Sound System Engineering & Optimization:
The effects of multiple arrivals on the intelligibility of reinforced speech

Timothy James Ryan

Sound Recording Area, Department of Music Research
Schulich School of Music
McGill University, Montreal
Initial Submission – February 2011
Final Submission – May 2011

A thesis submitted to McGill University in partial
fulfillment of the requirements of the degree of
Doctor of Philosophy

Abstract

The effects of multiple arrivals on the intelligibility of speech produced by live-sound reinforcement systems are examined. The intent is to determine if correlations exist between the manipulation of sound system optimization parameters and the subjective attribute speech intelligibility. Given the number, and wide range, of variables involved, this exploratory research project attempts to narrow the focus of further studies. Investigated variables are delay time between signals arriving from multiple elements of a loudspeaker array, array type and geometry and the two-way interactions of speech-to-noise ratio and array geometry with delay time.

Intelligibility scores were obtained through subjective evaluation of binaural recordings, reproduced via headphone, using the Modified Rhyme Test. These word-score results are compared with objective measurements of Speech Transmission Index (STI).

Results indicate that both variables, delay time and array geometry, have significant effects on intelligibility. Additionally, it is seen that all three of the possible two-way interactions have significant effects. Results further reveal that the STI measurement method overestimates the decrease in intelligibility due to short delay times between multiple arrivals.

Résumé

Les effets d'arrivées multiples sur l'intelligibilité de la parole produite sur systèmes de sonorisation sont examinés. Le but est de déterminer si des corrélations existent entre la manipulation des paramètres d'optimisation de systèmes de son et l'attribut subjectif de l'intelligibilité de la parole. Étant donné le nombre et la large gamme de variables impliquées, ce projet de recherche exploratoire cherche à mieux délimiter le champ d'études ultérieures. Les variables investiguées sont le temps de délai entre signaux en provenance de plusieurs éléments d'un système d'enceintes, le type d'agencement d'enceintes, la géométrie du système, ainsi que les interactions à deux voies entre le rapport parole-bruit et la géométrie du système avec le temps de délai.

Les scores d'intelligibilité ont été obtenus à travers l'évaluation subjective d'enregistrements binauraux reproduits sur casque d'écoute, en utilisant le test de rime modifiée (*Modified Rhyme Test* [*MRT*]). Ces résultats score-mot sont comparés avec des mesures objectives de l'index de transmission de la parole (*Speech Transmission Index*, [*STI*]).

Les résultats indiquent que les deux variables, temps de délai et géométrie, ont des effets significatifs sur l'intelligibilité. Par ailleurs, il a été constaté que les trois interactions à deux voies ont des effets significatifs. Des résultats révèlent par surcroît que la méthode de prise de mesure du *STI* surestime la décroissance de l'intelligibilité en raison des temps de délais courts entre arrivées multiples.

Dedication

To Erin and Loki.

and

For my parents and grandparents.

Preface

Considering the magnitude of this project, it was known from the onset that much help, advice and assistance would be required from many people – however it was not known just how many people and how much help would be needed. The author would like to take this opportunity to acknowledge, and extend his gratitude to the following people for their contributions to this project.

Many thanks go to:

Richard King for agreeing to join as project supervisor in the critical late stages. The author would like to personally thank Mr. King for his dedication to the successful completion of this dissertation, his collaboration in navigating the process and his willingness to do leg-work in the author's absence. **Dr. René Quesnel** for guiding the author through the initial phases of this research project. Personal thanks are extended to Dr. Quesnel for helping the author understand how to use his ears and for the many fascinating discussions during his technical ear training course. **Dr. Jonas Braasch** for helping the author to make the transition to audio research, and for his continued support throughout the doctoral journey. **Dr. Wieslaw Woszczyk**, who welcomed the author into the Sound Recording program at McGill, and for his invaluable insights, support and supervision in the early stages of the author's doctoral journey. **Dr. William Martens** for his help understanding the complicated fields of subjective evaluation and statistical design. **Dr. Laura Kretschmer**, the project's co-principal investigator at the University of Cincinnati, for her insights and for helping to get the project off the ground at UC. The author would also like to thank Dr. Kretschmer for letting a theatre student enroll in graduate-level Audiology classes. **Bob McCarthy** at Alignment & Design, for sharing his wisdom and experience, and for evaluating the potential for general interest in this study during the initial design phase. The author would also like to thank Mr. McCarthy for contributing images from his seminal tome on system optimization

to this dissertation. **Dr. Roger Schwenke** at Meyer Sound for providing the author with enough knowledge of the inner workings of their hardware, without divulging trade secrets, to effectively use it in this research. **Dr. Geoffrey Groocock**, at the College of Veterinary Medicine at Cornell University, for his contributions and assistance with statistical analysis using non-parametric testing methods. Also, for having been a dear friend for the better part of 20 years. **Dr. Terrell Finney, Rayburn Dobson** and **Patti Hall**, for permission to use, and assistance scheduling, the facilities of the College Conservatory of Music at the University of Cincinnati. **Stirling Shelton, Stevenson Miller** and **Richard Palmer**, at the College Conservatory of Music, for the use of rigging hardware and for their assistance flying loudspeakers. The **National Institute for Occupational Safety and Health** (NIOSH) for the loan of many pieces of equipment used in this research project, including the KEMAR, measurement microphones, preamplifiers, adaptors and calibrator. **Chucri A. (Chuck) Kardous**, at the Cincinnati division of NIOSH, for graciously offering his time, equipment and laboratory space. **Dr. Babette Verbsky**, at the Cincinnati division of NIOSH, for her assistance developing the author's body of literature on subjective testing. **Claudia Norman**, of the Institutional Review Board at the University of Cincinnati, for all of the meetings, for proofreading the research protocols and for answering so many questions about the process. **Lynda McNeil**, of the Research Ethics Board at McGill University, for her assistance understanding and compiling the submission for approval. **Sun Hee Kil** for her unwavering dedication and help during the stimulus recording sessions. **Martin Garneret** and **David Hunter**, for their logistical support of the pilot study. **Nik Tranby** & **Dr. Sungyoung Kim**, for their assistance with software programming. **Teri Waters** for the initial pass at proof reading this document, and for logistical support of the second half of phase 2 of the main study. **Velma Dawson** for the use of her basement, and for enduring the associated inconvenience. And last, but not least, thanks go to **all of the subjects** who participated in these studies - "Circle the thanks again."

Special thanks go to:

Jim and Betty Ryan, and **Barton and Mary Tomlinson**, for their support, encouragement, understanding and patience throughout 16 years of higher education. **Charles Hatcher**, at the College Conservatory of Music, for having the faith to entrust the education of his students to the author, both now and a decade ago. Also, thanks are due to Mr. Hatcher for the years of mentoring, friendship, patience, and discussions about pedagogical methodology and organizational structures. **Erin Waters** for being a guinea pig, beta tester, research assistant, editor, proof reader, sounding board and source of sanity. The author would be remiss if he did not profess his extreme gratitude for Ms. Waters' immutable support, and occasionally attenuated tolerance, of both the author and this project.

Table of Contents

List of Tables

List of Figures

1. Introduction

With few exceptions, sound has been an integral part of theatrical performance since the inception of the art form. The sound quality of music and the quality and intelligibility of speech that an audience perceives can significantly affect the ability of performers to communicate with their audience. Historically, theatrical and musical performance spaces have been designed in such a way that the architecture would provide sufficient natural acoustic amplification for a production [17]. In modern times, the push for greater audience capacity, the use of "multi-purpose" performance spaces and changes in performance type and orchestration have each necessitated the use of electronic amplification for the reinforcement of voice and music [103, 108].

Sound reinforcement systems containing microphones and loudspeakers are used to capture, amplify and distribute sound in an attempt to provide the desired sound quality, sound level and speech intelligibility to an entire audience. Several physical factors can affect speech intelligibility: Temporal distortion (smearing or loss of articulation), signal-to-noise ratio (the level of speech vs. the level of background noise/music) and frequency response irregularities (tonal coloration) [110]. One of the aims of sound reinforcement is to increase speech intelligibility by increasing its signal-to-noise ratio while adding minimal tonal coloration and temporal distortion.

In live theatrical performances, the preservation of the intelligibility of reinforced speech presents sound engineers with many challenges. Currently, several factors that affect speech intelligibility have been explored (e.g. [77, 107, 117, 143]). However, limitations of current measurement methods have hindered progress toward the quantification of certain effects of the sound reinforcement system itself on the intelligibility of the speech it is reinforcing [113, 114].

A single loudspeaker is often insufficient to meet the coverage and/or sound level requirements of a production [47, 124]. Thus, multiple loudspeakers, oriented into arrays, are often employed [56, 61, 174]. Whether grouped tightly

together or located some distance apart, there will be some difference in distance between each loudspeaker and an audience member [1, 124]. Differences in distance translate to differences in the travel time of sound, resulting in multiple arrivals: Time-delayed copies of the reinforced signal arriving and summing at a listener. The summation of multiple time-delayed sound signals results in a pattern of undesirable frequency response irregularities known as comb-filters [42, 46].

The negative effects of comb-filtering can be combated. Time-delays can be electronically added to speaker signals, in an attempt to align, or intentionally misalign, the timing of the arrivals of multiple loudspeaker broadcasts at the location of an individual listener in the audience [11, 94, 124]. As there are many listeners in an audience, it is physically impossible to achieve ideal loudspeaker time-alignment for every seat in an audience. Sound engineers are forced to select points within the audience where the signals from multiple loudspeakers will be aligned in time.

There are several differing theories regarding how to decide alignment locations [4, 11, 28, 42, 53, 124, 131] but to date, none have addressed the effects on intelligibility for the surrounding locations which are not time-aligned [115]. Research (e.g. [130, 170]) has shown that non-aligned multiple arrivals can negatively affect intelligibility. This same research has also shown that alignment through the use of electronic delay compensation can ameliorate the problem at specific audience locations. However, most studies used electronic intelligibility estimation methods and were limited in the number and types of loudspeaker arrays studied. Through the use of subjective evaluation and objective measurement methods, this series of studies will expand upon the body of literature, delineating the specific effects of delay time and two specific array geometries on the intelligibility of reproduced speech.

1.1 Motivation

In recent years, work has been done to examine the physical and electrical aspects of sound system design and optimization [28, 42, 124]. It is known that

physical factors such as comb-filtering have a negative impact on subjective factors such as tone quality. The problem, however, of identifying and quantifying the correlation between sound system resultant comb-filtering and the subjective factor of speech intelligibility has not been sufficiently addressed to date [110]. As such, sound engineers are forced to make sound system optimization decisions regarding multiple-loudspeaker time alignment based on an incomplete understanding of the effects on the audience's overall subjective experience. Through analysis of the impact of comb-filtering and loudspeaker time alignment, this research project presents a novel approach towards investigating, measuring and hopefully improving the intelligibility of reinforced speech.

The resulting data obtained in this research will provide sound engineers with a more complete understanding of the relationship between multiple-loudspeaker time alignment and speech intelligibility. Data from subjective testing will also be compared with results from electronic intelligibility estimation methods to determine whether these methods are in fact sensitive enough to account for the effects of comb-filtering.

1.2 Project Overview

While there are several physical measurement systems, such as Speech Transmission Index and Articulation Index, that are available to estimate the level of speech intelligibility produced by sound systems and other transmission methods [41, 51, 58, 104], each system contains inherent errors and no system is currently believed sensitive enough to accurately predict the effects of comb-filtering [110, 117]. The most accurate method for assessing speech intelligibility remains the administration of listening tests to human subjects [110].

The subjective testing method used for this project will be based on the American National Standards Institute (ANSI) *Method for Measuring the Intelligibility of Speech over Communication Systems* [5, 75]. The stimulus set used will be the Modified Rhyme Test (MRT). Intelligibility scores are determined based on the number of correct responses provided by the subject. As

the intelligibility score results will be of a quantitative data type, it is possible to use parametric statistical tests, such as analysis of variance (ANOVA), as well as non-parametric and log-linear analysis methods, to determine the significance of each of the experimental variables (described below), as well as the significance of possible interactions between experimental variables [14, 160, 180].

This project, and the studies contained within, represents a first attempt at exploring the effects of multiple arrivals of speech signals on intelligibility. The topic is quite complex and, given the sheer number of variables involved, prohibitively complicated to attempt to fully address with a single research project. Thus, while the results of this project will answer a number of questions, many more will remain.

In the interest of preserving the ecological validity of results, this research project will take an organic and exploratory approach – employing several commonly used sound system configurations to identify and study research questions that arise from real-world reinforcement scenarios. It is intended that the results from this project will guide and inform further research in the extended effort to correlate the physical parameters involved with sound system optimization with the subjective experience of an audience.

The results of these and further studies will be applicable to a wide variety of live sound applications both with and without music, ranging from dramatic and musical theatre productions to speeches, lectures, music concerts and multimedia performances.

1.3 Research Questions & Variables

When examining multiple arrivals of coherent audio signals, two properties of interest are the delay time between arrivals and the relative levels of the two arriving signals. The delay time determines the pattern of the resultant comb filter and the difference between relative levels determines the depth of the notches of the comb filter [1, 124]. As such, delay time and level difference would be included as independent variables in this research project.

For the cases of musical theatre and music concerts, sound engineers are faced with the challenge of maintaining an aesthetic balance of level between voice and music while still maintaining intelligibility. The presence of music amounts to an increase in the level of background noise, thus a decrease in the signal-to-noise ratio of the speech signal [126, 162]. This research project will therefore also include background noise level (signal-to-noise ratio) as an experimental variable, to investigate interactions between the level of non-speech sounds and speech intelligibility. Though the effect of signal-to-noise ratio (SNR) itself is known, the potential for SNR to obscure the effects of other variables is not.

While the use of music as the non-speech signal would be the most ecologically valid, previous research has identified that temporal variations in characteristics such as spectrum are a likely cause of unwanted variability in test results [73]. In the interest of controlling variables, and limiting potential contextual clues and listener preference biases, pink-weighted pseudo-random noise, exhibiting average spectral and power characteristics similar to popular music, will be used in the place of music.

To obtain the stimuli for this study, the MRT word lists will be reproduced by a sound system set up in Corbett Auditorium at the University of Cincinnati. A head and torso simulator (manikin with microphones in the ear canals) will be used to capture and record the sound produced by the sound reinforcement system. Stimulus sets corresponding to different experimental treatments will be obtained by delivering and capturing the MRT word lists using different sound system alignment conditions. These different alignment conditions correspond to variations in the four independent variables to be investigated: Delay time between the arrivals from two loudspeakers, level offsets between the arrivals, the orientation and focus of the two loudspeakers (vertical point-destination vs. horizontal point-source array) and the signal-to-noise ratio.

Though the associated stimuli were recorded, it should be noted that the variable “level offset” was ultimately not used in any of the studies in this project. It is included here, and mentioned throughout this dissertation, as it was present in

the original project design and its incorporation was considered during the design process of each study. While it was not possible to evaluate the effects of level offset within the scope of this project, it is the author's intention that this variable will be used in future studies.

1.3.1 Project Summary

In chapters 3–5 the acquisition and preparation of the stimulus sound files will be detailed, as will the headphone suitability study.

Chapter 3 – Stimuli were recorded for use in subjective evaluation and measurements were made for comparison with the results from listening tests.

Chapter 4 – A study was conducted to determine which, if any, available headphones would be acceptable for use in listening tests.

Chapter 5 – Stimuli were prepared and compensatory equalization was applied to counter the effects of the recording apparatus.

In chapters 6–9 a series of studies are detailed. As is standard scientific practice, research questions are posed in the form of null and alternative hypotheses. Statistical tests are performed on acquired subject data in order to determine whether there is significant evidence to reject each of the null hypotheses in favor of its alternative. As the alternative hypotheses are manifest, their declaration has been omitted for brevity.

Chapter 6 – A pilot study was conducted with two purposes: 1) Narrow the range of variable treatments to be used in the main studies and 2) verify the proper function of the testing methods employed.

Tested Hypotheses:

Ho1: Delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

Ho2: Signal-to-noise ratio does not affect the intelligibility of speech reproduced by a sound system.

Ho3: Array geometry does not affect the intelligibility of speech reproduced by a sound system.

Chapter 7 – The first phase of the main study was conducted. The goals of this phase were to investigate a reduced set of variable treatments, to test 6 hypotheses and to make note of effects and relationships observed in an attempt to further narrow the number of variable treatments before proceeding to the second phase of the main study.

Tested Hypotheses:

Ho1: Delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

Ho2: Signal-to-noise ratio does not affect the intelligibility of speech reproduced by a sound system.

Ho3: Array geometry does not affect the intelligibility of speech reproduced by a sound system.

Ho4: (interaction) Signal-to-noise ratio does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

Ho5: (interaction) Signal-to-noise ratio does not affect how array geometry affects the intelligibility of speech reproduced by a sound system.

Ho6: (interaction) Array geometry does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

Chapter 8 – In phase 2 of the main study, a further reduced set of variable treatments was studied, testing two hypotheses.

Tested Hypotheses:

Ho1: Delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

Ho2: (interaction) Array geometry does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

Chapter 9 – The results from the three studies involving subjective evaluation were compared with the results of objective measurements of Speech Transmission Index made in the original sound field.

Chapter 10 – Conclusions and discussion regarding all three phases of the project and the comparison of objective and subjective data, including directions for future research.

2. Review of Literature

As the studies included in this research project are intended to incorporate and expand upon knowledge and previous research in the field, it would be beneficial first to review the knowledge and research relevant to this dissertation. The following chapter is a combination of background information and a comprehensive review of the relevant literature.

2.1 Speech Intelligibility

Speech is one of the most commonly employed methods of communicating ideas. The full chain in the speech communication process spans from the formation of a thought in the mind of the talker, through the creation and transmission of sounds through a medium, to the detection and interpretation of the sounds by a listener [119]. At the tail end of the communication chain, a listener must successfully receive, deconstruct and interpret words in order to derive the correct meanings [119].

The basic element of speech is the phoneme, the smallest unit which carries a differential meaning. The English language is constructed from a set of 40 phonemes, composed of 24 consonant sounds and 16 vowel sounds [119]. Phonemes are combined to create syllables and a word is composed of one or more syllables.

The meaning of a word is encoded in the combination of its phonemes, from a word's context and from a listener's knowledge of the word. If one or more of a word's phonemes are not correctly perceived by a listener, the decoding of the combination of phonemes may yield a different meaning, or no meaning at all [145].

The context, meaningful or otherwise, in which a word is received plays additional importance as the perception of a phoneme is affected by neighboring phonemes, both within the same word and in adjacent words [119]. In addition to affecting factors such as listener expectations, meaningful context provides the

only method for differentiating between identical sounding words such as homonyms and homophones.

As Gelfand [62] points out, there is a difference between hearing speech and being able to understand it. Speech intelligibility is a term that describes the extent of a listener's ability to understand the meanings encoded in speech. The degree to which a listener is able to correctly identify phonemes, and thus words, when presented out of meaningful context is a measure of speech intelligibility [147]. This type of subjective evaluation method has been applied to the evaluation of the effectiveness of communication systems (e.g. [9, 75, 99, 122]). Additionally, objective methods have been developed for the estimation and prediction of the intelligibility of speech produced by communication systems. These methods attempt to assess the degree to which factors such as reverberation, background noise and distortion degrade the intelligibility of transmitted speech [59, 107, 143, 167].

2.1.1 Factors That Affect Speech Intelligibility

Before discussing methods for assessing degradation in speech transmission systems it would be helpful to understand what factors have the potential to degrade speech. In essence, factors that affect the intelligibility of reinforced speech can be grouped into categories relating to room acoustics, background noise, the sound system and its operation, the transmission path, the articulation of the talker and the skill of the listener [110, 116, 143]. Many of the specific factors that affect speech intelligibility are well established, while others are still under investigation (e.g. in this research project). Additionally, there are differing points of view regarding the specific contributions of each factor, as well as combinations of, and interactions between, multiple factors to the reduction of speech intelligibility [164]. The following is a list of factors, including indication of the areas that contain uncertainty.

1) Reverberation

Assuming that a single-loudspeaker speech reproduction system is operating indoors, sound may travel from the loudspeaker to the listener via many

paths. Energy that travels directly from the loudspeaker to the listener, and thus has the shortest path and travel time, is termed the direct sound (see chapter 2.2.2 for an explanation of sound propagation). Energy that arrives at the listener via paths including a small number of reflections will reach the listener a short time after the direct sound, and will generally be lower in level than the direct sound due to energy absorption by the reflecting surfaces. This group of arrivals is termed the early reflections, or early sound. Because there is a relatively small number of possible transmission paths that contain only a few reflections, the number of early reflections is also small and thus arrivals tend to be spaced. Also, due to room geometry, the majority of early reflections tend to be incident on the listener from angles in front of the lateral plane [17].

As the number of reflections in a transmission path increases, the number of potential paths increases. Eventually, some time after the reception of the direct sound, a condition is reached in which the number of reflections arriving at the listener is so numerous, and the reflections are so tightly packed in time, that individual reflections can not be distinguished from the mass. The sound energy of these later-arriving reflections is termed the reverberant energy. Due to the increased number of possible transmission paths, reverberant energy is not limited to frontal incidence, but rather tends to be incident from all angles [17].[1] The total sound reaching a listener will thus include the components of the direct sound, early reflections and reverberant energy.

Reverberant sound pressure in an impulse response usually decreases with time due to absorption during the increased number of reflections. Eventually, as the number of reflections increases, the level of the arrivals will drop below the audible threshold. In essence, the time required for the level of sound energy in a room to decay to a point below the audible threshold is referred to as the reverberation time of the room [42]. Lingering reverberant energy has the potential to mask the direct sound produced by a loudspeaker, depending on the

[1] A special condition, a diffuse reverberant field, exists when reverberant energy arrives from all directions, with equal level vs. direction. For the purposes of discussion in this dissertation, a reverberant sound field is not assumed to be diffuse.

relative level of the reverberant energy compared to the level of the direct sound. Thus both the reverberation time and the ratio of direct-to-reverberant sound energy (D/R ratio) are factors that affect intelligibility [106, 107, 110, 116, 143].

2) Discrete Echoes

With notable exceptions [67], the direct sound is generally perceived as being the source of the sound. However the auditory system incorporates a portion of the early sound into the direct sound. This process, termed fusion (or subjective masking), is a result of a phenomenon called post-masking, wherein a signal has the ability to mask other signals that arrive shortly after the original [107, 183]. As such, the portion of the early sound that arrives soon enough after the direct sound that it can be effectively fused is indistinguishable from, and serves to fortify the level of, the direct sound. The portion of the early sound that does not arrive soon enough to be fused will be perceived as individual reflections (echoes) [107]. Perceivable echoes have a negative impact on intelligibility, as they are not only a distraction for listeners but, like reverberant energy, serve to decrease the D/R ratio [67, 107, 116].

The question arises regarding the temporal location of the cutoff point between the direct and reverberant energy for calculation of the D/R ratio. There is considerable debate and data in the literature with regard to the specific integration time used in the fusion process (e.g. [17, 41, 67, 107]). There is further debate as to whether fusion affects different perceptual attributes, such as loudness and intelligibility, in the same manner or to the same degree. It is suggested by some [41] that fusion plays different roles for echo perception and intelligibility, in that even undetectable echoes can affect intelligibility. At this time, it seems clear that there is still uncertainty as to the effect on intelligibility of echoes with short delay time, which is why the author has focused on the topic as one of the primary research questions in this dissertation.

Haas [67] references earlier investigations on echo disturbance in telecommunications. These studies provide widely varying conclusions, placing the cutoff delay time required for an echo, with level equal to the original signal, to cause noticeable deterioration in sound quality between 40 ms and 100 ms.

Haas' own work [67] toward determining what he termed the "critical delay difference", concludes that 50% of listeners can detect an audible disturbance above around 40 ms for equal-level echoes incident from the front, with higher delay times for echoes incident from the side, rear and above, ranging up to 60 ms. Lochner and Burger [107] and Mapp [114] report similar relationships between critical delay difference and angle of incidence of reflections. Haas also determined that the delay time required for disturbance increases for echoes with level lower than the original signal.

Davis and Davis [41] cite previous works that have determined the cutoff for what "constitutes a detrimental reflection" to be anywhere from 0 ms to 95 ms. Conclusions from their own experiments comparing objective and subjective assessment results suggest that with regard to intelligibility, as opposed to the effects of fusion on loudness, no integration time is used. Klepper [94] cites previous works that have determined the cutoff to be between 10 ms and 60 ms, with shorter times corresponding to discrete echoes vs. echoes in the presence of reverberant "fill". Janssen [87], on the other hand, used the single value of 50 ms for his integration time.

Results from the multi-decade series of studies carried out by Peutz [144] indicate that, compared to the magnitude of effects due to other factors, echoes have a relatively small effect on intelligibility. However Peutz states that his results show that decreases in intelligibility resulting from a horizontally arriving echo are noticeable at delay times as short as 15 ms, with increasing detrimental effect up to 45 ms, above which point the effect becomes constant.

As mentioned previously, there is question as to whether fusion plays the same role in terms of echo detection and intelligibility. Lochner and Burger [107] state that level gains due to fusion, for delays of 95 ms and under, will increase intelligibility. However, Lochner and Burger also state that total fusion is only achieved up to 30 ms, beyond which the contribution of the echo diminishes [105]. In the work of Davis and Davis [41], it is acknowledged that fusion does occur up to around 50 ms to 80 ms, but the conclusions of their study suggest that, while

some portion (10 ms–20 ms) of the fused sound does cause an increase in perceived loudness, it does not lead to gains in intelligibility.

It has been further suggested [38, 111] that the auditory system could be viewed to employ various fusion integration times, depending on the perceptual attribute in question. In citing the findings of others, Mapp suggests that a delay time of 3 ms be applied to transient effects, 35-60 ms for information (such as speech and music) and 125-200 ms for perceived loudness.

3) Background noise

Background noise, like reverberant energy, has the potential to mask speech signals resulting in decreased intelligibility [18, 20, 107]. In essence, reverberant energy can be viewed as a type of background noise, though the two factors are addressed separately. Background noise is a blanket term used to describe many possible distracters to speech reception, including electronic noise (noise, buzz and hum in a sound system), noise due to air movement (heating and ventilation systems), other speech signals and music.

The metric used to describe the level of background noise, and its effect on intelligibility, is the signal-to-noise ratio (SNR, in dB) – the ratio of the level of the speech signal to the level of background noise. Results from the work of Peutz [144] indicate that variance in intelligibility can be found for values of SNR ranging from −5 dB to 25 dB.

4) Spectrum

There are several aspects of spectrum that relate to speech intelligibility, and can generally be broken into three categories: Production, reception and transmission. Figure 2.1 shows what could be considered the average spectrum of speech [59], though it is known that variations exist between male and female spectra [110]. It is also known that the formants of vowel sounds are relatively low in frequency (approximately 200 Hz–800 Hz) compared to consonants which extend to higher frequencies (approximately 1 kHz–8 kHz) [111, 117]. Comparing these frequency ranges to figure 2.1, it can be seen that the amount of energy produced by vowels is greater than the amount produced by consonants.

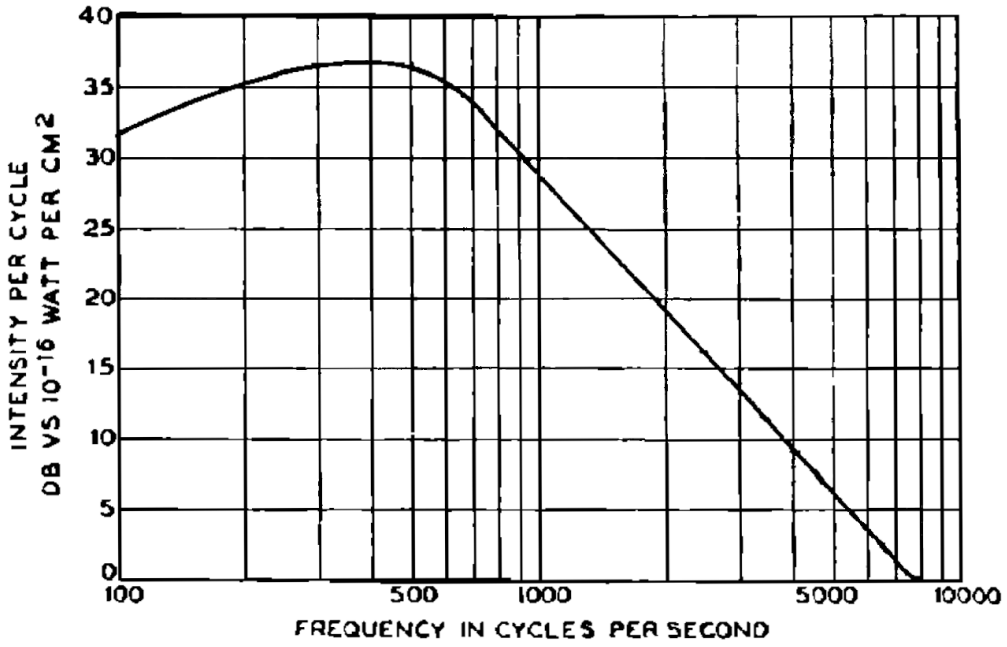


Figure 2.1 Idealized long-term average speech spectrum (Reprinted with permission from [59])

The effects of masking are frequency dependent [41, 117]. It is known from the study of critical bands [183] that individual frequency components of a signal can not only mask similar frequencies of equal level, but also higher frequencies of lower level. This is referred to as critical-band frequency masking, and the communication process is susceptible to its effects in several ways. Considering their lower frequency range and higher relative level at production, vowels have the potential to mask consonants, resulting in reduced perceived level in the 1 kHz to 8 kHz band.

As with production, the reception of speech is also frequency dependent. Figure 2.2 shows the contributions of different portions of the spectrum to the intelligibility of speech, indicating the importance of the 500 Hz–4 kHz region with emphasis on 2 kHz [110]. With the spectrum of speech production and spectral sensitivity of speech perception considered, it is clear that critical band frequency masking could have an effect on speech intelligibility.

With regard to transmission, Mapp [117] points out that loudspeaker frequency response and system equalization can also lead to critical-band

frequency masking. For example, if a speech reinforcement system were to have reduced high-frequency response, there would be further potential for reproduced consonant sounds to be masked by vowel sounds. As consonants carry the majority of speech information, it is therefore clear that sound system equalization is also a factor in speech intelligibility. Results of a study by Mapp [117] indicate that proper sound system equalization can increase speech intelligibility by more than 20%.

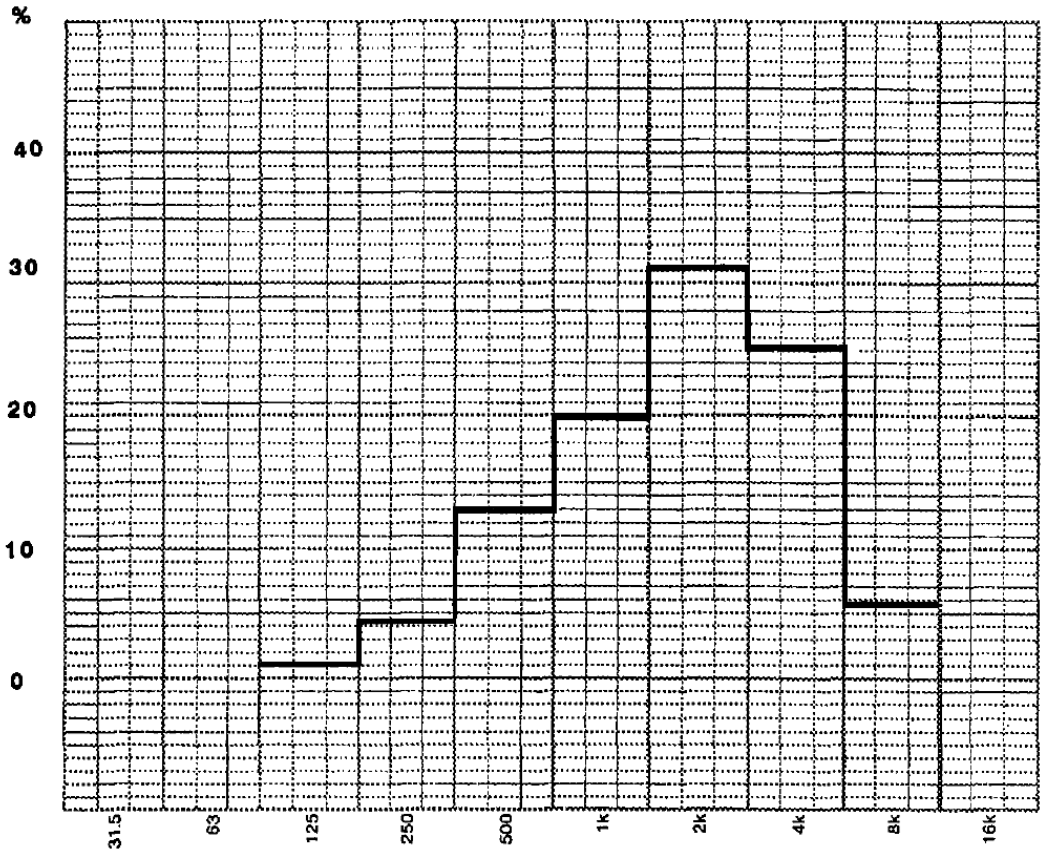


Figure 2.2 Octave band contributions to speech intelligibility (Reprinted with permission from [110])

One further consideration of transmission with respect to spectrum is the off-axis frequency response of directional loudspeakers. Shown in figure 2.3, the high-frequency response of directional loudspeakers is reduced at off-axis listening angles. Similar to the effects of equalization, a listener located off-axis from a loudspeaker may experience reduced intelligibility due to the rolling off of high-frequency components.

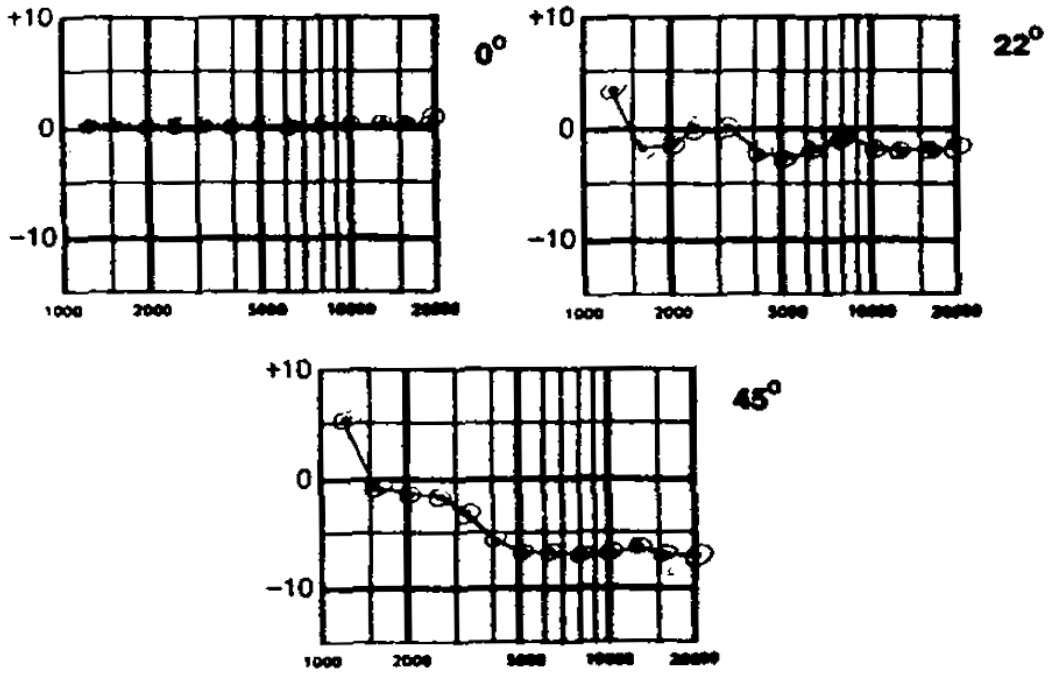


Figure 2.3 Magnitude vs. frequency response (1 kHz–20 kHz) of a constant directivity horn at several off-axis angles (Reprinted with permission from [57])

To date, knowledge of the effects of spectral anomalies on intelligibility is limited to wide-band effects on the order of 1-octave [37, 110], yet it is known that the critical bands of the human hearing mechanism are much smaller than one octave [183]. It is the intent of the current research project to investigate the effects of spectral phenomena of narrower bandwidth.

5) Loudspeaker Directivity

The term loudspeaker directivity, or Q, describes the radiation pattern of a loudspeaker.[2] Higher values of Q (as measured on-axis) indicate more directional loudspeaker radiation patterns [42]. The use of loudspeakers with high Q values allows for greater control of the distribution of sound in a room.

As D/R ratio is a factor that affects intelligibility, the ability to focus sound on audience listening areas, and away from all other areas in a room, will serve to reduce reverberant excitation and thus increase the D/R ratio [46].

[2] A rigorous definition of Q is beyond the scope of this text. The interested reader is directed to [142].

6) Distance between Listener and Loudspeaker

Similar to the effects of directivity, the distance between the listener and loudspeaker affects the D/R ratio. Results from studies by Peutz [143] indicate that intelligibility decreases as the listener-loudspeaker distance increases. The results also suggest that beyond a certain "critical" distance, intelligibility remains constant, as it is no longer a function of distance.

7) Rollover

As mentioned previously, inadequate speech level has a negative effect on intelligibility. Excessive level, however, can also reduce intelligibility. This effect, referred to as rollover, occurs at levels above that of normal conversation [59, 169]. Results from Lochner and Burger [107] show that intelligibility reaches a maximum in the range of 60 dB SPL. For levels above 60 dB, reduction in intelligibility is a factor of SNR. Their research indicates that the effect is most prominent for SNR's below 0 dB. With a speech level of 80 dB SPL, as was used in the current research project, rollover effects are negligible for SNR's above 0 dB.

8) Preservation of Modulation

Running speech can be viewed as a sequence of consonant and vowel sounds. Steeneken and Houtgast [167] point out that individual consonant and vowel sounds have different frequency spectra, and that an individual's ability to understand running speech is essentially a factor of the individual's ability to track spectral modulation (changes in spectrum over time).

Houtgast and Steeneken [77], in a review of their previous works, explain that running speech can be viewed as the amplitude modulation of different parts of the speech spectrum, with modulation rates ranging between 0.5 Hz and 16 Hz. This led to the concept of employing a modulation transfer function (abbreviated m(F) or MTF) to describe the effects of a speech communication system on the intelligibility of speech. As depicted in figure 2.4, properties of a communication system can affect a reduction in the depth of modulation of a speech signal, leading to a decrease in intelligibility.

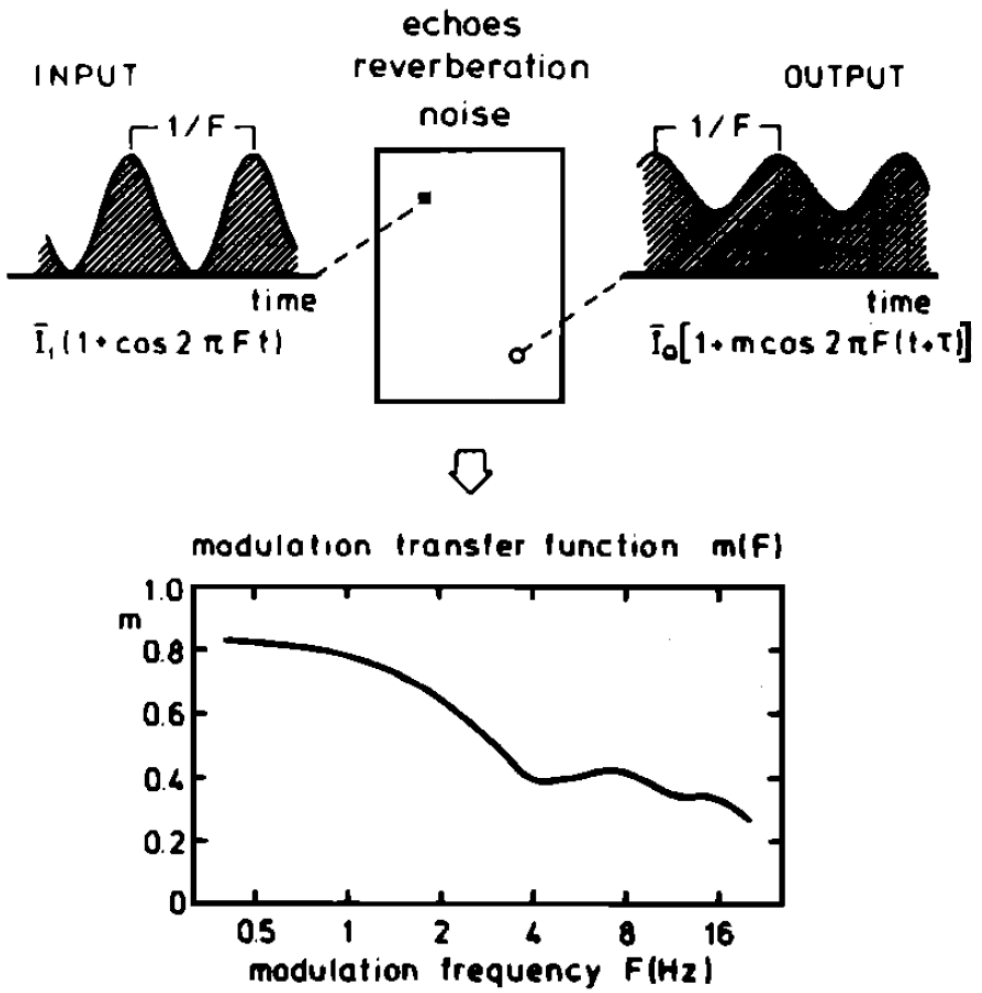


Figure 2.4 "*A speech transmission path in an enclosure is characterized by the modulation transfer function m(F), quantifying the degree of preservation of the original intensity modulations as a function of modulation frequency*." (Reprinted with permission from [77])

While this approach was key in the development of an objective intelligibility metric (see chapter 2.1.2), it also provides insight as to some of the factors that affect speech intelligibility. Reverberant energy, echoes and noise reduce modulation depth, and thus reduce intelligibility. It is noted however that modulation is a dynamic property and, as such, can be affected by the dynamic properties of a sound system, such as dynamic range, headroom/clipping, non-linear distortion and amplitude compression, as well as elements of digital systems such as data compression algorithms [110, 116].

9) Sound System Operation

The final factor mentioned in this text is perhaps the most unpredictable – the human factor. Sound system operation provides the potential for detriment to intelligibility from an amalgam of many of the factors previously mentioned [116]. The relative levels of speech versus music that a mix engineer creates amounts to a SNR. If the level of speech is not set high enough, insufficient intelligibility will be the result. Similarly, adjusting mix levels to produce excessive sound pressure levels can result in rollover.

Drastic or improper microphone and system equalization can have a marked impact on intelligibility. Poor gain structure can result in audible distortion and a decrease in SNR [149]. The use of effects processors can add echoes and artificial reverberation, and the overuse of dynamics processors can greatly reduce the modulation of speech [110].

2.1.2 Objective Methods for Estimating Speech Intelligibility

Over the past century, a variety of objective methods have been developed for the estimation of speech intelligibility over a communication system. These methods employ a combination of acoustical calculations and the electroacoustical measurement of many of the factors listed in the previous section. Through correlation with the empirical results from subjective testing, relationships have been determined that allow objective calculation-based and measurement-based methods to produce a result indicating intelligibility. In recent years, several of these methods have been included in measurement software as well as acoustical prediction programs [3, 49].

Over time the accuracy and precision of objective methods has increased, as more of the factors that affect intelligibility have been incorporated into the calculations and measurement protocols. However to date, no such method is believed to account for all of the factors that affect speech intelligibility [117]. Recent investigations have explored the integration of binaural mechanisms, higher frequency resolution and pre- and post-masking sensitivity [19, 37, 102, 155].

The following is a review of the more commonly used objective methods for estimating speech intelligibility produced by a speech transmission system.

1) C7, C35, C50 & C80

The family of "clarity" measurements is well recognized in the field of auditorium acoustics as an accurate predictor and measurement of sound quality [17, 114]. Each is a measure of the D/R ratio within a room, generally limited to the 1 kHz octave band. The difference between each of these measures is the temporal cutoff point between the direct and reverberant energy. C80, for example, compares the integration of the first 80 ms of early sound into the direct component to the integration of all sound after 80 ms. For C7 on the other hand, the integration time is set to 7 ms, resulting in a more transient-based measure. The existence of multiple clarity measures is a result of the ongoing debate regarding the integration time of the auditory system [38, 108, 110]. A similar metric was proposed by Lochner and Burger [107], citing 95 ms at the integration cutoff.

Reverberant energy is known to affect intelligibility and, as such, clarity measures are useful predictors. However as many of the other factors that affect intelligibility are not incorporated, clarity is a severely limited metric for prediction or measurement. This is particularly the case when sound reinforcement systems are involved, as the system has a large effect on intelligibility [110].

2) Articulation Index

Articulation Index (AI), originally developed by French and Steinberg [59], is a method based on the measurement of SNR in 20 frequency bands, producing a single-value result ranging between zero (unsatisfactory) and one (excellent). As the results of their studies clearly showed intelligibility to be frequency dependent – different bands providing different relative contribution to the overall subjective impression – frequency weighting was implemented in the calculation of AI.

As their work at Bell Telephone Laboratories was focused on intelligibility over single-channel telecommunication systems, factors such as

reverberation were not considered in the original design of the metric. Later adaptations by Kryter [97] and others incorporate the use of 1/3- and 1-octave frequency bands, and provide adjustment factors to account for reverberation. In the end, the treatment of reverberation has been found lacking [114] and the method does not allow for the separate analysis of intelligibility impairment due to the loss of articulation of vowels and consonants [143].

3) Articulation Loss of Consonants

The percent loss of the articulation of consonants ($\%AL_{cons}$), developed by Peutz [143] and later modified by Peutz [144] and Davis [40], is a predictive metric based primarily on acoustical parameters of a room and loudspeaker directivity. The metric focuses on consonants as Peutz found losses in consonants to be higher and a more accurate predictor than losses in vowels [143].

Though there are several variations of the $\%AL_{cons}$ equation, the general form includes reverberation time, room volume, listener-loudspeaker distance, loudspeaker Q and a factor to account for interference from other loudspeakers [41]. Originally developed as a calculated predictor of the intelligibility of speech, $\%AL_{cons}$ was later implemented as a direct measurement method in Techron TEF measurement devices [39]. Additionally, it is possible to determine $\%AL_{cons}$ via conversion from other measurements such as speech transmission index (by Farrel Becker, presented in [42]).

The predictive and measurement accuracy of the $\%AL_{cons}$ method has been shown to have high correlation to the results obtained through subjective assessment [41, 143], though correlation appears to decrease under conditions of diminished intelligibility [114]. The primary limitation of $\%AL_{cons}$ is that the formulae and measurements are limited to the 2 kHz 1-octave band, and is thus not sensitive to bandwidth, frequency response anomalies and critical band masking in other bands resulting from noise or equalization. Additionally, $\%AL_{cons}$ does not account for SNR, distinct echoes, distortion or non-linear effects, and it is difficult to correctly determine the effect of interference from other loudspeakers [110].

4) Speech Transmission Index

In 1972, Houtgast and Steeneken [76] published a description of the first intelligibility metric based on the modulation transfer function (MTF), which calculates the reduction in modulation of transmitted signals. The speech transmission index (STI) is derived from a matrix of 98 measurements using a speech-like stimulus signal. Measurement of 14 modulation frequencies (0.63 Hz–12.5 Hz, in 1/3-octave increments) are made in each of seven frequency bands (125 Hz–8 kHz, 1-octave wide each), replicating the low-frequency modulations of the human voice over its spectral range. The individual results within each octave band, representing the MTF of the band, are summed to create the transmission index for the band. The weighted average of the seven transmission indices is calculated for all bands. The resulting number is the STI, and ranges from zero (bad) to one (excellent).

Because its metric is based on the MTF, STI is sensitive to noise and reverberation, as well as non-linear and temporal distortions [168]. It is recognized that STI is a robust and highly accurate measurement, though it is believed that STI is incapable of accounting for certain sound system related distortions. Though STI is capable of sensitivity to critical band frequency masking, not all STI measurement devices include this feature and are thus not capable of accounting for certain equalization effects such as high-frequency reduction [117]. Also, as STI has a 1-octave frequency resolution, results are generally not sensitive to frequency response anomalies narrower than 1-octave in bandwidth [114].

In the 1980's, it was discovered that the MTF could be derived from an impulse response (IR) measurement. Mapp [109] points out that STI values derived in such a way show little variation between repeated measurements and suggests that, assuming that background noise remains constant, the averaging of multiple measurements should not be necessary. It should be noted however that IR derived STI does not actually use a speech-like modulated test stimulus and, as such, excitation of the sound system under test does not replicate natural conditions. If non-linear distortion (e.g. clipping) and compression are not

present in the sound system, differences in excitation should not affect test results [114]

Further developments with STI include variations in frequency weighting of the transmission indices to account for male and female speech spectra [112].

5) Rapid Speech Transmission Index

Full STI measurements are difficult to perform and the computational requirements were beyond the ability of the computers available at the time of its inception [110]. The Rapid Speech Transmission Index (RaSTI) was introduced and integrated into a hand-held measurement system, incorporating a nine measurement subset of the full STI. RaSTI measurements include four modulation frequencies in the 500 Hz octave band, and five frequencies in the 2 kHz band [77]. The RaSTI test signal is a speech-like modulated noise signal.

RaSTI suffers many of the drawbacks of STI, though it has long been recognized that the full STI measurement is more accurate than the smaller RaSTI measurement [114]. Also, as with $\%AL_{cons}$, the use of a reduced set of frequency bands leads to lack of sensitivity in many areas of the spectrum.

Recently, a version of the STI measurement for public address (STIPa) has been introduced, though it has yet to be standardized. Like RaSTI, STIPa uses a modulated speech-like test signal and a subset of the full number of STI measurements [114]. However unlike RaSTI, the calculation of STIPA uses measurements (14) in all seven of the STI frequency bands. Thus, though STIPa uses a reduced number of measurements, it should still be more sensitive to the effects of frequency anomalies than RaSTI. RaSTI and STIPa metrics can both be derived from IR measurements. However, STIPa results obtained in such a way are referred to as "STIPa equivalent" [49, 118].

2.1.2.1 Narrow Band Effects

It has been mentioned several times that frequency resolution, including number of frequency bands employed, is a drawback in objective intelligibility measurement systems. Figure 2.5 shows the frequency response of a system that Mapp [113] constructed to prove this point. RaSTI measurements were made on

the sound system and, despite the fact that this system would obviously greatly impair intelligibility, the RaSTI measurement returned near-perfect results (RaSTI = 0.98). The measurement scheme was unable to detect this impairment due to lack of resolution, as its measurements are based solely on information in the 500 Hz and 2 kHz bands. It is likewise conceivable that narrow-band frequency effects may elude other measurement methods with inadequate frequency response resolution (e.g. STI and $\%AL_{cons}$).

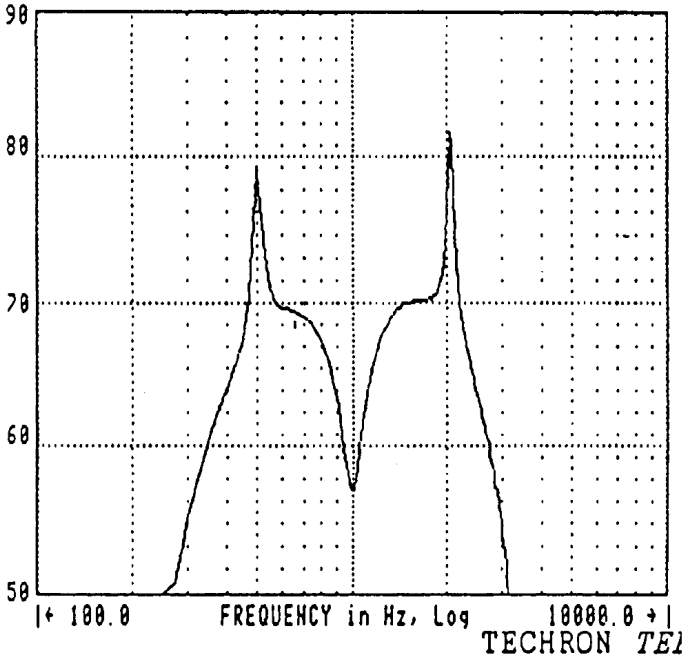


Figure 2.5 Frequency response of system used to confound RaSTI measurements (Reprinted with permission from [113])

An experiment reported by C. Davis [39] involved the measurement of a sound system under two conditions. The first condition resulted in a nominally flat frequency response, while the second produced a frequency response containing a deep notch near 2 kHz (due to loudspeaker misalignment). RaSTI and $\%AL_{cons}$ measurements indicated a clear reduction in intelligibility for the second condition. While the 2 kHz band has been shown to play a vital role in intelligibility, it is unclear whether the results from these measurements would correlate with subjective impression, given that more than 55% of the RaSTI data points and all of the data for the $\%AL_{cons}$ measurement lie in the 2 kHz band.

In addition to issues with resolution, measurement methods employing the MTF have an additional potential handicap. Figure 2.6 shows the effects of echoes on the magnitude of modulation vs. frequency response of a measured MTF. The result, in terms of frequency, is the production of a series of notches, as can be seen in the figure. (The effects of echoes and time-delayed signals are covered in depth in chapter 2.2.2) Such notches in MTF frequency response will lead to marked reductions in measured intelligibility when STI (or RaSTI) is calculated using this MTF. For example, an equal level echo, arriving 50 ms after the direct sound, would create notches in the measured MTF at 10 Hz in all 7 carrier frequency bands. Such notches would almost fully negate the results of the septet of measurements from the 10 Hz modulation, and would significantly impact the results in neighboring modulations.

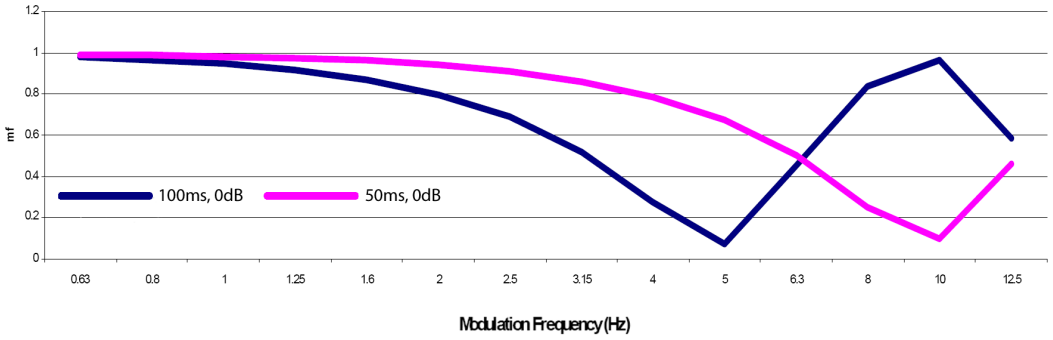


Figure 2.6 The effect of delayed signals on the magnitude vs. frequency response of a measured modulation transfer function (Reprinted with permission from [113]).

From the literature, it is reasonable to conclude that an echo arriving 50 ms after the direct sound could provide some detriment to intelligibility. As Peutz stated, though noticeable and measurable, when compared to other factors, the effects of echoes on intelligibility are relatively small [144]. However, it is again unclear as to whether measured effects would correlate with subjective impression, or if degradation shown in the results is merely a byproduct of the measurement method.

As the primary phenomenon studied in this dissertation, arrivals from multiple loudspeakers, involves narrow-band frequency effects and signal delays, it is clear that the use of objective measures for the assessment of intelligibility in

sound systems may lead to erroneous results. While this research may allow for further investigation of such error, the use of other measurement methods should be investigated.

2.1.3 Subjective Methods for Evaluating Speech Intelligibility

In contrast to objective methods for estimating speech intelligibility, subjective methods are also available to evaluate the intelligibility of speech. Such subjective methods involve the administration of listening tests to individual listeners or listening panels, and are usually conducted in the performance assessment of speech communication systems [14, 156]. With proper training and control of extraneous factors, it is possible for subjects to produce results that are consistent and repeatable within and between communication systems [179].

It is important to remember that, while subjective evaluation methods employ listeners rather than algorithms, the results of listening tests are still an estimation of the performance of a speech communication system. While it is widely accepted that subjective methods are the most accurate means for estimating speech intelligibility in a communication system [41, 114], it is suggested that the use of such evaluation methods could be complimented by the inclusion of other methods of system evaluation or measurement [156].

2.1.3.1 A Review of Various Testing Methods

A variety of listening tests exist, with objectives ranging from the identification of hearing disorders, to the diagnosis of speech and language skills in children, to the evaluation of communication systems. The core stimuli in these tests may contain nonsense syllables, single words or sentences.

An individual's ability to identify a word can be affected not only by the actual clarity of transmission of the word, but also by semantic and contextual clues [156]. Contrary to methods for developmental testing in children, successful isolation of the effects of a communication system on intelligibility requires that semantic and contextual clues are removed from testing stimuli.

Such clues could be present if sentences were to be used as the core stimuli for testing. Another issue is that of subject learning. Even with the use of sentences that are not highly predictable [50] or that are semantically anomalous [139], learning effects make it so that the same sentence cannot be used twice for the same listener, resulting in the need for very large stimulus sets [156].

The prohibitively large stimulus sets required, coupled with the ambiguity resulting from semantic and contextual clues, make sentence-based tests undesirable for use in communication system evaluation requiring repeated evaluations by each subject.

With regard to intelligibility, consonants carry more useful information and contain less energy than vowels, and are therefore more susceptible to degradation in a communication system [5]. Thus, intelligibility is primarily a byproduct of a listener's ability to differentiate between consonant-based phonemes [143]. As such, most single-word based listening tests that study intelligibility focus on the ability of subjects to distinguish between consonants [156].

Fairbanks [55] details the desire to create a speech intelligibility test based on phonemic differentiation, that would use words as the stimulus, test subjects' ability to identify said words and that would produce results relatable to the task of identifying everyday speech. The result of his work is the Fairbanks rhyme test (FRT): A testing method which employs monosyllabic English words, and tests subjects' ability to discriminate between consonantal and consonant-vowel transitional variations. The test contains 250 words which are grouped into 50 ensembles, each containing 5 rhyming words with identical stem spelling (e.g. -eel, -ook, -ale). The rhyming words differ only by the initial consonant.

Each test is constructed by selecting one word from each ensemble, resulting in a 50-word test list. Subjects are provided with a response sheet, and responses are indicated by writing the first letter of the word in the blank space preceding each of the 50 word stems. For each word stem, Fairbanks estimates that there are between 6 and 16 possible words in the English vocabulary from which a subject can choose.

The modified rhyme test (MRT), a six-alternative forced choice, closed-set adaptation of the FRT, was later developed [75]. The MRT contains 300 monosyllabic English words that take the form of consonant-vowel-consonant (CVC), consonant-vowel (CV) or vowel-consonant (VC). The majority of the words are of the CVC form. The words are grouped into 50 ensembles, each containing 6 rhyming words. Within each ensemble, differences exist between either the first consonants or final consonants, including absence of a consonant (in the case of CV and VC forms). The vowel, and associated vowel sound, remains constant within each ensemble.

Taking one word from each of the 50 ensembles, words are grouped into six 50-word lists. A test consists of the evaluation of one word list. Response sheets containing the 50 six-word ensembles are provided to subjects. Words are presented with or without a carrier sentence and subjects indicate responses by marking one word out of each six-word ensemble.

Initial tests conducted by House et al. [75] indicate that test sensitivity increases and the degree of error in results is reduced when subject population size and the number of word lists administered to each subject are increased. In their analysis, it is concluded that increases beyond a population of 12 or the evaluation of more than three word lists would not significantly improve test reliability.

Three testing methods are detailed in the American National Standards Institute's recommendation *Method for Measuring the Intelligibility of Speech over Communication Systems* [5]. These methods are the aforementioned Modified Rhyme Test [75], the Harvard phonetically balanced word test [50], and the Diagnostic Rhyme Test [176]. Each of these methods employs the evaluation of monosyllabic English words.

The phonetically balanced word test (PB) contains 1000 words, organized into 20 word lists each containing 50 words. The occurrence of the various phonemes found in the stimulus set have a frequency proportional to that found in everyday English [50]. Subjects are presented with a stimulus word embedded in

a carrier sentence. Like the FRT, the PB is an open format test – meaning that responses are generated by subjects, rather than selected from a list of alternatives.

One major disadvantage of the PB is the training time required to achieve stable results from listeners. Additionally, subjects continue to show improvement in scores even after sufficient training, sometimes taking weeks, is completed. As such, comparison of test results obtained at different phases of testing is difficult [156].

The diagnostic rhyme test (DRT) is the final test mentioned in the ANSI recommendation. The DRT employs a smaller set of stimuli as opposed to the similar MRT and FRT. The DRT is a closed format test, containing 192 stimulus words grouped into 96 rhyming pairs. Stimuli are delivered, without a carrier sentence, to subjects who are presented with the two-alternative forced choice task of identifying the correct word [176].

