speech plus noise, as described above, will produce a difference in the perceived spatial location of the speech and noise 3) the combination of the two processing techniques will produce even greater differences in the perceived locations of the speech and noise, and 4) the perceived spatial separation between the speech and noise will improve ease of listening to the speech.

It was also of interest to determine the effect of the processing technique on the perceived diffuseness of the speech and noise images.

SPATIAL SEPARATION:

The subjects identified the perceived locations of the speech and noise images on a computer-generated diagram. A typical set of data for one subject is shown in Figure 4.1.

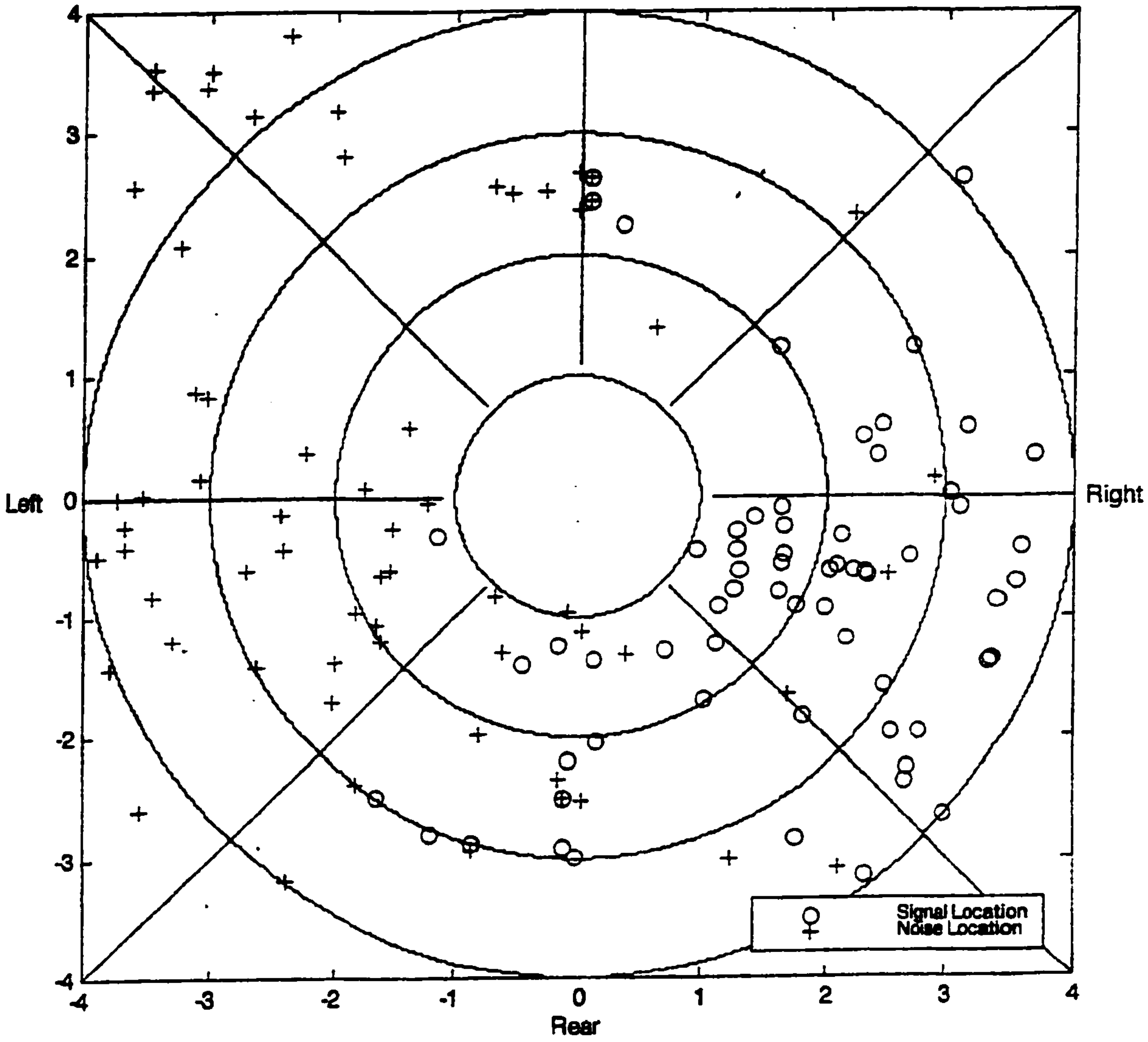


FIGURE 4.1: Set of data identifies the perceived location of the speech and noise on a computer-generated diagram. The circles show the perceived location of the speech image, the crosses represent the perceived location of the noise image. The scale for the X and Y- axes is in distance units (du) where 1 du is equal to the radius of the idealized head in the diagram.

The horizontal X-axis shows the perceived location of the sound image on the left-right dimension. The vertical Y-axis shows the perceived location of the sound image on the front-back dimension. The vertical and horizontal scales of the computer-generated diagram were selected arbitrarily using distance units (du) corresponding roughly to the radius of the head as shown in

Figure 4.1. A negative number on the horizontal axis indicates that the sound image was heard to the left of the head center. A negative number on the vertical axis indicates that the sound image was heard behind the center of the head. The circles show the perceived locations of the speech image for the various experimental conditions. The crosses show the perceived the locations of the noise image. The data in Figure 4.1 are for Subject nine. Figures showing the data for each of the 10 subjects appear in Appendix A.

As is evident from Figure 4.1, which is typical of almost all of the subjects, the speech image was usually located on the right side of the head while the location of the noise image was heard either near the center or on the left side of the head. The separation of the speech and noise is more evident on the X-axis than the Y-axis. There were also large differences in the locations of the speech and noise images between trials.

Of particular interest was the degree of separation between the speech and noise images on individual trials, and the effect of the experimental conditions on the degree of separation. This was analyzed by computing the difference in location between the speech and noise on each trial. The difference on the X-axis is given by /Xs-Xn/ for the X-axis and by /Ys-Yn/ for the Y-axis where the suffix 's' indicates speech and the suffix 'n' indicates noise.

In addition to judgments of perceived location, the subjects also rated the diffuseness of the speech images. These ratings were obtained on a 4-point scale with 1=very concentrated, 2=concentrated, 3=diffuse and 4= very diffuse.

In order to address the hypothesis relating to perceived spatial location, the following analyses of variance were performed.

- Perceived location of the speech image on the horizontal axis (Xs)
- Perceived location of speech image on the vertical axis (Ys)
- Perceived location of noise image on the horizontal axis (Xn)
- Perceived location of noise image on the vertical axis (Yn)
- The difference in spatial location of the speech and noise images on the horizontal axis (Xs-Xn)
- The difference in spatial location of the speech and noise images on the vertical axis (Ys-Yn)
- The perceived diffuseness of the speech image

A three-factor repeated measure analysis of variance was performed for each set of data. The factors were: Processing Technique, PT, (Interaural time difference, IT, Interaural intensity difference, II, Interaural time and intensity difference, ITI, and no processing, NP), Signal-to-Noise Ratio, SN, (+2 dB, 0 dB, -2dB and -4 dB), and Subjects, SU, (1 through 10).

4.1.1: LOCATION OF THE SPEECH IMAGE ON THE HORIZONTAL AXIS

The results of the analysis of variance are shown in Table 4.1. Processing Technique (PT) was found to be statistically significant (F (3,27)= 2.94,*p* <. 05). Signal to noise ratio (SN) was not significant. Similarly the PT x SN interaction was not significant. The mean scores for each processing technique are shown in Figure 4.2. A post-hoc analysis using the Tukey Honestly Significant Difference (HSD) method (Tukey, 1953) showed the mean locations for the II and ITI processing techniques did not differ significantly from each other and the mean location for the IT and NP did not differ significantly from each other but that the mean location for II and ITI did differ from IT and NP.

The deviation from the center of the head is not statistically significant for the unprocessed speech and the IT processing technique but is significant for the ITI and II processing technique.

TABLE 4.1: Analysis of Variance for the Location of the Speech on the X-axis.

	Sums of Squares	**Degrees of Freedom**	**Mean Squares**	**F**	**Significance Level**
Processing Technique (PT)	67.84	3	22.61	2.94	<. 05
Speech to Noise Ratio (SN)	1.68	3	.56	.279	NS
PT x SN Interaction	5.42	9	.602	.345	NS
SU	600.40	9	66.71	38.16	
SU x PT Interaction	208.15	27	7.71		
SU x SN Interaction	54.28	27	2.01		
SU x PT x SN	134.65	81	1.66		

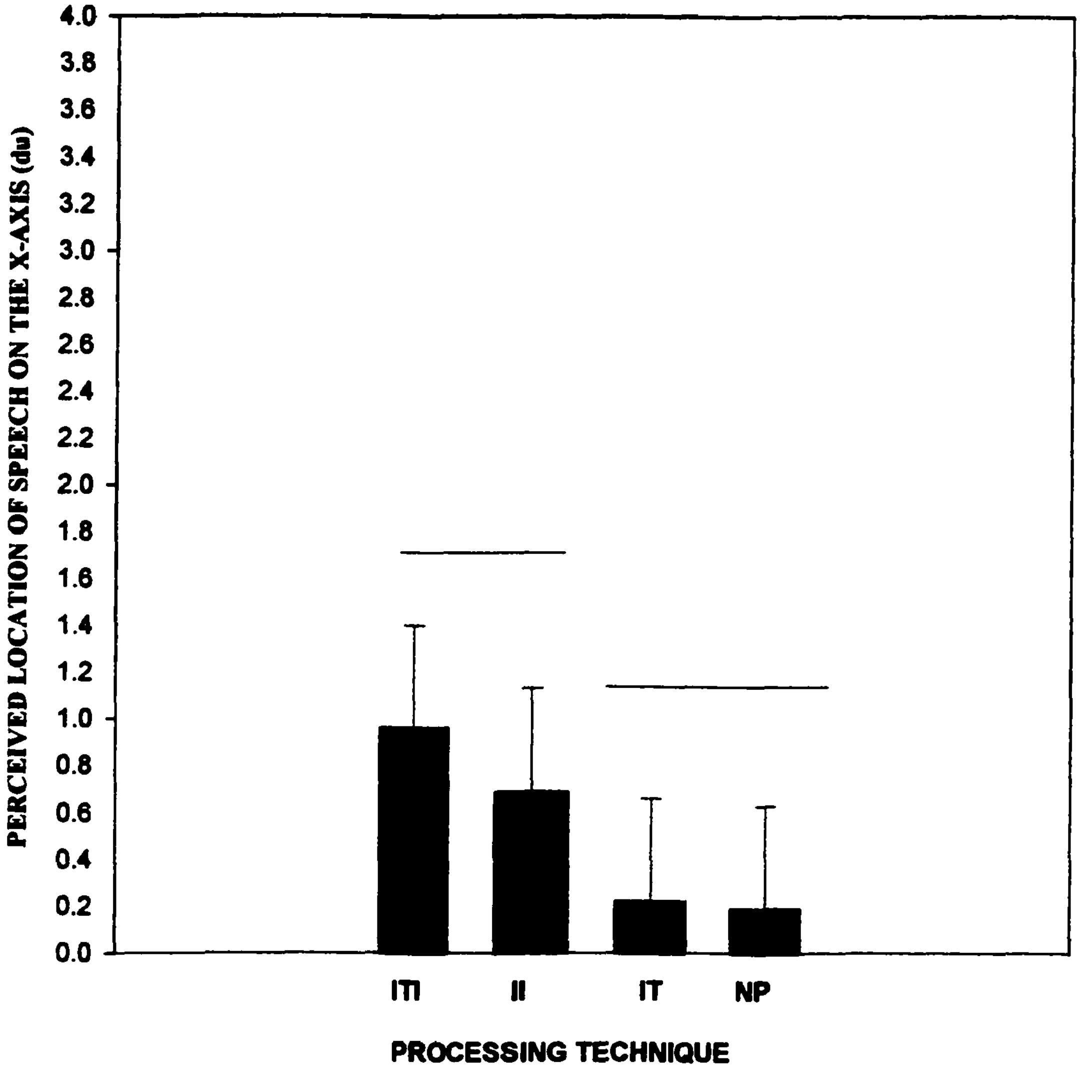


FIGURE 4.2: The effect of Processing Technique (PT) on the perceived location of the speech image on the X- axis. The perceived location is specified in terms of the distance units (du) used in the computer-generated diagram for recording the subjects' response. Each mean is the average over 10 subjects and 4 speech-to-noise ratios. The vertical lines show the standard error for each mean (se= .44). Means under a common superscript line do not differ significantly from each other, (*p*<*.05*).

4.1.2 LOCATION OF THE NOISE IMAGE ON THE X-AXIS

The results of the analysis of variance are shown in Table 4.2. Processing Technique (PT) was found to be significant (F (3,27)= 16.72,*p* <. 01). Signal to noise ratio (SN) was not significant. Similarly the PT x SN interaction was also not significant. The mean scores for each processing technique are shown in Figure 4.3. A post-hoc analysis using the Tukey Honestly Significant Difference method (HSD) (Tukey, 1953) showed that the mean locations for the ITI and IT processing techniques did not differ significantly. The mean locations for the II and NP (No Processing) techniques were significantly different from each other and the locations for the ITI and IT processing technique.

The deviation from the center of the head is not statistically significant for the noise image unprocessed (NP) but is significant for the II, IT and ITI processing techniques.

TABLE 4.2: Analysis of Variance for the Location of the Noise on the X-axis.

	Sums of Squares	**Degrees of Freedom**	**Mean Squares**	**F**	**Significance Level**
Processing Technique (PT)	376.78	3	125.59	16.72	<. 01
Speech to Noise Ratio (SN)	4.29	3	1.43	.82	NS
PT x SN Interaction	15.00	9	1.67	1.21	NS
SU	301.6	9	33.51		
SU x PT Interaction	202.89	27	7.51		
SU x SN Interaction	46.98	27	1.74		
SU x PT x SN	111.50	81	1.38		

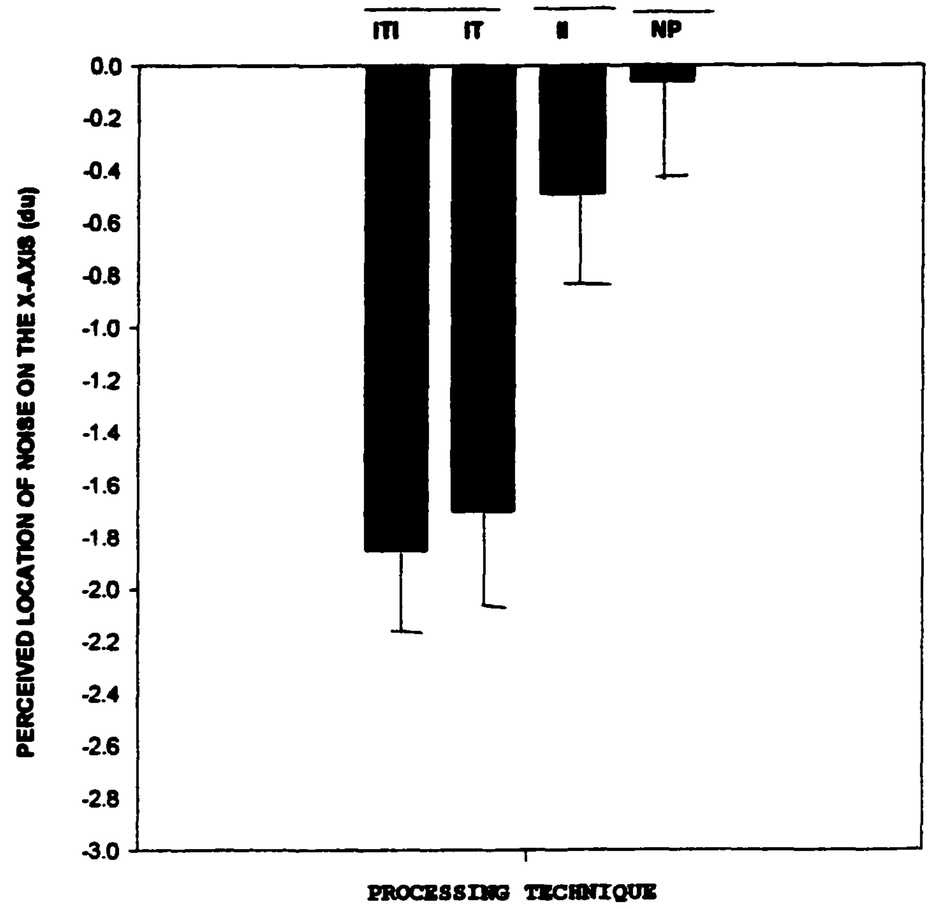


FIGURE 4.3: The effect of Processing Technique (PT) on the perceived location of the noise image on the X- axis. The perceived location is specified in terms of the distance units (du) used in the computer-generated diagram for recording the subjects' response. Each mean is the average over 10 subjects and 4 speech-to-noise ratios. The vertical lines show the standard error for each mean (se= .43). Means under a common subscript line do not differ significantly from each other, ($p<.05$).

4.1.3 LOCATION OF THE SPEECH IMAGE ON THE Y-AXIS

The results of the analysis of variance for speech on the Y-axis are shown in Table 4.3. The analysis revealed that neither of the main effects nor the interaction was significant at the .05 levels.

TABLE 4.3: Analysis of Variance for the Location of the Speech on the Y-axis.

	Sums of Squares	**Degrees of Freedom**	***Mean Squares***	***F***	***Significance Level***
Processing Technique (PT)	21.65	3	7.22	1.96	NS
Speech to Noise Ratio (SN)	26.33	3	8.78	2.41	NS
PT x SN Interaction	14.40	9	1.60	.92	NS
SU	588.49	9	65.39	40.48	
SU x PT Interaction	99.33	27	3.68		
SU x SN Interaction	98.25	27	3.64		
SU x PT x SN	139.99	81	1.73		

4.1.4 : LOCATION OF THE NOISE IMAGE ON THE Y-AXIS

The results of the analysis of variance for noise on the Y-axis are shown in Table 4.4. The analysis revealed that neither of the main effects nor the interaction was significant at the .05 levels.

TABLE 4.4: Analysis of Variance for the Location of the Noise on the Y-axis.

	Sums of Squares	**Degrees of Freedom**	***Mean Squares***	***F***	***Significance Level***
Processing Technique (PT)	2.27	3	.76	.40	NS
Speech to Noise Ratio (SN)	.78	3	.26	.14	NS
PT x SN Interaction	21.24	9	2.36	.97	NS
SU	494.76	9	54.97	29.1	
SU x PT Interaction	145.07	27	5.37	2.84	
SU x SN Interaction	49.22	27	1.82	.96	
SU x PT x SN	197.02	81	2.43	1.29	

4.1.5: DIFFERENCE IN LOCATION BETWEEN THE SPEECH AND NOISE IMAGES ON THE X-AXIS

The results of the analysis of variance for difference in location of speech-noise are shown in Table 4.5. Processing Technique (PT) was found to be significant (F (3,27)=30.85,*p* <.001). Signal-to-noise ratio (SN) was not significant. Similarly the PT x SN interaction was not significant. The mean scores for processing are shown in Figure 4.4. A post-hoc analysis using the Tukey Honestly Significant Difference test showed that the mean locations for the ITI, IT, II and NP processing techniques differed significantly from each other. As is evident from the diagram, the Interaural Time plus Interaural Intensity (ITI) condition showed the highest perceived difference in location (2.81 du) between the speech and noise image. The next highest difference in location (1.89 du) was shown for the Interaural Time Difference (IT) condition. The Interaural Intensity Difference (II) condition showed a significant difference in location of 1.18du. The observed difference in location between the speech and noise images for the no processing condition was 0.28 du and was not significant. The standard error for each of these means was 0.39 du.

Table 4.5: Analysis of Variance for the Location of the Speech -Noise on the X-axis

	Sums of Squares	Degrees of Freedom	Mean Squares	F	Significance Level
Processing Technique (PT)	553.68	3	184.56	30.85	<. 001
Speech to Noise Ratio (SN)	1.63	3	.54	.17	NS
PT x SN Interaction	7.32	9	0.81	.25	NS
SU	633.34	9	70.37		
SU x PT Interaction	161.54	27	5.98		
SU x SN Interaction	85.58	27	3.16		
SU x PT x SN	265.32	81	3.27		

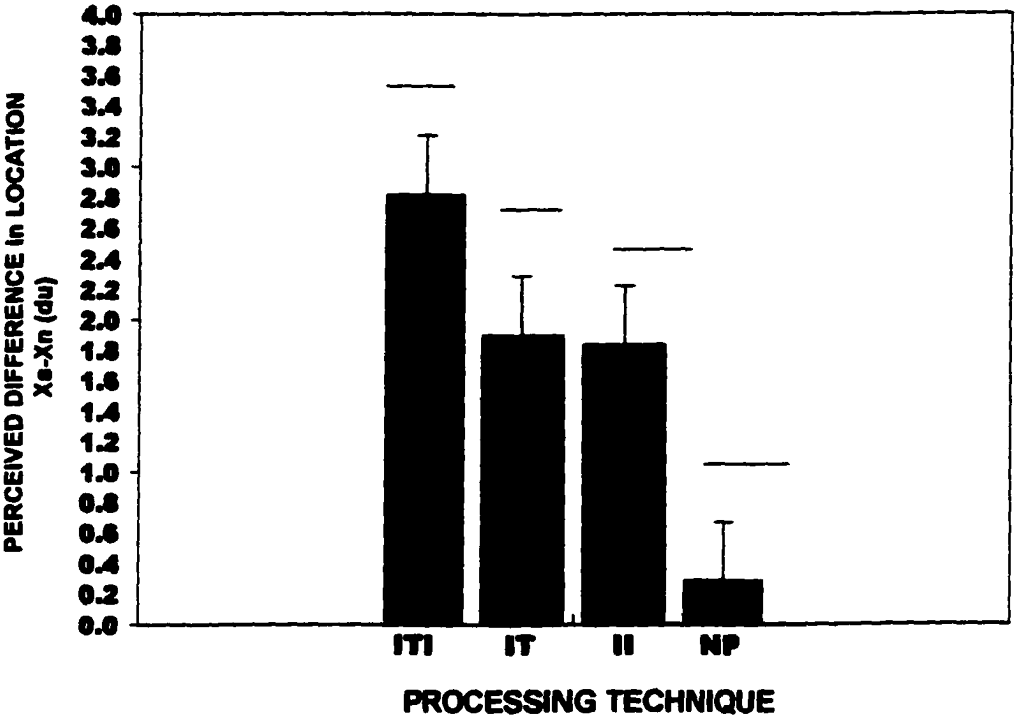


FIGURE 4.4: The effect of Processing Technique (PT) on the perceived difference in location between speech and noise images on the X- axis. The perceived location is specified in terms of the distance units (du) used in the computer-generated diagram for recording the subject's response. Each mean is the average over 10 subjects and 4 speech-to-noise ratios. The vertical lines show the standard error for each mean (se= .39). Each mean differed significantly from each other, (*p<.05*).

