2011

OBJECTIVE AND SUBJECTIVE EVALUATION OF DEREVERBERATION ALGORITHMS

Jonathan Mark Pietrobon

OBJECTIVE AND SUBJECTIVE EVALUATION OF DEREVERBERATION ALGORITHMS

(Spine title: EVALUATION OF DEREVERBERATION ALGORITHMS)

(Thesis format: Monograph)

by

Jonathan Pietrobon

Graduate Program in Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Engineering Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor

Dr. Vijay Parsa

Supervisory Committee

Dr. Ewan Macpherson

Dr. David Purcell

Examiners

Dr. Serguei Primak

Dr. Ilia Polushin

Dr. Susan Scollie

The thesis by

Jonathan Mark Pietrobon

entitled:

Objective and Subjective Evaluation of Dereverberation Algorithms

is accepted in partial fulfillment of the
requirements for the degree of
Master of Engineering Science

Date

Chair of the Thesis Examination Board

Abstract

Reverberation significantly impacts the quality and intelligibility of speech. Several dereverberation algorithms have been proposed in the literature to combat this problem. A majority of these algorithms utilize a single channel and are developed for monaural applications, and as such do not preserve the cues necessary for sound localization. This thesis describes a blind two-channel dereverberation technique that improves the quality of speech corrupted by reverberation while preserving cues that affect localization. The method is based by combining a short term (2ms) and long term (20ms) weighting function of the linear prediction (LP) residual of the input signal. The developed and other dereverberation algorithms are evaluated objectively and subjectively in terms of sound quality and localization accuracy. The binaural adaptation provides a significant increase in sound quality while removing the loss in localization ability found in the bilateral implementation.

Keywords: Binaural, dereverberation, reverberation, cue preservation, LP residual, localization

Acknowledgements

I would like to express my appreciation and gratitude to Dr. Vijay Parsa for his support and guidance as supervisor for my M.E.Sc. studies at the University of Western Ontario. I would also like to extend my thanks to Dr. Ewan Macpherson for his advice and support designing the localization task.

To my lab-mates and good friends, Julie Seto, Ben Morgan, and Arvind Venkatasubramanian, thank you for your help and support and for sharing this experience. As well, thank you to the staff at the National Centre for Audiology, specifically Steve Beaulac, David Grainger, Paula Folkeard, and Lucy Kieffer.

Finally to my family, thank you for your love and encouragement all these years.

Contents

List of Figures

List of Figures

List of Tables

List of Acronyms

AIR	Aachen impulse response
ANOVA	analysis of variance
APSD	auto-power spectral density
BAC	Beltone Anechoic Chamber
BMYM	binaural modified Yegnanarayana-Murthy
BRIR	binaural room impulse response
CELP	code-excited linear prediction
CPSD	cross-power spectral density
DSB	delay and sum beamformer
DSP	digital signal processing
FDR	false discovery rate
GCC-PHAT	generalized cross-correlation with phase transform
HATS	head and torso simulator
HSE	head shadow effect
IID	interaural intensity difference
ILD	interaural level difference
IPD	interaural phase difference
IS	Itakura-Saito
ITD	interaural time difference
ITF	interaural transfer function
LMS	least mean square
LP	linear prediction
LPC	linear prediction coding
MOS	mean opinion score
MUSHRA	multiple stimuli with hidden reference and anchor
NCA	National Center for Audiology
PESQ	perceptual evaluation of speech quality
RIR	room impulse response
RT60	60 dB reverberation decay time
SIMO	single input multiple output
SNR	signal to noise ratio
SRMR	speech to reverberation modulation energy ratio
SRR	speech to reverberation noise ratio
SS	spectral subtraction
SS+W	spectral subtraction with Wiener filtering
WSS	weighted spectral slope
YM	Yegnanarayana-Murthy

Chapter 1

Introduction

Speech is the most important and prominent form of communication in our society. Be it in person, over the phone, or via the internet, speech is involved in nearly any occupation or task. However, in a typical communication setting, speech is corrupted by background noise, interference, and/or reverberation which impact its perception. The presence of hearing loss compounds this issue, as it has been shown that noise and reverberation synergistically degrade the speech understanding capabilities of a hearing impaired listener [1]. There has been a growing interest in the development of digital signal processing (DSP) strategies to mitigate the effects of noise and reverberation in communication devices and assistive hearing instruments. The focus of this thesis is on reverberation, DSP algorithms that reduce reverberation ("dereverberation"), and the evaluation of their effect on speech perception and localization.

1.1 Effects of reverberation on speech perception

Under good conditions, speech is quite easily understood by a person with normal hearing. However, speech is almost always subjected to one or more forms of corruption. One of the most common causes of corruption in speech is reverberation. Reverberation is the reflection of sound waves off walls and other surfaces. The travel path of the reflected sound is longer than that of the direct sound, creating delayed versions of the original sound source at the listener.

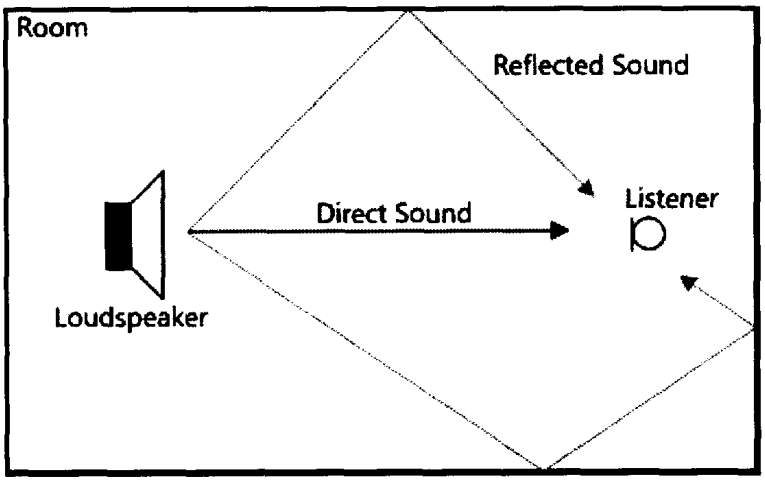


Figure 1-1: Model of reverberation in an enclosed space.

Reverberation can be modelled as

$$x(n) = h(n) * s(n) \tag{1.1}$$

where $s(n)$ is the uncorrupted source, $h(n)$ is the room impulse response, $x(n)$ is the input at the listener, and * is the convolution operator. From the perspective of the listener, it appears as if the signal has been smeared in the time-domain. This temporal smearing can make it difficult to understand words spoken in succession as the reverberation from one word may overlap and interfere with the next one. As well, reverberation tends to cause an increase in low frequency energy relative to higher frequency energy [1]. This effect is due to most materials having a lower coefficient of absorption at low frequencies than high frequencies. In other words, more high frequency energy is lost relative to low frequency energy with each reflection. When these reflections are summed at the listener, an overall increase of low frequency energy is observed. Due to the design of the auditory system, low frequencies are effective at masking higher frequencies, which can cause confusion or uncertainty for a listener trying to identify words, vowels, or other parts of speech [2].

Many studies have demonstrated the negative effects of reverberation on speech perception. Nabelek and Pickett demonstrate that small increases in reverberation can have profound effects on speech intelligibility, especially when coupled with background noise [1]. In a later study, Nabelek demonstrates that reverberation has a significant impact on vowel identification in listeners with sensorineural hearing loss, and that

performance in reverberation decreases with increased loss [3]. As well, age has been shown to have a negative effect on speech intelligibility in reverberation, even if the listener has otherwise normal hearing [4] [5]. Another study by Rogers *et al.* [6] suggests that elderly bilingual listeners may be more susceptible to the effects of reverberation than a monolingual listener with similar hearing loss. These studies demonstrate that reverberation has a significant impact on speech perception.

1.2 Effects of reverberation on localization

Reverberation is also known to have a negative impact on localization [7], the ability to determine the location of a sound source. Humans localize sound using a variety of cues which can be categorized as binaural and monaural cues [2]. Monaural cues, also known as spectral cues, are based on the filtering characteristics of the outer ear. These cues have been shown to aid in azimuth estimation, but contribute mostly to elevation estimation [2]. Binaural cues refer to the interaural level/intensity difference (ILD/IID) and the interaural time/phase difference (ITD/IPD). ILDs exist due the head shadow effect (HSE) in which an acoustic shadow is created on the side of the head opposite the sound source. This shadow is created by the diffraction of sound waves around the head. Higher frequencies, having smaller wavelengths, are more obstructed by the head than lower frequencies. As well, energy absorbed or reflected off the head and shoulders contributes to the difference in the sound pressure level at each ear. Hair and clothing absorb more energy at higher frequencies [8], contributing to greater ILDs at higher frequencies. Numerous studies have shown that the ILD as a localization cue is significant at frequencies greater than 1500 Hz, whereas the ITD is the dominant cue for frequencies below 1500 Hz [9] [10]. This idea has been termed the *duplex theory of localization.* More recent research has found that this frequency is not a hard cut-off and that the ILD and ITD can influence localization of sounds beyond this threshold [11] [12]. However, as a simple model of binaural localization, duplex theory is rather robust.

In a reverberant environment, reflections approach the listener from directions other than direct path from source to listener. This not only flattens the ILD, but also introduces

multiple ITDs for the auditory system to process. However, localizing in reverberant environments can still be accomplished with accuracy, using what is known as the *precedence effect* [13]. The auditory system applies much greater weighting to the first wavefront that passes by the ears for localization. This technique is effective since the first wave will almost always pass by both ears before any reflections are returned. However, the effectiveness of the precedent effect has been found to be dependent on whether early reflections agree with the direct path location [14]. If only localization along the horizontal plane is considered and if the first reflections come from the floor and the ceiling, the binaural cues of these reflections will strengthen the perceived direction towards the actual location. However, early reflections from a side wall will interfere with the direct path ITD and ILD, causing a decrease in localization performance.

1.3 Dereverberation algorithms

Dereverberation attempts to remove or lessen the negative effects that reverberation has on intelligibility and sound quality. Dereverberation is achieved in a multitude of ways, including channel inversion, maximization of second-order statistics through adaptive filtering, and spectral subtraction. Typically, dereverberation implementations are designed to operate on a single input channel. Since the signals from a sound source at the left and right ears are not the same due to the HSE, independently operating dereverberation processes may change the scale or time of arrival of the signals relative to each other. This change is a modification of the natural ITD and ILD, distorting a listener's perception of the actual sound location. As such, binaural processing, which strives to preserve these cues, is preferred.

1.4 Measures of speech perception

In order to validate the performance of any speech processing algorithm, it is necessary to use some measure of how it influences speech perception. Measures of speech perception can be categorized in several ways. This review will examine two ways of categorization; subjective or objective speech evaluation, and intelligibility or quality metrics.

1.4.1 Intelligibility versus quality measures

When evaluating speech perception, it is important to differentiate between intelligibility and quality and determine which is most appropriate for the system being tested. Intelligibility is defined as the degree to which speech can be understood whereas quality usually refers the perceived naturalness or pleasantness of a speech sample.

More specifically, intelligibility commonly refers to the percentage correct of speech units observed by a listener, where speech units may refer to sentences, words, phonemes, or other parts of speech [15]. Intelligibility tests are typically only appropriate when the system being tested produces considerable degradation. In systems with mild degradation, ceiling effects where subjects score near 100% are generally observed. This limitation can be circumvented in certain instances by introducing a known degradation to the system, such as additive noise, that makes the task more difficult.

Quality tests, which attempt to measure the perceived quality of a speech sample, do not suffer from ceiling effects as intelligibility tests do. As such, quality tests are useful for differentiating highly intelligible systems. In a quality test, subjects are asked to focus on a particular aspect of speech such as pleasantness or naturalness, or overall quality.

1.4.2 Subjective versus objective measures

Speech perception measures can be differentiated as being either subjective or objective. Subjective measures rely on feedback from a test subject, and are hence subject to preferences or ability at a task. Subjective testing can be costly and time consuming, however, subjective evaluation is important as it measures directly the effect of a system on people's perception of speech.

Objective metrics attempt to predict or estimate results from subjective testing by analyzing properties of a speech sample. Objective metrics carry several benefits associated with computer processing. For one, objective metrics are able to provide immediate feedback on a system which is useful for testing and fine tuning. Second, objective metrics can be calculated over a much larger set of speech samples than normally feasible for a subjective study. In this way, objective metrics are useful for

identifying dependencies on the talker, angle of presentation, and other parameters that can only be addressed in a large study. However, objective metrics tend to only be valid for the particular type of distortion that was used in their associated subjective test.

Objective metrics themselves can be classified as either intrusive or non-intrusive. An intrusive measure requires the original and enhanced signals to determine a score, usually by comparing the time-frequency differences between the input and output. In contrast, a non-intrusive measure determines a score based on the output only and is useful in situations where the input signal may not be readily available. When using objective metrics to measure the performance of speech enhancement algorithms, there are generally three signals of interest: the original (clean) speech, the corrupted speech, and the enhanced speech. The amount of enhancement can be characterized as the difference between the scores for the corrupted speech and the enhanced speech. For intrusive measures, the clean speech is used as the reference input in both cases.

1.5 Summary and statement of problem

Reverberation impacts speech quality and intelligibility as well as our ability to localize sounds. While dereverberation techniques have evolved over the past few years, they have been primarily designed for monaural applications. In a binaural application, two independently operating dereverberation algorithms may distort the cues necessary for sound localization, and may upset the naturalness of the processed speech, highlighting the need for a binaural dereverberation approach. Furthermore, typical hearing aids have been independent devices, individually programmed for the left and right ears with no communication between the two. With advances in wireless technologies, it is becoming more common for hearing aids to employ a wireless link between each aid. This link allows hearing aids to share information, leading to improved speech processing techniques. As such, binaural signal processing techniques are quickly becoming of great interest to hearing aid manufacturers as well. Binaural implementations of dereverberation techniques are relatively new and unexplored. There is little data on whether binaural implementations can provide significant improvements over their monaural counterparts, in terms of speech quality and intelligibility enhancement, as well

as localization ability. Intuitively, one would expect binaural implementations to offer an improvement in localization, but there is no formal evidence of this effect as of yet.

1.6 Research objectives

The primary objective of this research is to develop a binaural dereverberation method that meets the following criteria:

- Provides an improvement in speech quality over reverberant speech,
- Preserves or improves the cues that are used for localizing speech, and
- Has reasonable computational complexity to run in real-time on a portable device.

Secondary objectives of this research include:

- determining the relationship, if any, between binaural and bilateral dereverberation on sound source localization,
- examining the quality of objective metrics at predicting user preferences of speech in reverberation, and
- developing a binaural impulse response database for the purpose of generating reverberant speech.

1.7 Organization

The thesis is organized as follows. Chapter 2 begins by reviewing the effects of binaural versus monaural listening. It goes on to explore the variety of single and multi-channel dereverberation techniques that exist in literature, followed by a review of several objective speech evaluation metrics. Chapter 3 describes the development of a binaural room impulse response (BRIR) database created for the purpose of generating reverberant speech. The measurement procedure and characteristics of the measured impulse responses are all detailed. Chapter 3 concludes with an overview of another BRIR used for generating reverberant speech and evaluating dereverberation techniques [16]. Chapter 4 proposes a new binaural dereverberation technique using the linear predictive (LP) residual of the speech signal. An overview of the method described in

[17], from which the binaural implementation is derived, is presented as well as the motivations and changes involved in the new method. Chapter 5 evaluates the dereverberation techniques as well as another recent binaural dereverberation technique by Jeub *et al.* [18] in terms of sound quality and localization. The methodology of two subjective listening tasks by normal hearing listeners are explained and the results of the tasks are analyzed and compared against objective measures to determine correlations. Chapter 6 summarizes the conclusions and findings of this thesis and provides recommendations for future work.

Chapter 2

Literature Review

This section provides a review on research done on topics pertaining to this thesis. First, common methods of objective and subjective speech evaluation techniques and their appropriateness towards evaluating dereverberation algorithms are discussed. Second, the different types of dereverberation techniques found in literature are examined.

2.1 Electroacoustic speech evaluation

2.1.1 Sound quality

Effective evaluation of the intelligibility or quality of speech is of importance in many fields of speech research. The most common method of subjective evaluation of speech quality is the mean opinion score (MOS). The MOS is a five point scale of speech quality, described in Table 2-1.

Rating	Speech Quality	Level of Distortion
5	Excellent	Imperceptible
4	Good	Just perceptible, but not annoying
3	Fair	Perceptible and slightly annoying
2	Poor	Annoying, but not objectionable
1	Unsatisfactory	Very annoying and objectionable

Table 2-1: Mean Opinion Score description.

Electroacoustic measures are most commonly developed for evaluating speech coding systems [19] [20] [21] such as code excited linear prediction (CELP), however their effectiveness at evaluating speech enhancement systems varies. A comprehensive study on the effectiveness of typical speech quality metrics for evaluating speech enhancement algorithms was done by [22]. The score of 13 different speech enhancement algorithms in four noise conditions are compared against subjective data. It is demonstrated that certain measures, such as the segmental SNR [23], Itakura-Saito distance (IS) [24], and weighted spectral slope (WSS) [25], are poor to moderate predictors of overall quality with Pearson's correlation coefficients of 0.36, 0.60, and 0.64 respectively. However, other measures, like the cepstrum distance [19], log-likelihood ratio [23], and perceptual evaluation of speech quality (PESQ) [21], performed well with Pearson's correlation coefficients of 0.79, 0.85, and 0.89 respectively. While this study shows good performance for speech in noise, it does not address how preferences may change in reverberation. PESQ, the highest correlating measure of quality for speech in noise, suggests that it should not be used for evaluation when talker echo is present [21].