As the DRT stimulus set is smaller than those of its rhyming predecessors, this method has the advantage of shorter testing times. However, this method has the disadvantage that, within each word pair, differences are found only between the first consonants of the words.

Other examples of tests that employ the presentation of single monosyllabic English words include the speech identification test (SIT) [71] and the California consonant test (CCT) [141]. The SIT was devised as part of an experiment which studied the effects of context on both listeners' ability to identify words, and their certainty about said identifications [71]. Though similar to the MRT, the CCT was developed specifically for the purpose of clinical testing on hearing impairments [141].

2.1.3.2 Conclusions

The MRT seems the most appropriate test for use in this research project. The use of a test which employs monosyllabic stimuli will allow for repetitive testing on subjects, and will remove the potential for errors in results due to contextual and semantic clues. The subject training requirements for use of the MRT are far less than what is required for the PB.

Unlike the FRT and DRT, the MRT includes variation in both initial and final consonants. When the potential effects of post-masking are considered [183], it is clear that the MRT has the potential for sensitivity to a wider variety of factors that affect intelligibility (e.g. masking of the initial-consonant phoneme by the carrier sentence and the final-consonant phoneme by the stimulus word's own vowel).

2.1.3.3 Use of the Modified Rhyme Test

The MRT was originally developed as a method for the evaluation of communication systems [75]. Since then, it has been further modified for use in clinical Audiology [95] and tested for viability and reliability [15, 21, 138]. The test has been thoroughly established as a method for investigating differential effects of variable treatments on word recognition. Such testing has been done to compare groups of listeners with normal- and impaired-hearing (e.g. [52, 136]), communication transduction methods (e.g. [73, 123]), communication system effectiveness (e.g. [9, 122]), and communication enhancements (e.g. [99]).

In their recommendation [5], ANSI details procedures for the use of monosyllabic word tests for the evaluation of a wide variety of communication applications. It is stated that the recommendation is intended to cover systems ranging from face to face communication; to public address, radio and telecommunication; to underwater and outer space applications. The recommendation is also intended to deal with entire communication systems, including not only sound transmission and reproduction but also sound capture transducers and the talker's environment.

The scope of the recommendation is broad and so too are the prescriptions detailed therein. While the prescribed methods are well established for general communications systems testing, full adherence to the recommendations may not be necessary considering the limited scope of this dissertation. For this research project, the focus is placed on comparing, rather than measuring the absolute performance of a communication system, and is limited to the transmission and reproduction end of the communication path. This project is not concerned with

absolute pass/fail criteria, nor is it concerned with variations associated with the use of different talkers. For example, it is not the intent of this document to prove the effective function of an emergency address system.

Among other specifications, the recommendation indicates that a minimum of five talkers and five listeners should be used and that, as the talker has a greater effect on intelligibility, more talkers than listeners should be used. The document also recommends that a minimum of three MRT words lists should be evaluated by each listener for each variable treatment.

While a review of the literature involving practical application of the MRT does show a common testing method, it also shows a great number of differences between the specific implementations of the test. As the purposes and overall goals of an experiment will vary from study to study, so too it seems do the applications of testing methods. The obvious difference between the reviewed studies lies in the number of subjects used. Subject population size is governed by the experimental design of a study, type of subjects used (e.g. naïve vs. expert) and the amount of statistical strength needed to show significance [14]. However, other notable variations between uses of the MRT do exist, including:

1) Number of word lists used

In the literature, the number of MRT word lists used ranges from one list [73, 123] to the full set of six, as seen in Kreul et al. [95, 96], Atkinson and Catellier's study on the intelligibility of radio-communication systems for firefighters [9] and many others. Examples also exist of the use of three lists [99, 122].

2) Number of word lists used per variable treatment

Beyer et al. [21], studying the validity of the use of the MRT for clinical study and Nixon [138], studying the effectiveness of the MRT response foil words, used the original Stanford Research Institute (SRI) recordings made by Kreul et al. [95] which include an incomplete matrix of six lists and three talkers. Kusumoto et al. [99], in a study of speech enhancement in reverberant environments, used three word lists, but only one word list per reverberation time treatment. Nabelek and Mason [136] used six lists total, but only two per treatment, to study the

effects of reverberation time, signal-to-noise ratio and monaural vs. binaural hearing on normal- and hearing-impaired listeners.

3) Number of talkers used to deliver the word lists

As mentioned, the original SRI recordings [95] used three talkers. For an experiment studying differences between a normal talker and a laryngectomee talker, Holley et al. [73] used one talker for each talker type. McBride et al. [123] used one male and one female talker in a study of bone conduction transducers that included talker gender as an experimental variable. Atkinson [9] used six talkers. Kusumoto et al. [99] and Nabelek and Mason [136] both used one talker.

4) Number of ensembles presented

Though most examples of MRT use employ the full set of 50 ensembles per list, examples do exist where truncated lists are used. Matthews et al. [122], for example, used 30-word lists in a study on work load demands associated with the use of cell phones while driving.

5) Auditory display method

The MRT has been used in studies in which the auditory display method, to list a few, was a loudspeaker [136], pair of headphones [73, 75, 99], bone conduction transducer [123] and even cell phone [122].

6) Open vs. closed response set

Though originally developed as a closed-set test [75], examples exist where specific elements of an experiment suggest the administration of the MRT in an open-set form. In the experiment conducted by Matthews et al. for example [122], that involved subjects evaluating words while driving, the use of a closed-set test could have both been dangerous and compromised results by adding to the work load of subjects.

It is evident that the needs and goals of a particular experiment dictate the specific implementation of a testing method. Though credence should be given to the ANSI recommendation, it appears that the specific methodology detailed therein can be modified to suit a specific research endeavor.

As mentioned, the studies detailed in this dissertation are limited to examining effects caused by variations in the reproduction end of a sound system.

Also, as the focus of this study involves comparison of intelligibility impairments rather than absolute evaluation (pass/fail), the use of one talker (as opposed to five) should not have a negative effect on the validity of results. It was therefore decided that the use of one talker would be appropriate so long as the talker's speech was clear, steady in rate and well articulated.

The ANSI standard recommends the use of three MRT word lists, in part, to reduce the effects of potential listener learning over repeated testing. As has been seen from the literature [73, 123], the MRT has been employed with fewer than three lists. The use of three word lists instead of one equates to a 300% increase in testing time and costs. While it would be desirable to use only one list, it is unclear whether this would compromise test results. Three lists were therefore recorded for use in this research project. During the course of this project it would be possible to introduce "word list" as an experimental variable. If results indicate that word list does not have an effect on results, further testing may be able to reduce the number of lists, or number of lists per treatment, used.

An additional point found in the literature is in regard to restriction of the subject population. Non-native English speakers typically score lower on tests involving the identification of English words. Lexical considerations, such as lack of familiarity with particular test words, can produce effects on test scores that are unrelated to the speech communication system [62]. As such, and in the interest of controlling variance, the studies detailed in this dissertation will only employ native English speaking subjects.

2.2 Sound System Design & Optimization

The end goal of sound reinforcement is to deliver the desired program material to all members of an audience, with appropriate level and spectral content, with a minimum of variance in level and spectrum between audience members [124]. In modern times, complex sound reproduction systems, comprised of subsystems containing multiple loudspeaker drivers and enclosures, are often employed to achieve this goal [11, 56, 60, 174].

Fortunately, in addition to more complex sound systems, the modern era has also brought with it modern measurement tools and loudspeaker processing technology. Modern computer processing power makes it possible to employ software which can generate 2- and 3-dimensional predictions of sound system performance [2, 22, 48]. Modern measurement tools can utilize dual-channel fast Fourier transform analysis, with an accompanying battery of acoustical metrics and post-processing algorithms [3, 49]. Additionally, modern loudspeaker processing tools (drive systems) not only possess tools which were previously unavailable (e.g. digital delay lines), but can perform the functions of a conventional "drive rack" in a fraction of the space [30, 98]. Many drive system tools are even found within modern digital mixing consoles [182].

It is fortunate indeed that these tools are available. In order for a modern, complex sound system to adequately perform its task, it is necessary for the components of the system to be individually calibrated, and then integrated into the overall system in a manner such that the union functions as a cohesive whole. This process is called sound system optimization [124].

Generally speaking, the individual processes falling under the blanket term optimization are loudspeaker focus, equalization, level alignment and time alignment [124, 171]. Loudspeaker focus involves the spatial and angular positioning of a loudspeaker such that sound radiates in the desired direction. In most cases the desired result is the maximal ratio of sound radiated towards the audience vs. sound radiated towards reflective surfaces, with even level distribution across the audience areas [41]. Notable exceptions include indirect radiation methods designed to deliver diffuse reflected sound to listeners (e.g. effects and surround systems), and components used to affect radiation interference in loudspeaker arrays (e.g. [86, 88, 165]).

Equalization is employed to correct frequency-response irregularities and/or shape the spectrum of reproduced sound. In addition to improving sound quality, equalization is also used to increase the headroom of a system (peak removal), reduce electroacoustic regeneration (feedback) and also help shape the radiation patterns of loudspeaker arrays [25, 34, 93, 124].

The goal of level alignment of a loudspeaker, or array of loudspeakers, is to achieve adequate level, and equal level distribution, over the entire audience area. When viewed from the perspective of the entire sound system, level alignment attempts to maximize headroom and SNR through the creation of a proper structure of gain stages. The level of each loudspeaker, and corresponding electronic chain upstream, is adjusted such that electronic signals at nominal operating levels will create the desired sound pressure level in the audience. Levels of multiple loudspeakers are adjusted such that sound pressure levels in different areas of audience coverage, as well as regions of overlap between coverage areas, have minimal variance [124]. Additional level alignment techniques can be employed for use in loudspeaker arrays to affect an array's overall radiation pattern (e.g. [88, 175]).

Time alignment, which will be explored in the following sections, is a term used to describe the use of electronic delays to affect the summation of signals from multiple loudspeakers. Before discussing the alignment of components of loudspeaker arrays, it would perhaps be beneficial to review the need for, and variations of, loudspeaker arrays.

2.2.1 Types of Loudspeaker Arrays

In the early days of audio engineering, researchers attempted to develop a loudspeaker that could reproduce the full range of audible frequencies from a single source (see [16, 36], for example). In his exhaustive review of loudspeaker literature, Gander [61] describes the full-range, single source loudspeaker as a "holy grail" of audio engineering. He also points out that over time, as the physics and limitations of transducers were better understood, efforts to develop the mythic device were all but abandoned in favor of multi-way, multi-transducer systems.

Analogously, researchers have spent much time and effort in the pursuit of whole-room reproduction solutions that require only one loudspeaker per channel of reproduced information. While such solutions can often be found for small rooms, or small listening areas, as the size of the listening area increases the

effectiveness of single-loudspeaker solutions diminishes [124]. Thus, more complicated solutions are required. This section includes a review of reinforcement needs and solutions, array fundamentals and types, and a comparison between array solutions for small and large listening environments.

For effective sound reinforcement, two basic needs must be fulfilled. First, adequate level is required such that a minimum SNR can be achieved throughout the audience listening area. Appropriate loudspeaker directivity is likewise necessary, as direct sound from the reinforcement system must reach all listening areas [65, 124, 171]. As previously mentioned in the discussion of $\%AL_{cons}$, low loudspeaker directivity (wide radiation angles) can result in increased reverberant energy due to projection of sound energy towards reflective surfaces. In an attempt to increase the D/R ratio, and thus increase sound clarity, it is necessary to minimize the amount of sound radiated toward any location other than the listening areas [41].

From the 1930's through the development of the constant directivity horn in the mid 1970's to early 1980's, advances in horn technology have yielded increases in both acoustical power output and directivity control [47]. Despite these achievements, it is often the case that the output and coverage of a single loudspeaker proves insufficient for the needs of a production/installation. As such, combinations of multiple loudspeakers (i.e. arrays) are therefore employed.

In terms of power and pressure projected from (radiated by) an array, the addition of loudspeakers amounts to an increase in sound power output. By adjusting the angle between loudspeakers (splay) in an array, it is possible to affect the distribution of radiated sound pressure. In general, compared to a single loudspeaker, arrays with small splay angles result in higher pressure output over the same approximate coverage area, while arrays with wide splay angles distribute the equivalent pressure output of a single loudspeaker over a larger effective coverage area. Varying splay angle results in a trade-off between the coupling and spreading of projected sound pressure, allowing for many possibilities of total array output and radiation [57, 124, 128].

In addition to power/pressure projection, the summation of the sound energy output from multiple loudspeakers can create an interference pattern in an array's radiation (see chapter 2.2.2 for an explanation of summation). In the late 1980's, a variety of studies were conducted on the effects of interference on the total radiation pattern of "conventional", convex arc-style arrays composed of real sources. Studies by Fidlin and Carlson [57], Meyer and Seidel [128] and Gander and Eargle [60] all reveal the frequency-dependent, finger-like structures (lobes) that appear in array radiation patterns, and irregularities (comb filters) in resulting magnitude- vs. frequency-response measurements.

While the results of interference patterns are often undesirable (e.g. the formation of lobes), intentional manipulation of inter-loudspeaker interference can achieve desirable effects. Through the use of time delay, polarity inversion, level tapering, band limiting and spectral shading, it is possible to manipulate radiation patterns, enable beam steering and provide directivity control at low frequencies [101, 124, 127, 178]. Examples include directional woofer arrays [12, 24, 27, 68, 72], Bessel arrays [88], line source and column arrays [70, 93], Colinear arrays [10], discrete element line arrays [154, 165, 175], linear and planar CBT arrays [89] and striped panel arrays [8].

Despite the wide variety in type, orientation and manipulation of elements used in loudspeaker arrays, the vast majority of arrays used for sound reproduction and reinforcement can be represented by one, or a combination, of three different fundamental forms [124]. Figure 2.7 shows these three forms (line source, point source and point destination) and the two permutations of each (coupled and uncoupled). Most modern reinforcement systems employ a main system to cover the majority of the listening area, supplemented by one or more satellite systems to fill in gaps in the coverage of the main system.[3] When combined in a system, main and satellite (fill) systems – each containing either single loudspeakers, coupled arrays or uncoupled arrays – form uncoupled arrays [124]. In essence, a sound system can be viewed as a primary array composed of

[3] Multiple main and satellite systems can be used for multi-channel (e.g. Left-Right and Left-Center-Right) reproduction.

subsystem arrays. Each of these arrays, primary and subsystem, may be oriented vertically or horizontally.

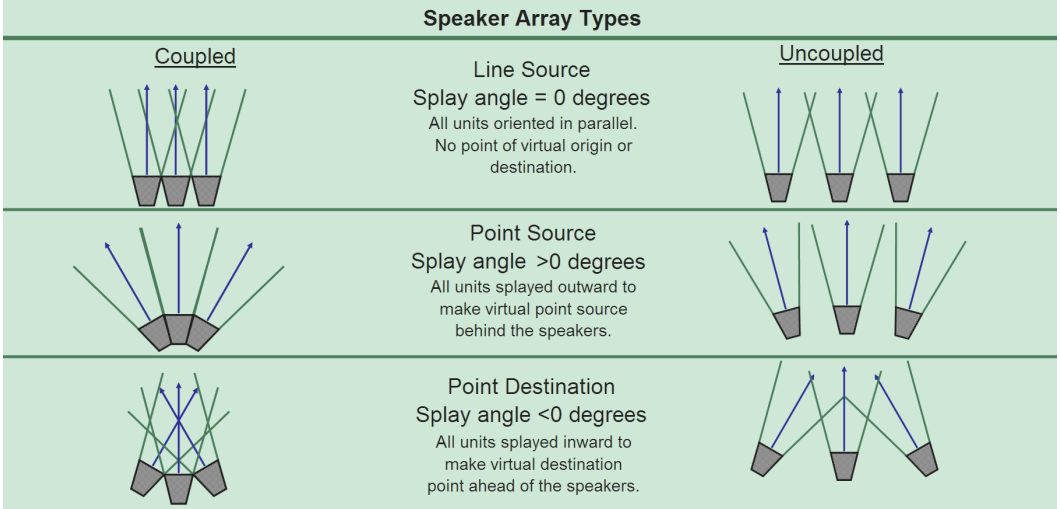


Figure 2.7 The six fundamental loudspeaker array types (Reprinted with permission from [124])

The studies detailed in this dissertation employ two examples of common arrays. The first comprises an uncoupled point-destination array of single loudspeakers, taking the form of a main subsystem and fill subsystem, as is typically used for supplemental coverage of the first few rows of an audience listening area. The second forms a typical uncoupled point-source array of front-fill loudspeakers: An example of a subsystem array, employed when a single satellite loudspeaker can not provide sufficient fill coverage.

As alluded to in chapter 1.2, the choice of these two array types stems from the use of an organic approach for the research project. Obvious differences exist between these two array types, making direct comparisons difficult. The intent of the series of studies, however, is not to compare intelligibility results for the two array types, rather to determine whether the effects of time alignment on intelligibility are different for different array types. If such differences are found, it would be possible in future studies to deconstruct and isolate the various variables within the complex variable “array type”. Chapter 3.2.2 further details the differences between the array types used in this project.

The result of the summation in coupled arrays is effectively the same as the result of summation in uncoupled arrays. The primary difference is the minimum distance from the array required before full contribution of all array elements is achieved [124].[4] Thus the results from the current research project could be applied to both coupled and uncoupled arrays and, as such, array coupling can be considered a controlled variable in this research.

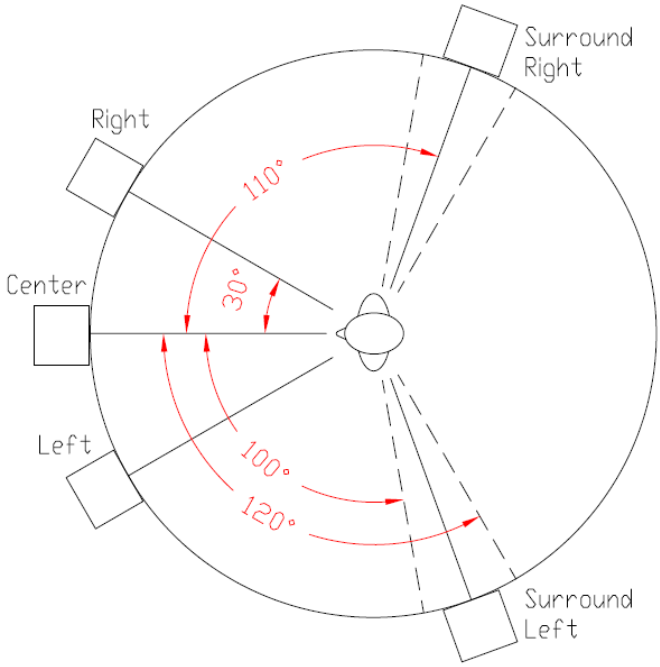


Figure 2.8 5.1-channel surround sound loudspeaker array, including left, center and right channels, found in [86]

While the focus of this dissertation is on loudspeaker systems for sound reinforcement, it would also be beneficial to consider systems that are used for sound reproduction. Inspection of the International Telecommunications Union recommendation [86] reveals two subsets of the 5.1-channel surround sound system: The 2-channel stereo and 3-channel stereo reproduction systems (see figure 2.8). These two systems are analogous to Left-Right (LR) and Left-Center-Right (LCR) systems used for sound reinforcement. The 2-channel and 3-channel

[4] A thorough discussion of individual Fresnel/Fraunhofer, chaotic and collective Fraunhofer regions of array radiation is beyond the scope of this dissertation. The interested reader is directed to [69].

systems are both uncoupled point-destination arrays, though the individual channels do not necessarily produce coherent signals.

Generally speaking, in television, movies and live performance, it is intended that speech signals are perceived to originate from the center in front of the listener. In the 2-channel system, centrally localized sound images are formed by the summation of equal-level contributions from the left and right channels. These are referred to as phantom images. In the 3-channel system, centrally localized sound images are created by a physical loudspeaker. Experiments by Holman [74] and Shirley et al. [162, 163] have shown that the use of a dedicated center channel results in increased intelligibility and clarity of speech when compared to phantom sound sources.

The orientation of the two arrays used in this research project results in a speech delivery method similar to those of the 2-channel and 3-channel systems. The main/fill primary array is oriented in the medial plane and provides a physical loudspeaker in front of the listener. The fill/fill subsystem array creates an auditory event in front of the listener due to summing localization [23]. As such, it is expected that there will be some difference in intelligibility scores between the two arrays.

2.2.2 Multiple Arrivals & Summation

When compared to single loudspeaker sound system, the use of multi-loudspeaker sound systems presents an additional layer of complexity. When multiple loudspeakers are used to reproduce sound signals, it is possible that a listener may hear sound from more than one loudspeaker. Equation 2.1 shows that the travel time of sound, for example from a loudspeaker to a listener, is proportional to the distance between the loudspeaker and the listener.

$$t = \frac{d}{c} \qquad \text{(Eq. 2.1)}$$

Where d is distance, c is the speed of sound and t is time.

If multiple loudspeakers are not the same distance from the listener, the projections from these loudspeakers will reach the listener at different times. The

listener will experience a form of electronically generated “echo”, referred to as multiple arrivals of a sound signal.

As mentioned in chapter 2.1.1, though it is debated whether echoes of various delay times can affect intelligibility, it is clear from the literature that the auditory system has some ability to fuse portions of the early sound with the direct sound. Analogous to the psychological results of auditory fusion, the phenomenon of summation is the combining of sound signals in the physical domain [42, 124].

The next few sections will describe some of the measurable and subjective effects of, and some of the methods that been developed to deal with, the summation of multiple arrivals.[5]

2.2.2.1 Background & Measurable Effects

The sine wave is one of the fundamental building blocks of sound. Figure 2.9 shows some of the basic properties of a sine wave, including amplitude,

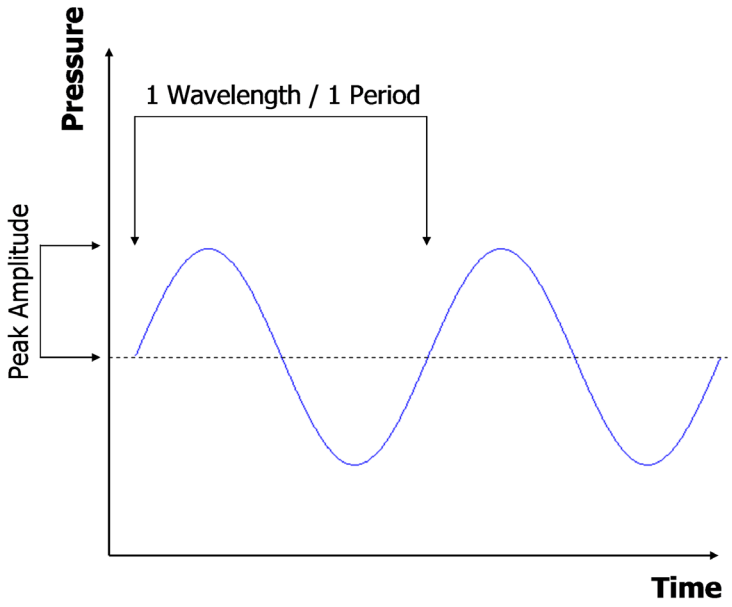


Figure 2.9 Properties of a sine wave

[5] It should be noted that the discussion of multiple arrivals and summation in this text is intended to describe the summation of direct, coherent sound radiated from loudspeakers and/or discrete echoes, not the summation of reverberant energy.

wavelength and the propagation of sound over time. As a sine wave propagates, its amplitude alternates back and forth between peak and minimum amplitude. The speed at which the wave completes its cycles is called frequency (*f*), and is measured in cycles per second (Hz). One full alternation (from peak to minimum back to peak) is called one cycle of the wave, and is also referred to as 360 degrees of rotation through a cycle. The time required to complete one full cycle, the period of the sine wave, is equal to the inverse of the wave's frequency, as shown in equation 2.2. The physical length (in air) of one cycle of the wave, the wavelength (λ), is determined by equation 2.3.

$$T = \frac{1}{f} \qquad \text{(Eq. 2.2)}$$

$$\lambda = f\,c = \frac{c}{T} \qquad \text{(Eq. 2.3)}$$

When comparing two sine waves, one can describe many things including differences in amplitudes, frequencies and a third term: Relative phase. Relative phase is expressed in terms of degrees, radians or wavelength, and describes whether two waves are at the same point in their respective cycles. For example, two waves that are at identical points in their cycles are deemed to have a zero-degree difference in relative phase (equal), while waves that are offset by ½-wavelength (figure 2.10) are deemed to have 180° difference in relative phase (inverted).

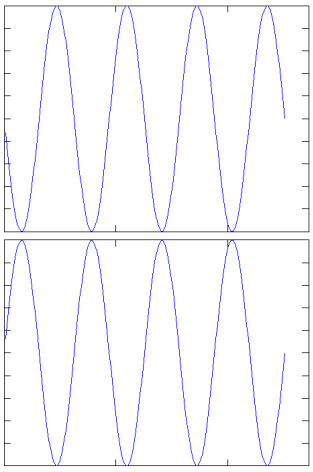


Figure 2.10 Example of two sine waves, 180° difference in relative phase

Basic Summation

When two sound waves arrive at a location in space they have the potential to constructively or destructively sum [42]. For example, if two sine waves of equal frequency, relative phase and amplitude arrive at a point in space, the summation of the two waves will produce a resulting wave with twice the amplitude of the original waves (figure 2.11). Alternatively, if the same two sine waves have a relative phase difference of 180°, the resulting summation will yield total cancellation (figure 2.12). The term summation refers to both the addition and cancellation of waves.

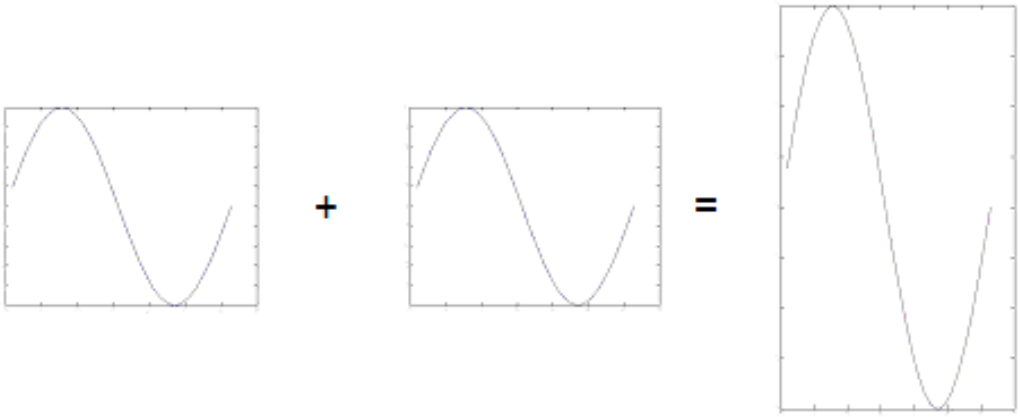


Figure 2.11 Summation of equal level sine waves with 0° relative phase

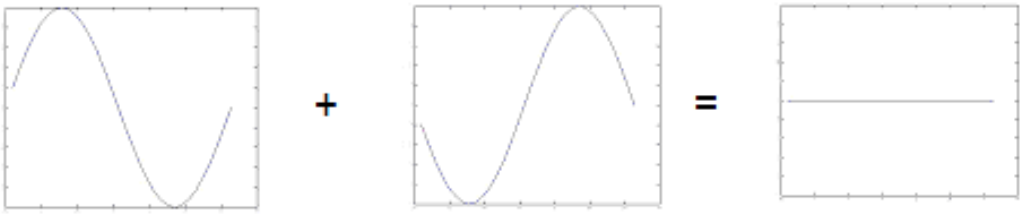


Figure 2.12 Summation of equal level sine waves with 180° relative phase

In addition to the 0° and 180° conditions, summation can occur at any degree of relative phase difference between two waves, resulting in varying degrees of constructive and destructive interference. Figure 2.13 shows the resulting gain or loss, compared to original amplitude, of the summation of two equal level sine waves for different relative phase offsets. As McCarthy [124]

points out in his book, *"1 + 1 = 1 (± 1)"*. In other words, the summation of two equal level signals can produce a result anywhere between double the original level (+6 dB) and no level at all (−∞ dB) depending on the relative phase of the two summed signals. When the summed sine waves are not of equal level, the potential for addition and cancellation is not as great.

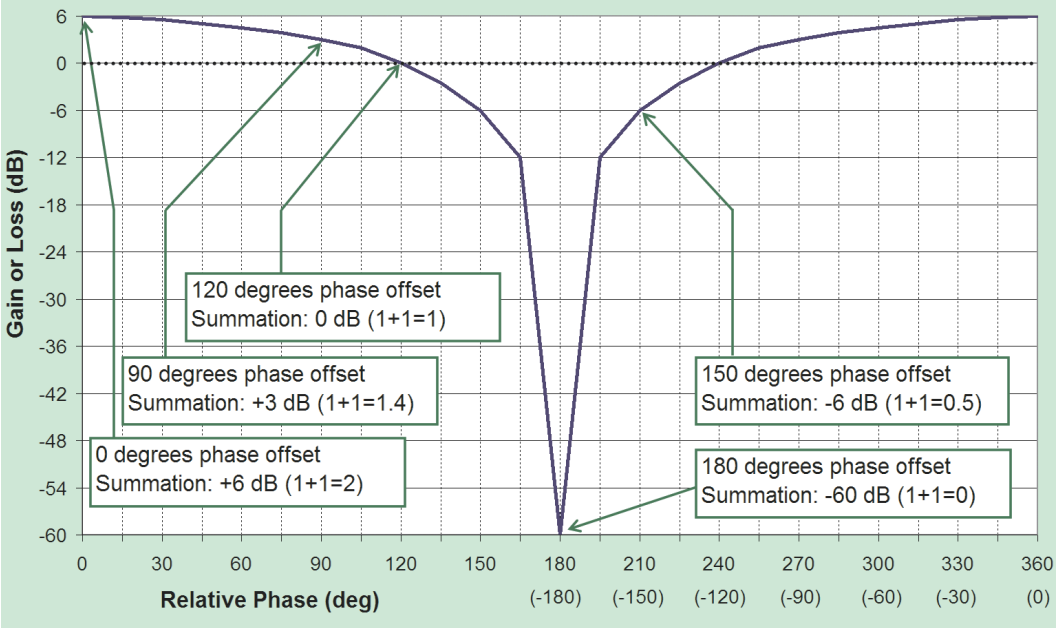


Figure 2.13 The effect of relative phase on the summation of two signals with equal level (Reprinted with permission from [124])

The summation of sine waves is independent of the wave's angle of incidence on a point in space, as described in figure 2.14. This is of interest when considering loudspeaker signal summation, as various angles of incidence would likely (though not necessarily) be the case with sounds produced by multiple loudspeakers.

Complex Summation

From harmonic analysis, it is known that complex sound signals, such as speech and music, can be represented as a combination of sine waves with varying amplitudes and relative phase [140]. As such, the properties of summation for complex sound signals are the same as those for simple sine waves [124], represented in figure 2.15.

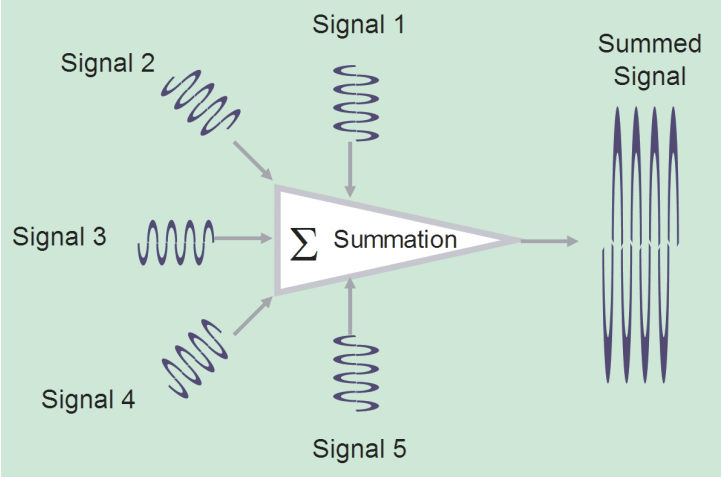


Figure 2.14 Equal relative phase summation of audio signals with non-identical angle of incidence to the receiver (Reprinted with permission from [124])

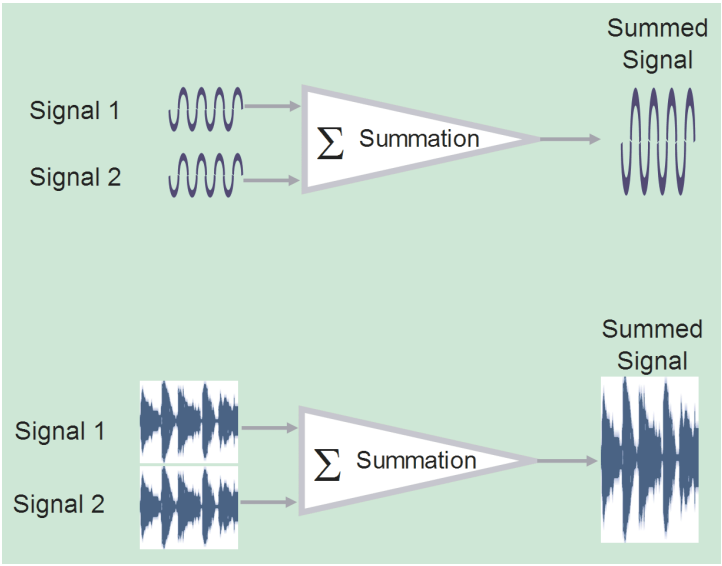


Figure 2.15 Equal relative phase summation of sine waves and coherent complex audio signals (Reprinted with permission from [124])

While the properties of summation are the same, the results are somewhat more complicated. As complex signals are effectively composed of a combination of sine waves, complex summation involves the many individual summations of these components. Figure 2.16 describes a simple example of the complex summation of identical signals that contain four frequency components each. The arrivals of the two signals at the spatial summation point differ by

some amount of delay. As each frequency has a different period, a shift in time will affect a different relative phase shift for each frequency. Thus, summation of each of these frequency components will yield different results. The example shown in figure 2.16, involves the summation of an original signal with a copy which has been delayed by 5 ms (approximately 1.7 m in air, 1 wavelength at 200 Hz). In this example, the 100 Hz components (½ λ offset) completely cancel, while the 125 Hz components (5/8 λ offset) only partially cancel. The 200 Hz components (1 λ offset) fully sum as, seen in figure 2.13, an offset of one full wavelength is effectively no offset: A 360° offset is equivalent to a 0° offset. It should be noted that, in practical application, signals rarely perfectly sum or cancel [124].

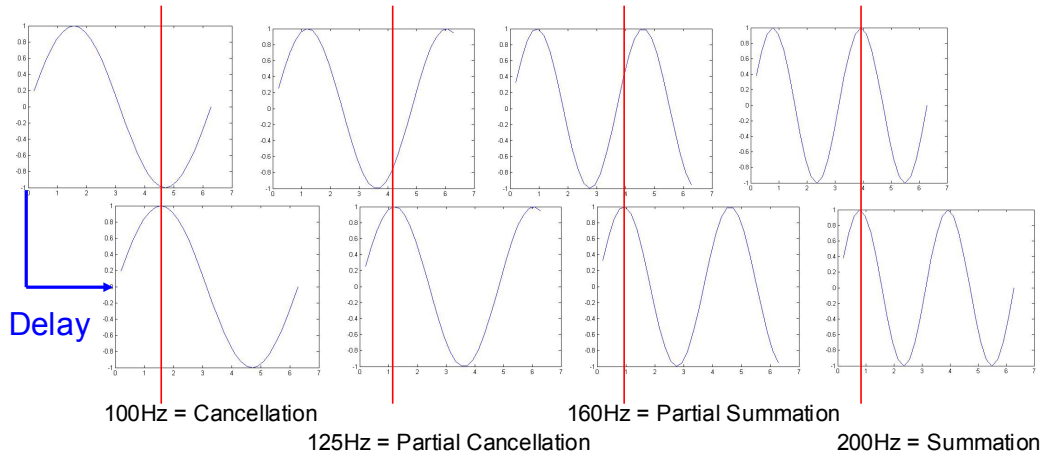


Figure 2.16 Summation with delay for a complex waveform with four different frequency components

Figure 2.17 shows the result of the summation of two real-world signals with the same delay time (5 ms), but for a more complex signal. The total cancellation at 100 Hz (½ λ) is still present, as are the effects at the other three frequencies used in figure 2.16. Proceeding higher in frequency, we note cancellation at 300 Hz (3/2 λ), summation at 400 Hz (2 λ), cancellation at 500 Hz (5/2 λ) and so on as wavelength offset increases vs. frequency. The result of the time-delayed summation of identical complex signals is a frequency response structure called a comb filter [46, 54].

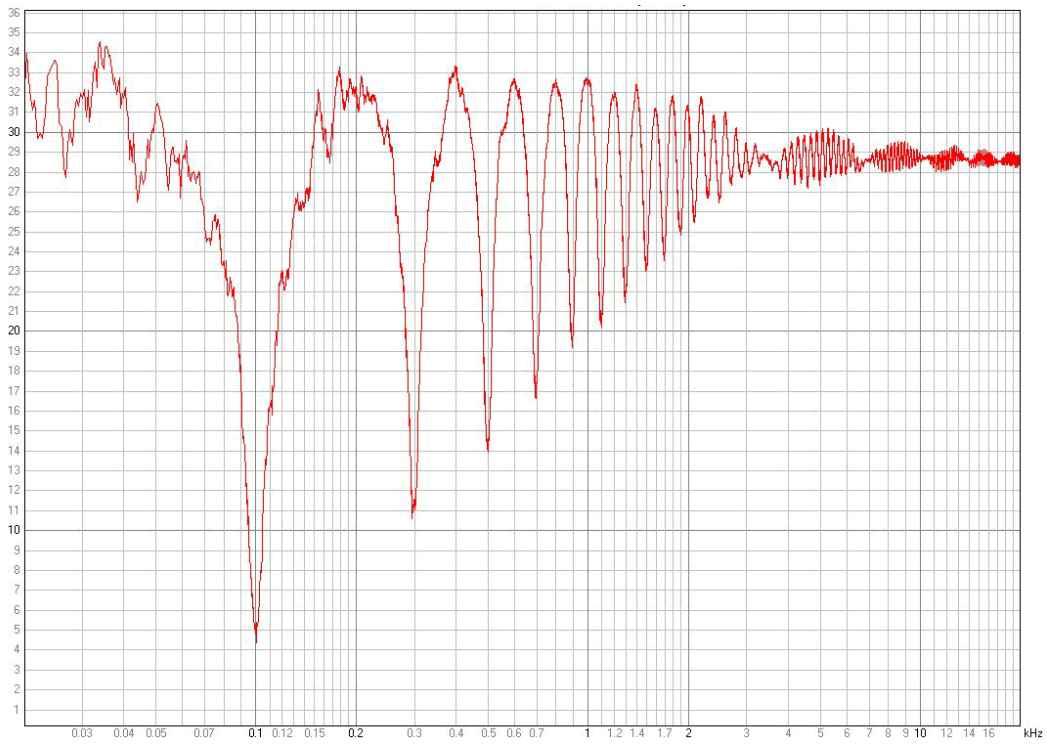


Figure 2.17 Summation with delay of a complex signal (5 sec pink noise). 1/12th octave smoothing. Note that the response seen in this graph was generated via electrical summation, thus the extreme depth of the notches.

The specific structure (spacing of peaks and dips) of a comb filter is determined by the time offset between summed signals [1]. For example, the summation of signals with a 0.1 ms delay time (approximately 0.034 m in air, 1 λ at 10,000 Hz) will see the first ½ λ offset (180° shift in relative phase) at 5,000 Hz (figure 2.18). The summation of signals with a 1 ms delay time (0.34 m in air, 1 λ at 1,000 Hz) will first cancel at 500 Hz (figure 2.19), and a 10 ms delay (3.4 m in air, 1 λ at 100 Hz) will cause canceling to start as low as 50 Hz (figure 2.20).

As the extent of summation is affected by the relative levels of the two signals being summed, the height of the peaks and depth of the notches in the resulting comb filter will be determined by a difference in level between signals. The most volatile effect is seen when summed signals are of equal level [124].

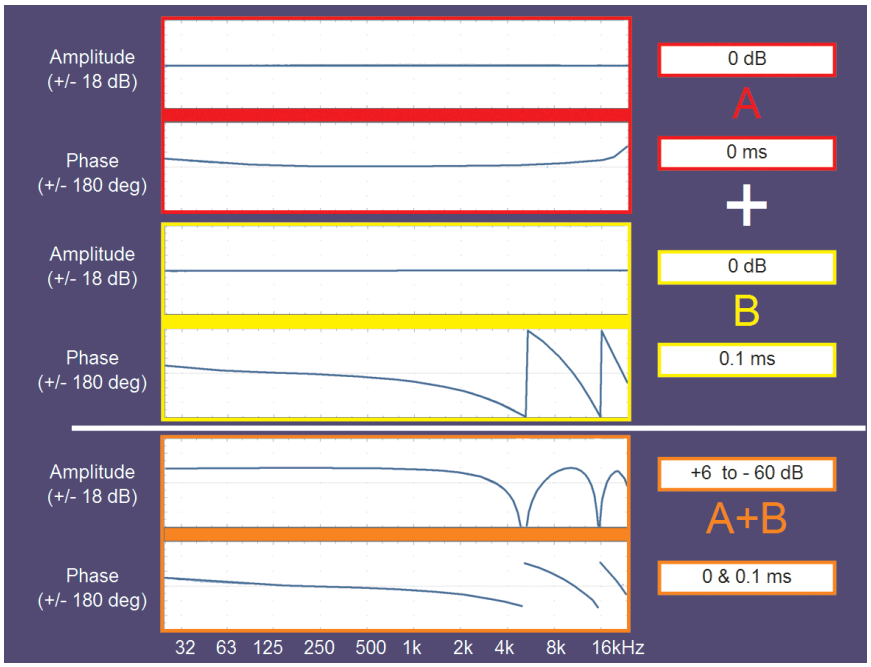


Figure 2.18 The effect of a 0.1 ms time offset on the frequency and phase response of summed coherent audio signals with equal level and relative phase (Reprinted with permission from [124])

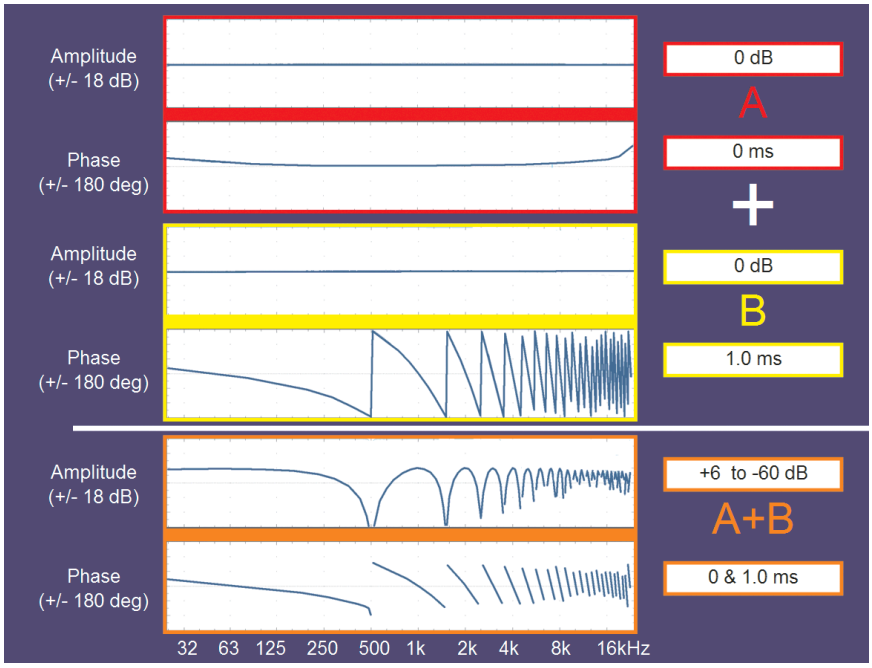


Figure 2.19 The effect of a 1 ms time offset on the frequency and phase response of summed coherent audio signals with equal level and relative phase (Reprinted with permission from [124])

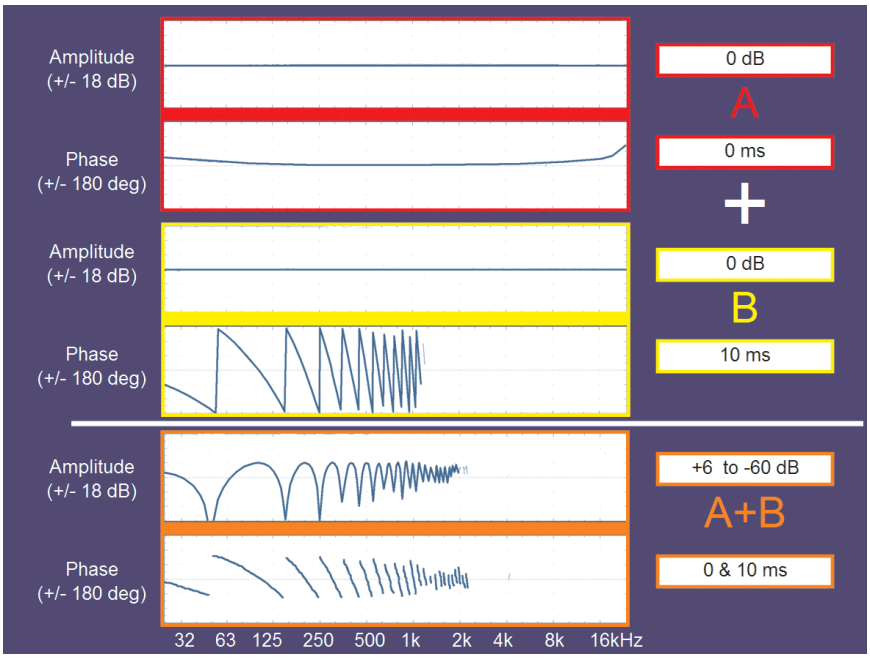


Figure 2.20 The effect of a 10 ms time offset on the frequency and phase response of summed coherent audio signals with equal level and relative phase (Reprinted with permission from [124])

2.2.2.2 Subjective Effects

In general, the subjective effects of multiple arrivals can be broken into several categories: Loudness, source localization, tonal coloration, echo perception and spatial enhancement [42, 67, 150]. In terms of the delay time between multiple arrivals, there are essentially three regions – short, medium and long – with different subjective effects contained in each. While researchers nominally agree on the different subjective effects, and that the three delay-time regions exist, there is still much debate regarding the specific delay times that correspond to the transition points between regions. There is also discussion regarding the role that the type of sound stimulus (e.g. speech, music, transients) plays on the temporal location of these transitions. The author therefore wishes to point out that the delay times listed in the following are approximate, and not intended to be construed as absolute transition points.

Multiple arrivals with short delays (0 ms–20 ms) create loudness, tonal coloration and spatial localization effects. Temporal fusion in the auditory system will integrate two arrivals falling in this region and affect an increase in perceived loudness. A short delay time can affect the perceived location of a sound source, known as the precedence effect [67], and can also create severe comb filters in the spectrum of the received, summed sound. As delay time increases, the strength of the precedence effect diminishes. Also, analogous to the phenomenon known as the Schroeder frequency [108], the frequency above which compactness of modal resonances in a room leads to inaudibility and statistical insignificance, higher delay times between multiple arrivals leads to greater notch density – eventually reaching a point wherein the resulting comb filter is inaudible.

Though there is some dissent, it is generally agreed that delay times of medium length (20 ms–80 ms) between multiple arrivals will still incur an increase in perceived level due to fusion [17, 41, 150]. Also in this time region, a spatial enhancement effect is seen, in the form of enlargement of the perceived width of the sound source (the auditory source width or ASW) [17, 150]. From the perspective of architectural acoustics, increased ASW is believed to add pleasantness and fullness of tone to sound sources.

The transition point between medium and long delay times is also vague, and depends on the nature and number of arrivals. In the case of many arrivals, e.g. reverberation, the perceived loudness of the original arrival is not affected. However, as arrivals tend to reach a listener from many directions, including the rear, the listener will experience an increase in the spatial enhancement known as listener envelopment (LEV) [150]. For the case of few arrivals with long delays, a listener will likely perceive these arrivals as separate distinct echoes.

2.2.3 Compensating for Multiple Arrivals

The creation of comb filters, through the summation of signals arriving from multiple loudspeakers with short delay times, results in an unnatural and unpleasant listening experience. Additionally, the path lengths between the loudspeakers and listener will likely vary over an audience area, as will the delay

between arrivals, resulting in different patterns of comb filters for different areas of the audience. The consequence of this is wide variance in the spectrum of sound received across the total audience that cannot be removed through loudspeaker equalization [124].

In order to preserve uniformity of sound system frequency response over the entire audience, the arrivals of signals from multiple loudspeakers would need to be aligned for all locations in the audience listening area. However this is rarely possible. As such, two distinctly different theories, each employing the use of electronically added delays, have developed regarding the method to compensate for the effects of multiple arrivals. Detailed below, these two theories are alignment and intentional misalignment.

Acknowledging that a single delay time cannot align the arrivals of two loudspeakers over an appreciable listening area, researchers have developed methods to intentionally misalign loudspeaker arrivals. As can be seen in figure 2.18 and figure 2.19, small offsets in arrival time (0.1 ms–1 ms) result in a sparse comb filter pattern with wide notches in the frequency band attributable to speech intelligibility (approximately 250 Hz–8 kHz). Conversely larger time offsets, such as 10 ms as seen in figure 2.20, result in a more compact structure of comb notches. The argument for misalignment is that the tight grouping of notches resulting from larger offsets would be less detrimental to sound quality than the sparse notch structure resulting from smaller offsets. It should be noted that while studies on misalignment have addressed sound and speech quality, speech intelligibility has been only peripherally discussed.

Mochimaru [131] points to the fact that, for acoustical summation, the level difference between peak and notch decreases to less than 6 dB SPL for frequencies above the 3/2 λ notch. He states that if the 3/2 λ notch falls at a frequency below the effective range of a loudspeaker, the resultant level variance due to combing would be less than 6 dB. Thus Mochimaru recommends that a minimum offset between arrivals, corresponding to 2 λ at the lowest reproducible frequency, should be ensured through the use of electronic delay.

For coupled point-source array configurations in which the low- and high-frequency transducers are separate, such as in "long-throw" and "near-far" horn arrays, the delay times required to create misalignment are quite short (2 λ at 500 Hz–2 kHz ranging from 4 ms–1 ms). However in more modern loudspeaker arrays, with the majority of frequencies being reproduced by transducers in the same enclosure, misalignment would require much larger delay times (e.g. 2 λ at 100 Hz would be 20 ms). Additionally, some sound system configurations make it difficult to ensure that a minimum arrival offset can be maintained. For example, when loudspeakers are located near the audience (e.g. an uncoupled array of front-fill loudspeakers), large delay offsets may be needed to maintain minimum offsets for listeners seated close to a loudspeaker whose signal is intended to arrive second.

El-Saghir and Maher [53] offered a similar misalignment technique called "milli-delay". Again citing that very small time offsets between arrivals can cause audible comb filtering, results of their study suggest the use of intentional time offsets in the range of 10 ms–35 ms to compact the pattern of combing to the point of inaudibility. As with Mochimaru's study, El-Saghir and Maher focus on coupled point-source arrays that are located some distance away from audience members.

Augspurger et al. [11] propose a different method of intentional misalignment. Their theory was that loudspeaker arrivals should be aligned as closely as possible and that the use of electronic stereo synthesis, to pre-comb loudspeaker signals, could effectively "scramble" the residual audible effects of combing due to multiple arrivals. Through subjective testing to evaluate fidelity, sharpness and brightness, it was found that this method could be effective in the 800 Hz–5 kHz range. However Augspurger et al. caution that the use of this alignment technique could result in timbral changes and that the frequency region of effectiveness is highly dependent on the ability to phase-match transducers. As with the previous studies mentioned, the loudspeaker arrays used in this study are located well away from audience members.

The intentional misalignment of loudspeaker signals in medium-to-large scale reinforcement appears to have fallen out of favor as, with one exception detailed below, more modern reports on signal alignment mentioned in the literature advocate actual alignment. Davis and Davis [42], Ahnert et al. [4] and McCarthy [124] have all indicated that the minimization of the effects of comb filtering is best accomplished by minimizing time offset between arrivals.

As mentioned previously, one of the benefits of loudspeaker arrays is increased power projection. In recent years, increases in available electrical (amplifier) power and the power capabilities of loudspeaker drivers have, to a large degree, reduced the need for high degrees of overlap in high-frequency coverage patterns in point-source arrays [47]. Considering the reductions in off-axis high-frequency radiation from a loudspeaker, larger splay angles between loudspeakers would result in lower levels of high-frequency components of one loudspeaker infiltrating the coverage area of an adjacent loudspeaker. As the levels of multiple arrivals within each coverage area would differ in level, summation of the arrivals would be less volatile yielding reduced notch-depth in the resultant comb filters [1]. However the point (or line) within the listening area where signals from multiple arrivals are equal in level is the location where comb filtering would be the most severe. As such, alignment of arrivals at this volatile location will result in minimum spectral variance over the listening area [124].

As Dickens notes [44], with regard to arrays used for sound field reproduction, the absolute alignment of equal-level sound sources with a high degree of overlap results in a small area for ideal listening. Reducing the degree of overlap and/or the level of the delayed arrival are two methods that can yield reduced spectral variance across a larger listening area [124].

Increased splay angle between loudspeakers is one design method wherein level differentials between arrivals from elements of point-source arrays reduce comb filtering and thus increase the size of the ideal, or at least useable, listening area. Additionally, level offset between arrivals due to differences in listener-loudspeaker distance can also increase the useable listening area. As a listener moves from a point equidistant from two loudspeakers to a point closer to one,

delay time between arrivals changes creating a combing structure – but level differential changes as well, resulting in reduced depth of comb notches [1]. For uncoupled arrays, the spacing of loudspeakers is an additional factor in determining the useable listening area. As is the case with coupled point sources, the location(s) of highest volatility for uncoupled arrays is found where the levels from two loudspeaker arrivals are equal. Again, alignment of arrivals at the point of maximum volatility results in acceptable levels of spectral variance throughout a larger listening area [124].

An additional viewpoint regarding the alignment of loudspeaker signals has been offered by Brown [28], with regard to the preservation of stereo imaging of reinforced sound in large rooms. With stereo reproduction, one must consider not only the effects of the summation of coherent loudspeaker signals, but also the summation of incoherent microphone signals. An example given is the use of a spaced pair of microphones to capture a two-channel stereophonic sound image of an instrument(s) (e.g. drum overhead). Each microphone will capture sound arriving from its side of the instrument as well as a delayed sound arriving from the opposite side. As time delays are involved, the summation of these microphone signals could result in comb filtering [28, 124].

In his discussion of electrical vs. acoustical summation, Brown [28] advocates the maintenance of stereo signals in supplementary satellite loudspeaker systems. More relevant to this discussion, Brown also states that the arrivals from satellite systems should lag the arrivals of the main sound system and that, by employing the precedence effect, the main sound system will be the perceived sound source.

2.2.4 Previous Experiments

As mentioned in previous chapters, a multitude of experiments have been conducted with regard to the reception of speech through a variety of transmission channels. From studying acoustic speech in a room to telecommunications and reinforcement sound systems, the goals of, and results from, the existing research

have been as diverse as the transmission methods studied. The works cited herein focus on determining the effects of multiple arrivals on intelligibility.

Steeneken & Houtgast [167] published a series of studies including the effects of SNR and a single echo on intelligibility as measured by the full STI and subjective tests (PB word list). Figure 2.21 shows the measured MTF's for the 500 Hz band as well as PB word scores. As expected, inspection of the graphs for the "+∞ dB" SNR (no noise) treatments reveals notches in the MTF at 10 Hz, 5 Hz and 2.5 Hz (for the 50 ms, 100 ms and 200 ms conditions respectively).

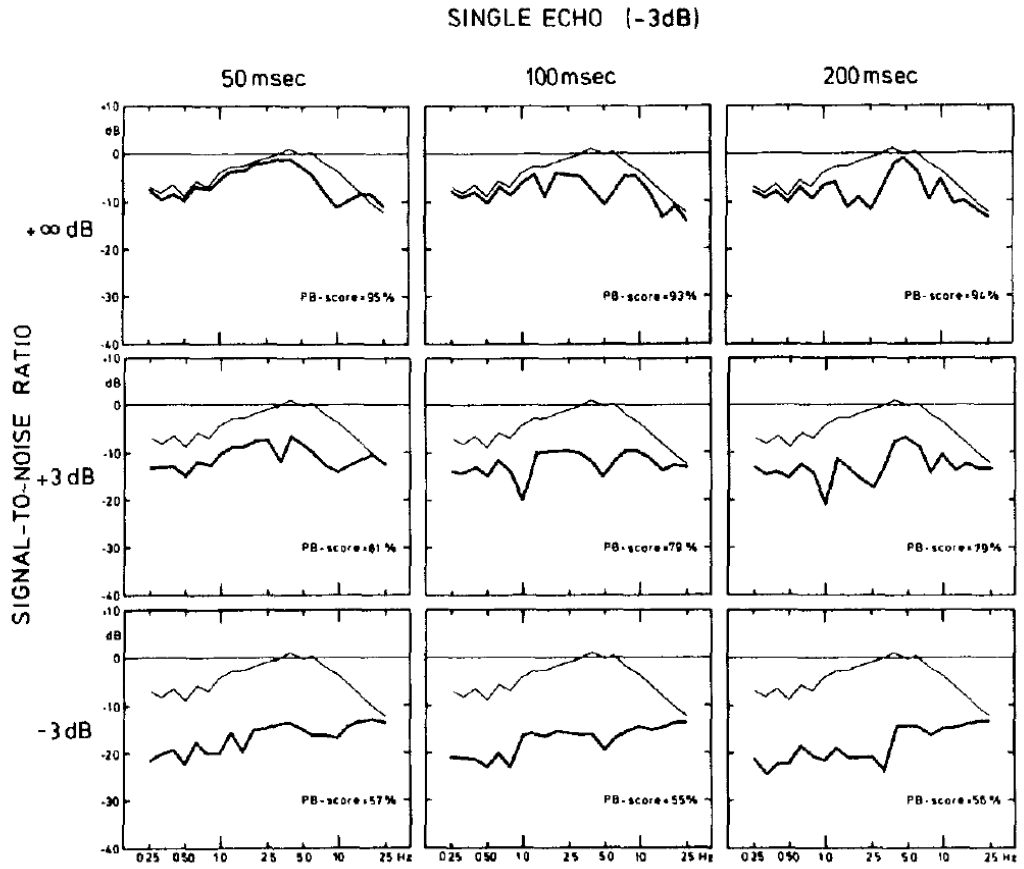


Figure 2.21 MTF graphs of the 500 Hz band, showing the interaction effects of SNR and delay time on intelligibility (dark lines), compared to the MTF graphs of the unaffected transmission system (Reprinted with permission from [167])

However, two interesting things can be found in these graphs. The first is that a higher noise level (lower SNR) appears to mitigate the effects of the echo – the notches found in the no noise treatments are less defined in the graphs of the

higher noise conditions. Also the PB word scores, again for the no noise treatments, are 96%, 93% and 94% respectively. What is not shown is that the word score for the unaffected transmission system was 99%. It is interesting that, though the reductions in word score were minimal, it would appear that there is at least some correlation between the measurements and the subjective results. If the echo were equal in level to the direct signal, it follows that the effect of the delay would have been greater. Angle of incidence of the echo is not mentioned, but is assumed to be from the same direction as the direct sound.

An investigation by Peutz [144] using $\%AL_{cons}$ indicates that, for a horizontally arriving echo, intelligibility appears to be affected (loss of 0.5 $\%AL_{cons}$) with delay times as short as 15 ms. The results of the study further indicate that intelligibility will decrease for delay times up to 45 ms (loss of 3% AL_{cons}), remaining constant beyond that delay time. Again it is unclear as to how well this calculation correlates with subjective impression, and whether a difference of 0.5–3 $\%AL_{cons}$ would be detectable in practical application. In a study by Davis and Davis [41] it is stated that echoes arriving within 3 ms of the direct sound can affect intelligibility by creating deep comb filter notches in the spectrum of speech. Though RaSTI and $\%AL_{cons}$ measurements were made for such delay times, it is unclear as to whether these findings are supported by results from subjective evaluation.

Teuber and Völker [170] conducted a series of studies involving the use of RaSTI to determine the effects of single and multiple echoes, and the use of compensatory delay, on intelligibility. In a laboratory study, employing a digital delay device to create echoes of the RaSTI test signal, Teuber and Völker measured the effects of varied delay times and echo level. The results, shown in figure 2.22, indicate that RaSTI values drop significantly between 20 ms and 60 ms, with a continued overall decrease as delay time increases. Similar results were found when multiple echoes were injected into the measurement. As mentioned previously, echoes create comb filters with spectra determined by delay time. A summed echo with delay time of 40 ms will have its ½ λ notch at 12.5 Hz, which is very close to one of the modulator frequencies used by RaSTI

(see [110] for RaSTI MTF matrix). Similarly, echoes between 40 ms and 70 ms will have an effect on the RaSTI MTF at the upper two modulator frequencies. Echoes with longer delay times will result in ½ λ cancellations at lower modulator frequencies with 3/2 λ cancellations eventually entering the MTF spectrum. Considering that RaSTI only uses 9 measurements, it is debatable whether these data are indicative of actual intelligibility.