4.1.6: DIFFERENCE IN LOCATION BETWEEN THE SPEECH AND NOISE IMAGES ON THE Y-AXIS

Table 4.6: Analysis of Variance for the Location of the Speech –Noise on the Y-axis

	Sums of Squares	**Degrees of Freedom**	**Mean Squares**	**F**	**Significance Level**
Processing Technique (PT)	17.70	3	5.82	.60	NS
Speech to Noise Ratio (SN)	32.24	3	.10.75	1.71	NS
PT x SN Interaction	18.60	9	2.07		NS
Subject (SU)	333.96	9	37.10		
SU x PT Interaction	259.10	27	9.60		
SU x SN Interaction	85.58	27	3.16		
SU x PT x SN	266.72	81	3.30		

The perceived spatial separation between the speech and noise images on the Y-axis was also subjected to a repeated measures analysis of variance (ANOVA). The results of this analysis are shown in Table 4.6. In this case neither of the main effects or interaction were significant. Processing Technique was not statistically significant.

4.1.7: SUMMARY

Processing Technique (PT) was found to have a significant effect on the location of both the speech and noise images on the X-axis but not on the Y-axis. The largest difference in spatial location was produced by the ITI condition followed by the IT condition. The smallest difference in location was produced by the II condition. Neither the speech or noise image differed significantly from the center of the head for the unprocessed condition.

There were large between subject differences. These differences were analyzed on a subject-by-subject basis, as described in the next section.

4.1.8: INDIVIDUAL RESULTS

4.1.8.1: ANALYSIS OF INDIVIDUAL DIFFERENCES OF THE SPEECH IMAGE ON THE X-AXIS

A separate fixed effects analysis of variance was carried out for each subject. The results of the analyses are summarized in Table 4.7. Figure 4.5a shows the effect of Processing Technique on the location of the speech image on the X-axis. The data have been averaged over four signal-to-noise ratios. The ITI condition showed the largest displacement for 6 of the 10 subjects. Although there were large between subject differences in the average location of the speech image, 8 of the 10 subjects (all but Subjects 4 and 8) showed essentially the same pattern. The speech image was perceived to the right of the center of the head (i.e. away from the side receiving the interaural delay for speech peaks). The subjects showing a significant displacement from the center of the head showed a bias in that direction for all of the experimental conditions. The shift was shown to the right side of the head for 8 of the 10 subjects. Subject 4 had a bias to the left side of the head.

Speech-to-noise ratio (SN) was significant for Subject 4. A post-hoc analysis using the Tukey Honestly Significant Difference test showed the mean locations for the 2dB, 0dB, and -2 dB speech-to-noise ratios differed significantly from the -4dB speech-to-noise ratio.

The data for Subject 8 were erratic. If the data from this subject are omitted, then the mean displacement averaged over subjects and Speech-to-noise ratios for the ITI, II, IT, and NP conditions are 1.14, 0.61, 0.44, .07 respectively. If subject 4 and 8 are omitted the mean displacement averaged over subjects and Speech-to-noise ratios for the ITI, II, IT, and NP conditions are 1.49, 0.89, 0.76, 0.37 respectively.

Table 4.7: Fixed Effects Analysis of Variance for the Dimension of the Speech Image on the X-Axis for each Subject.

Individual ANOVAS Xs	F - RATIO										
	df	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Processing Technique (PT)	3/27	7.54 ***	4.43 ***	.63	1.24	1.28	12.3 ***	3.70 ***	27.7 ***	3.21 ***	1.11
Speech to Noise Ratio (SN)	3/27	.69	1.22	1.45.	4.41 ***	1.37	.25	.19	1.25	.14	.15
PT x SN Interaction	9/27	.44	.72	1.31	1.03	.44	.73	.51	1.10	1.92	.99
Standard Error of Mean (PT)		.21	.30	.42	.31	.50	.16	.40	.33	.30	.14

***= Significance level <. 05

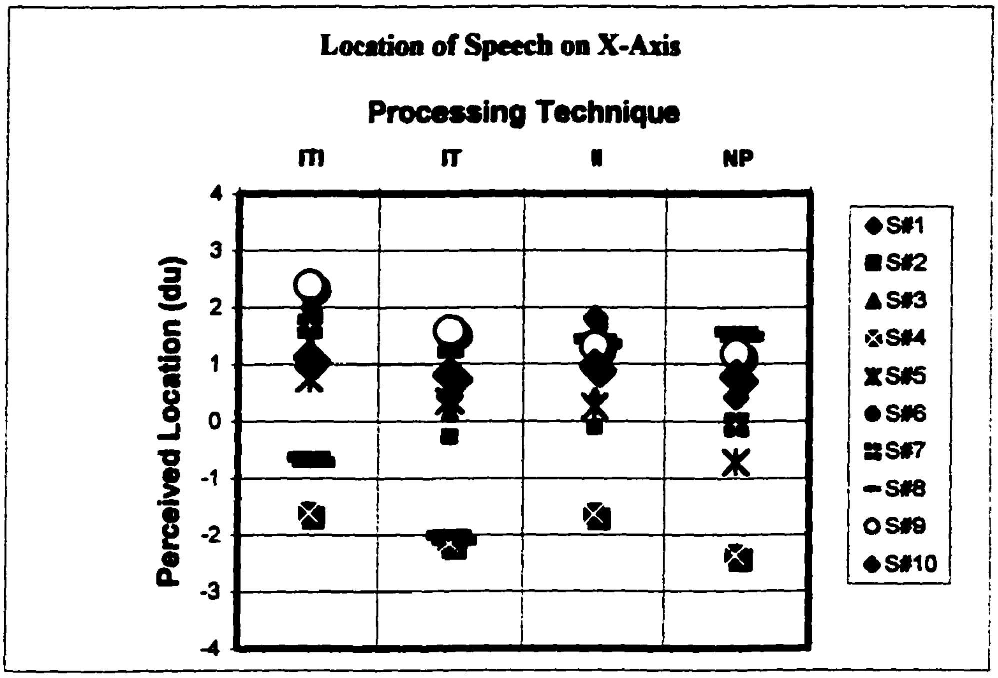


FIGURE 4.5: LOCATION OF SPEECH IMAGE ON THE X-AXIS FOR EACH SUBJECT

4.1.8.2: ANALYSIS OF INDIVIDUAL DIFFERENCES OF THE NOISE IMAGE ON THE X AXIS

A separate fixed effects analysis of variance was carried out for each subject. The results of the analyses are summarized in Table 4.8. Processing Technique was significant for 8 out of 10 subjects. The ITI condition showed the largest displacement for 5 of these subjects and the IT showed the largest displacement for the remaining 3 subjects. The II condition showed a small but significant displacement of the noise images to the left of the center of the head (8 of the subjects). Although there were large between subject differences in the average location of the noise image 9 of the 10 subjects with one exception (Subject 4) showed essentially the same pattern. The noise image was perceived to the left of the center of the head (negative numbers), i.e. away from the side receiving the delayed and/or attenuated speech peaks. The NP condition did not show a significant displacement of the noise image, on average, although individual subjects showed significant displacement to the right or left of the head center. Subject 4 differed from the other subjects in that this subject showed essentially the same displacement for all processing conditions.

Speech-to-noise ratio (SN) was significant for Subject 1. A post-hoc analysis using the Tukey Honestly Significant Difference test showed the mean locations for the 2dB, 0dB, speech-to-noise ratio differed significantly from the -2dB and -4dB speech-to-noise ratio condition.

Table 4.8: Fixed Effects Analysis of Variance for the Dimension of the Noise Image on the X-Axis for each Subject.

Individual ANOVAS Xn		F- RATIO									
	df	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Processing Technique (PT)	3/27	59.03 ***	7.74 ***	4.16 ***	.38	.1.37	9.90 ***	6.27 ***	62.4 ***	6.05 ***	31.2 ***
Speech to Noise Ratio (SN)	3/27	4.0 ***	.80	.360	.50	1.49	.15	.23	.66	1.31	2.32
PT x SN Interaction	9/27	.66	1.16	.90	.87	.99	.22	.86	1.16	.63	.81
Standard Error of Mean		.01	.23	.42	.32	.58	.24	.38	.28	.40	.22

***= Significance level < 05

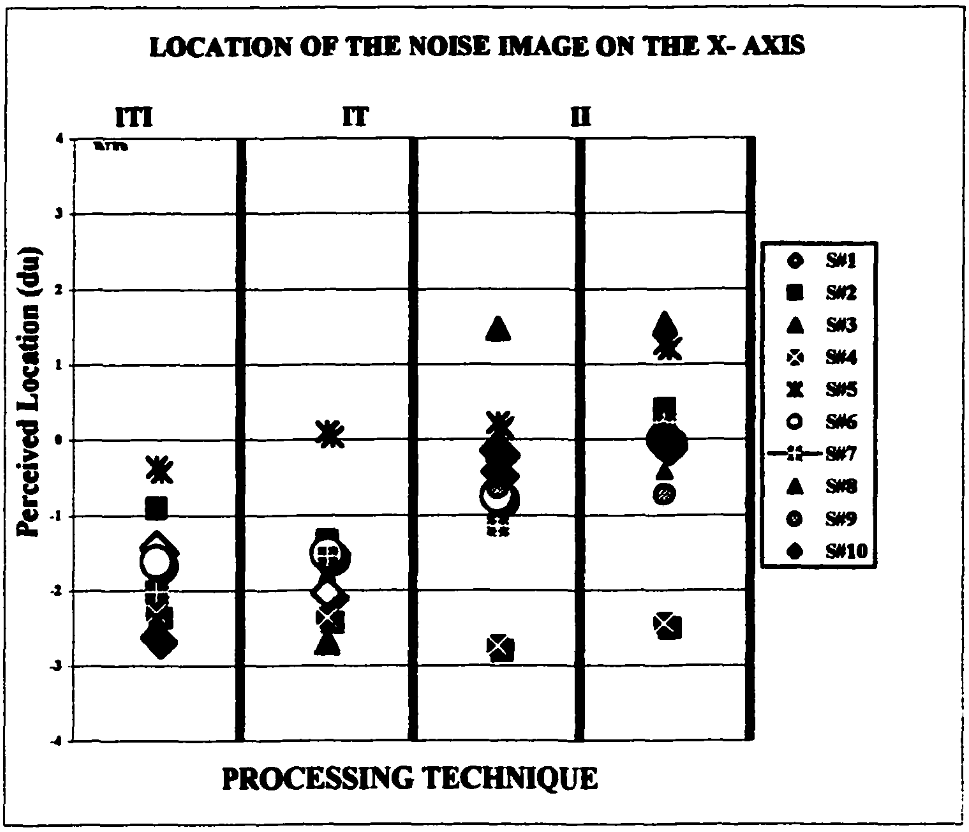


FIGURE 4.6: LOCATION OF NOISE IMAGE ON THE X-AXIS FOR EACH SUBJECT

4.1.8.3: ANALYSIS OF INDIVIDUAL DIFFERENCES OF THE SPEECH IMAGE ON THE Y-AXIS

A separate fixed effects analysis of variance was carried out for each subject. The results of the analysis are summarized in Table 4.9. Unlike the speech image on the X- axis there was no consistent pattern of results for the speech image on the Y-axis. Processing Technique was significant for only 2 out of 10 subjects. Figure 4.7 shows the effect of Processing Technique on the location of the speech image on the Y-axis. The data have been averaged over four speech-noise-ratios. Speech-to-noise ratio was significant for 1 of the 10 subjects. None of the subjects showed a PT x SN interaction.

Table 4.9: Fixed Effects Analysis of Variance for the Dimension of the Speech Image on the Y-Axis for each Subject.

Individual ANOVAS Ys	df	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
		F - RATIO									
Processing Technique (PT)	3/27	4.45 ***	3.30	.59	3.63	.22	1.57	.04	.37	1.19	9.36 ***
Speech to Noise Ratio (SN)	3/27	1.9	4.17 ***	.55	2.50	7.57 ***	3.08	.90	1.72	.45	.59
PT x SN Interaction	9/27	1.02	.20	2.99	1.99	.91	.61	1.07	2.09	.62	1.29
Standard Error of Mean		.38	.45	.19	.35	.31	.27	.28	.31	.36	.18

***= Significance level <.05

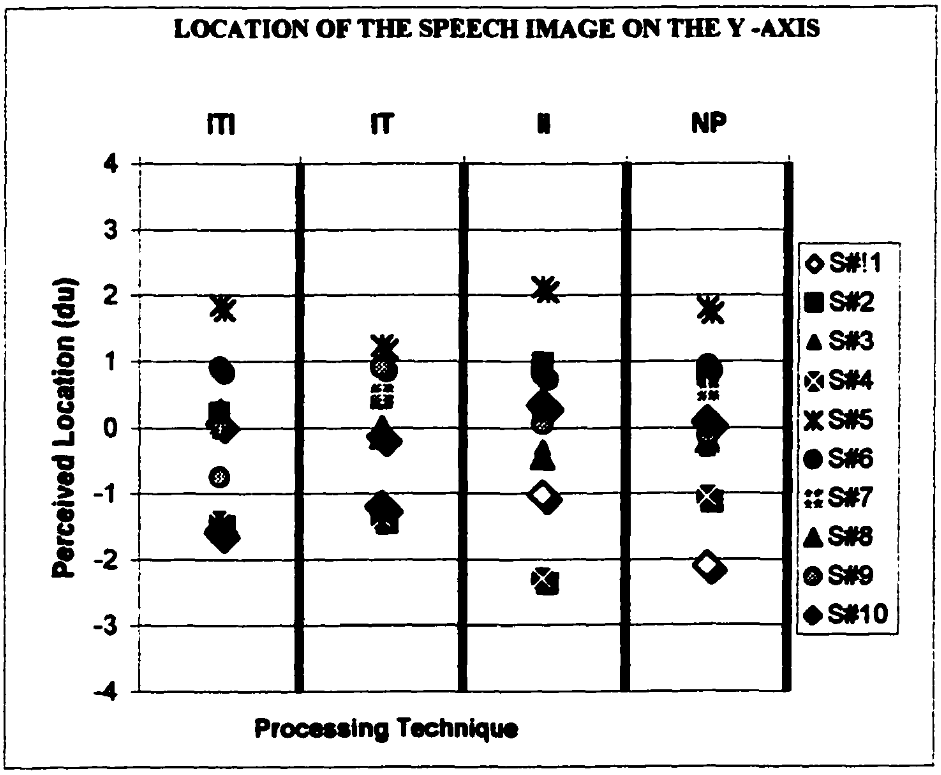


FIGURE 4.7: LOCATION OF SPEECH IMAGE ON THE Y-AXIS FOR EACH SUBJECT

4.1.8.4: ANALYSIS OF INDIVIDUAL DIFFERENCES OF THE NOISE IMAGE ON THE Y-AXIS

Data on the location of the noise image on the Y-axis was analyzed in the same way. The analysis of variance for each subject is shown in Table 4.10. Unlike the location measurements on the X-axis, the data do not show a consistent pattern on the Y-axis. This result is consistent with the analysis of variance, which did not show a significant processing effect for the noise image on the Y-axis for 8 of 10 subjects. Figure 4.8 shows the effect of Processing Technique on the location of the noise image on the Y-axis. The data has been averaged over four speech-to-noise ratios.

TABLE 4.10: Fixed Effects Analysis of Variance for the Dimension of the Noise Image on the Y-Axis for each Subject.

Individual ANOVAS Ys	F - RATIO										
	df	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Processing Technique (PT)	3/27	22.4 ***	.98	1.6	1.50	.77	.08	1.42	1.34	1.85	4.18 ***
Speech to Noise Ratio (SN)	3/27	2.68	1.98	.32	1.79	.12	.40	1.11	1.82	.26	.23
PT x SN Interaction	9/27	2.15	.53	.70	2.39	.77	1.63	1.18	.46	1.80	.63
Standard Error of Mean		.21	.34	.15	.44	.42	.20	.24	.25	.50	.46

***= Significance level <. 05

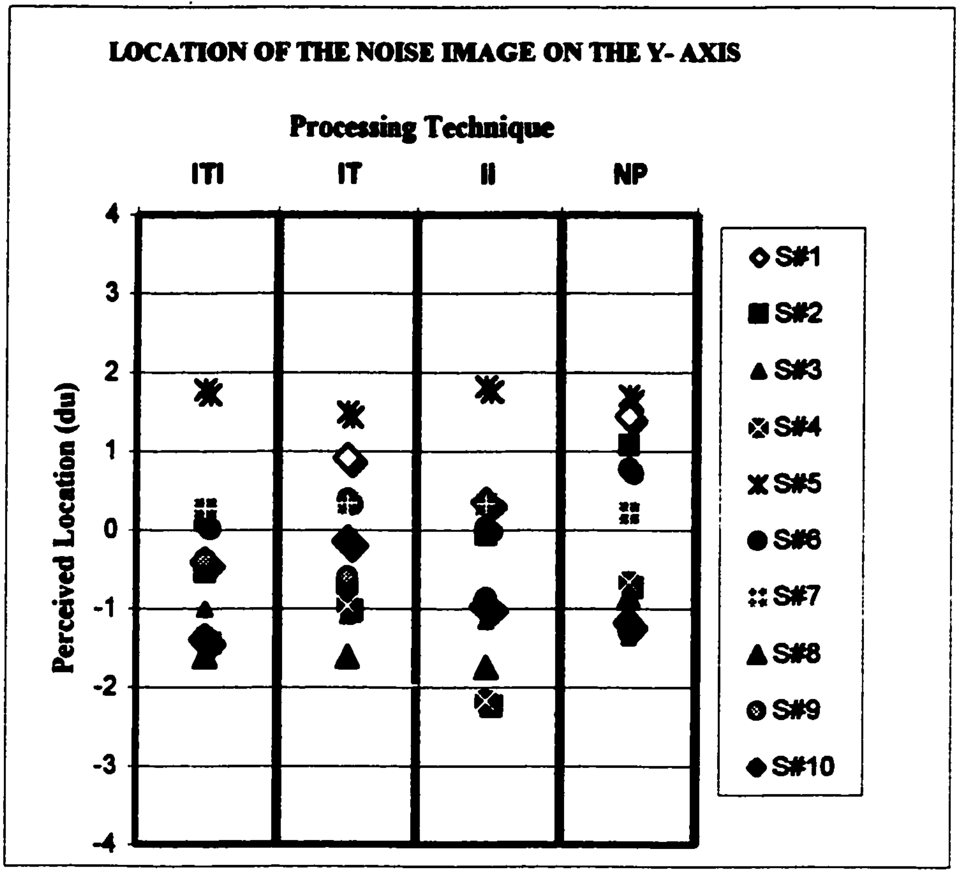


FIGURE 4.8: LOCATION OF NOISE IMAGE ON THE Y-AXIS FOR EACH SUBJECT

4.1.8.5: ANALYSIS OF INDIVIDUAL DIFFERENCES ON THE Xs-Xn AXIS

The large between subject differences observed for the speech and noise images appear to be similar. In addition, there were between trial differences in perceived location. It was thus of interest to examine the difference in location between the speech and noise images on the X-axis for each subject on each trial, thereby reducing both between-subject and between-trial effects. Xs-Xn was calculated for each trial and subjected to an analysis of variance. Bear in mind that 4 replications were obtained for each experimental condition. This allowed for a relatively sensitive fixed-effect analysis of variance for each subject.

The results of these analyses are summarized in Table 4.11. Only 1 of the 10 subjects (Subject 5) did not show a significant processing effect. This subject had a relatively large replication error variance and although the observed difference between the ITI and NP condition was as large as that obtained for other subjects the difference was not statistically significant. This subject also differed from the remaining subjects in other ways as shown in Figure 4.9.

Figure 4.9 shows the effect of Processing Technique on Xs-Xn. With one exception (Subject 4), every subject showed the largest separation between the speech and noise images for the ITI condition. Even in the case of Subject 4, Xs-Xn for the ITI condition was not significantly different from that for the II condition, which for this subject showed the largest value of Xs-Xn. Similarly with the exception of Subject 4, all of the subjects showed the IT condition to yield the second highest value of Xs-Xn. The II condition showed the smallest effect of the processing conditions, except for Subject 4 as noted above. The NP condition did not show significant difference in location between the speech and noise images (i.e. Xs-Xn = 0 except for subject 5 and 9).