A study by [26] evaluates objective metrics in reverberant environments with RT_{60}'s between 291 – 447 ms. Their results indicate that the correlation of the cepstral distance to overall speech quality drops considerably ($\rho = 0.17$) from the speech in noise conditions in [22]. This study demonstrates that metrics that perform reasonably well in speech in noise cannot be assumed to have similar performance in reverberation. Falk and Chan [27] introduced a new, non-intrusive metric, the speech to reverberation modulation energy ratio (SRMR). The SRMR measures the ratio of low frequency envelope modulations over higher frequency envelope modulations in a speech signal. First, the input signal is filtered by a 23-channel gammatone filterbank. The Hilbert transform is used to compute the envelope within each band in 256 ms frames with 87.5% overlap. Each envelope frame is multiplied by a 256 ms Hamming window. The modulation spectral energy for each band j is computed as the squared magnitude of the discrete Fourier transform of the temporal envelope frames. $\bar{\varepsilon}_{j,k}$ denotes the average modulation energy over all frames of the jth band, grouped by the kth modulation filter, where

$j = 1, \ldots, 23$ and $k = 1, \ldots, 8$. The average per-modulation band energy $\bar{\varepsilon}_k$ is then defined as

$$\bar{\varepsilon}_k = \frac{1}{23}\sum_{j=1}^{23} \bar{\varepsilon}_{j,k} \tag{2.1}$$

and the SRMR is defined as

$$SRMR = \frac{\sum_{k=1}^{4} \bar{\varepsilon}_k}{\sum_{k=5}^{K^*} \bar{\varepsilon}_k}, \tag{2.2}$$

where K^* is the modulation band that accounts for 90% of the total modulation energy.

Falk and Chan [27] demonstrate that envelope modulations greater than 20 Hz are associated with reverberation. Thus, the ratio of envelope modulations to modulations due to reverberation should provide a measure of the severity of reverberation. The SRMR achieves a correlation of 0.80 for overall quality with normal hearing subjects, outperforming other state-of-the-art algorithms [27].

2.1.2 Localization

Unlike sound quality, there are few electroacoustic predictors of localization ability beyond simple measures such as the ILD error. However, Faller and Merimaa [28] present an improved method of selectively picking the ILD and ITD in time frames above a certain interaural coherence. This method models the precedence effect in that localization in reverberant conditions relies heavily on the first wavefront. For determining the ILD (our primary concern), the input signals are divided into 20 ms (320 samples at a sampling rate of 16 kHz) frames with 99.9% overlap (319 samples). Each frame is filtered using a 24 band Gammatone cochlear filterbank [29] [30] with center frequencies according to the Glasberg and Moore model [31]. The interaural coherence is calculated for each sub-band frame and those above a threshold are used for the ILD determination. Jeub *et al.* [18] extend this method by introducing a variable threshold, in which only the 10% most reliable values in any given sub-band are used for ILD determination. They then examine the mean ILD difference between different dereverberation algorithms and reverberant speech and the variance of this error and

conclude that large variations in the ILD difference affects source localization. However, this relationship has not been thoroughly explored.

2.2 Dereverberation techniques

In the last decade, dereverberation has been an area of great interest for researchers. A wide variety of techniques have been used to reduce the effects of reverberation on speech. This section will review many of these solutions, organized by the general approach taken.

2.2.1 Direct RIR inversion

A popular method for dereverberation attempts to find the inverse of the room impulse response. Using the reverberation model in (1.1), the original source signal can be recovered by convolving $x(n)$ with the inverse of the room impulse response.

$$s(n) = h^{-1}(n) * x(n) \tag{2.3}$$

The problem is a matter of efficiently and accurately determining $h^{-1}(n)$ without *a priori* knowledge of $s(n)$. This problem is known as blind channel estimation and has a wide variety of applications in communications. Huang and Benesty [32] consider the estimation of a single input multiple output (SIMO) FIR system. By employing multiple system outputs, or for the purposes of dereverberation, a multi-microphone system, a set of error functions can be defined by taking the difference between each combination of inputs convolved with an estimated impulse response. These error functions are combined in a cost function which can be solved using a least mean squares (LMS) or Newton algorithm. However, this method is not demonstrated for practical dereverberation. Room impulse responses are considerably larger than those tested in this study (500 samples to convergence for a length 15 two-channel FIR system, versus a RIR which could range from 800 to 8000 samples at a sampling frequency of 8 kHz). As well, the source signal used was an uncorrelated binary phase shift keying (BPSK) sequence which has a white power spectrum. If speech is assumed to be the source signal, the convergence time will increase due to its coloured power spectrum.

Attempts have been made to address these issues. Studies [33] and [34] suggest that the complexity of the LMS algorithm can be reduced by updating a fraction of the total filter coefficients. Kokkinakis and Loizou [34] report improvement to reverberant speech with computational times feasible in portable systems. Zhang *et al.* [35] take an alternate approach, instead adapting the length of the adaptive filter to minimize the mean-square deviation at each iteration. Tong Zhou [36] provides a variable step size implementation of the variable tap-length filter, offering moderate improvements to convergence time and steady state error. However, for all presented solutions, the convergence time with speech as an input remains unknown. With all these systems, care must be taken when determining the inverse RIR. Real acoustic spaces tend to have mixed-phase impulse responses, and as such do not have stable inverses [37] [38]. However, Miyoshi and Kaneda [37] have demonstrated that it is possible to determine the exact inverse of a mixed-phase system using multiple inputs/outputs.

2.2.2 LP residual and statistical methods

The linear predictive residual of speech has many properties that are beneficial towards effective dereverberation. Yegnanarayana and Murthy [17] looked at the LP residual and its properties. Among their findings was that the LP residual has strong peaks that correspond with the glottal cycle and that in reverberant speech, the regions between glottal peaks have added noise. Building on this discovery, they developed a single channel method that weights the LP residual to enhance the peaks of the glottal cycles while suppressing the smaller peaks caused by reverberation. In [39], enhancement of the glottal cycle is achieved by coherently adding the LP residual from a microphone array. Since the peaks due to reverberation occur randomly across the array, the overall signal to noise ratio is increased.

Gillespie *et al.* [40] also exploit the properties of the LP residual through a multi-channel technique that maximizes LP residual kurtosis using a gradient-descent adaptive filter. A similar method was implemented in [41] using the skewness of the LP residual instead. Wu and Wang [42] found that while such methods were capable of reducing the colouration distortion caused by early reverberation, they do little to reduce the effects of long-term reverberation. To tackle this issue, a second stage was employed by [43],

which uses pitch periodicity as a measure of reverberation time and enhances the speech through spectral subtraction. Nakatoni and Miyashi [44] describe a similar technique where the harmonic structure is estimated and weighted to enhance dominant harmonics while reducing noisy components. However, this technique requires an extensive training period, making it unsuitable in situations where the RIR is not constant.

2.2.3 Coherence based post-filtering

A common method of speech enhancement in noise uses the multi-channel Wiener filter to estimate the desired signal. Zelinski [45] applied the same theory to dereverberation and presents an adaptive post-filtering scheme based on the Wiener filter fed by a 4-microphone delay and sum beamformer. However, in this scheme, the determination of the filter coefficients assumes that the noise signal is uncorrelated with the source, which is false for realistic conditions. Recognizing this inaccuracy, McCowan and Bourlard [46] replace the assumption of incoherent noise with a known spatial coherence function. The spatial coherence function is a normalized measure of the correlation between two points in space. Specifically, McCowan and Bourlard [46] use the coherence function of a diffuse noise field, given in [47] as

$$\Gamma_{ij}(f) = sinc\left(\frac{2\pi f d_{ij}}{c}\right), \tag{2.4}$$

where d_{ij} is the distance separating two positions i and j, f is frequency in Hz, and c is the speed of sound. While more accurate than the assumption of complete incoherency, this coherence function is not ideal for binaural use as it ignores the head shadow effect. Thus, Jeub and Vary [48] developed a new coherence function that models the complex geometry of the head as two spatially separated circular plates at each side of the head. Assuming an ear separation of 17 cm, an approximation of the noise field coherence is given by

$$\hat{\Gamma}_{ij}^{head}(f) = \sum_{p=1}^{P} a_p e^{-\left(\frac{f-b_p}{c_p}\right)^2}, \tag{2.5}$$

where $P = 3$ and the model parameters are as given in Table 2-2.

p	α_p	b_p	c_p
1	1	18.97	291.1
2	$14.5 * 10^{-3}$	875.2	105.7
3	$2.38 * 10^{-3}$	1371	151.5

Table 2-2: Coefficients of the binaural coherence model for a microphone separation of 17 cm.

Post-filtering using this coherence function shows an improvement in the SRMR scores over the diffuse and incoherent noise fields [48].

2.2.4 Binaural speech enhancement

Binaural processing is a recent area of interest among researchers and remains relatively unexplored. The goal of binaural processing is to preserve the ITD and ILD cues that account for much of our localization ability. Klasen *et al.* [49] present a binaural noise-reduction algorithm that preserves the inter-aural transfer function (ITF), and thereby the ITD and ILD, by adding terms representing the error from the desired ITF into the cost function of the Wiener filter. While not a dereverberation algorithm, the modifications to the cost function can be applied to any technique incorporating the Wiener filter, such as those described in section 2.2.3. The disadvantage of such a method is the inherent trade-off between ITF preservation and noise reduction. Demanding greater ITF preservation comes at the cost of SNR improvement. In [50], it is demonstrated that when a subject is presented noise and speech from two different angles, the bilateral multichannel Wiener filter "moves" the noise source to that of the speech. Intelligibility of speech increases when the noise and speech are spatially separated [51], suggesting that the reduction in SNR improvement from preserving the ITF will be offset by some degree by the spatial release from masking. However, this improvement is shown to not be great enough to overcome the negative effects of the ITF preservation [50], and is only relevant when the noise is initially spatially separated from the speech. Since reverberation can be considered to be correlated noise, it is reasonable to expect similar effects with Wiener based dereverberation techniques.

Jeub *et al.* [18] present a two-stage binaural dereverberation method that will be used as a comparison against the binaural technique developed in this thesis. The first stage

consists of a spectral subtraction rule designed to remove late reverberation. Using a statistical model of late reverberation, the variance of the late reverberant speech is estimated. Time-frequency frames where the estimated reverberant variance is high are suppressed. The second stage uses the circular plate coherence model described in 2.2.3 for dereverberation of early reflections. Preservation of the binaural cues is achieved in the first stage by employing a delay-and-sum beamformer. The output of the beamformer is used to determine the spectral subtraction weights which are applied to the left and right channels. Similarly, the second stage uses the dual-channel Wiener filter and produces one set of gains which are applied to both left and right channels. Therefore, the ILD is unaffected. The phase of the input signal is kept, ensuring no changes in the ITD. The Two-Stage method is shown to have significant improvement in SRMR over the other coherence based methods. Furthermore, subjective evaluation with 17 experienced listeners revealed that the Two-Stage binaural method is preferred over the bilateral version, both in terms of preservation of ITD/ILD cues and speech quality.

2.3 Summary

Single-channel dereverberation has a rich history, employing a wide variety of techniques. The effects these techniques have on localization remains relatively unknown. As well, reverberation has only recently become the focus for developing objective sound quality and intelligibility metrics. The SRMR measure shows promise, but has yet to reach widespread use.

Chapter 3

Development of Binaural Impulse Response Database

In order to evaluate the effects of dereverberation algorithms on speech quality and localization, it is necessary to first generate reverberant speech. This task can be accomplished in several ways. The most straightforward approach is to simply make a recording of speech in a reverberant environment. This approach, however, requires recordings to be made for each stimulus, and becomes unfeasible when large amounts of stimuli are required. As well, one must have the proper presentation and recording equipment to make such measurements and an appropriate room to make them in. To avoid these constraints, considerable work has been done on creating computer simulations of reverberant transfer functions. There are two main methods of acoustic simulation, the image source method and the ray tracing method.

The image-source model simulates reverberation in a rectangular space by creating mirrored copies of the room and speaker position around the original room with the listener [38]. The contribution of each copied sound source is adjusted for the distance to the original listener and the absorption at each reflecting wall. The image source method suffers from being restricted to rooms with simple geometry. As well, the reflections it produces tend to be too perfect from what would be seen in a real room, especially at longer reverberation times [52]. The ray tracing method models sound as particles scattering from a source location. The particles travel in straight lines, reflecting off of

surfaces a set number of times. To determine the impulse response at a point, a certain volume or area must be taken to catch rays that pass through. Due to finite particles and reflections, ray tracing has the chance of missing reflections. Hybrid methods exist which can produce realistic simulations [53], but these solutions are expensive.

In order to achieve the flexibility of simulation methods with the realism of real room recordings, a binaural impulse response database was created. Using recordings made in reverberant environments, it is possible to extract the speaker to ear impulse response which can then be convolved with any stimulus. The recordings were made through a Head and Torso Simulator (HATS) within the Beltone Anechoic Chamber (BAC) and the reverberation chamber at the National Centre for Audiology (NCA). A brief description of the room properties along with the experimental procedure to record the impulse responses is given below.

3.1 Room specifications

3.1.1 NCA reverberation chamber

The dimensions of the reverberation chamber are given in Table 3-1. It conforms to specifications in ISO 3741:1999 for measuring sound power levels. Two removable heavy fabric curtains and a piece of acoustic foam were used to adjust the properties of the chamber to achieve varying degrees of reverberation. The different configurations are depicted in Figure 3-1 and Table 3-2 lists their respective reverberant properties. These configurations will be referred to as R1, R2, R3, and R4 for the setups depicted in Figure 3-1 a, b, c, and d respectively.

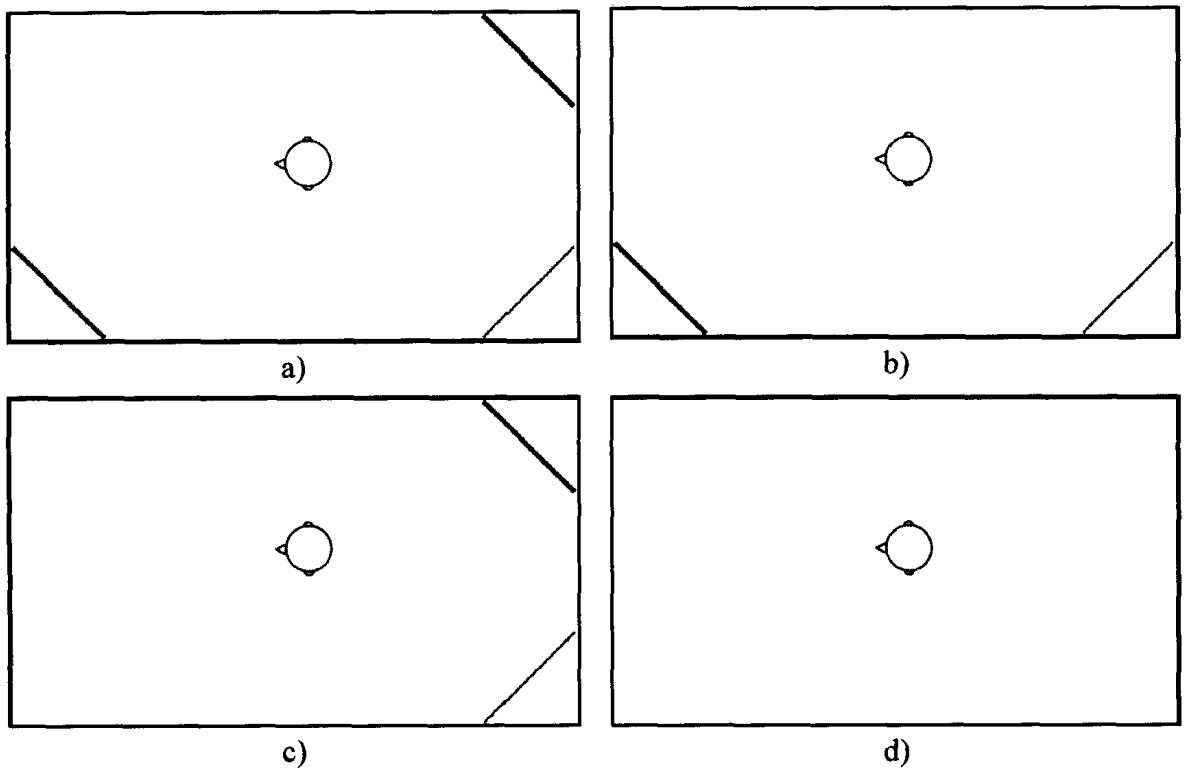


Figure 3-1: Four configurations of the reverb chamber with respect to manikin (HATS) positioning. Thick line indicates curtain, thin line indicates acoustic foam.

3.1.2 Beltone Anechoic Chamber

The BAC is a semi-anechoic chamber. The ceiling and walls are constructed with a 125 Hz cut-off wedge system. Unused floor space is filled with moveable sound absorbing floor panels and acoustic foam. The dimensions of the BAC are given in Table 3-1. See Appendix A for pictures of the HATS setup in each chamber.