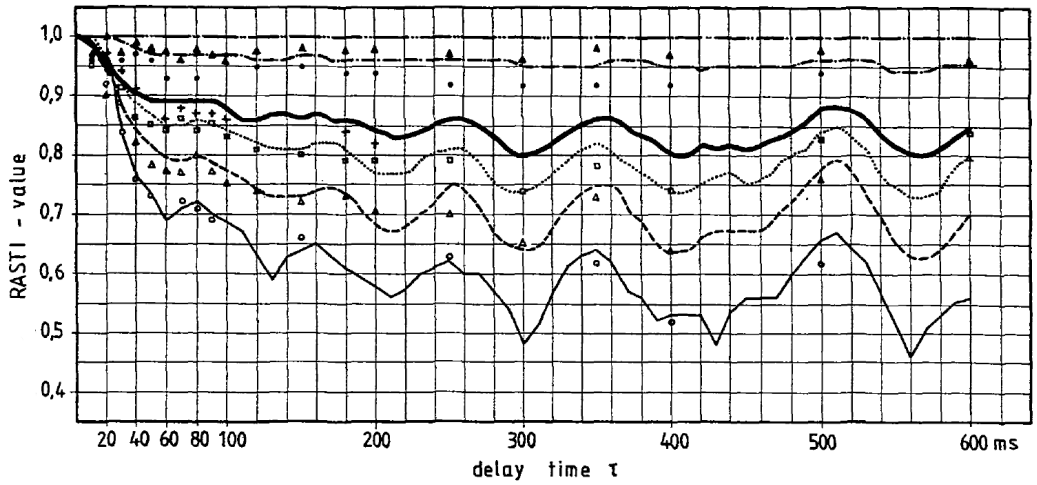


Figure 2.22 Measured RaSTI values vs. delay time and echo level obtained by injecting an electronically delayed echo into the measurement process. The bottom line corresponds to an echo with level equal to the direct signal (Reprinted with permission from [170])

A second experiment by Teuber and Völker [170], employing offset loudspeakers in the outdoors, found similar RaSTI measurement results with regard to echo delay time. Finally, a series of case studies on distributed sound systems in halls found that the addition of electronic delays to compensate for multiple arrivals produced improved RaSTI scores. Examination of the ground plan and sectional drawings of the sound systems tested indicate that the minimum distance offset between main and satellite sound systems (columns and ceiling loudspeakers) was over 15 meters. The nominal travel time of 15 meters being 45 ms, and thus in the range of times corresponding to comb filtering in the RaSTI MTF, the results are unfortunately inconclusive.

Another interesting aspect of these case studies involves array geometry. Each hall studied contained two types of satellite sound systems: Horizontally oriented, uncoupled line source arrays, composed of column loudspeakers hung from the ceiling, and over-head uncoupled planar (2-dimensional linear) arrays of single loudspeakers. Because of this, the incidence of sound from the satellite systems, at each measurement location, was either from the same direction as that of the main system or from above.

Mochimaru [130] experimented with the effects of both single echoes and multiple echoes on the intelligibility of reproduced speech. Comparison of STI measurements with subjective evaluations showed good correlation for echoes ranging from 50 ms to 100 ms. Further studies, employing only STI measurements, indicate that STI values can be reduced significantly (0.05) for a delay time of 40 ms, when multiple echoes are involved. His research on multiple echoes also indicates that delay times as short as 20 ms may have a significant effect on reducing intelligibility.

Clearly, questions remain as to the true contribution of multiple arrivals to, and their effect on, speech intelligibility. From previous experiments it is seen that different methods, setups and research focus can yield different results.

From the literature, it appears generally accepted that multiple arrivals, with delay difference greater than 40 ms, will have an effect on intelligibility. As effects in the region between 0 ms and 40 ms have not been delineated, the current research project focuses on effects in this range. In a conversation with McCarthy regarding variable ranges [125], the value of studying short delay times was discussed. Given that head movement of a seated listener could affect a change of up to a few milliseconds in the difference between multiple arrivals, for this research project it was decided not to study delay times less than 5 ms.

While the work of Haas [67] and others address the issue of the angle of echo incidence in terms of listener distraction, the author has been unable to find previous experiments that address this in terms of intelligibility. As modern loudspeaker arrays are generally oriented in vertical or horizontal configurations, and can take the form of several array types, it would be of interest to know

whether array geometry would affect the degree to which delay times between multiple arrivals degrade intelligibility.

While it is questionable whether objective measurement methods are capable of dealing with the types of phenomena addressed in this text, it is not the author's intention to discredit the use of objective measurements. On the contrary, such methods are extremely useful, and further research can only serve to improve these well established procedures. As such, one goal of the current research project would be to add to the body of knowledge regarding the absolute effects of multiple arrivals, evaluating the correlation between measurements and subjective impression and incorporating effects due to variables such as array orientation. However, a second more practical goal could be to determine the weight, or credence, that should be afforded these potential effects. A sound system designer is often, for a variety of reasons, forced to make compromises that result in deviation from an ideal design. It would be valuable to know the relative contributions of misalignment and array geometry to the reduction of intelligibility, so that such knowledge could be factored into design decisions.

2.3 Binaural Recording

For this research project, it was decided to employ the use of binaural recording to capture the stimuli needed for subjective testing. Binaural recording is a 2-channel recording method that relies on the creation and capture of sound field modifications caused by the human anatomy [92, 132]. Binaural recordings capture inter-aural level differences (ILDs), inter-aural time differences (ITDs) and the shadowing effects and resonances caused by the geometry of the ears, head and (if applicable) torso.

In theory, if all of the auditory cues found in an actual sound field are present in, and delivered to a listener by, a binaural recording, the listener will experience all of the cues found in the original sound field. As such, the listener would be virtually transported to the original sound field [146, 173].

For subjective evaluations of sound system performance in a room, it is necessary for listeners to experience the sound field which exists in the natural

acoustic environment. Due to logistical concerns (switching time between variable treatments, maintaining a blind study, availability of subjects/resources, etc.) it is often impossible to carry out subjective evaluations in-situ. For these situations, binaural recordings, delivered to subjects via headphone display, are a suitable alternative [132, 146, 173].

Binaural recording technology has been implemented for a variety of applications, including evaluation of architectural acoustics and acoustical measurement [22, 146] and intelligibility testing [78, 116, 149, 181]. Additionally, binaural technology has been implemented for the recording of music [63, 66, 92] and simulation (auralization) of the performance of architectural acoustics [2].

2.3.1 Limitations

There are of course limitations to the effectiveness of these methods, including issues with static and dynamic cues and modal conflicts. Additional concerns arise with regard to transduction and coupling errors (see chapter 2.3.3) [173].

While binaural recordings can have the ability to bring a sound field to a listener, the recordings alone can not bring the environment to the listener. The lack of visual and haptic cues can lead to modal conflicts, resulting in listener confusion and a decreased sense of immersion in the virtual environment. While this is a concern for those working with virtual reality and immersive entertainment, it should have negligible impact on studies of speech intelligibility as modal conflicts will have little impact on hearing thresholds or perceived loudness [173].

When evaluating or listening to binaural recordings, a listener's ability to accurately localize sound sources is diminished for several reasons. First, the listener does not have the ability to utilize dynamic auditory cues, such as those created by head motion/rotation, to aid in localization. Also, it has been shown that when comparing localization performance with real sources vs. binaural recordings of real sources, performance with binaural recordings exhibits an

increase in both localization errors for sound sources in the medial plane and errors in the judgment of sound source distance [129, 134, 135]. Given that localization of sound sources in the medial plane does not rely on ILDs or ITDs, and that ILD and ITD are not effective distance cues (for sources beyond 1 meter), the aforementioned errors are most likely caused by differences in the head-related transfer functions between the recording ear and the listener's ear [150, 161]. In other words, the individual listener's ear geometry is different from that of the ear used for recording. The result is the presentation of inaccurate static cues to the listener [134].

An additional complication is variance in the size and shape of the head and torso between listeners. Though the design of head-and-torso simulators has been based on anthropometric measurements of many thousands of people, the resulting dimensions used for construction have been determined by averaging measured data [31, 32, 35]. As with differences in ear geometry, differences in head and torso size between the simulator and real listener will lead to inaccurate presentation of static cues [64].

The question remains, would the use of binaural recording be viable for the research project detailed in this document. The project focuses solely on the intelligibility of speech reproduced by a sound system. As the entire sound reproduction system is nominally in front of the listening position, and reverberation is not a studied variable, the maintenance of sound cues resulting from non-frontal incidence would be of little concern. Sound source distance cues would also be of little concern. As seen in figure 2.23 from Møller et al., most localization errors are due to front-back and above confusions and distance confusions [135]. Very few localization errors (when using a KEMAR manikin, for example) occur between front and front high, or between ±45 degree locations on the forward half of the horizontal plane.

Though care should be taken with regard to transduction and coupling issues, the literature suggests that binaural recording would be acceptable for this research project.

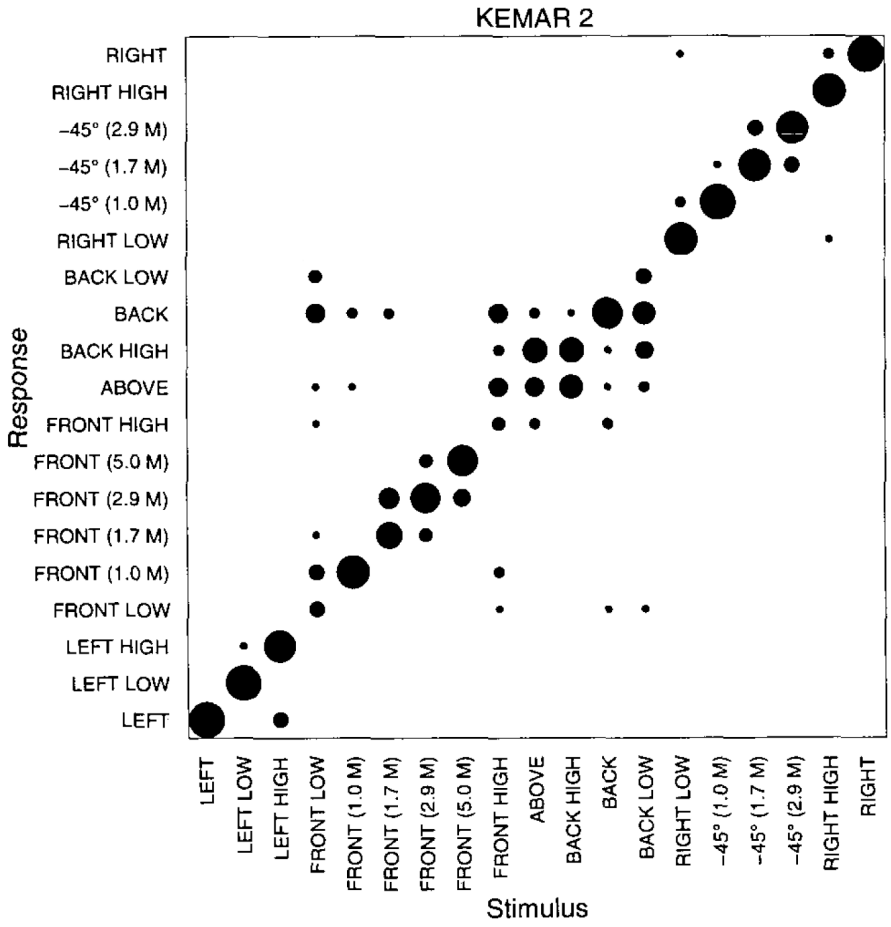


Figure 2.23 Results of localization experiment involving binaural recording (with blocked ear canal) and headphone playback (Reprinted with permission from [135])

2.3.2 Types of Capture Devices

Since the 1960's, a variety of modern binaural recording capture devices have become available for researchers, ranging from spheres [158] to artificial heads [137] to head-and-torso simulators (HATS) [13, 29, 32, 35, 64]. Focusing on HATS, significant differences exist between the various available apparati. Head, torso and pinna dimensions vary between devices, and the recording microphones can be positioned at the opening of the blocked ear canal or at the eardrum location. Variations also exist, for HATS with microphones located at the eardrum, between the types of ear canal simulators used. Several international standards exist to establish consistency between HATS devices (e.g. [81, 85]),

however differences exist even between standards. Burandt et al. [31] point out that the dimensions listed in the standards are not representative of the size of the average person. He also states that headphones often do not stay properly situated on the HATS.

The HATS used for the capture of stimuli for this research project was the Knowles Electronics Manikin for Acoustic Research (KEMAR). The KEMAR HATS conforms to the geometrical and acoustical requirements of the ANSI and ITU recommendations [7, 85, 129]. More detail on the specific KEMAR setup used is included in chapter 3.1.4.

2.3.3 Ear Simulators

A HATS can create differences in ILD, ITD and spectrum between the locations of the ears on the manikin. The addition of pinnae serves to simulate the spectral shadowing and resonance functions of portions of the outer ear. However if recording microphones are located at the eardrum position, ear canal simulators are required to recreate the resonance and acoustical coupling functions of the remainder of the outer ear [33, 159, 173].

A flat-plate coupler, with measurement microphone flush-mounted in a hole in the center, is one way to measure the response of headphones [159, 172]. However, if it is desired to know the response of the headphone at the eardrum, an ear-canal simulator is required.

An ear-canal simulator provides a bridge between the recording microphone (artificial eardrum) and headphone driver. The simulator attempts to replicate the resistive and reactive components of the acoustical impedance of the human outer ear, thus providing a realistic acoustical load for the driver. As with transfer gain of voltage through electrical circuits (e.g. a voltage divider), the impedance of the headphone-driven load affects the transfer of sound pressure from the entrance of the ear canal to the eardrum [132, 133, 177]. Thus, an ear canal simulator that better approximates the loading characteristics of the outer ear will provide more accurate sound pressures at the artificial eardrum. Note that

as the complex component of impedance (reactance) is present in this system, the transfer of sound pressure to the eardrum will be frequency dependent.

Several variations of simple eardrum simulators exist. The metal, single-cavity, two cubic centimeter (2 cc) coupler, mentioned by Bauer et al. [13] approximates the volume of the ear canal and eardrum (1.8 cc). Comparison of the 2 cc coupler to real ears has shown that the 2 cc coupler is highly inaccurate at higher frequencies [151]. Similar to the 2 cc coupler, the 6 cc coupler is also a single-chamber metal cavity. The 6 cc coupler, also called an artificial ear, is often used for air-conduction calibration of headphones used for audiological testing [62]. In the early 1970's, Zwislocki presented a different type of ear canal simulator [184, 185]. The Zwislocki coupler is composed of a central cylinder, similar in length to the human ear canal, and four branches. Each branch functions as one of four parallel impedances, contributing to the total acoustical impedance of the device [32, 33, 152]. The Zwislocki-type ear canal simulator is an improvement over simple, single-cavity simulators, as it more closely emulates measured eardrum impedance [33].

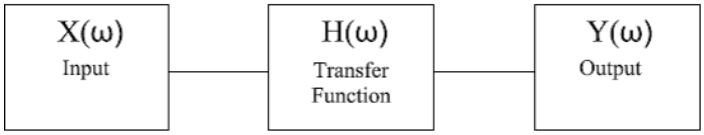
The Zwislocki DB-100 ear canal simulators used in the KEMAR for this research project conform to IEC 711 and ANSI S3.25 [6, 80, 129], and provide a good approximation of the impedance of a real ear canal up to around 4 kHz [177].[6] Also, for calibration of headphone levels, a 6 cc coupler was constructed employing a flat headphone mounting plate.

2.3.4 Headphone Equalization

The goal of binaural recordings in this project would be to create a virtual replica of an original sound field, thus eliminating the need for listeners to be in the actual sound field. When using binaural recordings in this way, it is necessary to consider the effects of the capture and playback systems on the recorded sound. Figure 2.24 details the concept of a transfer function. Sound is input into a

[6] The effective frequency range of IEC 711 compliant ear canal simulators is limited to frequencies below 8 kHz.

system and the resulting output of the system is the frequency-domain product of the effects of the system (the transfer function) with the input.



$$Y(\omega) = H(\omega) * X(\omega)$$

Figure 2.24 Block diagram and equation detailing the relationship between input, output and transfer function

If sound passes sequentially through more than one system, the resulting output is the frequency-domain product of all of the individual transfer functions with the input. Figure 2.25 details this type of compound transfer function for the case of an individual listening to a sound system in a room.[7]

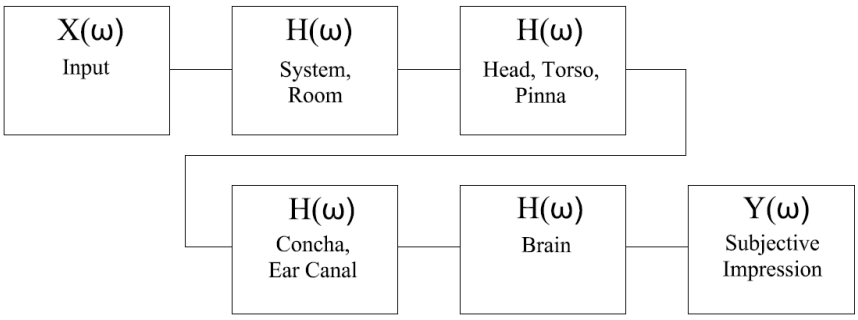


Figure 2.25 Simplified block diagram of compound transfer function of in situ subjective evaluation

For the case of binaural recording and headphone playback, capture and playback are two discrete processes (see figure 2.26) [146]. Comparison reveals that the compound transfer functions of in situ evaluation and evaluation via binaural recording/playback are quite similar. However the binaural method contains several additional individual transfer functions: The concha and ear canal of the manikin, the recording system and the headphone [64, 100]. If these three elements were to be completely removed from the compound transfer

[7] The author acknowledges that these are highly simplified diagrams, and that the effects of the sound system and room, and the neurological and cognitive processes of the auditory mechanism, are vastly more complicated than single transfer functions would suggest.

function of the binaural recording/playback system, the only remaining difference between the binaural and in situ transfer functions would be the difference between the individual transfer functions of the head, torso and pinna of the manikin and the listener. If the transfer functions of the recording microphone, headphone and manikin concha and ear canal are not removed, the result would be inaccuracies in the playback. Also, as there would be both manikin and listener ear canal/concha transfer functions in the chain, the result would be essentially a doubling of the effects of the concha and ear canal [23, 90, 172].

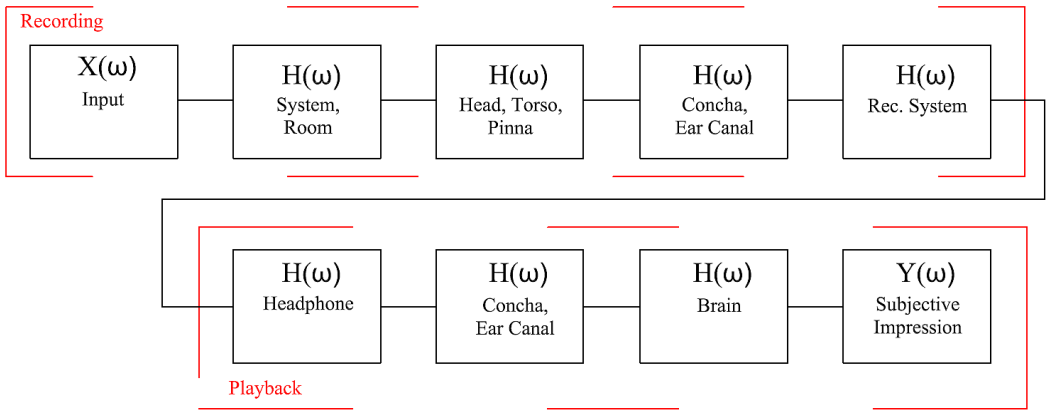


Figure 2.26 Simplified block diagram of compound transfer function of subjective evaluation using binaural recording and headphone playback systems

It is generally agreed that the effects of the recording microphone, headphone and manikin concha and ear canal should be removed from the compound transfer function [100, 132]. The theories behind, and methods for, accomplishing this are however quite different, though most focus on corrective equalization (e.g. [66, 90, 100, 132]).

The approach to equalization is basically divided into compensating for either free- or diffuse-field responses [14, 100]. The free-field response of the manikin corresponds to the magnitude vs. frequency response of the system, for sound sources located directly to the front, as measured in a free field (non-reflective, non-reverberant environment). The diffuse-field response of the manikin is the magnitude vs. frequency response of the system averaged over all possible angles of incidence [100]. As the diffuse-field response is not

directionally dependent, it is generally the preferred approach to equalization [90, 100].

Larcher et al. [100] describe two methods, based on the work of Møller et al. [132, 133], for measuring the response of the recording and playback systems. The decoupled method involves measuring the two responses separately using a reference sound field. The non-decoupled method involves measuring the playback system using the recording system. It is stated that this method is viable as long as the headphones used supply a free-air equivalent coupling to the ear (FEC) – i.e. the headphone does not contribute to acoustical impedance found between the eardrum and the air. Most non-sealed cavity (open) headphones can be considered FEC [100, 132]. Further review of specific examples of corrective equalization in the literature can be found in chapters 4 and 5.

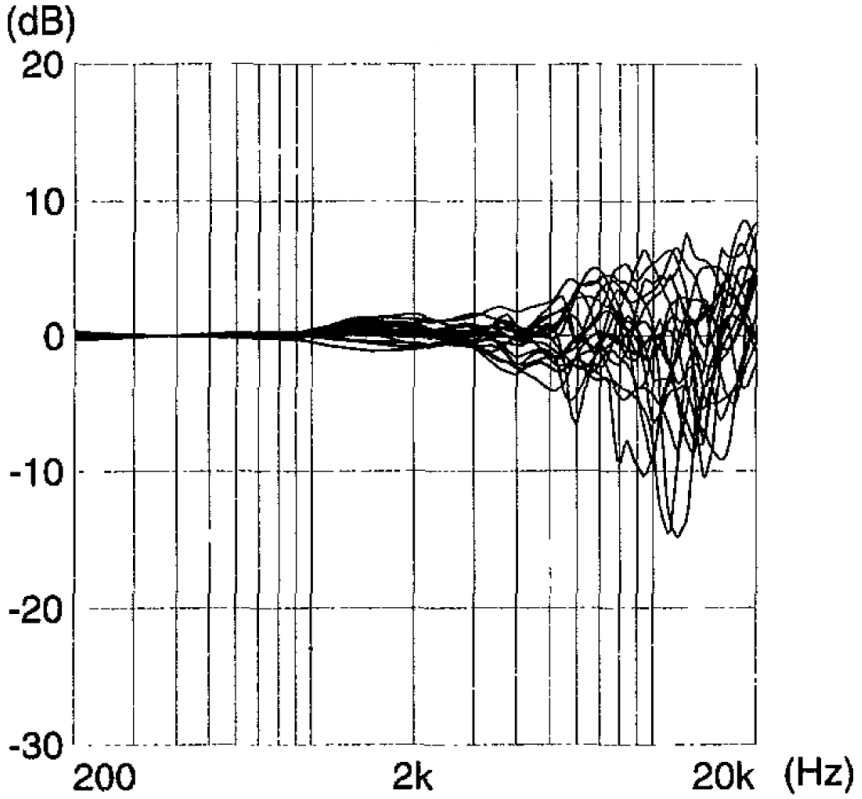


Figure 2.27 Response error for 20 subjects with non-individual equalization filters used for headphone compensational equalization (Reprinted with permission from [134])

While corrective equalization can effectively remove many of the artifacts introduced by the transfer functions of the transducers in the recording/playback systems, a caveat exists with regard to removing the effects of the manikin's outer ear. The results from a study by Møller et al. [134] indicate that differences exist between the responses of different individuals' ears, and that response errors will arise if a single average response is used for equalization. Figure 2.27 shows that this error increases with frequency, reaching +2/−4 dB by 4 kHz, with extremely large differences (+5/−10 dB) noted above 8 kHz. A similar study of the DB-100 Zwislocki coupler by Voss and Allen [177] shows similar trends above 4 kHz. While it would appear possible to determine a single equalization filter for frequencies below 4 kHz, the same can not be said for frequencies above.

3. Stimulus Creation & Acquisition

Before work could begin on the series of studies detailed in this document, it was first necessary to create stimuli for the research subjects to evaluate. It was decided to employ the Modified Rhyme Test (MRT) for these studies and, as such, the MRT word lists would need to be processed through a transmission line – transmitter(s), medium and receiver(s) – in order to create the project stimuli. As specific properties of the transmission line were the focus of these studies, the different combinations of these properties were defined as treatments before the recording process began. Once recording began, it was necessary to have a procedure for manipulating and verifying that the properties of the transmission line matched the desired properties of each treatment.

This chapter details the audio systems and equipment, variable treatments and procedures used to obtain the stimuli which were used in both the pilot and main studies of this project.

3.1 Sound Systems

The transmission line used consisted of an auditorium (medium), a sound measurement system, a sound reproduction system (transmitter) and a sound capture system (receiver). See appendices A, B and C for drawings and schematics.

3.1.1 Corbett Auditorium

The venue used for recording stimuli was Corbett Auditorium. Containing approximately 650 seats in the orchestra level and 200 in the balcony, it is the largest of four performance spaces at the College Conservatory of Music (CCM), University of Cincinnati. The auditorium is rectangular in shape, with a moderate wall splay from the middle to the rear of the house, and has an approximate volume of 16,000 m^3. The volume is reduced to around 9,000 m^3 when its large orchestral shell is in place (as it was for this study), producing a measured average

reverberation time of 1.55 s (see Figure 3.1).[8] The ambient background noise level ranges from a low of 30 dB SPL(A) to a high of 37 dB SPL(A) when the HVAC system is active.

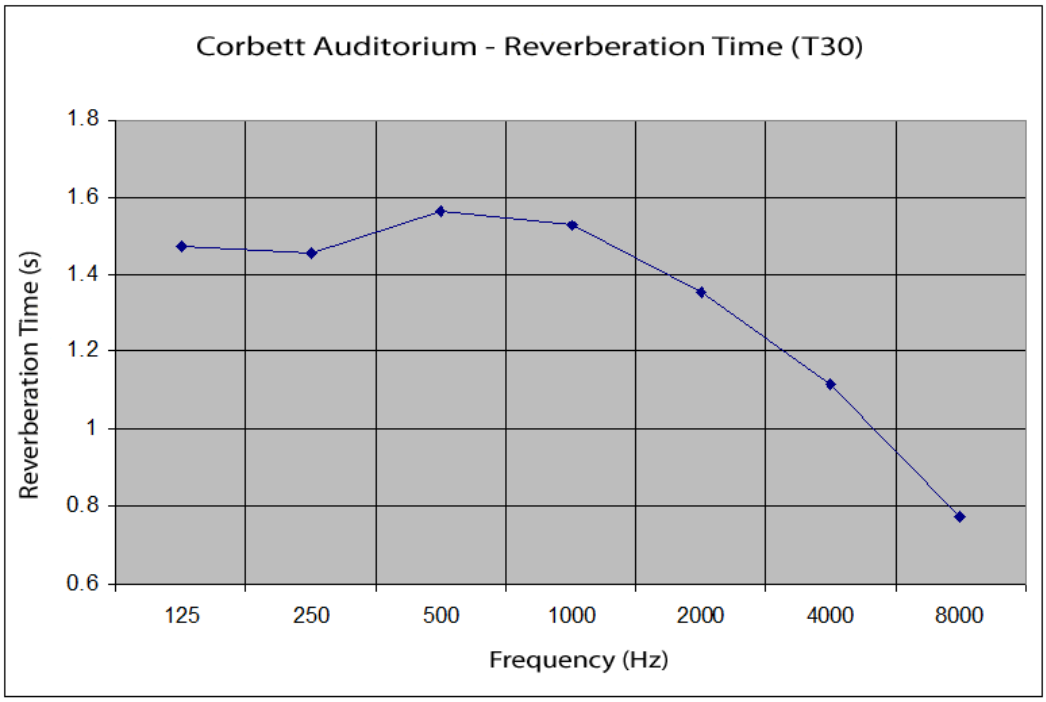


Figure 3.1 Reverberation Time vs. Frequency for Corbett Auditorium (Octave Bands). Note: The steep roll-off above 2 kHz is partially due to the directional characteristics of the sound source used.

The construction of the theatre is primarily of wood, however the side and rear walls contain quadratic-residue diffusers and the flat fascia of the balcony contains absorptive paneling. The floating ceiling is composed of semi-reflective, perforated acoustic panels. The seats, which were in the upright position for all measurements, are composed of wood covered in porous upholstery. The floors

[8] Reverberation time measurements conformed to ISO 3382:1997 [83] for "Low Coverage" conditions, with the exception that a directional sound source (Genelec 1032a) was used. As all of the stimulus reproducing sound sources involved in this study would be nominally directional, and the impulse responses generated would resemble those from a directional source, the Genelec was considered to be an acceptable substitute for an omni-directional sound source in these measurements.

are made of non-porous painted concrete and are covered in industrial-type carpeting in the aisles.

3.1.2 Measurement System

The audio system used for measurement consisted of a Mark of the Unicorn (MOTU) 896HD audio interface, a PC-based laptop computer, the EASERA software package, a measurement microphone (Brüel & Kjaer type 4166 (½" diaphragm) capsule with PCB Piezotronics 426A30 preamplifier) and a microphone power supply (Larson Davis 2221). The measurement system was level-calibrated in EASERA using a Larson Davis CAL-200 acoustic calibrator (IEC 60942-1:2003 Class 1 compliant) and a true RMS digital multi-meter (Fluke 87-III). Verification of calibrated levels was performed at the end of each day.

Audio signals in this system, as well as the other two audio systems, were routed through a Yamaha PM5D mixing console. The purpose of the mixing console was to provide a single programmable device in which to control signal gains. The console also allowed for the automation of signal routing between the three audio systems when switching back and forth between measuring and recording.

The stimulus used for measurement was a pink-weighted Maximum-Length Sequence (MLS) of order 18 (262,143 samples). At the sampling rate of 44.1 kHz, each sequence had a duration of 5.94 seconds. This length was sufficient to prevent time aliasing, as it was longer than the total sound system's impulse response [148].

All measurements were made using the dual-channel Fast Fourier Transform capabilities of EASERA. Initial calibration of the various sound systems was performed with EASERA in "live-mode" using flat-weighting. Live-mode was set such that the FFT size was 32,768 (743 ms in duration, resolution of 1.3 Hz) using an integration time of 3 seconds (4 averages). "Measure-mode" was set to time and frequency resolutions of 22.7 μs and 168 mHz, using 1 pre-send and 4 averages, and was used for loudspeaker equalization, level and time alignment. No psophometric weighting filters were used to

compensate for perceived loudness during equalization (i.e. flat weighting). However, given that the stimulus recordings (and later the stimulus evaluations) would have a sound pressure level of 77 to 83 dB, it was decided to use C-weighting for level calibration [126].

3.1.3 Reproduction Sound System

The sound reproduction system was essentially a 3-channel Left-Center-Right (LCR) sound system: Speech signals were reproduced by the Center channel while noise signals were reproduced by the Left and Right channels. The system (see Appendices A, B and C for drawings) consisted of a main loudspeaker, two front-fill loudspeakers and, on each side of the stage, a mid/high- and low-frequency loudspeaker (2-channel, 3-way noise subsystem). The main loudspeaker was a Meyer UPA-1p, which was fed from a Midas XL-88 matrix mixer to provide a variable gain stage at the loudspeaker. The front fill loudspeakers were Meyer UPM-1's, each processed by a Meyer P-1A and individually amplified by a Crown Macro-Tech 1200. Each channel of the noise subsystem contained one Meyer UPJ-1p and one JBL 4648 woofer enclosure. Each JBL 4648 was powered by an individual channel from a Crown Macro-Tech 2400. All loudspeakers in the reproduction system were controlled by one of two multifunction loudspeaker control processors (BSS FDS-355 Omni-Drive).

The main loudspeaker was located at the center of the proscenium arch 8.4 meters above the stage, 0.9 meters downstage from the measurement origin and 11.1 meters (direct line) from the center listening position. The main loudspeaker was the only loudspeaker that was elevated above the stage floor. The center front-fill loudspeaker was located on the center line of the stage, 3 meters downstage from origin and 5.8 meters upstage from the center listening position. The right front fill was located 4.5 meters house-right of the center line, 2.6 meters downstage from the origin and with an angle of 8 degrees from center. The loudspeakers contained in the noise subsystem were located on the horizontal (up/downstage) origin line and oriented symmetrically about the center line. The woofer enclosures were located 5.9 meters from center. The crossover frequency

for the noise subsystem was set to 100 Hz, which eliminated the need to angle the woofer enclosures. The mid-high-frequency loudspeakers were located 5.1 meters from center with a 30 degree angle towards center, yielding a 60 degree subtended angle at the center listening position.

3.1.3.1 Loudspeaker Calibration

Calibration (initial optimization) of the reproduction system entailed loudspeaker equalization, time alignment and level alignment. Though conversations with engineers at the manufacturer indicated that it would be unlikely in this case, it is the author's experience that some loudspeaker systems have the potential to exhibit differing performance and room-interface reaction at different drive levels [157]. As such, an initial (rough) level alignment was performed using a pink-weighted MLS pseudo-random noise and a hand-held sound level meter (IEC 651 Type II) prior to equalization [79].

The first step in the equalization of each loudspeaker was the observation of the 24th-octave resolution response of the loudspeaker. The next step was correction of spectral tilt (as viewed with 1/3rd-octave resolution), followed by less broad adjustments (1/6th-octave resolution) and completing with fine, cut-only adjustments (12th- and 24th-octave resolution). All equalization was done whilst bearing in mind the 24th-octave resolution response curve and the potential for errors in procedure and decisions due to graphical averaging and reduced resolution. Note: Equalization of the noise subsystem was performed after crossovers had been set in the BSS Omni-Drive (100 Hz, 12 dB/oct Bessel) and phase aligned.

Time alignment was carried out using transfer function measurements viewed as time-domain impulse response measurements and verified using the squared energy-time curve (ETC). The center UPM was aligned to the Meyer UPA – the most distant loudspeaker. The house-right UPM was aligned to the center UPM in the same manner but using a different measurement location (see below). The impulse response of a loudspeaker was measured and the arrival time of the impulse was noted. Electronic delay was added to loudspeaker signals

in the loudspeaker control processors such that all impulse responses were aligned to the UPA. Time alignment was not performed on the noise subsystem aside from phase alignment of the woofers to the mid-high loudspeakers.

Prior to level alignment, all input channel faders on the console were set at unity and all power amplifier input attenuators (where applicable) were set to −0 dB. The level of each loudspeaker was adjusted only at the outputs of the mixing console. The level was set such that a MLS signal delivered to the input of the mixing console, with pre-fader level of −12 dB FS in the channel strip, would produce a signal that measured 80 dB(C) SPL at the measurement position. Verification of level calibration was performed at the end of each day.

The Meyer UPA and center UPM (vertical array configuration) were calibrated using the measurement system's microphone in the manikin-center position (see appendices A, B and C for details). The center and house-right Meyer UPMs (horizontal array configuration) were equalized and initially level and time aligned with the measurement microphone located on-axis with each loudspeaker. Calibration of the left and right channels (noise reproduction) was carried out with the measurement microphone located at the manikin-center position.

3.1.3.2 Speech Level Calibration

For the studies in this project, it will be of interest to know the ratio of levels between speech signals and noise signals (signal-to-noise ratio). In order to calculate these figures, it will be necessary to know both the level of the noise distracter and speech signals.

Procedures for measuring sound pressure level in a sound field are well documented (e.g. [126]). Psophometric filters can be used to produce measurement results which more closely relate to perceived loudness. Measurement integration times can be adjusted so that results represent instantaneous or continuous (peak vs. average) sound pressure levels, and many levels of averaging in between. Several problems arise, however, when attempting to measure the level of running speech [118].

Figure 3.2 shows a fairly typical waveform for a piece of modern rock-and-roll music.[9] Upon visual inspection, a clear difference is noticed between the peak and average levels. In contrast, figure 3.3 shows the waveform of running speech (MRT word list A). While there is still a clear difference between the peak and average levels of the speech signal, there is also a good deal of nominally silent space between utterances. These silent spaces are a result of the pace at which the talker delivered the list of words.

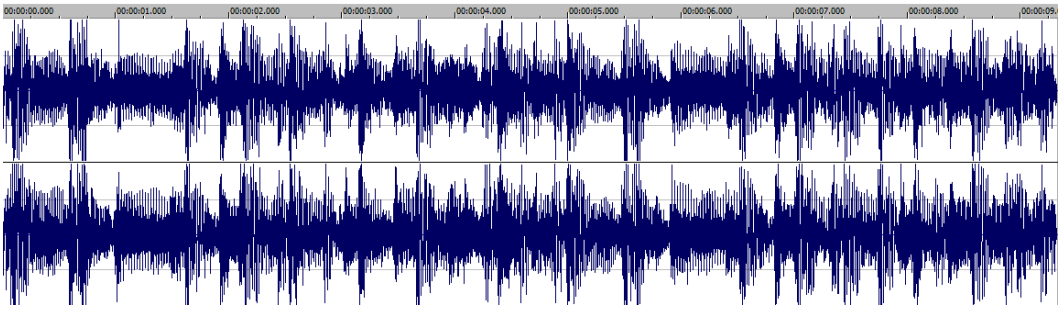


Figure 3.2 Waveform of a normalized 9 sec clip of a modern rock song

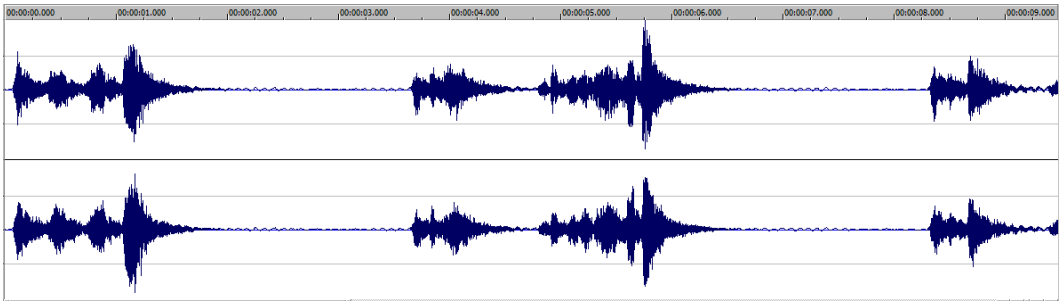


Figure 3.3 Waveform of a normalized 9 sec clip of captured MRT word list A (viewed with same amplitude and time resolution as figure 3.2).

Measurement of this signal using a fast integration time (small averaging window) would produce widely varying numbers depending on what part of the signal falls within the window. If this signal were to be measured with a slow integration time (large window), the results would have less variance, but the averaging function would integrate the silent spaces. This would have the effect

[9] This excerpt is taken from the intro section of "Sunday Morning After" by Amanda Marshall, from the album *Everybody's Got a Story* (as used in the Technical Ear Training course at McGill University).

of reducing the level of the measured signal – like adding a large number of zeros to the list of numbers to be averaged.

As the end goal is to compare the level of a noise signal with that of the speech signal, the silence between speech signals is not of interest and should somehow be removed from the measurement. This can be accomplished in one of two ways: 1) Measure only while speech signal is present or 2) remove the silence from the signal prior to measurement.

In figure 3.4, we see the waveform of the same MRT list, but with the silent spaces removed through editing (referred to as the condensed word list). Measurement of this signal should produce a result which indicates the level of the actual speech signals, rather than the level of the overall signal. Measurement of this modified signal is in accordance with ITU-T P.56, wherein the level of speech is measured including the brief low- to zero-levels (down to at least 16 dB below the average speech level), but excluding noise that is not part of the speech signal [84].

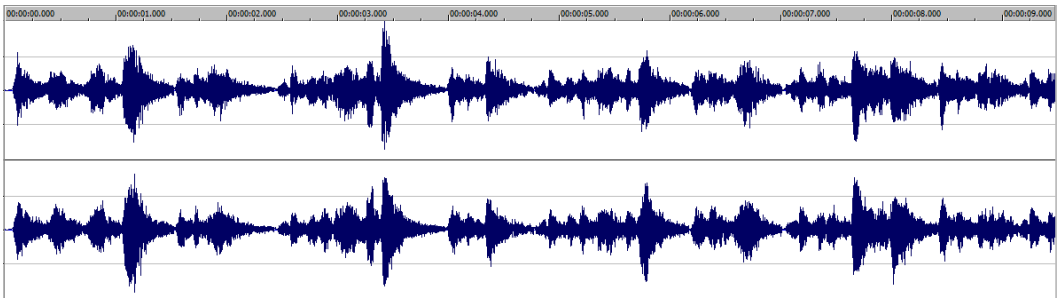


Figure 3.4 Waveform of a normalized 9 sec clip of captured condensed MRT word list A (viewed with same amplitude and time resolution as figure 3.2).

Table 3.1 details four of the methods used to measure and calibrate the level of speech signals. These methods differ in integration time (FFT size), the number of measurements that are averaged together to produce the result and speech signal used (stimulus). They also differ in measurement method: Methods 3 and 4 rely solely on the measurement equipment, while methods 1 and 2 required the author to "only measure while the speech signal is present."

The results of methods 1 and 2 have the greatest degree of variance. This was to be expected considering that the unaltered word list was used. Also,

considering that the condensed word list was used for measurement methods 3 and 4, it is not surprising that the results of method 3 indicate levels which are slightly higher. The results from method 4 however indicate that this increase in level disappears with a sufficient number of averages.

It is clear that the different methods produce different results, both in terms of overall level and range of variance. What is interesting is that the mid points in the range of results for each method are quite close.

Parameters	**Method 1**	**Method 2**	**Method 3**	**Method 4**
Measurement Equipment	EASERA & Researcher	EASERA & Researcher	EASERA	EASERA
FFT Size	8192 (186 ms)	32768 (743 ms)	32768 (743 ms)	32768 (743 ms)
Number of Averages	2 (372 ms)	1 (743 ms)	4 (3 sec)	8 (5.9 sec)
Stimulus	MRT List A	MRT List A	Condensed & Looped	Condensed & Looped
Result (dB C-weighted)	78.6-81.6	78.9-80.8	80.5-81.7	79.5-80.6
Mid Point	80.1	79.9	81.1	80.1

Table 3.1 Methods used to measure the level of running speech with results.

A fifth method of measurement was attempted using the condensed word list. A hand-held SPL meter was set for C-weighting, 1-second integration time and 50 dB–100 dB range. The result from this method showed levels ranging from 77.1 dB–82.3 dB (79.7 dB midpoint).

Results from these measurement methods are all similar, though not identical. As such, even after level calibration, one can only determine that the average level of the speech signal is approximately 80 dB (C-weighted). Signal-to-noise ratio (SNR) in future studies must therefore be considered approximate as well.

3.1.4 Capture Sound System

Binaural sound capture was accomplished by use of a head and torso simulator: The Knowles Electronics Manikin for Acoustic Research (KEMAR)

[32]. The KEMAR unit used for these recordings was fitted with two neck rings (two rings being mean shoulder-to-ear height for males, mean height for a female is zero rings) and the "two-sigma-ears" (auricles). The ears were further fitted with Industrial Research Products (IRP) model DB-050 ear canal extensions, connecting to IRP model DB-100 Zwislocki-style occluded ear simulators. The transducers used were Brüel & Kjaer type 4165 (½" diaphragm) microphone capsules, which were connected to a Brüel & Kjaer type 2807 power supply via type 2660 preamplifiers and type UA-0122 microphone adaptors (see appendix A for schematic). Figure 3.5 shows the location and orientation of transducers, couplers and adaptors inside of the KEMAR head.

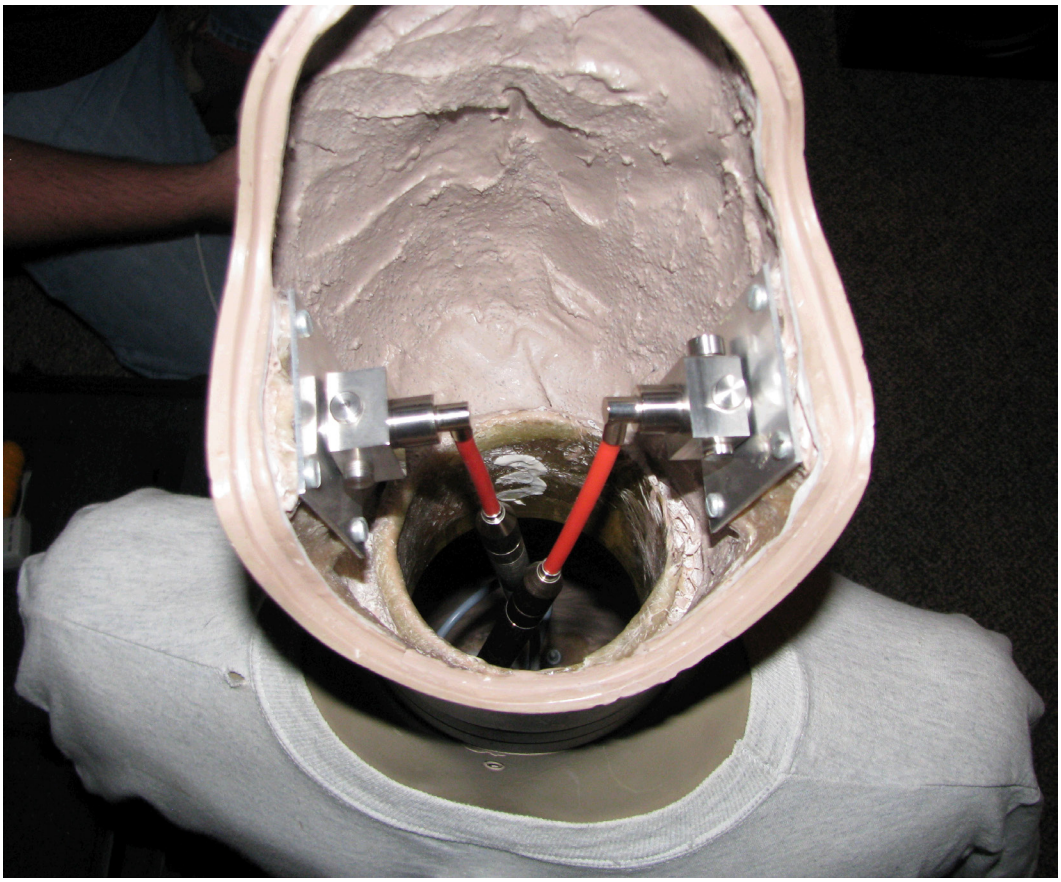


Figure 3.5 View of Zwislocki-style couplers, microphone elements and adaptors inside of the KEMAR head.

Microphone signals were sent to the Yamaha PM5D mixing console for fine level adjustment and A/D conversion. Signals were then were routed to a

MOTU 896HD MKII for connection to a Macintosh MacBook and were recorded using Cubase LE 4. All recordings were captured at a resolution of 96 kHz, 24 bit.

Prior to physical mounting in the KEMAR unit, the level of each microphone was calibrated (re: 94 dB SPL @ 1 kHz) using a Larson Davis CAL-200 acoustic calibrator and the measurement system in "live mode" (detailed in chapter 3.1.2). Level adjustment was performed at input channel head amplifiers in the mixing console. This calibration was repeated every time the KEMAR unit was moved or if microphone signals were interrupted (power down of power supply). Verification of calibrated levels was performed at the end of every day.

The KEMAR manikin was located in a different location for each of the two different loudspeaker array geometries studied (see appendices B and C for plan and section views). The manikin-center position was located on the theatre's center line, 8.7 m downstage from the measurement origin and was used for recording the vertical array configuration. The manikin mid-right position, used to record the horizontal array configuration, was located 2.9 m house-right of center and 8.6 m downstage from the origin. When in the mid-right position, the manikin was angled slightly to face halfway between the two front fill loudspeakers.

3.2 Stimuli

As mentioned, the Modified Rhyme Test (MRT) was selected for this series of studies. Three of the six word lists (lists A, D and F) were used, employing one trained, native English-speaking, male talker with no discernable accent. Each of the test stimuli (target words) was embedded in a carrier sentence.

The original recordings were purchased through a distribution company, but appear to have been made in a non-reverberant vocal recording booth.[10] It should be noted that in the set of 50 six-word ensembles that were used in these studies, several ensembles differ from those detailed in the ANSI standard [5].

[10] The raw MRT recordings were purchased through Cosmos Distributing, 4744 Westridge Dr, Kelowna, B.C. Canada V1W 1A1.

The three MRT lists were delivered through the reproduction sound system under a variety of variable combination treatments.

3.2.1 Selection of Variables & Treatments

Without knowing where the results from the series of studies would lead, and considering the difficulty of coordinating equipment and facilities, it was decided to record stimuli using a wide variety of variable treatments. The experimental variables in question for these studies would be 1) delay time between the arrivals from two loudspeakers, 2) level offset between the multiple arrivals and 3) array geometry (vertical vs. horizontal array). The three MRT lists were recorded for each of the variable combinations (treatments) in the 6×2×2 matrix formed by delay time, level offset and array geometry. The fourth experimental variable, Signal-to-Noise Ratio (SNR), would be synthesized (via electronic mixing) at a later date (see chapter 5.2).

Delay Time (ms)	0	5	10	20	30	40	
Level Offset (dB SPL)	0	6					
Array Geometry	Vert	Hor					
Noise Level (dB SPL)	≈ 30	68	71	74	77	80	83

Table 3.2 System optimization variable values used during the capture process

3.2.2 More on Array Types

One possible approach toward studying the variable array type would be to isolate all of the associated variables (e.g. on- vs. off-axis listening, monaural vs. binaural listening and measurement/equalization location). However, as mentioned in chapters 1.2 and 2.2.1, the ecological/organic approach of this research project lead to the inclusion of two commonly employed loudspeaker array types. The result of this approach is that array geometry becomes a complex, or compound variable – an amalgam of several variables.

On one hand, one could view array geometry as an insufficiently controlled or nuisance variable. Alternatively, one could view this as a study of

real-world scenarios, intended to determine whether the use of different array types necessitates different optimization approaches. This question has not been previously addressed. As such, and considering that the identification of differences could guide future (more controlled) studies, it was decided that the proposed approach would be most appropriate for this series of studies. Still, the specific differences between these array types require elucidation.

The first, and most obvious, difference between the two array types is that they are oriented in different planes. The received signals from the horizontal array type will differ at each ear, while the signals from the loudspeakers in the vertical array type will not differ between the ears. This amounts to a difference between monaural and binaural listening methods. Given that objective measurements (STI) will be used for comparison with subjective results (see chapter 9), and that the STI measurement method is inherently monaural, it could be possible to identify the effects of listening method.

The second difference between the array types lies in the fact that the vertical array type offers on-axis listening of both loudspeakers, while the listener is off-axis from both loudspeakers for the horizontal array type. Figure 3.6 shows the measured on-axis magnitude vs. frequency response of the three loudspeakers used to create the two arrays. While the responses are relatively flat for the on-axis conditions, the off-axis response of the front fill loudspeakers, as seen in figure 3.7, are not. Thus, even if loudspeaker arrivals are time aligned (delay time = 0 ms), a comb filter will still be present in the summed response of the horizontal array type.

The third difference builds on the on- vs. off-axis response issue. In typical deployment scenarios, loudspeaker response is equalized at a point that is either on a loudspeaker's axis or in the center of the desired coverage area [124]. Also, time alignment is generally performed at the point of maximum volatility – the point where equal level is received from each loudspeaker. For the vertical (point-destination) array type, it is conceivable that these locations will coincide. For the horizontal (point-source) array type, however, the point of equal level will generally not be in the center of either loudspeaker's coverage area and will be

inherently off-axis from both loudspeakers. Thus, while the comb filters observed in the off-axis response of the front fills could be corrected via equalization, this would not be done in practical application. As such, the equalization of all loudspeakers was performed using on-axis measurements.

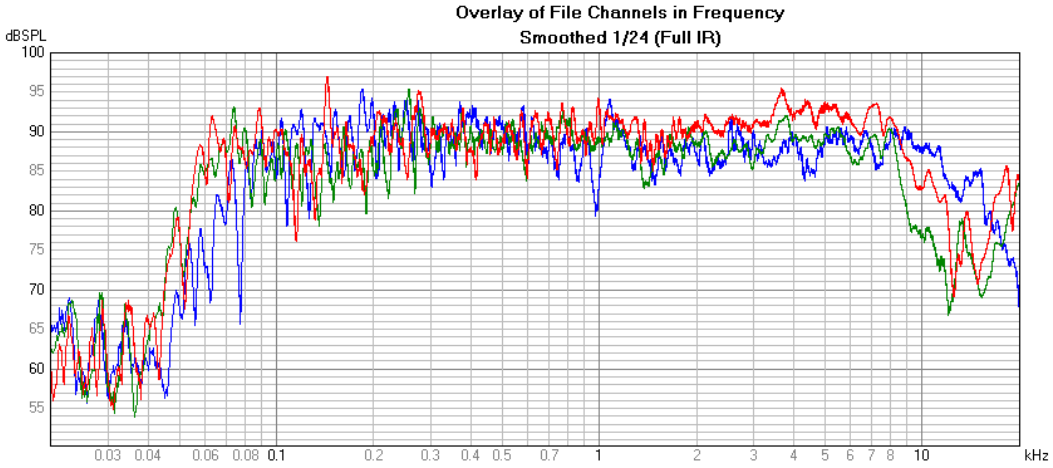


Figure 3.6 Magnitude vs. frequency response of the UPA and two UPM loudspeakers, as measured on the axis of each loudspeaker

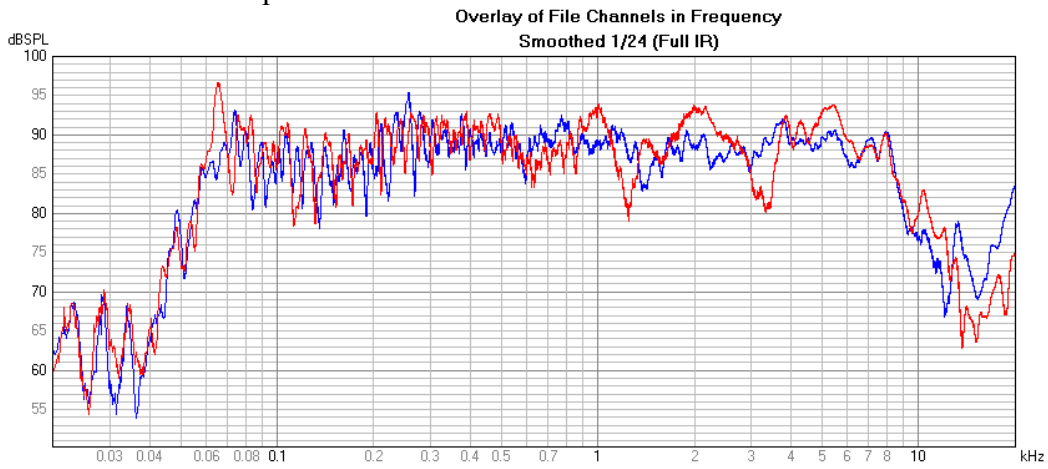


Figure 3.7 Magnitude vs. frequency response of the house-center UPM loudspeaker. The blue trace is measured on the loudspeaker's axis; the red trace is measured at the manikin mid-right position (30° off-axis).

3.3 Procedures

For each treatment condition, the reproduction sound system was recalibrated for the appropriate time and level alignment/misalignment. Once the recalibration was measured and data recorded, the three MRT word lists were reproduced and recorded using the capture sound system. Vertical and horizontal array geometries were recorded on separate days to minimize the need to recalibrate the binaural recording system. For vertical geometry, the main loudspeaker was held constant while the temporal and level properties of the center front fill loudspeaker were manipulated. For the horizontal geometry, it was the center front fill that was held constant and the house-right front fill that was adjusted.

3.3.1 Measurement, Verification and Capture

Each treatment used two of the speech reproduction loudspeakers. For each recalibration, both loudspeakers were measured separately to verify proper values for time and level offsets. The loudspeakers were then measured together; first with no additional noise, then six additional times with different levels of noise supplied by the noise reproduction sound system. While noise would not be added to the MRT recordings until later (see chapter 5.2), for analysis and comparison purposes it was desired to measure certain parameters (STI, RaSTI, etc.) of the reproduction sound system under various noise conditions for each treatment condition.

For variable treatments which included the vertical array geometry, the measurement microphone was located at the manikin-center position. For those which included the horizontal geometry, the measurement microphone was located at the manikin mid-right location.

After recalibration for each treatment condition, the three MRT lists (A, D & F) were recorded using the KEMAR system. Once all MRT recording was complete, the KEMAR system was used to record pink-weighted noise at six

Figure 3.8 KEMAR and measurement microphone in manikin-center recording position

different sound pressure levels, as delivered by the noise reproduction system and recorded at each of the manikin locations (12 noise recordings total). If unwanted noises were detected during any of the recordings (doors, airplanes, piano movers, etc.), said recording was restarted.

4. Headphone Suitability Study

The listening tests conducted in the pilot and main studies would use a head-mounted auditory display to deliver the stimuli to subjects. This raised the obvious question of which specific auditory display to use. Toole [173] suggests that the preferred solution is to use earphones which are inserted into the ear canal. This type of device generally offers greater response consistency as 1) it altogether avoids reflections caused by the concha and pinna, and 2) because it is easier to achieve consistent device placement on/in the head, thus more consistent coupling of the drivers to the eardrum [173].

While in-ear devices are preferable, for this study it was decided to focus on circum-aural and supra-aural headphones for the sake of practicality, comfort, equipment availability and hygiene. The goal therefore was to identify the pair of available headphones which provide the most consistent mounting on a listener's head and eardrum, and which also have a reasonably flat magnitude versus frequency response.

4.1 Headphones Tested

The headphones studied ranged from professional audio to audiophile quality and varied in mounting type (circum-aural, supra-aural) and diaphragm ventilation (open, closed). The four pairs of headphones used were the Sony MDR-7506 (pro, supra, closed), Sony MDR-V600 (pro, circum, closed), Grado RS-1 (audiophile, supra, open) and Sennheiser HD-650 (audiophile, circum, open).

4.2 Equipment & Procedure

The measurement apparatus used was identical to the measurement system detailed in chapter 3.1.2 except that the measurement microphone, preamplifier and power supply were replaced with a KEMAR manikin (see chapter 3.1.4 for

KEMAR setup and specification). Measurements were performed in the Sound Design Studio at CCM.

Measurements were performed using the non-decoupled measurement method, as detailed by Larcher et al., and used by Minnaar et al. [100, 129]. Headphones were placed on the KEMAR manikin and measured using the left ear of KEMAR. Headphones were removed, remounted on the manikin and re-measured for a total of ten measurements each for the HD-650 and RS-1 and five measurements each for the MDR-V600 and MDR-7506.

4.3 Results

The results from these measurements can be seen in figures 4.1 through 4.4. The figures show the results for each headphone, with repeated measurements superimposed, on a graph of magnitude versus frequency using 1/24th-octave smoothing.