Subject 4 was the only subject to show a significant effect for speech-to-noise ratio. A post-hoc analysis using the Tukey Honestly Significant Difference test showed the mean locations for the 2dB, 0dB, -2dB speech-to-noise ratios differed significantly from the -4dB speech-to-noise ratio condition. The mean location for the -4dB speech-to-noise ratio condition was

-2.75du. This was substantially further to the left of the center of the head than mean location for the other three speech-to-noise ratios (+2dB, 0dB and -2dB).

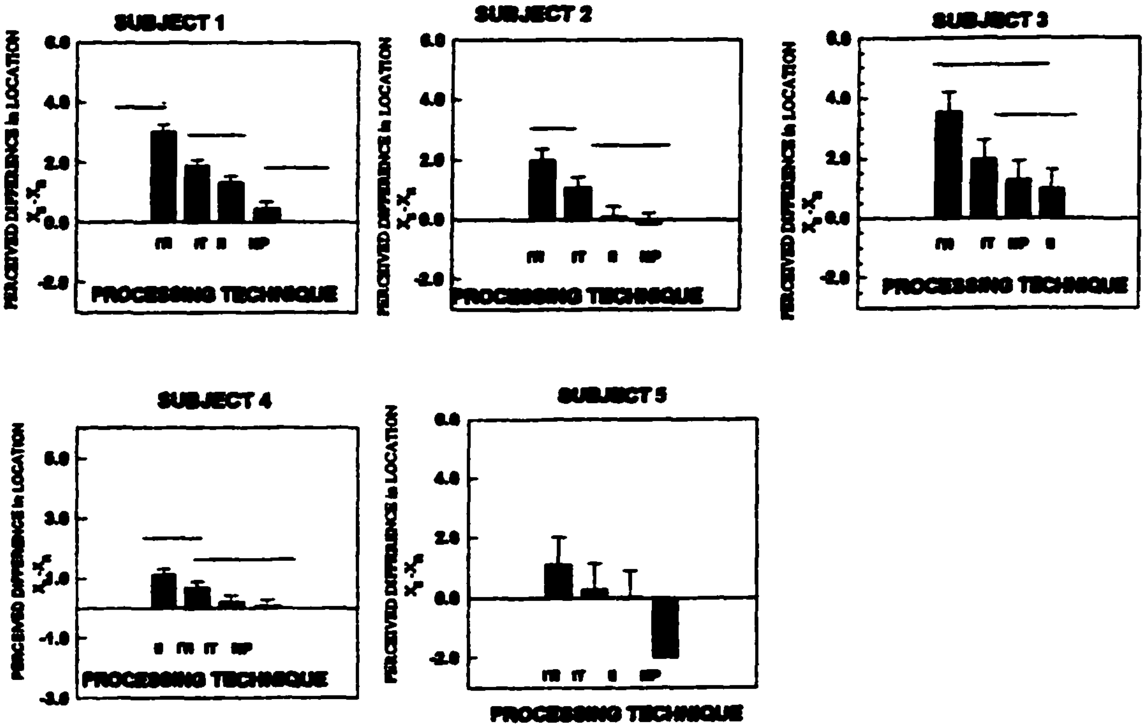
The PT x SN (Processing Technique x Signal to Noise Ratio) interaction was not significant for any of the subjects. This is consistent with the observation that all of the subjects showed the same general pattern of results. It is noted that the standard error of the mean for subjects 1,2,4,6,8,and 10 was relatively small, while the standard error of the mean for subjects 3,5,7 and 9 was fairly large.

Table 4.11: Fixed Effects Analysis of Variance for the Dimension of the Speech-Noise on the X-Axis for each Subject.

Individual ANOVAs Xs-Xn		**F - RATIO**									
	Df	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
Processing Technique (PT)	3/27	23.2 ***	6.95 ***	3.09 ***	5.5 ***	2.11	25.38 ***	5.5 ***	5.81 ***	6.95 ***	41.9 ***
Speech to Noise Ratio (SN)	3/27	1.13	1.36	0.78	9.03 ***	0.52	0.1	0.15	0.7	0.68	1.47
PT x SN Interaction	9/27	0.78	0.63	0.95	0.86	0.60	0.54	0.55	0.68	1.2	0.85
Standard Error of Mean		.22	.37	.65	.20	.90	.26	.73	.36	.56	.21

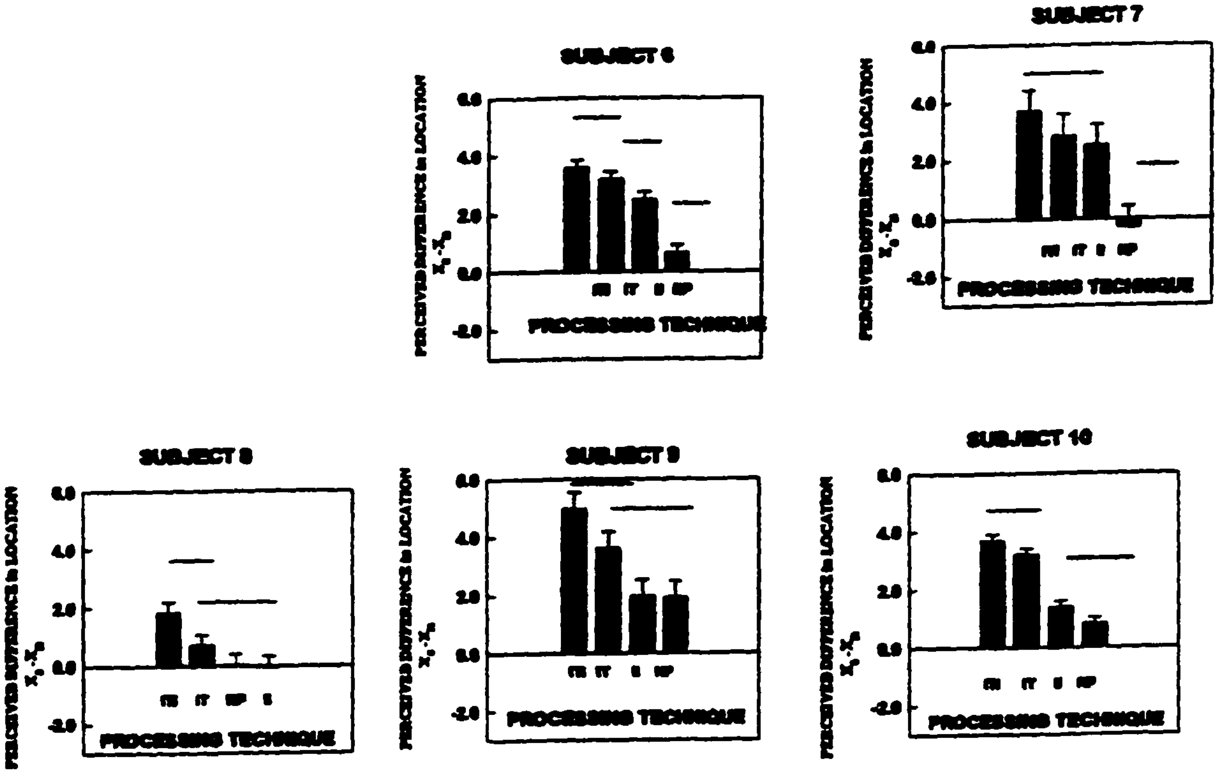
***= Significance level <. 05

FIGURE 4.9:The Effect of Processing technique (PT) on X_s-X_a. Means that do not differ significantly from each other ($p<0.01$) are grouped together by a superscript line.



IT- Interaural Time Difference
II- Interaural Intensity Difference
ITI- Interaural Time and
Intensity Difference
NP- No Processing

FIGURE 4.9: The Effect of Processing technique (PT) on Xs-Xn. Means that do not differ significantly from each other ($p<0.01$) are grouped together by a superscript line.



IT- Interaural Time Difference
II- Interaural Intensity Difference
ITI- Interaural Time and Intensity Difference
NP- No Processing

4.2.1: DIFFUSENESS RATING RESULTS

A four-point rating scale with descriptive adjectives (1= very concentrated, 2= concentrated, 3= diffuse, 4= very diffuse) was used. It was assumed that the ratings approximated an interval scale in order to perform a conventional analysis of variance, as recommended by Gabrielson (1976).

4.2.2: ANALYSIS OF DIFFUSENESS RATINGS

Ratings of diffuseness were subjected to the same repeated measures analysis of variance (ANOVA) as for the perceived differences in location.

TABLE 4.12
Analysis of Variance for Ratings of Diffuseness

	Sums of Squares	Degrees of Freedom	Mean Squares	F	Significance Level
Processing Technique (PT)	6.06	3	2.02	2.35	NS
Speech to Noise Ratio (SN)	3.75	3	1.25	2.46	NS
PT x SN Interaction	.75	9	.08	.57	NS
Subject (SU)	8.01	9	.89		
SU x PT Interaction	23.30	27	.86		
SU x SN Interaction	13.73	27	.51		
SU x PT x SN	11.9	81	.15		

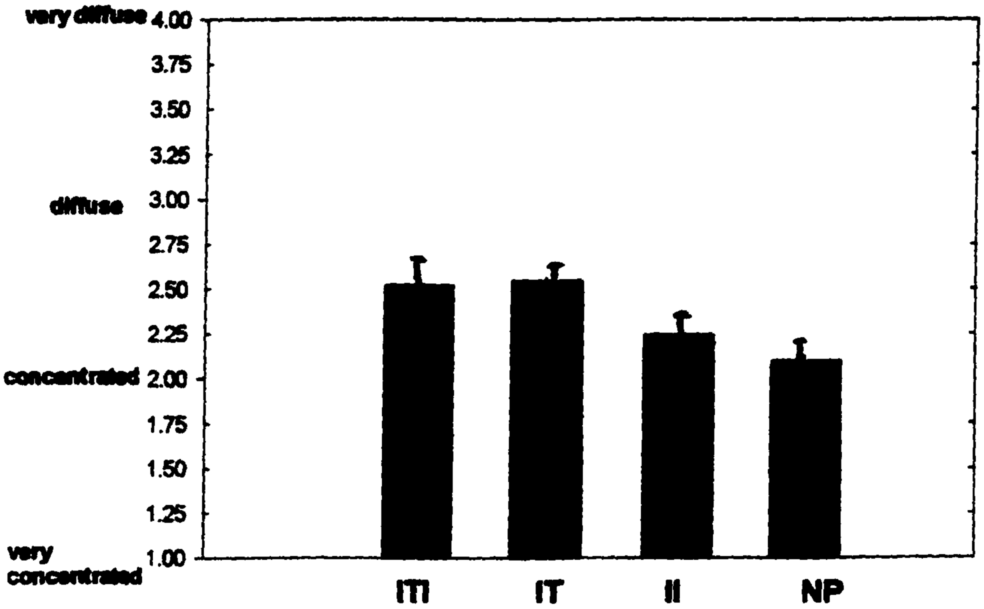


FIGURE 4.8: The Effect of Processing Technique (PT) on the ratings of diffuseness of speech. Mean and standard deviation (± sd) for the factor of processing technique. The vertical arrows show the standard error for each mean (se=.14).

The results of this analysis are shown in Table 4.12. Neither of the two main effects or the PT x SN interaction was found to be significant. As before, it should be noted, that there were significant inter-subject differences in the pattern of responses. Although these differences were not statistically significant in a repeated measures analysis, they are significant for individual subjects using a fixed effects analysis, as described in the next section.

4.10.2: ANALYSIS OF INDIVIDUAL RESULTS

A fixed- effect ANOVA was performed on the data for each subject. The results of these analyses are summarized in Table 4.13 and show that for, nine of the ten subjects Processing

Technique is significant. The mean scores averaged over speech-to-noise ratio are shown in Figure 4.11. Post hoc analyses using the Tukey Honestly Significant Difference Test showed that for 7 of the 10 subjects the mean diffuseness ratings for both the ITI and IT conditions are significantly higher than for the NP condition. These subjects, for the most part, also showed large differences in the spatial location of the speech and noise images. In contrast, the subjects who rated the NP condition as most diffuse showed relatively small differences in the spatial locations of the speech and noise images.

Table 4.13: Fixed Analysis of Variance for the Rating of Diffuseness for each Subject

Individual ANOVAS	F- RATIO										
	df	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Processing Technique (PT)	3/9	7.52 ***	5.24 ***	1.4	3.67 ***	8.25 ***	22.21 ***	6.94 ***	18.35 ***	3.21 ***	4.54 ***
Speech to Noise Ratio (SN)	3/9	12.61 ***	12.29 ***	1.8	8.25 ***	2.30	1.76	1.40	1.23	1.39	5.61 ***
PT x SN Interaction	9/27	.31	1.76	.86	.57	2.04	.81	.58	1.22	.89	.69
Standard Error of Mean		.16	.20	.28	.16	.28	.16	.14	.19	.21	.13

***= Significance level <.05

FIGURE 4.9: The Effect of Processing Technique (PT) on the ratings of diffuseness of the image. Means that do not differ significantly from each other ($p<0.01$) are grouped together by a superscript line.

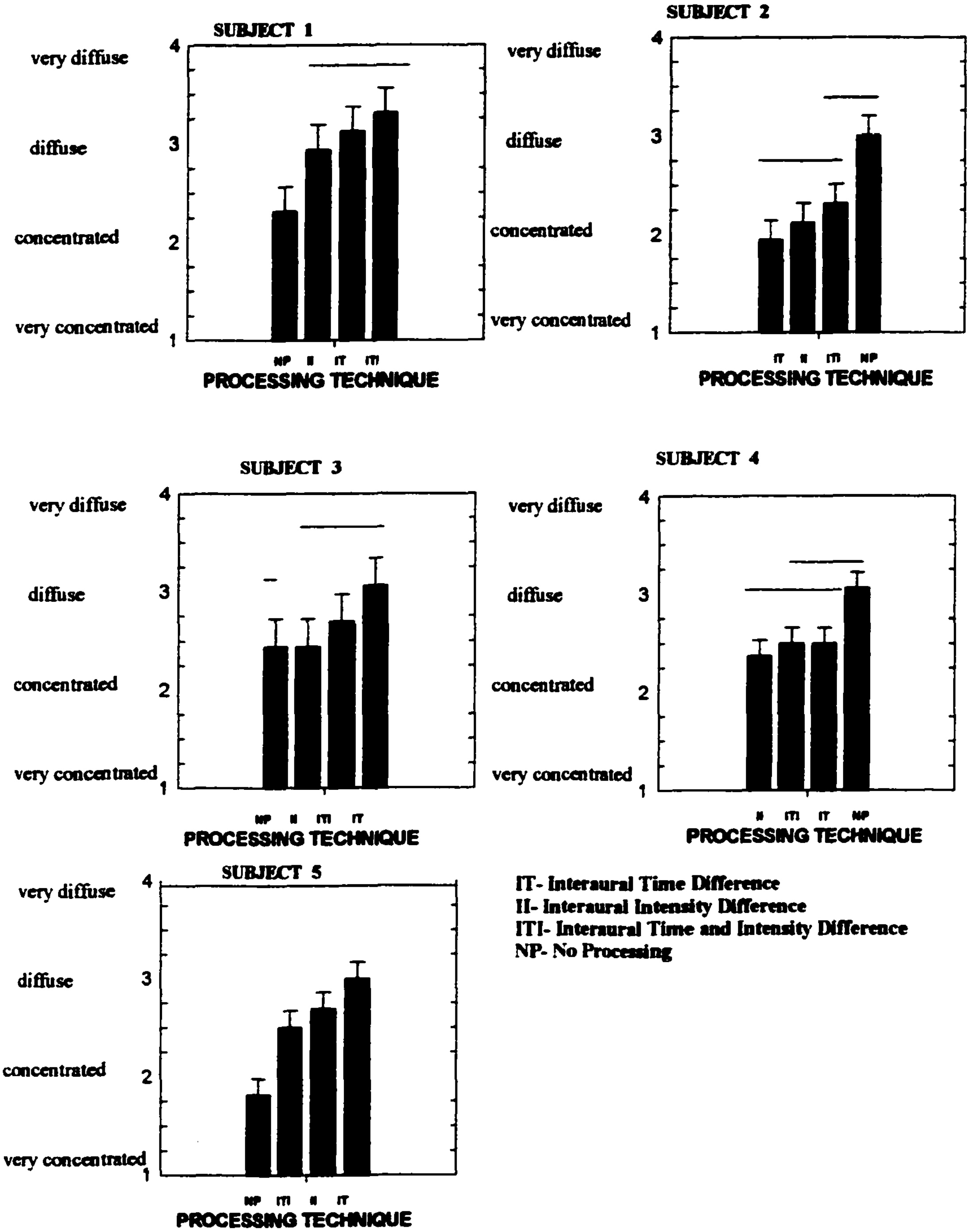
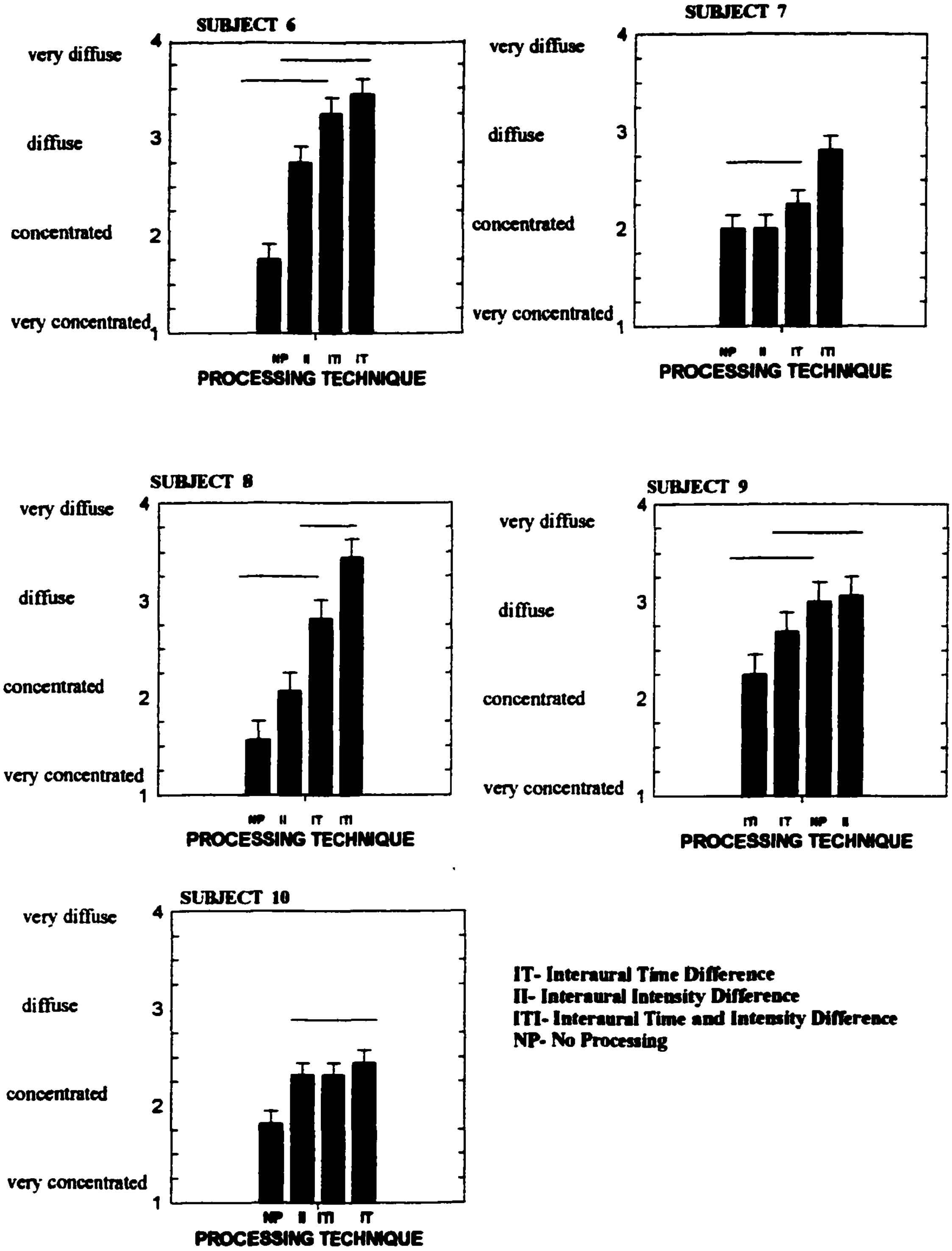


FIGURE 4.9 The Effect of Processing Technique (PT) on the ratings of diffuseness of speech. Means that do not differ significantly from each other ($p<0.01$) are grouped together by a superscript line.



4.3.1: EASE OF LISTENING RESULTS

The ease of listening data took the form of paired-comparisons. All possible paired-comparisons among the four Processing conditions (ITI, IT,II and NP) were tabulated for each subject at each speech-to-noise ratio. The subject was presented with two stimuli in sequence. The processing used for the first stimulus of the pair is identified by the rows. The columns identify the processing used for the second stimulus of the pair. A typical table of this type is shown in Table 4.14. Each cell contains either 1 or 0. An entry of 1 indicates that the stimulus heard first was judged to be easier to listen to than the second stimulus. An entry of 0 indicates the opposite judgment. For example, a stimulus processed by the ITI technique is followed by a stimulus processed by the IT technique. If the subject judges the former to be easier to listen to, a 1 is entered in the cell corresponding to Row ITI and Column II. The proportion of 1's in a given row is a measure of the relative ease of listening for the processing condition represented by that row. In this illustrative example, the row corresponding to the ITI condition has the highest proportion of 1's (3 out of 3 paired comparisons) and receives the highest score for ease of listening (3/3 = 1.0). The row corresponding to the IT condition has the second highest proportion of 1's (2 out of 3) and receives a score of .67 for the condition, and so on.