Room	Length	Width	Height	Volume (cu ft)
Beltone Anechoic Chamber	12'	23'	18'	4,968
Reverberation Chamber	20.3'	13.3'	8.7'	2,349

Table 3-1: Anechoic and reverberant room dimensions.

3.2 Procedure

3.2.1 Equipment and measurement setup

The measurements were carried out using quality audio equipment available at the National Centre for Audiology. All playback and recording was done at a sampling rate of 44.1 kHz via an Echo AudioFire 12 sound card. The output channels of the AudioFire 12 were connected to a Soundweb 9088iis DSP unit which performs speaker equalization, channel switching, and level adjustments. Samples were played out through an array of 16 Tannoy Di5 DC loudspeakers, which were situated in a circle with radius 1.4 m spaced equally apart by 22.5 degrees. Lab.Gruppen C Series and QSC CX 168 power amplifiers were used as the interface between the Soundweb system and the loudspeakers in the reverberation and anechoic chambers respectively.

The recordings were made through a Brüel & Kjær Type 4128C HATS and connected to a G.R.A.S Type 12AA Power Module. The height of the HATS was adjusted such that the ear canal was level with the vertical centre of the surrounding loud speakers.

3.2.2 Impulse response measurement

The binaural impulse responses were constructed using the swept sine method as described in [54]. In this method, a signal $s(n)$ is constructed as a sine wave that exponentially increases in frequency over time T.

$$s(n) = sin\left(K\left(e^{-n/Lf_s} - 1\right)\right), \tag{3.1}$$

$$where\ K = \frac{\omega_1 T}{ln\frac{\omega_2}{\omega_1}}\ and\ L = \frac{T}{ln\frac{\omega_2}{\omega_1}}$$

$$and\ \omega_1 = 200\pi,\ \omega_2 = 32000\pi,\ T = 3$$

$s(n)$ is repeated to form the stimulus that is presented from the loudspeaker. Let $x(n)$ be the recorded signal from one ear of the HATS. The first half of $x(n)$ is discarded such that only the recording of the second sine sweep is kept, and the Fourier Transform of

$s(n)$ and the truncated $x(n)$ is taken. Dividing $S(\omega)$ by $X_T(\omega)$ element-wise and taking the real portion of the inverse Fourier Transform results in the impulse response from the loudspeaker to the ear.

$$x_T = x\left(\frac{M}{2} + 1 : M\right), \text{ where M is the length of x}$$

$$S(\omega) = \mathcal{F}[s(n)], \quad X_T(\omega) = \mathcal{F}[x_T(n)]$$

$$h(n) = real\left(\mathcal{F}^{-1}\left[\frac{X_T(\omega)}{S(\omega)}\right]\right) \tag{3.2}$$

Recordings were made at each of the 16 loudspeakers around the HATS. This procedure was repeated for the left and right channels of each recording to produce binaural impulse responses from a stimulus at each angle. For all azimuth angles referred to in this thesis, 0 degrees refers to the loudspeaker positioned directly in front of the HATS, with azimuth increasing in the clockwise direction.

3.3 Aachen impulse response (AIR) binaural database

For a more accurate comparison of the dereverberation algorithms, reverberant speech was also generated using the binaural database defined in [16]. This database was also designed specifically for the evaluation of dereverberation techniques. The AIR database has impulse responses measured in real environments (a meeting room, office, lecture hall, and stairwell). Using the AIR Database in conjunction with the RIR's of the NCA chambers allows for a greater range of reverberation times to be tested. As well, it can be seen how evaluation metrics differ for similar reverberation times between a test chamber and a real environment.

3.4 RT_{60} estimation

RT_{60} is commonly used to categorize the degree of reverberation in an environment. It is defined as the amount of time required for a source that is switched off to decay by 60 dB. A modified version of the Schroeder method [55] was used to estimate the RT_{60}.

Let $h(n)$ be a channel of a binaural impulse response where $n = 0, 1, 2 \ldots M - 1$. To determine the Schroeder integral $s(n)$, $h(n)$ is squared and normalized such that the sum of all elements of $h^2(n)$ equals 1. This normalization is not necessary, but aids in visualization of the Schroeder integral. Next, the power decay function is found by taking the cumulative sum of the reversed normalized squared impulse response. Finally, $s(n)$ is converted into dB.

$$h_{norm}^2(n) = \frac{h^2(n)}{\sum_{i=0}^{M-1} h^2(i)} \tag{3.3}$$

$$s(n) = \sum_{i=n}^{M-1} h_{norm}^2(i) \tag{3.4}$$

$$s_{db}(n) = 10\, log_{10}\, s(n) \tag{3.5}$$

Because impulse responses rarely have the dynamic range to directly measure a 60 dB decrease in power, the RT_{60} is more commonly calculated by taking a linear fit of a 10, 20, or 30 dB drop and extrapolating. For the calculations in this thesis, a 10 dB drop between -5 dB and -15 dB was used to find the linear fit. Figure 3-2 depicts the Schroeder integral for a sample RIR. The dashed red lines show the region between -5 dB and -15 dB at which the linear fit is calculated. The green line shows the extension of the linear fit to -60 dB where the RT_{60} is read.

To verify the RT_{60} measurements of the NCA chambers, the RT_{60}'s of the impulse responses provided by the AIR database were measured using the above procedure and compared to their reported values. Figure 3-3 shows the results across all rooms in the AIR database except for the stairway, as no reverberation times for those RIRs were provided.

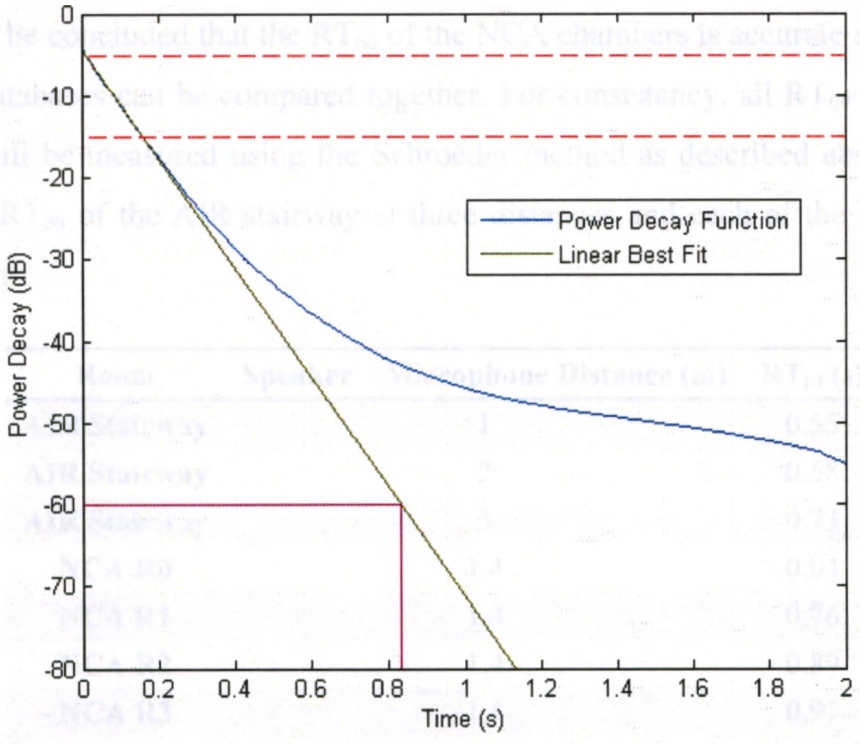


Figure 3-2: Example of calculating the RT_{60} of a RIR using the Schroeder method.

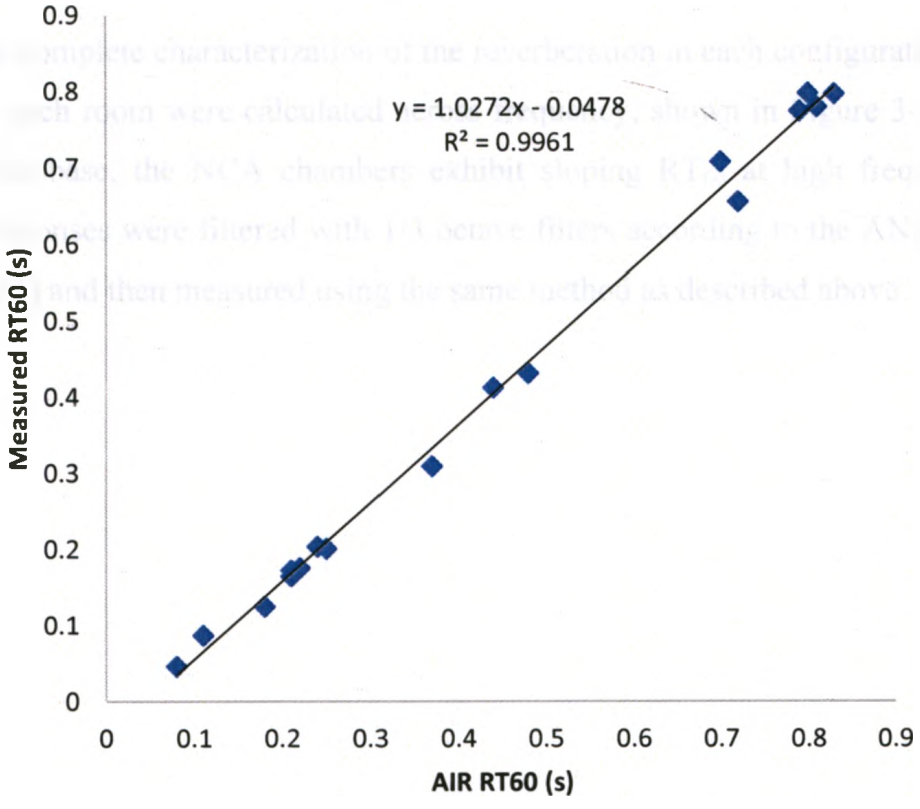


Figure 3-3: Comparison of RT_{60} reported by [16] and measured using the Schroeder method.

As seen, the correlation between the measured and reported RT_{60} is quite high ($R^2 = 0.99$). It can be concluded that the RT_{60} of the NCA chambers is accurate and that results from both databases can be compared together. For consistency, all RT_{60} values used in this thesis will be measured using the Schroeder method as described above. Table 3-2 outlines the RT_{60} of the AIR stairway at three distances and each of the NCA chamber configurations.

Room	Speaker – Microphone Distance (m)	RT_{60} (s)
AIR Stairway	1	0.55
AIR Stairway	2	0.68
AIR Stairway	3	0.73
NCA R0	1.4	0.04
NCA R1	1.4	0.76
NCA R2	1.4	0.89
NCA R3	1.4	0.97
NCA R4	1.4	1.39

Table 3-2: Overview of measured reverberation times for each room. All values are averaged between the left and right channels and for all available angles.

For a more complete characterization of the reverberation in each configuration, the RT_{60} values for each room were calculated across frequency, shown in Figure 3-4. Similar to the AIR database, the NCA chambers exhibit sloping RT_{60} at high frequencies. The impulse responses were filtered with 1/3 octave filters according to the ANSI S1.1-1986 standard [56] and then measured using the same method as described above.

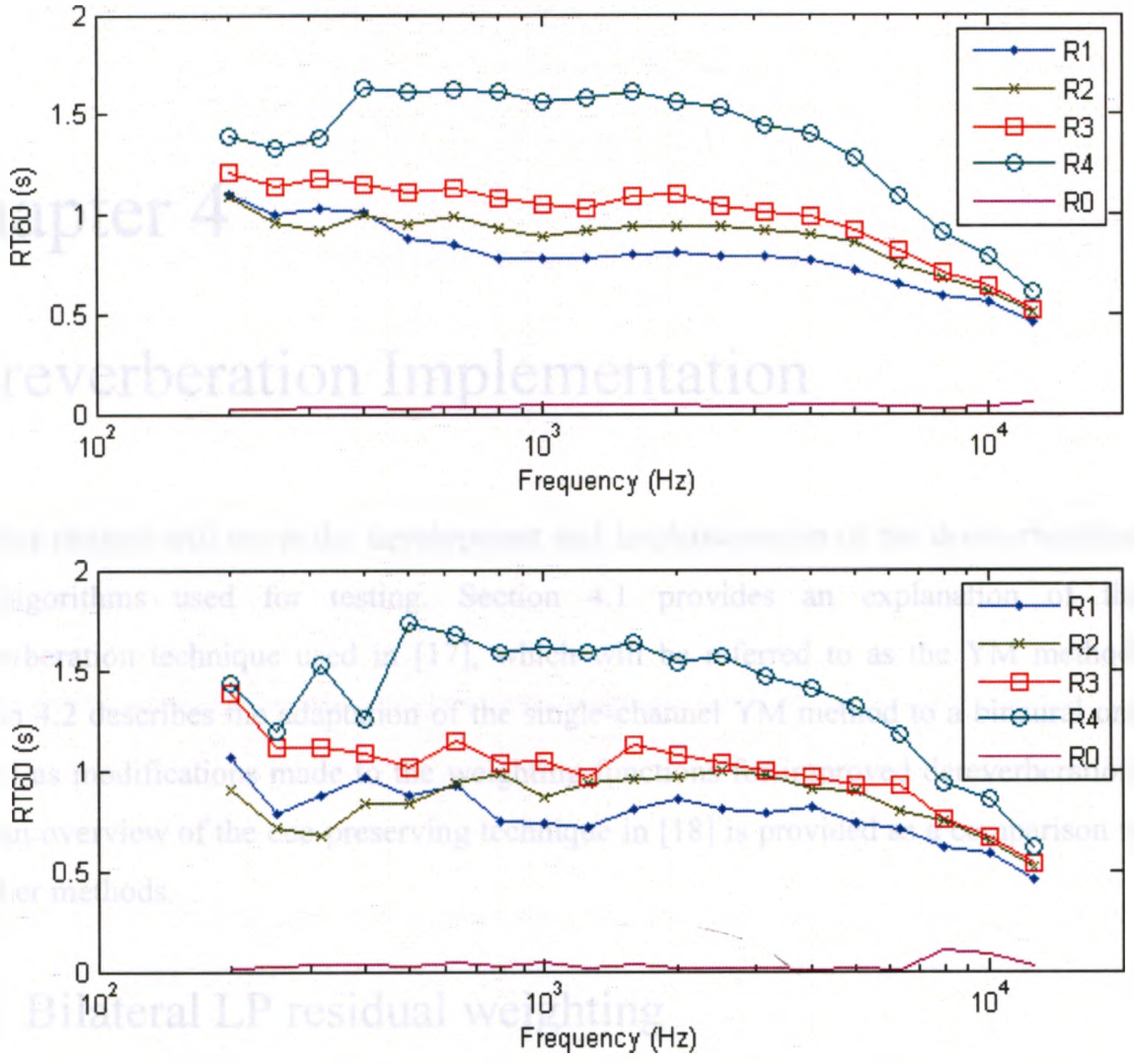


Figure 3-4: Reverberation time across frequency averaged over left and right channels and a) all angles and b) zero degrees azimuth

3.5 Summary

A binaural room impulse response database has been developed for generating stimuli in realistic conditions. A total of five room conditions are made available, one in an anechoic setting and four at different levels of reverberation. The presented database is fully characterized by RT_{60} and extends the available range of testing conditions offered by other BRIR databases.

Chapter 4

Dereverberation Implementation

This chapter will cover the development and implementation of the dereverberation algorithms used for testing. Section 4.1 provides an explanation of the dereverberation technique used in [17], which will be referred to as the YM method. Section 4.2 describes the adaptation of the single-channel YM method to a binaural one as well as modifications made to the weighting functions for improved dereverberation. Last, an overview of the cue-preserving technique in [18] is provided as a comparison to the other methods.

4.1 Bilateral LP residual weighting

The basis of the technique presented in [17] is that the direct component of speech should be enhanced while suppressing regions where the reverberant component dominates. Weighting the data in this way gives the effect of enhancement since perception of speech is highly influenced by regions with higher energies. Thus, it is necessary to determine which regions of the data signal contain the direct component and which areas are dominated by the reverberant component. Yegnanarayana and Murthy [17] perform an in-depth investigation into the statistics and spectra of clean and reverberant speech segments to determine how these regions can be effectively determined. The results are summarized as follows.

Processing should be performed on the LP residual, rather than on the speech signal directly. The LP residual is more suitable as distortions caused by processing will be

smoothed out due to the all-pole synthesis filter upon reconstruction [17]. Let $x(n)$ be the input signal to be processed. The predicted signal, $\hat{x}(n)$, is given by

$$\hat{x}(n) = \sum_{i=1}^{p} a_i x(n-i), \tag{4.1}$$

where p is the order of the linear predictor and a_i are the predictor coefficients solved using the autocorrelation method. The residual is then given as error between the original and the predicted signal.

$$e(n) = x(n) - \hat{x}(n) \tag{4.2}$$

The LP residual will be modified in short (2 ms) and long (20 ms) segments. Weighting of the long segments (gross weight function) is achieved by estimating the entropy of the LP residual, defined in (4.3). Similarly, weighting of the short segments (fine weight function) is computed from the normalized prediction error, defined in (4.6).