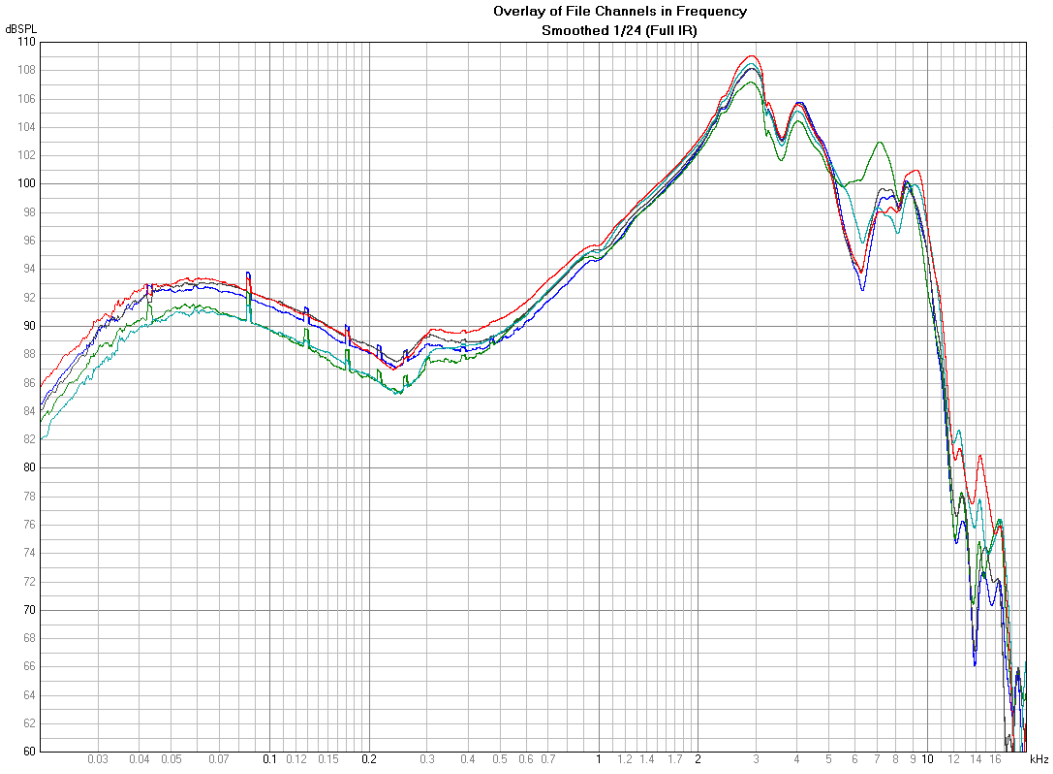


Figure 4.1 Magnitude vs. Frequency Response of Sony MDR-7506

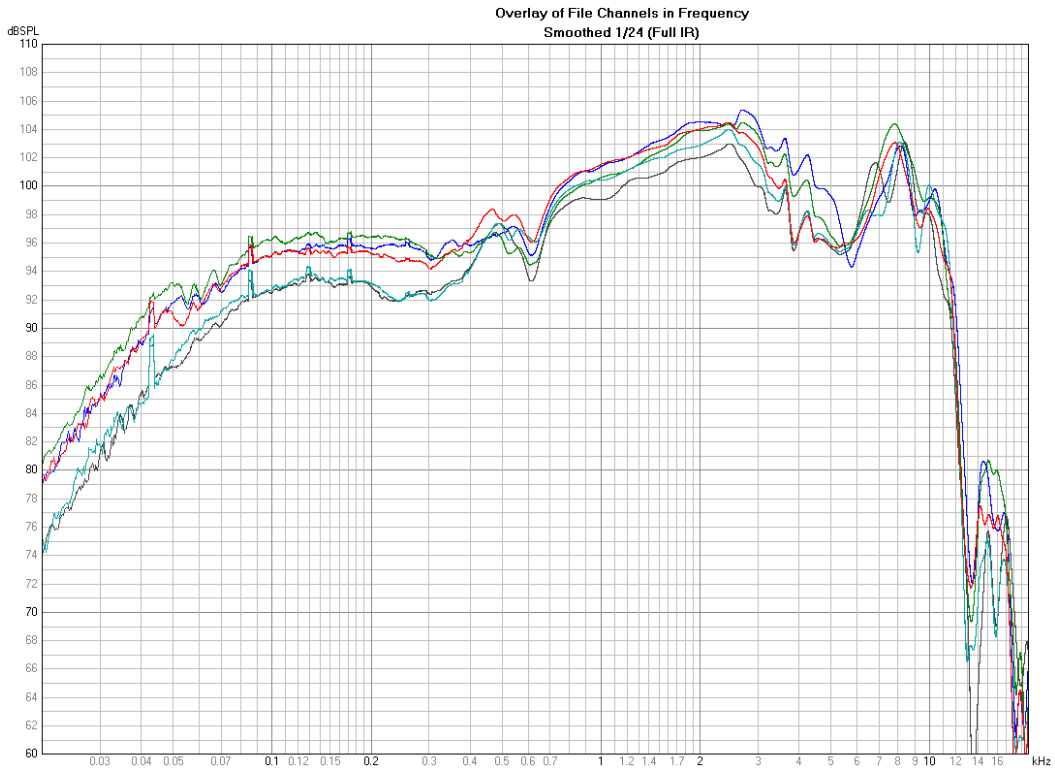


Figure 4.2 Magnitude vs. Frequency Response of Sony MDR-V600

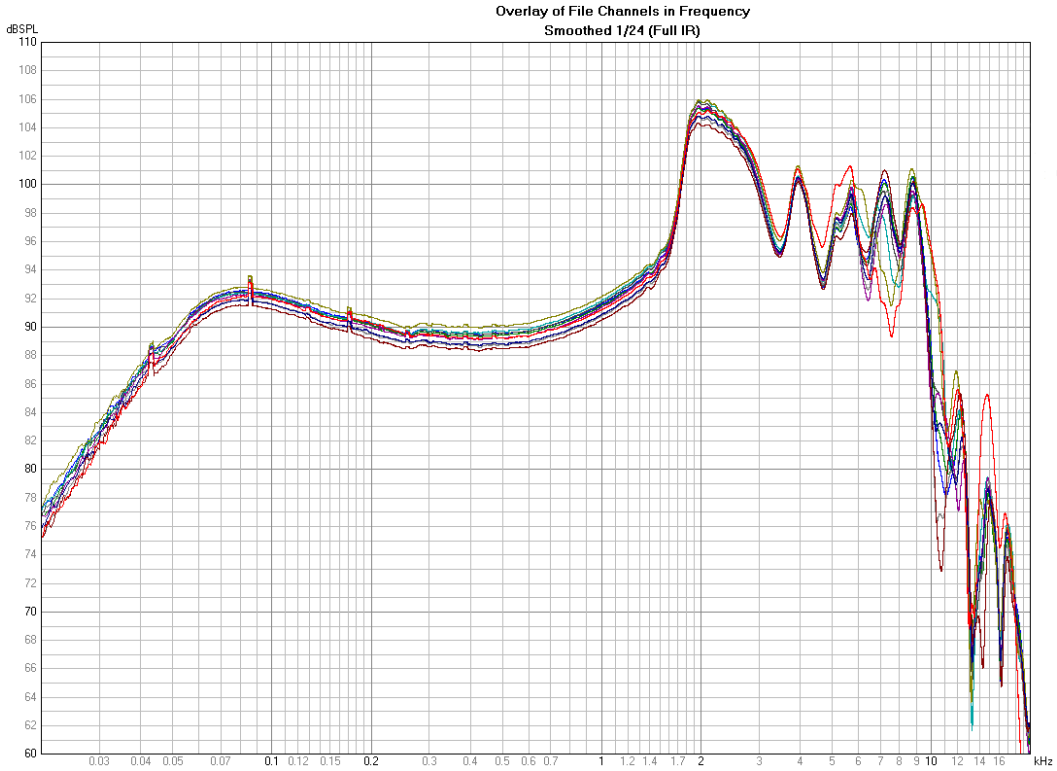


Figure 4.3 Magnitude vs. Frequency Response of Grado RS-1

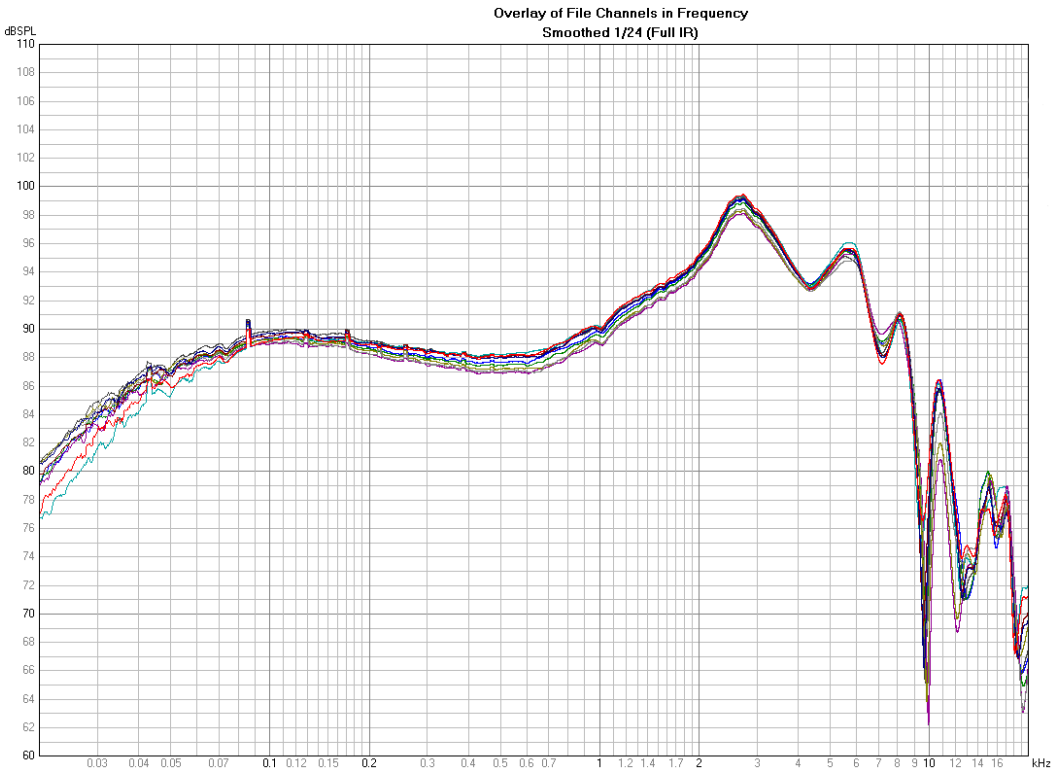


Figure 4.4 Magnitude vs. Frequency Response of Sennheiser HD-650

One can see from these graphs that several of these devices do indeed fail to be consistently mounted on the manikin's head.[11] In the graph for the supra-aural MDR-7506 (figure 4.1), for example, there is variance in low-frequency response indicating variation in the quality of seal between the headphone and the pinna. While fairly consistent in the middle frequencies, note the variance due to reflections in the outer ear at and above 2 kHz. The MDR-V600 (figure 4.2) produced similar results though the degree of variance was much higher. The graphs for these two headphones were evaluated during the measurement process and, after only five repeated measurements, it was concluded that these two headphones would be inadequate.

In figure 4.3, one can see that the response of the RS-1 is consistent (±1 dB) up to approximately 4.7 kHz, indicating an acceptable seal between the headphone and the manikin's pinna. However, the degree of variance above this

[11] The harmonic noise found in the low-frequency region of each graph was caused by the power supply of the measurement laptop.

point is an indication of changes in the pattern of reflections in the outer ear. This suggests inconsistency in the placement of the headphone, likely due to the fact that the headphone is supra-aural.

While there are some irregularities at the very low and very high frequencies, the response curves for the HD-650 (figure 4.4) are consistent (±1 dB) in the range from 40 Hz–8 kHz, which is the functional range of the Zwislocki-type ear canal simulator [80]. This indicates that this circum-aural headphone is making a consistent seal with the manikin's head. It also indicates that though there is evidence of some degree of change in the pattern of reflections in the outer ear, these changes are not affecting the frequency range (125 Hz–8 kHz) that impacts speech intelligibility [110].

4.4 Discussion

At the beginning of this study, the criteria for an acceptable headphone were stated:

1) Consistent reflection pattern in the outer ear
2) Consistent coupling of the headphone to the head and driver to the eardrum
3) Reasonably flat magnitude versus frequency response

The two pro-audio quality headphones (Sony MDR-7506 and V600) both fail to meet the first two criteria. The RS-1 meets the second but fails the first. The HD-650 was the only headphone to meet both of the first two criteria (for 40 Hz–10 kHz).

Though the various headphones produced varying degrees of consistency, none of the headphones tested produced a response curve which could be considered flat. Remembering that measurements were made at the eardrum of the manikin, it would be useful to consider the transfer function of the path between the headphone driver and the eardrum of the manikin.

Figure 4.5 shows the response of a KEMAR manikin with a Zwislocki-type occluded ear canal simulator [90, 185]. When comparing this response with

that of the HD-650 one notes that both contain a prominent narrow peak at approximately 2.7 kHz and a wider peak centered at roughly 4 kHz. As noted by Klepko [92], the peak between 2 kHz and 3 kHz is a result of resonance in the concha and is not directionally dependent. This frequency response trend is also reported in the ITU-T P.58 recommendation [85] regarding free and diffuse field response tolerances for head-and-torso simulators.

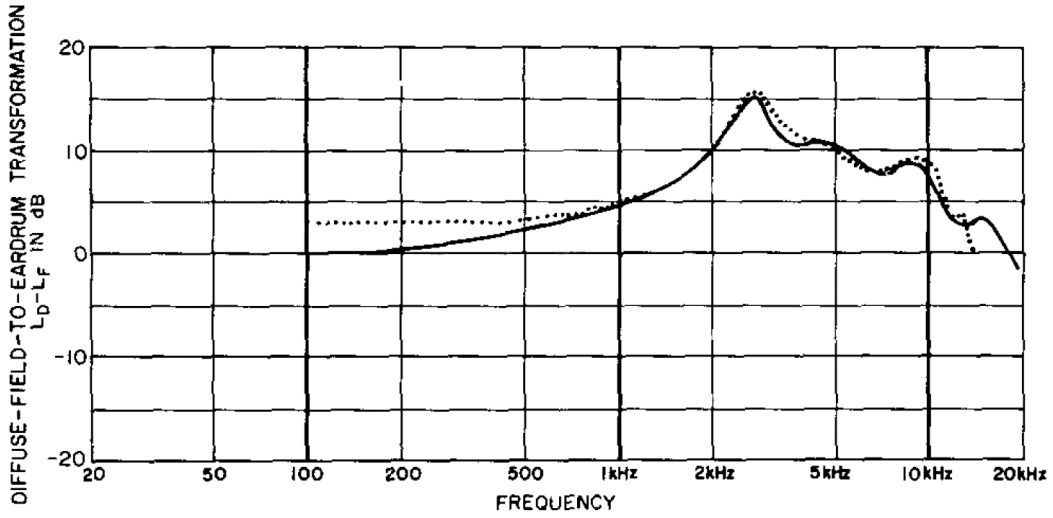


Figure 4.5 Random-incidence eardrum-pressure response of KEMAR manikin (Reprinted with permission from [90])

By visually removing the effects of the manikin, the resulting response for the headphone would contain 1) a slight elevation in the low frequencies, 2) high-frequency roll-off starting at around 9 kHz, and 3) the effects of outer ear reflections. Given that the headphone response will show evidence of outer ear reflections for any headphone-ear coupling, and that the pattern of reflections will change from ear to ear, it was decided to focus on the components of the headphone response that are not reflection-based.

Figure 4.6 shows the response for the HD-650 with 1/3rd-octave smoothing. This level of smoothing, though inappropriate for many applications, is used here to remove many of the effects of outer ear reflections through averaging. What remains is more a spectral tilt than a response, but it allows for better inspection of the point of high-frequency roll-off. Again, visually removing the effect of KEMAR from the response, one can see that the response

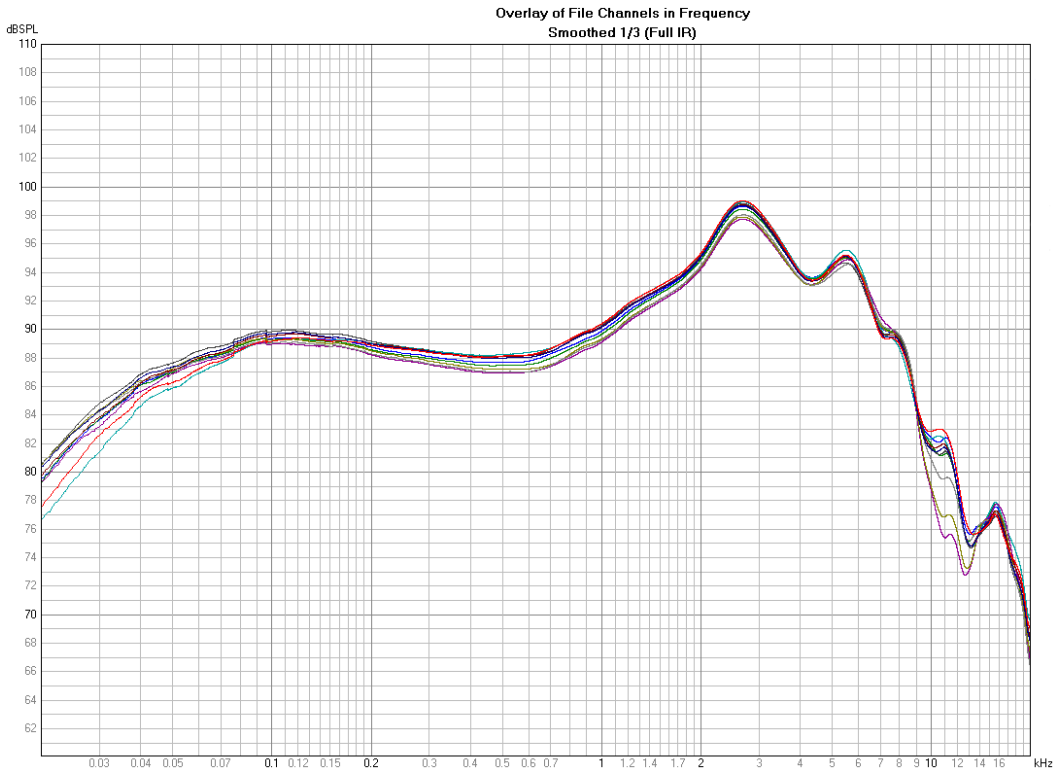


Figure 4.6 Magnitude vs. Frequency Response of Sennheiser HD-650, 1/3rd-Octave Resolution

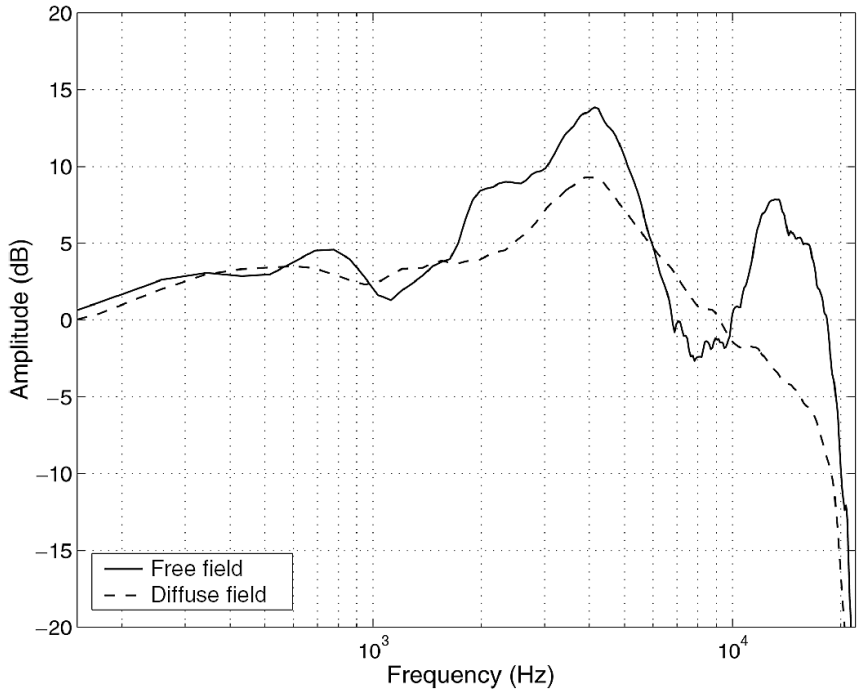


Figure 4.7 Free- and diffuse-field responses for blocked ear canal (Reprinted with permission from [14])

begins to roll-off at around 8 kHz for high-frequencies and 50 Hz for low-frequencies.

With regard to free- vs. diffuse field response, figure 4.7 shows the two responses measured under blocked ear canal conditions. When compared to these measurements, the response of the HD-650, as obtained by the non-decoupled measurement method (shown in figure 4.6), closely resembles the diffuse field measurement.

4.5 Conclusions

The Sony MDR-7506, Sony MDR-V600 and Grado RS-1 do not meet the declared criteria of suitability for use in this research project.

The Sennheiser HD-650, through 10 repeated measurements, met the first two criteria. The device produced results which indicated consistent reflection patterns in the outer ear, coupling of the headphone to the head and coupling of the driver to the eardrum in the frequency range from 40 Hz–10 kHz, which is acceptable for this research project.

As for the third criterion, the Sennheiser HD-650 has an acceptably flat frequency response from 50 Hz–8 kHz. Though the high-frequency limit is lower than would be desired, a reasonably conservative high-shelf filter, applied during stimulus equalization, could conceivably extend the response to 10 or 12 kHz with minimal degradation to the integrity of the stimuli. As such, the HD-650 would meet all of the criteria and is thus acceptable for use in this research project.

5. Preparation of Stimuli

The stimuli that were acquired from the Corbett Auditorium recordings would need to be processed before they could be used in listening tests. In their captured form, the stimuli consisted of 24 wave files, each containing three 50-word MRT lists recorded sequentially and without additional noise, and twelve wave files containing noise recorded at a different levels and locations.

The stimulus files that would be used for this project's studies would need to be equalized, mixed with noise recordings and then parsed. The result would be 108 wave files used for the pilot, 2400 files for phase 1, and 1200 files for phase 2.

5.1 Equalization

From a review of the literature, it is apparent that there are many approaches to the equalization of binaural recordings; ranging from the implementation of Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) equalization filters (e.g. [134, 135]), to parametric equalization filters [90, 92, 173], or to no equalization at all [26]. However, the question arises of which method of corrective equalization is most suitable and appropriate for this specific series of studies.

As mentioned in chapter 2.3.4, it is evident that magnitude versus frequency response anomalies, resulting from the pattern of reflections between the headphone driver and the listener's eardrum, can vary greatly (±3 dB and greater) above 4 kHz between listeners [134]. Thus if filters were to be used to equalize the measured response of the headphone-KEMAR combination, narrow peaks and dips generated by such corrective equalization filters in the region above 4 kHz could yield highly undesirable results. A peak or dip used to counter the effects of a resonance in the KEMAR ear could, given a different individual ear, result not in a cancellation but rather in a peak or dip with larger magnitude. As such, it was determined that equalization, if used, should focus on the areas of

the measured headphone-KEMAR response which would be common to all headphone-ear pairings, thus narrow-band equalization in the region above 4 kHz would be inappropriate.

5.1.1 Determining an Equalization Method

The use of IIR filters was rejected due to this type of filter's potential for issues arising from causality and instability [140]. If an FIR filter were to be used to invert the impulse response of the headphone-KEMAR combination, this would lead to narrow-band notches and peaks in the region above 4 kHz which, as stated above, would not be appropriate. The remaining options are parametric equalization or no equalization.

As seen in figures 4.5 and 4.6, both responses have a prominent peak in the mid- to high-frequency region. It is reasonable to conclude that some sort of equalization would be needed, and this was verified upon listening to the raw stimulus recordings. It was decided that the solution would be to employ parametric equalization filters to cancel the narrow peak at 2.7 kHz and the wider peak centered at 3 kHz, and to smooth the overall spectral tilt of the response. Though slight differences exist (likely due to the response of the headphones), this is essentially the diffuse-field corrective equalization method used by the Etymotic ER-11 KEMAR microphone preamplifier, with the addition of a high-frequency boost similar to that used by Toole [173].

The corrective equalization detailed in table 5.1 was applied, before merging and parsing, to all sound files obtained from the Corbett Auditorium recordings using Sony's Sound Forge software package. A graph comparing the original and corrected transfer functions is shown in figure 5.1.

Filter Type	Frequency (Hz)	B.W. (Octaves) / Slope (dB / Oct.)	Gain	Purpose
PEQ	2700	0.4	−6 dB	Cancel Narrow Peak
PEQ	1600	1.0	−2.5 dB	Cancel Wide Peak
PEQ	3000	2.5	−5 dB	Cancel Wide Peak
PEQ	150	1.6	−3 dB	Spectral Tilt
H-Shelf	6500	6	+6 dB	Spectral Tilt

Table 5.1 Corrective equalization settled upon for use on all stimuli

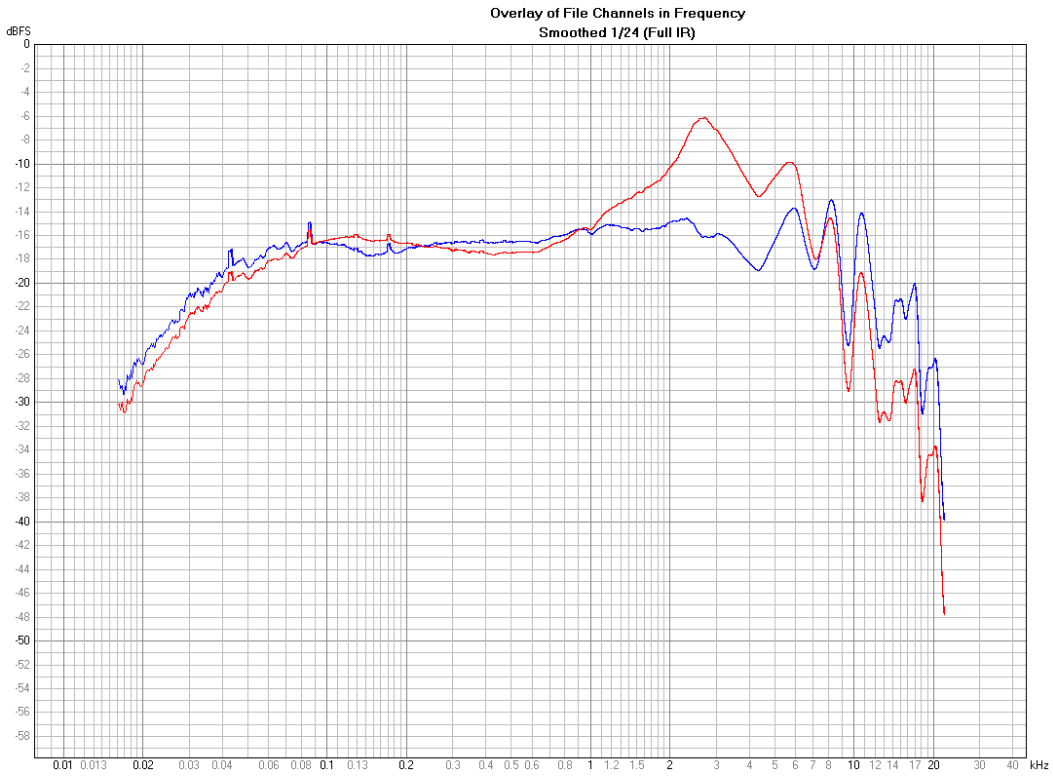


Figure 5.1 Magnitude vs. Frequency Response of the Sennheiser HD-650 before and after the application of corrective equalization. 1/24th-Octave smoothing.

5.2 Merging Noise

As mentioned in chapter 3.3.1, the MRT word lists (recorded under each of the 24 treatments) and twelve different noise distracter recordings were captured separately. Considering that the ambient noise level of the recording space was 30–40 dB SPL below the level of the speech and noise distracter signals, the electronic summation of two such signals would produce a negligible increase in system noise level. It was also considered that speech and noise distracter signals would be produced by separate loudspeakers, thus there was no possibility for inter-modulation distortion between signals within a loudspeaker. As such it was determined, through discussions with advisors, that it would be viable to use electronic rather than acoustic summation of speech and noise

signals to produce the experimental stimuli [91, 125]. Additionally, similar methods have been used by other experimenters [181].

For the creation of the stimuli required for treatments (see chapters 6.2, 7.2 and 8.2 for specific treatments), noise was added either to a whole MRT word list (pilot study) or to individual-word files (main studies). Following equalization and noise merging, all stimulus audio files were converted from their original 96 kHz, 24 bit resolution to 44.1 kHz, 16 bit due to limitations of the intended playback devices.

5.3 Parsing

After the appropriate noise level was added, the stimulus files would need to be parsed into separate and new files of the desired size/number of MRT words. The stimulus files that would be used for the pilot study would contain a single 50-word MRT list. Each file was approximately three minutes long.

The stimulus files that would be used for both phases of the main study would be further parsed, such that each 3.5 sec wave file contained a single word from an MRT wordlist. The composition of the files used for both phases of the main study was as such: Noise fades up over 0.75 sec, the target word plays within the carrier sentence and then the noise fades out over 0.5 sec.

6. Pilot Study

Given that it takes approximately 3–5 minutes for one subject to evaluate one MRT word list, it could take a single subject over 40 hours to fully evaluate a 7×6×2×2 matrix of variable treatments using three word lists per treatment. It is fairly obvious that this type of study would be foolish to attempt and impossible to complete. It was therefore decided to employ a pilot study in this research project, the goal of which would be to reduce the total number of variable treatments to a more manageable figure.

After the stimulus preparation process was completed, the author (and others) listened to several of the MRT word lists as recorded under several variable treatments. It was noted that word identification was extremely easy for treatments that contained no added noise, and that identification was extremely difficult for treatments that contained the highest level of added noise. This preliminary evaluation, coupled with data from the literature [45, 107, 144], led to the conclusion that it would be important to identify the range of the variable Signal-to-Noise Ratio (SNR) that would yield variance in intelligibility scores without overpowering the effects of the other experimental variables.

The pilot study would have three main purposes:

1) Identify a useful range for the variable Signal-to-Noise Ratio
2) Identify a range of values for the other experimental variables that would be of interest for subsequent studies
3) Determine if there exists any issues or problems with testing procedures or principles.

6.1 Hypotheses

In addition to potentially reducing the size of the variable treatment matrix, the pilot study would also offer a preliminary chance to explore the research questions posed by this project. The number of subjects used in the pilot would be small, and thus the number of data points acquired would also be small. As

such the hypotheses that could be tested would likely have to be limited to first-order effects if there would be any hope of finding statistical strength or significance in the results. The resulting hypotheses were as follows:

Ho1: Delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

Ho2: Signal-to-noise ratio does not affect the intelligibility of speech reproduced by a sound system.

Ho3: Array geometry does not affect the intelligibility of speech reproduced by a sound system.

It should be noted that while the second null hypothesis has been rejected many times in many different studies (e.g. [45, 107, 144]), its presence in this study would serve as a test of the methods and procedures of the study itself.

6.2 Study Design

At 40 hours per subject, the full set of variable treatments would be too large to evaluate even in a pilot study. As such it was necessary to construct a reduced set of variable treatments – one that would allow for the fulfillment of the pilot's three main purposes and also provide insight towards evaluating its' three hypotheses.

As can be seen in table 6.1, it was decided to use all but one of the possible values for the variable SNR. During the author's preliminary evaluation of the stimuli (noted above), it was determined that the highest SNR value that still contained added noise (SNR = 12 dB) would not have a significant impact on intelligibility scores. While the highest SNR value (SNR = 50 dB, the condition with no added noise) was also believed to be inadequate to provide significant impairment for subjects, it was left in this study as a proof of concept to show the need for the manipulation of SNR.

Delay Time (ms)	5	20	40			
Approx. SNR (dB)	50	9	6	3	0	-3
Array Geometry	Vert	Hor				

Table 6.1 Variable values used in the pilot study.

As mentioned in chapter 5.2, the SNR conditions were synthesized by electronically mixing the MRT word list recordings with noise recordings. To obtain the recordings of the two different array types, the KEMAR manikin was positioned in two different locations in the auditorium (see appendices B and C for drawings). A separate set of noise recordings was captured for each of these positions and, for the pilot study, it was decided that these separate noise recordings should be used for treatments that include the corresponding array types.

With regard to the variable delay time, the objective of this study was to determine whether the experimental variable would have an effect on intelligibility scores and, if so, determine the range in which the effect begins to take place. The variable values of 5 ms and 40 ms were chosen as they likely constitute short and medium length delay times, thus representing both comb-filtering and spatial enhancement effects, without introducing audible echoes [42, 67, 150]. The value of 20 ms was also included such that, in the event that a significant effect is seen, the results would have greater resolution towards guiding further study.

While it is unclear whether a delay between multiple arrivals will have an effect on intelligibility scores, what is clear is that a level offset between multiple arrivals will serve to mitigate the potential effects of a delay. While such an effect may be of interest in further studies, reducing the effect of an experimental variable would be counterproductive for this preliminary study. As such, the variable level offset was excluded from the pilot.

The variable treatments used for this study formed a 3×6×2 matrix, comprising a total of 36 treatments. Three MRT word lists were used for each treatment, resulting in a total of 108 stimulus sets. Each subject evaluated all of the 108 stimulus sets. The presentation order of the stimulus sets was randomized for each of the subjects using the Matlab function randperm [121].

6.3 Equipment

Subjects in the pilot study listened to stimuli via headphones and recorded their answers on provided response sheets (see appendix J). The stimulus playback system consisted of a MacBook laptop computer, MOTU 896 HD audio interface and a pair of Sennheiser HD-650 headphones. Stimulus audio files were played using iTunes. As mentioned in chapter 5.2, all audio files had a resolution of 44.1 kHz, 16 bit. The track labels for each stimulus indicated the order in which the files should be played, and made no mention of the underlying experimental design.

The level of playback was calibrated using a hand-held sound level meter (IEC 651 Type II). The meter was attached to a 6 cc coupler to approximate ear canal loading effects, and positioned on the left headphone using a flat-plate coupler. Playback level was adjusted at the MOTU such that playback of a stimulus set which contained the loudest noise level (SNR = −3 dB) would produce a measured result of 83 dB SPL (C-weighted, slow integration) – the level of the original sound field, as recorded in the original acoustical environment.

6.4 Subjects

This study used four native English-speaking subjects, 3 male and one female, ranging from 24 to 31 years of age. All subjects were professional audio engineers with a Bachelor's degree or higher in fields relating to sound engineering. All subjects were verified to have unimpaired hearing (re: 25 dB HL at octave frequencies from 250 Hz to 8 kHz) through the administration of hearing acuity tests [62]. Subjects were compensated $5 USD for each listening session.

6.5 Locations

Listening tests were conducted at two recording studios: GAHS studios in Union, KY, USA and the Sound Design Studio at the College Conservatory of Music, University of Cincinnati, OH, USA. Both spaces were found to have background noise levels corresponding to NC-30 or lower [82] when measured (see chapter 3.1.2 for measurement equipment specifications).

6.6 Procedures

All four of the subjects evaluated the 108 stimulus sets in nine sessions, each session containing 12 stimulus sets and taking approximately 36 minutes (45 minutes including breaks) to complete. At the beginning of the first session, each subject was given written and oral instructions regarding the types of sounds they would be evaluating, operation of the playback device and the method of response (see appendix H for instructions).

The subject would then undergo a training process to become familiarized with the stimuli and testing procedures. The training used for this study involved the evaluation of 6 stimulus sets comprised of: All three word lists delivered under variable treatment 1 (5 ms, 50 dB, Vertical Array), list A under treatment 19 (5 ms, 50 dB, Horizontal), list D under treatment 18 (40 ms, −3 dB, Vertical) and list F under treatment 29 (20 ms, 3 dB, Horizontal). Though this is obviously not all-inclusive of the total number of treatments used in this study, this combination follows established recommendations by providing subjects with the opportunity to hear all of the individual words that will be presented (under one of the most intelligible variable treatments), and experience the magnitude of the differences between auditory attributes of the various treatments to be used in the study [14]. This level of training was deemed appropriate as this study employs identification rather than discrimination or scaling tasks, and does not involve the study of listener preference.

Once the training process was complete, the test administrator spoke with the subject to verify that they understood the instructions and operation of the

apparatus, and to remind the subject of the importance of taking breaks to minimize fatigue and distraction. The subject then began the first session of stimulus evaluation.

At the completion of the listening session, the subject was debriefed to determine if they had any concerns about the testing procedure and if they had experienced any perceived hazards or issues with the testing apparatus. No such concerns were noted by any of the subjects during the testing procedure.

The subject was then presented with the post-session oral script document, and their next session was scheduled (see appendix I for document). Subsequent listening sessions proceeded in the same manner as the initial session. However, if the time between an individual subjects' sessions was less than 48 hours, the subject was not required to complete the training process before beginning a session.

6.7 Results & Analysis

The response sheets for the listening tests were scored and the results recorded as the number of correct responses out of 50. Though the probability of a subject randomly guessing the correct answer is quite low (16.7%) for a six-alternative forced choice task, it was decided that the results should be adjusted to account for this probability, as is recommended by ANSI standard S3.2 [5], using the following equation:

$$\textit{Adjusted Score (Ra)} = \frac{\textit{Correct Responses} - \textit{Incorrect Responses}}{\textit{Number of Choices} - 1} \qquad \text{(Eq. 6.1)}$$

The range of Ra would therefore be: $-10 \leq \text{Ra} \leq 10$

As mentioned earlier, the first main purpose of this pilot study was to determine a useful range of the variable SNR. Figure 6.1 shows a box and whisker plot of the adjusted scores vs. SNR. For the case of the variable treatments where SNR = 50 dB, a distinct ceiling effect can be seen in the data, resulting from the test method's inherent upper score limit. Of course, in order to evaluate whether differences exist in adjusted scores between treatment groups, such an effect must be avoided in order to achieve adequate variance in the scores.

Thus the 50 dB SNR variable value should be excluded from future studies and, in fact, from much of the analysis in this study. Similarly the distribution of the data for the 9 dB case should be considered. Though not exhibiting an obvious ceiling effect, the data shows that perfect scores (10 out of 10) were recorded, suggesting that it would be possible to find such an effect if a larger subject population were used.

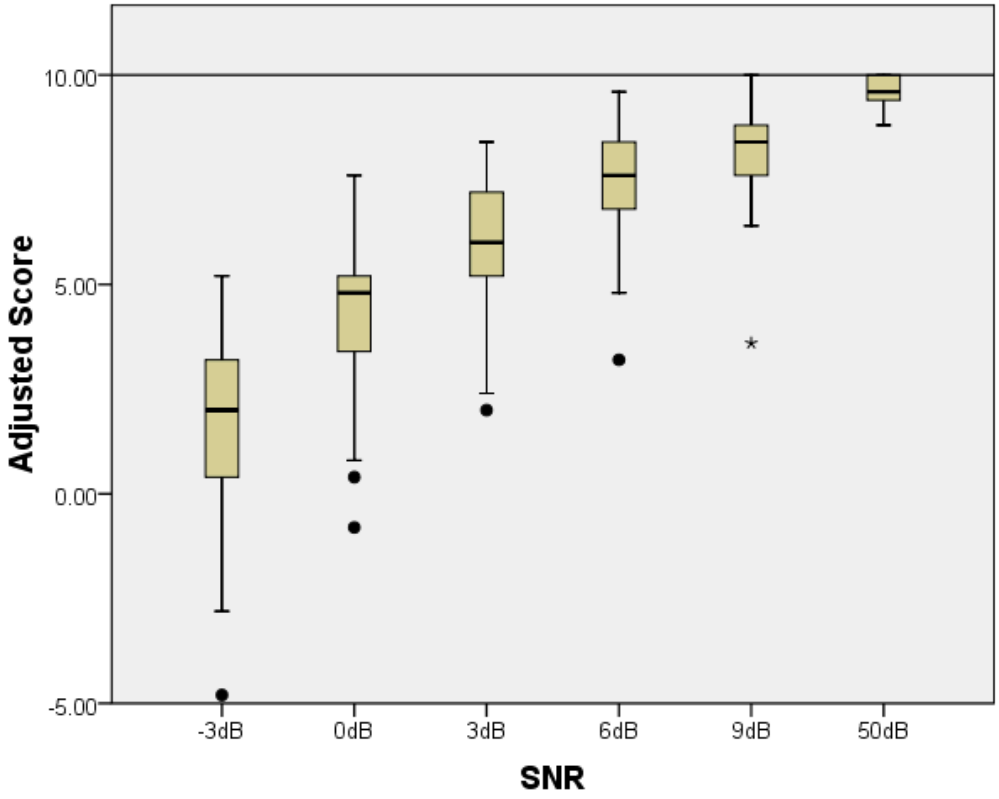


Figure 6.1 Box and whisker plot of adjusted score vs. SNR for all tested levels of SNR.

For the data sets corresponding to the remaining SNR values, there appears to be variance between and within groups with no evidence of a ceiling effect. For the set of treatments where SNR = −3 dB however, note that the highest recorded adjusted score is a 5.2, which in terms of percentage of words correctly identified corresponds to a score of 76%. This suggests that the level of the noise present in these treatments could be too high, making it too difficult for subjects to perform the necessary identification tasks. This data also suggests that the strength of the effect of SNR at this level may overpower the effects of the

other experimental variables. Also, considering that variability is higher for lower scores [156], as can be seen from the results in figure 6.1, the SNR value of −3 dB should likely be excluded from future studies.

The data from the remaining three groups of SNR values show that there is variance within each group and between groups, allowing for both the study of the effect of SNR and the effects of the other experimental variables within SNR groups. For future studies it would therefore make sense to exclude the SNR values of −3 dB, 9 dB and 50 dB. With the exception of investigating the effectiveness of listener training and within-study learning, the remainder of the pilot study analysis will also exclude the data from treatments including these three SNR values. The remaining data is referred to as the *SNR stratified data set*.

6.7.1 Homogeneity of Variance

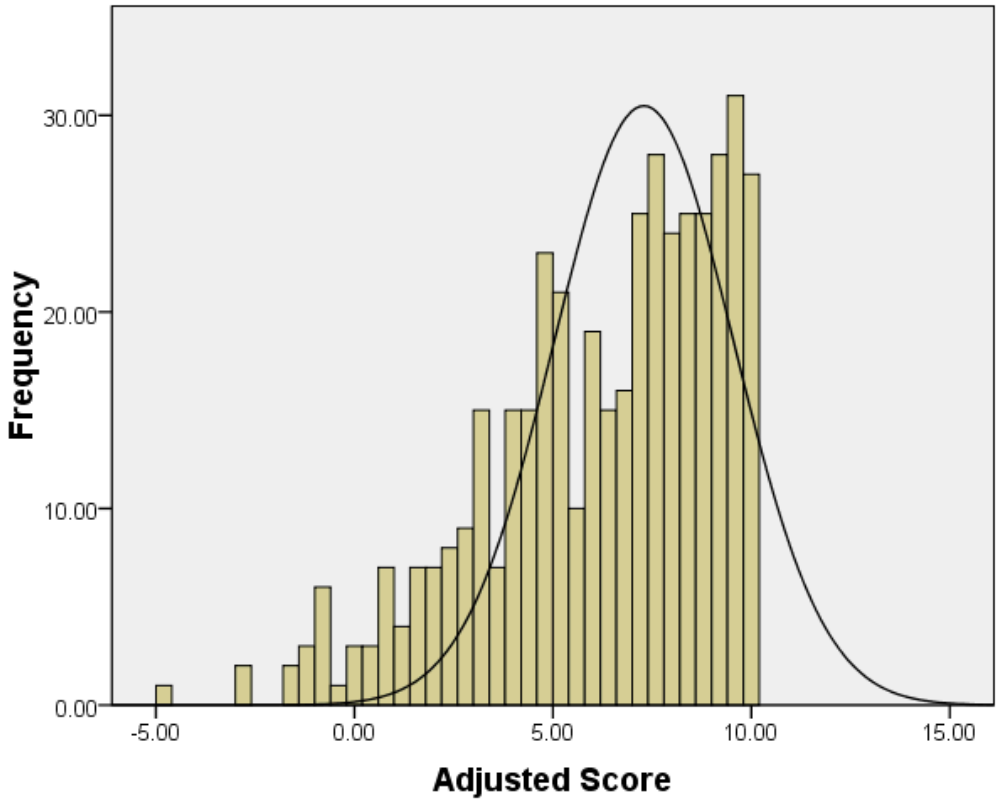


Figure 6.2 Histogram comparing the distribution of adjusted scores for all data collected in pilot study to a normal distribution.

A frequency plot (figure 6.2) of the original data set confirms that the data is left-skewed and that the upper limit of the range of adjusted score may be the culprit. A frequency plot of the SNR stratified data set shows that ceiling effects have been eliminated and that degree of skew has been reduced (figure 6.3). The result is a distribution that more closely resembles a normal distribution, though still exhibiting a bi-modal shape.

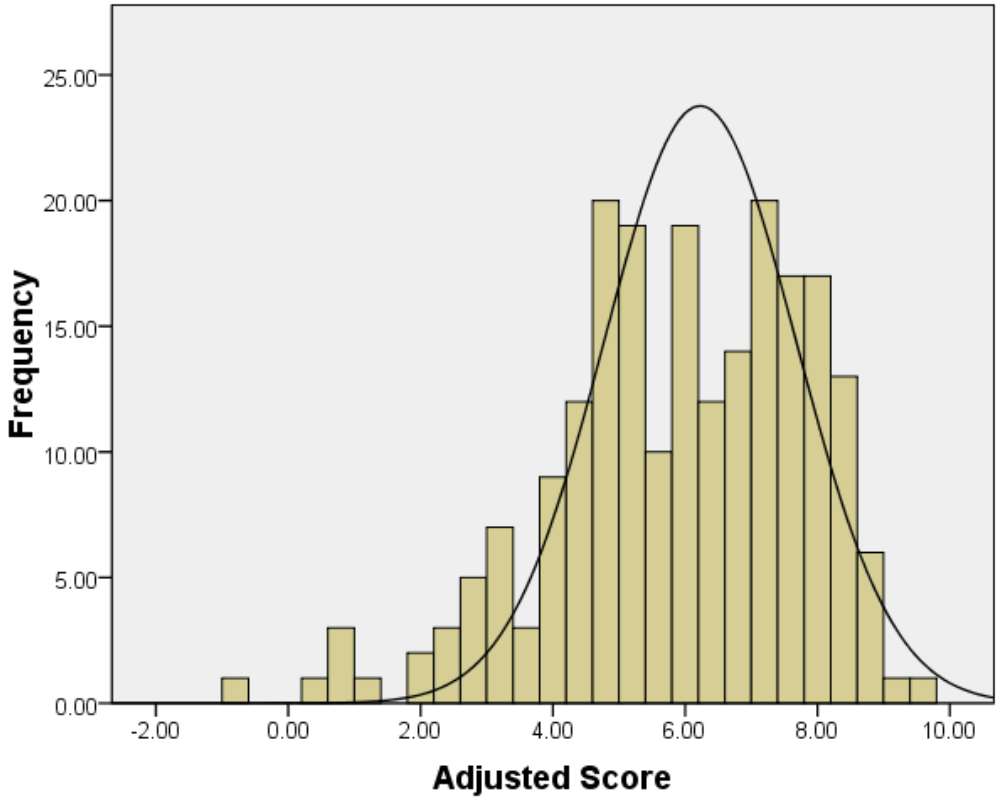


Figure 6.3 Histogram comparing the distribution of adjusted score for SNR stratified data set (SNR values 0 dB, 3 dB, and 6 dB) to a normal distribution.

The standard method for the analysis of numerical data obtained from listening tests is analysis of variance (ANOVA) [14]. The ANOVA process however makes several assumptions, including assumptions about the homogeneity of variance of the data to be analyzed [153]. Table 6.2 shows the results of two standard tests for homogeneity of variance for both the original and SNR stratified data sets. As can be seen from the probability statistics from each of these tests ($p < 0.001$), it is clear that neither of these data sets contains

normally distributed data. This is confirmed by inspection of the detrended normal Q-Q plot of the SNR stratified data set (figure 6.4). As such, it is questionable whether the use of a parametric test such as ANOVA, which is based on the comparison of means, is appropriate.

Data Set	K-S Stat.	K-S Sig.	S-W Stat.	S-W Sig.
Original	0.122	<0.001	0.932	<0.001
SNR Strat	0.098	<0.001	0.962	<0.001

Table 6.2 Results of tests for homogeneity of variance: Kolmogorov-Smirnov and Shapiro-Wilk statistics

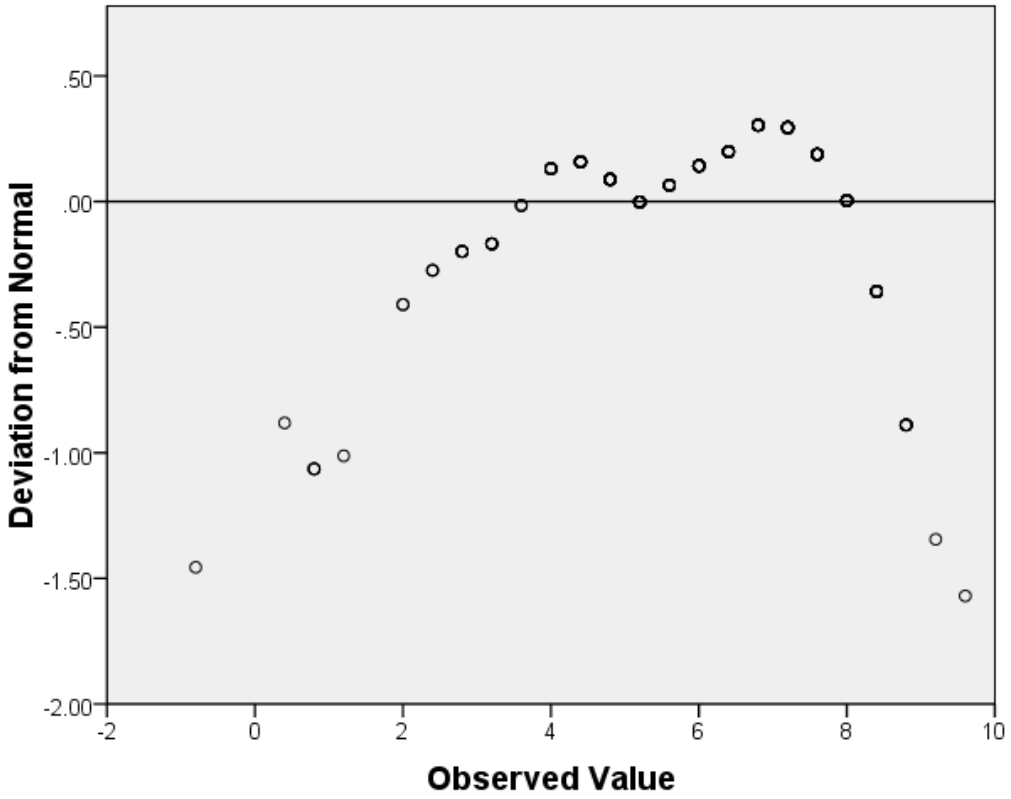


Figure 6.4 Detrended normal Q-Q plot of adjusted score in the SNR stratified data set.

Non-parametric tests – tests which are based on the comparison of medians rather than means – are the accepted solution for analysis of data sets which lack homoskedasticity (normality of distribution) [160]. The Kruskal-Wallis test, an extension of the non-parametric Wilcoxon rank-sum test, will therefore be used for the analysis of the data obtained in the pilot study [160]. As

it is the standard (and expected) analysis method, ANOVA results will also be presented.

6.7.2 Analysis of SNR Stratified Data Set

The results of the first-order analysis of the data are shown in table 6.3. Not surprisingly, SNR and subject have a very significant effect on the results. As mentioned previously, the effect of SNR on the intelligibility of speech through transmission systems is well documented (e.g. [45, 107, 144]). It is also easy to conceive that different people, having different listening skills and hearing acuity, might have differing performance in this type of test. The performance of one subject was of particular interest and will be detailed in chapter 6.8.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Subject	7.277	<0.001	14.642	0.002
Word List	2.396	0.094	4.289	0.117
Delay Time (5 – 40 ms)	3.622	0.028	7.104	0.029
Delay Time (5 – 20 ms)	0.237	0.627	0.201	0.654
Delay Time (20 – 40 ms)	4.151	0.043	4.316	0.038
Array Type	9.262	0.003	9.120	0.003
SNR	86.029	<0.001	104.528	<0.001

Table 6.3 Results of ANOVA and Kruskal-Wallis tests for first-order effects of experimental variables on adjusted score (SNR stratified data set)

It would appear that there is no significant difference in the results obtained from the three different MRT word lists (Word List). This could be an indication of one of the areas where test-size reduction could be achieved.

Delay time and array type both appear to have significant effects. Of particular interest is that, in this study, delay time did not produce a difference (strength or significance) between the 5 ms and 20 ms treatment groups. This could be an indication of the direction for subsequent studies.

As mentioned at the beginning of this chapter, the author felt that it was doubtful that the number of data points acquired in this study would be sufficient

to adequately perform tests for significance of interaction effect (second-order and higher effects). Table 6.4 confirms that this is in fact the case. The tests show that the interactions of SNR with delay time, and delay time with array type, do not indicate strength or significance of effect.

Variable	ANOVA F-Stat	ANOVA Sig.
SNR × Delay Time	0.771	0.545
Delay Time × Array Type	0.611	0.544

Table 6.4 Results of ANOVA for second-order effects of experimental variables on adjusted score (SNR stratified data set)

6.7.3 Further Stratification by Array Type

Another way of viewing the question of interaction effects is to ask whether array geometry modulates (or mitigates) either the magnitude of the effect of delay time or the amount of delay time required to produce a noticeable effect. By dividing the SNR stratified data set into data sets corresponding to array geometry (SNR-Vertical and SNR-Horizontal stratified), we can examine whether delay time has a noticeable effect within these groups. Tables 6.5 and 6.6 show the results of the analysis for these new data sets.

Stratification reduces the number of data points available for an analysis, and dividing a stratified data set compounds the issue. For example, the Kruskal-Wallis test on the SNR stratified data set shows that, over the range of 5 to 40 ms, delay time has a significant effect on adjusted scores ($\chi^2 = 7.104$, Sig. = 0.029). However after dividing the data sets, the same test fails to find a significant effect for either of the array types. This suggests that the lack of significance may be due to the reduced power stemming from the lack of data points (Type-II error), rather than a lack of effect. This could be of interest in future studies.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (5 – 40 ms)	2.152	0.121	3.503	0.173
Delay Time (5 – 20 ms)	0.219	0.641	0.204	0.651
Delay Time (20 – 40 ms)	3.973	0.050	3.008	0.083

Table 6.5 Results of ANOVA and Kruskal-Wallis tests for first-order effects of delay time on adjusted score (SNR-Vertical stratified data set)

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (5 – 40 ms)	2.210	0.115	3.888	0.143
Delay Time (5 – 20 ms)	1.191	0.279	0.806	0.369
Delay Time (20 – 40 ms)	0.986	0.324	1.343	0.247

Table 6.6 Results of ANOVA and Kruskal-Wallis tests for first-order effects of delay time on adjusted score (SNR-Horizontal stratified data set)

6.8 Discussion

The first, and probably most obvious, point that might be made is that more data points are required before this project's research questions can be adequately addressed. Yet even with the small data set generated by this preliminary study, there are several things that can be learned and applied to the remainder of this project.

6.8.1 A Deeper Look at Array Type

From table 6.3 it can be seen that array geometry has a fairly strong and very significant effect on scores (K-W test: χ^2 Stat = 9.120, Sig. = 0.003) in the SNR stratified data set. If one were to take a moment to consider the actual array geometries, the cause of this effect will become clear.

As mentioned in chapter 2.2.1, the vertical array geometry presents the listener (via binaural recordings) with a 3-channel reproduction of speech and noise, similar to the type of display found in a 5.1-channel surround sound system [86]. The horizontal array, on the other hand, provides two 2-channel displays. The latter is analogous to a single 2-channel stereophonic reproduction system.

Also mentioned in chapter 2.2.1, the use of a dedicated center channel has been shown to result in increased intelligibility and clarity of speech when compared to 2-channel systems that rely on channel summation and phantom sound sources to create a center image. Both of the arrays used in this study employ channel summation, and thus create a phantom sound source image for speech signals. However the vertical array geometry employs speech channel summation in a plane which is perpendicular to the plane of noise signal channel summation. In addition, this geometry contains a physical loudspeaker located in front of the listening position. One would expect, then, that intelligibility scores would be higher for treatments including the vertical array geometry. The results shown in tables 6.7 and 6.8, and in figure 6.5, confirm that this was the case for the data collected in this study.

Array Type	Mean	Std. Dev.
Vertical	6.300	1.832
Horizontal	5.526	1.906

Table 6.7 Comparison of means, from ANOVA on SNR stratified data set

Array Type	Mean Rank
Vertical	121.31
Horizontal	95.69

Table 6.8 Comparison of mean ranks, from Kruskal-Wallis test on SNR stratified data set

The choice to use separate sets of noise recordings for the different array types is another possible factor that could have contributed to the observed effect of array type. As two different sets of noise recordings were used, this ultimately should be viewed as an insufficiently controlled variable. The possibility must therefore be considered that this may have been the cause of the observed effect.

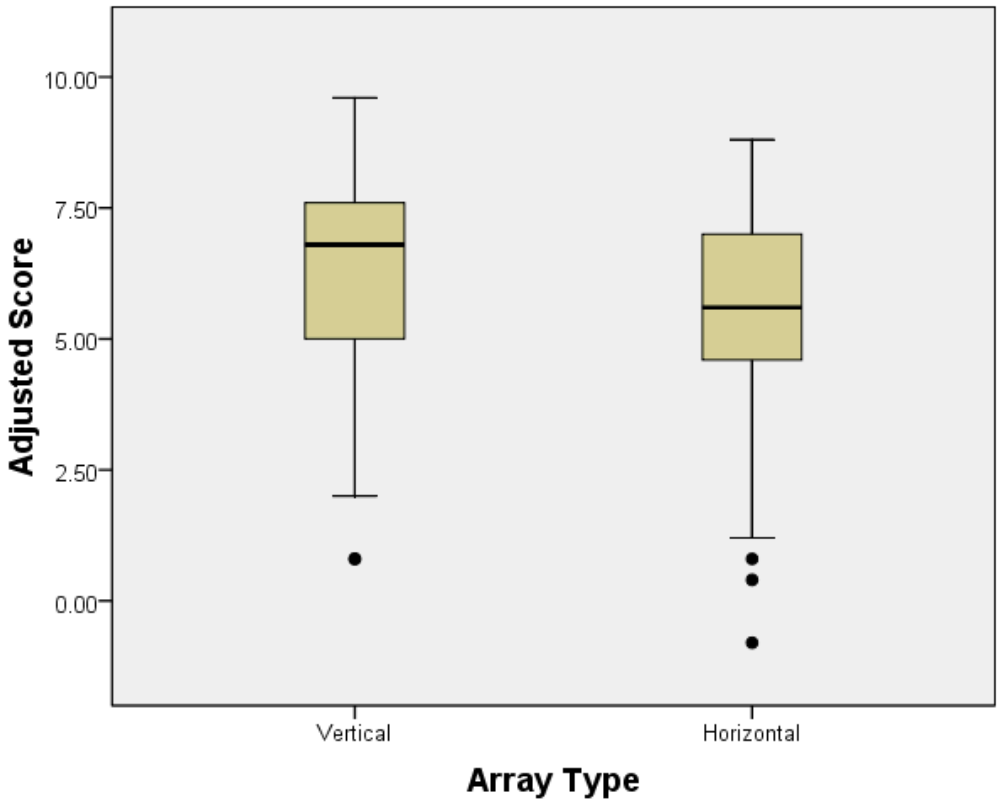


Figure 6.5 Box and whisker plot of adjusted score vs. array type for SNR stratified data set

As mentioned earlier, one of the main purposes of this pilot study was to identify problems with the experimental design and methods before proceeding to the main studies. The use of two different noise sets is a design element that would apparently fall into that group.

6.8.2 Training & Effects of Presentation Order

In determining whether there were issues with the testing procedures and principles, one area of particular interest was the training of listening test subjects. Subjects need to engage in sufficient training such that they understand how to use the testing apparatus, understand the method and/or scales used to indicate responses, and are familiar with the magnitude of variance of auditory attributes between the variable treatments that will be presented during the tests. This raises the question of, "How much training is enough?" A subject's time is limited and,

given the tradeoff between training time and testing time available in a testing session, it is important to find an appropriate balance between the two.

If a subject's training is insufficient, one way that this may manifest itself is in a noticeable improvement in subject performance over the course of the experiment. Table 6.9 shows the results of the analysis of the effect of presentation order on adjusted scores. Remembering that the presentation order of variable treatments was randomized for each individual subject, a presentation order effect would likely indicate some form of learning on the part of the subjects. As can be seen, the results indicate that presentation order does not have a significant effect on scores. Considering the large number of word lists that each subject was asked to evaluate, and the fact that evidence of a learning effect could be obscured by such a large data set, the analysis was repeated using only the first 10 lists evaluated by each subject. Again, the results indicate that there is no significant effect. Combined with positive feedback from subjects, the data leads to the conclusion that the amount of training provided was adequate.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Order (all)	0.888	0.765	97.429	0.735
Order (1-10)	0.846	0.581	9.473	0.395

Table 6.9 Results of ANOVA and Kruskal-Wallis tests for the effect of stimulus presentation order on adjusted score (original data set). Results are shown for all word lists evaluated and for the first 10 sets evaluated by each subject.

6.8.3 Test Duration & Subject Fatigue

Bech and Zacharov [14] discuss listening test duration and the importance that a researcher must place on preventing subjects' loss of attention or boredom. They note that 20 minutes appears to be a suitable duration for a listening session, and that sessions lasting 30–40 minutes are also acceptable if subjects are able to take breaks when they feel themselves getting fatigued or bored.

The listening sessions for the study detailed in this chapter, which lasted approximately 36 minutes, fall into the latter of these groups. As such, subjects

were advised, verbally and in writing, regarding the importance of breaks. As can be seen in figure 6.7, there is some question as to whether this session duration is appropriate for all subjects.