The set of 12 paired comparisons was repeated 4 times for each subject and each speech-to-noise ratio. A typical set of data showing the row sums for four replications is shown in Table 4.15. The row proportions are then computed by dividing the row sums by 12 since, in this case, each row contains the results of 4 x 3 paired comparisons.

TABLE 4.14:
An illustration of a typical table of tabulation for the paired- comparison task. This table is for one subject's responses for 1 set of paired comparisons among the four Processing Techniques at the -2dB signal to noise ratio level.

SECOND STIMULUS OF THE PAIR

FIRST STIMULUS OF PAIR	IT	II	ITI	NP	ROW SUM	PROPORTION OF 1's
IT	X	1	0	1	2	2/3 = .67
II	0	X	0	1	1	1/3 = .33
ITI	1	1	X	1	1	3/3 = 1.00
NP	0	0	0	X	0	0/3 = 0

TABLE 4.15:
A typical set of responses for 4 replications of the paired- comparison task at a given speech-to-noise ratio(-2 dB).

REPLICATIONS	R1	R2	R3	R4	ROW SUMS FOR 4 REPLICATIONS	ROW PROPORTIONS
ITI	2	3	3	2	10	.83
IT	3	2	2	1	8	.67
II	1	0	2	2	5	.42
NP	1	0	0	0	1	.08

4.3.2: ANALYSIS OF EASE OF LISTENING RESULTS

A data set of the form shown in Table 4.16 was obtained for each subject at each speech-to-noise ratio. A repeated measures analysis of variance (ANOVA) was then performed on the row proportions. As in the previous analyses, the factors were 1) Processing Technique (PT), 2) Speech-to-Noise Ratio (SN) and Subjects (S). Note, that in this analysis the data involved paired comparisons among processing conditions at a common speech-to-noise ratio. For each of the 4x4 paired- comparison matrices, the expected number of 1 response's is 6/12, i.e., half of the cells on average will receive a score of 1, and half will receive a score of 0 (on average). As a consequence, the score for each SN condition will be .5 since half of the cells for each matrix will be 1, on average. As a consequence, this experiment does not allow for a test of signal to noise ratio. It does, however, allow for a test of the PT x SN interaction, that the effect of PT at each SN level can be tested. Since the data took the form of proportions, an arc-sine transformation was used to stabilize the error variance (Winer et al, 1991). The results of the analysis are shown in Table 4.16.

TABLE 4.16:

ANALYSIS OF VARIANCE FOR THE JUDGMENT OF EASE OF LISTENING

	Sums of Squares	**Degrees of Freedom**	**Mean Squares**	**F**	**Significance Level**
Processing Technique (PT)	10.64	3	3.55	10.70	.001
Speech to Noise Ratio (SN)	1.09	3	.363	2.251	NS
PT x SN Interaction	.96	9	.107	1.23	NS
SU x PT Interaction	8.95	27	.332		
SU x SN Interaction	4.35	27	.161		
SU x PT x SN	7.06	81	.087		

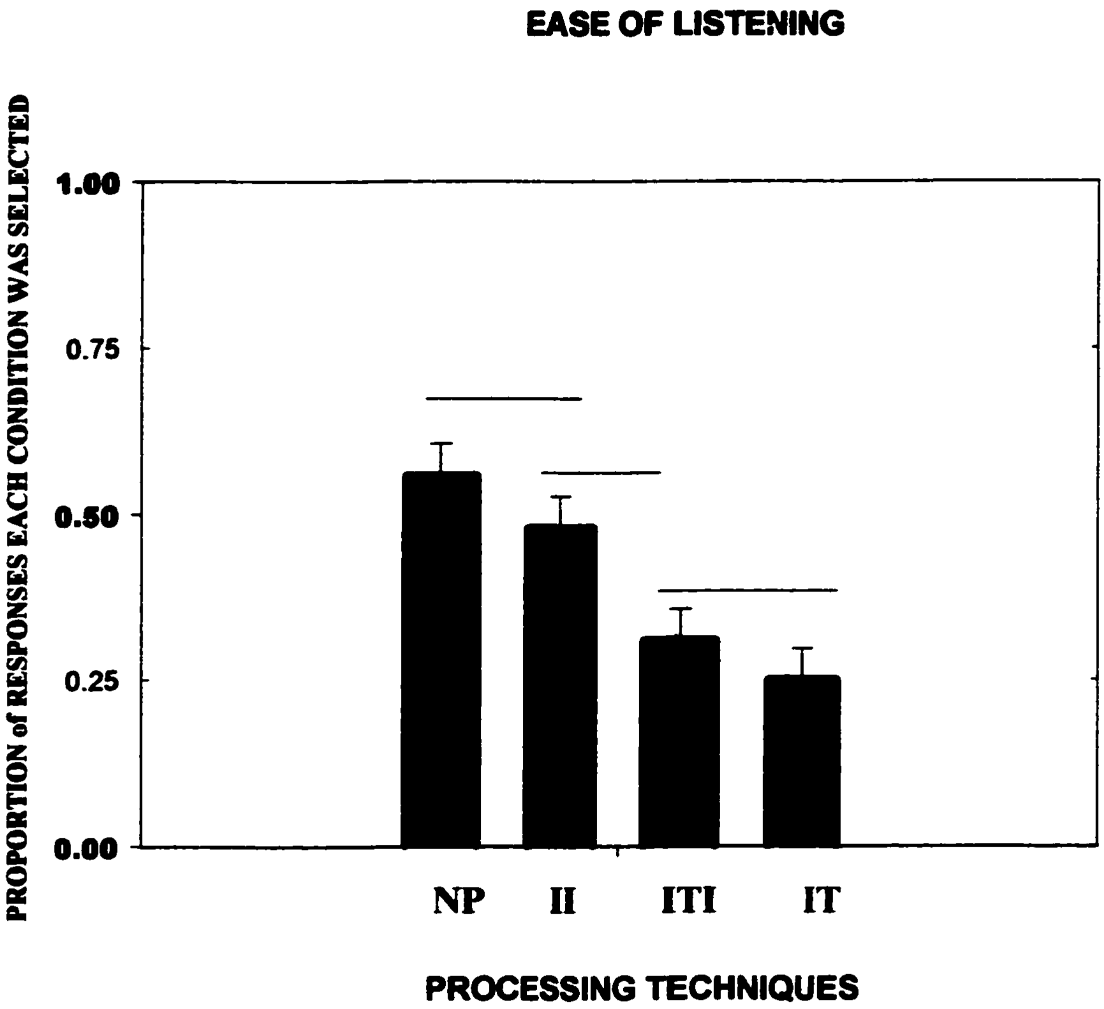


FIGURE 4.12:
The profiles of ease of listening with 4 signal processing techniques for 4 speech-to-noise ratios of 10 normal hearing subjects. Means that do not differ significantly from each other (*p<0.01*) are grouped together by a superscript line.

Processing Technique (PT) was found to be significant (F (3,27)=10.70, *p* <. 001). The mean scores for the four processing techniques are shown Figure 4.12. A post-hoc analysis using the Tukey Honest Significant Difference test showed the NP condition to be significantly different (easier to listen to) than either the ITI or the IT conditions. The mean ease of listening score for the II condition was not significantly different from the NP condition.

4.3.3: ANALYSIS OF INDIVIDUAL EASE OF LISTENING RESULTS

As before, separate fixed ANOVA's were performed for each subject. The mean scores for each subject and the results of the post hoc analysis are shown in Figure 4.13. Processing techniques was not shown to be significant for 4 of 10 subjects. On average, the Processing Technique did not improve ease of listening. However, subjects 2,3, and 9 found either the II or the IT conditions easier to listen to than the NP condition. The subjects were asked their subjective impression of the stimuli. Several subjects commented that the speech and noise image fluctuated during a test sentence. Some subjects noted that they found listening easier in cases in which the stimuli were fixed.

TABLE 4.17:

FIXED ANALYSIS OF VARIANCE FOR EASE- OF- LISTENING JUDGMENTS OF EACH SUBJECT

Individual ANOVAS		F- RATIO									
	df	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Processing Technique (PT)	3	10.83 ***	1.38	2.85	1.57	12.21 ***	4.43	30.55 ***	8.32 **	3.22 *	27.27 ***
Speech to Noise Ratio (SN)	3	7.74 **	1.39	.40	.98	.97	3.11	.37	1.28	.33	7.80 **
Standard Error of Mean		.05	.07	.08	.08	.07	.06	.04	.07	.06	.04

***= Significance level <. 05

FIGURE 4.13 The Effect of Processing Technique (PT) on the profiles of ease of listening for each subject. Means that do not differ significantly from each other (*p<0.05*) are grouped together by a superscript line.

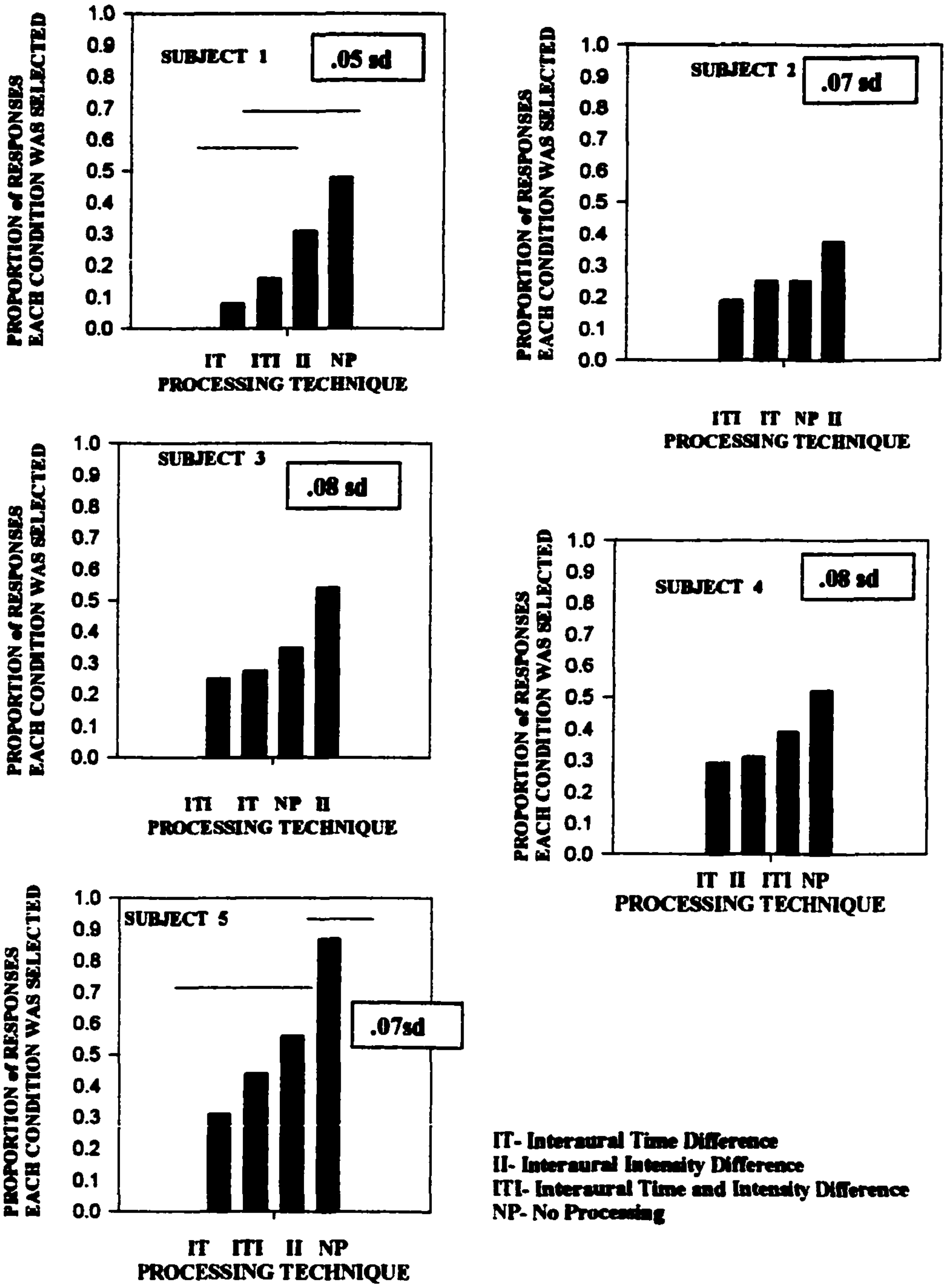
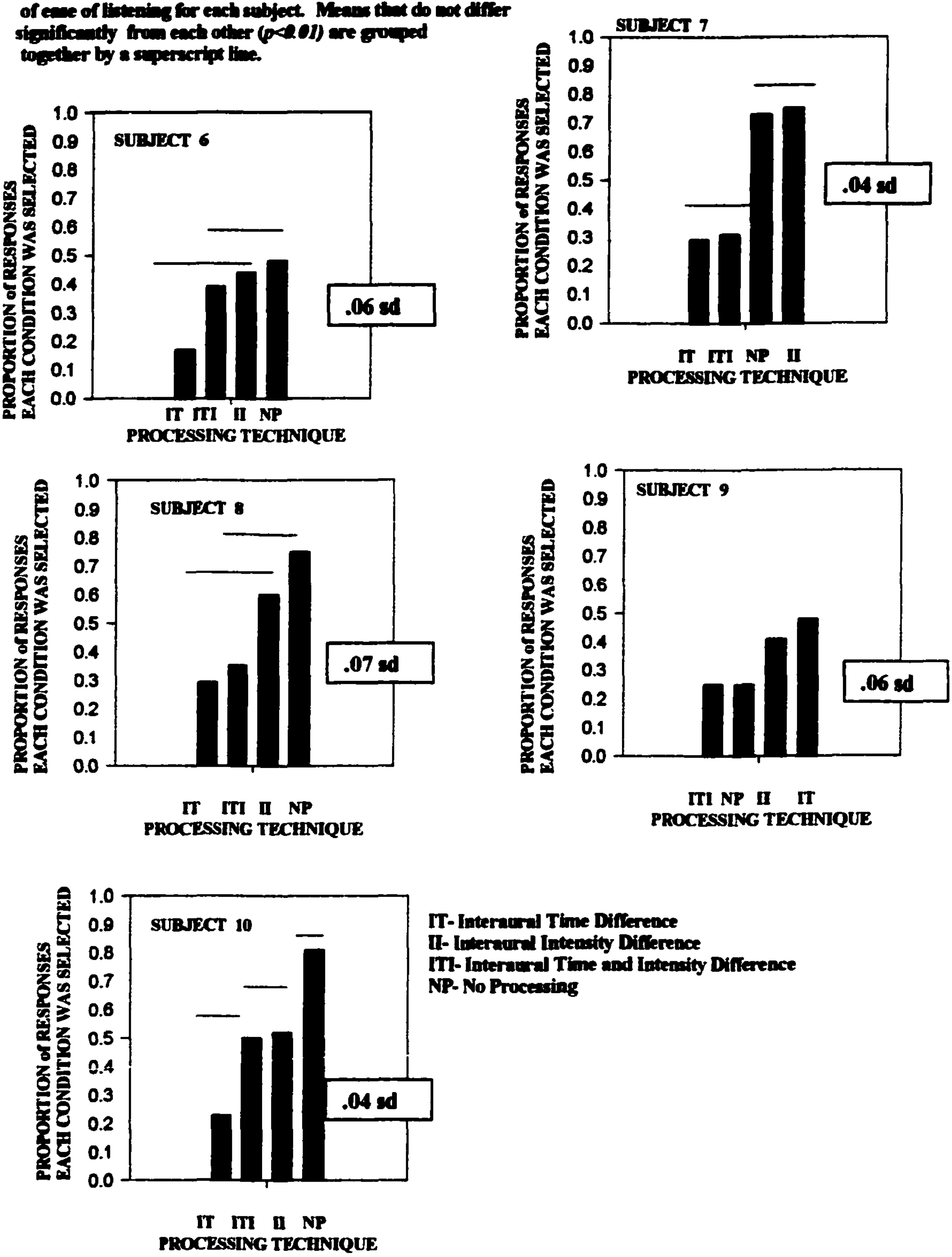


FIGURE 4:13: The Effect of Processing Technique (PT) on the profiles of ease of listening for each subject. Means that do not differ significantly from each other ($p<.01$) are grouped together by a superscript line.



CHAPTER V

DISCUSSION

This study evaluated a new method of signal processing for converting a monotic signal to a dichotic signal for the purpose of separating speech from background noise. Of particular interest was the effect of the technique on the apparent spatial locations of speech and noise and also on ease of listening. Three variations of the processing technique were investigated with no processing (NP) serving as the control condition. These were:

1) Interaural Intensity Differences, (II), applied differently to the spectral peaks and residual,
2) Interaural Time Differences, (IT), applied differently to the spectral peaks and residual,
3) Interaural Time and Intensity Differences (ITI) applied differently to the spectral peaks and residual.

Four Hypotheses were tested.

Experimental Hypothesis #1

Experimental Hypothesis #1 states, "Interaural intensity differences applied to the spectral components of the speech plus noise, as described in the Methods Section, produce differences in the apparent spatial locations of the speech and noise."

The experimental data support this hypothesis. The method of signal processing using interaural intensity-differences (II) produced a statistically significant difference in the locations of the speech and noise images. For interaural intensity differences, the speech image was located near the center of the head or slightly to the right while the noise image was located near the center of the head or slightly to the left.

On average, the magnitude of the effect of II processing on the spatial location of the speech image on the X-axis was a shift to the right of 0.47 distance units (du) from the control condition (no processing, NP). A shift of 0.47 du to the right corresponds to 11.75% of the distance from the head center to the right ear. The magnitude of the average shift in location for the noise image was 0.45du to the left, or an 11.25% shift to the left ear. The overall difference in location produced by the II condition was thus .92du (.47du + .45du), which corresponds to 23.0% of the head's radius (from head center to each ear).

The magnitude of the shift in spatial location is less than that which would have been obtained had either the speech or noise been presented alone and subjected to the interaural intensity differences used in this study. Tobias and Curtis (1959) observed that an interaural intensity difference of 15 to 20 dB resulted in the sound image being heard almost at the ear receiving the more intense signal. An average shift of 11.5% to one ear would thus correspond to an II difference of 1.7 to 2.3 dB. A more precise comparison is possible in that previous studies have indicated for interaural intensity differences of less than 10 to 12 dB there is almost a linear relationship between perceived lateral location and the interaural intensity difference expressed in decibels (Yost, 1981; Watson and Mittler 1965). Linear interpolotion can thus be used to estimate the equivalent interaural intensity differences that would produce the observed spatial location for speech or noise alone.

In order to compare the current findings to previous studies we need to account for the bias that some individuals exhibited. By definition, the bias is the amount by which the sound image deviates from the center of the head for the No Processing (NP) condition is, (i.e., a signal presented diotically is heard at the center of the head if the two ears are equally sensitive and there is no judgmental bias). The bias observed for each subject is reported in Table 5.1

TABLE 5:1: SUMMARY OF INDIVIDUAL RESULTS ACCOUNTING FOR BIAS

Speech Spatial Separation				**Noise Spatial Separation**			
SUBJECT	**II**	**IT**	**ITI**	**SUBJECT**	**II**	**IT**	**ITI**
1	0.76	0.02	1.2	*1*	-0.3	-1.45	-1.27
2	-0.01	-0.17	1.17	*2*	-0.22	-1.38	-0.95
3	-0.39	-0.69	0.48	*3*	-0.1	-1.39	-1.79
4	0.72	0.21	0.74	*4*	-0.3	0.08	0.14
5	0.98	1.08	1.47	*5*	-1.04	-1.17	-1.62
6	1.09	1.02	1.31	*6*	-0.75	-1.52	-1.62
7	1.42	1.4	1.73	*7*	-1.37	-1.74	-2.25
8	-0.12	-3.58	-2.19	*8*	-0.06	-4.24	-3.4
9	0.13	0.41	1.2	*9*	0.08	-1.3	-1.88
10	0.18	0.04	0.32	*10*	-0.4	-2.34	-2.54
MEAN	.47	-.08	.74		-.45	-1.64	- 1.78
MEAN BAIS	***.042***					***.073***	

Note that the mean bias, averaged over subjects does not differ significantly from zero; i.e., there is no bias on average, although several subjects show significant bias to either the left or right.

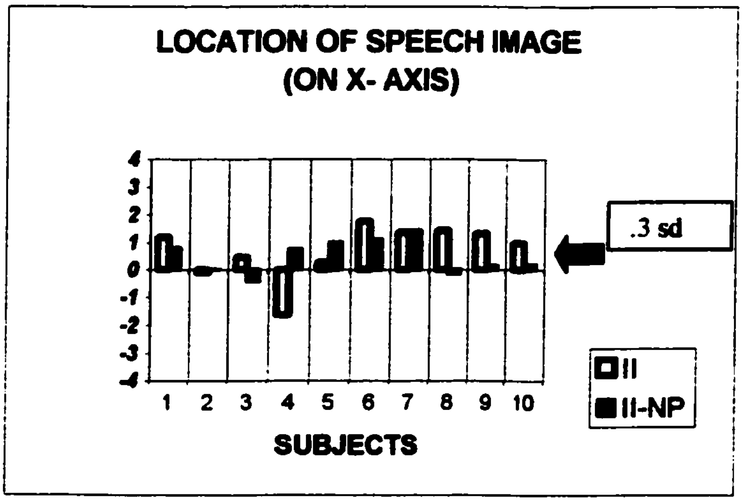


Figure 5.1a: The effect of II Processing Technique on Speech. Spatial location for each subject shown with the open bar and Corrected spatial location (for observer bias) shown with the solid bar. The standard deviation for the corrected spatial location was .3 du.