4.1.1 Gross weight function

The linear prediction residual for the gross weighting function is developed using a frame size of 20 ms (160 samples at 8 kHz sampling rate), Hamming window, and a 10th order LP analysis using the autocorrelation method. The residual signal is segmented into 20 ms frames with 10 ms overlap. To identify the regions of high and low signal to reverberation component ratio (SRR), the entropy of the kth frame, H_k, is calculated as,

$$H_k = -\sum_{i=1}^{M} p_i \log p_i \tag{4.3}$$

where p_i is the probability of a sample falling in the ith bin, and M is the total number of bins. To obtain a good estimate of p_i, the number of bins is chosen as $M = 7$. The bins are evenly distributed between the minimum and maximum sample values within each frame. In the event that no samples fall in a bin, $p_i \log p_i$ is taken as zero. It is necessary to obtain a weight for each sample of the LP residual. Thus, the entropy determined for each frame is replicated by 10 ms worth of samples (80 samples at 8 kHz sampling rate) and concatenated to form an extended entropy function with a length equal to the LP

residual. Next, this expanded entropy function is smoothed with a 75 ms moving average window to remove discontinuities from the expansion procedure.

In regions where there is a strong direct component, the probability density function of the samples is skewed, reducing the entropy. Likewise, in regions with reverberant tails or no speech, the probability density function of the samples is closer to a Gaussian distribution, increasing the entropy. Thus, a weighting function that is inversely proportional to the entropy is needed. The non-linear weighting function proposed by [17] is given by,

$$w^{gross} = \left(\frac{w_{max}^{gross} - w_{min}^{gross}}{2}\right) tanh\left(\alpha_g \pi (H_k - a)\right) + \left(\frac{w_{max}^{gross} + w_{min}^{gross}}{2}\right) \tag{4.4}$$

where $w_{max}^{gross} = 1$, $w_{min}^{gross} = 0.1$, $\alpha = -1.5$, and $a = 1.55$. α defines the slope of the weight function and a is the entropy at which $w^{gross} = \frac{w_{max}^{gross} + w_{min}^{gross}}{2}$. The mapping function of the entropy to the gross weight function is depicted in Figure 4-1.

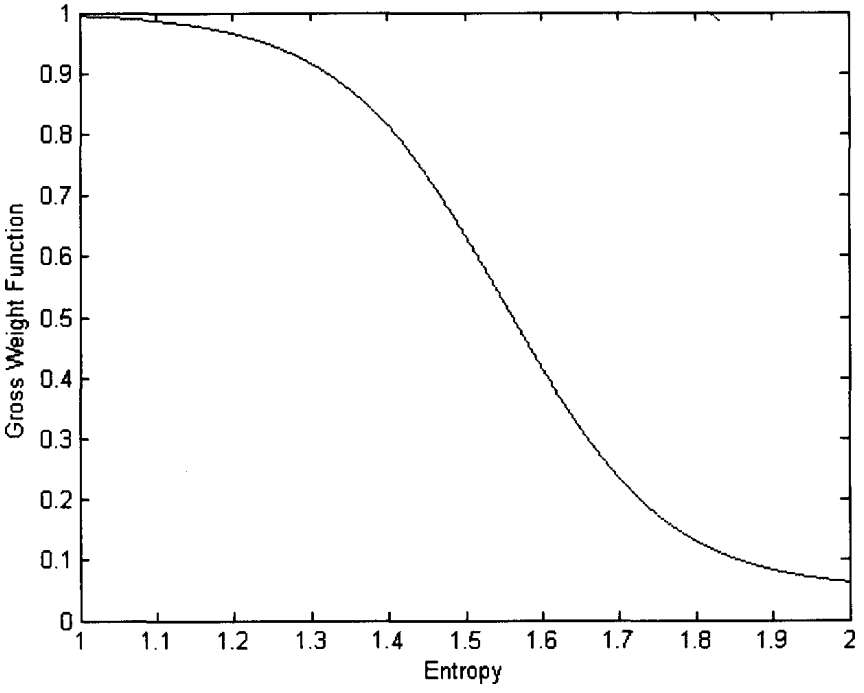


Figure 4-1: Gross weight mapping function of the entropy of the LP residual.

4.1.2 Fine weight function

The fine weight function is intended to improve the definition of the glottal cycle, thereby reducing the perceived reverberation. The fine weight function is formed using a mapping function similar to that of the gross weight function, but instead is calculated using the normalized prediction error. Letting $x(n)$ be the input speech, the 5th order LP coefficients are computed in 2 ms frames. For any frame k, the linear prediction estimate is given by,

$$\widehat{x_k}(n) = \sum_{i=1}^{5} a_{ki} x(n-i) \tag{4.5}$$

And the normalized prediction error, $e_{norm}(n)$, is found at each sample.

$$e_{norm}(n) = \frac{|\hat{x}(n) - x(n)|}{max(|\hat{x}(n) - x(n)|)} \tag{4.6}$$

Next, the trend in the prediction error is removed by smoothing the prediction error with a 10ms Hamming window and subtracting it from the normalized prediction error. The fine weight function, w^{fine}, is given by,

$$x = \left(\frac{w_{max}^{fine} - w_{min}^{fine}}{2}\right) tanh\left(\alpha_g \pi(\eta_k')\right) + \left(\frac{w_{max}^{fine} + w_{min}^{fine}}{2}\right) \tag{4.7}$$

where η' is the de-trended normalized LP error at frame k, $w_{max}^{fine} = 1$, $w_{min}^{fine} = 0.6$, and $\alpha = 1.5$. The mapping function of the normalized LP error to the fine weight function is depicted in Figure 4-2. A sample of the fine weight function for the sentence "The birch canoe slid on the smooth planks" is provided in Figure 4-3.

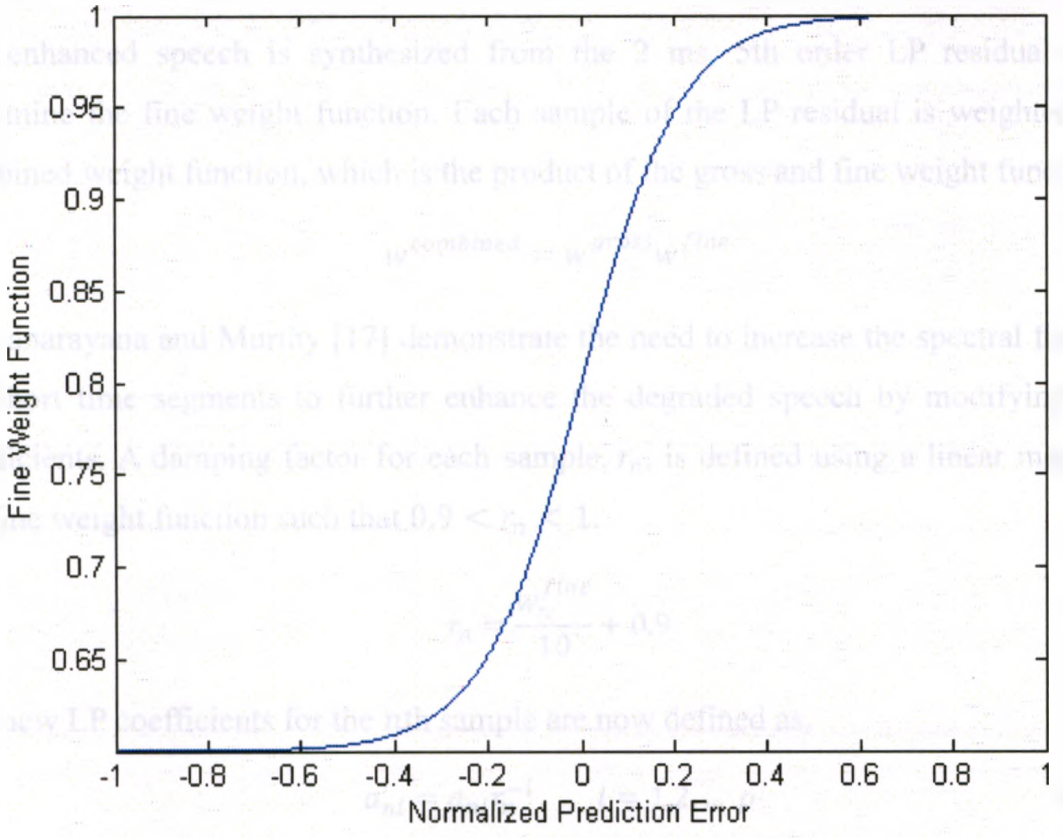


Figure 4-2: Fine weight mapping function of the normalized prediction error.

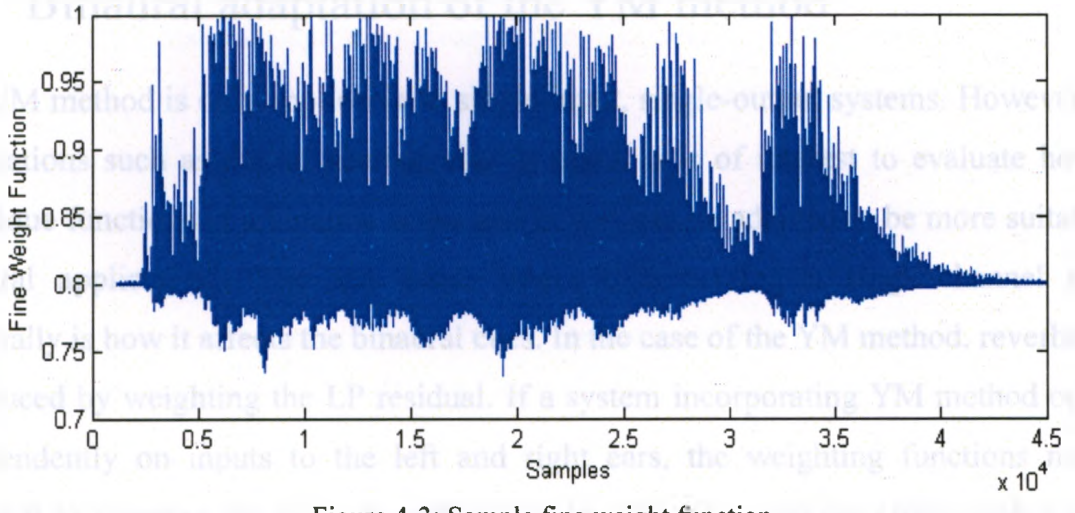


Figure 4-3: Sample fine weight function.

4.1.3 Synthesis

The enhanced speech is synthesized from the 2 ms, 5th order LP residual used to determine the fine weight function. Each sample of the LP residual is weighted by the combined weight function, which is the product of the gross and fine weight functions.

$$w^{combined} = w^{gross} w^{fine} \tag{4.8}$$

Yegnanarayana and Murthy [17] demonstrate the need to increase the spectral flatness of the short time segments to further enhance the degraded speech by modifying the LP coefficients. A damping factor for each sample, r_n, is defined using a linear mapping of the fine weight function such that $0.9 < r_n < 1$.

$$r_n = \frac{w_n^{fine}}{10} + 0.9 \tag{4.9}$$

The new LP coefficients for the nth sample are now defined as,

$$a'_{ni} = a_{ni} r_n^{-i} \qquad i = 1,2, \ldots \rho \tag{4.10}$$

where a_{ni} is the ith LP coefficient for the nth frame.

4.2 Binaural adaptation of the YM method

The YM method is only applicable to single-input, single-output systems. However, with applications such as digital hearing aids in mind, it is of interest to evaluate how this technique functions in a binaural sense and how it can be adapted to be more suitable for binaural applications. The key issue when implementing a single-channel system bilaterally is how it affects the binaural cues. In the case of the YM method, reverberation is reduced by weighting the LP residual. If a system incorporating YM method operates independently on inputs to the left and right ears, the weighting functions must be identical to preserve the ILD. A preliminary investigation into the effect such a system would have on the ILD was to examine the gross weighting function of the left and right ears for speech presented at different angles. Figure 4-4 shows the gross weight functions for the left and right inputs for the sentence "The birch canoe slid on the smooth planks."

In this example, the sentence was presented from an azimuth of 45 degrees. As seen, the right is weighted higher than the left channel at nearly all points. This effect becomes more pronounced at greater angles. Thus, the YM method exaggerates the natural ILD.

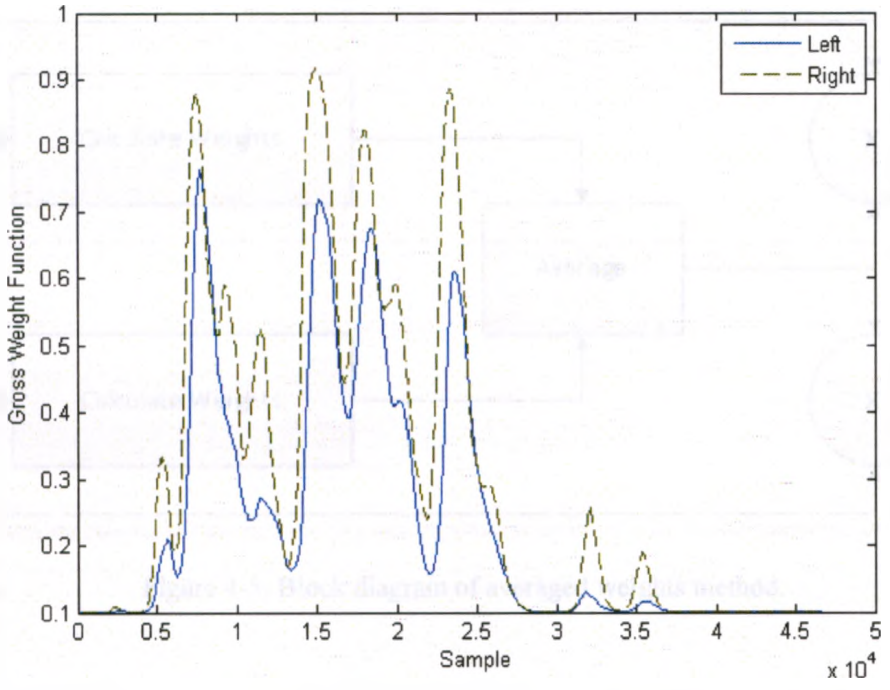


Figure 4-4: Gross weight function for a sentence presented from 45 degrees.

4.2.1 Binaural cue preserving schemes

The goal of a binaural implementation of the YM method is to preserve the ILD. The only way this goal can be achieved is if the same weighting is applied to both channels. In this way, the ILD is preserved and no loss in localization ability should be observed. Only the gross weighting function is considered, as disparity in the gross weighting functions contributes most of the ILD difference. This synchronization can be achieved in several ways. A simple method is to simply choose a channel and apply its weightings to both channels. However, only through analysis of an entire word or sentence can it be determined which weight would result in better performance. For the sentence shown in Figure 4-4, one would expect that choosing the left coefficients would result in overall poorer performance than the right channel. The right weights more accurately reflect the clean envelope since the source location was closer to the right.

Another method would be to average the resulting weight functions from both channels. It avoids solving the question of which weight results in better performance by accepting weights that guaranteed to be neither best nor the worst. Figure 4-5 shows the block diagram of such a system.

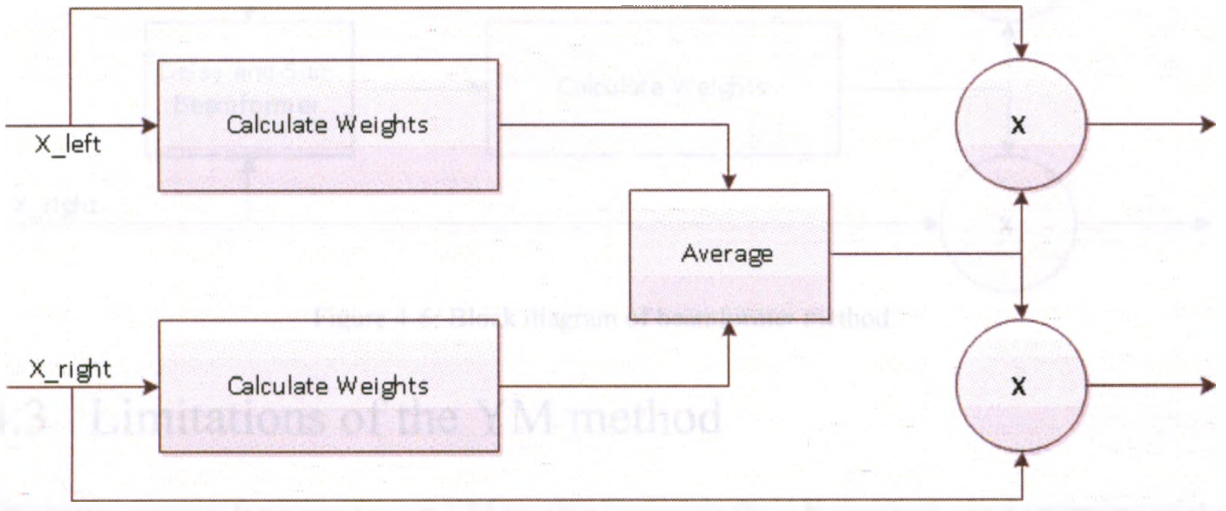


Figure 4-5: Block diagram of averaged weights method.

The final method investigated implements a delay-and-sum beamformer (DSB) to combine the input channels and then calculate the weights of the resulting signal. The implemented beamformer uses the generalized cross-correlation with phase transform method of time-delay estimation (GCC-PHAT) (see [57]). This technique presents several advantages. Since the DSB provides a degree of dereverberation by coherently adding direct speech components while incoherently adding reverberation, the entropy function of the DSB output will more closely resemble the anechoic entropy. This in turn will cause better determination of the weighting function since reverberation tends to flatten the entropy function. As well, the DSB lends itself well to scaling the system. This implementation assumes one microphone at each ear, however, it is quite common in hearing aid design to have multiple microphone inputs on each device. Thus, the DSB method should improve in quality as the number of inputs increases, while preserving binaural cues. Another advantage of the DSB method is that the weights only ever need to be calculated for one signal, reducing computational complexity for multi-microphone systems. For these reasons, the beamformer implementation was selected as the binaural

implementation of choice for the modified YM dereverberation algorithm presented in section 4.4.