Figure 6.6 is a reprise of figure 6.1 from chapter 6.7, with the addition of outlier identification. It is interesting to note that all of the outliers come from the data set of subject #4, whose data point indices range from 325 to 432. During the debriefing after the subject's final session, the subject indicated to the author that there were several points during the overall testing process when the subject was aware of the onset of fatigue and probably waited too long before taking a break. Figure 6.7 further shows that the data set acquired from subject #4 contains a significantly different range of variance towards the lower bounds of the scale. The outlier seen in the data for subject #3 suggests that this individual may have experienced a similar period of fatigue.

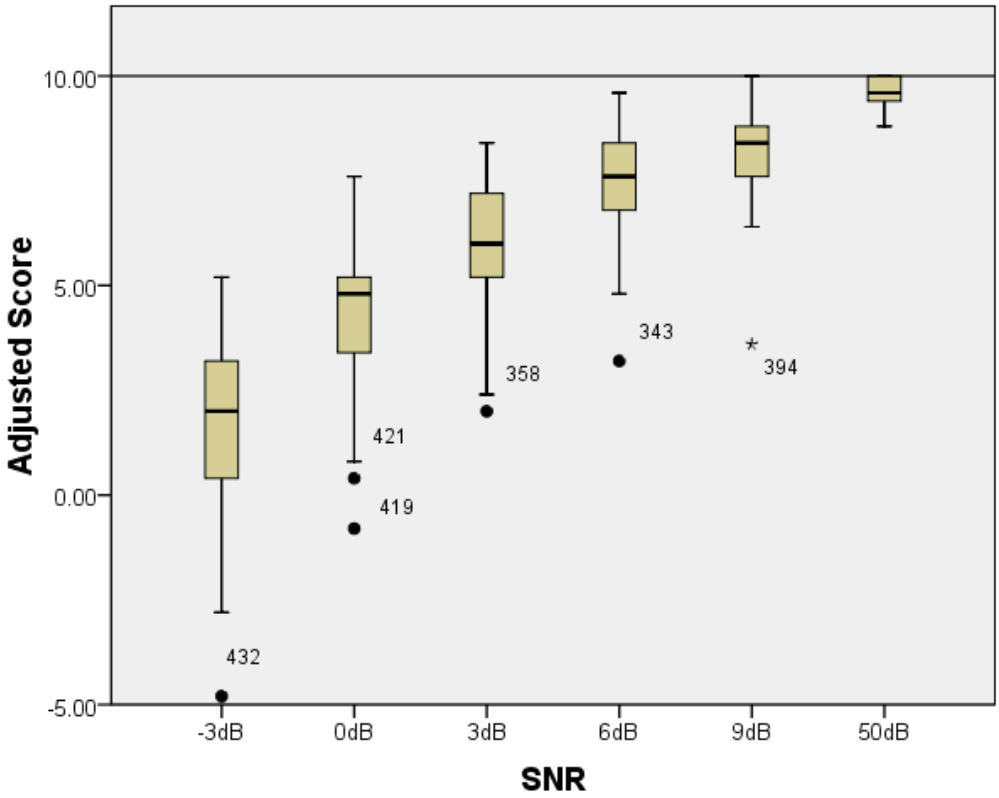


Figure 6.6 Box and whisker plot of adjusted score vs. SNR for all tested levels of SNR. Outliers are identified by data point index number.

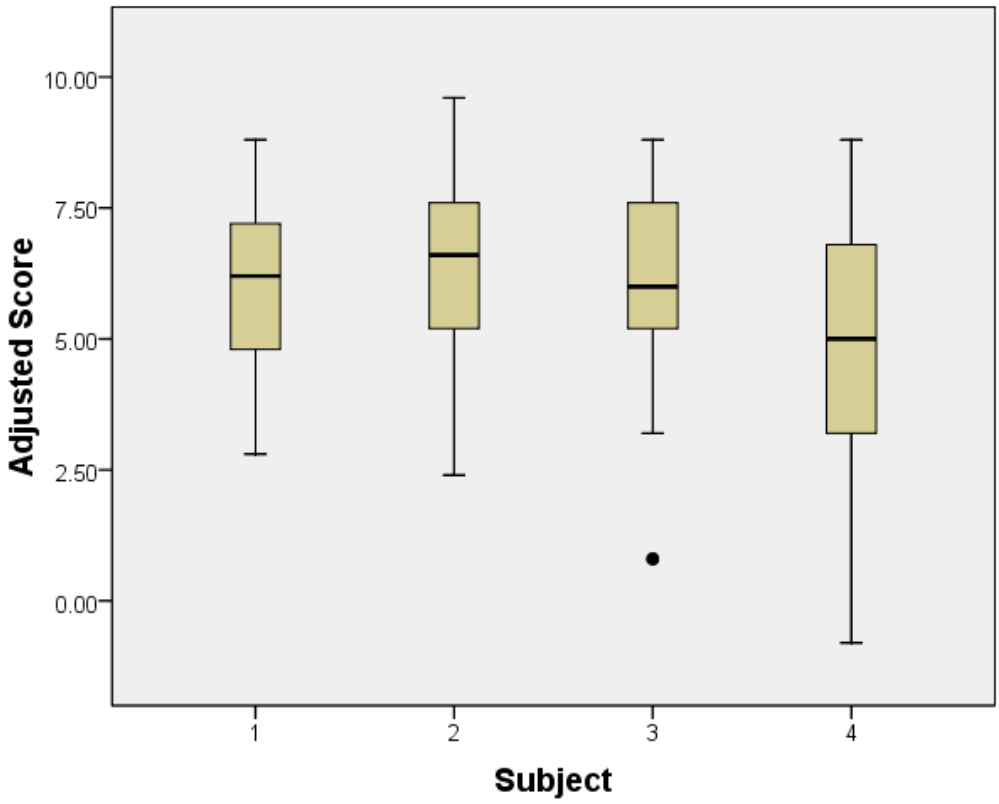


Figure 6.7 Box and whisker plot of adjusted score vs. subject (SNR stratified data set)

The information from the debriefing and from figures 6.6 and 6.7 reinforces the importance of breaks in the testing process. It is clear that for future studies session duration and break spacing should be reevaluated, and possibly adjusted, to avoid issues arising from listener fatigue.

6.8.4 Effects of Word List

As mentioned in chapter 2.1.3.3, many studies which have employed the MRT have elected to use a reduced set of stimuli as opposed to the set recommended in the ANSI standard [5]. The reduction of the number of lists used can drastically reduce the overall size and duration of a study and, to that end, was included as a variable of interest in this study. Table 6.10 once again shows the results of analyses regarding an effect of word list on scores. While the F- and χ^2-statistics suggest that there may be an effect present, the probability values for

both statistical tests indicate that the tests fail to rule out the possibility that the effect is solely due to chance.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Word List	2.396	0.094	4.289	0.117

Table 6.10 Results of ANOVA and Kruskal-Wallis tests for the effect of MRT word list on adjusted score (SNR stratified data set)

A similar result is found when figure 6.8 is examined. While it does appear that there is a slight upward shift in medians from list 1 to 3, the degree of overlap between inter-quartile ranges and variance between extrema make it difficult to draw any conclusions.

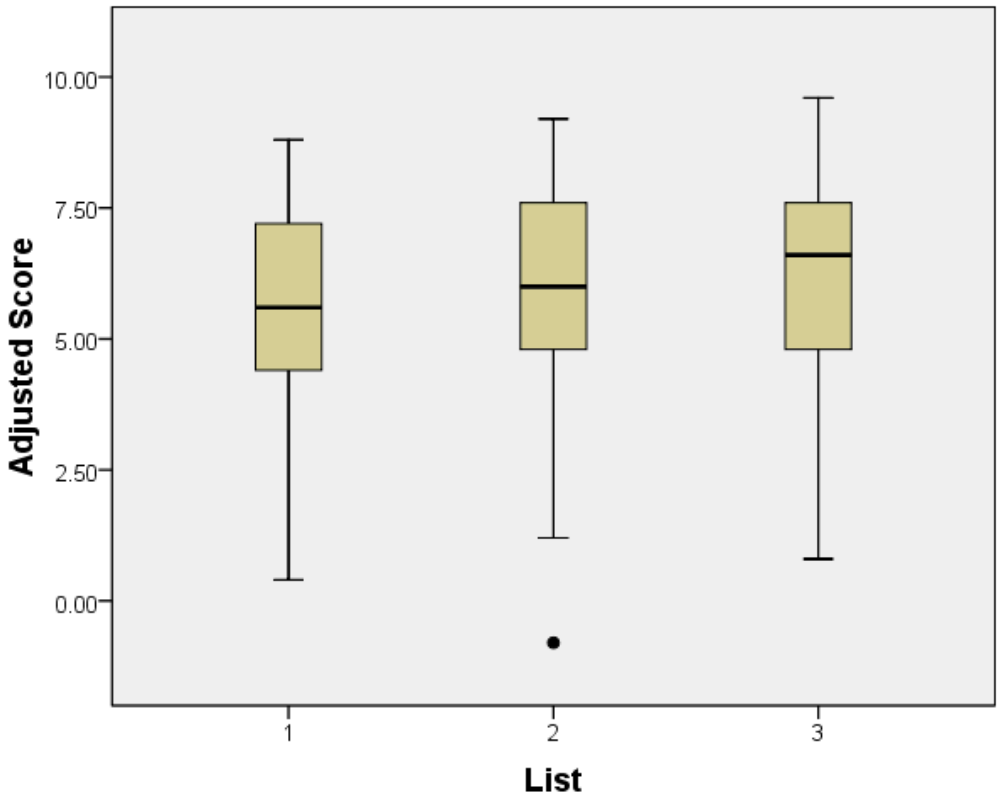


Figure 6.8 Box and whisker plot of adjusted score vs. word list (SNR stratified data set)

The relevant conundrum is that a variety of word lists should be used in order to minimize the likelihood that subject learning will defeat the listening test. If, for example, after evaluating 5 stimulus sets, a subject realizes that the 5th word

will always be “SIT”, the challenge is lost and the evaluation of further stimulus sets becomes moot. However if the choice of word list used for evaluation of a particular treatment will have no effect on the results, it would be possible to have subjects evaluate stimulus treatments using only one list, randomized by treatment. This would amount to a 66% reduction of total test size and time.

It has been mentioned that, given the meager amount of data, the results of this study should be viewed as a preliminary venture. Significance seen (or more likely, not seen) in these results may be purely a result of the size of the study. After all, the data contains only 12 data points for any given treatment. However, with 144 data points per word list, confidence in the results can be somewhat higher. When one considers the potential for design reduction that this variable presents, it is certainly worth further investigation.

6.9 Conclusions

As stated at the beginning of this chapter, the three main purposes of this study were to find a useful range for the variable SNR, attempt to reduce the number of treatments needed for future studies and determine if there are any problems with the design of the study or testing procedures. Additionally, three research questions were posed in the form of hypotheses.

6.9.1 Hypotheses

Ho1: Delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

From the results of both parametric and non-parametric tests, there is sufficient evidence to reject this null hypothesis. The results indicate that the variable delay time does have a significant effect on intelligibility scores. The tests also indicate that this variable begins to have an effect somewhere in the region between 20 ms and 40 ms.

Ho2: Signal-to-noise ratio does not affect the intelligibility of speech reproduced by a sound system.

The results indicate that this hypothesis is clearly rejected.

Ho3: Array geometry does not affect the intelligibility of speech reproduced by a sound system.

The results clearly indicate that, in this study, array geometry does have a significant effect on intelligibility scores. What is unclear is whether the observed effect was actually due to array geometry, or whether it is due to an error in experimental design. Given this lack of clarity, it would be unwise to reject this null hypothesis at this juncture.

6.9.2 Main Points of the Pilot Study

1) Identify a useful range for the variable SNR

It was determined that the SNR values of 50 dB, 12 dB and 9 dB would not provide sufficient impairment to intelligibility. The result was an observed ceiling effect in the data. It was further determined that the SNR value of −3 dB provided too much impairment to intelligibility. The result was excessive levels of variance in intelligibility scores. It was therefore concluded that the useful range of SNR values was between 0 dB and 6 dB.

2) Identify a range of values for the other experimental variables that would be of interest for subsequent studies

While a useful range for the variable SNR has been identified, the question remains regarding how many values of this variable should be used in future studies. The three values that remain differ by only 3 dB each and, while such a difference would likely be large enough to have a significant effect on results, it may be possible to eliminate one or more of these variable values.

In the scope of this project, it is desired to determine whether the interaction between SNR and delay time has an effect on intelligibility scores. As can be seen from table 6.4, it is not possible to draw any conclusions about

interactions at this time. If it is desired to examine a SNR and delay time interaction in future studies, it would be necessary to employ at least two levels of the variable SNR. It is recommended that the following study use the SNR values of 0 dB and 6 dB, as this will allow for the investigation of said interaction, and because these values represent the largest SNR difference within the determined useful range.

The results from this study suggest that the effect of delay time on scores begins to become significant somewhere in the range from 20 ms to 40 ms. Future studies therefore have several options regarding which set of delay time values to use. The most obvious course of action would be to use only delay time values in the range between 20 ms and 40 ms, in an attempt to better specify the point of significance. A less obvious course becomes manifest when the size of the pilot study is again considered. The lack of effect significance found in the 5 ms to 20 ms range may be due to the diminutive number of data points rather than an actual lack of effect. If a future, larger study were to concentrate on values within this range, it could provide more convincing results than were possible from this study. If this is to be attempted, it is recommended that the 40 ms value be left in the study as a dummy value to test experimental methods.

Given the uncertainty arising from the use of two different noise sets for the variable array type, the effects of this variable, and interactions involving this variable, should be reevaluated in future studies. As such, both values for this variable will need to be retained.

Though not technically an experimental variable, analysis failed to prove that the variable word list had a significant effect on scores. As this could lead to a substantial reduction in the size of future studies, it is recommended that the first phase of the main study include this as an experimental variable. If this is done, said study should evaluate whether list-based study size reductions present issues or problems to the experimental design.

3) Determine if there exists any issues or problems with testing procedures or principles.

The first problem identified with the design of this study was the use of scoring sheets rather than computerized data acquisition. The process of scoring response sheets is extremely time consuming and does introduce the potential for human error. Future studies should (and would) employ a computer program and graphical user interface developed using Matlab.

One potential issue encountered involved the degree of training that subjects were given before their initial listening test session. Analysis of the effects of presentation order on intelligibility scores revealed that no evidence of learning was present in the results. It was thus concluded that the level of training was sufficient to prevent the invalidation of the results from a subject's first few stimulus evaluations.

Another potential issue was identified during the analysis of the effects of array type on intelligibility scores. It is clear that using different sets of noise files, to manufacture the SNR variable for treatments involving different array types, introduces an unwanted level of uncertainty into the results. While it may have the effect of reducing the ecological validity of future results to a small degree, it is clear that one set of noise files should be used on all variable treatments.

Listening session duration and breaks are other areas of the project design that need addressing. It is clear from the results and subject debriefing that at least one subject who took part in this study did not take breaks at appropriate intervals during the testing process. There is no evidence to prove that the duration of the testing sessions was excessive, however it is imperative that the test administrator convey the importance of breaks to subjects in future studies. Relying on a subject's judgment of his or her own level of fatigue may not be the best way to determine break frequency. Developing a fixed policy regarding the number of breaks per listening session should be considered.

6.9.3 Parting Thoughts

This research project was charged with finding correlations between sound system optimization parameters and speech intelligibility. What originally resulted was a four-dimensional matrix of variables containing 168 possible variable treatments. This pilot study was implemented as a way to cast a broad net over the research questions, in an attempt to reduce the size and complexity of the task at hand. To that end, the study was a success. In addition, the net that was cast has returned some information that will serve as veritable guideposts for the following main study. Issues with study design and testing methodology have been identified and addressed, and preliminary data has indicated that there is indeed value in continuing the project.

The pilot study began with three purposes and three questions. Heading into the next study, some answers, and many more questions, have been discovered.

7. Main Study: Phase 1

The pilot study was able to significantly reduce the size of the project's variable matrix from 7×6×2×2 (168 treatments) to 3×6×2×2 (72 treatments). Additionally the results from the pilot, though not conclusive, do suggest a range of interest for the variable delay time, which could further reduce the size of the matrix to 3×3×2×2 (36 treatments). Finally, the results from the pilot suggest that it may be possible to reduce testing time by reducing the number of MRT word lists needed to evaluate each treatment.

36 treatments would be a manageable set of stimuli if the number of required word lists were reduced. However, suggestions are not conclusions and, at this point, it has not been conclusively shown that it would be appropriate to carry out the requisite reductions in the treatment matrix. Though it is an improvement over the size of the original treatment matrix, 72 treatments remains a prohibitively large number of stimuli to attempt conclusive subjective evaluation. It is clear that further reduction is necessary.

This first phase of the main study will serve as an intermediary stage between the pilot study and the second phase of the main study. It could be considered a method of successive approximation: Each study in this project seeks to evaluate hypotheses and facilitate greater clarity and statistical strength in subsequent studies. Casting a less broad net over the variable treatments, the main goal of this phase would therefore be the further reduction of the number of variable treatments such that the size of the treatment matrix used in the second phase would be manageable.

This first phase of the main study would have 4 main points:

1) Attempt to further reduce the number of treatments of interest
2) Determine if the effects of delay time values in the region less than 20 ms are still found insignificant with a larger test group.
3) Test the validity of using only one MRT word list per treatment
4) Evaluate hypotheses

7.1 Hypotheses

The number of subjects used in this study would be substantially larger than that used in the pilot study. As such the results would have a greater potential to determine the significance (or lack thereof) of effects of the experimental variables on the dependent variable intelligibility score.

The effects of delay time remain of particular interest and, as such, will be examined here. Also, given the greater number of subjects, it may be possible to see significance in interaction effects between delay time and the variables SNR and array type.

The question remains whether array type itself has an effect on intelligibility scores. Given the results from previous research [74, 163], it is expected that a significant effect will be noticed. Further, inclusion of a null hypothesis regarding array type could function as a dummy hypothesis to test the validity of the experimental design. For this same reason, a null hypothesis regarding the effect of SNR will also be included.

The resulting hypotheses were as follows:

Ho1: A delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

Ho2: Signal-to-noise ratio does not affect the intelligibility of speech reproduced by a sound system.

Ho3: Array geometry does not affect the intelligibility of speech reproduced by a sound system.

Ho4: (interaction) Signal-to-noise ratio does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

Ho5: (interaction) Signal-to-noise ratio does not affect how array geometry affects the intelligibility of speech reproduced by a sound system.

Ho6: (interaction) Array geometry does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

7.2 Study Design

As mentioned, conclusions from the previous study suggest that the incorporation of more subjects into the design of the current study could yield better resolution in the investigation of the effects of delay time.

The results of the previous study suggest that the variable delay time begins to have a significant effect on intelligibility scores in the range between 20 ms and 40 ms. As was seen in the studies by Teuber and Völker and Peutz [144, 170], electronic estimation methods indicate that the degree of impairment to intelligibility due to multiple arrivals is directly related, up to a point, to the amount of delay time between the arrivals. While results from these studies do not suggest a point of critical significance, reductions in measured (via RaSTI) speech transmission index and $\%AL_{cons}$ in the region do suggest that it is possible for delay time to have a significant effect in the range less than 20 ms. In the combined effort to evaluate the first hypothesis and reduce the size of the variable treatment matrix, it was decided to focus on the region between 0 ms and 10 ms. The 40 ms variable value was also included to verify the results of the pilot study; that the point of critical significance is found in the region less than 40 ms.

Following the recommendations of the pilot study, SNR variable values would be confined to the region between 0 dB and 6 dB, and should include at least two values, such that the study of interactions would be possible. As such, two values (0 dB and 6 dB) were chosen for this study. As recommended in chapter 6.8.1, all of the noise recordings used to manufacture SNR values were taken from those recorded at the manikin-center recording position.

Both types of array geometry would be studied. While investigating the first-order effect of array type on intelligibility could be deemed a trivial pursuit, the retention of both types in the study will allow for several investigations. Acknowledging the differences mentioned in chapter 3.2.2, if the horizontal and

vertical array types are sufficiently analogous to 2-channel and 5.1-channel reproduction systems (respectively), it is reasonable to expect that intelligibility scores will be higher for the vertical array geometry. Analysis of the first-order effect could therefore provide insight as to the successfulness of this study's design. Such analysis could also serve to clarify the degree of uncertainty injected into results of the pilot study due to the use of two different noise sets. Additionally, examination of the two-way effect of array type × delay time could yield further insight into the role of the complex variable array type.

As can be seen from table 7.1, the design of this study would employ a 4x2x2 matrix of variable treatments containing 16 total treatments. The presentation order of the variable treatments was randomized for each of the subjects using the Matlab function "randperm". For each of these 400 variable treatment presentations (16 treatments by 25 subjects), an MRT word list was randomly assigned using the Matlab function "randi", which generates pseudo-random integers from a uniform discrete distribution [121].

Delay Time (ms)	0	5	10	40
Approx. SNR (dB)	6	0		
Array Geometry	Vert	Hor		

Table 7.1 Variable values used in the first phase of the main study.

7.3 Equipment

As with the pilot study, subjects would evaluate binaural recordings via headphone display. The audio playback equipment used in the current study includes an IBM Lenovo S10 Ideapad (PC-based) netbook computer, Lexicon Lambda USB audio interface and Sennheiser HD-650 circum-aural headphones. All audio files had a resolution of 44.1 kHz, 16 bit.

The level of playback was calibrated using a hand-held sound level meter (IEC 651 Type II). The meter was attached to a 6 cc coupler to approximate ear canal effects, and positioned on the left headphone using a flat-plate coupler. Playback level was adjusted at the audio interface such that playback of a stimulus set which contained the loudest noise level used in this study (SNR = 0

dB) would produce a measured result of 80 dB SPL (C-weighted, slow integration) – the level of the original sound field, as recorded in the original acoustical environment. If it was necessary to perform a hearing acuity test on a subject (as was done during each subject's first listening session), level calibration for the playback software was performed after the completion of the acuity test.

A program, including graphical user interface (GUI), was written using Matlab (see figure 7.1). Subjects would enter their user number and the test number (1–16) in the main Matlab window, initiating the test and launching the GUI. In contrast to the testing method employed for the pilot study, the program would allow subjects to evaluate stimuli at a pace determined by the subject. A Subject would press the "PLAY" button, listen to the stimulus, and then attempt to identify the target word from the ensemble of six possible choices. The program would only allow an individual stimulus file to be played once.

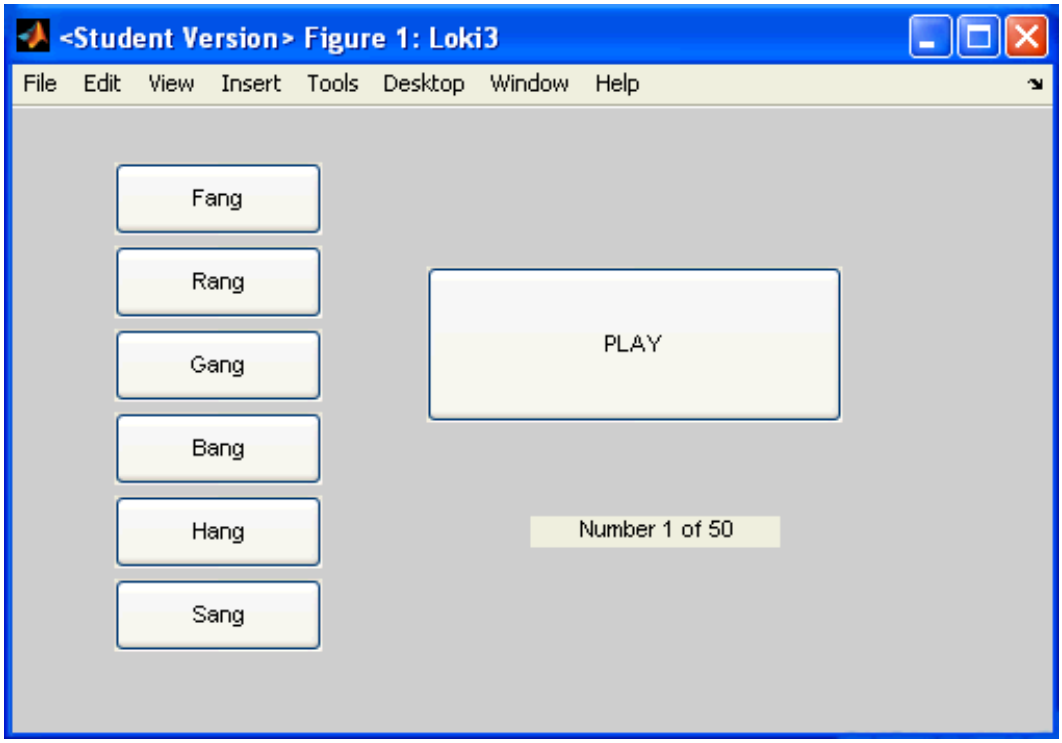


Figure 7.1 Graphical user interface for the associated program ("Loki3") used for electronic presentation of stimuli and recording of subject responses.

7.4 Subjects

This study used 25 native English-speaking subjects, 10 male and 15 female, ranging from 19 to 39 years of age. In terms of familiarity with the field of audio engineering, 3 subjects were professionals in the field, 8 were students and 14 indicated no experience with the discipline. Two of the subjects had participated in the pilot study. All subjects were verified to have unimpaired hearing (re: 25 dB HL at octave frequencies from 250 Hz to 8 kHz) through the administration of hearing acuity tests [62]. Subjects were compensated $5 USD for each listening session.

7.5 Locations

Listening tests were conducted at two locations. For the first 14 subjects, the location used was the Sound Design Studio at the College Conservatory of Music, University of Cincinnati, OH, USA. For the remainder of the subjects, tests were carried out during off-hours in the main backstage area of the Norton Opera Hall at the Chautauqua Institution, Chautauqua, NY, USA. Both spaces were found to have background noise levels corresponding to NC-30 or less [82] when measured (see chapter 3.1.2 for measurement equipment specifications).

7.6 Procedures

All 25 of the subjects evaluated one MRT word list for each of the 16 variable treatments. For each subject, this was completed in two sessions, each session containing 8 stimulus sets and taking approximately 36 minutes (45 minutes including breaks) to complete. At the beginning of each subject's first session, a hearing acuity test was administered to verify that the subject did not have a hearing impairment. Each subject was then given written and oral instructions regarding the types of sounds they would be evaluating, operation of the playback device and the method of response (see appendix H for instructions).

The subject would then undergo a training process to become familiarized with the stimuli and testing procedures. The training used for this study involved the evaluation of 4 stimulus sets comprised of: List A delivered under variable treatment 1 (0 ms, 6 dB, Vertical Array), list D under treatment 4 (40 ms, 6 dB, Vertical Array) and list F under treatments 9 (0 ms, 6 dB, Horizontal) and 16 (40 ms, 0 dB, Horizontal). This training set provided subjects with the opportunity to hear all of the individual words that would be presented, and experience the magnitude of the differences between auditory attributes of the various treatments to be used in the study.

Once the training process was complete, the test administrator spoke with the subject to verify that they understood the instructions and operation of the apparatus, and to remind the subject of the importance of taking breaks to minimize fatigue and distraction. Acknowledging the pilot study's recommendation, a fixed policy regarding the spacing of breaks was implemented for this study. The subject was instructed that, while they were free to pause during the testing process at any point, they would be required to take a 1- to 2-minute break after the completion of every two 50-word sets (approximately every 8–10 minutes).

The subject then began the first session of stimulus evaluation. At the completion of the listening session, the subject was debriefed to determine if they had any concerns about the testing procedure and if they had experienced any perceived hazards or issues with the testing apparatus. The only issue, noted by one subject, was in regard to the uncomfortable size of the "travel mouse" which was connected to the netbook.

The subject was then presented with the post-session oral script document, and their next session was scheduled (see appendix I for document). The second listening session proceeded in the same manner as the initial session. However, if the time between an individual subjects' sessions was less than 48 hours, the subject was not required to complete the training process before beginning the second session.

7.7 Results & Analysis

Results of the listening tests were stored by the test software, scored as the number of correct responses out of 50. As was the case in the previous study, the results were adjusted to account for the probability of chance-guessing using the following equation:

$$\textit{Adjusted Score (Ra)} = \frac{\textit{Correct Responses} - \textit{Incorrect Responses}}{\textit{Number of Choices} - 1} \qquad \text{(Eq. 7.1)}$$

The range of Ra would therefore be: $-10 \leq \text{Ra} \leq 10$

7.7.1 Defining Exclusionary Criteria

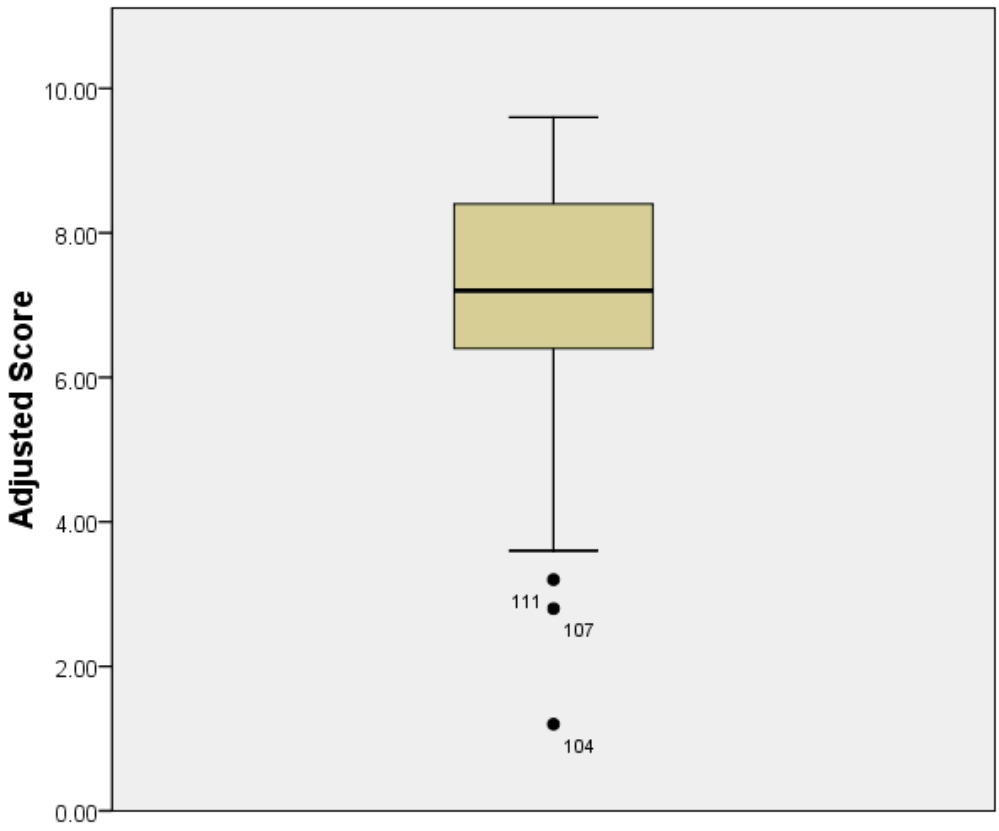


Figure 7.2 Box and whisker plot of adjusted score (full data set)

Figure 7.2 shows the range and general distribution of the adjusted scores obtained from all subjects for all treatments used in this study. Identified in the plot are three outliers, falling more than 1.5 times the box length from the 25th

percentile. While outliers are not uncommon, it is important to ascertain the cause of detected outliers prior to commencing a full statistical analysis. As is noted, the three outliers are found at data points 104, 107 and 111. All three of these points are found in the results of subject number seven.

Examination of the range and distribution of results from all individual subjects (figure 7.3) reveals abnormally wide variance in the results for subject seven. As subject seven was not available for interview at the time of the analysis, the author was unable to determine the underlying cause of this variance. However, prior to the first listening session, subject seven had asked whether problems with maintaining attention would be cause for exclusion from the study. Though at the time the principal investigator did not exclude the subject from the study, the data seems to indicate that exclusion may be prudent prior to analysis. Not only is there wide variance in the subject's scores, but it can also be seen that the median of these scores is below the 25th percentile of any other subject.

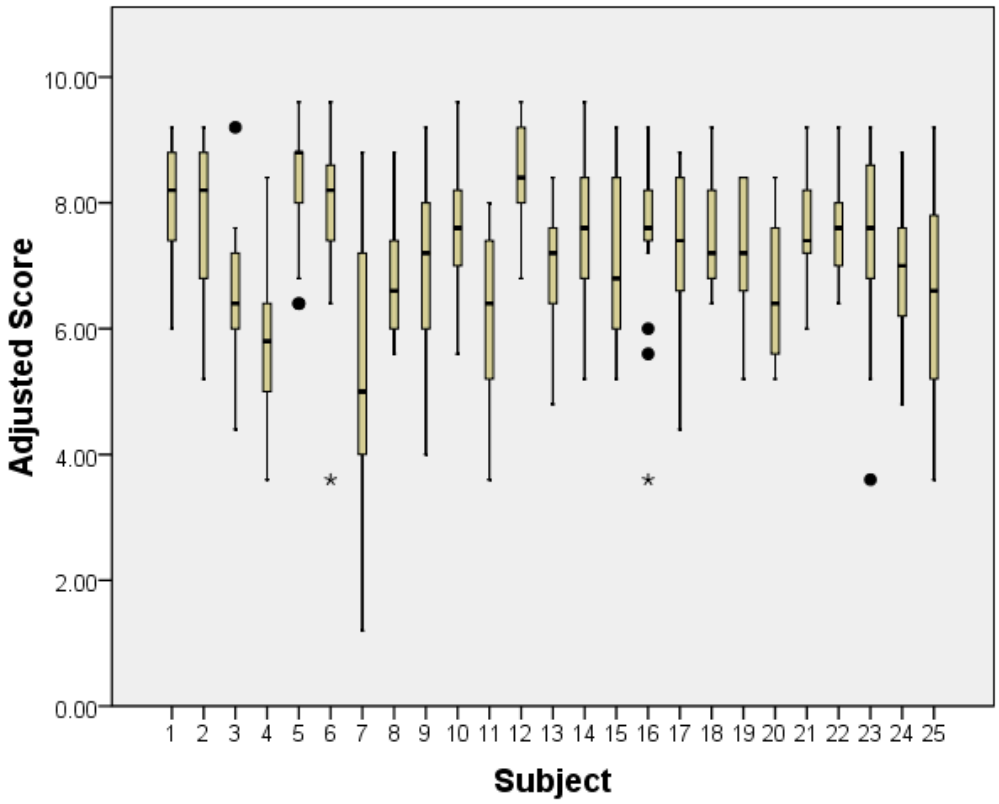


Figure 7.3 Box and whisker plot of adjusted score vs. subject (full data set)

As it was not possible to definitively determine the causes of the variance or low scores, two sets of analyses were performed as recommended in [14]. The results from both sets are quite similar, with only minor changes in statistical strength and significance between. As such, the results from the data set that excludes subject seven will be reported in detail (Strat_7 data set); though differences between the two sets of results will be noted in the text. Results from further stratification of the Strat_7 data set, for array type and SNR, will also be reported.

7.7.2 Homogeneity of Variance

Figures 7.4 and 7.5 show the histogram and detrended Q-Q plot for the full set of data. As was the case in the pilot study, the data does not conform to a normal distribution. This conclusion is confirmed by the results of two tests for the homogeneity of variance (Table 7.2) as, for the Kolmogorov-Smirnov and

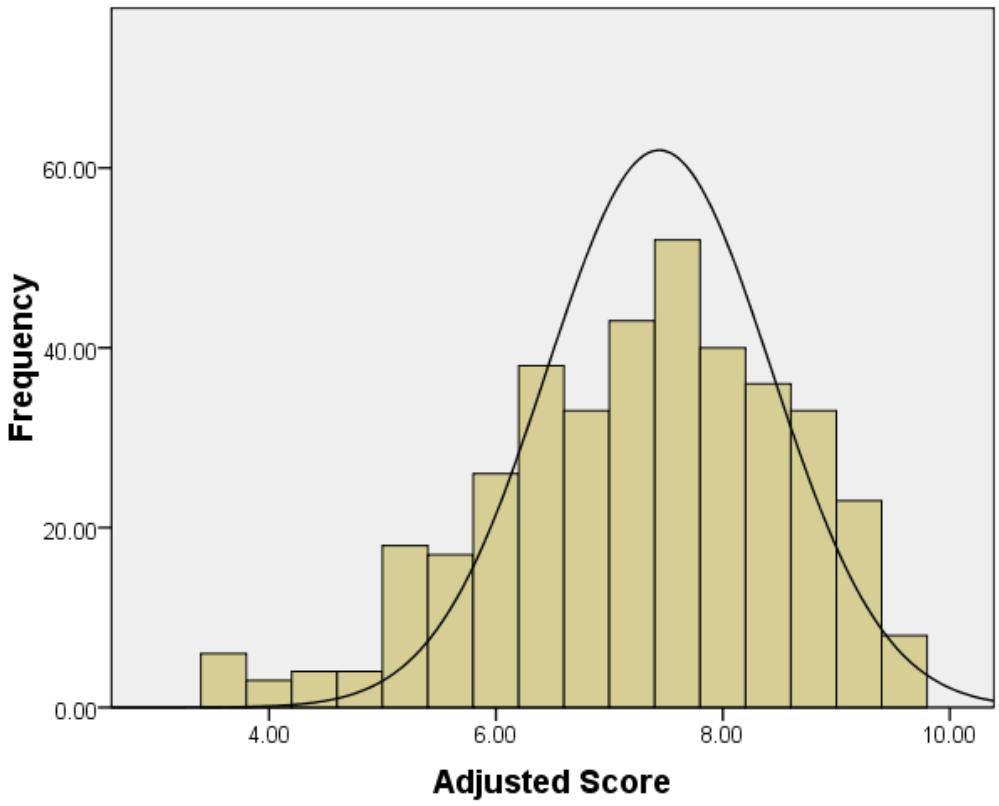


Figure 7.4 Histogram of adjusted score (Strat_7 data set)

Shapiro-Wilk tests, a high significance value would indicate normality while a low value (e.g. $p < 0.001$) indicates a non-normal distribution.

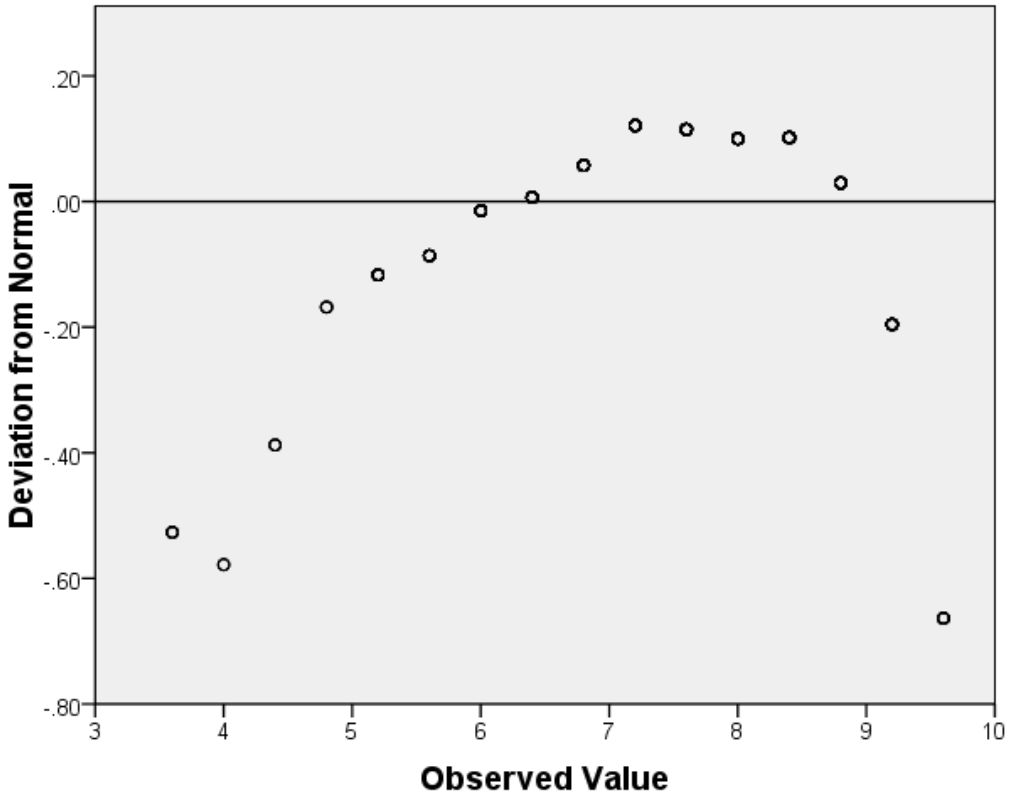


Figure 7.5 Detrended normal Q-Q plot of adjusted score (Strat_7 data set)

Data Set	K-S Stat.	K-S Sig.	S-W Stat.	S-W Sig.
Original	0.104	< 0.001	0.962	< 0.001
Strat_7	0.102	< 0.001	0.969	< 0.001
Strat_7-Vert	0.094	< 0.001	0.968	< 0.001
Strat_7-Hor	0.111	< 0.001	0.959	< 0.001
Strat_7-SNR0	0.085	0.002	0.978	0.004
Strat_7-SNR6	0.012	< 0.001	0.946	< 0.001

Table 7.2 Tests for homogeneity of variance for the six data sets: Kolmogorov-Smirnov and Shapiro-Wilk statistics

It is clear that none of these data sets is normally distributed. As such, statistical analysis using non-parametric tests is indicated. The Kruskal-Wallis test will therefore be used, though the results of ANOVA will also be included.

Additionally the author has been made aware of, and will thus employ, a form of log-linear analysis called multiway contingency tables analysis (MCTA) [120, 165, 180].

7.7.3 Analysis of Strat_7 Data Set

The results of the first order analysis of the Strat_7 data are shown in Table 7.3. Subject is once again, and unsurprisingly, found to have an effect on results.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Subject	4.811	< 0.001	89.946	< 0.001
Delay Time (0 – 40 ms)	4.326	0.005	11.329	0.010
Delay Time (0 – 10 ms)	0.565	0.569	1.094	0.579
Delay Time (10 – 40 ms)	9.239	0.003	7.747	0.005
Array Type	3.789	0.052	3.488	0.062
SNR	156.175	< 0.001	119.157	< 0.001

Table 7.3 Results of ANOVA and Kruskal-Wallis tests for first-order effects of experimental variables on adjusted score (Strat_7 data set)

Also, the variable SNR shows high strength and significance. As the first order effect of SNR was included as a dummy variable, to check the function of the experiment, confidence in the testing methodology and execution is raised. The first order effect of array type was also included as a dummy variable. Though the box plot (figure 7.6) of adjusted score vs. array type does indicate some difference between the two arrays, results from this initial analysis of the effect of array type do not quite meet the standard for sufficient statistical significance ($p = 0.062 > 0.05$).[12] It is possible that said effect may become significant when further stratified data sets are analyzed.

[12] For the analysis of the data set that included subject seven, both the ANOVA and Kruskal-Wallis tests found array type to be a significant variable ($p = 0.044$ and 0.043 respectively))

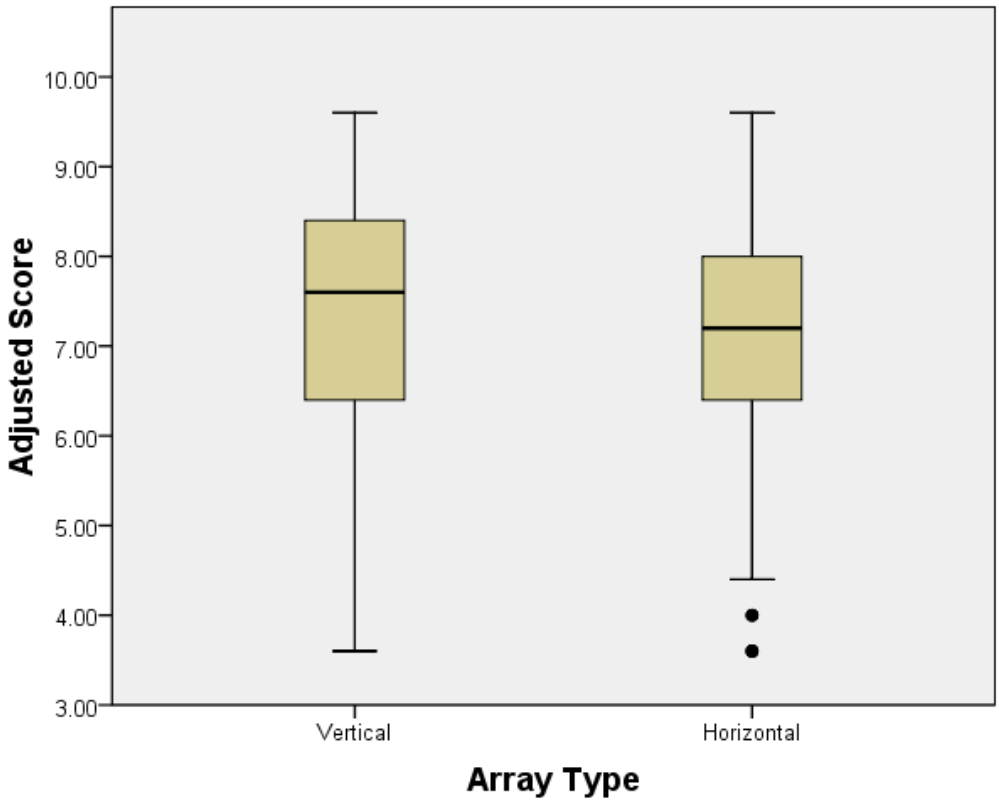


Figure 7.6 Box and whisker plot of adjusted score vs. array type (Strat_7 data set)

Delay time is seen to have a significant effect for the ranges of 0 ms–40 ms and 10 ms–40 ms, however no significant effect is found in the range 0 ms–10 ms. As can be seen in figure 7.7, the results for delay times of 0 ms and 5 ms are nearly identical. The results for the 10 ms treatments show a slight reduction in overall variance, though the median value is lower than those found for the 0 ms and 5 ms treatments.

As seen in table 7.4, neither strength nor significance is found for any of the three 2-way interactions. It was suggested in chapter 6.7.2 that the acquisition of more data points could possibly reveal significant interaction effects, however 24 points per treatment (as opposed to twelve) has not yielded further clarity.

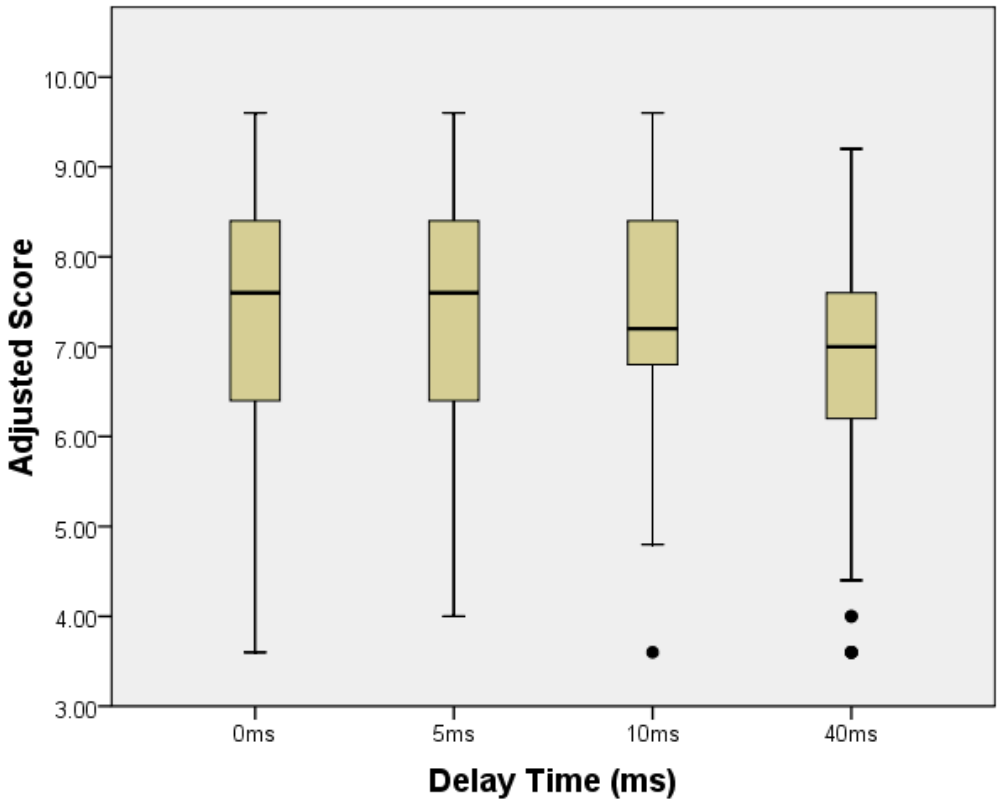


Figure 7.7 Box and whisker plot of adjusted score vs. delay time (Strat_7 data set)

Variable	ANOVA F-Stat	ANOVA Sig.
SNR × Delay Time	0.592	0.621
Delay Time × Array Type	1.776	0.151
SNR × Array Type	2.582	0.109

Table 7.4 Results of ANOVA for second-order effects of experimental variables on adjusted score (Strat_7 data set)

7.7.4 Stratification by SNR

The work of Lochner and Burger [107] showed that the effects of rollover interact with SNR (see chapter 2.1.1). Also, the work of Steeneken and Houtgast [167], with regard to the measured effects of delay and SNR on the MTF, indicates that the effect of delay may be mitigated when higher SNR values are

used (see chapter 2.2.4). Based on the findings from these studies, and considering the relative strength of the effect of SNR, it has been proposed by the author that said effect could serve to obscure the effects of array type and delay time. Thus the Strat_7 data set was further stratified by SNR into two data sets: Strat_7-SNR6 and Strat_7-SNR0. The results of the analyses for these data sets are shown in tables 7.5 and 7.6.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (0 – 5 ms)	3.354	0.070	3.000	0.068
Delay Time (0 – 10 ms)	1.83	0.164	3.331	0.189
Delay Time (0 – 40 ms)	4.599	0.004	14.821	0.002
Delay Time (5 – 10 ms)	0.707	0.403	1.050	0.305
Delay Time (5 – 40 ms)	7.373	0.001	14.974	0.001
Delay Time (10 – 40 ms)	8.069	0.006	7.397	0.007
Array Type	9.437	0.002	8.210	0.004

Table 7.5 Results of ANOVA and Kruskal-Wallis tests for effects of delay time and array type on adjusted score (Strat_7-SNR6 data set)

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (0 – 5 ms)	0.014	0.907	0.066	0.797
Delay Time (0 – 10 ms)	0.067	0.935	0.223	0.895
Delay Time (0 – 40 ms)	2.656	0.05	6.819	0.078
Delay Time (5 – 10 ms)	0.130	0.719	0.282	0.595
Delay Time (5 – 40 ms)	03.550	0.031	6.055	0.048
Delay Time (10 – 40 ms)	5.950	0.017	5.604	0.018
Array Type	0.213	0.645	0.096	0.757

Table 7.6 Results of ANOVA and Kruskal-Wallis tests for effects of delay time and array type on adjusted score (Strat_7-SNR0 data set)

The effect of delay time is more easily identified in the higher SNR condition. Both strength and significance are increased for the 0 ms–40 ms, 5 ms–40 ms and 10 ms–40 ms ranges. An illustration of these increases can be seen in figure 7.8. For the 6 dB SNR data, a comparison of medians shows a downward trend in intelligibility scores as delay time increases. This trend is not seen for the 0 dB SNR data.[13]

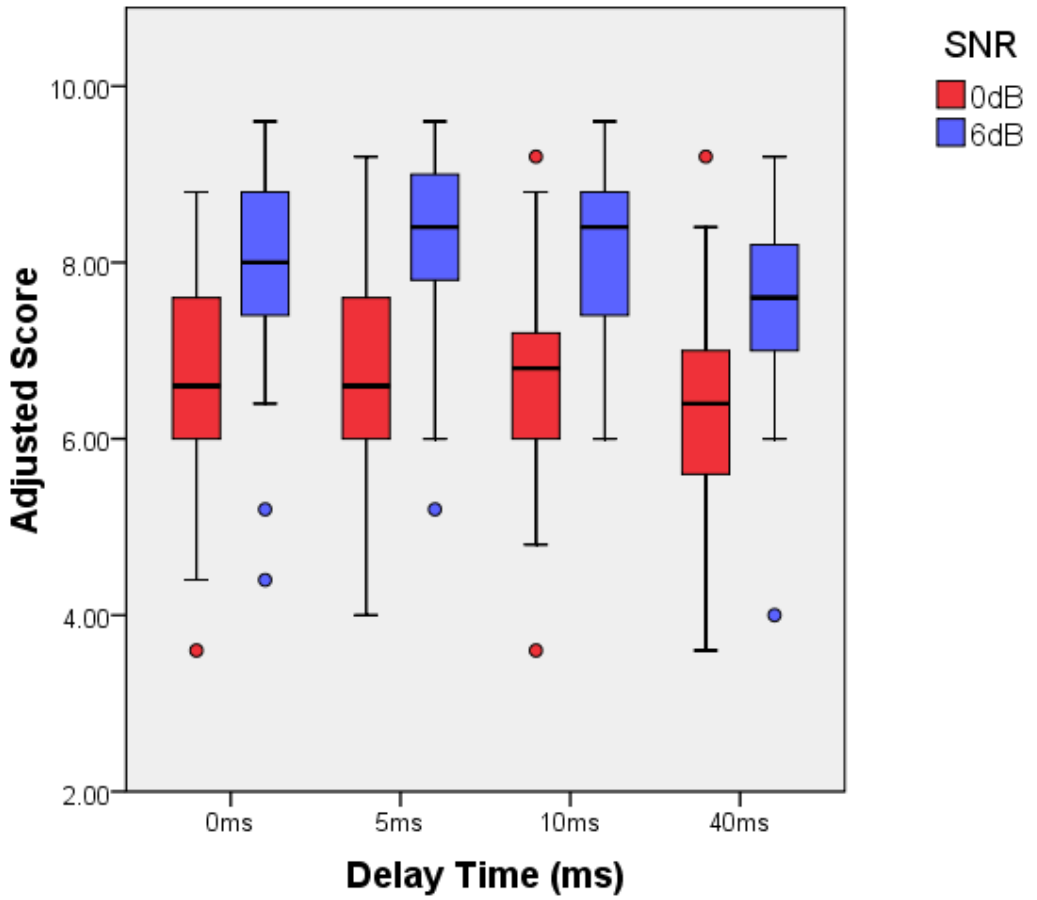


Figure 7.8 Box and whisker plot of adjusted score vs. delay time, by SNR (Strat_7 data set)

It is possible that the addition of more data points could refine the results for the lower SNR data, removing some degree of variance, thus revealing a relationship between delay time and scores. However, as was noted in chapter 6.7,

[13] The analysis of the data set that includes subject seven revealed that, for the SNR0 condition, the only significant effect was delay time in the range 10 ms–40 ms)

variance generally increases as SNR decreases [156]. It is therefore conceivable that the obfuscation of the effects of other variables is the result of the same phenomenon that causes the inherent variance found in low SNR conditions. In other words, with regard to speech intelligibility, it appears that the effects of higher noise levels mask the effects of delayed multiple arrivals.

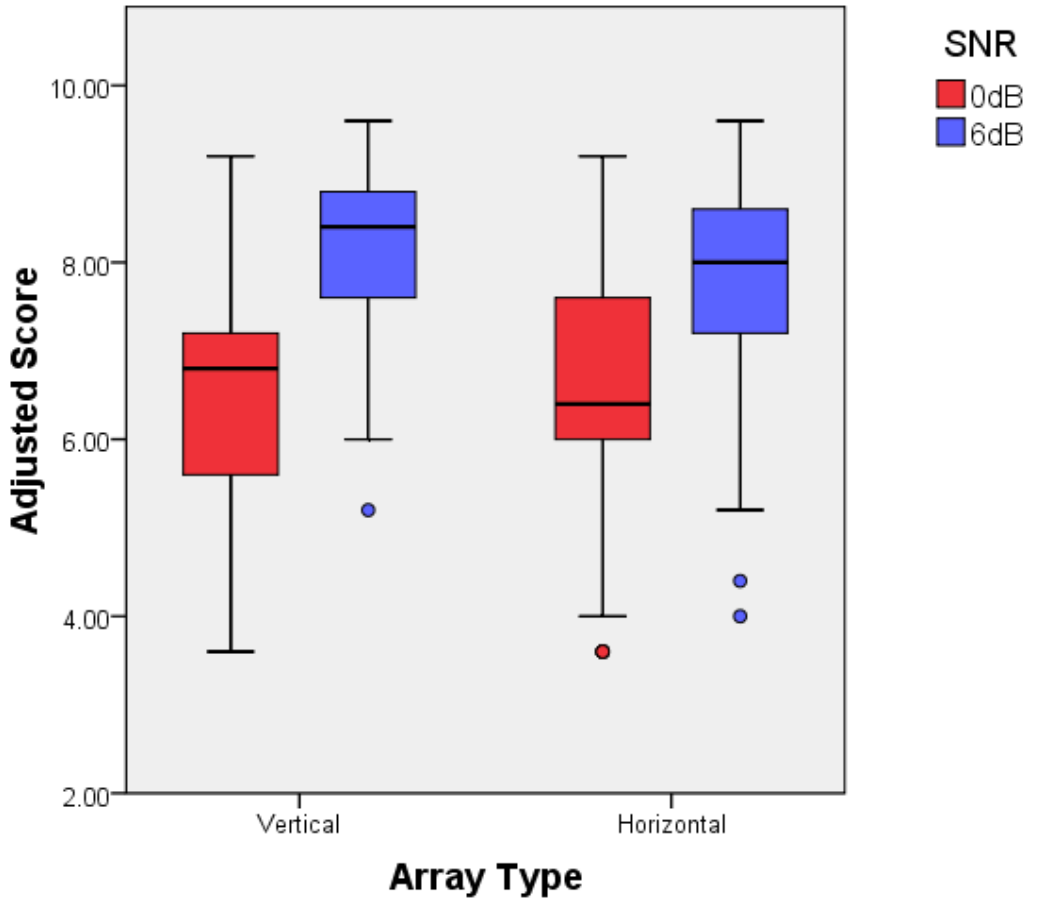


Figure 7.9 Box and whisker plot of adjusted score vs. array type, by SNR (Strat_7 data set)

A similar trend is found to exist for the effects of array type, shown in figure 7.9. Contrary to the results from the analysis for the full Strat_7 data set, array type is seen to have a clear effect at the higher SNR condition ($\chi^2 = 8.210$, $p = 0.004$). From a review of the works of Holman [74] and Shirley et al. [162, 163], it was anticipated that the scores obtained from the vertical array type would be higher. Of particular interest in these results is that, while the scores for the vertical array are higher for the 6 dB SNR, there is no distinguishable difference

in scores for the 0 dB SNR. These results would suggest that the gains in intelligibility usually afforded by the use of a center channel are reduced if not negated by the presence of higher noise levels.

7.7.5 Stratification by Array Type

The results of the various two-way ANOVA tests for the complete Strat_7 data set did not reveal significant interactions between variables. However, as seen in the previous section, it can be possible to divine knowledge of variable relationships through stratification of the data set by a single variable.

The Strat_7 data set was again divided into two data sets, this time according to array type: Strat_7-Vert and Strat_7-Hor. Tables 7.7 and 7.8 show the results of analyses carried out on these two data sets.

First, one can see that the strength of the effect of SNR is greater for the vertical array. This is in agreement with the results from the previous section, shown in figure 7.9.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (0 – 5 ms)	3.565	0.062	3.292	0.070
Delay Time (0 – 10 ms)	1.842	0.162	3.142	0.208
Delay Time (0 – 40 ms)	1.931	0.126	5.182	0.159
Delay Time (5 – 10 ms)	0.735	0.394	0.749	0.387
Delay Time (5 – 40 ms)	2.312	0.103	4.242	0.120
Delay Time (10 – 40 ms)	1.635	0.204	1.286	0.257
SNR	107.941	< 0.001	72.526	< 0.001

Table 7.7 Results of ANOVA and Kruskal-Wallis tests for effects of delay time and SNR on adjusted score (Strat_7-Vert data set)

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (0 – 5 ms)	0.308	0.058	0.147	0.702
Delay Time (0 – 10 ms)	0.229	0.796	0.222	0.895
Delay Time (0 – 40 ms)	4.144	0.007	10.031	0.018
Delay Time (5 – 10 ms)	0.346	0.558	0.186	0.666
Delay Time (5 – 40 ms)	4.951	0.008	8.484	0.014
Delay Time (10 – 40 ms)	9.189	0.003	8.084	0.004
SNR	56.628	< 0.001	48.420	< 0.001

Table 7.8 Results of ANOVA and Kruskal-Wallis tests for effects of delay time and SNR on adjusted score (Strat_7-Hor data set)

From these results it would appear that delay time does not have a significant effect on intelligibility scores for the vertical array configuration. Results for the horizontal configuration, on the other hand, indicate that there are significant differences in scores for the delay time ranges of 0 ms–40 ms, 5 ms–40 ms and 10 ms–40 ms.