Figure 5.1a shows the spatial locations for speech after correcting for individual biases. The open bars show the spatial locations as reported by each subject. The solid bars show the spatial locations after correcting for observer bias. These bars represent the difference in location between the II and NP conditions, i.e., they show the change in spatial location of the speech image resulting from the processing. The effect of II processing on the spatial location of the speech image is thus seen to be a shift to the right Subjects 1,4, 5, 6,7,and 9. The remaining subjects (2,3,8,and 10) did not show a significant shift in the location of the speech image with II processing.

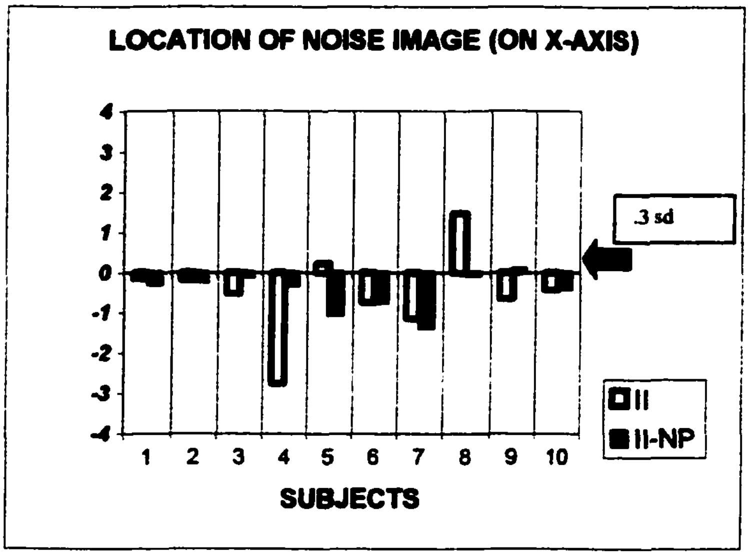


Figure 5.1.b: The effect of II Processing Technique on Noise. Spatial location for each subject shown with the open bar and Corrected spatial location (for observer bias) shown with the solid bar. The standard deviation for the corrected spatial location was .3du.

The effect of II processing on the noise image is summarized in Figure 5.1b. The II processing condition showed a shift to the left for Subjects 1,4,5,6,7and 10. With the exception of Subject 8 who showed a shift to the right the remaining subjects showed no significant change in location of the noise image with II processing.

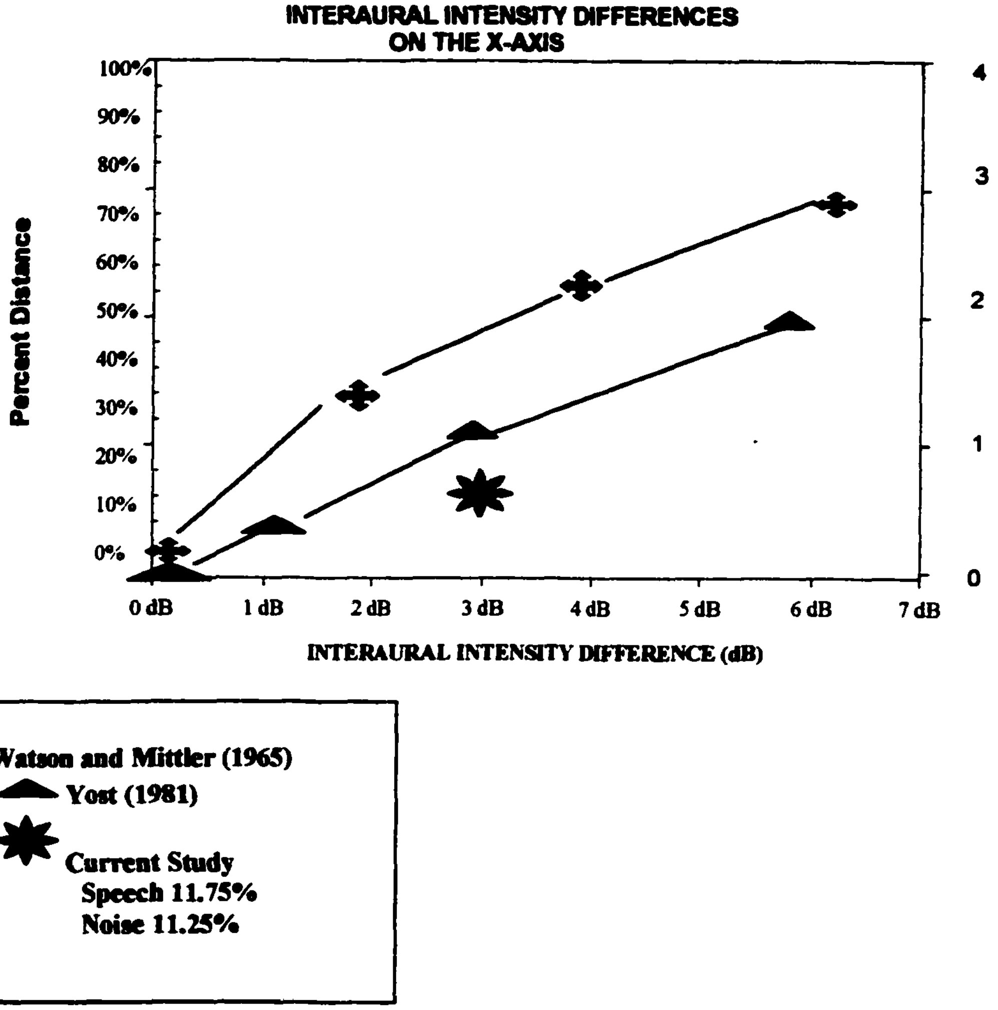


FIGURE 5.1c: The perceived location of pure tones (Watson and Mittler, (1965); Yost (1981) as a function of the interaural intensity difference. The location of speech and noise image for the current study as a function of II condition. The perceived location is normalized across the studies in terms of the percent distance from an image lateralized at midline (0%) and one lateralized entirely toward one ear (100%).

Figure 5.1c compares the results of the current study with those of previous studies. A 500-Hz tonal pulse was used in the Watson and Mittler (1969) study and a 500 Hz pure tone in the Yost (1981) study. The average data for these two studies are shown in Figure 5.1c. The average locations for the speech and noise image in the current study (after correction for individual subject bias) are shown by the S and N symbols, respectively in Figure 5.1c. Note that these locations are about half the distance from the head center for tonal stimuli in isolation (no background noise) as reported by Yost (1981) and Watson and Mittler (1965).

The II processing technique did not have a significant effect on spatial location of the speech and noise on the front-back axis (the Y-axis). This is consistent with previous studies of Licklider (1948, 1962).

Experimental Hypothesis #2

Experimental Hypothesis #2 states, "Interaural time differences applied to the spectral components of the speech plus noise, as described in the Methods Section, produce differences in the apparent spatial locations of the speech and noise."

Overall, the findings support this hypothesis. The IT processing technique had a significant effect on the spatial locations of the speech and noise images on the horizontal axis, but not on the front-back axis.

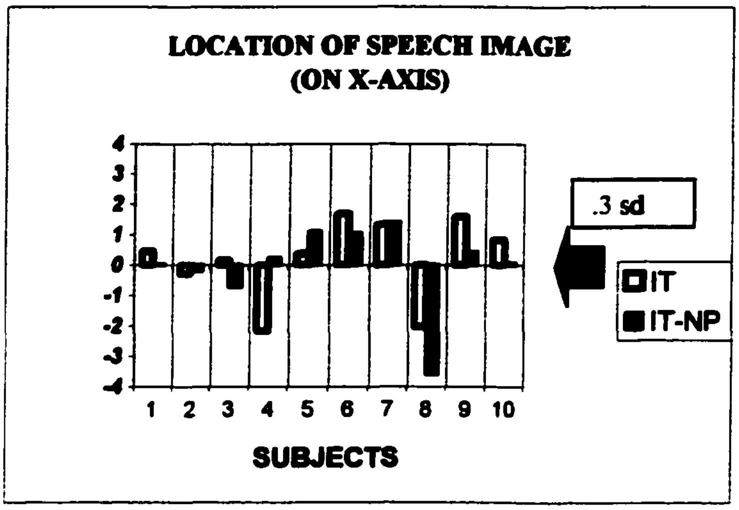


FIGURE 5.2a: The effect of IT Processing on Speech.
Spatial location for each subject shown with the open bar and
Corrected spatial location (for observer bias) shown with the solid bar.
The standard deviation for the corrected spatial location was .3du.

The magnitude of the effect with IT processing on the speech image was less than the II processing technique, as shown in Table 5.1 and Figure 5.2a. Figure 5.2a and Figure 5.2b show the effect of the IT processing on the speech and noise images, respectively. As before, the open bars show the locations as reported by each subject and the solid bars show the shift in location after correcting for observer bias.

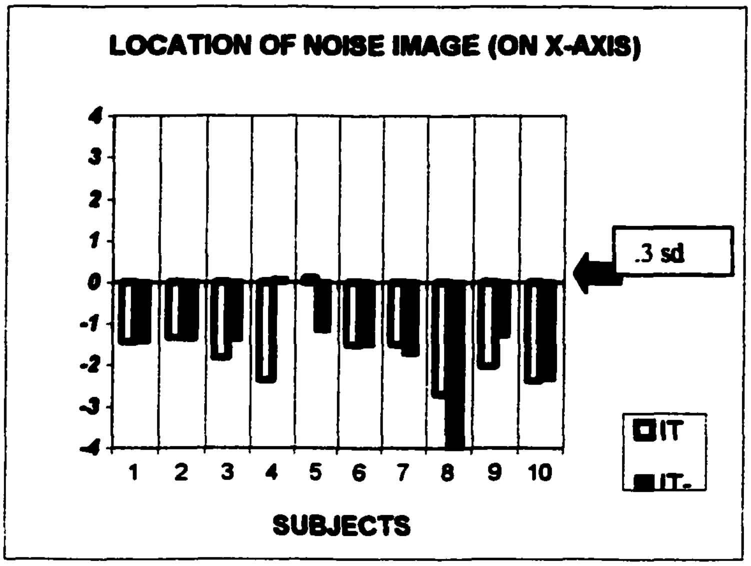


**FIGURE 5.2b: The effect of IT Processing on Noise.
Spatial location for each subject shown with the open bar and
Corrected spatial location (for observer bias) shown with the solid bar.
The standard deviation for the corrected spatial location was .3du.**

The magnitude of the effect with IT processing on the noise image was much greater than for the II condition (-1.64du on average after correcting for observer bias, or -1.85du without this correction) as shown in Figure 5.2b. This shift of the noise image for the IT condition was almost 4 times greater than that for the noise image for the II processing condition (0.45 du). The shift of the noise image was also substantially larger than that for the speech image (for the IT condition).

It is again interesting to compare the findings with previous investigations in which sound images are shifted by means of interaural time differences. Figure 5.2c compares the results of the current study with other published research.

Similar to the II processing technique, the magnitude of the shift of the speech and the noise images with the IT processing is about half of that reported for 500Hz tonal signals in isolation (i.e., no background noise) as reported by Sayers (1964), Watson (1965), Yost (1981).

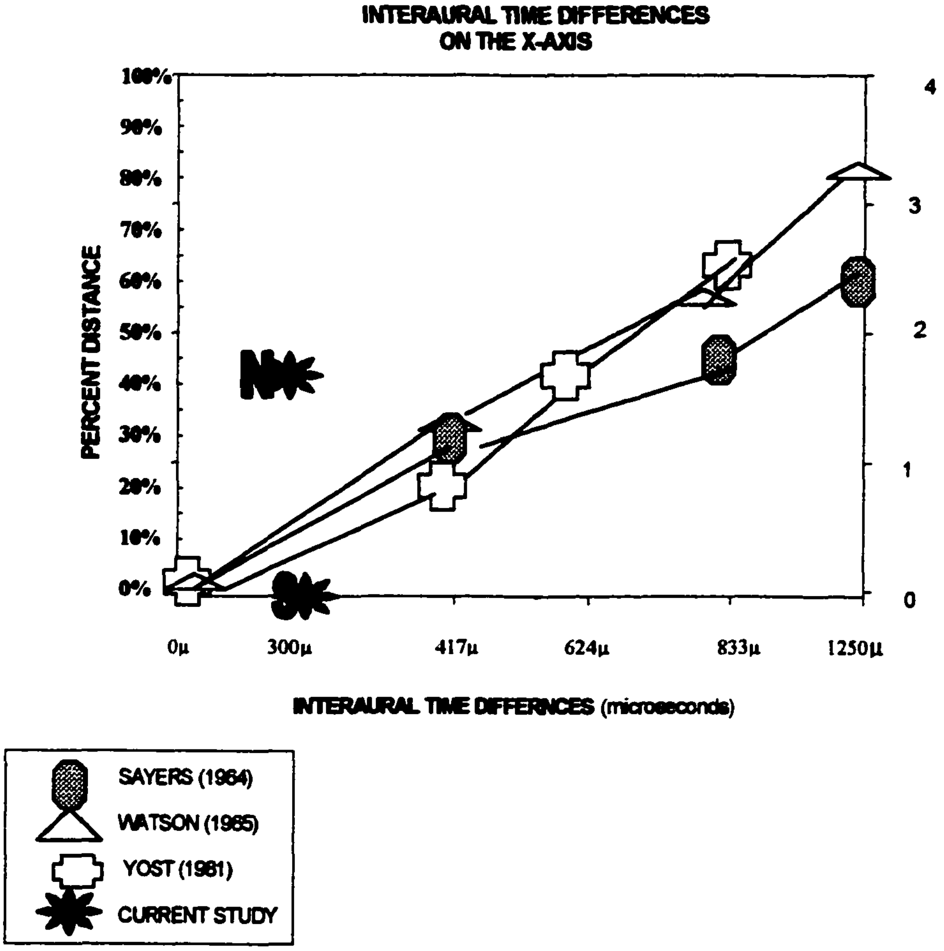


FIGURE 5.2c: The perceived location of pure tones (Sayers (1964), Watson and Mittler, (1965); Yost (1981) as a function of the interaural time difference. The location of speech and noise image for current study as a function of IT condition. The perceived location is normalized across the studies in terms of the percent distance from an image lateralized at midline (0%) and one lateralized entirely toward one ear (100%).

Experimental Hypothesis #3

Experimental Hypothesis #3 states, "The combination of the two processing techniques will produce even greater differences in the apparent locations of the speech and noise."

The experimental data supports this hypothesis. The shifts in location on the X-axis for speech, for noise and for the difference between the speech and noise were significantly larger than for either the II or IT conditions.

On average, the spatial separation between the speech and noise images for the II condition was .92 du, and 1.75 du for the IT condition. For the ITI condition it was 2.52 du. The magnitude of the shift in spatial separation for the ITI processing is roughly equal to the sum of shift on location for II and IT processing. The sum of the shifts for II and IT processing conditions does not differ significantly from the shift in location for the ITI condition, except for Subject 8 whose data were erratic.

Figure 5.3a and Figure 5.3b show the effect of the ITI processing on the speech and noise images, respectively. As before, the open bars show the locations as reported by each subject and the solid bars show the shift in location after correcting for observer bias.

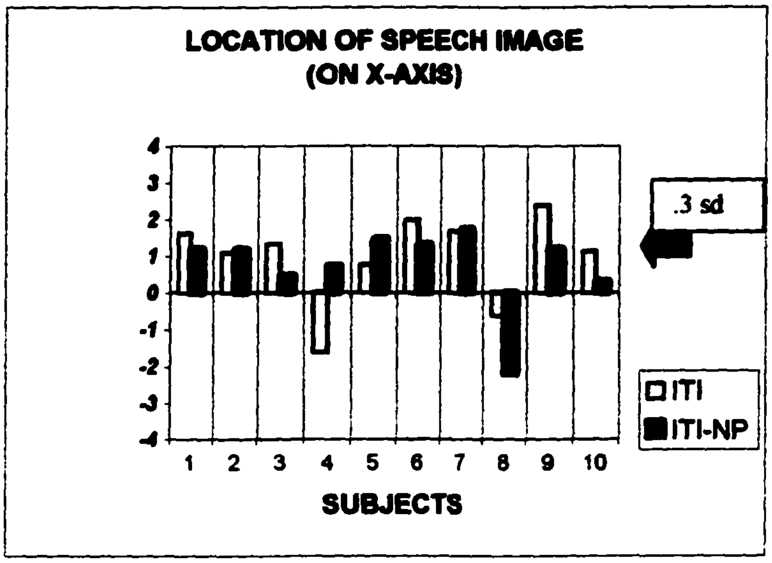
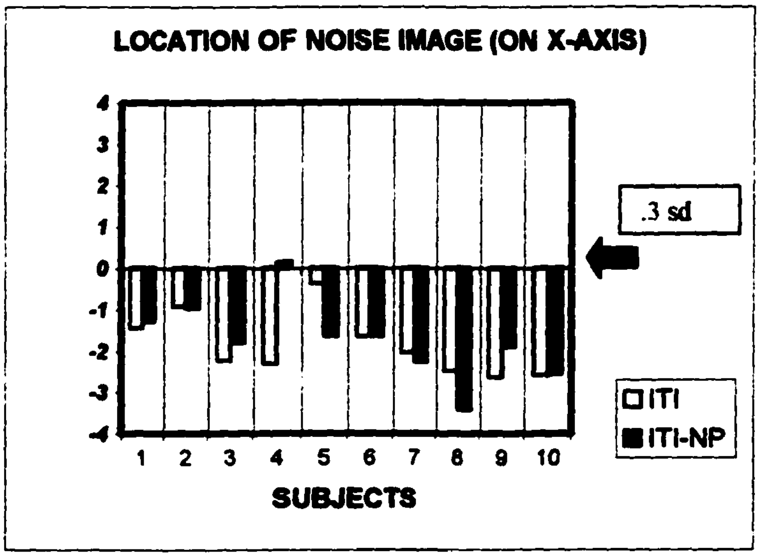


FIGURE 5.3a and 5. 3b : (above and below) The effect of ITI Processing on Speech and Noise. Spatial location for each subject shown with the open bar. Corrected spatial location (for observer bias) shown with the black bar. The standard deviation for the corrected spatial location was .3du.



Of the four processing conditions considered in this study (II, IT, ITI and NP), ITI processing was the most promising for separating speech from noise. The difference in spatial location between the speech and the noise images, the ITI processing conditions was 2.91 du, or almost three-fourths the distance between the center of the head and the ear.

Similarly to the II and IT processing conditions, the ITI processing condition also did not have a significant effect on spatial location of the speech and noise on the Y- axis. This is consistent with previous studies of Licklider (1948) and others.

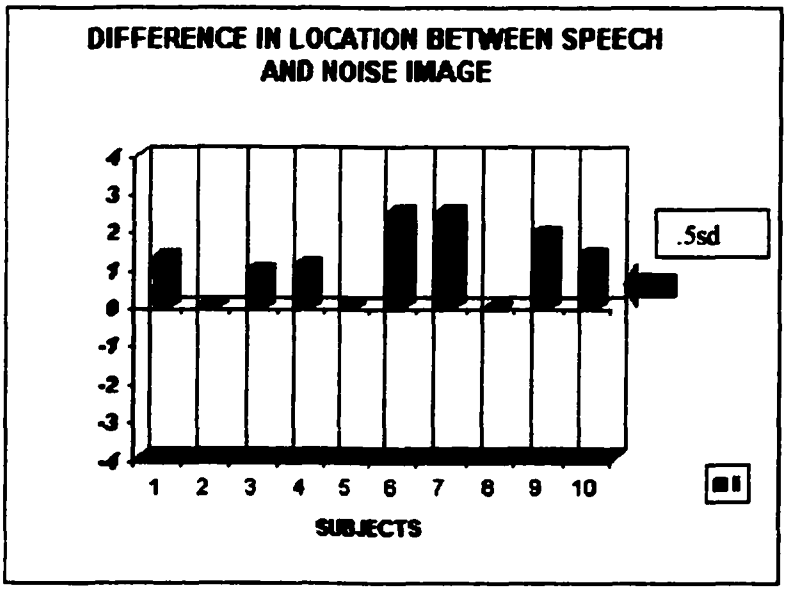
INDIVIDUAL RESULTS

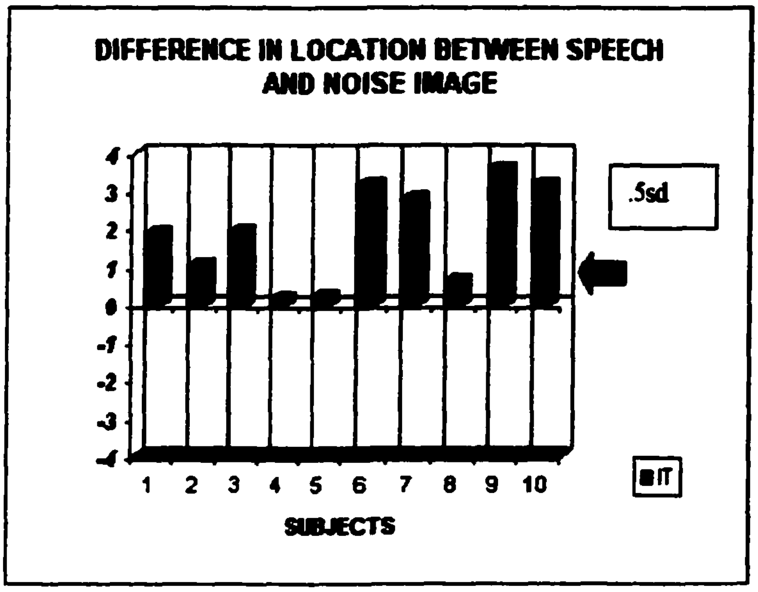
As noted above, there were several subjects who exhibited a large bias. Specifically, a significant bias (greater than 3sd) was shown for Subjects 4 and 8 when judging the location of the speech image and the noise image separately. Interestingly, subject 4 had a significant bias to the left side of the head for both speech and noise. As previously noted subject 8 differed significantly from the other subjects. No obvious reasons could be found for this subject's performance other than perhaps a misunderstanding of the instructions.