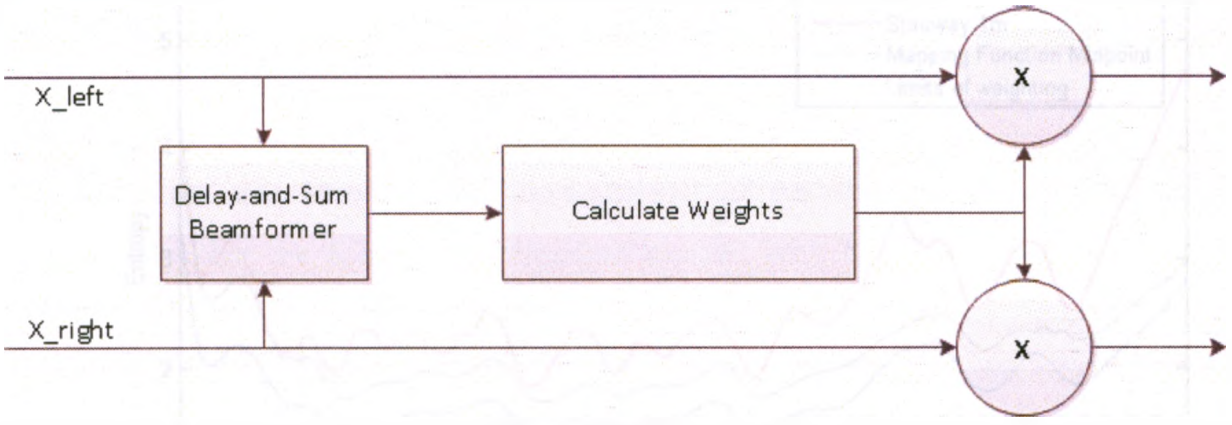


Figure 4-6: Block diagram of beamformer method.

4.3 Limitations of the YM method

The gross weight function in the YM method weights the LP residual via a mapping of its entropy. The significance of the weighting is based on the difference between the entropy of a given frame and the chosen midpoint of the mapping function. For maximum effectiveness, the mapping function should be centered on the average value of the entropy. However, an examination of the entropy of speech in different reverberant environments shows that a static mapping may not yield ideal results.

Figure 4-7 plots the smoothed entropy function of a sentence in three reverberant conditions along with the upper and lower limits of the gross weight mapping function. As seen, nearly all the entropy values fall within the range of the mapping function in the R2 condition, but in the stairway most of the entropy exceeds the upper limit.

As seen, the parameters are well suited for the R2 condition, but are ill-suited for the same sentence presented in the stairway. The effect this has on the weighting function is demonstrated in Figure 4-8. The YM method is incapable of determining useful weights for speech with entropy that falls outside the range of the defined mapping function.

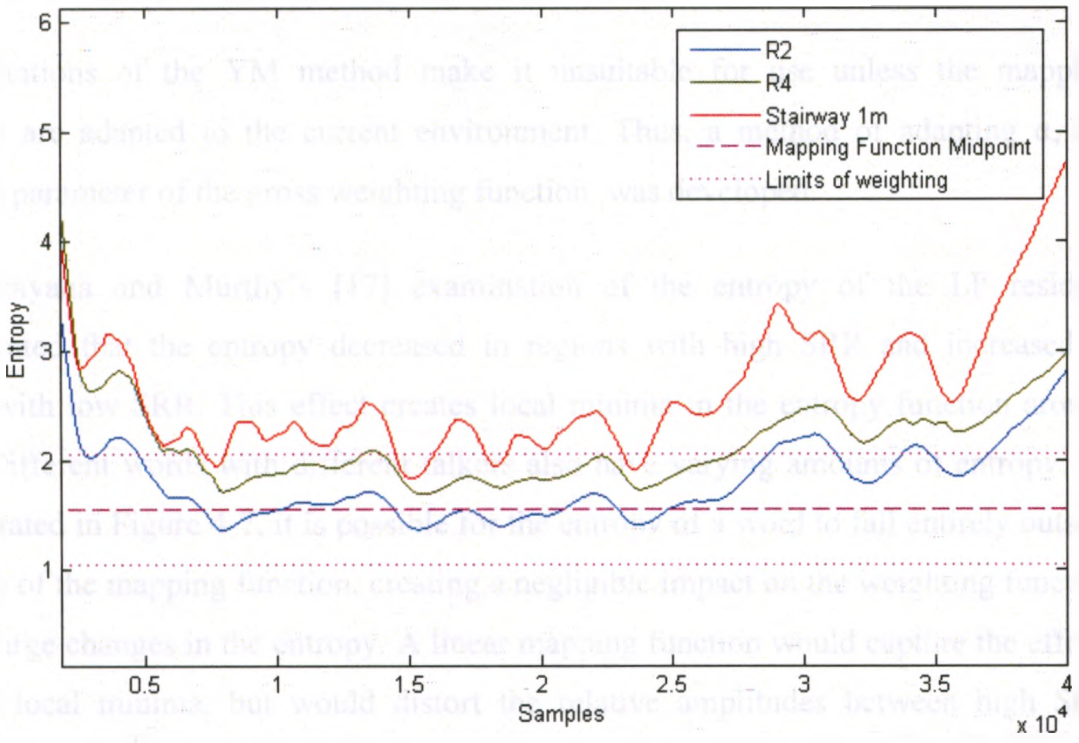


Figure 4-7: Comparison of the entropy and gross weighting parameters in several reverberant conditions.

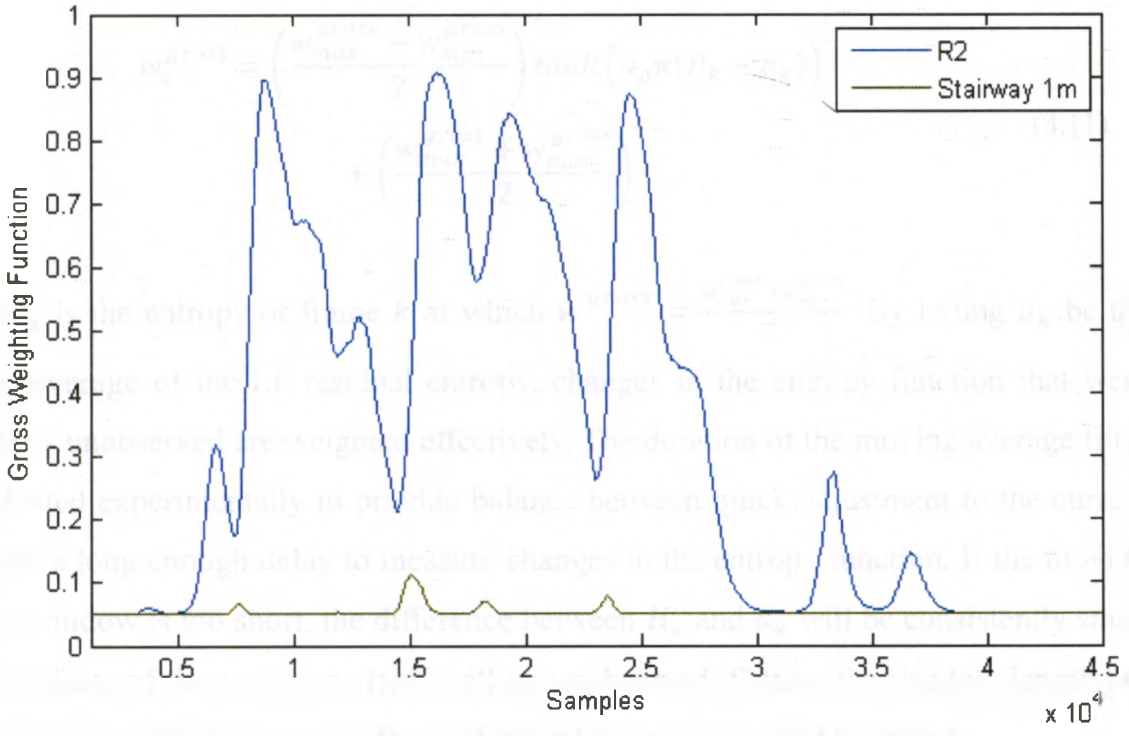


Figure 4-8: Gross weighting function using the YM method of a sentence in the R2 and stairway environments

4.4 Modifications to the YM method

The limitations of the YM method make it unsuitable for use unless the mapping functions are adapted to the current environment. Thus, a method of adapting a, the midpoint parameter of the gross weighting function, was developed.

Yegnanarayana and Murthy's [17] examination of the entropy of the LP residual demonstrated that the entropy decreased in regions with high SRR and increased in regions with low SRR. This effect creates local minima in the entropy function around words. Different words with different talkers also have varying amounts of entropy. As demonstrated in Figure 4-7, it is possible for the entropy of a word to fall entirely outside the range of the mapping function, creating a negligible impact on the weighting function despite large changes in the entropy. A linear mapping function would capture the effects of these local minima, but would distort the relative amplitudes between high SRR regions. Instead, it would be ideal to calculate the LP residual weights on a per word basis. Let a new gross weighting function be defined as

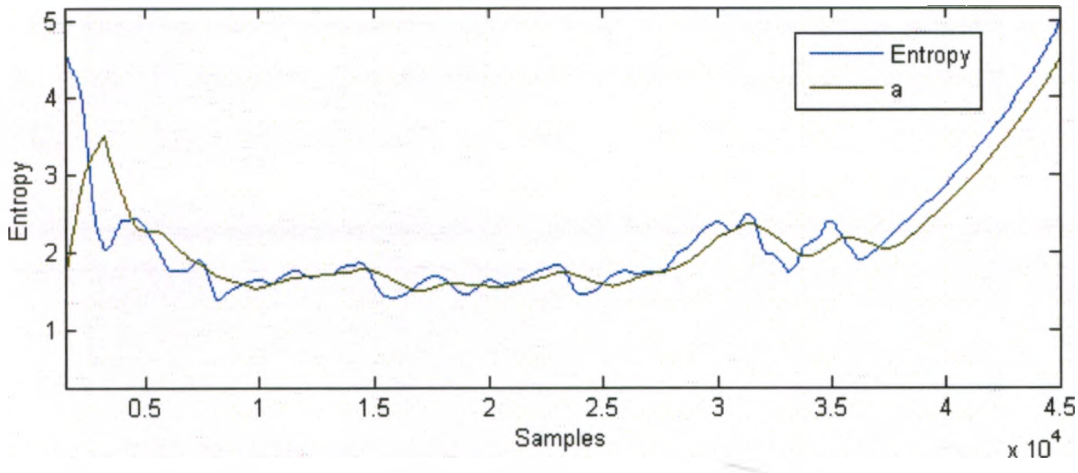
$$w_k^{gross} = \left(\frac{w_{max}^{gross} - w_{min}^{gross}}{2}\right) tanh\left(\alpha_g \pi (H_k - a_k)\right) + \left(\frac{w_{max}^{gross} + w_{min}^{gross}}{2}\right) \tag{4.11}$$

where a_k is the entropy of frame k at which $w^{gross} = \frac{w_{max}^{gross} + w_{min}^{gross}}{2}$. By letting a_k be the running average of the LP residual entropy, changes in the entropy function that were previously unobserved are weighted effectively. The duration of the moving average filter was selected experimentally to provide balance between quick adjustment to the current word and a long enough delay to measure changes in the entropy function. If the moving average window is too short, the difference between H_k and a_k will be consistently small and the effects of the variations in H will go unobserved. Setting the window length too large will cause effects similar to those observed in the original YM method.

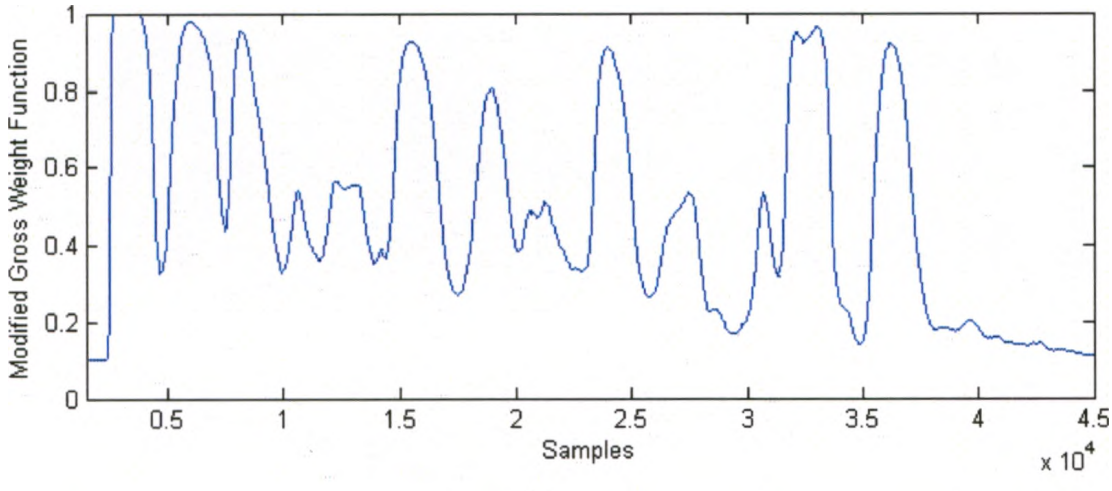
By letting a_k be the mean entropy of the previous 150 ms of H,

$$a_k = \sum_{n=0}^{0.15\text{fs}} \frac{H_{k-n}}{0.15\text{fs}} \tag{4.12}$$

This definition of a_k tracks the short time (150ms) mean of the entropy function H. When the entropy function exceeds this mean, the gross weighting function will decrease. Likewise, the gross weighting function will increase when the entropy falls below a_k. Figure 4-9 demonstrates how the modified gross weight function is derived with respect to the variable a_k.



a)



b)

Figure 4-9: Comparison of a) the variable midpoint parameter a and the smoothed entropy function and b) the resulting weighting function.

Figure 4-10 illustrates the difference in the modified method versus the original proposed by Yegnanarayana and Murthy [17]. As can be seen, there is greater definition between words and word sounds using the modified method than with the original. As well, the trailing reverberant tail is considerably reduced. Even in higher reverberation (Figure 4-11), the modified method has comparable performance, if not better.

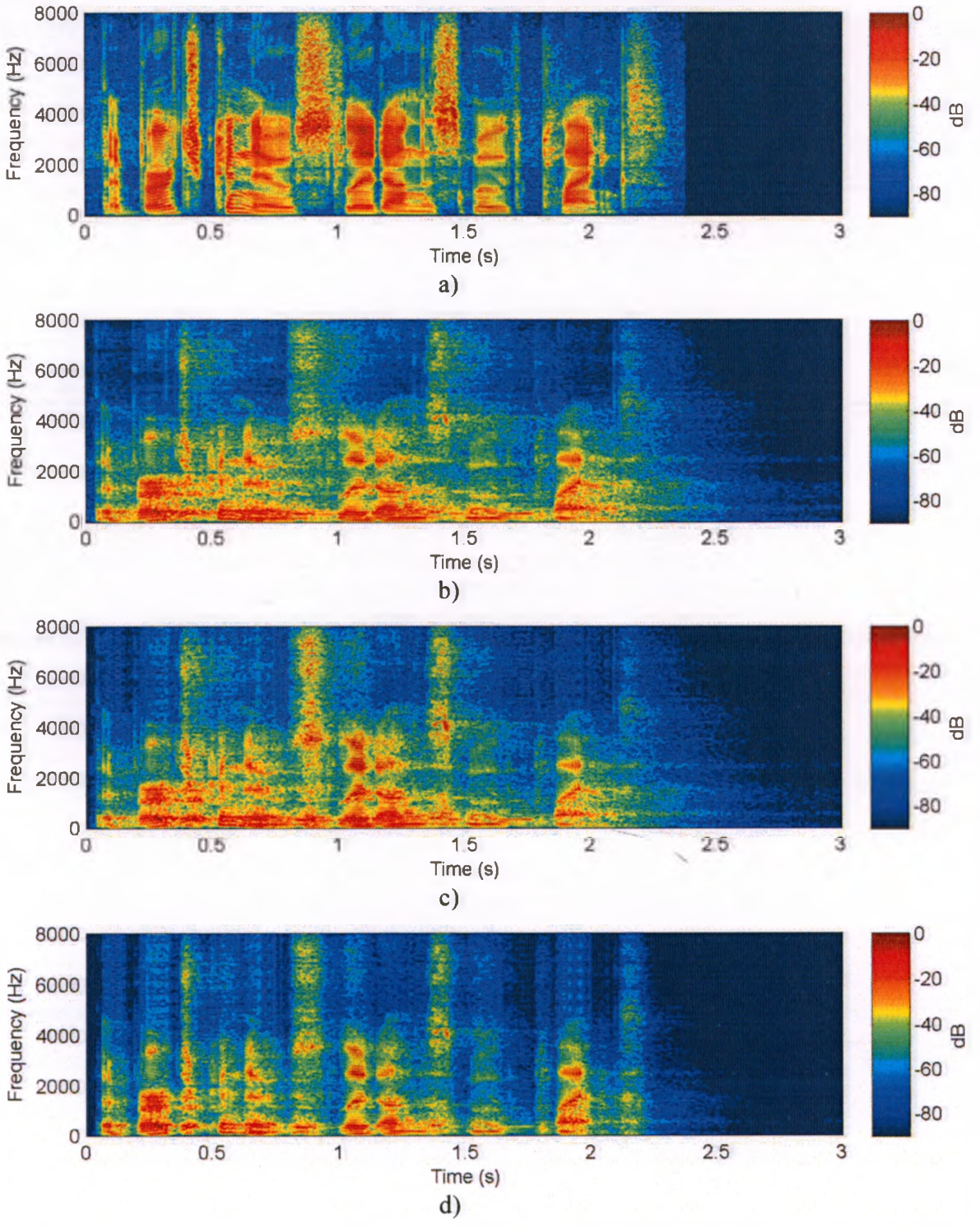


Figure 4-10: Normalized spectrograms of the sentence "The birch canoe slid on the smooth planks" presented from 0 degrees as recorded in the left ear of a HATS in the a) anechoic chamber, b) lecture hall, c) lecture hall after bilateral YM dereverberation, and d) lecture hall after binaural modified YM dereverberation