A graphical comparison of the intelligibility scores from both arrays is shown in figure 7.10. There appears to be good agreement between the scores for both array types for the 0 ms and 5 ms conditions. At 40 ms it is clear that there is a difference in scores for the two arrays. While the statistical analysis of the effect of the 5 ms–10 ms delay range for the horizontal array did not show significance ($\chi^2 = 0.186$, $p = 0.666$), an inference could be made from the plot. It appears that the separate effects of delay time on the two arrays begin to diverge for delay times greater than 5 ms, with effects becoming significant somewhere between 10 ms and 40 ms.

These results are surprising. The experiments reported by Haas [67] indicates that the critical delay difference, required for an echo to disturb listening, increases as the angle of echo incidence deviates from front/center. While the vertical array configuration does employ a loudspeaker located at an elevated angle, Haas' results indicate that elevation has less impact on lengthening the

critical delay difference than lateral angular-offset. As such, one would expect to see the scores for the vertical array decline before those from the horizontal array.

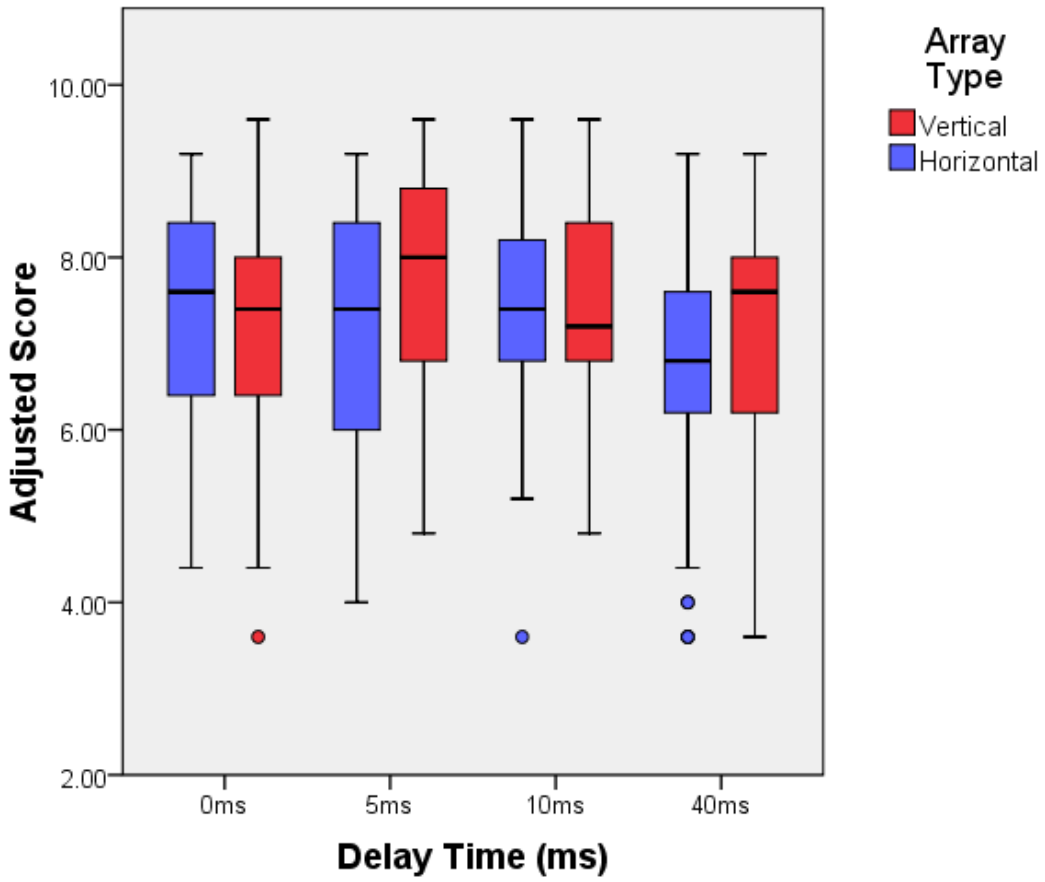


Figure 7.10 Box and whisker plot of adjusted score vs. delay time, by array type (Strat_7 data set)

One possible explanation for this disparity lies in the difference between the goals of the studies of Haas and the current research project. The investigation of echo detection/disturbance is not the same as the investigation of intelligibility. As mentioned in chapters 2.1.1 and 2.2.2.2, there is debate surrounding the relationship between the fusion (post-masking) of early reflections and intelligibility. The results of the current study would seem to agree with other researchers [41, 111] indicating that fusion plays different roles for echo perception and intelligibility.

Another possible explanation, along the same lines, has to do with the fact that the current study does not focus explicitly on angle of incidence; rather it

examines two different types of arrays – each differing in orientation and focus. Rather than being purely an issue of monaural vs. binaural hearing, it is equally likely that the observed effect of array type may indicate a difference in the compound effect that also includes effects from array orientation and focus. This is an area of interest for further study.

Also, based on the work of Mochimaru [130] and others [11, 41, 170], it is unlikely that delay time would have no significant effect for the vertical array type. What is more likely is that the effect of delay time is stronger for the horizontal geometry, and that tests on the data from this study were merely unable to find significance for the less-strong effect in the vertical geometry.

7.7.6 Multiway Contingency Tables Analysis (MCTA)

MCTA is a form of log-linear analysis commonly used with categorical variables. As the name would suggest, MCTA is a method for analyzing relationships between multiple independent and dependent variables, wherein the variables are organized into a categorical table and the results can be expressed in terms of frequencies (e.g. times passed vs. failed). Though similar in principle to the χ^2-test for association, which is often performed on two-way contingency tables (i.e. a 2-by-2 square), MCTA is capable of handling the analysis of larger contingency tables [165, 180].

The goal of MCTA is to form a model that predicts variance in the results using the least number of factors. Starting with the highest order factor (e.g. a 4-way interaction), and working backward through the hierarchy of lower-order factors, the MCTA process eliminates factors that do not have a significant effect on the model's prediction accuracy. For the example of a 4-way contingency table, the process removes the 4-way interaction from the model and tests to see whether a significant change is seen in the accuracy of prediction. If the 4-way interaction were found to be significant it would be retained in the model, otherwise it would be removed from the model and the process would then examine the significance of the four 3-way effects. The backward iteration

repeats until no factors can be removed without affecting the prediction accuracy of the model. The remaining factors are referred to as the generating class.

For the current study, the MCTA would contain one 4-way contingency table including array type, SNR, delay time and "pass/fail". Several 3-way tables were also constructed using the four stratified data sets (by SNR and by array type) previously mentioned. An example of a 3-way contingency table is shown in table 7.9. Note that all independent variables have been converted in to categorical form.

Array Type **Vert = 1,** **Hor = 2**	**SNR** **SNR0 = 1** **SNR6 = 2**	**Delay Time** **0 ms = 1** **5 ms = 2** **10 ms = 3** **40 ms = 4**	**Pass / Fail** **Fail = 1** **Pass = 2**	**Frequency**
1	2	1	1	12
1	2	1	2	13
1	2	2	1	6
1	2	2	2	19
1	2	3	1	7
1	2	3	2	18
1	2	4	1	11
1	2	4	2	14
2	2	1	1	12
2	2	1	2	13
2	2	2	1	7
2	2	2	2	18
2	2	3	1	11
2	2	3	2	14
2	2	4	1	20
2	2	4	2	5

Table 7.9 Multiway contingency table for Strat_7-SNR6 data set, pass criterion: adjusted score ≥ 8 (90%)

In order for MCTA to be possible, the scalar result data from this study would need to be transformed into categorical frequency data. In order to do this, a pass/fail criterion must be set. Additional restrictions of MCTA dictate that the choice of this criterion must result in no frequencies being less than one, and not more than 20% of frequencies should be less than 5. Analysis of adjusted score by treatment revealed that scores of 7.6 (88% correct) and 8 (90% correct) would

fulfill these requirements. Analysis was performed using both of these criteria and results are reported in table 7.10.

MCTA of the Strat_7 data set, using the 88% criterion, did not reveal any significant effects aside from the first-order effect of the dummy variable SNR.[14] Analysis of the SNR-stratified data sets (Strat_7-SNR0 and Strat_7-SNR6) revealed that array type was removed from the model early in the process for the SNR0 set, yet was included in the generating class for the SNR6 data set. This result is in agreement with the results of the parametric and non-parametric tests detailed in tables 7.5 and 7.6 and figure 7.9. The results of MCTA (88%) on the array-stratified data determined that SNR was the only effect in the generating class for the vertical array type, while SNR and delay time composed the generating class for the horizontal array type. These results are also in accord with the results from parametric and non-parametric tests (shown in tables 7.7 and 7.8 and figure 7.10).

Data Set	**88% Crit.**	**90% Crit.**
Strat_7	SNR	SNR Delay Time
Strat_7-SNR0	None	None
Strat_7-SNR6	Array Type	Delay Time Array Type
Strat_7-Vert	SNR	SNR
Strat_7-Hor	SNR Delay Time	SNR Delay Time

Table 7.10 Results showing generating class for MCTA using 88% and 90% pass criteria

Using the 90% pass criterion, MCTA of the full Strat_7 data set revealed that both SNR and delay time compose the generating class. Analysis of the two SNR-stratified data sets revealed that delay time and array type were significant for the SNR6 condition but not for SNR0. Analysis of the array-stratified data sets found delay time significant for the horizontal but not vertical array type.

The use of the two different pass criteria has shown that the choice of the criterion point has an effect of the sensitivity of the MCTA test. The specific

[14] For all MCTA tests, the maximum number of iterations was set to 10 and the criterion for significance was set to 0.05.

implications of these differences are unclear (e.g. whether delay time alone is incapable of reducing scores below 88%). What is clear is that the results of the 90% pass criterion tests are in agreement with the results found earlier in this chapter, namely: 1) Delay time and array type are significant effects, though these effects are reduced / obscured at low SNR levels, and 2) delay time is found to have a significant effect for the horizontal array geometry.

7.8 Discussion

When compared to the results of the pilot study, it can be seen that the increase in size of the subject population has led to greater statistical strength and significance in the results of the current study. The increased number of data points has also allowed for significant results to be found in stratified analyses.

In addition to evaluating the first- and second-order effects of the experimental variables, it was also of interest to evaluate the effects of potential nuisance variables.

7.8.1 Training & Effects of Presentation Order

As was the case with the pilot study, it was desired to know whether subjects received a sufficient amount of training prior to beginning the battery of subjective evaluations. Considering that the presentation order of variable treatments was randomized for each subject, if the analysis of the effects of presentation order on scores were to reveal an improvement in subject performance, this would be an indication of learning. As mentioned in chapter 6.8.2, learning during the first few sessions would indicate inadequate training, and learning over the course of all 16 sessions could indicate that an insufficient number of word lists were used. The results of this analysis are shown in table 7.11 and figure 7.11.

Inspection of the plot reveals no trend over the entire span of sessions. However there is a slight upward trend in medians for the first three sessions, suggesting that learning may be present over the short term. One might also infer

some degree of a decrease in variance beyond session number nine, which could be an additional indicator of the quality and sufficiency of subject training.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Order (all)	0.939	0.521	13.004	0.602
Order (1-4)	0.614	0.608	1.574	0.665

Table 7.11 Results of ANOVA and Kruskal-Wallis tests for effects of presentation order on adjusted scores (Strat_7 data set)

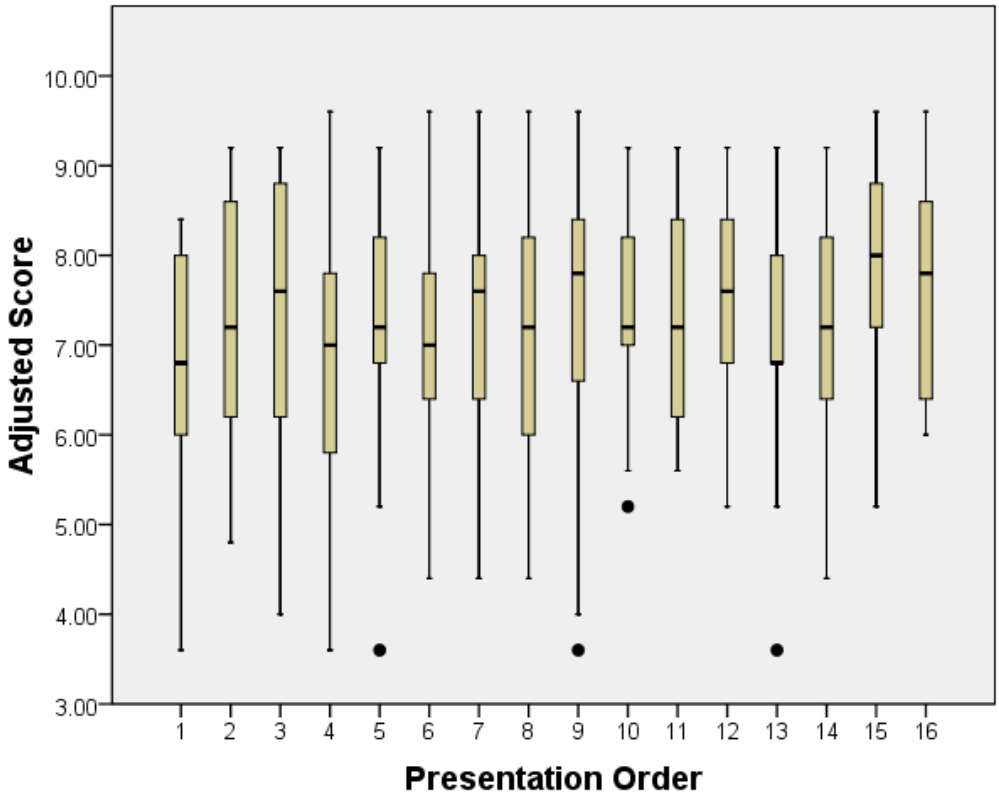


Figure 7.11 Box and whisker plot of adjusted score vs. presentation order (Strat_7 data set)

Through examination of results of statistical tests (shown in table 7.11), it is clear that there is not sufficient evidence to indicate that learning, whether short or long term, was a factor in this study. This analysis was also performed for presentation order ranges of Order(1-8), Order(1-3), Order(1-2) and Order(9-16) yielding similarly insignificant results.

7.8.2 The Effects of Word List

As mentioned previously, it was desired to study whether the choice of word list would have an effect on scores, as this could affect large reductions in testing time and cost. Results from the pilot study in this regard were inconclusive however the results from the current study are not.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Word List	9.879	< 0.001	20.213	< 0.001

Table 7.12 Results of ANOVA and Kruskal-Wallis tests for effects of word list on adjusted scores (Strat_7 data set)

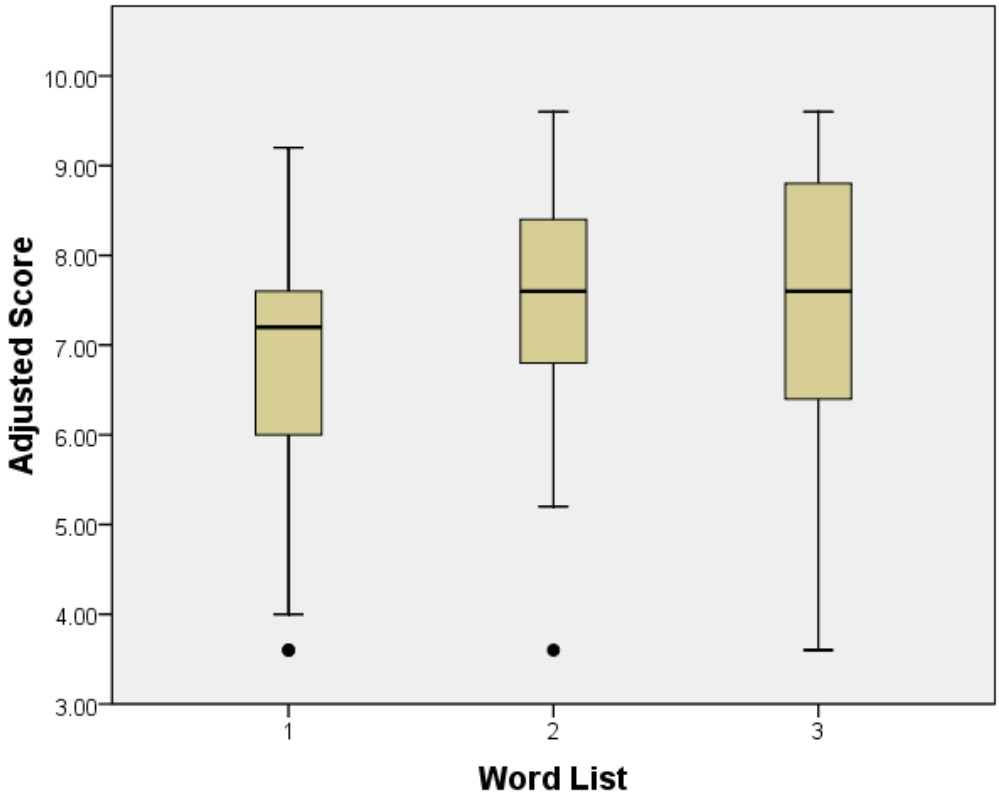


Figure 7.12 Box and whisker plot of adjusted score vs. word list (Strat_7 data set)

As seen in table 7.12, there is little doubt that word list does indeed have an effect. Specifically, as seen in figure 7.12, though the results from lists two and three (MRT lists D and F, respectively) appear quite similar, there is a clear difference between those and the results from list one (MRT list A). This result is

curious considering that studies noted in the literature have used reduced sets of word lists (e.g. [9, 73, 122, 123]) or assigned different word lists to different treatments (e.g. [21, 95, 99, 136, 138]).

Inspection of the matrix of subject responses showed that subjects consistently selected the incorrect response for word number 48 in list A. The correct word "BAT" was mistaken for the word "BATH" - the correct answer for list 3 (MRT list F). The author listened to the two sound files used to create these stimuli and indeed, even with no added noise, the two words were virtually indistinguishable. As it was possible, though unlikely, that an error was made during stimulus preparation, a visual inspection of the two waveforms and spectrographs was performed on samples from the same treatment, verifying that the two words were indeed different.

On the adjusted score scale ($-10 \leq Ra \leq 10$), each word missed corresponds to a 0.4 point drop in score. Such a drop, as would be found when subjects routinely miss the word *bat*, could account for the differences in scores seen between word list number one and the other two lists. From the analysis, it is evident that the assignment of different word lists to different treatments constituted an insufficiently controlled variable or, at best, imbued the additional error associated with a fractional factorial design. Thus, unfortunately, future studies should employ a fully populated word list by treatment matrix.

The question remains as to the validity of the results of the current study. As Bech and Zacharov explain, the lack of control of a variable constitutes a disturbing (or nuisance) variable [14]. They state that the method for dealing with such variables it either to control them or employ randomization that will break relationships between the nuisance variable and independent variables. They further state that such randomization increases a statistical model's error component and/or residual variance.

As mentioned, for this study word lists were randomly assigned to treatments for each subject, using a uniform distribution. Thus any increase in error and/or variance due to the difference between word lists would be spread

across all treatments. Even so, an attempt at post-hoc control of this nuisance variable is mandated if one is to have faith in the results of the study.

The offending data, word number 48, was censored from the results of all word lists and the full statistical analysis presented in this chapter was repeated. The data resulting from the removal of the 48th word contained 49 words each, and were thus identified with the marker "49w" – compared to the original 50 word ("50w") data.

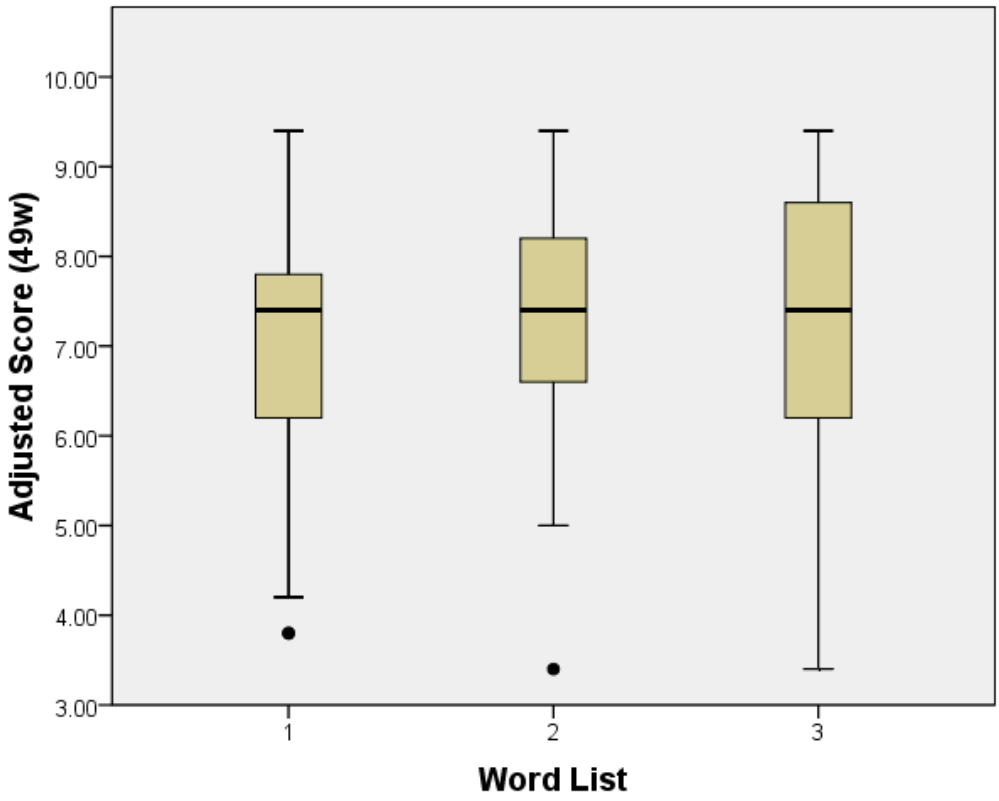


Figure 7.13 Box and whisker plot of adjusted score vs. word list (Strat_7, 49w data set)

Figure 7.13 shows the box and whisker plot of the effect of word list for the 49w data set. It would appear that the censoring of the results for the 48th word in all data effectively compensates for the differences observed between word lists, given that the means for the three word lists are now identical. The results shown in table 7.13 confirm that the effects of word list are no longer found to be significant. Differences do exist in result variances between the three word lists and, as the statistical strength and confidence are not negligible (χ^2 =

4.625, p = 0.099), word list could still be viewed as an insufficiently controlled nuisance variable. This should be taken into account in further studies.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Word List	1.926	0.147	4.625	0.099

Table 7.13 Results of ANOVA and Kruskal-Wallis tests for effects of word list on adjusted scores (Strat_7, 49w data set)

Aside from the effects of word list, the majority of results obtained did not differ between the 50w and 49w data sets in the current study. For effects that were found to be significant with the 50w data set, the removal of the residual variance caused by the 48th word merely increased the statistical strength and confidence.[15] For most of the effects that were not found significant in the analysis of the 50w data set, the same was found for the 49w data set. Two notable exceptions to this were the effect of array type for the Strat_7 data set and the effect of delay time (0 ms–5 ms) in the Strat_7-Vert data set.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Array Type (50w)	3.789	0.052	3.488	0.062
Array Type (49w)	6.365	0.012	6.323	0.012

Table 7.14 Differences between statistics for array type between the 50w and 49w data sets (Strat_7 data set)

As seen in table 7.14, the effect of array type is found to be significant in the results of the 49w data set. While this result is different from that obtained through analysis of the 50w data set, it is not entirely surprising. As noted in chapter 7.7.4, it is known from the works of Holman [74] and Shirley et al. [162, 163] that the scores obtained from the vertical array type should be higher than those for the horizontal array type. The statistics obtained for the 50w data set suggest that an effect is present, though the absolute confidence criterion (5%) was not met. It is clear that the reduction of the error component due to the

[15] For example, the analysis of the full data set revealed that the statistics for delay time (10 ms–40 ms) shifted from χ^2 = 7.877, p = 0.005 (50w) to χ^2 = 9.121, p = 0.003 (49w).

nuisance variable has increased the clarity of the statistical test, rendering a clearly significant effect where borderline significance was previously found.

Likewise, the reduction in residual variance has yielded significance regarding the effect of short delay times. As seen in table 7.15, there is sufficient evidence to conclude that there is a difference between scores for the 0 ms and 5 ms conditions for the vertical array type.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (0 – 5 ms) (50w)	3.565	0.062	3.292	0.070
Delay Time (0 – 5 ms) (49w)	4.007	0.048	3.925	0.048

Table 7.15 Differences between statistics for 0 ms–5 ms delay times between the 50w and 49w data sets (Strat_7-Vert data set)

7.8.3 Very Short Delay Times

It can be seen in figure 7.10 that scores are generally lower for the 0ms condition than for the 5 ms condition. What is interesting is that these differences are only found to be significant for the vertical array type. In fact, as shown in table 7.16, it would appear that this effect is completely negligible for the horizontal array type.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (0 – 5 ms) (Horizontal)	0.058	0.810	0.005	0.941
Delay Time (0 – 5 ms) (Vertical)	4.007	0.048	3.925	0.048

Table 7.16 Differences between statistics for 0 ms–5 ms delay times between the Strat_7-Hor and Strat_7-Vert sets (49w data sets)

This result, a difference in scores between the 0 ms and 5 ms delay times, was wholly unexpected and sparked a thorough review of the research project. It

was discovered that an endemic calibration error was made during the stimulus capture process. As seen in figure 3.8 (chapter 3.3), the calibration microphone used for time alignment was placed just above the head of KEMAR. While this would not affect the time alignment of loudspeakers in the horizontal array geometry, it would affect alignment for the vertical array.

The difference in distance between the source and receiver would be different for the measurement microphone and KEMAR's ears. The distances between the source and the two receivers would be approximately equal for the front-fill loudspeaker, however the distance to the main loudspeaker (Meyer UPA-1p) would be approximately 10 cm (≈ 0.3 ms travel time) shorter for the case of the measurement microphone. Thus for the 0 ms condition, the summation of the signals from the main and fill loudspeakers at the ears would actually have a 0.3 ms offset, resulting in a ½ λ notch near 1.6 kHz, 3/2 λ notch at 4.8 kHz, 5/2 λ notch at 8 kHz, etc.

It is somewhat unfortunate that this error makes it impossible to determine whether there would be a significant difference in intelligibility scores between the 0 ms and 5 ms conditions for the vertical array type. However, if a fortuitous error is indeed possible, this may have been just such a folly. Based on previous studies and personal communicae, it was never anticipated that a 5 ms offset in arrivals would have a significant impact on intelligibility [11, 28, 131, 125, 144]. As mentioned in chapter 2.2.3, the two popular methods for loudspeaker alignment are absolute alignment and intentional misalignment. What the results found in this study indicate is that, at least for vertically oriented point-destination arrays, very short delay times (e.g. 0.3 ms) do indeed have a greater negative impact on intelligibility than only moderately short (e.g. 5 ms) delay times. These subjective results confirm the findings of C. Davis [39] that were obtained through objective (RaSTI and $\%AL_{cons}$) methods. Coincidently, the calibration error found in this study corresponds to the exact delay offset used by Davis.

7.9 Conclusions

As with the pilot study, this first phase of the main study had several main points, including the evaluation of a variety of hypotheses. This section will address these questions and tasks, as well as indications of possible directions for the final phase of the main study.

7.9.1 Hypotheses

Ho1: Delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

From the results of both parametric and non-parametric tests, there is sufficient evidence to reject this null hypothesis. The results indicate that the variable delay time does have a significant effect on intelligibility scores. The tests indicate that this variable begins to have an effect somewhere in the region between 10 ms and 40 ms. The tests further indicate that the effects of very short delay times also have a negative impact on intelligibility scores.

Ho2: Signal-to-noise ratio does not affect the intelligibility of speech reproduced by a sound system.

As mentioned, SNR has a well known effect on intelligibility scores. The results found in this study concur, allowing for the clear rejection of this null hypothesis.

Ho3: Array geometry does not affect the intelligibility of speech reproduced by a sound system.

For the data set including subject number seven, results of parametric and non-parametric test indicate that there is sufficient evidence to reject this null hypothesis. When the data from the 48^{th} word in each word list is censored, the results from the Strat_7 data set also substantiate the rejection of the null hypothesis for this first-order effect.

Ho4: (interaction) Signal-to-noise ratio does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

The second-order ANOVA results do not indicate that an interaction effect exists between SNR and delay time. However independent analysis of the two SNR-stratified data sets reveals that an interaction does exist. It was observed that higher noise levels (lower SNR) obscure much of the effect of delay time on intelligibility. The results found in this study provide sufficient evidence to reject this null hypothesis.

Ho5: (interaction) Signal-to-noise ratio does not affect how array geometry affects the intelligibility of speech reproduced by a sound system.

The second-order ANOVA results do not indicate that an interaction effect exists between SNR and array type. Again, through stratification by SNR, it was found that there is sufficient evidence to reject the null hypothesis. The effects of array type on intelligibility scores is clearly diminished for lower SNR's.

Ho6: (interaction) Array geometry does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

Once again, the second-order ANOVA was unable to detect an interaction. Through stratification by array type, it was found that the effect of longer delay times on intelligibility is greater for the horizontal array geometry. This result differs from the suggested, though not significant, trend seen in the pilot study. The result also would seem to deviate from the expected, given the results of the studies by Haas [67]. It was also seen that the effects of very short delay times are also significant, at least for the vertical array geometry. Indubitably, this relationship warrants additional investigation.

7.9.2 Main Points of this Study

1) Attempt to further reduce the number of treatments of interest

At this point, the first- and second-order effects involving SNR appear well established. For further study it is recommended that only one SNR value be used, thus removing one dimension from the variable treatment matrix. As the effects of SNR appear strong enough to overwhelm the effect of the other experimental variables, it is also recommended that this variable value be excluded from future studies. From the set of three useful SNR values found in the pilot study, the remaining two would be 6 dB and 3 dB. While it would likely be easier to see other variable effects if the 6 dB value were to be used, use of the 3 dB value could provide additional data for use in a between-subjects/studies analysis.

Given the propensity for low SNR values to obscure the effects of other variables, a greater number of subjects and/or data points will likely be required if the 3 dB SNR is to be used in further studies.

2) Determine if the effects of delay time values in the region less than 20 ms are still insignificant with a larger test group.

It has been seen that delay time begins to have an effect somewhere in the range between 10 ms and 40 ms. These results are in agreement with those of the pilot study. It is clear that the 5 ms variable value can be excluded from further study. It would however be useful to include the 30 ms value, as this would provide increased time resolution for determining the amount of delay required to affect intelligibility.

The effects of very short delay times (e.g. under 5 ms) remain of interest. However, as stimuli for this range of delay times were not captured for this research project, such effects can not be addressed at this time.

3) Test the validity of using only one MRT word list per treatment

Results indicate a clear difference between the scores obtained using MRT list A and those obtained from using lists D and F. Analysis of the results from each of these word lists indicated one possible source of variance. However as not all of the variance can be accounted for, it is recommended that a full factorial design (list by treatment) should be used in future studies in the interest of controlling this potential nuisance variable.

4) Evaluate hypotheses

As can be seen in the previous section, all six of the hypotheses posed at the beginning of this study have been evaluated, resulting in the discovery of several interesting variable interactions.

7.9.3 Parting Thoughts

The original design of this research project included 168 possible treatments. The pilot study was successful in reducing this number significantly. The first phase of the main study has also been successful in that variable effects and relationships have been identified and the number of variable treatments of interest has been further reduced.

Several potential nuisance variables have been found insignificant. The level of training provided for subjects continues to appear adequate. No signs of learning are detected and the randomization of treatments would seem to prevent any biases due to presentation sequence. One significant nuisance variable (word list) has been identified. The control of this variable could lead to further refinement of the experimental design as the research project progresses.

It has been determined that SNR does interact with both delay time and array type. There are opportunities for further research into these interactions – specifically, to determine the critical levels of SNR below which the effects of other factors are masked.

It has also been determined that an interaction exists between array type and delay time. This would be another area of interest in further research,

particularly as the results found were unexpected. In this arena, the question arises regarding differences in the critical delay time required for each type of array to produce a significantly adverse effect on intelligibility.

An additional avenue for potential future research would be to investigate the extent to which a difference in intelligibility exists between the 0 ms and 5 ms delay time conditions. It would appear that, as delay time increases from 0 ms, a region of clear impairment exists, followed by a region of minimal impairment, and followed by another region of clear impairment. As this relates directly to the question of alignment vs. intentional misalignment of loudspeaker arrays, it seems worthy of extended study.

At the end of this, the penultimate study, the ratio of questions answered to new questions encountered has begun to tilt in the favor of the researcher. While several questions remain, the final study will focus in one direction, leaving many of these questions unaddressed.

8. Main Study: Phase 2

After the completion of the pilot study and Phase 1 of the main study, it was clear that there were many potential avenues for further research. In the final study of this research project, it was decided to follow one of these paths – the investigation of the interaction between array type and delay time. By focusing on fewer variable treatments, it would be possible to obtain a greater number of data points per treatment using the same number of subjects, yielding greater power in the statistical tests.

This second phase of the main study would have 2 main points:

1) Evaluate hypotheses
2) Identify potential future research questions that the current study would be unable to address.

It was discovered during the first phase that an issue existed with one of the words in word list number one. It should be noted that this issue was not uncovered prior to the commencement of the current study. As such, analysis of the data obtained in this study will include results for the censored (49w) data set, denoting any differences found between the censored and non-censored data sets.

8.1 Hypotheses

In phase 2 of the main study, a further reduced set of variable treatments was studied, testing two hypotheses.

Ho1: Delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

Employing a different range for the variable delay time, this hypothesis will be tested. If the null hypothesis is rejected, it will allow for the determination of the amount of delay required to affect a negative change in intelligibility.

Ho2: (interaction) Array geometry does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

If the first null hypothesis is again rejected, it will be possible for the second hypothesis to be tested. Using the new range of delay times, this study will attempt to confirm the findings of the first phase of the main study. Further, this study will attempt to determine whether the amount of delay required to affect a negative change in intelligibility differs by array type.

8.2 Study Design

As mentioned, conclusions from the previous study indicate the value of employing a full factorial design with regard to word lists and treatments. Results from the previous study also indicate that the effect of delay time becomes significant somewhere in the range of 10 ms to 40 ms. Thus, values for the variable delay time will span the range of 10 ms to 40 ms, in 10 ms increments.

As it is integral to the evaluation of the second null hypothesis, both values of array geometry will be incorporated into the study.

It has been seen that the effects of delay time and array type are obscured when lower SNR values are used. While this interaction would be of interest for future studies, it was decided to employ only one SNR value for this study. As recommended in the conclusions of the previous study, the 0 dB SNR value will not be used. From the two values that remain, 6 dB and 3 dB, it was decided to use the 3 dB SNR value. As the effects of the lower SNR could have a masking effect on the effects of array type and delay time, it would be likely that a greater number of subjects would be required to observe significant effects. However, as the 3 dB SNR was not used in the previous study, the observation of results similar to those found in the previous study could provide further validation of findings. Additionally, it could be possible in future analysis to compare the results from the current and previous studies in an attempt to garner a preliminary understanding of the effects of SNR on the interaction between array type and delay time.

The variable values to be used for this study are detailed in table 8.1. These values form a 4×1×2 matrix of eight total treatments. As all subjects would evaluate each treatment with each word list, the total number of evaluations (sets) per subject would be 24. The presentation order of these sets was randomized for each of the subjects using the Matlab function "randperm".

Delay Time (ms)	10	20	30	40
Approx. SNR (dB)	3			
Array Geometry	Vert	Hor		

Table 8.1 Variable values used in the second phase of the main study.

8.3 Equipment

Once again, subjects would evaluate binaural recordings via headphone display. The audio playback system used in the current study was the same as in phase one: An IBM Lenovo S10 Ideapad (PC-based) netbook computer, Lexicon Lambda USB audio interface and Sennheiser HD-650 circum-aural headphones. All audio files had a resolution of 44.1 kHz, 16 bit.

The level of playback was calibrated using a hand-held sound level meter (IEC 651 Type II). The meter was attached to a 6 cc coupler to approximate ear canal effects, and positioned on the left headphone using a flat-plate coupler. Playback level was adjusted at the Lambda such that playback of a stimulus set used in this study (SNR = 3 dB) would produce a measured result of 77 dB SPL (C-weighted, slow integration) when no speech signal was present. Again this was the level of the original sound field, as recorded in the original acoustical environment. If it was necessary to perform a hearing acuity test on a subject (as was done during each subject's first listening session), level calibration for the playback software was performed after the completion of the acuity test.

A Matlab program, including graphical user interface (called Loki3), was used for the administration of listening tests (see figure 7.1). As was the case in the previous study, the program would allow subjects to evaluate stimuli at a pace determined by the subject. A subject would press the "PLAY" button, listen to the stimulus, and then attempt to identify the target word from the ensemble of six

possible choices. The program would only allow an individual stimulus file to be played once.

8.4 Subjects

This study used 35 native English-speaking subjects, 25 male and 10 female, ranging from 19 to 37 years of age. In terms of familiarity with the field of audio engineering, 2 subjects were audio professionals, 9 were students in the field and 24 indicated no experience with the field (though some were professional or amateur musicians). Four of the subjects had participated in the previous study. All subjects were verified to have unimpaired hearing (re: 25 dB HL at octave frequencies from 250 Hz to 8 kHz) through the administration of hearing acuity tests [62]. Subjects were compensated $5 USD for each listening session.

Of all of the subjects that began the experiment, two of the subjects were unable to complete the full three sessions. One subject was unable to schedule a third session and the other opted not to continue citing unpleasant nausea resulting from the listening sessions. In the interest of maintaining a full factorial design of subject vs. treatment, the results from this study will include only the data from the 33 subjects that completed all three listening sessions.

8.5 Locations

Listening tests were conducted at two locations. For the first 16 subjects, the location used was the Sound Design Studio at the College Conservatory of Music, University of Cincinnati, OH, USA. For the remainder of the subjects, tests were carried out at a listening facility in Merriam, KS, USA. Both spaces were found to have background noise levels corresponding to NC-30 or less [82] when measured (see chapter 3.1.2 for measurement equipment specifications).

8.6 Procedures

All 33 of the subjects evaluated three MRT word lists for each of the 8 variable treatments. For each subject, this was completed in three sessions, each session containing 8 stimulus sets and taking approximately 36 minutes (45 minutes including breaks) to complete. At the beginning of each subject's first session, a hearing acuity test was administered to verify that the subject did not have a hearing impairment. Each subject was then given written and oral instructions regarding the types of sounds they would be evaluating, operation of the playback device and the method of response (see appendix H for instructions).

The subject would then undergo a training process to become familiarized with the stimuli and testing procedures. The training used for this study involved the evaluation of 4 stimulus sets comprised of: List A delivered under variable treatment 1 (10 ms, Vertical Array), list D under treatment 4 (40 ms, Vertical Array) and list F under treatments 5 (10 ms, Horizontal) and 8 (40 ms, Horizontal). This training set provided subjects with the opportunity to hear all of the individual words that would be presented, and experience the magnitude of the differences between auditory attributes of the various treatments to be used in the study.

Once the training process was complete, the test administrator spoke with the subject to verify that they understood the instructions and operation of the apparatus, and to remind the subject of the importance of taking breaks to minimize fatigue and distraction. As was the case in the previous study, a fixed policy regarding the spacing of breaks was implemented. The subject was instructed that, while they were free to pause the testing process at any point, they would be required to take a 1- to 2-minute break after the completion of every two 50-word sets (approximately every 8–10 minutes).

The subject then began the first session of stimulus evaluation. At the completion of the listening session, the subject was debriefed to determine if they had any concerns about the testing procedure and if they had experienced any perceived hazards or issues with the testing apparatus. As mentioned, one subject

noted unpleasant nausea from listening to binaural recordings. Additionally, some subjects indicated that they had accidentally selected the wrong answer "a few" times.

The subject was then presented with the post-session oral script document, and their next session was scheduled (see appendix I for document). Subsequent listening sessions proceeded in the same manner as the initial session, except with regard to the training process. If the time between an individual subjects' sessions was less than 24 hours, the subject was not required to complete the training process before beginning a session. If the time between sessions was greater than 24 hours, the subject was required to complete one word list from the training process to re-acclimate themselves with the testing apparatus and listening process.

8.7 Results & Analysis

Results of the listening tests were stored by the test software and, as both 50w and 49w data sets would be analyzed, the results were scored as the number of correct responses out of both 50 and 49. As was the case in the previous study, the results were adjusted to account for the probability of chance-guessing using the following equation:

$$\textit{Adjusted Score (Ra)} = \frac{\textit{Correct Responses} - \textit{Incorrect Responses}}{\textit{Number of Choices} - 1} \qquad \text{(Eq. 8.1)}$$

The range of Ra would therefore be: $-10 \leq \text{Ra} \leq 10$ (50w)

$-9.8 \leq \text{Ra} \leq 9.8$ (49w)

As was the case in both of the previous studies, the results obtained from this study were not normally distributed. Table 8.2 shows the results of tests for homogeneity for all of the full and stratified data sets reported (49w data sets). Consequently, the preferred method of analysis would be to use non-parametric tests. Again, the results of the parametric ANOVA tests will also be reported.

Data Set	K-S Stat.	K-S Sig.	S-W Stat.	S-W Sig.
Original	0.106	< 0.001	0.967	< 0.001
Strat_6	0.091	< 0.001	0.983	< 0.001
Strat_6-Vert	0.091	< 0.001	0.985	< 0.001
Strat_6-Hor	0.096	< 0.001	0.984	< 0.001

Table 8.2 Tests for homogeneity of variance for the four data sets generated (49w censored data)

8.7.1 Exclusionary Criteria

As was the case in the previous study, the variance and means were examined for each subject. Figure 8.1 shows the range and general distribution of the adjusted scores obtained from all subjects for all treatments used in this study. Identified in the plot are a number of outliers, falling more than 1.5 times the box length from the 25th percentile.

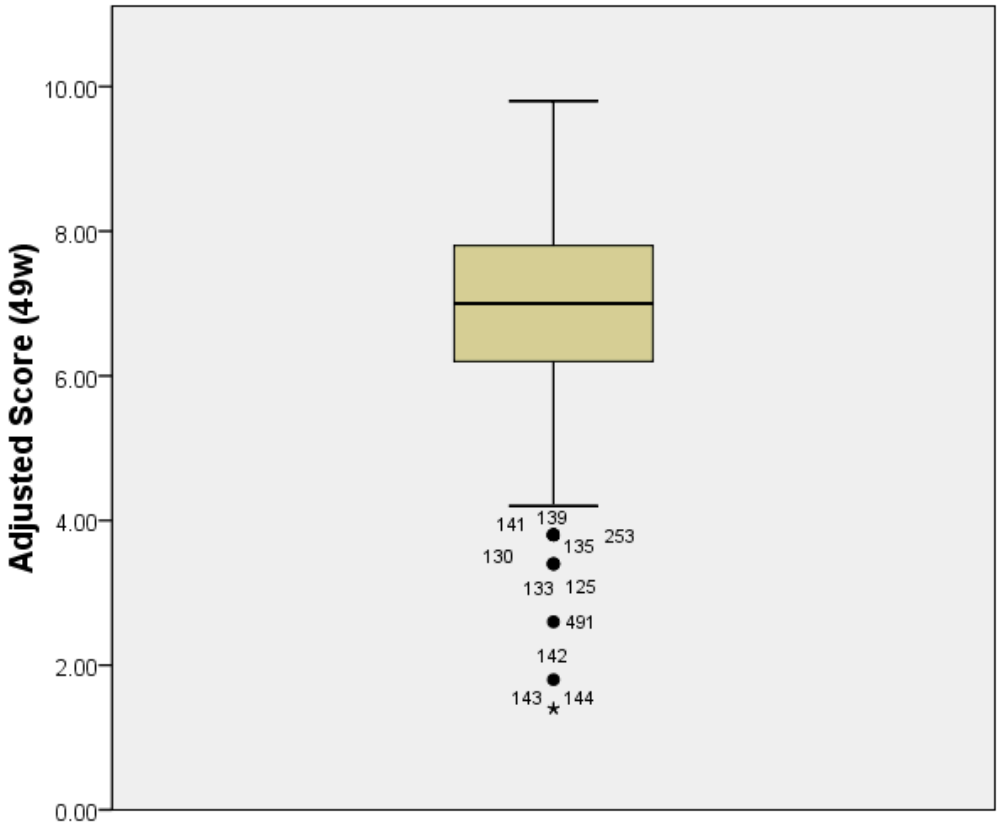


Figure 8.1 Box and whisker plot of adjusted score (full 49w data set)

As previously mentioned, it is important to ascertain the cause of detected outliers prior to commencing a full statistical analysis. As is noted, the majority of outliers are found between data points 125 and 144, corresponding to the results from subject number six.

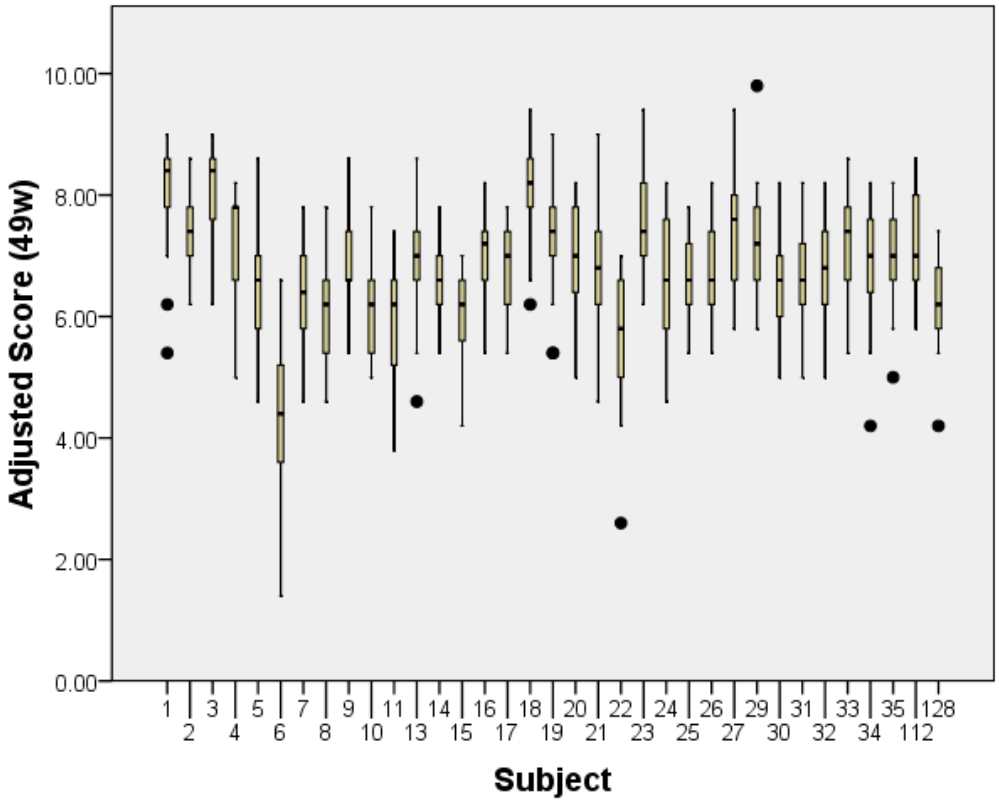


Figure 8.2 Box and whisker plot of adjusted score vs. subject (full 49w data set). Note that, as subjects 12 and 28 did not finish all of the testing sessions, their user numbers have been shifted to 112 and 128 for ease of identification and exclusion.

Examination of the range and distribution of results from all individual subjects (figure 8.2) reveals abnormally wide variance in the results for subject six. Subject six was not available for interview at the time of the analysis. While the subject had indicated no known hearing impairments and passed the hearing acuity test, the author is aware that the subject has a predominant speech impediment. Though at the time the subject was not excluded from the study, the data seems to indicate that exclusion may be prudent prior to analysis. Not only is

there wide variance in the subject's scores, but it can also be seen that the median of these scores is well below the 25th percentile of any other subject.

As it was not possible to definitively determine the causes of the variance or low scores, two sets of analyses were performed as recommended in [14]. The results from both sets were quite similar, with only minor changes in statistical strength and significance between. As such, the results from the data set that excludes subject six, as well as the aforementioned exclusion of subjects 12 and 28, will be reported in detail (Strat_6 data set). Results from further stratification of the Strat_6 data set by array type will also be reported.

8.7.2 49w Results and Analysis

As mentioned in chapter 7.8.2, an issue was found with the 48th word in the first word list. As such, the results of the previous study were reanalyzed using a data set that excluded the results for the 48th words from each word list (49w data set). For the current study, the same data censoring method was applied and statistical analysis was performed on both the censored and uncensored data sets.

As can be seen from tables 8.3 and 8.4, there is little difference between the analysis results from the two data sets. The added control of the nuisance variable word list provided by the 49w data set does not change which effects are or are not significant. For significant effects, statistical strength and confidence are increased. Little change is noted for effects that were not found significant. The effect of word list, which will be discussed further in chapter 8.8.2, is reduced but not eliminated.

The results shown here confirm the findings from the previous study as delay time has a clearly significant effect in the range of 10 ms–40 ms. As significant effects were found in the ranges of 20 ms–40 ms and 30 ms–40 ms, yet no significant effect was found in the ranges of 10 ms–30 ms or 20 ms–30 ms, these results indicate that delay time begins to have a significant effect on intelligibility somewhere above 30 ms.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Subject	9.652	< 0.001	214.195	< 0.001
Word List	56.75	< 0.001	100.793	< 0.001
Delay Time (10 – 20 ms)	0.049	0.0484	0.406	0.524
Delay Time (10 – 30 ms)	0.7.3	0.496	0.789	0.674
Delay Time (10 – 40 ms)	3.695	0.012	9.792	0.020
Delay Time (20 – 30 ms)	1.422	0.234	0.704	0.401
Delay Time (20 – 40 ms)	5.127	0.006	9.247	0.010
Delay Time (30 – 40 ms)	3.636	0.057	3.828	0.050
Array Type	14.953	< 0.001	14.721	< 0.001

Table 8.3 Results of ANOVA and Kruskal-Wallis tests for first-order effects of experimental variables on adjusted score (Strat_6, 50w data set)

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Subject	10.989	< 0.001	232.411	< 0.001
Word List	21.913	< 0.001	43.023	< 0.001
Delay Time (10 – 20 ms)	0.699	0.404	0.498	0.480
Delay Time (10 – 30 ms)	0.749	0.473	0.757	0.685
Delay Time (10 – 40 ms)	4.366	0.005	11.470	0.009
Delay Time (20 – 30 ms)	1.463	0.227	0.609	0.435
Delay Time (20 – 40 ms)	6.130	0.002	10.950	0.004
Delay Time (30 – 40 ms)	4.617	0.032	4.813	0.028
Array Type	17.070	< 0.001	17.877	< 0.001

Table 8.4 Results of ANOVA and Kruskal-Wallis tests for first-order effects of experimental variables on adjusted score (Strat_6, 49w data set)

Another interesting finding is that array type has a clear effect on intelligibility scores (figure 8.3). While, as mentioned previously, this trend

would generally be expected, the underlying reasons for the difference would not. Said reasons will be explored in the next section.

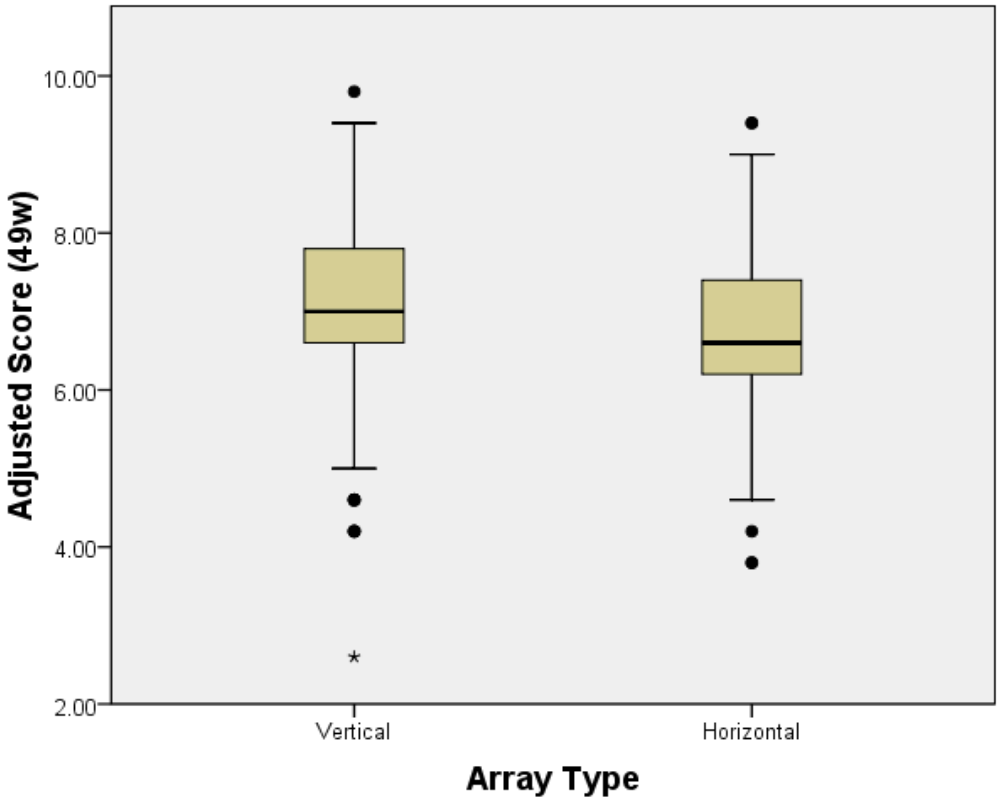


Figure 8.3 Box and whisker plot of adjusted score vs. array type (Strat_6, 49w data set)

	Out of 50		Out of 49	
Variable	**ANOVA F-Stat**	**ANOVA Sig.**	**ANOVA F-Stat**	**ANOVA Sig.**
Array Type × Delay Time (10 – 40 ms)	2.241	0.082	2.200	0.087
Array Type × Delay Time (20 – 40 ms)	3.309	0.037	3.172	0.043
Array Type × Delay Time (30 – 40 ms)	3.905	0.049	3.116	0.078

Table 8.5 Results of ANOVA for second-order effects of experimental variables on adjusted score (Strat_6 data set)

As for the second-order effect of array type on delay time, minor differences were found in the ANOVA results for the 50w and 49w data sets, as shown in table 8.5. Given the non-normal distribution of the data, it is unclear whether the differences between these observed results are indicative of actual differences between the data sets or rather a product of statistical illusion. As such, both the 50w and 49w versions of the Strat_6 data set were further stratified by array type.

8.7.3 Stratification by Array Type

The results from the 49w and 50w data sets showed no difference when stratified by array type. The results from the two array types however displayed great difference. As can be seen in tables 8.6 and 8.7, several delay time ranges have an effect for the horizontal array type, while no significant effect on intelligibility scores can be found for the vertical array.

The increased clarity provided through stratification shows the same effects to be significant (vs. the analysis of the Strat_6 data set); however the strength and confidence are greatly increased. From these analyses it is clear, though again unexpected, that delayed multiple arrivals have a greater negative impact on speech intelligibility when delivered from a pair of loudspeakers oriented in a horizontal point-source array. It is also clear that noticeable detriment is found for delay times above 30 ms. This is not to say that shorter delay times have no effect on intelligibility. Merely, the findings suggest that, for delay times shorter than 30 ms, the injurious effects are difficult to identify. Other known factors, such as SNR and reverberation time, will likely carry more weight in the ultimate determination of the intelligibility of a speech reinforcement system. In other words, if the time offset between multiple arrivals is kept to less than 30 ms and intelligibility is still found lacking for a sound system, one should examine other factors for the underlying cause of the significant impairment.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (10 – 20 ms)	0.192	0.661	0.260	0.610
Delay Time (10 – 30 ms)	0.388	0.679	0.417	0.812
Delay Time (10 – 40 ms)	0.54	0.655	0.953	0.813
Delay Time (20 – 30 ms)	0.204	0.652	0.000	0.993
Delay Time (20 – 40 ms)	0.291	0.748	0.262	0.877
Delay Time (30 – 40 ms)	0.091	0.764	0.200	0.654

Table 8.6 Results of ANOVA and Kruskal-Wallis tests for first-order effects of experimental variables on adjusted score (Strat_6-Vert, 49w data set)

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Delay Time (10 – 20 ms)	2.783	0.097	2.413	0.120
Delay Time (10 – 30 ms)	1.378	0.254	2.552	0.279
Delay Time (10 – 40 ms)	6.184	< 0.001	18.872	< 0.001
Delay Time (20 – 30 ms)	1.605	0.207	1.346	0.246
Delay Time (20 – 40 ms)	9.338	< 0.001	18.720	< 0.001
Delay Time (30 – 40 ms)	7.687	0.006	7.406	0.006

Table 8.7 Results of ANOVA and Kruskal-Wallis tests for first-order effects of experimental variables on adjusted score (Strat_6-Hor, 49w data set)

8.7.4 Multiway Contingency Tables Analysis (MCTA)

As was done in the previous study, multiway contingency tables analysis (MCTA) was employed to confirm or refute the findings of the parametric and non-parametric analyses (see chapter 7.7.6 for an explanation of MCTA). As before, two pass/fail criteria were chosen. Due to the greater number of data

points per treatment and the use of only one SNR value, the requirements for MCTA could be fulfilled by a greater number of criteria.[16] As such, criteria that lie on or in between the medians of scores for the various treatments were selected.

The results of MCTA (table 8.8) were the same for the 50w and 49w data sets, and are both in agreement with the results obtained through the other analysis methods – namely that delay time has a significant effect on intelligibility scores for the horizontal, but not vertical, array geometry. It should be noted that for the 49w data set, the second-order interaction of array type × delay time showed borderline significance ($p = 0.074$) for the 85% pass criterion. This is an indication that the upper pass/fail criterion point did not provide sufficient statistical resolution to detect the relationship.

Data Set	**83% Crit.**	**85% Crit.**
Strat_6	Array Type × Delay Time	Array Type & Delay Time
Strat_6-Vert	None	None
Strat_6-Hor	Delay Time	Delay Time

Table 8.8 Results showing generating class for MCTA using 83% and 85% pass criteria.

8.8 Discussion

It is clear from the various statistical tests used that the effect of delay time on intelligibility scores is different for the two types of arrays studied. Figure 8.4 shows these differing relationships. It is clear from the graph that, as delay time between arrivals increases, scores for the horizontal array decrease while scores for the vertical array do not.

One can see that the median scores in the vertical array are identical for the various levels of the variable delay time. Variance seen in the scores could be an indication that an effect does exist. However it is also possible that, as the statistical tests detailed in this study were unable to find significance, the power of any such effect would pale in comparison to the effects of other factors. This is a

[16] MCTA requires that all frequencies must be greater then 0 and no more than 20% of the count frequencies should be less than five.

good indication of the amount of weight one should give to this specific factor when designing or optimizing sound systems for intelligibility. The variance seen in these scores could also merely be caused by factors such as differences between test subjects and/or word lists.

The same could be said about the observed variance in scores for the horizontal array type. Specifically, comparison of the score distributions for the 20 ms and 30 ms conditions shows a drop in median score as well as increased variance and a larger inter-quartile range. While this may be an indication that delay time begins to have an effect in the 20 ms–30 ms range, the effect was not prevalent enough to be found significant. Again, this suggests the amount of weight that should be given this factor with regard to the intelligibility of reinforced speech.