If observer bias is taken into account, the ITI processing technique produced a significant change in separation of the speech and noise image for 7 out of 10 subjects (*p=<.05*). Of the remaining three subjects, two were erratic (Subject 4 and Subject 8) in their judgments with poor consistency between conditions. Subject 5 showed a relatively large effect although it was not statistically significant. This subject, however, showed significant changes in location for the speech and noise images, but not for speech-noise.

Figure 5 (a, b and c) shows the spatial location for the speech-noise for each of the processing conditions of II, IT and ITI.

FIGURE 5.5a: The effect of Processing Technique (II) on the Spatial Separation of the Speech and Noise Image. Spatial location for each subject shown with the gray bar. The standard deviation for the corrected spatial location was .5du.





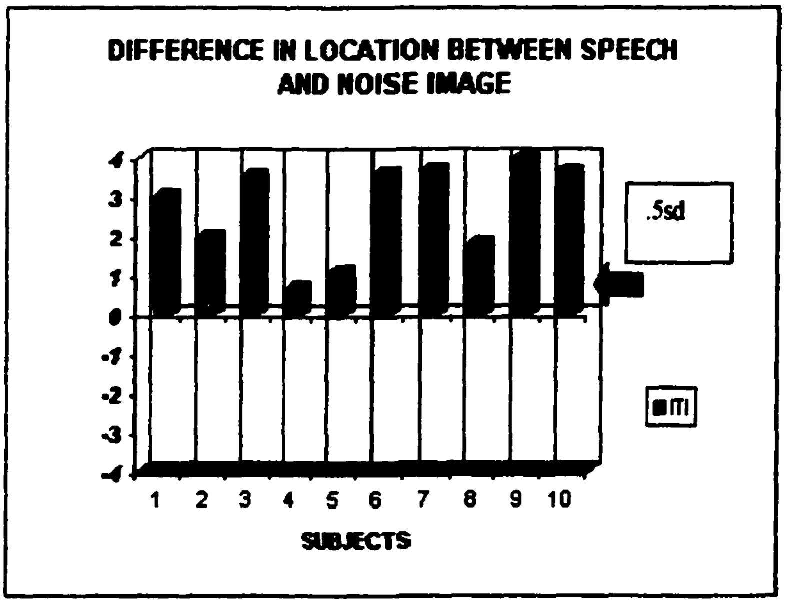


FIGURE 5.5b, and 5.5c: The effect of Processing Technique (IT and ITI) on the Spatial Separation of the Speech and Noise Image. Spatial location for each subject shown with the gray bar. The standard deviation for the corrected spatial location was .5du.

DIFFUSENESS

To compliment the data on the spatial separation, judgments were also obtained on the diffuseness of the overall sound image. Seven of the ten subjects indicated the sound image to be the most diffuse with either the IT or ITI processing technique. The diffuseness of the sound image and the separation of the speech and noise images had a similar trend. The processing condition (ITI) that demonstrated the greatest magnitude of separation between the speech and noise image also showed the greatest diffuseness of the sound image. The scale of diffuseness that the subjects selected on the average was quite narrow. A difference of less than half a unit of the rating scale was obtained for the two most extreme conditions (i.e., 0.47 rating scale units on a 5-point scale for the difference between the ITI condition, and the no processing (NP) condition, which was the most concentrated.

THE STIMULI

Interestingly the magnitude of the effect of IT and the ITI processing technique was greater on the noise image compared to the speech image. Since the spectral peaks of the signal are delayed to the left ear it is to be expected that the sound image associated with the spectral peaks (the speech) would be heard on the right side of the head. Similarly, it is to be expected that the sound image associated with the less intense spectral components (the noise), which are delayed to the right ear, would be heard on the left side of the head. Since only the largest spectral peaks were delayed to the left ear, not all of the speech signal was delayed to the left side. In contrast almost all of the noise was delayed to the right ear. This could explain why the magnitude of the effect for the speech image was less than that for the noise image for the IT and ITI condition

Experimental hypothesis #4

Experimental hypothesis #4 states, " The apparent spatial separation between the speech and noise improves ease of listening to the speech."

The ease of listening data showed a significant effect, across all subjects and signal to noise ratios. Of the four processing techniques considered, the No Processing condition showed the highest average rating for ease of listening. Analysis of the data for individual subjects, however, showed that least four subjects (2,3,7,and 9) showed significantly higher ratings for either the II or IT processing technique.

Contrary to expectation, the improvements in ease of listening did appear to be not correlated with degree of spatial separation, since the conditions showing improved ease of listening also showed the smallest differences in spatial separation between the speech and noise images. A possible explanation for this finding is that there was some instability in the location of the image. Several subjects's commented after the test that the speech image for the condition they preferred was more stable than for the other conditions. This was clearly the NP condition since it did not involve any processing.

Processing research on speech enhancement methods have shown that an improvement of the signal- to -noise ratio does not necessarily result in an enhancement of speech intelligibility (Lim, 1983). Previous studies using sine wave modeling as a noise reduction technique have shown to have a benefit but at the expense of an audible signal-processing distortion, Kates (1991, 1994). Kates, (1991 and 1994) using sine wave modeling as a noise reduction technique showed that although the speech to noise ratio was increased the process of discarding the low-level spectral components resulted in an audible-processing distortion that counteracted the effect of an improved speech to noise ratio.

The signal processing technique used in the present investigation did not discard the remaining spectral components, but instead introduced interaural time and/or intensity differences. The perceived effect was that of a change in the apparent spatial locations of the speech and noise, without the processing distortions of the type reported by Kates et al. (1991, 1994). A new kind of distortion was reported, however, that of instability in the spatial location of the sound images; i.e., several subjects reported that the sound images seemed to "wobble" or move about their average location. Some of the subjects found this spatial instability distracting, which may account for the No Processing condition receiving higher ease of listening ratings for most (but not all) of the subjects.

FUTURE RESEARCH

The data from this study reveal the complexity of the relationship between the experimental factors. Whereas the processing technique was effective in separating speech from noise in terms of their apparent spatial locations, it was not effective in improving ease of listening for most of the subjects. As noted above, this was believed to be the result of instability in the spatial locations of the sound images. Further research on methods of signal processing that will stabilize the image location is called for, followed by experiments to determine if a stable spatial separation between speech and noise improves ease of listening and its effect on speech intelligibility.

While beyond the scope of the present investigation, the processing technique already developed can be further refined to alleviate the movement of the image.

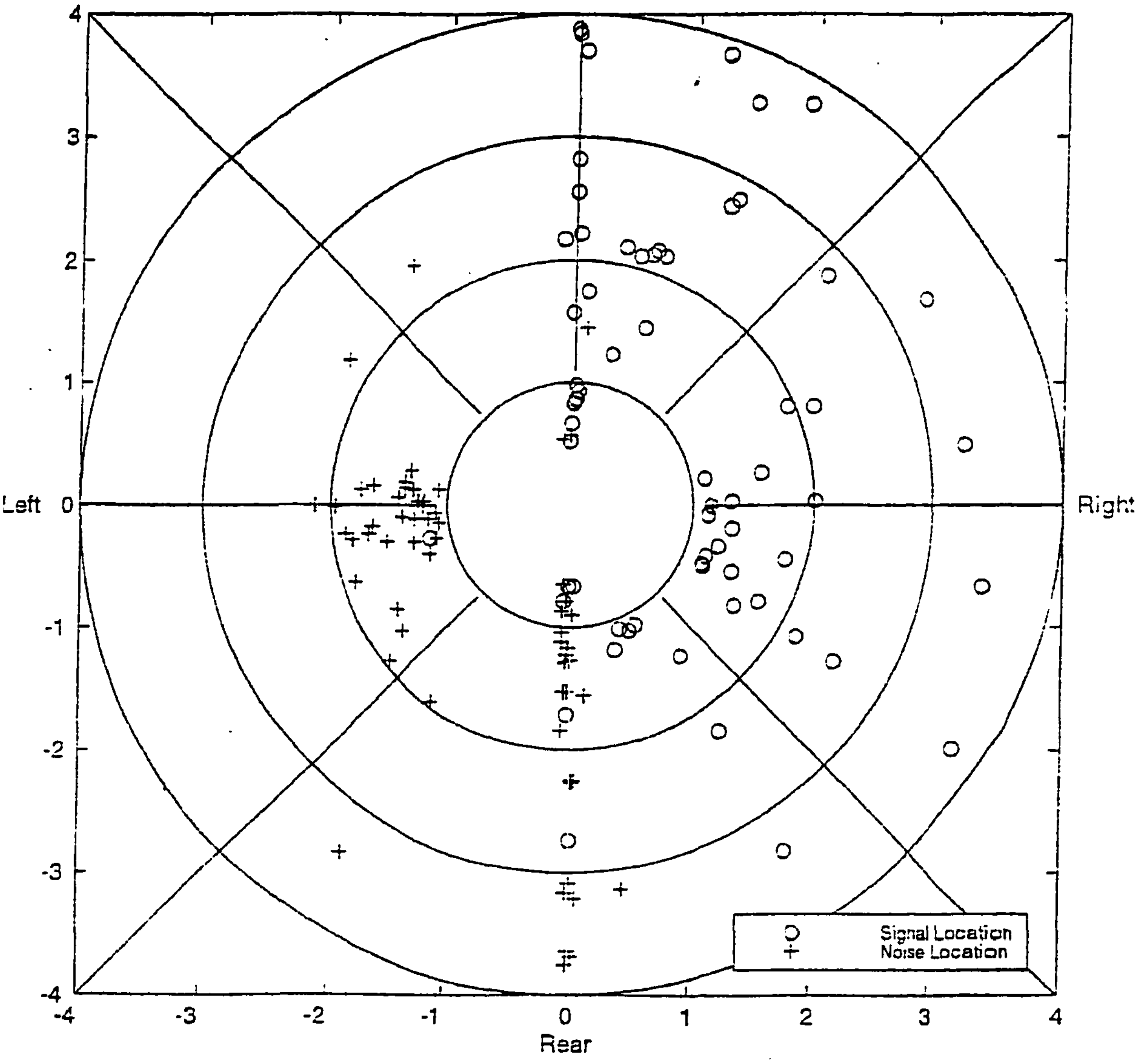
CONCLUSION

The primary focus of this investigation was to evaluate the effect of simulated binaural separation on the perception of speech against a background of noise. The processing conditions had a highly significant effect on the location of the sound image on the X-axis. The data supported that the II and IT processing conditions produced a difference in the spatial location of the speech and noise on the X-axis. The combination of the two processing conditions (ITI) produced an even a greater difference on the X-axis. Although the processing conditions did not have an effect on ease of listening subjects' comments support further investigation.

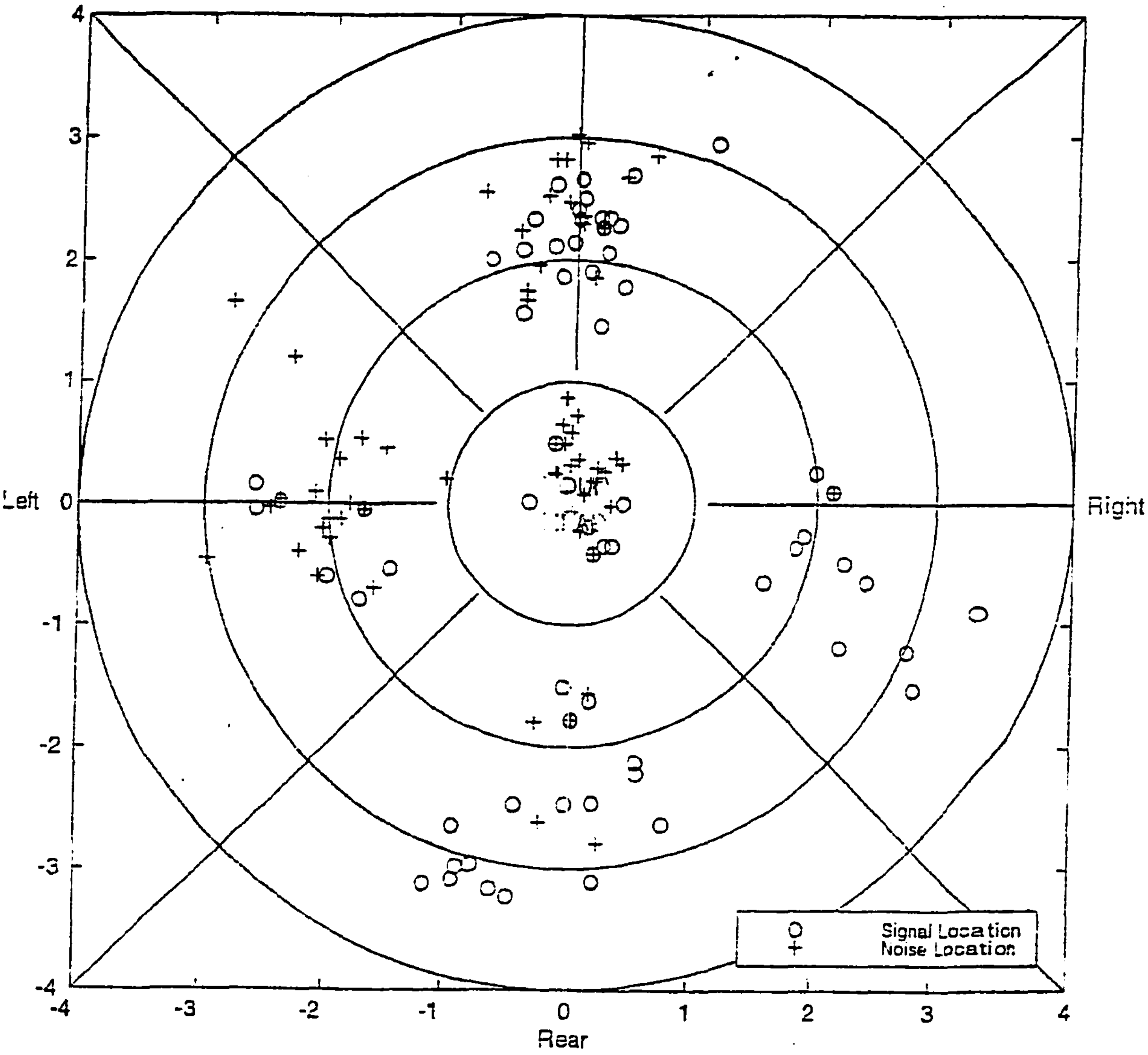
APPENDIX A

SUBJECT SPATIAL LOCATION RESULTS

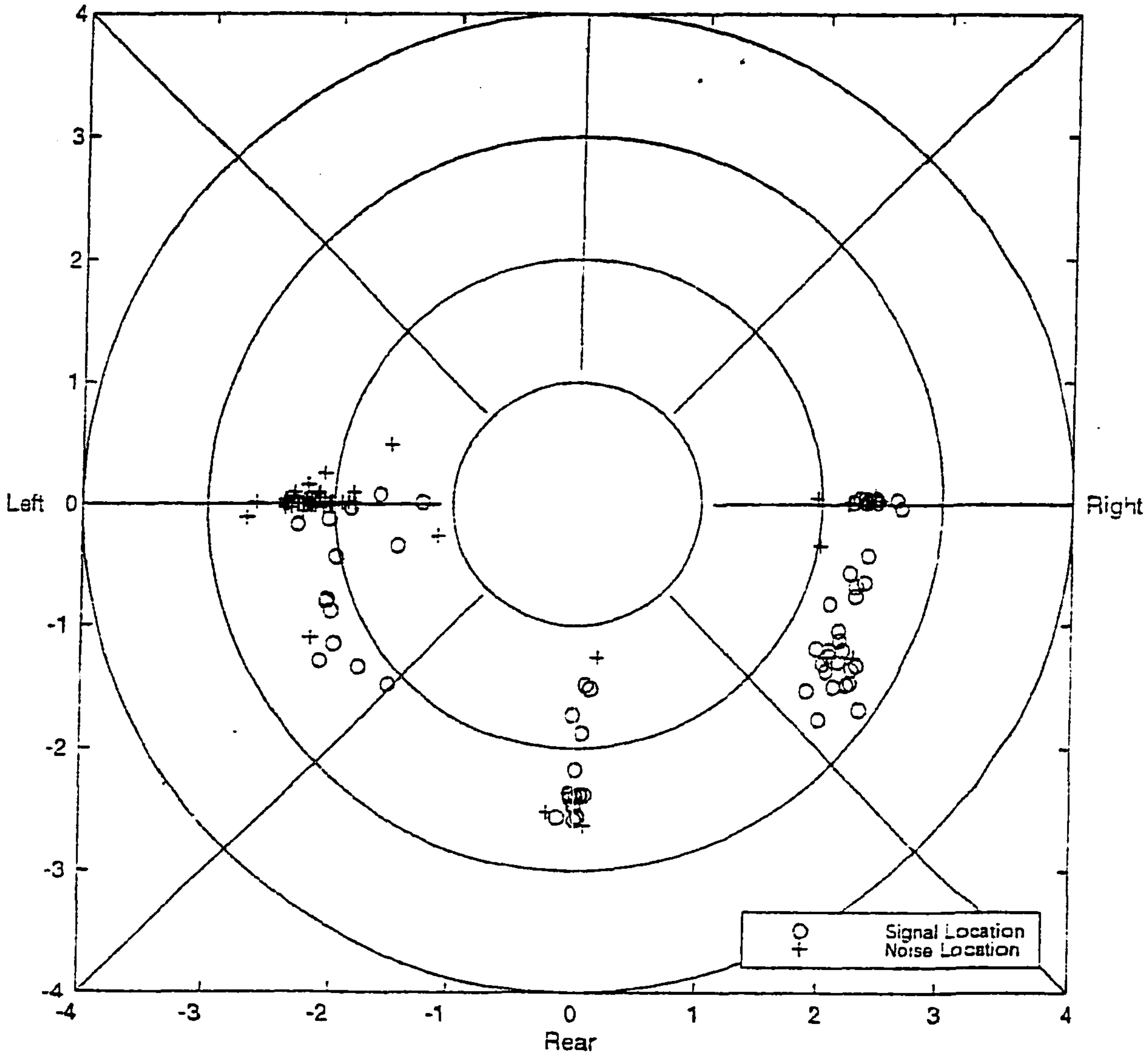
S1


Left
Right
Rear
4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Signal Location
Noise Location

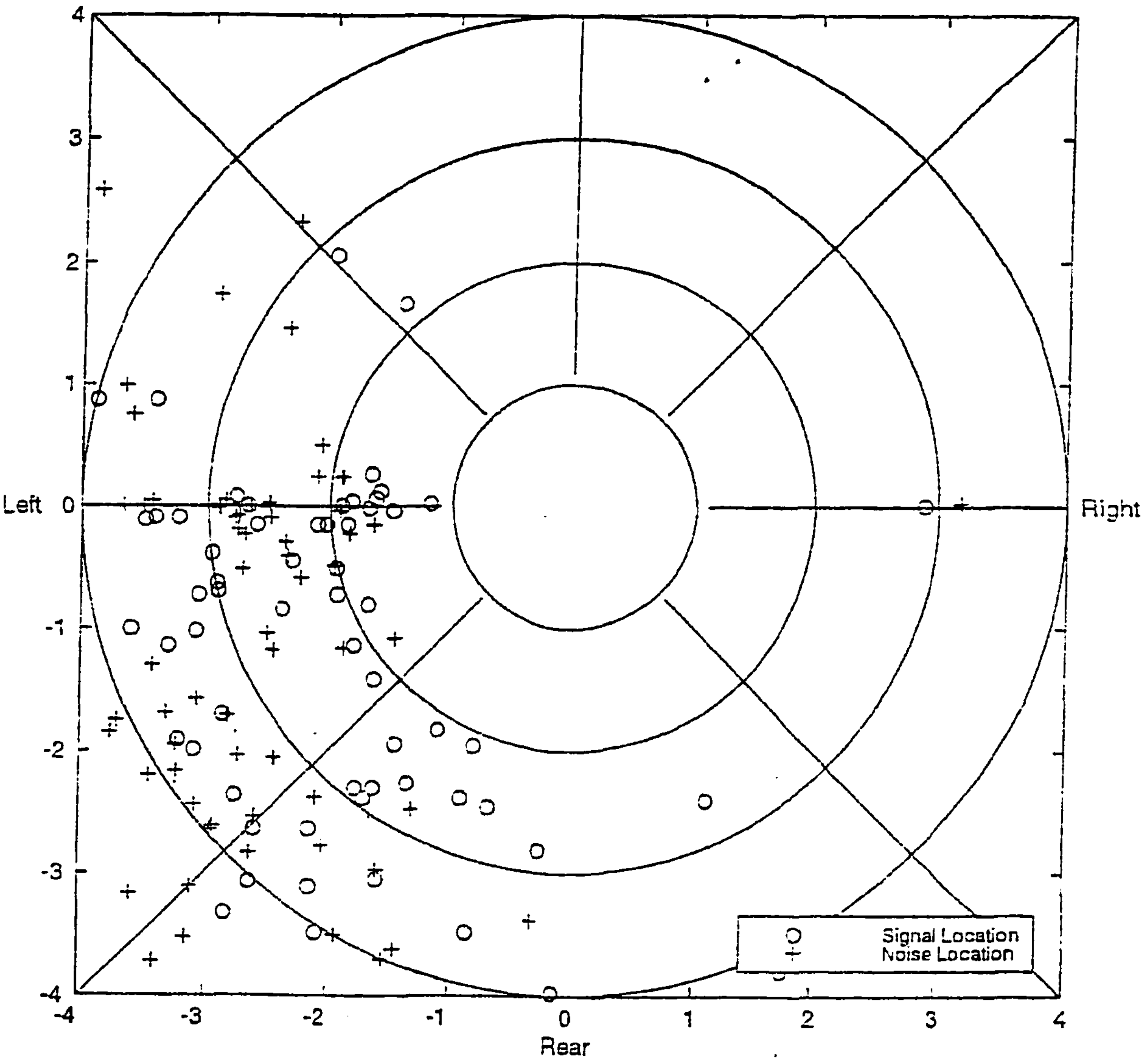
S2


4
3
2
1
Left 0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Rear
Right
Signal Location
Noise Location