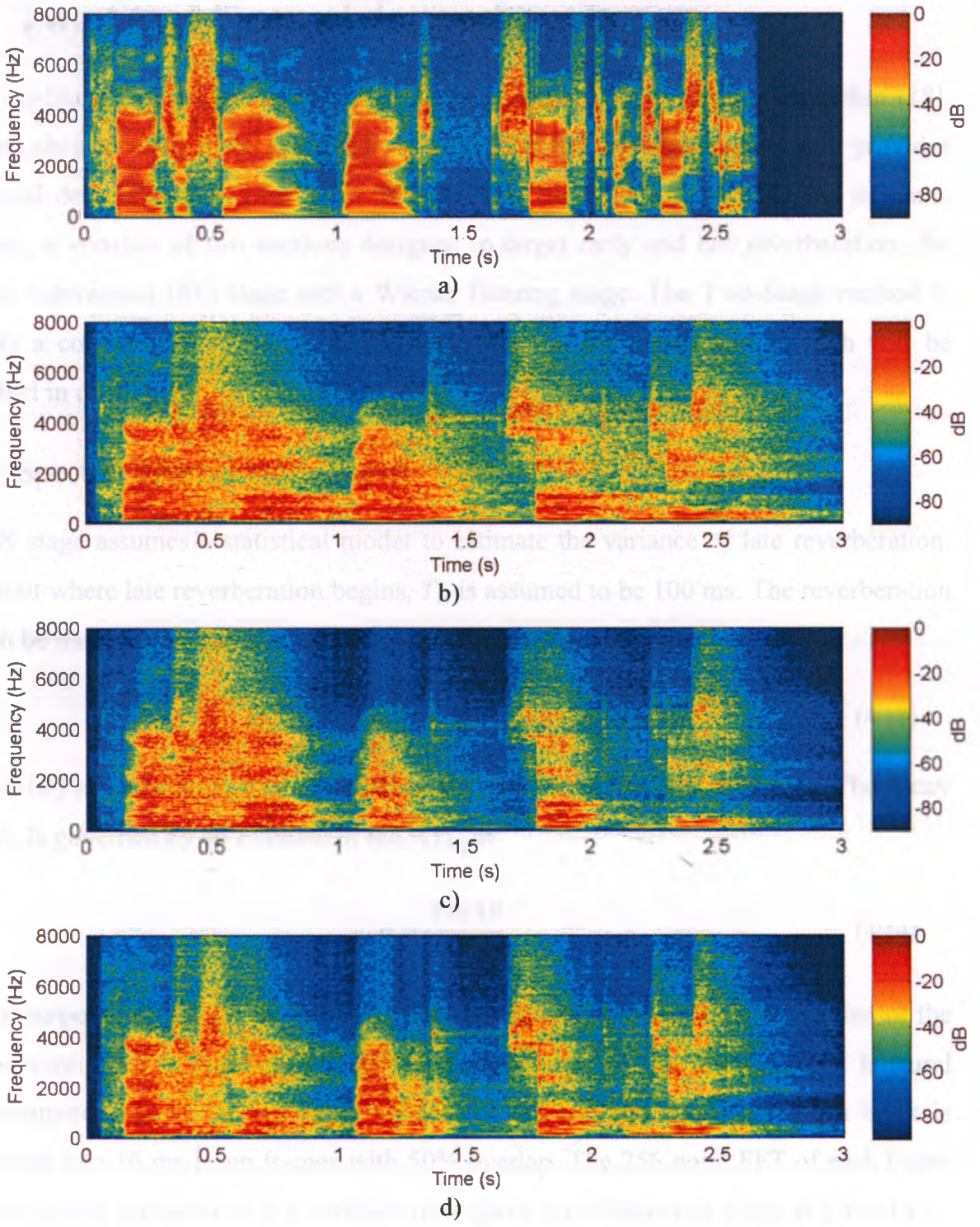


Figure 4-11: Normalized spectrograms of the sentence "Clams are small, round, soft, and tasty" presented from 0 degrees as recorded in the left ear of the HATS in the a) anechoic chamber, b) R2, c) R2 after bilateral YM dereverberation, and d) R2 after binaural modified YM dereverberation

4.5 Two-Stage binaural dereverberation

The Two-Stage method is a recently developed binaural dereverberation algorithm [18]. Results showed that it is preferred over a bilateral implementation and provides additional dereverberation over other Wiener filter based techniques [18]. As its name suggests, it consists of two sections designed to target early and late reverberation: the spectral subtraction (SS) stage and a Wiener filtering stage. The Two-Stage method is used as a comparison to the developed algorithm in this thesis, and as such will be described in detail here.

4.5.1 Spectral subtraction stage

The SS stage assumes a statistical model to estimate the variance of late reverberation. The point where late reverberation begins, T_l, is assumed to be 100 ms. The reverberation tail can be modelled as the exponentially decaying random sequence

$$\hat{h}_{late}(k) = n(k)e^{-\rho k f_s^{-1}} \; for \; k \geq 0, \tag{4.13}$$

where $n(k)$ is zero-mean, normally distributed sequence of random variables. The decay time, ρ, is governed by an estimate of the RT_{60} as

$$\rho = \frac{3\,ln\,10}{RT_{60}}. \tag{4.14}$$

For the purposes of this thesis, the RT_{60} is assumed to be known. Blind estimation of the reverberation time is possible, but is beyond the scope of this thesis. For a binaural implementation, there are two input signals for the left and right ears. Each input is segmented into 16 ms Hann frames with 50% overlap. The 256 point FFT of each frame is taken and the estimator of late variance for a given time-frequency frame is given by

$$\sigma^2_{x_{late}} = e^{-2\rho T_l}\sigma_x^2(\lambda - N_l, \mu), \tag{4.15}$$

where λ is the frame index, μ is the FFT index, and N_l is the number of frames corresponding to the time T_l. σ_x^2, the spectral variance of the reverberant speech is calculated using recursive averaging with a smoothing factor $\alpha = 0.9$ as

$$\sigma_x^2(\lambda,\mu) = \alpha \cdot \sigma_x^2(\lambda-1,\mu) + (1-\alpha)\left|X_{ref}(\lambda,\mu)\right|^2, \tag{4.16}$$

where $X_{ref}(\lambda,\mu)$ is the output of a delay and sum beamformer of the time-aligned left and right input channels. The suppression weights are then given as

$$G'_{late}(\lambda,\mu) = 1 - \frac{\sqrt{\sigma^2_{x_{late}}(\lambda,\mu)}}{|X(\lambda,\mu)|}. \tag{4.17}$$

The values of G'_{late} are then bounded by a lower bound of 0.3 to prevent overestimation of the late spectral variance. To reduce processing distortions in regions of high SNR, a measure of the power ratio between the input and enhanced signals is found as

$$\zeta(\lambda) = \frac{\sum_{\mu=0}^{M-1}|G'_{late}(\lambda,\mu)\cdot X(\lambda,\mu)|^2}{\sum_{\mu=0}^{M-1}|X(\lambda,\mu)|^2}, \tag{4.18}$$

where M is the FFT size. Based on this ratio, an adaptively sized smoothing window is applied to the weights within each time frame. The length of the window is given by

$$N_s(\lambda) = \begin{cases} 1 & \zeta(\lambda) \geq \zeta_{thr} \\ 2 \cdot round\left[\left(1 - \frac{\zeta(\lambda)}{\zeta_{thr}}\right)\cdot \Psi\right] + 1, & else, \end{cases} \tag{4.19}$$

where the threshold for windowing $\zeta_{thr} = 0.4$ and the scaling factor $\Psi = 25$. The moving average window can be expressed as an impulse response as

$$H_s(\lambda,\mu) = \begin{cases} \frac{1}{N_s(\lambda)}, & \mu < N_s(\lambda) \\ 0 & else \end{cases}. \tag{4.20}$$

Finally, the smoothed gains are applied equally to the left and right input channels.

$$\begin{aligned} S_l(\lambda,\mu) &= X_l(\lambda,\mu)\cdot\left(G'_{late}(\lambda,\mu) * H_s(\lambda,\mu)\right) \\ S_r(\lambda,\mu) &= X_r(\lambda,\mu)\cdot\left(G'_{late}(\lambda,\mu) * H_s(\lambda,\mu)\right) \end{aligned} \tag{4.21}$$

4.5.2 Wiener filter stage

In this section, only the implementation of the Wiener filter is examined. For derivation of the optimal Wiener filter gains, see [18].

First, the left and right input signals are segmented as in the SS stage (16 ms Hann frames, 256 point FFT). Next, the auto power spectral densities (APSDs) and cross power spectral density (CPSD) are recursively estimated from the input signals as follows.

$$\widehat{\Phi}_{x_l x_l | x_r x_r}(\lambda,\mu) = \alpha \widehat{\Phi}_{x_l x_l | x_r x_r}(\lambda-1,\mu) + (1-\alpha)\left|X_{l|r}(\lambda,\mu)\right|^2 \tag{4.22}$$

$$\widehat{\Phi}_{x_l x_r}(\lambda,\mu) = \alpha \widehat{\Phi}_{x_l x_r}(\lambda-1,\mu) + (1-\alpha) X_l(\lambda,\mu)\cdot X_r^*(\lambda,\mu) \tag{4.23}$$

The smoothing factor α is set as 0.8. The estimated APSD of the original uncorrupted signal is given by

$$\widehat{\Phi}_{ss}(\lambda,\mu) = \frac{Re\{\widehat{\Phi}_{x_l x_r}(\lambda,\mu)\} - \frac{1}{2} Re\{\Gamma_{x_l x_r}(\Omega)\}\left(\widehat{\Phi}_{x_l x_l}(\lambda,\mu) + \widehat{\Phi}_{x_r x_r}(\lambda,\mu)\right)}{1 - Re\{\Gamma_{x_l x_r}(\Omega)\}}, \tag{4.24}$$

where $\Gamma_{x_l x_r}(\Omega)$ is the head coherence function given in (2.5) and $Re\{\cdot\}$ returns the real part of its argument. The coherence function is bounded to $0 \le \Gamma_{x_l x_r}(\Omega) \le 0.99$ to prevent division by zero errors. The Wiener filter gains are then determined as

$$G(\lambda,\mu) = \frac{\widehat{\Phi}_{ss}(\lambda,\mu)}{\frac{1}{2}\cdot\left(\widehat{\Phi}_{x_l x_l}(\lambda,\mu) + \widehat{\Phi}_{x_r x_r}(\lambda,\mu)\right)}. \tag{4.25}$$

A lower limit of 0.3 is applied to the Wiener filter gains to control against overestimation of errors. The spectral weights are then applied evenly to each of the two channels.

$$\begin{aligned} \hat{S}_l(\lambda,\mu) &= X_l(\lambda,\mu)\cdot G(\lambda,\mu) \\ \hat{S}_r(\lambda,\mu) &= X_r(\lambda,\mu)\cdot G(\lambda,\mu) \end{aligned} \tag{4.26}$$

Chapter 5

Subjective and Objective Evaluation

In this Chapter, subjective and objective evaluations are used to analyze the performance of bilateral and binaural dereverberation in terms of sound quality and localization. First, the dereverberation algorithms are evaluated by overall sound quality using the objective SRMR metric and through a subjective sound quality study. Second, the algorithms are evaluated by their ability to preserve binaural localization cues. Objectively, the ILDs are measured using the variable selection threshold method described in section 2.1.2. Subjective performance was measured through a localization task.

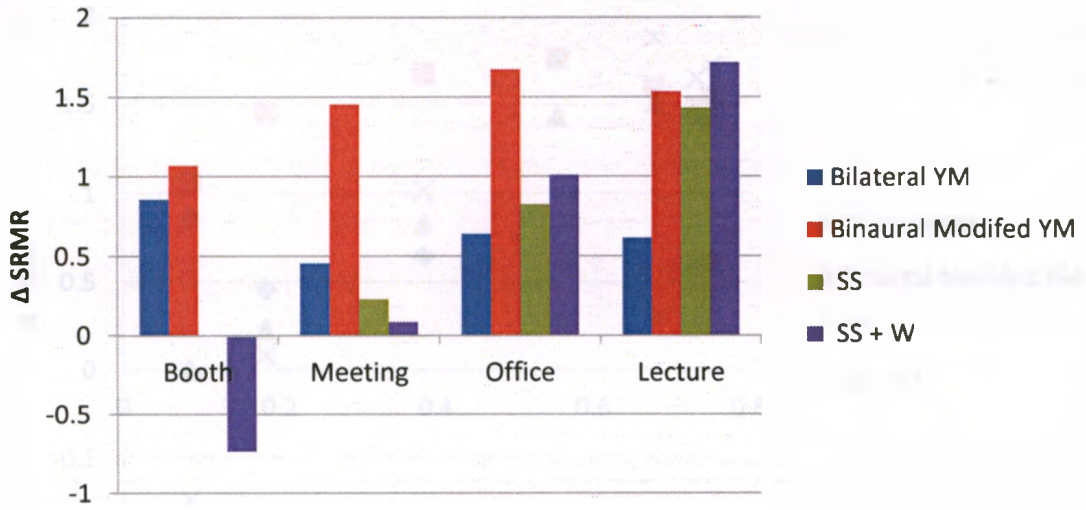
5.1 Sound quality evaluation

Overall sound quality was evaluated objectively and subjectively for five different conditions: unprocessed (reverberant), bilateral YM, binaural modified YM (BMYM), binaural spectral subtraction (SS) (first stage of the Two-Stage method [18]), and binaural SS + Wiener (SS+W) filtering (Two-Stage method in [18]).