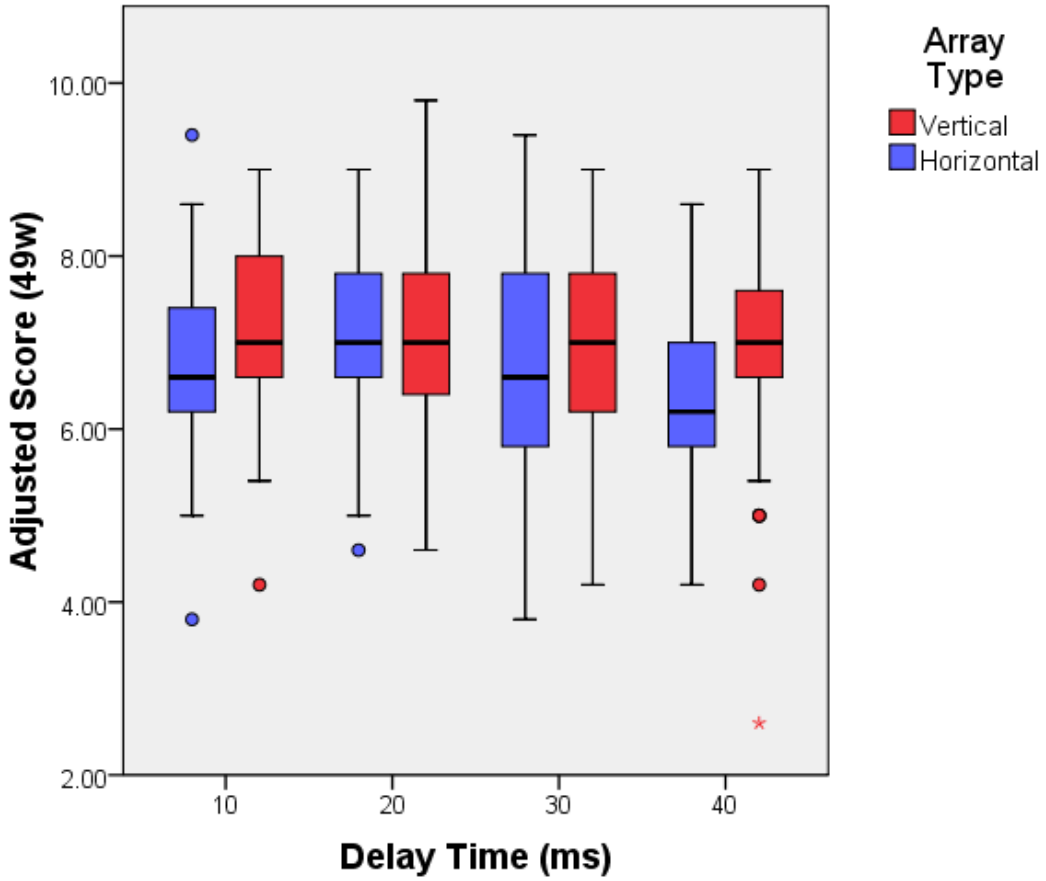


Figure 8.4 Box and whisker plot of adjusted score vs. delay time, by array type (Strat_6 data set)

8.8.1 Training & Effects of Presentation Order

As was the case with the previous two studies, it was desired to know whether subjects received a sufficient amount of training prior to beginning the battery of subjective evaluations. Considering that the presentation order of variable treatments was randomized for each subject, if the analysis of the effects of presentation order on scores were to reveal an improvement in subject performance, this would be an indication of learning. As mentioned in the previous chapters, learning during the first few sessions would indicate inadequate training, and learning over the course of all 24 sessions could indicate that an insufficient number of word lists were used. The results of this analysis are shown in figure 8.5 and table 8.9.

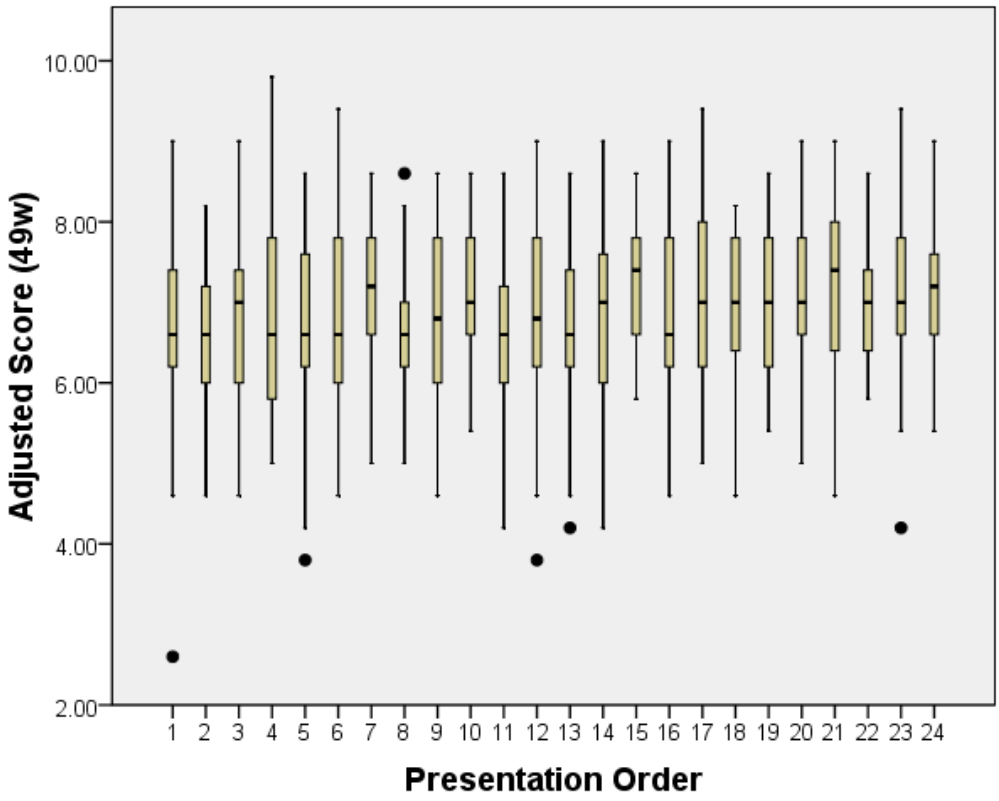


Figure 8.5 Box and whisker plot of adjusted score vs. presentation order (Strat_6 data set)

Inspection of the plot reveals no trends. There is no indication of learning over the first few sessions, nor is there any noticeable increase in means over the

span of 24 sessions. The only element of interest is the one outlier found in the results of the first session.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Order (1-24)	1.116	0.321	25.573	0.321
Order (1-8)	0.732	0.645	5.635	0.583
Order (1-4)	0.333	0.81	0.504	0.918

Table 8.9 Results of ANOVA and Kruskal-Wallis tests for effects of presentation order on adjusted scores (Strat_6, 49w data set)

Through examination of results of statistical tests (shown in table 8.9), it is clear that there is not sufficient evidence to indicate that learning, whether short or long term, was a factor in this study. As with the previous studies, it appears that the amount of training employed and the number of total word lists used in this study were adequate.

8.8.2 The Effects of Word List

As mentioned in the previous chapter, an issue regarding the 48th word in word list number one was discovered during the first phase of the main study. Also during that study, removal of the results for the 48th word in each word list was found to reduce the both the strength and significance of the nuisance variable to a point where the effect was not considered significant. The same censoring of data was applied to the results of the current study and, as table 8.10 shows, an equivalent outcome was not reached.

Variable	ANOVA F-Stat	ANOVA Sig.	K-W χ^2 Stat	K-W Sig.
Word List (50w)	56.750	< 0.001	100.793	< 0.001
Word List (49w)	21.913	< 0.001	43.023	< 0.001

Table 8.10 Results of ANOVA and Kruskal-Wallis tests for effects of word list on adjusted scores (Strat_6 data set)

The strength of the effect was found to be considerably reduced in the 49w data set, indicating that the removal of the offending data points was prudent. However word list remains a significant factor. While an effect is present it

would have no more impact on the results of the study than the "subject" variable, as a full factorial design of treatment vs. word list was used.

The results do indicate that, for future studies involving the same MRT word lists used here, it would be necessary to treat word list as a nuisance variable to be controlled through full factorial design.

8.9 Conclusions

As with the previous two studies, this final phase of the main study involved the addressing of main points and the evaluation of hypotheses. However unlike the previous studies, the scope and the number of research questions had been reduced, allowing for greater focus and increased clarity. This section will address the research questions posed and points of the study, as well as possible directions for future study.

8.9.1 Hypotheses

Ho1: Delay time between multiple arrivals does not affect the intelligibility of speech reproduced by a sound system.

The results of the statistical tests used in this study provide sufficient evidence to reject this null hypothesis. Negative effects on intelligibility were found to be significant for multiple arrivals separated by a 40 ms delay. The specific amount of delay required to cause significant detriment was found to lie in the range of 30 ms–40 ms.

Ho2: (interaction) Array geometry does not affect how delay time between multiple arrivals affects the intelligibility of speech reproduced by a sound system.

The results of this study also provide sufficient evidence to reject this null hypothesis. It was found that time offsets between multiple arrivals, in the range of 10 ms–40 ms, have no significant effect on intelligibility for the vertical array

type. However a significant effect was found for the horizontal array type under the same conditions.

8.9.2 Directions for Future Study

It is clear that delayed multiple arrivals have an effect on intelligibility when delivered from horizontally oriented loudspeakers. As the delay time resolution of this study was limited to steps of 10 ms, it may be of interest for future studies to employ smaller delay steps in the interest of determining the actual point between 30 ms and 40 ms where the effect of delay becomes significant.

On a related note, for this series of studies, the choice of which loudspeaker to hold constant and which to delay was considered a potential variable. Considering the size of the original variable treatment matrix, it was decided that it would not be feasible to include this additional variable. Thus this variable was controlled.

For the case of the horizontal array, the right loudspeaker was selected to be the variable-time source. This raises the question of whether the results found in these studies, namely the interaction between array type and delay time, would also be found if the left loudspeaker had been delayed.

As stated previously, all of the subjects who participated in the studies were native English speakers. Also, it is known that most speakers of western languages tend to be right-ear dominant for the purposes of speech perception [42, 62]. Thus, it would be interesting to know whether the detriment to intelligibility observed for delayed arrivals in the horizontal array are indeed a result of the array geometry, or rather indicative of a higher-order perceptual process.

8.9.3 Parting Thoughts

Reduction in the number of evaluated variable treatments has allowed for the critical study of the first- and second-order effects of two experimental

variables on the intelligibility of speech produced by a sound system. This has yielded the ability to successfully evaluate the two posed hypotheses.

Over the course of the three studies detailed in this research project, a number of factors that negatively affect speech intelligibility have been identified, as have several interesting relationships between these factors. What remains from the original charge of this project is to determine whether the effects and interactions observed through subjective testing will correlate with objective measurements made on the original sound systems used to create the subjective testing stimuli.

9. Comparison of Subjective and Objective Results

As mentioned in chapter 2.1.2, a variety of objective measurements and measurement methods are available for the assessment of the intelligibility of a speech communication system. While objective measurements do provide a fast and nominally effective method for system assessment, it is generally agreed that the correlations between measurable properties and subjective impression have yet to be absolutely defined.

In this chapter, the results from objective measurements are compared to the results from subjective testing with the hope of further delineating the relationship between objective and subjective assessment methods.

9.1 Overview of Measurements

Objective measurement scores were obtained using EASERA and the measurement system detailed in chapter 3.1.2. All scores were obtained from the measured impulse response. As such, metrics such as $\%AL_{cons}$ and STIPa should be viewed as "equivalent" scores.

For the STI family of metrics, "standard" octave weighting was selected in EASERA (shown in table 9.1), thus employing no redundance weights. In addition to weighting, it was selected that the effects of critical band frequency masking and noise be taken into account in the calculations.

Octave	Oct. Weight
125 Hz	0.130
250 Hz	0.140
500 Hz	0.110
1 kHz	0.120
2 kHz	0.190
4 kHz	0.170
8 kHz	0.140

Table 9.1 Octave weights used for STI calculations in EASERA

Measurements were made for all variable treatments and noise levels as indicated in table 3.2. As the impulse response measurement method was dual-

channel FFT using maximum length sequences (MLS), the measurements had an inherently high immunity to noise. For example, the measured frequency response of the system varied little between very low- and very high-noise conditions. An example of the differences between these measurements can be seen in figure 9.1. Also seen in the figure, due to the fact that the frequency response of vocal reproduction system rolls off in the lower-frequency range, differences become significant in the frequency region below 100 Hz. Due to the measurement method's immunity to noise, it was necessary to manually input the signal level and SNR (per octave) prior to intelligibility calculations.

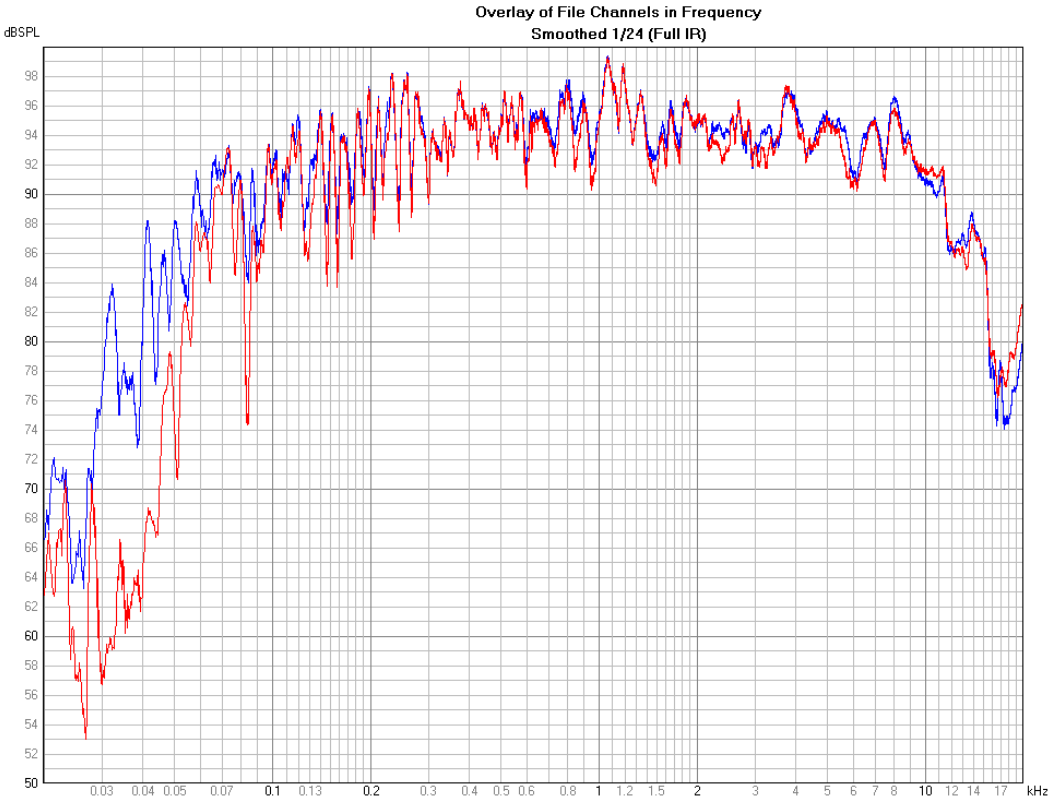


Figure 9.1 Impulse responses for project variable treatment 1 (0 ms delay time, vertical array type, 0 dB level offset) for the 50 dB (red) and −3 dB SNR (blue) conditions.

As mentioned in chapter 2.1.2, STI measurements derived from impulse response measurements do not use a modulated, speech-like test signal. However, as the sound system was not operating outside of its linear region, and no

compression or limiting was involved, differences between the two excitation signals should not affect the measurements [114].

Also, as pointed out by Mapp [118], in order to obtain absolute measures of speech intelligibility provided by a communication system, it is necessary to shape the spectrum of the test signal to match the spectrum of speech (e.g. figure 2.1, chapter 2.1.1). This was not the case with the measurements made for this research project. Mapp mentions that this common oversight can affect SNR and absolute level calculations, as well as corrections for the effects of masking. However Mapp also mentions that even this type of measurement can be considered a guideline and can be used for verifying base predictions.

When one considers the noise immunity of the MLS measurement method, as well as the fact that SNR and absolute level were manually measured and entered into EASERA, the only variable unaccounted for is the effect of masking. Considering that the test signal used had a nominally flat frequency response and thus contained more high-frequency energy than that of a speech signal, the effects of masking were likely underestimated. As such it is probable that the scores reported herein are higher than one would expect to find.

For the purposes of this study, only relative measures and changes/differences between measured conditions of the system are of interest. Also, considering that wide-band equalization is not being studied, the absolute error in results caused by the underestimation of the effects of masking should have little impact on the analysis. However it should be considered that the effects of masking on narrow-band frequency anomalies above 1 kHz may also be underestimated.

9.2 Verification of Findings

Through subjective testing, significance was found for each of the 1-way and 2-way effects of delay time, array type and SNR. The following is a comparison of the objective and subjective findings for each of these effects. As the variable level offset was not used in the subjective testing processes, only the

results from treatments using the level offset condition used in the testing (0 dB) will be reported.

Delay Time

Delay time was found to begin to have a significant first-order effect for offsets greater than 30 ms. As seen in figure 9.2, the STI scores remain fairly constant over the range of 5 ms–30 ms, with a considerable drop between 30 ms and 40 ms. While this is in agreement with the results of subjective testing, a curiosity is noted with regard to the drop in STI score between the 0 ms and 5 ms delay times.

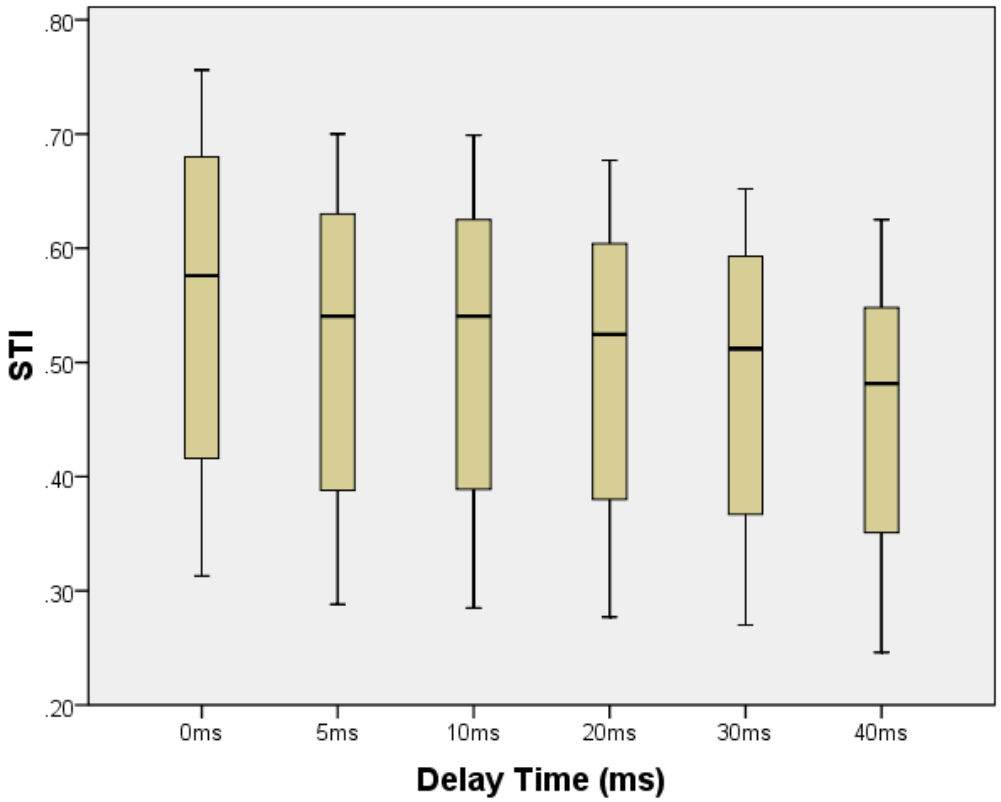


Figure 9.2 Box and whisker plot of STI vs. delay time (all treatments)

Revisiting the concept of the modulation transfer function (MTF), as seen in figure 2.4 (chapter 2.1.1), the cause of the drop in measured STI between the 0 ms and 5 ms conditions becomes manifest. Any reduction in measured modulation of the various test signals will result in a drop in STI score. As with the ordinary summation of signals (as detailed in chapter 2.2.2.1), the summation

of modulated signals will result in some combination of constructive and destructive interference. However with modulated signals, there is an added layer of complexity with regard to constructive interference. As seen in figure 9.3, the summation of two delayed modulated signals gives rise to the possibility of great reductions in modulation for certain combinations of modulation frequency vs. delay time.

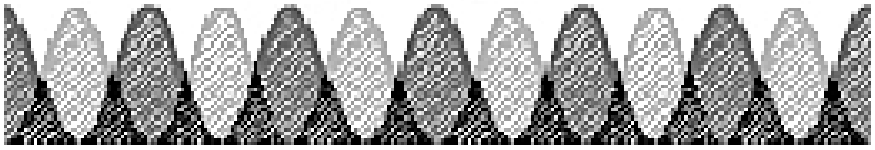


Figure 9.3 Superimposed, delayed modulated signals (Reprinted with permission, adapted from [77])

Given a specific delay time, the summation of delayed signals at different modulation frequencies will result in different degrees of preservation/increase vs. reduction of modulation. Thus, the relatively large number of modulation frequencies used in the STI measurement ensures that the transmission index for a frequency band will not be reduced to zero. However as the STI measurements clearly indicate a reduction in intelligibility for the 5 ms delay condition, and these results are not supported by the results of subjective testing, it seems clear that the STI measurement method overestimates the detriment caused by multiple arrivals with a 5 ms offset.

SNR

The variable SNR was found to have an extremely powerful effect on intelligibility scores. As can be seen in figure 9.4, this relationship was also recognized by the STI measurement. When compared to figure 6.1 (chapter 6.7), one can see similarities in the upward trend of means for higher SNR values. This correlation is not startling considering the known effects of SNR on both intelligibility and the MTF.

An interesting point is that one can also see differences in the trend of variance between the plots. Whereas for the human listeners, variance increases for lower SNR conditions, the opposite is true for the measured results. As the

cause of variance is different in in the two assessment methods, it is not surprising that the amount of variance would also differ. However the results of the measurements are in agreement with the interaction effects found through subjective testing – namely that the effects of other factors are obscured under lower SNR conditions.

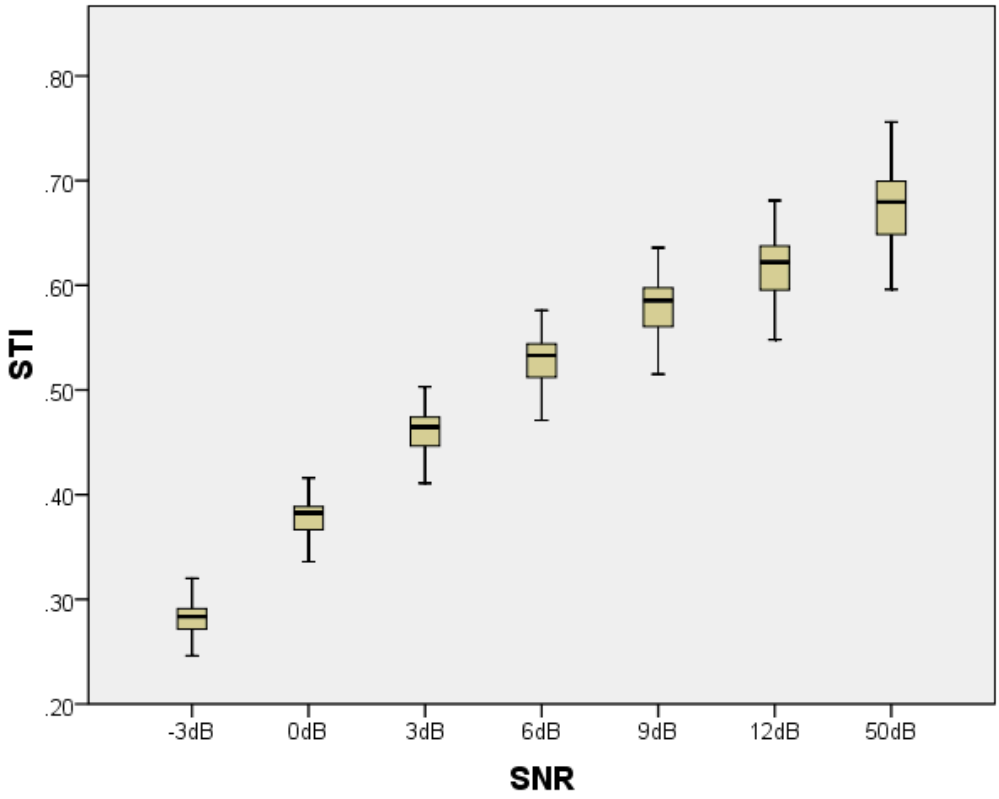


Figure 9.4 Box and whisker plot of STI vs. SNR (all treatments)

Array Type

The first-order effects of array type were found to be significant for the 49w data sets in both the first and second phases of the main study. As seen in figure 9.5, the nearly coincident ranges of values and large degree of overlap between the inter-quartile ranges of the STI scores for the two array types make it is difficult to definitively establish that a difference exists. Though some difference can be noted, the equivalent plot of results from subjective testing (seen in figure 8.3, chapter 8.7.2) shows a clearer difference, and thus a greater apparent first-order effect.

As mentioned previously, the differences between array types are an amalgam of three distinct variables – monaural vs. binaural hearing, point-source vs. point-destination array configuration and on- vs. off-axis listening location. The STI measurement is a single-channel measurement and it would appear that the results of the measurements, as shown in figure 9.5, do not completely account for the perceived difference in intelligibility between array types. This would seem to indicate one of two things. First, it is possible that at least some portion of the difference in subjective impression found between array types was due to a binaural mechanism. Second, it is equally possible that the STI measurement method is not sensitive enough to detect the cause of the difference. An analysis of variable interactions may provide more insight.

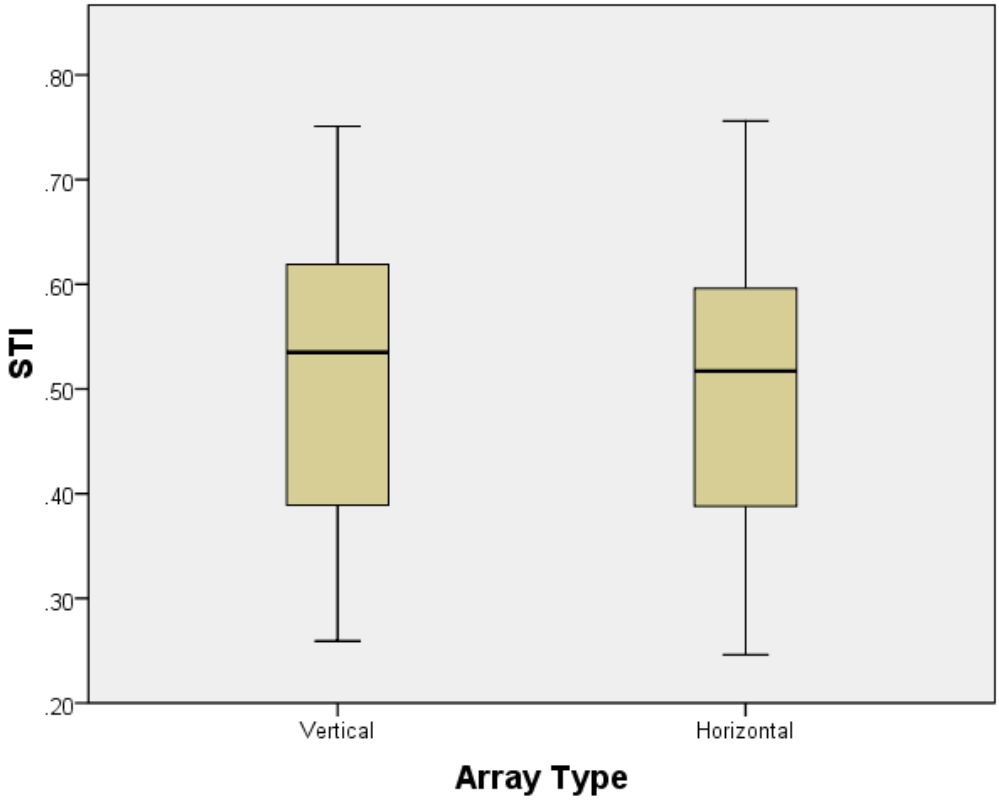


Figure 9.5 Box and whisker plot of STI vs. array type (all treatments)

SNR × Delay Time

It was found through subjective testing that the effects of delay time are obscured by the more powerful effects of SNR, for treatments containing lower

SNR conditions. The results from objective measurements, shown in figure 9.6, appear to confirm these findings. A comparison of the STI scores for the 50 dB and −3 dB SNR values shows that differences in score due to delay time are significantly reduced for the lower SNR condition. As such, it is reasonable to conclude that, in terms of detecting the interaction effects of SNR × delay time, the STI measurement method functions in a manner similar to human perception.

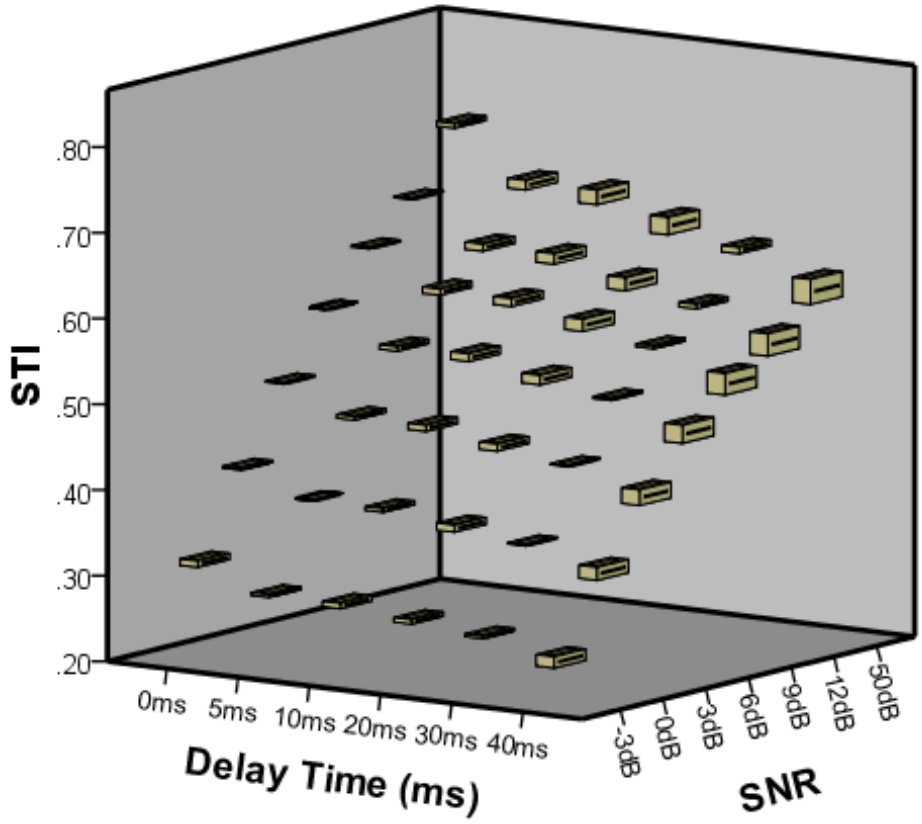


Figure 9.6 3-dimensional box and whisker plot of STI vs. delay time, by SNR (all treatments)

SNR × Array Geometry

As was the case with the interaction effect of SNR × delay time, the STI measurement method appears capable of detecting at least some of the interaction effects of SNR × array type. In figure 9.7, one can see that the difference between array types noted for the 50 dB SNR condition slowly diminish as the value of SNR decreases. Also, as was seen in figure 9.4, the results shown in figure 9.7

continue to show a reduction in the size of the range of scores for lower SNR conditions.

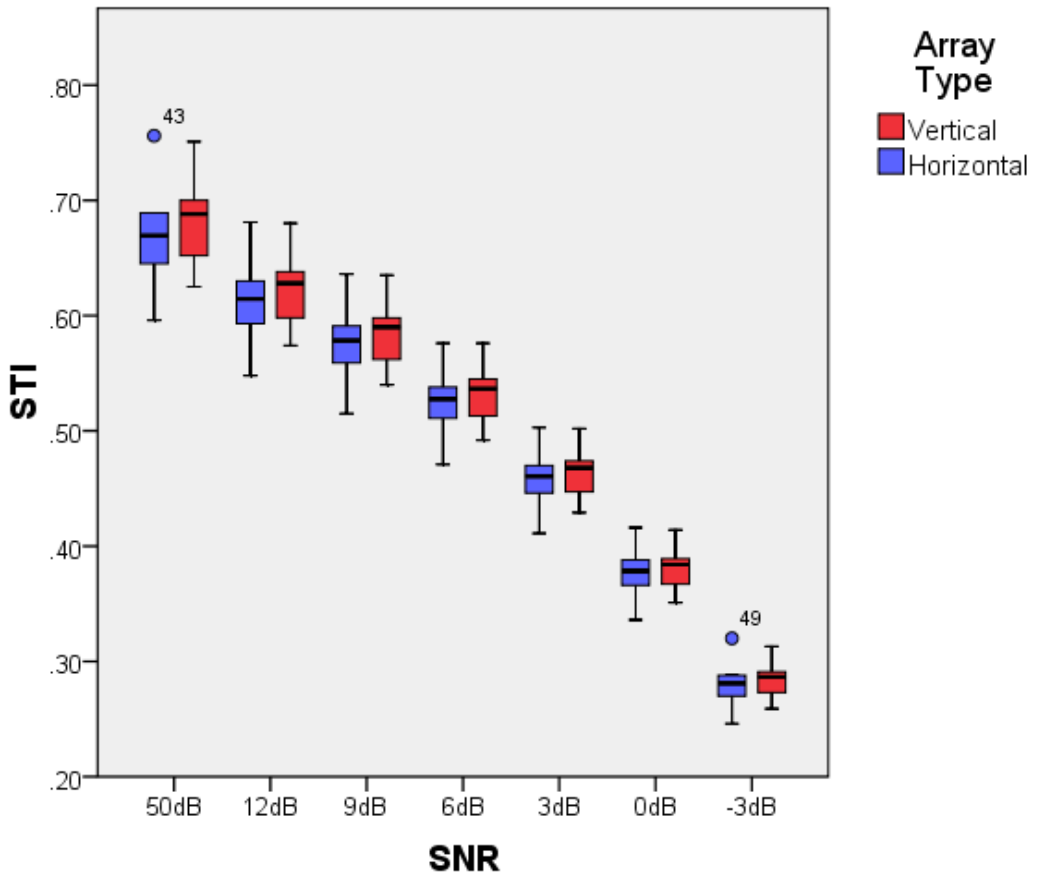


Figure 9.7 Box and whisker plot of STI vs. SNR, by array type (all treatments)

It remains unclear whether STI is sensitive to all aspects of the effect of array type, as it is possible that the reductive qualities seen in the interaction with SNR may only serve to mask the effects that the measurement method is capable of detecting. Thus, while it is clear that the STI measurement method is to some degree sensitive to the 2-way effects of SNR × array type, it is not clear that this sensitivity is absolute.

Figure 9.8 shows a different view of the data, using the format of figure 7.9 from chapter 7.7.4 (with the inclusion of the 3 dB SNR condition). As can be seen in the figure, the results of objective measurement show a less distinct difference in scores for the higher SNR condition. While this observed difference

may be due to differences between the STI and word score scales, it is also possible that this data indicates that the STI measurement method is not fully sensitive to the effects of array type.

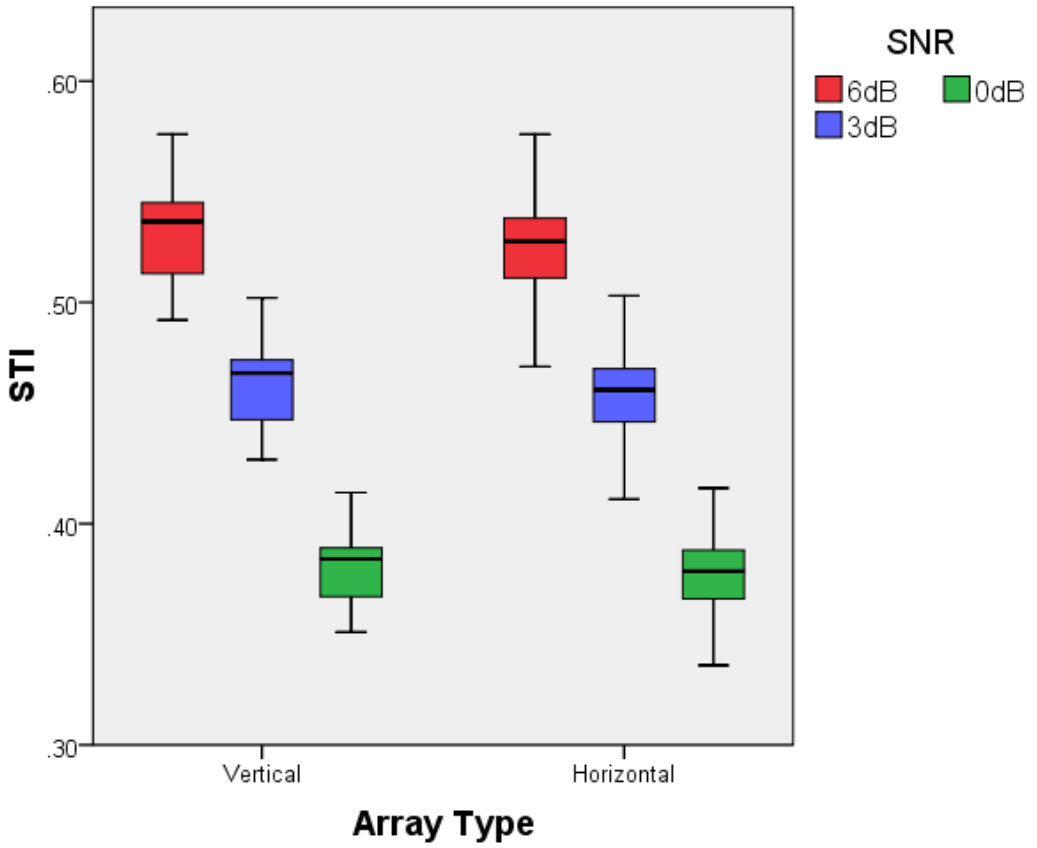


Figure 9.8 Box and whisker plot of STI vs. SNR, by array type (all treatments containing SNR conditions 6 dB, 3 dB and 0 dB)

Array Geometry × Delay Time

Several interesting aspects of this interaction have been noted in previous chapters. The first is that the effects of very short delay times (around 0.3 ms) were found to be significant for the vertical array type.[17] The second is that the effects of delay times ranging from 5 ms–40 ms were only found to be significant in the range of 30 ms–40 ms, and only for the horizontal array type.

The results of objective measurement are equally interesting. As seen in figure 9.9, the first thing to note is that there is virtually no difference between

[17] As previously mentioned, the effects of very short delay times were not tested for the horizontal array type.

scores for the array types for the 0 ms condition. It should be reiterated that the short (0.3 ms) delay time found in the audio recordings was not present in the signals received by the measurement microphone, yet the notches in the frequency response of the loudspeakers in the horizontal array type (seen in figure 3.7, chapter 3.2.2) would be captured by the measurement microphone. While it is possible that the underestimated effects of masking, due to the lack of spectral shaping of the test signal, could reduce the effect of the higher-frequency notch on the STI score, the notch at around 1.2 kHz would be virtually unaffected by spectral shaping. Given that there is no difference between the scores for the two array types under the 0 ms condition, it is reasonable to conclude that the STI measurement method is not sensitive to the narrow-band frequency response anomalies caused by very short delay times.

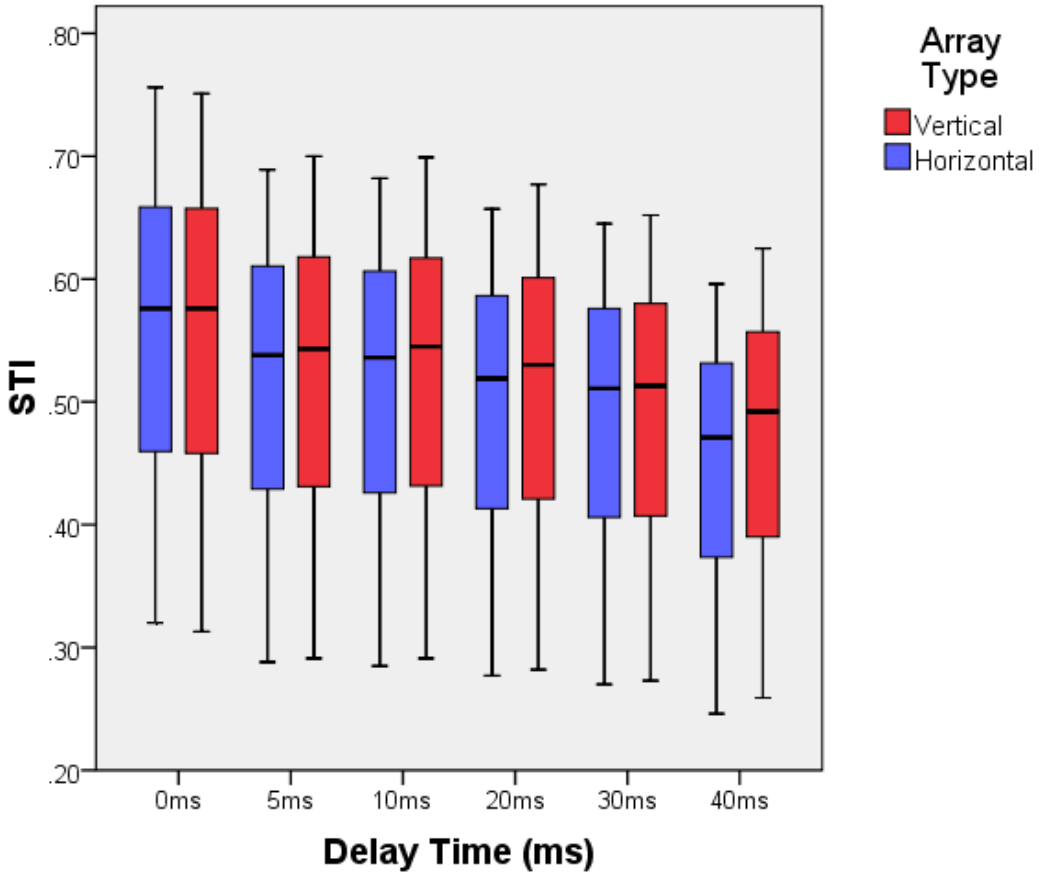


Figure 9.9 Box and whisker plot of STI vs. delay time, by array type (all treatments)

Another pattern of note is that scores remain very similar between the two array types in the range of 5 ms–30 ms. Given the results shown in figures 7.10 (chapter 7.7.5) and 8.4 (chapter 8.8), one would expect to see large differences between scores for the 30 ms condition as well as less overall change in the scores for the vertical array. In the interest of a more direct comparison, a plot of the STI scores for the 0 dB–6 dB SNR conditions has been included in figure 9.10.

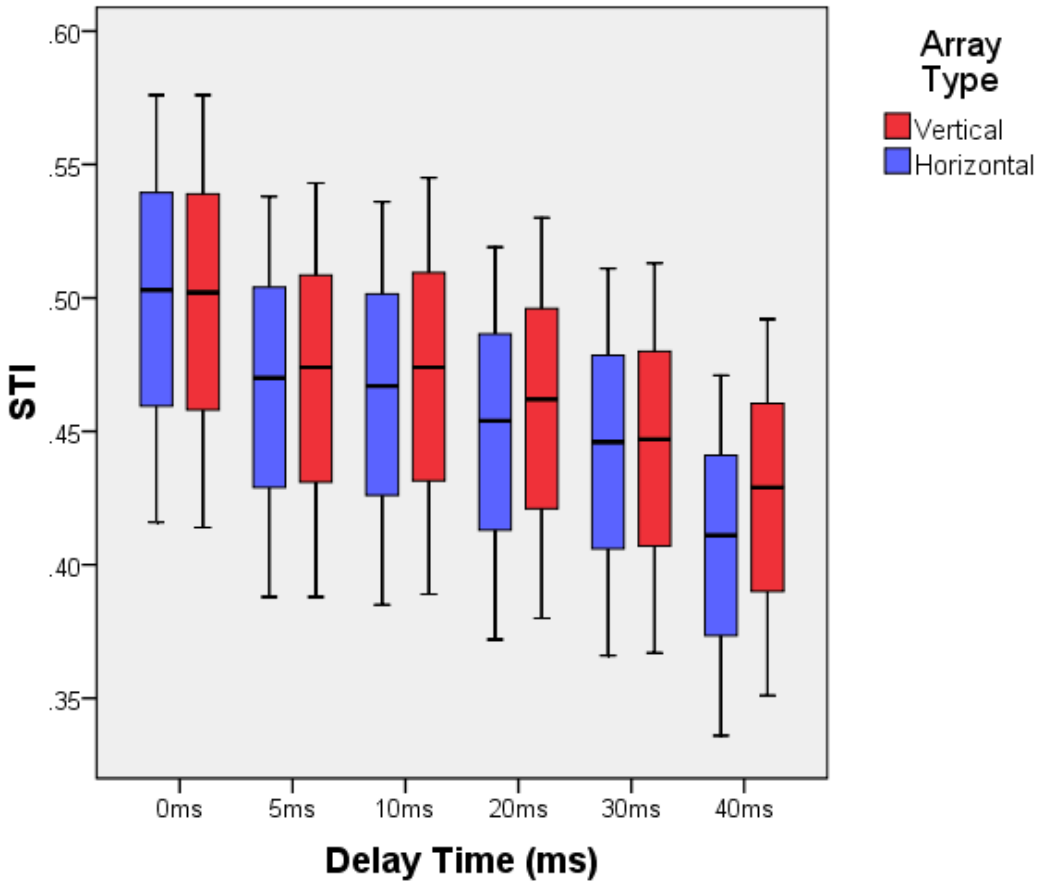


Figure 9.10 Box and whisker plot of STI vs. delay time, by array type (all treatments containing SNR conditions 6 dB, 3 dB and 0 dB)

Differences found between subjective and objective assessment methods, with regard to the interaction of array type × delay time, could indicate one of two things. First, it is possible that differences in scores do exist between the different delay times, but that the effect is essentially negligible. The second possibility is that there is a binaural vs. monaural hearing mechanism at work, creating

differences in the scores between array types, which is not detected by the STI measurement.

The third thing of note is that the STI measurements correctly identify that a difference in intelligibility exists between the two arrays at the 40 ms level. Again, as the measurement method is single-channel, the results suggest that, at least some portion of, the difference between scores is due to a monaural effect.

9.2.1 Conclusions

The STI measurement method does appear sensitive to the effects of delayed multiple arrivals on speech intelligibility. Curiously, it does not appear that measurements of the 40 ms delay value are skewed by the presence of comb filter notches in the MTF, as these results are in agreement with word score results. However the results for short delay values (e.g. 5 ms) do appear to diverge from the results of subjective testing, possibly due to an inherent limitation of the measurement method.

STI is obviously capable of accounting for the effects of SNR. It also appears capable of accurately predicting the subjective effects of the interaction of SNR with both delay time and array type.

With regard to array type, it is unclear whether STI is entirely sensitive to the first-order effect. However the measurement method does appear to detect some of the effects of the interaction of array type and delay time, specifically for longer delays.

10. Discussion

"E = MC² ± 3 dB"

- David Engstrom (in [43])

Perhaps a bit whimsical, but this observation reminds one that the acumination of knowledge often carries with it the introduction of new unknowns. It also reminds one that certainty is rarely certain. Research, however, is a process of successive approximation, wherein each step has the potential to yield a greater understanding of the world around.

The research project detailed in this dissertation focused on the complex question of how sound system optimization affects the intelligibility of reinforced speech. As an early attempt in the deliberation process, the goal of the project was to deconstruct this larger question, identifying noteworthy research avenues and following several of the identified paths. As many of the potential research questions were unknown at the onset of the project, an ecological approach was adopted, studying real-world reinforcement scenarios in an actual performance space.

Through the use of subjective and objective testing methods, many potential research paths were discovered and a variety of hypotheses were evaluated. The following two sections discuss the findings of this series of studies, including both questions answered and unanswered.

10.1 Questions Answered & Implications

Through subjective testing, it was found that all three of the experimental variables used in this research project (SNR, delay time and array geometry) had significant first-order effects on the intelligibility of reinforced speech. Through stratified analysis of the various data sets, it was found that all three of the second-order interaction effects between these variables also had significant effects. The observed effects and relationships are as follows:

1) Delay times greater than 30 ms have an effect for the horizontal array geometry.
2) Very short delay times (0.3 ms) have an effect for the vertical array geometry.
3) Scores for the horizontal array geometry were found to be lower than scores for the vertical array geometry.
4) As SNR decreases, the magnitude of the effects of delay time and array geometry also decreases.

When the results of subjective tests were compared with the results of objective measurements (STI), both correlations and discrepancies were uncovered:

1) STI appears to accurately identify the effects of delay times greater than 30 ms.
2) STI also accurately predicts that scores are generally lower for the horizontal array type.
3) STI predicts a larger negative effect, versus the results of subjective testing, for delay times in the 5 ms–30 ms range.

System engineers have been employing compensational delay to align loudspeaker arrivals since the 1960's. Likewise, the impairment to intelligibility due to echoes has been studied for some time. These are not new or innovative concepts. However, the explicit delineation of the amount of delay offset, geometrical relationship of loudspeakers and relationship to the background sound field that are required to make a noticeable difference in speech intelligibility are novel endeavors.

During the last part of the 20th century, debate formed regarding whether the greatest degree of intelligibility would be achieved through the absolute alignment or the intentional misalignment of loudspeaker arrivals. The issue was treated in a global manner, and most often assessed by objective measurement.

The results of this research project, however, indicate that all loudspeaker arrays may not be equal.

For the vertically-oriented point-destination array, the comb filter created by the very short time offset had a noticeable effect on intelligibility scores. For this same array, the addition of a 5 ms delay (4.7 ms total offset) improved scores, indicating that misalignment improves intelligibility. For the horizontally-oriented point-source array, a very similar comb filter was present in the frequency response for the 0 ms condition. For this array, the addition of a 5 ms delay did not improve intelligibility scores. The conditions studied with these two arrays are analogous to the difference between acoustical and electrical summation in a sound system. In one case, the comb filtering is created through summation at the listener; in the other, the comb filter exists in the signal received from each loudspeaker.[18] The results of this study suggest that misalignment is effective at combating the effects of comb filters created by acoustical summation, but are not effective at improving the intelligibility of "pre-combed" signals. Additionally, the trend of lower scores for the horizontal vs. vertical arrays may be due to the inherent comb filter in the response of the horizontal array, or to the existence of a physical sound source in front of the listener. Further investigation is, of course, required to rule out the effects of the complex variable array geometry.

Though this research project was not charged with determining the appropriate value for the integration time (if any) to be used with regard to fusion, the results do suggest one conclusion. Given that delay times begin to have a significant effect in the region between 30 ms–40 ms, and the fact that very short delay times can have an effect, the author proposes that fusion may indeed play a role in speech intelligibility for delay times less than 30 ms–40 ms, and that the effects of very short delay times on intelligibility could be due to frequency response anomalies rather than a lack of fusion.

---

[18] For the case of the horizontal array, the comb filter is created via acoustical summation of the signals from multiple drivers within each loudspeaker. As the individual drivers were not delayed relative to each other, the analogy of electrical summation still holds.

In addition to results regarding very short delay times, the studies in this research project uncovered an interaction between delay time and array geometry for longer delay times. Results indicate that multiple arrivals have a greater potential to negatively affect intelligibility for the horizontal array geometry. As this relationship was also, to some degree, detected by STI measurements, it is unlikely that binaural listening is the sole cause of this interaction.

Finally, results from the comparison of subjective and objective assessment methods clearly indicate that objective measurements can provide inaccurate results for conditions involving multiple arrivals. In terms of measurement reliability, these results implicate that objective measures should not be used to evaluate the intelligibility of time-delayed multiple arrivals. Further, prior to performing objective measurements, impulse response measurements are needed to ensure that time-delayed multiple arrivals are not present in the received test signal.

10.2 Questions Unanswered & Suggestions for Further Study

Through the course of this project, many potential research questions were uncovered. Several of these questions have been addressed. Alas, owing to a variety of factors, a number of questions have not. The charge of research, however, is not restricted to the answering of questions. Rather, research should create questions as well – both to sustain, and to spark new interest. This section contains a number of prospective research questions for future study.

The first, and perhaps most obvious, avenue for future study is the deconstruction of the compound variable array geometry. This research project has shown that differences do exist between the relative intelligibilities of the two array geometries. The question remains as to what factors cause these differences. Isolation and control of the variables plane (medial vs. horizontal) and array focus (point source vs. point destination) could aid in determining whether, or to what degree, the observed effect of array geometry is due to hearing method (binaural vs. monaural), on- vs. off-axis listening, equalization location or the existence of a

physical center channel. It is recommended that point-destination arrays be used to isolate plane, and the medial plane used to isolate array focus.

If hearing method is found to play a part in the observed differences between array geometries, it is suggested that "delay side" be investigated in the future. In this series of studies, for the horizontal array geometry, the left loudspeaker was held constant while the right loudspeaker was shifted in time. In an informal survey of the test subjects, nearly all participants indicated right-ear dominance with regard to speech.[19] As delays in the 0 ms–20 ms range have the spatial effect of shifting the apparent sound source, delaying the right loudspeaker would shift the source away from the listener's dominant ear. This raises the question of whether delay time added to the left loudspeaker, affecting a source shift toward the dominant ear, would have a different effect on intelligibility.

With regard to delay time, the effect of delays in the region of 0 ms–5 ms warrants further investigation. The existence of negative effects of the acoustical summation of signals with such short time offsets is clear. The remaining research question involves the delineation of the amount of delay required to affect a noticeable detriment, and the minimum amount of further delay required to overcome the detriment.

Finally, questions remain regarding the interactions of SNR on delay time and array geometry. It was observed that low SNR conditions have a mitigating effect on the effects of the other two variables. However, the specific SNR thresholds required to affect this phenomenon have not been determined.

10.3 Final Thoughts

The motivation for this research project manifested from the author's own work as a sound system designer and system engineer. Often encountering situations in which design and optimization decisions were required, yet no information was available to guide these decisions, best-fit solutions were usually found through a process of trial, error and modification. While much of the

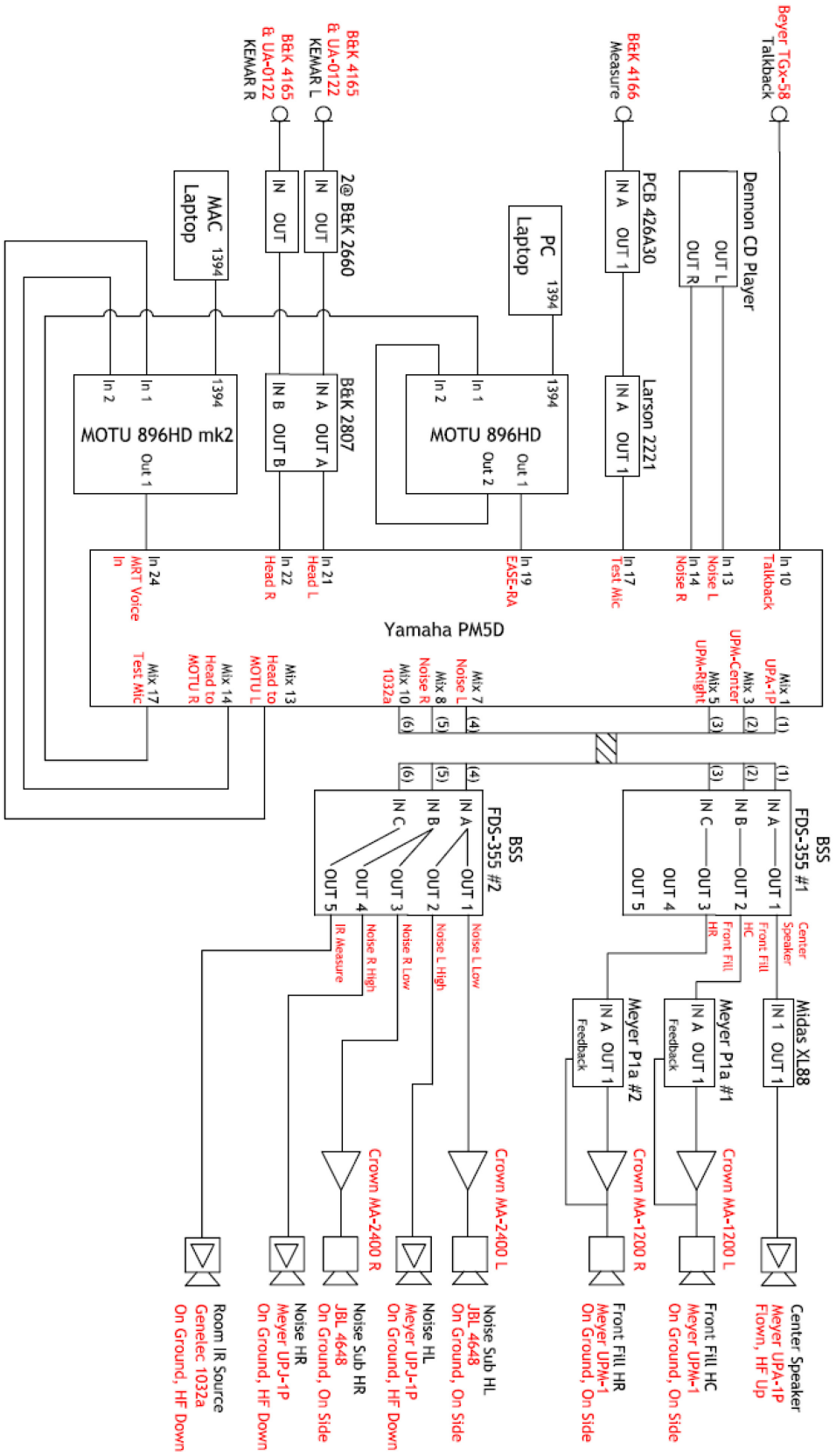
[19] The question was asked during the hearing acuity test: "What ear do you use when talking on the telephone?"

available knowledge in the field of live reinforcement has been garnered by such empirical means, a point can be reached wherein these methods are incapable of providing further illumination. It is at that point that structured, scientific research is required.

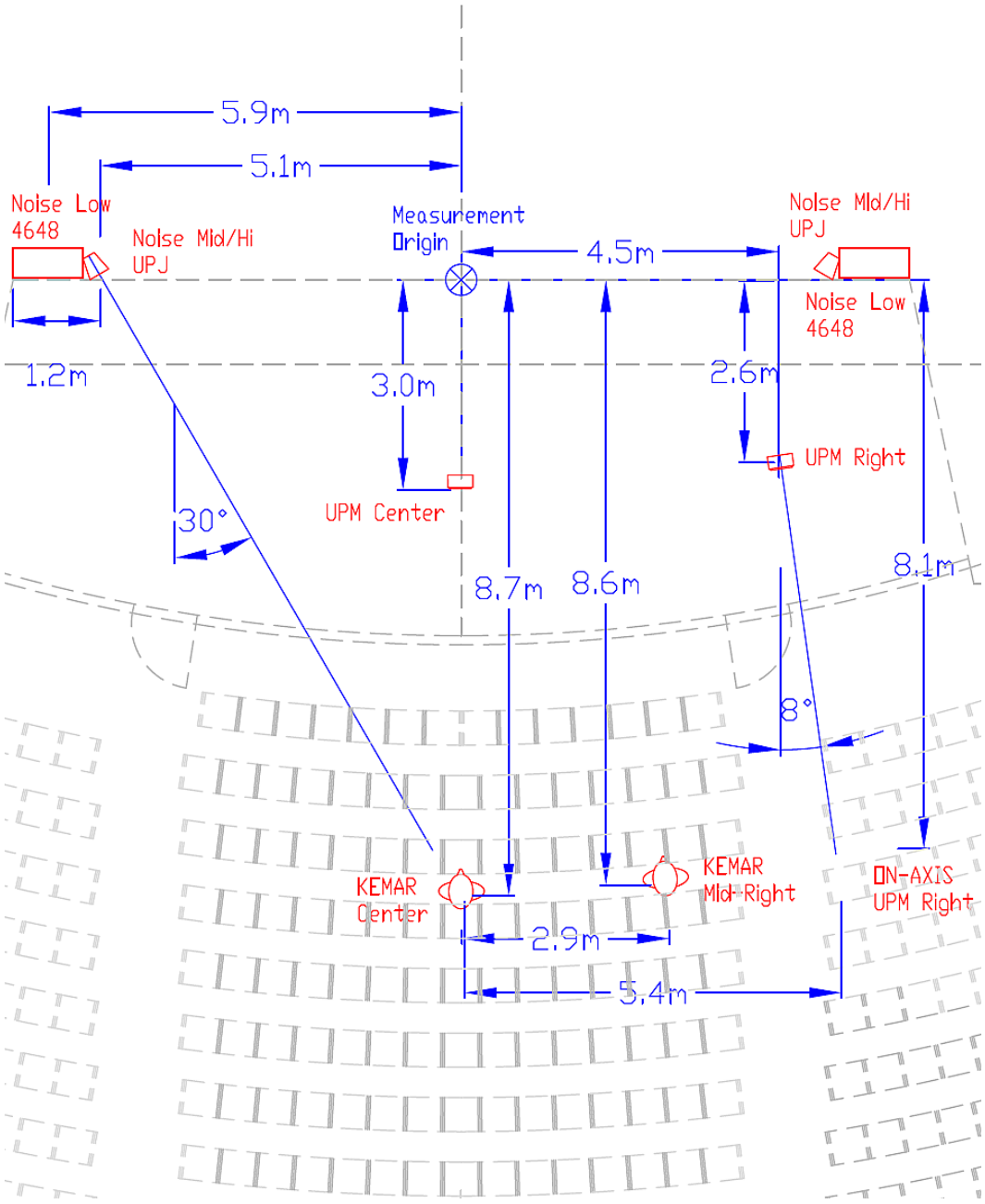
The author was recently asked how the knowledge gained from this research project will affect his future design and optimization decisions. While it will take time to fully cogitate and assimilate the actionability of these findings, several decision-making aids are already apparent with regard to alignment. Multiple arrivals with delay times ranging between 5 ms–30 ms do not significantly affect intelligibility, though they do affect sound quality. Thus, if intelligibility is of paramount importance in a given reinforcement situation, intentional misalignment of vertically oriented point-destination arrays is warranted for the preservation of intelligibility at the cost of the overall sound quality of the vocal reinforcement system. Conversely, one may place higher, though not sole, priority on sound quality. Though the effects of very short delay times are not yet known for this type of array, due to the smaller region of overlap and volatility, the use of aligned, horizontally oriented point-source arrays will yield better overall quality and negatively affect intelligibility for the smallest number of people.