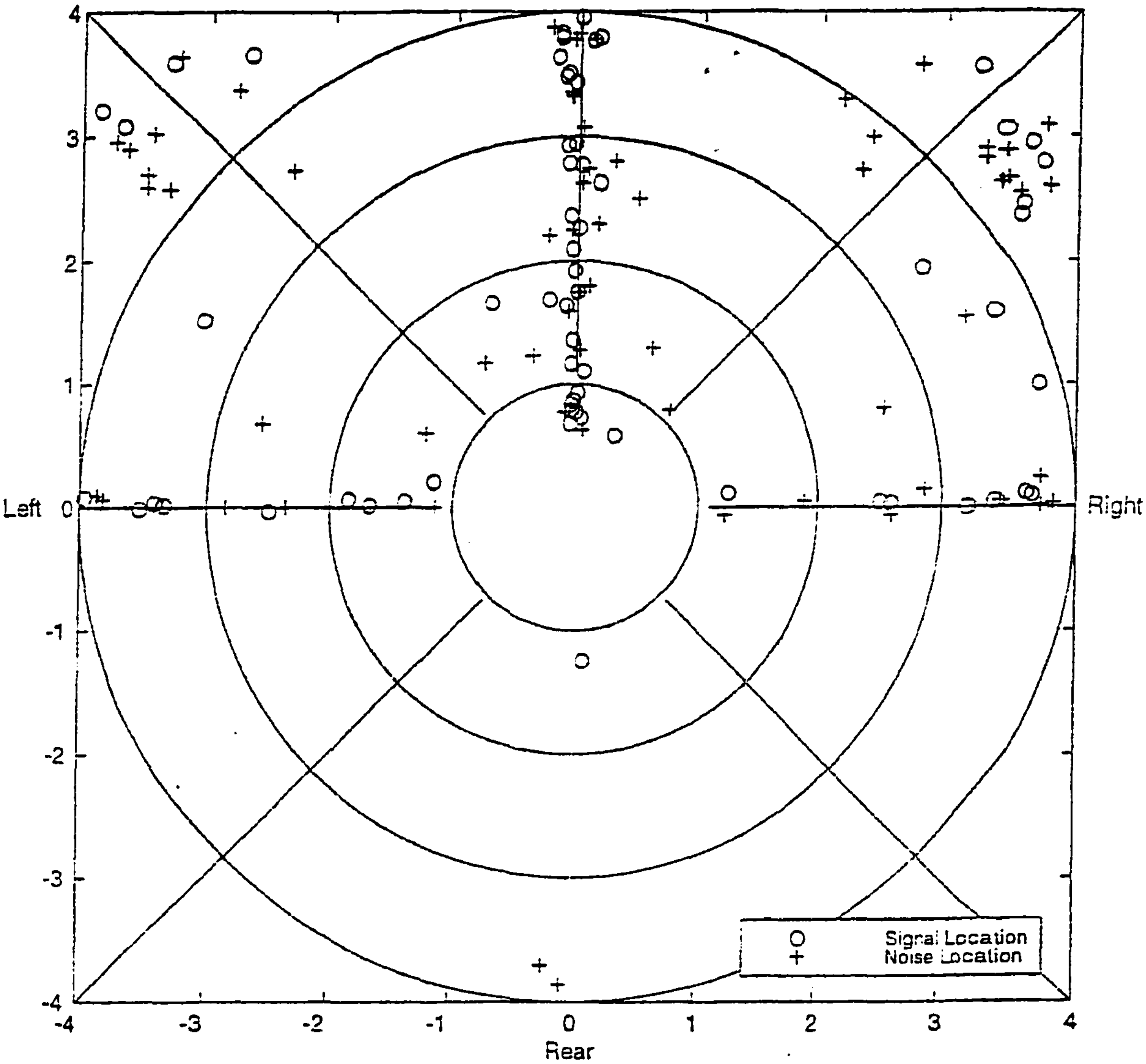
S3


Left
Right
Rear
4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Signal Location
Noise Location

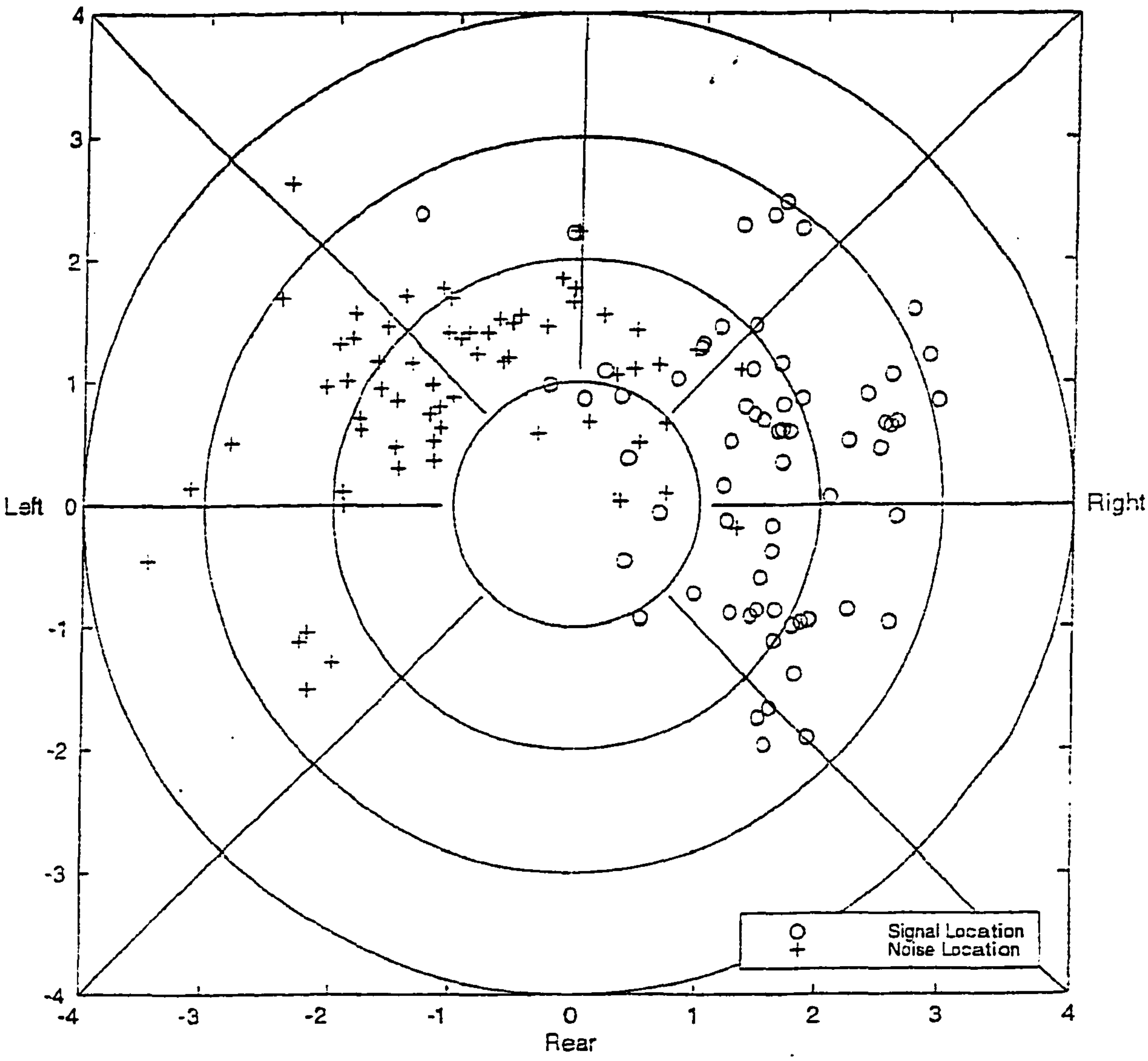
S4


4
3
2
1
Left 0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Rear
Right
Signal Location
Noise Location

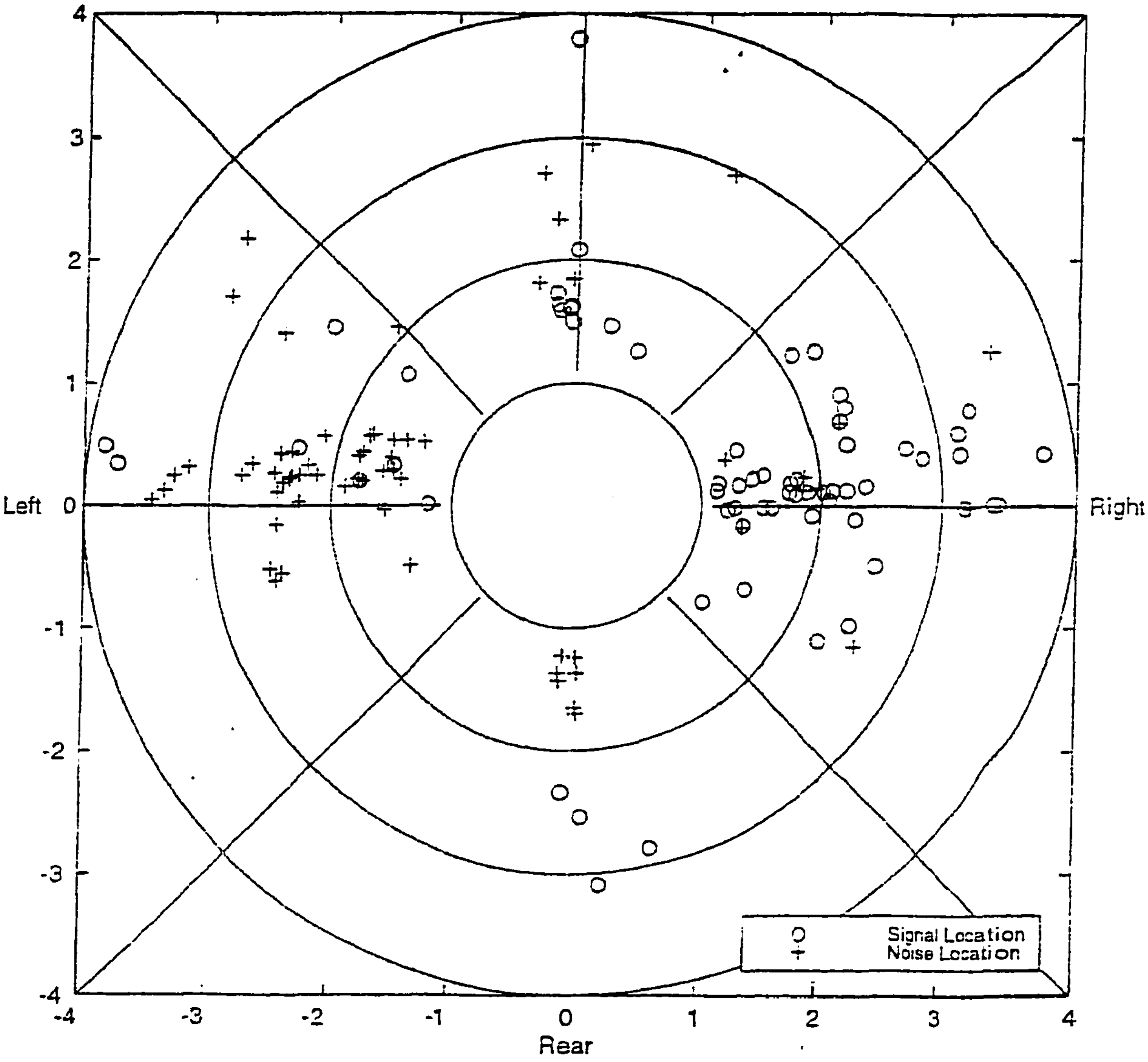
S5


4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Left
Right
Rear
Signal Location
Noise Location

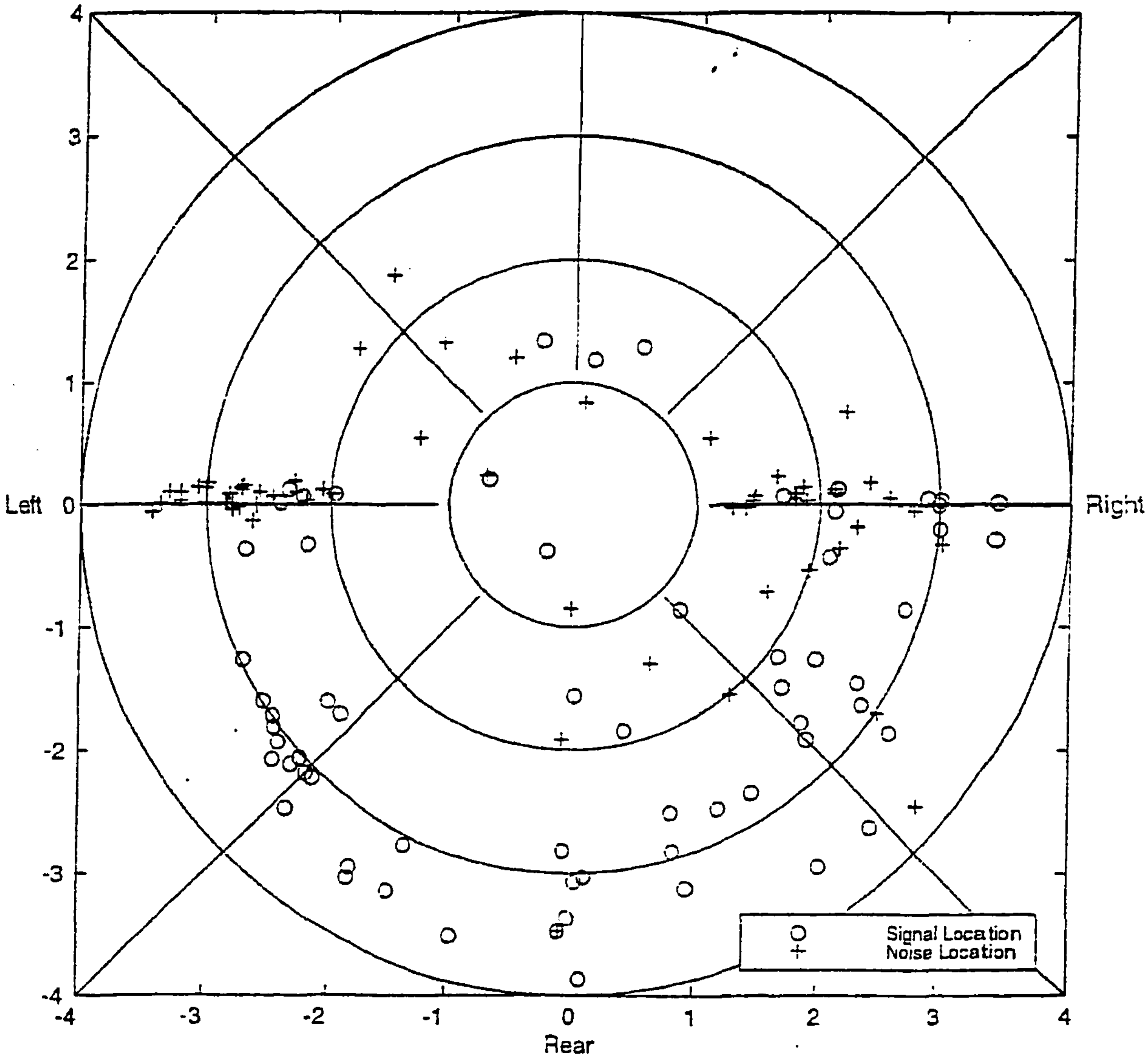
S6


4
3
2
1
0
-1
-2
-3
-4
Left
Right
-4
-3
-2
-1
0
1
2
3
4
Rear
Signal Location
Noise Location

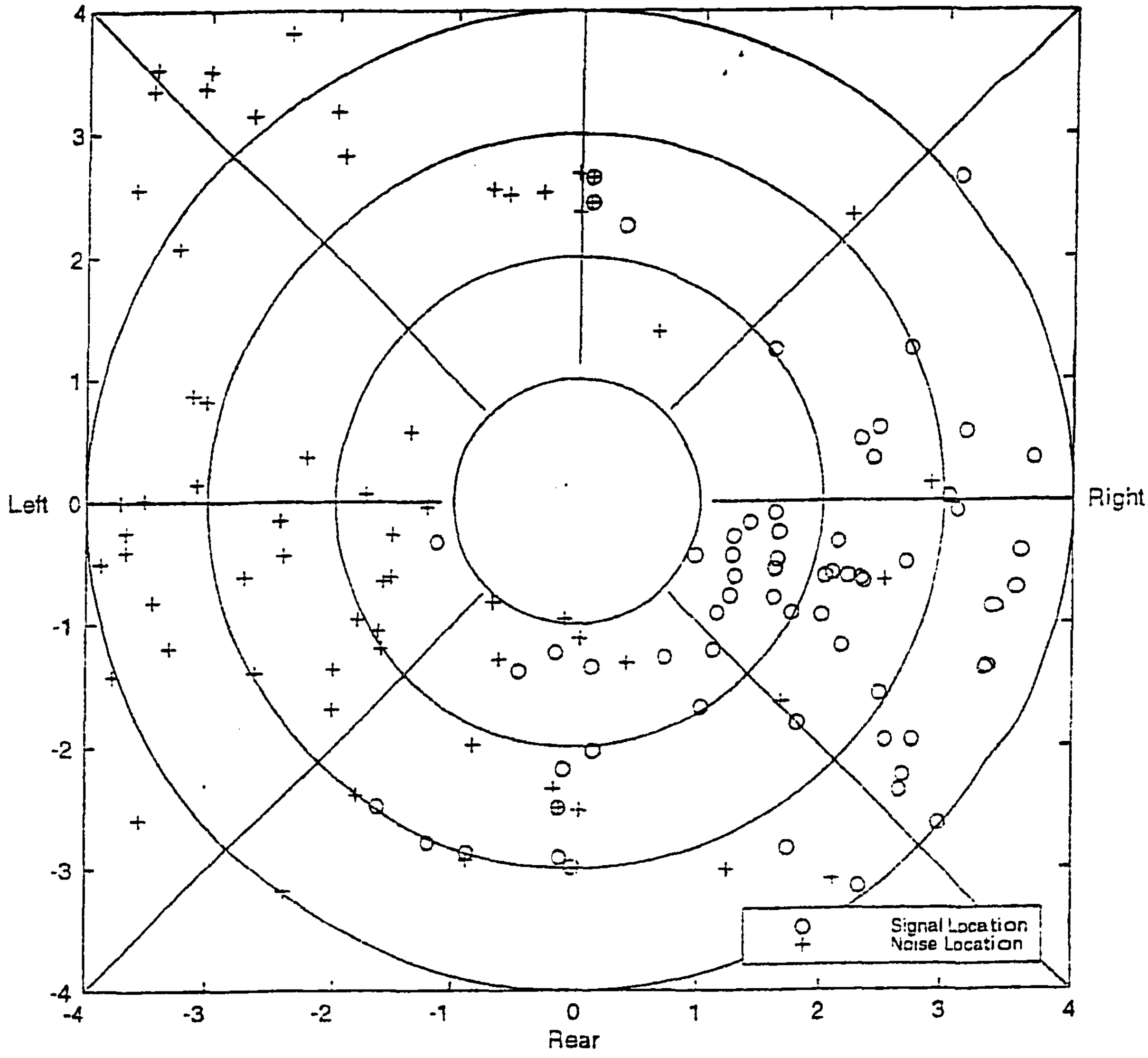
S7


4
3
2
1
Left 0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Rear
Right
Signal Location
Noise Location