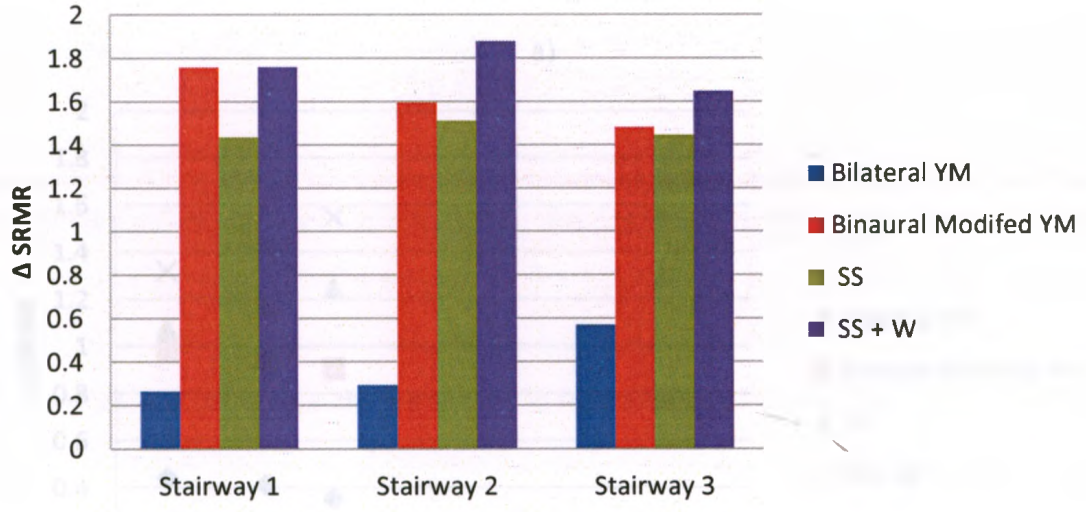
5.1.1 Objective evaluation

Objective evaluation of each condition was performed using 16 sentences (8 male, 8 female) from the TIMIT corpus of speech [58]. For each room condition in the AIR and NCA BRIR databases, the BRIR was downsampled to 16 kHz to match that of the clean speech. The speech samples were then convolved with the downsampled BRIRs to

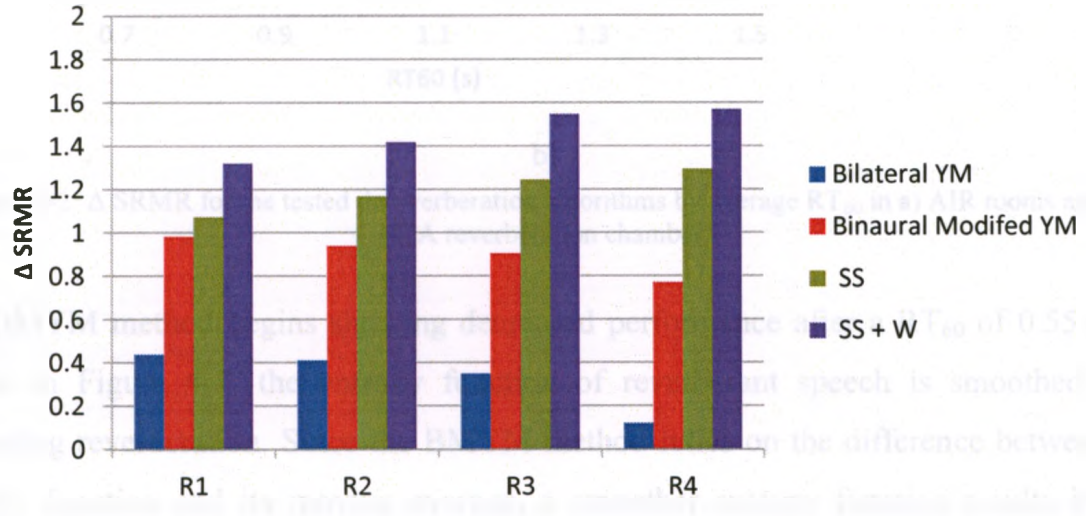
generate reverberant speech. The resulting corrupted speech was then processed using the aforementioned dereverberation techniques. The SRMR score is averaged across the left and right channels of the reverberant/processed speech and across all talkers. In rooms where angular data is available (AIR Stairwell and all NCA conditions), the SRMR is averaged across available positions from -90 to 0 degrees azimuth. Figure 5-1 shows the improvement in SRMR score from each dereverberation technique with respect to the reverberant condition. In the sound booth, meeting room, and office, the BMYM method outperforms the Two-Stage method. In the stairwell conditions they have comparable performance. However, in the NCA reverberation chambers, the Two-Stage method outperforms the BMYM method. In all but the least reverberant conditions, the YM method demonstrated the least SRMR improvement. Obviously, the reverberant conditions have an effect on the SRMR improvement. To investigate this effect, the SRMR improvement scores were plotted against the RT_{60} of each room condition in Figure 5-2. The results are separated between the AIR and NCA rooms because the characteristics of the NCA reverberation chamber caused an overall lower SRMR improvement in all algorithms. The BMYM method significantly outperforms the Two-Stage method until an RT_{60} of 0.55 seconds (AIR Stairway 1). After this point, the Two-Stage method shows greater SRMR improvement. This effect can be explained by the nature of the Two-Stage method, namely the SS block, and the effects of reverberation on the YM method. The SS stage of the Two-Stage method is designed specifically to target late reverberation in the signal. To do so, it requires an estimate of the RT_{60} to statistically model the late reverberation, which for the purposes of this test was assumed to be known. As the reverberation time increases, and hence the impact late reverberation has on the SRMR, the SS block should offer greater improvement.



a)

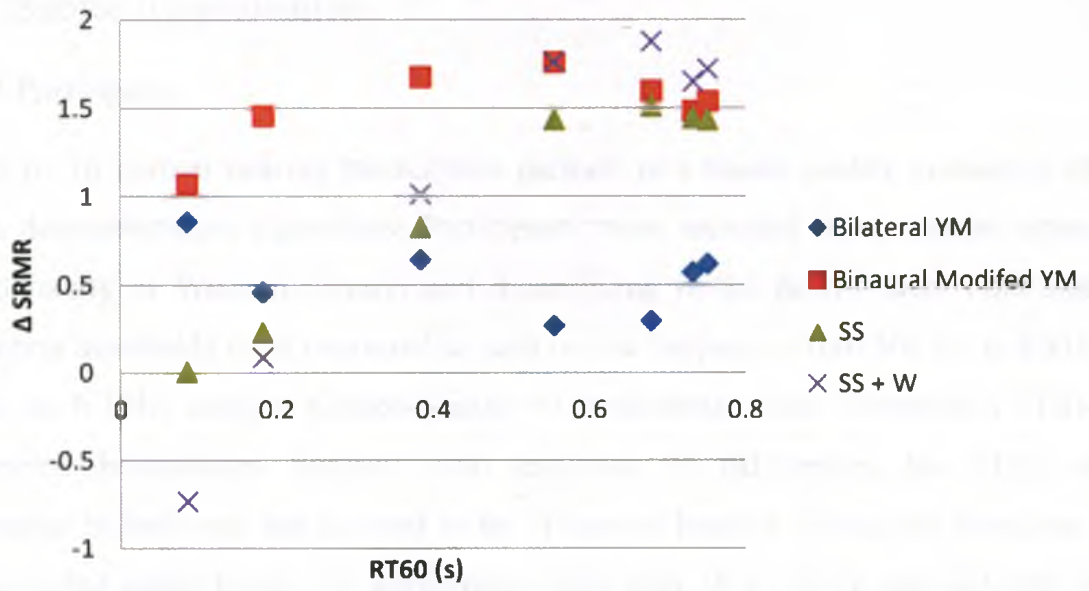


b)

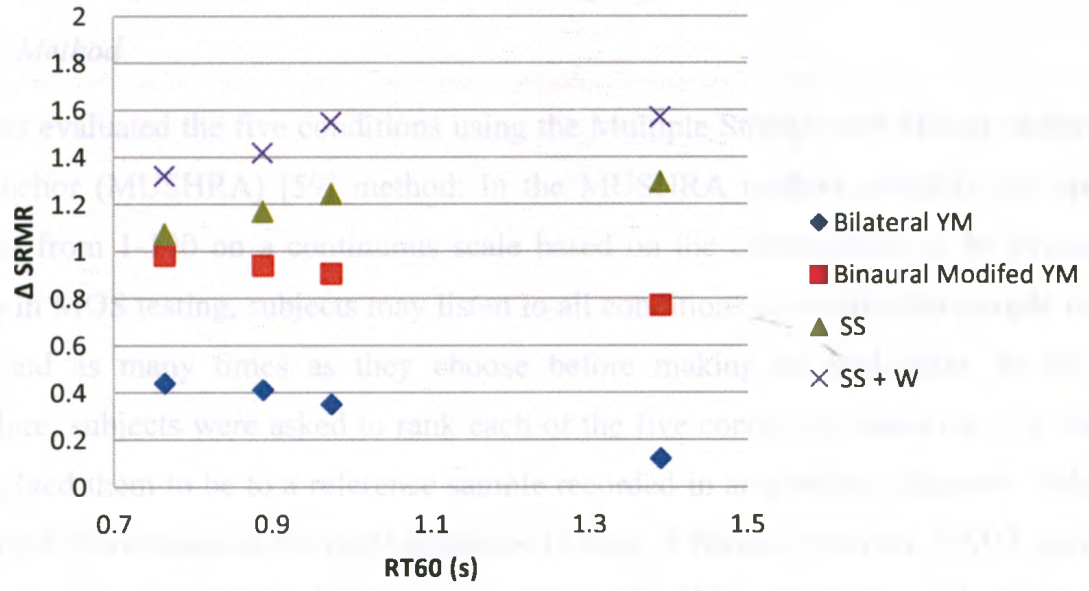


c)

Figure 5-1: Improvement in SRMR score in the a) AIR rooms, b) AIR stairwell, c) NCA anechoic and reverberation chamber.



a)



b)

Figure 5-2: Δ SRMR for the tested dereverberation algorithms by average RT_{60} in a) AIR rooms and b) NCA reverberation chamber.

The BMYM method begins showing decreased performance after a RT_{60} of 0.55 s. As shown in Figure 4-7, the entropy function of reverberant speech is smoothed with increasing reverberation. Since the BMYM method relies on the difference between the entropy function and its moving average, a smoother entropy function results in less profound changes in the gross weighting function. Thus, a decrease in SRMR improvement is expected as the RT_{60} is increased.

5.1.2 Subjective evaluation

5.1.2.1 Participants

A total of 16 normal hearing participants partook in a sound quality evaluation of the chosen dereverberation algorithms. Participants were recruited from students attending the University of Western Ontario and those living in the nearby area. Pure-tone air conduction thresholds were measured at each octave frequency from 500 Hz to 8 kHz, as well as at 6 kHz, using a Grason-Stadler 61 audiometer with Telephonics TDH-50P audiometric headphones. Subjects with less than 25 dB hearing loss (HL) at all frequencies in both ears are deemed to be of normal hearing. Screening was done in a double-walled sound booth. All participants were ages 18 or above and had little to no prior experience in sound quality evaluation.

5.1.2.2 Method

Subjects evaluated the five conditions using the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) [59] method. In the MUSHRA method, subjects rate speech samples from 1-100 on a continuous scale based on the characteristic to be evaluated. Unlike in MOS testing, subjects may listen to all conditions of a particular sample in any order and as many times as they choose before making an evaluation. In the test procedure, subjects were asked to rank each of the five conditions based on how similar they judged them to be to a reference sample recorded in an anechoic chamber. Subjects performed this evaluation for eight sentences (4 male, 4 female) from the TIMIT database [58] generated using the NCA R2 room condition. The R2 condition was selected to get a reverberation time between the worst case scenario (R4) and more commonly experienced reverberation times (AIR stairwell and lecture hall). The ordering of the five conditions for each sentence is randomly generated. The order in which subjects rate each sentence is also randomized. The speech sentences were equalized to have equivalent power and presented to the subjects over Sennheiser HDA 200 headphones while in a double-walled sound booth.

5.1.2.3 Results

First, the subjective results were tested for reliability. Cronbach's alpha for the test data was found to be 0.959, indicating a high degree of inter-subject agreement. Figure 5-3 displays the speech quality scores averaged across all talkers and all subjects for each condition, with error bars representing one standard deviation. It can be seen that the unprocessed reverberant speech received the highest rating and the bilateral YM algorithm the lowest. Statistical analyses were completed to further probe these results. Repeated measures ANOVA was performed on the subjective scores using SPSS statistical software package to investigate the effect of talker gender, algorithm, and their interaction. No statistical significance was found between the scores of male and female speech samples ($F = 2.708$, $p = 0.121$). Statistically significant differences were found for all algorithmic conditions with respect to the unprocessed condition with the exception of SS ($p = 0.064$). Post-hoc tests of significance of contrasts between the algorithms were performed using the False Discovery Rate (FDR) control procedure [60]. Results revealed that all of the binaural algorithms were statistically better than the bilateral YM algorithm. Furthermore, the SS method was statistically better than both BMYM and SSW method, and there was no statistical difference between the BMYM and SSW scores ($p = 0.451$). These results agree with those reported in the literature. For example, in [18], 17 experienced listeners judged the speech quality of bilateral and binaural dereverberation algorithms in comparison to that of the unprocessed sample. They reported that on average 73.6% preferred the binaural implementation of the SS algorithm over a bilateral implementation.

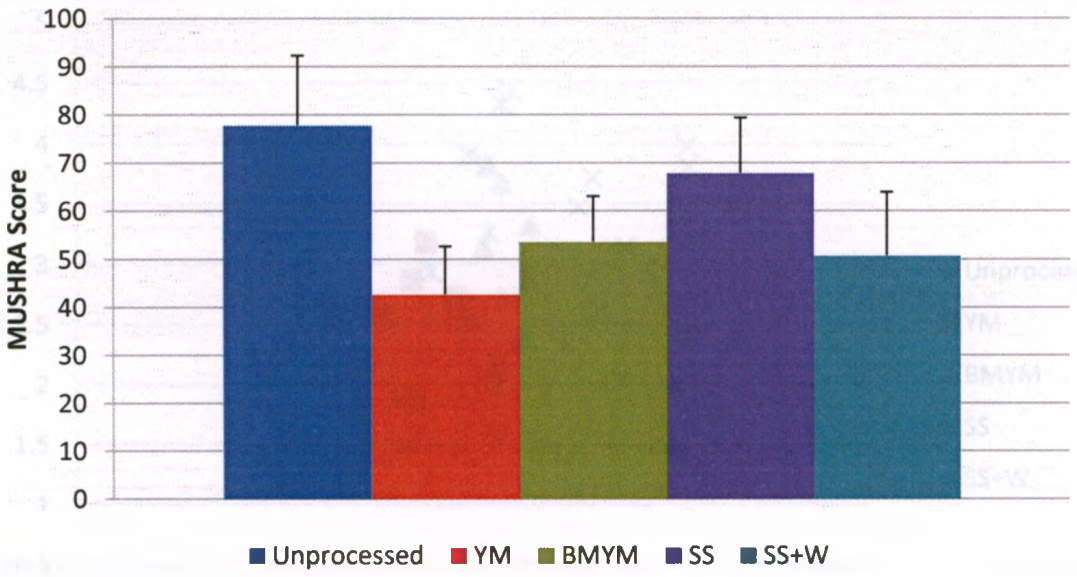


Figure 5-3: Overall scores across all subjects and sentences for each condition.

Another notable feature of Figure 5.3 is the fact that subjects ranked the unprocessed speech more highly despite an expected decrease in reverberation with the dereverberation algorithms. It is thought that normal hearing listeners, as used in this study, are more perceptive to unnaturalness caused by speech enhancement than the apparent reduction in reverberation. In the YM and BMYM methods, unnaturalness might stem from the changing amplitude of the waveform caused by the gross weighting function. The SS and SS+W methods introduce musical noise due to the spectral filtering. In [26], similar discrepancies between objective and subjective results were reported and thought to be due to processing distortions.

[18]Figure 5-4 shows the results of the subjective study compared against the SRMR predictor. The Pearson's correlation coefficient for the entire data set is -0.16, indicating weak negative correlation. This result is counter-intuitive and implies that the SRMR is not a useful predictor of subjective results. The discrepancy in the objective and subjective scores for the unprocessed speech is an obvious contributor to this poor correlation. Recalculating the correlation coefficient after removing the unprocessed condition resulted in a r value of 0.35. Here we see the correlation has improved significantly, but still not to the degree reported in [27].

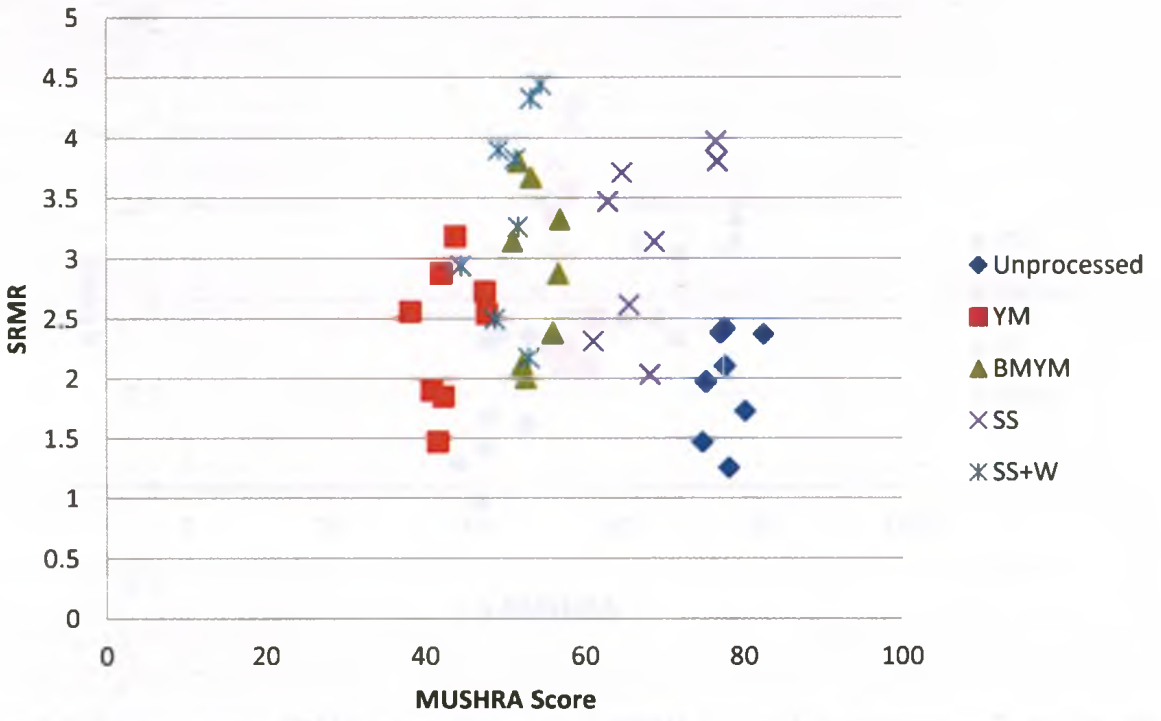


Figure 5-4: Average MUSHRA scores across listeners vs SRMR scores for each sentence in the subjective study.

To mitigate these effects, the difference in subjective score from the reverberant condition is compared against the Δ SRMR in Figure 5-5. The correlation between the Δ SRMR and Δ MUSHRA is 0.46, an improvement over the correlation of the absolute scores. To explain the reduced correlation of the SRMR with respect to [27], the differences in testing must be recognized. Falk and Chan [27] determined the correlation using reverberant speech and dereverberated speech from a DSB at a variety of RT_{60}s. Unprocessed speech clearly will have no processing distortions, and DSBs introduce little distortion [26]. This suggests that while the SRMR is a good predictor of undistorted reverberant speech, results indicate that it is not robust towards predicting the perception of processing distortions.