The question of how to optimally optimize a sound system is far from answered. While many aspects of the greater question have been clarified, many questions remain. It is with great pleasure, and great reverence for the many scientists, researchers and sound engineers who have framed the grimoire of sound system engineering, that the author offers this dissertation as a contribution to the ongoing effort to answer these questions, and to increase the body of knowledge in the field of live sound reinforcement.

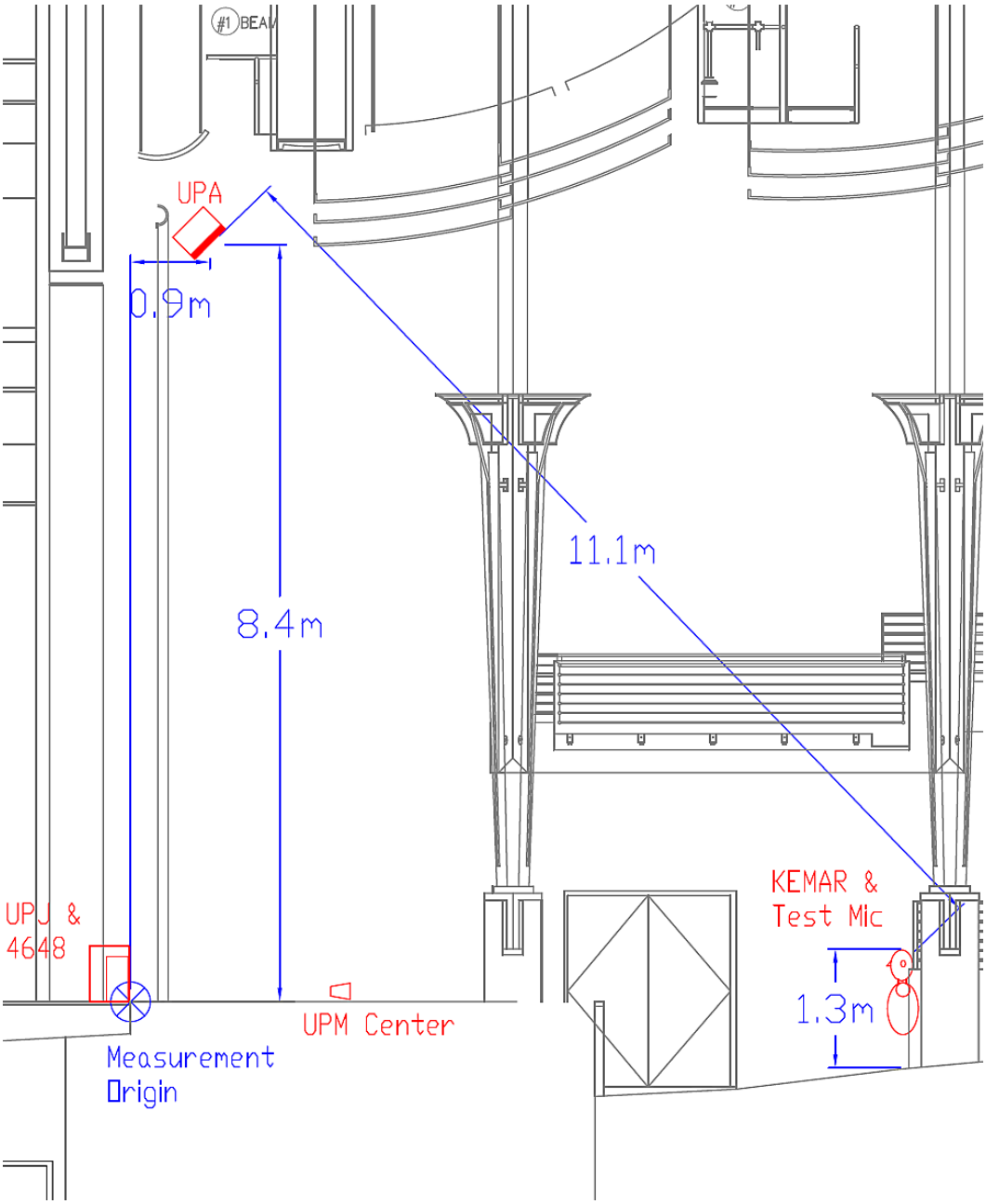
Appendix A: Recording Sound Systems - Schematic



Appendix B: Recording Sound Systems - Plan View



Appendix C: Recording Sound Systems - Section View



Appendix D: Recruitment Flyer

Looking for:

Participants for a Research Study

On

Speech Intelligibility and Live Sound Reinforcement

WHO?

Ages 18-40, Male or Female
Native English Speakers
No Known Hearing Damage/Loss

WHAT?

Listening to Speech with Headphones
2-3 Sessions - About 1 Hour per Session
No Obligation – Can Withdraw from Study at Any Time
Confidential – Your Identity Will Not Be Released
Compensation Provided

WHEN?

Between **June 1** and **June 12**, 2010
Sessions will be Scheduled Individually

Questions? Interested?

Contact:
Tim Ryan
CCM.audio.research@gmail.com

Sound Design Area
Theatre Design and Production
College Conservatory of Music
University of Cincinnati

Sound Recording Area
Schulich School of Music
McGill University

Supervisor: René Quesnel Sound Recording Area, McGill University
Co-PI: Laura Kretschmer Department of Communication Sciences and Disorders, University of Cincinnati

*** Dates varied for different recruitment periods

Appendix E: Standard Email Response to Inquiry

Dear _____

Thank you very much for your interest in our research project.

This is a standard form that I am using to respond to all inquiries. I am sorry if this seems a bit impersonal, but it is important that all participants in the project receive the same information and the same instructions.

At this point, we are looking for new participants with any level of audio engineering or audio evaluation experience, including people with no experience in either of these areas.

For this project, you will be asked to participate in sessions in which you listen to speech with headphones and enter responses into a computer. These sessions will take place in the CCM Sound Design Studio at the University of Cincinnati.

I have attached a copy of our Listening Test Consent Form. For now, please use this document for informational purposes. I hope it will explain the purpose and nature of the research project and answer any questions that you have. If you wish to participate in this study you will be asked to sign a copy of the Listening Test Consent Form at the beginning of your first listening session.

For this project, it is important that you are between the ages of 18 and 40 years old, that you are a native English speaker and that you have no known hearing impairments.

If you decide to participate in this project, you will be asked to participate in two-to-four 1-hour sessions. You will be able to withdraw from the project at any time, but if you think in advance that you will not be able to complete all of the sessions I ask that you do not sign up for the project.

If you are still interested in participating, please contact me via email to schedule a time for your first listening session.

Thank you for your time and interest.

-Tim Ryan

Timothy Ryan
Visiting Assistant Professor
Theatre Design and Production
College Conservatory of Music
University of Cincinnati

Doctoral Candidate
Sound Recording Area
Schulich School of Music
McGill University

Contact:
CCM.audio.research@gmail.com

Supervisor: René Quesnel — Sound Recording Area, McGill University
Co-PI: Laura Kretschmer — Department of Communication Sciences and Disorders, University of Cincinnati

Appendix F: Listening Test Consent Form

Sound System Engineering and the Intelligibility of Reinforced Speech

Principal Investigator (PI): Timothy Ryan
774-238-8930 CCM.audio.research@gmail.com

Visiting Assistant Professor
Department of Theater, Design and Production
College Conservatory of Music
University of Cincinnati

Ph.D. Candidate
Sound Recording Area
Schulich School of Music
McGill University

Purpose: You are being asked to participate in a research study which will investigate how several factors associated with sound system tuning and optimization affect how well an audience can understand reinforced speech. This study will include approximately 60 participants, all native English speakers, between the ages of 18 and 40 years old and with no known hearing impairments. Participants will be divided into three categories based on individual critical listening and subjective assessment experience level: Naïve, Experienced and Expert.

Listening Sessions: You are being asked to participate in a series of two to four separate listening sessions which will last approximately one hour each. In each session, you will listen to a series of sound clips via headphones and enter your responses into a computer. Each sound clip includes a recording of one English word, and may or may not include background noise. Your task is to identify the word in question and enter your response via keyboard into the computer. You will receive specific written and oral instructions, as well as a brief training session at the beginning of each session. If you are a student at either the University of Cincinnati or McGill University, know that your participation in this study is not related to your university courses, and that you are free to withdraw from the study at any time with no penalty.

Potential Coercion: The project principal investigator is faculty at the University of Cincinnati. If you are a student at the University of Cincinnati, it is possible that you are one of his students. If you are one of his students you will not be able to participate in this research project. Your decision to participate in, not to participate in or to withdraw from the study will have no positive or negative effect on your academic standing, your grades or production assignments. If you feel you are being pressured in any way to be in this study, please contact the Institutional Review Board at the University of Cincinnati, the project supervisor Prof. René Quesnel or the co-investigator Prof. Laura Kretschmer (see below for contact information).

Potential Risks: This research study presents you, the participant, with a minimal level of risk. You will experience sound pressure levels in the range of 70 – 80dB SPL. This is equivalent to the level of loud, un-amplified speech and is not believed to pose any threat to your hearing. However this sound pressure level range is loud enough that you may experience auditory fatigue – a temporary condition. To minimize auditory fatigue, you are encouraged to take breaks every 30 minutes or more frequently if you wish. The sound output level of the headphones is not adjustable. This prevents you from accidentally turning up the level and minimizes the potential harm to your ears. If you have any questions or concerns regarding sound exposure and its effect on hearing loss, the project principal investigator will be happy to answer any questions and/or supply you with literature references on the subject.

The sound clips that you will hear were created using a process called binaural recording. Binaural recording is one of many methods used to create surround-sound recordings. You may find that you have a surround-sound listening experience, or that it sounds like you are in a different location than the testing room. As such, some people find listening to binaural recordings disorienting at first. Taking periodic breaks will help minimize any disorienting effects. If you experience any adverse effects (dizziness, nausea, etc.) please stop your session immediately.

Your Hearing: For this research study it is important that participants have "normal" hearing. If you have any known permanent hearing impairment, you need not reveal this fact but the project principal investigator asks that you do not take part in this study. As part of your first listening session, you will take part in a brief hearing test to verify your hearing acuity.

If you are experiencing a temporary physical or psychological hearing impairment (including effects due to head cold, ear infection, ringing in your ears from a recent concert, sleep deprivation or alcohol/drug use) on the day of a scheduled listening session, the project principal investigator asks that you reschedule your session. For legal purposes, do not explain the specific reason for rescheduling. Stating that you have a "hearing conflict" as a reason is all that is requested. No record will be kept of this information. If you volunteer information about illegal activities, the project principal investigator is obligated by law to report your activities to the proper authorities.

If you are not an expert in critical listening, it is recommended that you not engage in a critical audio mixing activity for several hours after completing a listening session. As with all such experiences, auditory fatigue and/or listening to "strange" sounds for an extended period of time can affect your personal reference points for what "sounds good". If you plan to hit the studio or you have to mix a show right after a scheduled session you are encouraged to reschedule your session on grounds of a "hearing conflict".

Compensation: You will be given compensation of $5 (US or Canadian dollars depending on session location) for your time for each listening session at the beginning of each session. If you decide for whatever reason that you do not wish to complete the session, you may keep the compensation. If you do not pass the brief hearing acuity test, you may keep the compensation for your first listening session. You will be asked to sign a receipt each time you participate in a listening session and receive monetary compensation.

You will not receive any direct benefit from participating in this study. There are no alternative versions of this study. Your alternative is not to participate.

Results: While you are involved in this study, you will not be told what was specifically done to create the differences between the sound clips that you will hear. This is a standard policy employed to prevent certain factors and biases from interfering with the validity of your session results. In other words, the project principal investigator is investigating what you hear, not what you might think you hear. At the conclusion of the study the project principal investigator will be happy to provide you with the published results and an explanation of the sound system optimization procedures involved.

Confidentiality: This research represents the core phase of the writing of a doctoral dissertation. Like all such works, the finished text will be available to the public. The information may also be published in the *Journal of the Audio Engineering Society*, and presented at meetings of the Audio Engineering Society. Your listener experience category, age and sex are the only potentially identifying pieces of information that will be included in publications and presentations. All other information obtained about you will be held strictly confidential. Your session answers are stored in a data file which will be stored on a password-protected computer which is disconnected from the internet. Your data files will be coded so that only the project principal investigator knows which are yours. Your name does not appear in your data file. Signed consent forms will be kept in a locked safe in the project principal investigator's office. Two years after the conclusion of the study, all documentation which contains personal identifiable information will be destroyed (shredded or deleted). This consent form will be retained for a period of 3 years after the conclusion of this study. Coded data files will be kept for potential use in future studies, however the codes which could be used to identify individual participants will be destroyed. Files will be renamed with anonymous identifiers (Subject #1, etc.)

While you are involved with this study, the project principal investigator asks that you do not discuss details of the study with anyone else who may be participating in the study. Any details that you share with another participant may affect the results of their sessions. Of course issues, complaints and problems should be directed to the appropriate authorities.

Contact: If you have any questions about this research study, you may contact the following:
Principal Investigator (PI): Timothy Ryan – 774-238-8930, CCM.audio.research@gmail.com
Co-Principal Investigator (Co-PI): Prof. Laura Kretschmer – 513-558-8514
Project Supervisor: Prof. René Quesnel – 514-398-4535 x089496
Research Ethics Officer (McGill): Lynda McNeil – 514-398-6831, lynda.mcneil@mcgill.ca
Institutional Review Board (UC): Chairperson – 513-558-5784

For the portion of this study that takes place at the University of Cincinnati:
The University of Cincinnati Institutional Review Board – Social and Behavioral Sciences reviews all non-medical research projects that involve human participants to be sure the rights and welfare of participants are protected. If you have questions about your rights as a participant, you may contact the Chairperson of the University of Cincinnati Institutional Review Board – Social and Behavioral Sciences at 513-558-5784. If you have a concern about the study you may also call the UC Research Compliance Hotline at (800) 889-1547, or you may write to the Institutional Review Board-Social and Behavioral Sciences, G-08 Wherry Hall, ML 0567, 3225 Eden Avenue, PO Box 670567, Cincinnati, OH 45267-0567, or you may email the IRB office at irb@ucmail.uc.edu.

No Obligation: Participation in this study is voluntary. You do not have to participate in this research study. Refusal to participate will involve no penalty or loss. You have the right to withdraw from this study at any point, for any reason, without penalty or loss.

Consent: This document is designed to inform you of the nature of this study, the procedures involved, potential risks/benefits to you, the degree of confidentiality involved and the potential conflicts of interest involved/associated with you participation. If you agree to be involved with this study, the project principal investigator is required to obtain your informed consent. This means that you understand and agree with everything that is written in this document, and you give your permission for the project principal investigator to use your data for his research as specified in the above section "Confidentiality".

A signed copy of this form will be retained for project records as detailed above. An additional copy will be given to you to keep for your records. Do not sign this form until you receive that copy.

Signatures: By signing below you indicate that you understand everything that has been said in this document, and you give your consent to be a participant in this study, and you give your consent to have your results published as specified above and you attest that you have no known permanent hearing damage.

_____ _____
Participant Date

_____
Participant's Printed Name

_____ _____
Principal Investigator (Timothy Ryan) Date

Appendix G: Initial Session Questionnaire

Please answer the following questions to the best of your ability:

GENERAL:

This information will be used to determine if you are eligible to participate in this study. The only information about you that may be used for publication is age, sex and whether you believe you have difficulty understanding speech (general questions 2, 4, and 7). Your name and all other information will be held confidential. You will be assigned a code number, and all data from you will be referenced by that number (e.g. The data from Subject #5…)

1) Name: _____ 2) Age: _____

3) Native English speaker? Y ____ N ____ 4) Sex: M ____ F ____

5) Do you have any known permanent hearing impairment? N ____
NOTE: "No" is the only available answer here. If you do have a permanent hearing impairment, you can not take part in this study.

6) Do you experience temporary hearing impairments often? Y ____ N ____

7) Do you believe that you have more difficulty understanding speech than other people? Y ____ N ____

AUDIO / EVALUATION EXPERIENCE:

This information will be used to determine which experience level category your data will be assigned to. The only information about you that will be used in publication is experience level category. Publications may however include a general description of the group of participants in a category (e.g. "Participants in Category 3 include live sound engineers, broadcast engineers and professional musicians.")

1) Which of the following describes you (check all that apply)

____ Professional live sound engineer
____ Professional broadcast engineer
____ Student of live sound or recording engineering
____ Professional (or college major) musician
____ I listen to music on a regular basis
____ Professional recording engineer
____ Student of broadcast engineering
____ I Listen to loud music often (iPods count)
____ Amateur musician
____ I listen to speech or "books on tape" regularly
____ Professional or student of any type of subjective assessment or evaluation
____ I have no experience with sound systems, sound engineering or subjective assessment and evaluation of sound

2) How familiar are you with concepts such as *reverberation*, *sound field* and *signal summation*?

____ Very familiar (I can define and use them in a sentence)
____ Somewhat familiar (I have heard the terms before, and I have some idea of what they are)
____ Unfamiliar (I do not know much or anything about these concepts)

3) How familiar are you with subjective evaluation (listening tests)?

NOTE: A "listening test" is not to be confused with a "hearing evaluation" which is a procedure performed by a doctor or audiologist to determine if you have hearing impairments.

____ Very familiar (I have administered listening tests/I am a researcher in subjective evaluation)
____ Familiar (I have been a participant in listening tests in the last 6 months)
____ Somewhat familiar (I have participated in a listening test in the past)
____ Unfamiliar (I have never participated in a subjective evaluation listening test)

4) How often have you listened to binaural recordings via headphones?

____ Often ____Rarely ____ Once ____ Never

END Thank you for completing this questionnaire.

PI: Timothy Ryan Sound Design Area, University of Cincinnati CCM.audio.research@gmail.com
Sound Recording Area, McGill University
Supervisor: René Quesnel Sound Recording Area, McGill University
Co-PI: Laura Kretschmer Department of Communication Sciences and Disorders, University of Cincinnati

Appendix H: Listening Test Instructions

Pilot Study Instructions

Overview:
You will be working with a pair of headphones, a laptop computer and a set of pre-numbered response sheets.
You will do a brief warm-up, then you will do the listening test.

Training / Warm-up:

The warm-up has two purposes:
1) To familiarize you with the different types of sounds that you will hear
2) To familiarize you with the pace of testing and response method

Sounds:
Playback of sounds is done through iTunes.
All files are in the playlist.
The file name for each set corresponds to a pre-numbered response sheet.

For each repetition, you will hear a target word embedded within a sentence:
"Number (x / 50): Circle the _____ again."
Your task is to identify the target word and select it from a list of 6 possible choices.

Listen to the sounds carefully – get familiar with them – try to get a feel for this type of listening. These are binaural recordings; they may take a little time to get accustomed to.

Do not over-think your level of sureness. React swiftly, but not frantically.

This is not a test of how well YOU hear. This is not a contest. This is not a test of your intelligence. Don't think. Don't look for patterns. Listen, respond, repeat… that's it.

Listening Test:

A **rep** is one word. A **set** is 50 reps. A set should take you 3-4 minutes.
A **session** contains 12 **sets** (approximately 45 minutes with breaks).
You should take a break whenever you feel tired or distracted (just hit the **PAUSE button**).

I will be right next door if you have any problems.

If you feel yourself getting tired or distracted, take a break!

PI: Timothy Ryan Sound Design Area, University of Cincinnati CCM.audio.research@gmail.com
Sound Recording Area, McGill University
Supervisor: René Quesnel Sound Recording Area, McGill University

Main Study Instructions

Overview:

You will be working with a pair of headphones and a laptop computer.
You will do a brief warm-up, then you will do the listening test.

Training / Warm-up:

The warm-up has two purposes:
1) To familiarize you with the different types of sounds that you will hear
2) To familiarize you with the computer program

Sounds:
For each repetition, you will hear a target word embedded within a sentence: "Circle the _____ again."
Your task is to identify the target word and select it from a list of 6 possible choices.

Listen to the sounds carefully – get familiar with them – try to get a feel for this type of listening. These are binaural recordings; they may take a little time to get accustomed to.

Computer Interface:
To **play** a rep, press the **PLAY button**.
Listen to the test sentence.
Press a **word button** to select one of the 6 words.
You must select one of the 6 choices.

You can only PLAY the sentence once.

Do not over-think your level of sureness. If you find yourself taking more than 3-5 seconds, you are over-thinking. React swiftly, but not frantically.

This is not a test of how well YOU hear. This is not a contest. This is not a test of your intelligence. Don't think. Don't look for patterns. Listen, respond, repeat… that's it.

Listening Test:

A **rep** is one word. A **set** is 50 reps. A set should take you 5-6 minutes.
A **session** contains 8 **sets** (approximately 60 minutes with breaks).
You should take a short break after every 2 sets (more often if you like… just don't hit the **PLAY button**).

To begin a set, press the **UP ARROW** on the keyboard
Enter the appropriate **USER** and **SESSION** numbers, then press **ENTER**
The test window will appear and you can begin

I will be right next door if you have any problems.

If you feel yourself getting tired or distracted, take a break!

Please Note: This equipment is very expensive. Please treat it with the utmost care.

PI: Timothy Ryan Sound Design Area, University of Cincinnati CCM.audio.research@gmail.com
Sound Recording Area, McGill University
Supervisor: René Quesnel Sound Recording Area, McGill University
Co-PI: Laura Kretschmer Department of Communication Sciences and Disorders, University of Cincinnati

Appendix I: Post Session Oral Script

Provide the participant with a copy of this script after each session:

Administrator to Participant:

1) Did everything go well?

2) How do your ears feel?

3) Did you experience any problems, risks or discomforts?

4) Do you or did you have any concerns?

5) Did you have any issues with the testing apparatus?

6) (If applicable) Would you like to schedule your next session at this point? You can schedule by email if you like. Remember of course that you do not have to continue participating in this study if you do not wish to.

Notes to Administrator:

If the participant had issues with the testing apparatus please make note of the type of issue. Do not make record of the identity of the participant.

If the participant experienced auditory fatigue or mild dizziness, recommend to the participant that they take more frequent breaks in the future.

If the participant experienced any serious problems or ailments, or if the participant has any concerns, please direct the participant to contact the appropriate individual / agency.

For test-related issues: Timothy Ryan
For ethics-related issues (UC): Laura Kretschmer or UC IRB
For ethics-related issues (McGill): René Quesnel or McGill Research Ethics Officer
For major health issues: Seek medical attention immediately

Note to Participant:
It is very important that you report ANY and ALL problems. The test administrator is obligated to assist you in reporting problems. If the test administrator attempts to discourage you from reporting a problem, THAT is a problem. Report it.

Principal Investigator: Timothy Ryan – 774-238-8930, CCM.audio.research@gmail.com
Co-Principal Investigator: Prof. Laura Kretschmer – 513-558-8514
Project Supervisor: Prof. René Quesnel – 514-398-4535 x089496
Research Ethics Officer (McGill): Lynda McNeil – 514-398-6831, lynda.mcneil@mcgill.ca
Institutional Review Board (UC): Chairperson – 513-558-5784

PI: Timothy Ryan Sound Design Area, University of Cincinnati CCM.audio.research@gmail.com
Sound Recording Area, McGill University
Supervisor: René Quesnel Sound Recording Area, McGill University
Co-PI: Laura Kretschmer Department of Communication Sciences and Disorders, University of Cincinnati

Appendix J: MRT Response Sheet

1.		2.		3.		4.		5.	
FANG	BANG	MARK	BARK	PEEL	KEEL	TANG	TAB	SICK	SIT
RANG	HANG	PARK	HARK	FEEL	EEL	TAM	TAP	SING	SIN
GANG	SANG	LARK	DARK	REEL	HEEL	TACK	TAN	SILL	SIP
6.		**7.**		**8.**		**9.**		**10.**	
MASS	MAP	PUB	PUG	HOP	POP	BEST	WEST	CUFF	CUP
MAD	MAN	PUTT	PUFF	TOP	COP	NEST	REST	CUD	CUT
MAT	MATH	PUN	PUP	SHOP	MOP	TEST	VEST	CUB	CUSS
11.		**12.**		**13.**		**14.**		**15.**	
SALE	SAKE	DUST	RUST	HEAVE	HEAL	DIM	DIN	TOOK	LOOK
SAFE	SAVE	JUST	GUST	HEATH	HEAP	DID	DIG	COOK	HOOK
SANE	SAME	BUST	MUST	HEAR	HEAT	DIP	DILL	BOOK	SHOOK
16.		**17.**		**18.**		**19.**		**20.**	
SAP	SAT	GUN	RUN	PAGE	PALE	GOT	HOT	TICK	WICK
SAG	SASS	BUN	NUN	PANE	PAY	TOT	POT	PICK	SICK
SACK	SAD	SUN	FUN	PAVE	PACE	LOT	NOT	KICK	LICK
21.		**22.**		**23.**		**24.**		**25.**	
WIT	FIT	KITH	KIT	FOIL	OIL	FIG	RIG	PEACH	PEAS
SIT	HIT	KISS	KID	COIL	TOIL	PIG	WIG	PEAL	PEAK
BIT	KIT	KING	KILL	SOIL	BOIL	BIG	JIG	PEAT	PEACE
26.		**27.**		**28.**		**29.**		**30.**	
PILL	PIP	SUP	SUNG	FIZZ	FIT	BENT	TENT	PAT	PANG
PIG	PIN	SUN	SUM	FILL	FIB	WENT	DENT	PASS	PAN
PIT	PICK	SUD	SUB	FIG	FIN	SENT	RENT	PAD	PATH
31.		**32.**		**33.**		**34.**		**35.**	
TEACH	TEAR	DUD	DUN	BEAK	BEAM	WAY	SAY	THEN	HEN
TEAK	TEAM	DUB	DULL	BEAT	BEAD	MAY	DAY	PEN	MEN
TEAL	TEASE	DUG	DUCK	BEACH	BEAN	GAY	PAY	TEN	DEN
36.		**37.**		**38.**		**39.**		**40.**	
PAW	SAW	LANE	LACE	PALE	TALE	TILL	BILL	BED	WED
THAW	LAW	LAKE	LAY	MALE	GALE	HILL	KILL	FED	LED
JAW	RAW	LAME	LATE	BALE	SALE	FILL	WILL	RED	SHED
41.		**42.**		**43.**		**44.**		**45.**	
HOLD	GOLD	BUN	BUFF	SEED	SEEM	SIN	TIN	NEAT	HEAT
FOLD	COLD	BUG	BUCK	SEEP	SEEN	WIN	DIN	BEAT	MEAT
SOLD	TOLD	BUT	BUS	SEETHE	SEEK	FIN	PIN	SEAT	FEAT
46.		**47.**		**48.**		**49.**		**50.**	
FAME	NAME	SIP	RIP	BATH	BACK	CAKE	CAPE	RACE	RATE
CAME	SAME	TIP	DIP	BAN	BAD	CASE	CANE	RAKE	RAY
GAME	TAME	LIP	HIP	BASS	BAT	CAVE	CAME	RAZE	RAVE

Reference List

[1] Ahnert, W. "Comb-filter Distortions and their Perception in Sound Reinforcement Systems," from the proceedings of the 84th convention of the Audio Engineering Society, Paris, France, Preprint 2565, March 1988.

[2] Ahnert, W., Feistel, R., "EARS Auralization Software," J. Audio Eng. Soc., 41(11), 894-904, 1993.

[3] Ahnert, W., Feistel, S., Schmitz, O., "Implementation of Intelligibility Algorithms into EASE 4.0," from the proceedings of the 21st international conference of the Audio Engineering Society, St. Petersburg, Russia, June 2002.

[4] Ahnert, W., Feistel, S., Maier, T., Miron, A. R., "Loudspeaker Time Alignment using Live Sound Measurements," from the proceedings of the 124th convention of the Audio Engineering Society, Amsterdam, the Netherlands, Preprint 7433, May 2008.

[5] ANSI, "Method for Measuring the Intelligibility of Speech Over Communications Systems," (ANSI S3.2-1989), American National Standards Institute, New York, NY, 1989.

[6] ANSI, "American National Standard for an Occluded Ear Simulator," (ANSI S3.25-1979), American National Standards Institute, New York, NY, 1979.

[7] ANSI, "Specification for Manikin for Simulated in situ Airborne Acoustic Measurements," (ANSI S3.36-1985), American National Standards Institute, New York, NY, 1985.

[8] Antilla, M., Kataja, J., Valimaka, V., "Sound Directivity Control Using Striped Panel Loudspeakers," from the proceedings of the 110th convention of the Audio Engineering Society, Amsterdam, The Netherlands, Preprint 5306, May 2001.

[9] Atkinson, D. J., Catellier, A. A., "Intelligibility of Selected Radio Systems in the Presence of Fireground Noise: Test Plan and Results," Technical Report TR-08-453, U.S. Department of Commerce, National Telecommunications and Information Administration, June 2008.

[10] Augspurger, G., Brawley, J., "An Improved Colinear Array," from the proceedings of the 74th convention of the Audio Engineering Society, New York, NY, Preprint 2047, Oct. 1983.

[11] Augspurger, G., Bech, S., Brook, R., Cohen, E., Eargle, J., Schindler, T. A., "Use of Stereo Synthesis to Reduce Subjective/Objective Interference Effects: The perception of Comb Filtering, Part 2," from the proceedings of the 87th convention of the Audio Engineering Society, New York, U.S.A., Preprint 2862, Oct. 1989.

[12] Augspurger, G., "Near-Field and Far-Field Performance of Large Woofer Arrays," J. Audio Eng. Soc., 38(4), 231-236, 1990.

[13] Bauer, B. B., Rosenheck, A. J., Abbagnaro, L. A., "External-Ear Replica for Acoustical Testing," J. Acoust. Soc. Am. 42(1), 204-207, July 1967.

[14] Bech, S., Zacharov, N., *Perceptual Audio Evaluation – Theory, Method and Application*, John Wiley & Sons Ltd., West Sussex, England, 2006.

[15] Bell, D. W., Kreul, E. J., Nixon, J. C., "Reliability of the Modified Rhyme Test for Hearing," J. of Speech and Hearing Research, 15, 287-295, 1972.

[16] Beranek, L. L., "Loudspeakers and Microphones," J. Acoust. Soc. Am. 26(5), 618-629, Sept. 1954.

[17] Beranek, L. L., *Concert and Opera Halls: How They Sound*, Acoustical Society of America, Woodbury, New, York, 1996.

[18] Beutelmann, R., Brand, T., "Prediction of Speech Intelligibility in spatial Noise and Reverberation for Normal-Hearing and Hearing-Impaired Listeners," J. Acoust. Soc. Am. 120(1), 331-342, July 2006.

[19] Beutalmann, R., Brand, T., Kollmeier, B., "Prediction of Binaural Speech Intelligibility with Frequency-Dependent Interaural Phase Differences," J. Acoust. Soc. Am. 126(3), 1359-1368, Sept. 2009.

[20] Beutelmann, R., Brand, T., Kollmeier, B., "Revision, Extension, and Evaluation of a Binaural Speech Intelligibility Model," J. Acoust. Soc. Am. 127(4), 2479-2497, April 2010.

[21] Beyer, M. R., Webster, J. C., Dague, D. M., "Revalidation of the Clinical Test Version of the Modified Rhyme Words," J. of Speech and Hearing Research, 12, 374-378, 1969.

[22] Blauert, J., Pösselt, C., "Application of Modeling Tools in the Process of Planning Electronic Room Acoustics," from the proceedings of the 6th international conference of the Audio Engineering Society, Nashville, TN, May 1988.

[23] Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised edition, The MIT Press, Cambridge, MA, 1997.

[24] Boone, M., Ouweltjes, O, "Design of a Loudspeaker System with a Low-Frequency Cardioidlike Radiation Pattern, "J. Audio Eng. Soc., 45(9), 702-707, 1997.

[25] Boner, C. P., Boner, C. R., "A Procedure for Controlling Room-Ring Modes and Feedback Modes in Sound Systems with Narrow-Band Filters," J. audio Eng. Soc., 13, 297-299, 1965.

[26] Braasch, J. Personal Communication, January 2010.

[27] Breshears, V., Hinz, R., "An Integrated 3-Way Constant Directivity Speaker Array," from the proceedings of the 101st convention of the Audio Engineering Society, Los Angeles, CA, Preprint 4323, Nov. 1996.

[28] Brown, J., "Systems for Stereophonic Sound Reinforcement: Performance Criteria, Design Techniques, and Practical Examples," from the proceedings of the 113th convention of the Audio Engineering Society, Los Angeles, California, Preprint 5666, Oct. 2002.

[29] Brüel & Kjaer, Head and Torso Simulator Type 4128. Product Information.

[30] BSS, Omnidrive FDS-355. Product Information.

[31] Burandt, U., Pösselt, C., Ambrozus, S., Hosenfeld, M., Knauff, V., "Anthropometric Contribution to Standardizing Manikins for Artificial-Head Microphones and to Measuring Headphones and Ear Protectors," Applied Ergonomics, 22.6, 373-378, 1991.

[32] Burkhard, M. D., Sachs, R. M., "Anthropometric Manikin for Acoustic Research," J. Acoust. Soc. Am., 58(1), 214-222, 1975.

[33] Burkhard, M. D., "Measuring the Constants of Ear Simulators," J. Audio Eng. Soc., 25(12), 1008-1015, 1977.

[34] Cabot, R. C., "Equalization, Current Practice and New Directions," from the proceedings of the 6th international conference of the Audio Engineering Society, Nashville, Tennessee, May 1988.

[35] Christensen, F., Jensen, C. B., Møller, H., "The Design of VALDEMAR – An Artificial Head for Binaural Recording Purposes," from the proceedings of the 109th convention of the Audio Engineering Society, Los Angeles, California, Preprint 5253, Sept. 2000.

[36] Cohen, A. B., "Mechanical Crossover Characteristics in Dual Diaphragm Loudspeakers," J. Audio Eng. Soc., 5(1), 11-17, Jan. 1957.

[37] Dau, T., Kollmeier, B., "Modeling Auditory Processing of Amplitude Modulation. I. Detection and Masking with Narrow-Band Carriers," J. Acoust. Soc. Am. 102(5), Pt. 1, Nov. 1997.

[38] Dau, T., Kollmeier, B., Kohlrausch, A., "Modeling Auditory Processing of Amplitude Modulation. II. Spectral and Temporal Integration," J. Acoust. Soc. Am. 102(5), Pt. 1, Nov. 1997.

[39] Davis, C., "Measurement of %Alcons," J. Audio Eng. Soc., 34(11), 1986.

[40] Davis, D., "Equivalent Acoustic Distance," J. Audio Eng. Soc., 21(8), 646-649, 1973.

[41] Davis, D., Davis, C., "Application of Speech Intelligibility to Sound Reinforcement," J. Audio Eng. Soc. 37(12), 1002-1019, 1989.

[42] Davis, D., Davis, C., *Sound System Engineering*, 2nd Edition, Focal Press, Newton, MA, 1997.

[43] Davis, D., Davis, C., *If Bad Sound Were Fatal, Audio would be the Leading Cause of Death*, 1stBooks, Bloomington, IN, 2004.

[44] Dickens, D., "Automated Time Alignment and Equalization of a Speaker Array for Sound Field Reproduction," from the proceedings of the 6th Australian regional conference of the Audio Engineering Society, Melbourne, Australia, Preprint 4317, Sept. 1996.

[45] Dubbelboer, F., Houtgast, T., "A Detailed Study on the Effects of Noise on Speech Intelligibility," J. Acoust. Soc. Am., 122(5), 2865-2871, 2007.

[46] Eargle, J., *Loudspeaker Handbook*, 2nd Edition, Kluwer Academic Publishers, Norwell, MA, 2003.

[47] Eargle, J., Gander, M., "Historical Perspectives and Technology Overview of Loudspeakers for Sound Reinforcement," J. Audio Eng. Soc., 52(4), 412-432, 2004.

[48] EASE 4.2 Software Manual, Acoustic Design Ahnert, 2009.

[49] EASERA 1.0.6 Software Manual, Software Design Ahnert, 2007.

[50] Egan, J. P., "Articulation Testing," Laryngoscope, 58, 955-991, 1948.

[51] Eggenschwiler, K., Machner, R., "Intercomparison Measurements of Room Acoustical Parameters and Measures for Speech Intelligibility in a Room with a Sound System," J. Audio Eng. Soc. 53(3), 199-204, Mar. 2005.

[52] Elkins, E. F., "Evaluation of Modified Rhyme Test Results from Impaired- and Normal-Hearing Listeners," J. of Speech and Hearing Research, 14, 589-595, 1971.

[53] El-Saghir, E., Maher, M., "Virtual Shifting of Speaker Array Components," from the proceedings of the 106th convention of the Audio Engineering Society, Munich, Germany, Preprint 4893, May 1999.

[54] Everest, F. A., *Master Handbook of Acoustics*, 4th Edition, McGraw-Hill, New York, NY, 2001

[55] Fairbanks, G., "Test of Phonetic Differentiation: The Rhyme Test," J. Acoust. Soc. Am., 30(7), 596-600, Jul. 1958.

[56] Fidlin, P., Carlson, D., "The Basic Concepts and Problems Associated with Large-Scale Concert Sound Loudspeaker Arrays," from the proceedings of the 86th convention of the Audio Engineering Society, Hamburg, Germany, Preprint 2802, Mar. 1989.

[57] Fidlin, P., Carlson, D., "Comparative Performance of Three Types of Directional Devices Used as Concert-Sound Loudspeaker Array Elements," J. Audio Eng. Soc., 38(4), 271-295, 1990.

[58] Fletcher, H., Galt, R. H., "The Perception of Speech and its Relation to Telephony," J. Acoust. Soc. Am. 22(2), 89-151, Mar. 1950.

[59] French, N. R., Steinberg, J. C., "Factors Governing the Intelligibility of Speech Sounds," J. Acoust. Soc. Am., 19(1), 90-119, 1947.

[60] Gander, M. R., Eargle, J. M., "Measurement and Estimation of Large Loudspeaker Array Performance", J. Audio Eng. Soc., 38(4), 204 - 220, 1990.

[61] Gander, M. R., "Fifty Years of Loudspeaker Developments as Viewed Through the Perspective of the Audio Engineering Society," J. Audio Eng. Soc. 46(1/2), 43-58, Feb. 1998.

[62] Gelfand, S. A., *Essentials of Audiology*, 3rd Edition, Thieme Medical Publishers, New York, NY, 2009

[63] Gierlich, H. W., Genuit, K., "Processing Artificial-Head Recordings," J. Audio Eng. Soc., 37(1/2), 34-39, 1989.

[64] Gierlich, H. W., "The Application of Binaural Technology," Applied Acoustics, 36, 219-243, 1992.

[65] Green, I., Maxfield, J., "Public Address Systems", J. Audio Eng. Soc., 25(4), 184-195, 1977.

[66] Griesinger, D., "Equalization and Spatial Equalization of Dummy Head Recordings for Loudspeaker Reproduction," from the proceedings of the 85th convention of the Audio Engineering Society, Los Angeles, California, Nov. 1988.

[67] Haas, H., "The Influence of a Single Echo on the Audibility of Speech," J. Audio Eng. Soc., 20(2), 146-159, 1972.

[68] Harrell, J., "End-Fire Line Array of Loudspeakers," J. Audio Eng. Soc., 43(7/8), 581-591, 1995

[69] Heil, C., "Sound Fields Radiated by Multiple Sound Source Arrays," from the proceedings of the 92nd convention of the Audio Engineering Society, Vienna, Austria, Preprint 3269, Mar. 1992.

[70] Hilliard, J., "Unbaffled Loudspeaker Column Arrays," J. Audio Eng. Soc., 672-673, 1970.

[71] Hochhaus, L. and Antes, J. R., "Speech identification and 'knowing that you know'." Journal of Perception and Psychophysics, 13, 131-132, 1973.

[72] Holland, K. R., Fahy, F. J., "A Low-Cost End-Fire Acoustic Radiator," J. Audio Eng. Soc., 39(7/8), 540-550, 1991.

[73] Holley, S. C., Lerman, J., Randolph, K., "A Comparison of the Intelligibility of Esophageal, Electrolaryngeal, and Normal Speech in Quiet and in Noise," J. Communication Disorders, 16, 143-155, 1983.

[74] Holman, T., "New Factors in Sound for Cinema and Television," J. Audio Eng. Soc., 39(7/8), 529-539, 1991.

[75] House, A. S., Williams, C. E., Hecker, M. H. L., Kryter, K. D., "Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set," J. Acoust. Soc. Am., 37(1), 158-166, 1965.

[76] Houtgast, T., Steeneken, H. J. M., "The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility," Acustica, 28, 66-73, 1973.

[77] Houtgast, T., Steeneken, H. J. M., "A Review of the MTF Concept in Room Acoustics and its use for Estimating Speech Intelligibility in Auditoria," J. Acoust. Soc. Am., 77(3), 1069-1077, Mar. 1985.

[78] Humes, L. E., Boney, S., Loven, F., "Further Validation of the Speech Transmission Index (STI)," J. Speech and Hearing Research, 30, 403-410, 1987.

[79] IEC, "Sound Level Meters," (IEC 651), International Electrotechnical Commission, Geneva, Switzerland, 1979.

[80] IEC, "Occluded-Ear Simulator for the Measurement of Earphones Coupled to the Ear by Ear Inserts," (IEC 711, First Edition), International Electrotechnical Commission, Geneva, Switzerland, 1981.

[81] IEC, "Provisional Head and Torso Simulator for Acoustic Measurements on Air Conduction Hearing Aids," (IEC 959, First Edition), International Electrotechnical Commission, Geneva, Switzerland, 1990.

[82] ISO, "Background Acoustic Noise Levels in Theatres, Review Rooms and Dubbing Rooms," (ISO-9568), International Standards Organization, Geneva, Switzerland, 1993.

[83] ISO, "Acoustics – Measurement of the Reverberation Time of Rooms with reference to Other Acoustical Parameters," (ISO-3382), International Standards Organization, Geneva, Switzerland, 1997.

[84] ITU, "Objective Measurement of Active Speech Level," (ITU-T P.56), International Telecommunications Union, Geneva, Switzerland, 1993.

[85] ITU, "Head and Torso Simulator for Telephonometry," (ITU-T P.58), International Telecommunications Union, Geneva, Switzerland, 1996.

[86] ITU, "Multichannel Stereophonic Sound System with and without Accompanying Picture," (ITU-R BS.775-2), International Telecommunications Union, Geneva, Switzerland, 2006.

[87] Janssen, J. H., "A Method for the Calculation of Speech Intelligibility under Conditions of Reverberation and Noise," Acustica, 7, 305-310, 1957.

[88] Keele, D. B., "Effective Performance of Belles Arrays," J. Audio Eng. Soc., 38(10), 723-748, 1990.

[89] Keele, D. B., "Implementation of Straight Line and Flat Panel Constant Beamwidth Transducer (CBT) Loudspeaker Arrays using Signal Delays," from the proceedings of the 113th convention of the Audio Engineering Society, Los Angeles, CA, Preprint 5653, Oct. 2002.

[90] Killion, M. C., "Equalization Filter for Eardrum-Pressure Recording Using a KEMAR Manikin," J. Audio Eng. Soc., 27(1/2), 13-16, 1979.

[91] King, R. Personal Communication, September 2010.

[92] Klepko, J., "5-Channel Microphone Array with Binaural-Head for Multichannel Reproduction," doctoral thesis, Faculty of Music, McGill University, Sept. 1999.

[93] Klepper, D. L., Steele, D. W., "Constant Directional Characteristics from a Line Source Array," J. Audio Eng. Soc., 11(3), 198-202, 1963.

[94] Klepper, D., "Time-Delay Units for Sound Reinforcement Systems," J. Audio Eng. Soc., 15(2), 176-179, 1967.

[95] Kreul, E. J., Nixon, J. C., Kryter, K. D., Bell, D. W., Lang, J. S., Schubert, E. D., "A Proposed Clinical Test of Speech Discrimination, J. Speech and Hearing Research, 11, 536-552, 1968.

[96] Kreul, E. J., Bell, D. W., Nixon, J. C., "Factors Affecting Speech Discrimination Test Difficulty," J. of Speech and Hearing Research, 12, 281-287, 1969.

[97] Kryter, K. D., "Methods for the Calculation and use of the Articulation Index," J. Acoust. Soc. Am., 34(11), 1689-1697, 1962.

[98] Kuriyama, J., Tokko, T., Suzuki, S., Hashiguchi, Y., "A Compact Digital Signal Processing system Consisting of DSP Modules,", from the proceedings of the 86th convention of the Audio Engineering Society, Hamburg, Germany, Preprint 2771, Mar. 1989.

[99] Kusumoto, A., Arai, T., Kinoshita, K., Hodoshima, N., Vaughn, N., "Modulation Enhancement of Speech by a Pre-Processing Algorithm for Improving Intelligibility in Reverberant Environments," Speech Communication, 45, 101-113, 2005.

[100] Larcher, V., Vandernoot, G., Jot, J., "Equalization Methods in Binaural Technology," from the proceedings of the 105th convention of the Audio Engineering Society, San Francisco, California, Sept. 1998.

[101] Leembruggen, G., Packer, N., Goldburg, B., Backstrom, D., "Development of a Shaded, Beam Steered Line Array Loudspeaker with Integral Amplification and DSP Processing," from the proceedings of the 105th convention of the Audio Engineering Society, San Francisco, California, Sept. 1998.

[102] Leembruggen, G., Hippler, M., Mapp, P., "Further Investigations into Improving STI's Recognition of the Effects of Poor Frequency Response on Subjective Intelligibility," from the proceedings of the 128th convention of the Audio Engineering Society, London, UK, May 2010.

[103] Leonard, J., *Theatre Sound*, Routledge, New York, NY, 2001.

[104] Li, F. F., "Speech Intelligibility of VoIP to PSTN Interworking – A Key Index for the QoS," from the proceedings of the IEE Telecommunications Quality of Services Conference, London, UK, March, 2004.

[105] Lochner, J. P. A., Burger, J. F., "The Subjective Masking of Short Time Delayed Echoes by Their Primary Sounds and Their Contribution to the Intelligibility of Speech," Acustica, 8, 1-10, 1958.

[106] Lochner, J. P. A., Burger, J. F., "Intelligibility of Speech under Reverberant Conditions," Acustica, 11, 195-200, 1961.

[107] Lochner, J. P. A., Burger, J. F., "The Influence of Reflections on Auditorium Acoustics," J. Sound Vib., 1(4), 426-454, 1964.

[108] Long, M., *Architectural Acoustics*, Elsevier Academic Press, Burlington, MA, 2006.

[109] Mapp, P., "A Comparison between STI and RASTI Speech Intelligibility Measurement Systems," from the proceedings of the 100th convention of the Audio Engineering Society, Copenhagen, Denmark, Preprint 4279, May 1996.

[110] Mapp, P., "Practical Limitations of Objective Speech Intelligibility Measurements of Sound Reinforcement Systems," from the proceedings of the Audio Engineering Society 102nd Convention, Munich, Germany, Preprint 4410, Mar. 1997.

[111] Mapp, P., "From Loudspeaker to Ear – Measurement and Reality," from the proceedings of the 12th UK conference of the Audio Engineering Society, London, UK, April 1997.

[112] Mapp, P., "Relationships between Speech Intelligibility Measures for Sound Systems," from the proceedings of the 112th convention of the Audio Engineering Society, Munich, Germany, Preprint 5604, May 2002.

[113] Mapp, P., "Modifying STI to Better Reflect Subjective Impression," from the proceedings of the 21st international conference of the Audio Engineering Society, St. Petersburg, Russia, June 2002.

[114] Mapp, P., "Limitations of Current Sound System Intelligibility Verification Techniques," from the proceedings of the 113th convention of the Audio Engineering Society, Los Angeles, CA, Preprint 5668, Oct. 2002.

[115] Mapp, P., "The Acoustic and Intelligibility Performance of Assistive Listening & Deaf Aid Loop (AFILS) Systems," from the proceedings of the 114th convention of the Audio Engineering Society, Amsterdam, The Netherlands, Mar. 2003.

[116] Mapp, P., "Intelligibility – Winning the Acoustics Battle," from the proceedings of the 18th U.K. conference of the Audio Engineering Society, London, England, Apr. 2003.

[117] Mapp, P., "Some Effects of Equalization on Sound System Intelligibility and Measurement," from the proceedings of the Audio Engineering Society 115th Convention, New York, U.S.A., Preprint 5986, Oct. 2003.

[118] Mapp, P., "Systematic & Common Errors in Sound System STI and Intelligibility Measurements," from the proceedings of the 117th convention of the Audio Engineering Society, San Francisco, CA, Preprint 6271, Oct. 2004.

[119] Marcus, S. M., Syrdal, A. K., "Speech: Articulatory, Linguistic, Acoustic and Perceptual Differences," In Syrdal, A., Bennett, R. Greenspan, S. (Ed.), *Applied Speech Technology*, CRC Press, Boca Raton, FL, 1995.

[120] Martens, W. Personal Communication, September 2010.

[121] Matlab Software Manual, Mathworks, 2010.

[122] Matthews, R., Legg, S., Charlton, S., "The Effect of Cell Phone Type on Drivers Subjective Workload During Concurrent Driving and Conversing," Accident Analysis and Prevention, 35, 451-457, 2003.

[123] McBride, M., Hodges, M., French, J., "Speech Intelligibility Differences of Male and Female Vocal Signals Transmitted Through Bone Conduction in Background Noise: Implications for Voice Communication Headset Design," Int. J. of Ind. Ergonomics, 38, 1038-1044, 2008.

[124] McCarthy, B, *Sound Systems: Design and Optimization*, 1st Edition, Focal Press, Burlington, MA, 2007.

[125] McCarthy, B. Personal Communication, March 2009.

[126] Metzler, B., *Audio Measurement Handbook*, Audio Precision, Beaverton, OR, 1993.

[127] Meyer, D., "Digital Control of Loudspeaker Array Directivity," J. Audio Eng. Soc., 32(10), 747-754, 1984.

[128] Meyer, J., Seidel, F., "Large Arrays: Measured Free-Field Polar Patterns Compared to a Theoretical Model of a Curved Surface Source," J. Audio Eng. Soc., 38(4), 260–270, 1990.

[129] Minnaar, P., Olsen, S., Christensen, F., Møller, H., "Localization with Binaural Recordings from Artificial and Human Heads," J. Audio Eng. Soc., 49(5), 323-336, 2001.

[130] Mochimaru, A., "An Evaluation of the Accuracy of MTF-STI Measurements by Comparison to Japanese Three-Syllable Listening Tests," from the proceedings of the 89th convention of the Audio Engineering Society, Los Angeles, California, Preprint 2941, Sept. 1990.

[131] Mochimaru, A., "An Advanced Concept for Loudspeaker Array Design in Sound Reinforcement Systems," from the proceedings of the 91st convention of the Audio Engineering Society, New York, NY, Preprint 3139, October 1991.

[132] Møller, H., "Fundamentals of Binaural Technology," Applied Acoustics, 36, 171-218, 1992.

[133] Møller, H., Hammershøi, D., Jensen, C., Sørensen, M., "Transfer Characteristics of Headphones Measured on Human Ears," J. Audio Eng. Soc., 43(4), 203–217, 1995.

[134] Møller, H., Jensen, C., Hammershøi, D., Sørensen, M., "Using a Typical Human Subject for Binaural Recording," from the proceedings of the 100th convention of the Audio Engineering Society, Copenhagen, Denmark, Preprint 4157, May 1996.

[135] Møller, H., Hammershøi, D., Jensen, C., Sørensen, M., "Evaluation of Artificial Heads in Listening Tests," J. Audio Eng. Soc., 47(3), 83-100, 1999.

[136] Nabelek, A. K., Mason, D., "Effect of Noise and Reverberation on Binaural and Monaural Word Identification by Subjects with Various Audiograms," J. of Speech and Hearing Research, 24, 375-383, 1981.

[137] Neumann, Artificial Head Type KU100. Product Information.

[138] Nixon, J. C., "Investigation of the Response Foils of the Modified Rhyme Hearing Test," J. of Speech and Hearing Research, 16, 658-666, 1973.

[139] Nye, P. W., Gaitenby, J., "Consonant Intelligibility in Synthetic Speech and in a Natural Control (Modified Rhyme Test Results)," Haskins Laboratories Status Report on Speech Research, SR-33, 77-91, 1973.

[140] Oppenheim, A. V., Schafer, R. W., Buck, J. R., *Discrete-Time Signal Processing*, 2nd Edition, Prentice Hall, Upper Saddle River, New Jersey, 1999.

[141] Owens, E., Schubert, E. D., "Development of the California Consonant Test," J. Speech and Hearing Research, 20, 463-474, 1977.

[142] Peterson, A. P. G., Gross, E. E., *Handbook of Noise Measurement*, Ninth Edition, General Radio Company, 1980.

[143] Peutz, V. M. A., "Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room," J. Audio Eng. Soc., 19(11), 915-919, 1971.

[144] Peutz, V. M. A., "Designing Sound Systems for Speech Intelligibility," from the proceedings of the 48th convention of the Audio Engineering Society, Los Angeles, CA, May 1974.

[145] Peutz, V. M. A., "Speech Information and Speech Intelligibility," from the proceedings of the 85th convention of the Audio Engineering Society, Los Angeles, CA, Preprint 2732, Nov. 1988.

[146] Pompetzki, W., "Binaural Recording and Reproduction for Documentation and Evaluation," from the proceeding of the 8th international conference of the Audio Engineering Society, Washington D.C., U.S.A., May 1990.

[147] Ralston, J. V., Pisoni, D. B., "Perception and Comprehension of Speech," In Syrdal, A., Bennett, R. Greenspan, S. (Ed.), *Applied Speech Technology*, CRC Press, Boca Raton, FL, 1995.

[148] Rife, D., Vanderkooy, J., "Transfer Function Measurement with Maximum Length Sequences," J. Audio Eng. Soc., 37(6), 419-444, 1989.

[149] Rijk, K., Breuer, F., Peutz, V., "Speech Intelligibility in Some German Sports Stadiums," J. Audio Eng. Soc., 39(1/2), 37-46, 1991.

[150] Rumsey, F., *Spatial Audio*, Focal Press, Oxford, England, 2001.

[151] Sachs, R. M., Burkhard, M. D., "Earphone Pressure Response in Ears and Couplers," from the proceedings of the 83rd meeting of the Acoustical Society of America, Buffalo, NY, Apr. 1972.

[152] Sachs, R. M., Burkhard, M. D., "Zwislocki Coupler Evaluation with Insert Earphones," Technical Report 8, Knowles Electronics, Nov. 1972.

[153] Scheaffer, R. L., McClave, J. T., *Probability and Statistics for Engineers*, Duxbury Press, Belmont, California, 1995.

[154] Scheirman, D., "Practical Considerations for Field Deployment of Modular Line Array Systems," from the proceedings of the 21st international conference of the Audio Engineering Society, St. Petersburg, Russia, June 2002.

[155] Schlesinger, A., Ramirez, J. P., Boone, M. M., "Evaluation of a Speech-based and Binaural Speech Transmission Index," from the proceedings of the 40th international conference of the Audio Engineering Society, Tokyo, Japan, Oct. 2010.

[156] Schmidt-Nielsen, A., "Intelligibility and Acceptability Testing for Speech Technology," In Syrdal, A., Bennett, R. Greenspan, S. (Ed.), *Applied Speech Technology*, CRC Press, Boca Raton, FL, 1995.

[157] Schwenke, R. Personal Communication, March 2009.

[158] Schoeps, Stereo Sphere Microphone type KFM 6. Product Information.

[159] Shaw, E. A. G., Thiessen, G. J., "Acoustics of Circumaural Headphones," J. Acoust. Soc. Am., 34(9), 1233-1246, 1962.

[160] Sheskin, D., *Handbook of Parametric and Non-parametric Statistical Procedures*, CRC Press, New York, 2004.

[161] Shinn-Cunningham, B., "Distance Cues for Virtual Auditory Space," from the proceedings of the 1st Pacific Rim conference on multi media of the Institute of Electrical and Electronics Engineers, Sydney, Australia, Dec. 2000.

[162] Shirley, B., Kendrick, P., "Measurement of Speech Intelligibility in Noise: A Comparison of a Stereo Image Source and a Central Loudspeaker Source," from the proceedings of the Audio Engineering Society 118th Convention, Barcelona, Spain, Preprint 6372, May 2005.

[163] Shirley, B., Churchill, C., "The Effect of Stereo Crosstalk on Intelligibility: Comparison of a Phantom Stereo Image and a Central Loudspeaker Source," J. Audio Eng. Soc., 55(10), 852-863, 2007.

[164] Smith, H. G., "Acoustic Design Considerations for Speech Intelligibility," J. Audio Eng. Soc., 29(6), 408-415, 1981.

[165] Smith, D., "Discrete Element Line Arrays – Their Modeling and Optimization," J. Audio Eng. Soc., 45(11), 949-964, 1997.

[166] SPSS Software Manual, IBM Corporation, 2010.

[167] Steeneken, H, Houtgast, T., "A Fast Method for the Determination of the Intelligibility of Running Speech," from the proceedings of the 44th convention of the Audio Engineering Society, Rotterdam, The Netherlands, Feb. 1973.

[168] Steeneken, H. J. M., Houtgast, T., "A Physical Method for Measuring Speech-Transmission Quality," J. Acoust. Soc. Am., 67(1), 318-326, 1980.

[169] Summers, V. Cord, M., "Intelligibility of Speech in Noise at High Presentation Levels: Effects of Hearing Loss and Frequency Region," J. Acoust. Soc. Am., 122(2), 1130-1137, 2007.

[170] Teuber, W., Völker, E., "Improved Speech Intelligibility by Use of Time Delays in Sound Systems," from the proceedings of the Audio Engineering Society 84th Convention, Paris, France, Preprint 2562, Mar. 1988.

[171] Thurmond, B., "The Scope of Sound System Performance Optimization," from the proceedings of the 6th international conference of the Audio Engineering Society, Nashville, Tennessee, May 1988.

[172] Toole, F. E., "The Acoustics and Psychoacoustics of Headphones," from the proceedings of the 2nd international conference of the Audio Engineering Society, Anaheim, U.S.A., Preprint 1006, May 1984.

[173] Toole, F. E., "Binaural Record/Reproduction Systems and Their Use in Psychoacoustic Investigations," from the proceedings of the Audio Engineering Society 91st Convention, New York, NY, Preprint 3179, Oct. 1991.

[174] Ureda, M., "'J' and 'Spiral' Line Arrays," from the proceedings of the 111th convention of the Audio Engineering Society, New York, NY, Preprint 5485, Sept. 2001.

[175] Ureda, M., "Line Arrays: Theory and Applications," from the proceedings of the 110th convention of the Audio Engineering Society, Amsterdam, the Netherlands, Preprint 5304, May 2001.

[176] Voiers, W. D., "Diagnostic Evaluation of Speech Intelligibility," In Hawley, M. E. (Ed.), *Speech Intelligibility and Speaker Recognition*, Dowden, Hutchinson and Ross, Stroudsburg, PA, 1977.

[177] Voss, S. E., Allen, J. B., "Measurement of Acoustic Impedance and Reflectance in the Human Ear Canal," J. Acoust. Soc. Am., 95(1), 372-384, 1994.

[178] Van der Wal, M., Start, E. W., De Vries, D., "Design of Logarithmically Spaced Constant-Directivity Transducer Arrays," J. Audio Eng. Soc., 44(6), 497-507, 1996.

[179] Williams, C. E., Hecker, M. H. L., "Relation between Intelligibility Scores for Four Test Methods and Three Types of Speech Distortion," J. Acoust. Soc. Am., 44(4), 1002-1006, 1968.

[180] Wickens, T. D., *Multiway Contingency Tables Analysis for the Social Sciences*, Psychology Press, New York, 1989.

[181] Van Wijngaarden, S., Drullman, R., "Binaural Intelligibility Prediction Based on the Speech Transmission Index," J. Acoust. Soc. Am. 123(6), 4514-4523, 2008.

[182] Yamaha, PM5D Digital Mixing Console. Product Information.

[183] Zwicker, E., Fastl, H., *Psychoacoustics: Facts and Models*, 2nd updated Edition, Springer Publishing House, Berlin, Germany, 1999.

[184] Zwislocki, J., "An Acoustic Coupler for Earphone Calibration," Report LSC-S-7, Laboratory of Sensory Communications, Syracuse University, Syracuse, NY, 1970.

[185] Zwislocki, J., "An Ear-Like Coupler for Earphone Calibration," Report LSC-S-9, Laboratory of Sensory Communication, Syracuse University, Syracuse, NY, 1971.