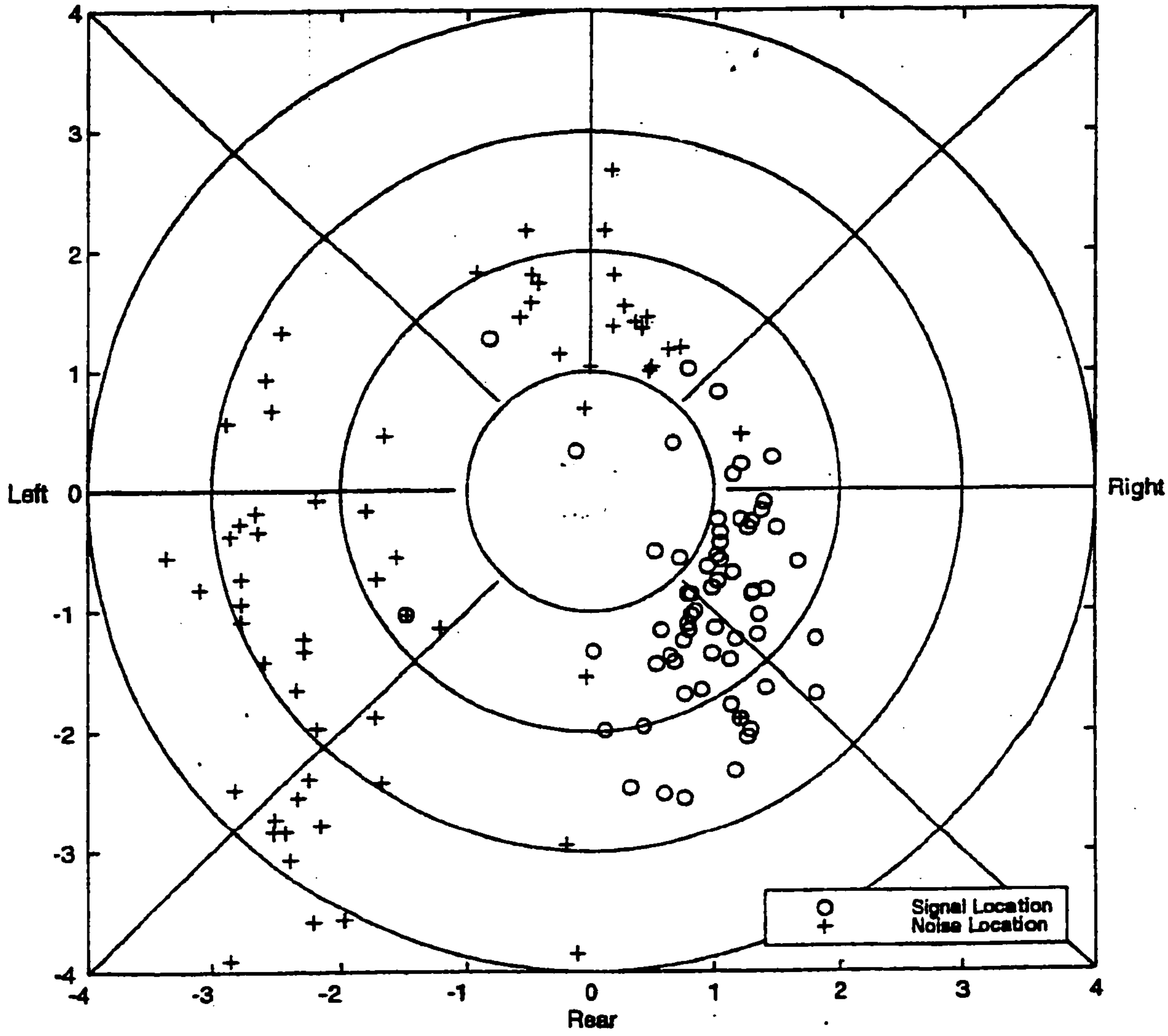
S8


Left
Right
Rear
4
3
2
1
0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Signal Location
Noise Location

S9


4
3
2
1
Left 0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Rear
Right
Signal Location
Noise Location

S10


4
3
2
1
Left 0
-1
-2
-3
-4
-4
-3
-2
-1
0
1
2
3
4
Rear
Right
Signal Location
Noise Location

APPENDIX B

SUBJECT INSTRUCTIONS AND CONSENT FORM

SUBJECT INSTRUCTIONS

SPATIAL LOCATION

You will be asked to make three judgments on each stimulus. You may listen to each stimulus as many times as you wish BEFORE you begin to make your three judgments. Press the REPEAT SENTENCE button to do this. The first two judgments you will be asked to do are:

1. **Identify the perceived location of the speech**
2. **Identify the perceived location of the noise**

Each time you will identify the speech and the noise by clicking the mouse button. An 'O' will appear on the screen to represent the speech. An '+' will appear on the screen to represent the noise.

YOU MUST IDENTIFY THE SPEECH FIRST AND THEN THE NOISE FOR EACH STIMULUS

THE THIRD JUDGMENT WILL BE TO RATE DIFFUSE OF THE SPEECH USING THE SCALE BELOW

Very Concentrated	Concentrated	Diffuse	Very Diffuse

SUBJECT INSTRUCTIONS

LISTENING EFFORT

During the test you will be listening to speech along with background noise. You will listen to a man stating one sentence (#1) and then stating another sentence (#2). We would like you to tell us which of the two samples (#1 or #2) takes less effort in order to understand as much of the speech as you possibly can.

Look at the monitor. Use the mouse to move the arrow to the center of the box marked No. 1 and hit the left mouse button. This will turn on sample #1. The box will turn from green to red to indicate that you are listening to sample #1. When you are ready to listen to the second sample, use the mouse to move the arrow to the center of the box marked No. 2 and hit the left mouse button. This will turn on sample #2. Box No. 2 will now change from green to red, indicating that you are listening to sample #2.

You may listen to each sample as long as you would like, and as many times as you would like.

Once you have decided which sample takes more effort, use the mouse to move the arrow to the center box labeled "choose" and hit the left mouse button. Then use the mouse to move the arrow to the center of either box No. 1 or Box No. 2 and hit the left mouse button.

Once you have made this decision you will go on to a new pair of samples.

Research Consent Form

I am being asked if I wish to participate in a research study. My decision to take part in this study is voluntary. If I decide not to participate or if I choose to withdraw after beginning the study, I will not lose any benefits associated with my medical care. I am encouraged to ask questions before deciding whether I wish to participate and at any time during the course of the project. I will be told of any new findings that may influence my decision to continue to participate in this research project.

Title of Protocol

EVALUATION OF A BINAURAL LISTENING SYSTEM USING A MONOPHASIC INPUT

Investigator(s)

MARYROSE MCINERNEY, MA, M.Phil
HARRY LEVITT, Ph.D.

Expected Duration of the Subject's Participation

This experiment is expected to take no more than three 1 to2 hour-long sessions to complete.

Purpose of the Project

An advantage of binaural hearing is that different sources of sound can be spatially separated however the same benefit is not shown with a single source of sounds. The purpose of this experiment is to evaluate if a simulated binaural system with input from a single source can produce separation of speech and noise. The separation of signals, clarity of the speech, ease of listening and speech intelligibility will be measured.

Description of Procedures

Before beginning the experiment, I will be asked to take a hearing test. I will receive my audiological test results. If I do not have normal hearing at that point I will be counseled and referred appropriately according to American Speech Language Hearing Association Guidelines. During the experiment I will listen to speech at a comfortable level under headphones and rate the clarity and the effort it took to listen to the speech. I will also identify the apparent location of the speech and noise. I will write down the words that I hear. Under no circumstances will I be required to listen to sounds that are known to be potentially dangerous. All testing will be done under the supervision of a licensed audiologist

Possible Discomforts or Risks

None

Possible Benefits of Participation

I will receive no individual benefit by participating in this research. Society in general and people with hearing loss in particular may benefit from the development of improved signal processing techniques. Subjects will be reimbursed for their participation at the rate of $8 per hour plus the cost of local transportation.

Alternative Treatment Available

Participation is entirely voluntary, and I may decline to participate at any time. I do not require any alternative treatment.

Costs

The cost of my health care and audiological services will not increase as a result of my participation in this research.

My identity and participation are confidential.

I hereby consent to participate.

Subject's Name ____

Signature ____________________________________ Date: ____________

If you have any questions about this research, you can call me at 201 996-4110 or my advisor, Dr Harry Levitt at 212-642-2395. If you have questions about your rights as a participant in this study, you can contact Hilry Fisher, Sponsored Research, Graduate School, City University of New York, 212-642-2059.

BIBLIOGRAPHY

Allen, J.B. (1977). "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *Journal of the Acoustical Society of America, 62, 912-915.*

Balchandran, R. and Levitt, H. (1997). "Sine wave modeling of speech for normal hearing and hearing-impaired listeners," *presented at the National Institutes of Health, IInd Conference on hearing and Research and Development,* Washington D.C..

Bench, J., Bamford, J. (1979). *Speech-hearing tests and the spoken language of hearing-impaired children.* New York: Academic Press.

Bilger, R. (1984). *Manual for the clinical use of the revised SPIN test.* Champaign-Urbana, University of Illinois.

Bodden, M. (1993). "Modeling human sound source localization and the cocktail-party-effect," *Acta Acustica, 1, 43-55.*

Broadbent, D. and Ladefoged, P. (1957). "On the fusion of sounds reaching different sense organs," *Journal of the Acoustical Society of America, 29, 708-710.*

Boll, S. (1979) "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans Acoustic Speech, 27, 113-20.*

Carhart, R., Tillman, T., and Johnson, K. (1967). "Release of masking for speech through interaural time delay," Journal *of Acoustical Society of America, 42, 124-138.*

CHABA (1989). *Removal of noise from noise-degraded speech signal.* (report of the) Committee on Hearing, Bioacoustics and Biomechanics, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC: National Academy Press. 1989.

Cherry, C., (1953). "Some experiments on the recognition of speech, with one and two ears," *Journal of the Acoustical Society of America, 25, 975-979.*

Cherry, C. and Taylor, W., (1954). "Some further experiments upon recognition of speech, with one and with two ears," *Journal of Acoustical Society America, 26, 554-559.*

Corbett, C., (1986). "Filtering competing messages to enhance mutual intelligibility," *MS Thesis, MIT, Cambridge, MA*

Dirks, D., and Wilson, R., (1969). "The effect of spatially separated sound sources on speech intelligibility," *Journal Speech Hearing Research, 12, 5-38.*

Durlach N. (1963). "Equalization and cancellation theory of binaural masking level differences," *Journal of the Acoustical Society of America, 35, 1206-1218.*

Durlach N. (1972). Binaural signal detection: equalization and cancellation theory. In J. Tobias (Ed.) *Foundations of Modern Auditory Theory* (pp.371-461). New York: Academic Press.

Eisenberg, L., Levitt, H., (1991). "Paired comparison judgments for hearing aid selection in children," *Ear and Hearing, 12, 6, 417-430.*

Eisenberg, L., Dirks, D., Bell, T., (1995). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing" *Journal of Speech and Hearing research, 38, 222-233.*

Eisenberg, L., Dirks, D., (1995). "Reliability and sensitivity of paired comparisons and category rating in children," *Journal of Speech and Hearing Research, 38, 1157-1167.*

Eisenberg, L., Dirks, D., (1997). "Subjective Judgments of Speech clarity measured by paired comparisons and category rating," *Ear and Hearing, 18, 4, 294-305.*

Eisenberg, L., Dirks, D., Takayanagi, S., and Martinez, A. (1998). "Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions," *Journal of Speech and Hearing Research, 38, 327-339.*

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

Feuerstein, J. (1992). "Monaural versus binaural hearing: ease of listening, word recognition, and attentional effort," *Ear and Hearing, 13, 2, 80-6.*

Gabrielson, A. and Sjogren, H. (1979). 'Perceived sound quality of hearing aids," *Scandinavian Audiology,8, 159-169.*

Gagne, J. (1988). "Excess masking among listeners with a sensorineural hearing loss," *Journal of Acoustical Society of America, 83, 2311-21.*

Gebhardt, C. and Goldstein, D. (1972). "Frequency discrimination and the M.L.D.," *Journal of Acoustical Society of America, 51, 1228-1232.*

Geller, D., Margolis, R. (1984). "Magnitude estimation of loudness I: application to hearing aid selection,"*Journal of Speech Hearing Research, 27, 20-27.*

Gilkey, R. and Anderson, T. (1997). (eds) *Binaural and Spatial Hearing in Real and Virtual Environments,* Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Graupe, D.,Grosspietsch, J., Basseas, S. (1987). A single-microphone based self adaptive filter of noise from speech and its performance evaluation," *Journal Rehabilitative Research Development, 24, 119-26.*

Guttman, N., van Bergeik, W. and David, E. (1960). "Monaural temporal masking investigated by binaural interaction," Journal *of Acoustical Society of America, 32, 1329-1336.*

Hafter, E. Schlauch, R. (1992). "Cognitive factors and selection of auditory listening bands. In A.L. Dancer, D. Henderson, R.J. Salvi, & P. Hameernik (eds)., *Noise-induced Hearing Loss.* St.Louis, MO.: Mosby-Year Book.

Haas, H. (1951). (English translation, 1972) "The influence of a single echo on the audibility of speech," *Journal Audio Engineer Society, 20,146-159.*

Hirsh, I. (1948). "The Influence of interaural phase on interaural summation and inhibition," Journal *of Acoustical Society of America, 20, 536-544.*

Hirsh, I. (1948). "The role of interaural phase in loudness," Journal *of Acoustical Society of America, 20, 761-766.*

Hirsh, I. (1950). "The relation between localization and intelligibility," Journal *of Acoustical Society of America, 22, 196-200.*

Hoffman, I. And Levitt, H. (1978). "A note on simultaneous and interleaved masking," *Journal of Communication Disorders, 11, 207-213.*

Humes, L., Christensen, L., Bess, Hedley-Williams, A. (1997). "A Comparison of the Benefit Provided by Well-Fit Linear Hearings and Instruments With Automatic Reductions of Low-Frequency Gain," *Journal of Speech and Hearing Research, 40, 666-685.*

Humes, L., Christensen, L., Bess, Hedley-Williams, A. (1999). "A Comparison of the Benefit Provided by a Linear and a Two-Channel Wide Dynamic Range Compression Hearing Aid," *Journal of Speech and Hearing Research, 42, 65-79.*

Jeffres, L., (1961). " Lateralization vs localization," *Journal of Acoustical Society of America, 33, 482-483.*

Jeffres, L., (1961). "Masking and interaural phase.II. 167 Cycles," *Journal of Acoustical Society of America, 34, 1124-1126.*

Jerger, S. and Jerger, J. (1984). *Pediatric Speech Intelligibility Test.* St Louis, Auditec of St Louis.

Kates, J. (1991). "A simplified representation of speech for the hearing impaired," *Journal of Acoustical Society of America, 89 pt2,1961.*

Kates, J. (1994). " Speech Enhancement Based on a Sinusoidal Model," *Journal of Speech and Hearing Research, 37, 449-464.*

Koenig, W. (1950). "Subjective effects in binaural hearing," *Journal of Acoustical Society of America, 22, 61-62.*

Kollmeir, B. and Peissing, J. (1990). "Speech intelligibility enhancement by interaural magnification," *Acta Otolaryngol. Suppl. 469, 215-223.*

Kollmeir, B., and Koch, R. (1994). "Speech enhancement based on physiological and psychoacoustic models of modulation perception and binaural interaction," *Journal of Acoustical Society of America, 95, 1593-1602.*

Kuhn, G (1977). Model for the interaural time differences in the azimuth plane. *Journal of Acoustical Society of America, 62,1,157-167.*

Kuhn, G (1987). "Physical acoustics and measurements pertaining directional hearing. 3-25. In Yost, WA & Gourevitch, G, *Directional Hearing,* SpringerVerlag, NewYork.

Libby, E. (1980). (ed) *Binaural Hearing Amplification, Volume I and II,* Zenetron, Inc, Chicago, Ill.

Lim, J. and Oppenheim, A. (1979). "Enhancement and bandwidth compression of noisy speech," *Proceedings IEEE, 67, 1586-604.*

Kuk, F. and Lau, C. (1995). "The Application of Binomial Probability Theory to Paired Comparison Judgments," *American Journal of Audiology, 4, 37- 42.*

Lehnert, H. and Blauert, J. (1992). "Principles of binaural room simulation," *Journal of Applied Acoustics, 36,259-291.*

Levitt, H., Bakke, M., Kates, J., Neuman, A., Schwander, T., Weiss, M. (1993). Ignal processing for hearing impairment," *Scandinavian Audiology,Suppl39, 7-19.*

Levitt, H., Neuman, A. and Sullivan, .J. (1990). "Studies with digital hearing aids," *Acta Oto-laryngology, Suppl. 469, 57-69.*

Levitt, H. and Voroba, B. (1989). "Binaural Hearing", in *Introductory Hearing Science, ed.* Gerber, J., Saunders, W.B. Publishing House.

Levitt, H. and Rabiner, L. (1967). "Binaural release from masking for speech and gain intelligibility," *Journal of Acoustical Society of America, 42, 601-608.*

Levitt, H. and Rabiner, L. (1967). "Predicting binaural gain in intelligibility and release from masking for speech," *Journal of Acoustical Society of America, 42, 820-829.*

Licklider, J. (1948). "The influence of interaural phase relations upon masking of speech by white noise," *Journal of Acoustical Society of America, 20, 150-159.*

MacAlister, P.V. (1990). "The effects of hearing aids on speech discrimination in noise by normal-hearing listeners." *Journal of Rehabilitation Research, 27 (1), 33-42.*

Markides, A. (1977). *Binaural Hearing Aids,* Academic Press, New York, NY.

McAulay, R. and Quatieri T. (1986). "Speech analysis/synthesis based on a sinusoidal representation," *IEEE transactions on Acoustics, Speech and Signal Processing, 28, 137-145.*

Mills, A. (1958). "On the minimum audible angle. *Journal of the Acoustical Society of America, 30, 4, 237-246.*

Middlebrooks, J (1992). "Narrow-band sound localization related to external ear acoustic," *Journal of Acoustical Society of America, 92,5, 2607-2624.*

Musicant, A., Butler, R. (1984). "The influence of pinnae-based spectral cues on sound localization," *Journal of the Acoustical Society of America*,82,5, 1631-1636.

Neuman, A., Eisenberg, L. (1991). "Evaluation of a dereverberation technique," *Journal of Communication Disorders, 24, 211-21.*

Neuman, A., Levitt, H., Mills,R., and Schwander, T. (1987). "An evaluation of three adaptive hearing aid stragegies. *Journal of the Acoustical Society of America, 82, 1967-1976.*

Neuman, A, Bakke, M., Mackersie, C., Hellman, S., Levitt, H. (1995). "Effects of Release time in compression hearing aids: paired comparison judgments of quality," *Journal of Acoustical Society of America, 98, 3182-3187.*

Nilsson, M., Soli, S., Sullivan, J. (1994). "Development of the Hearing in Noise test for the Measurement of Speech Reception Thresholds in Quiet and in Noise," *Journal of the Acoustical Society of America, 95, 2, 1085-1095.*

Preminger, J. and VanTasell, D. (1995). "Quantifying the Relation Between Speech Quality and Speech Intelligibility. *Journal of Speech and Hearing Research, 38, 714-726.*

Preminger, J. and VanTasell, D. (1995). "Measurement of Speech Quality as a Tool to Optimize the Fitting of a Hearing Aid. *Journal of Speech and Hearing Research, 38, 726-736.*

Purdy, S. and Pavlovic, C. (1992). "Reliability, sensitivity and validity of magnitude, estimation, category scaling and paired-comparison judgments of speech intelligibility by older listeners." *Audiology,31,254-271.*

Quatieri T. and McAulay, R. and (1991). "Peak-to-RMS reduction of speech based on a sinusoidal model," *IEEE transactions on Acoustics, Speech and Signal Processing, 39, 272-287.*

Rakerd, B, Hartmann, W (1985). "Localization of sound in rooms II: The effects of a single reflecting surface. *Journal of the Acoustical Society of America, 78, 2, 524-533.*

Robinson, D. and Jeffress, L. (1962). "Effect of varying the interaural noise correlation on the detectability of tonal signals," Journal *of Acoustical Society of America, 35, 1947-1952.*

Ruotolo, B. (1979). "Discrimination of symmetric time-intensity traded binaural stimuli," *Journal of the Acoustical Society of America, 66, 1733-1737.*

Sayers, B. (1964). "Acoustic-image lateralization judgments with binaural tones," *Journal of the Acoustical Society of America, 36, 923-926.*

Spieth, W., Curtis, J., and Webster, J. (1954). "Responding to one of two simultaneous messages," *Journal of Acoustical Society of America, 26, 391-396.*

Schweitzer, H., Terry, A., and Grim, M. (1996). "Three experimental measures of a digital beamforming signal processing algorithm," *Journal of American Academy of Audiology, 7, 4, 230-9.*

Shelton, B, Searle, C (1980). "The influence of vision on the absolute identification of sound-source position". *Perception and Psychophysics,* 28, 589-596.

Saunders, G. Levitt, H. (1991). "An automated true-false response time test for evaluating noise reduction systems (abstract). *ASHA, 33, 164.*

Teas, D. *(1962). "Lateralization of acoustic transients," Journal of the Acoustical Society of America, 34, 1460-1465.*

Thurstone, L. (1927). "Psychophysical analysis". *American Journal of Psychology, 38, 369-389.*

Tobias and Curtis (1959). "Binaural masking of noise by noise," *Journal of Acoustical Society of America, 31,127A.*

Watson and Mittler (1965). "Time-intensity equivalence in auditory lateralization:a graphical method," *Psychonometric Science, 2, 219-220.*

Weiss, M. and Aschkenasy, E. (1975). "Automatic detection and enhancement of speech signals," *Report No RADC-TR*-75-77. *Rome Air Devlopment Center, Grifiss Air Force Base, NY.*

Wiener, N. (1949). *Interpolation, extrapolation and smoothing of stationary time series.* J Wiley, New York.

Wightman, F. and Kistler, D. (1989). " Headphone simulation of free-field listening: I: Stimulus synthesis," *Journal of Acoustical Society of America, 85, 858-867.*

Wightman, F. and Kistler, D. (1989). " Headphone simulation of free-field listening: II: Psychophysical validation," *Journal of Acoustical Society of America, 85, 868-878.*

Wightman, F., (1992). "The dominant role of low-frequency interaural time differences in sound localization," *Journal of the Acoustical Society of America, 91,3, 1648-1661.*

Winer, B., Brown, D. Michels, K. (1991). *Statistcal principles in experimental design, McGraw-Hill, New York.*

Yost, W. (1981). "Lateral position of sinusoids presented with interaural intensive and temporal differences," *Journal of the Acoustical Society of America, 70,2, 397-409.*

Yost, W. (1996). "A simulated 'cocktail party' with up to three sound sources," *Perception and Psychophysics, 58, 1026-1036.*

Zwislocki, J, and Feldman, R. (1956). "Just noticeable differences in dichotic phase," *Journal of the Acoustical Society of America, 28, 860-864.*