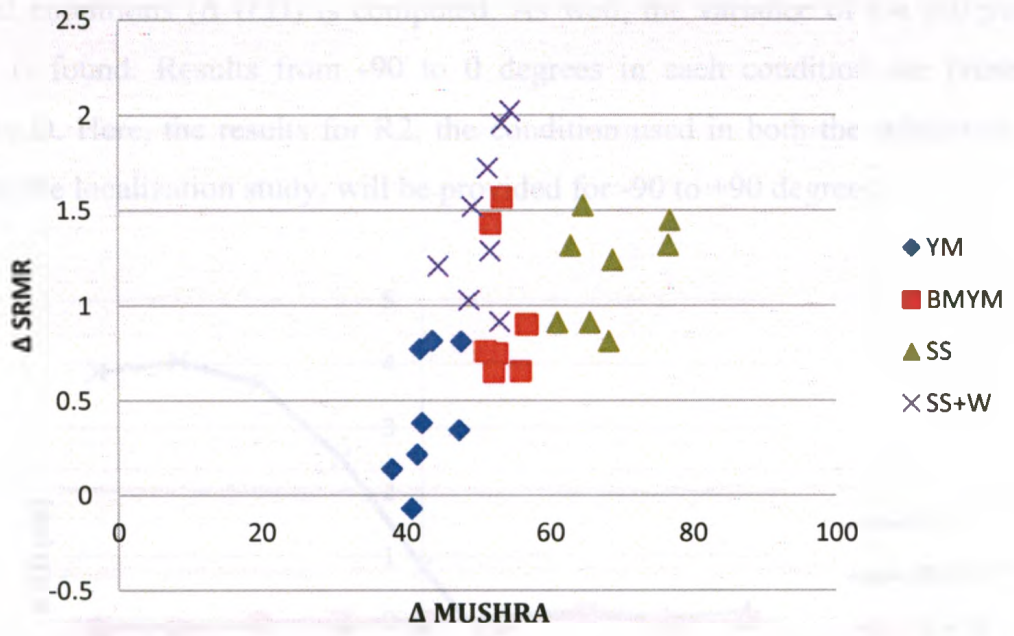


Figure 5-5: Improvement in SRMR vs Improvement in MUSHRA score for each sentence in the subjective study.

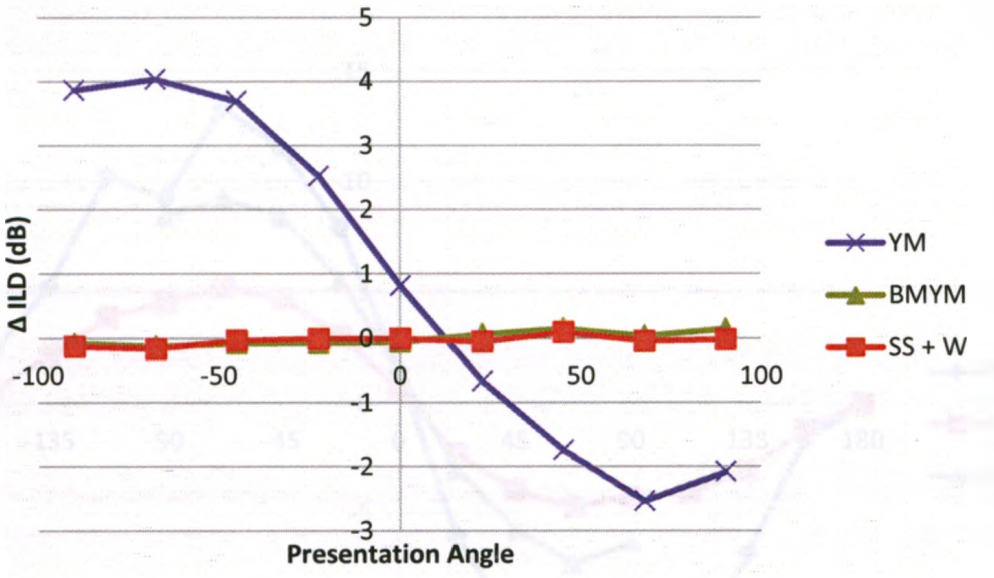
5.2 Localization evaluation

To assess the degree to which bilateral processing disturbs localization, an objective and subjective study is presented. Objectively, the ILD cue is calculated using the selective technique described by [18]. The ITD is not investigated as the dereverberation algorithms under investigation all preserve the ITD. Subjectively, a localization task is presented to accurately measure localization performance in subjects. The unprocessed, YM, BMYM, and SS+W methods are examined. The SS method is not evaluated independently because the SS+W method uses the same method of binaural synchronization in both stages.

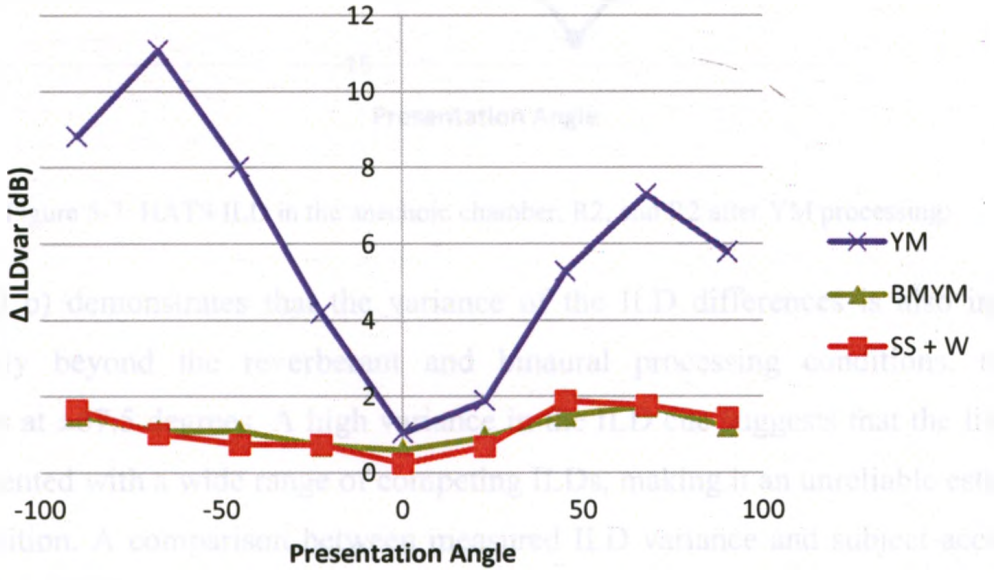
5.2.1 Objective evaluation

The ILD was calculated for the same 16 sentences from the TIMIT database as used in the sound quality evaluation. For the AIR Stairwell conditions and the NCA reverberation conditions, the mean difference in ILD between the reverberant and

processed conditions (Δ ILD) is computed. As well, the variance of the differences (Δ ILDvar) is found. Results from -90 to 0 degrees in each condition are presented in Appendix D. Here, the results for R2, the condition used in both the subjective quality study and the localization study, will be provided for -90 to +90 degrees.



a)



b)

Figure 5-6: a) Mean Δ ILD and b) Δ ILD variance in R2 by angle.

Figure 5-6 a) shows that binaural algorithms are effective at preserving the ILD while the bilateral YM method significantly changes the ILD from the reverberant condition. The changes in the mean ILD for the binaural algorithms are all less than 0.25 dB, below the

minimum audible difference in ILD of 0.5 dB [14]. Since the ILD determined here is the difference of the right from the left channel, the YM Method is exaggerating the ILD beyond what would normally be heard in the reverberant environment. To illustrate this effect, Figure 5-7 shows the ILD measured in the anechoic chamber and R2, along with a composite ILD formed by adding the mean Δ ILD to the R2 ILD.

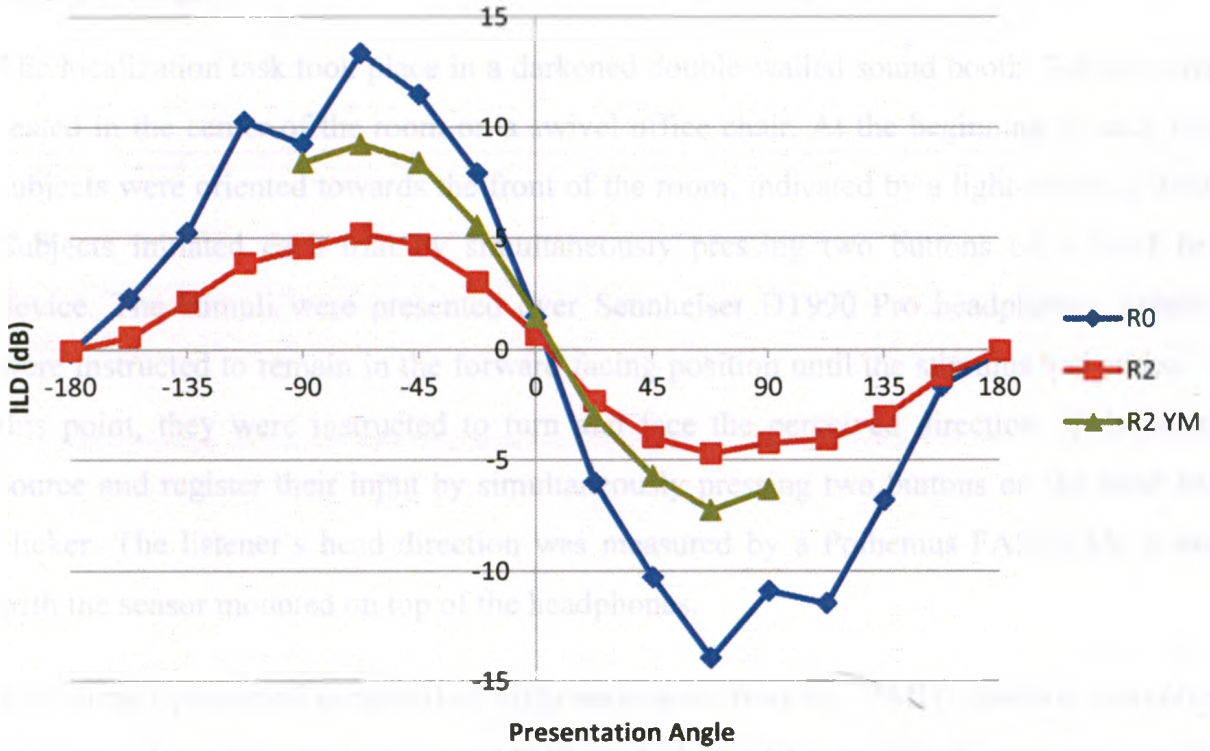


Figure 5-7: HATS ILD in the anechoic chamber, R2, and R2 after YM processing.

Figure 5-6 b) demonstrates that the variance of the ILD differences is also increased significantly beyond the reverberant and binaural processing conditions, reaching maximums at ±67.5 degrees. A high variance in the ILD cue suggests that the listener is being presented with a wide range of competing ILDs, making it an unreliable estimate of source position. A comparison between measured ILD variance and subject accuracy is performed in 5.2.2.3.

5.2.2 Subjective evaluation

5.2.2.1 Participants

Participant selection and screening was performed identically to that of the sound quality study (see section 5.1.2.1). Six participants took part in both studies.

5.2.2.2 Method

The localization task took place in a darkened double-walled sound booth. Subjects were seated in the center of the room on a swivel office chair. At the beginning of each trial, subjects were oriented towards the front of the room, indicated by a light-emitting diode. Subjects initiated each trial by simultaneously pressing two buttons on a hand held device. The stimuli were presented over Sennheiser DT990 Pro headphones. Subjects were instructed to remain in the forward facing position until the stimulus had ended. At this point, they were instructed to turn and face the perceived direction of the sound source and register their input by simultaneously pressing two buttons on the hand held clicker. The listener's head direction was measured by a Polhemus FASTRAK system with the sensor mounted on top of the headphones.

The stimuli presented consisted of eight sentences from the TIMIT database convolved with the R2 condition and processed with the YM, BMYM, and SS+W methods at angles from -90 to +90 degrees. Three second bursts of wideband white noise in the R0 condition were used to determine anechoic accuracy. All stimuli were originally sampled at 16 kHz, and then up-sampled to 48828 Hz to work with the head-tracking system. To avoid front-back confusions, subjects were informed that the possible range of positions was from -90 to +90 degrees.

Subjects were first trained on the system, completing 45 trials with anechoic white noise presented in random order. Next, subjects began the actual localization task consisting of 288 speech stimuli (8 sentences x 9 angles x 4 conditions) and 45 (9 angles x 5 repetitions) wideband white noise stimuli. The order of presentation was randomized for each subject. Subjects completed the localization task in five sets, each consisting of 60-70 trials, and were allowed breaks in between sets to reduce listening fatigue.

5.2.2.3 Results

Figure 5-8 plots the average response azimuth against the target azimuths for each condition. Table 5-1 includes the standard deviation of each condition at each target azimuth. The range of response azimuth is condensed, with the average response azimuth being ±70 degrees when the target azimuth was ±90. Since this effect is seen in both the anechoic and reverberant conditions, the cause must not be external. A possible cause could be that during the instruction, participants were informed of the range of target azimuths, and not wanting to respond inaccurately, made conservative estimates when evaluating the outer angles. Since the error in azimuth estimation from the target azimuth would be greatly biased in the outer angles because of this trend, the error is instead computed from a linear fit of the anechoic white noise responses for each subject, seen in Figure 5-9.

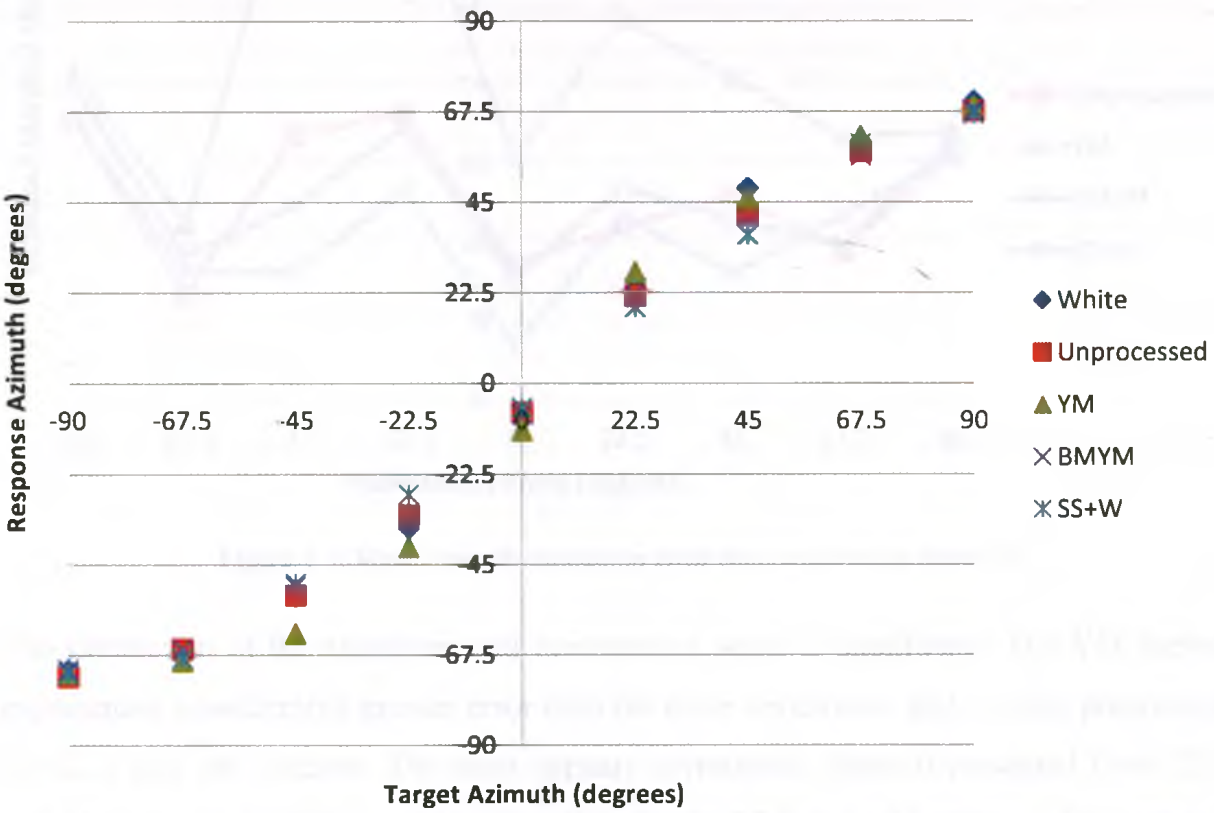


Figure 5-8: Overall response of azimuth across 16 subjects.

Angle of Presentation	White	Unprocessed	YM	BMYM	SS+W
-90	13.48	11.27	12.15	11.23	11.26
-67.5	11.83	11.04	12.62	12.39	11.70
-45	15.04	11.09	14.41	10.43	9.73
-22.5	13.27	10.56	13.85	10.95	11.00
0	8.83	9.61	11.06	9.73	8.90
22.5	7.89	11.12	14.88	11.80	13.69
45	19.18	13.29	14.65	12.26	11.92
67.5	16.52	14.82	15.79	13.53	13.90
90	15.22	15.69	15.39	15.52	15.20

Table 5-1: Standard deviation of overall response azimuths from all 16 subjects.

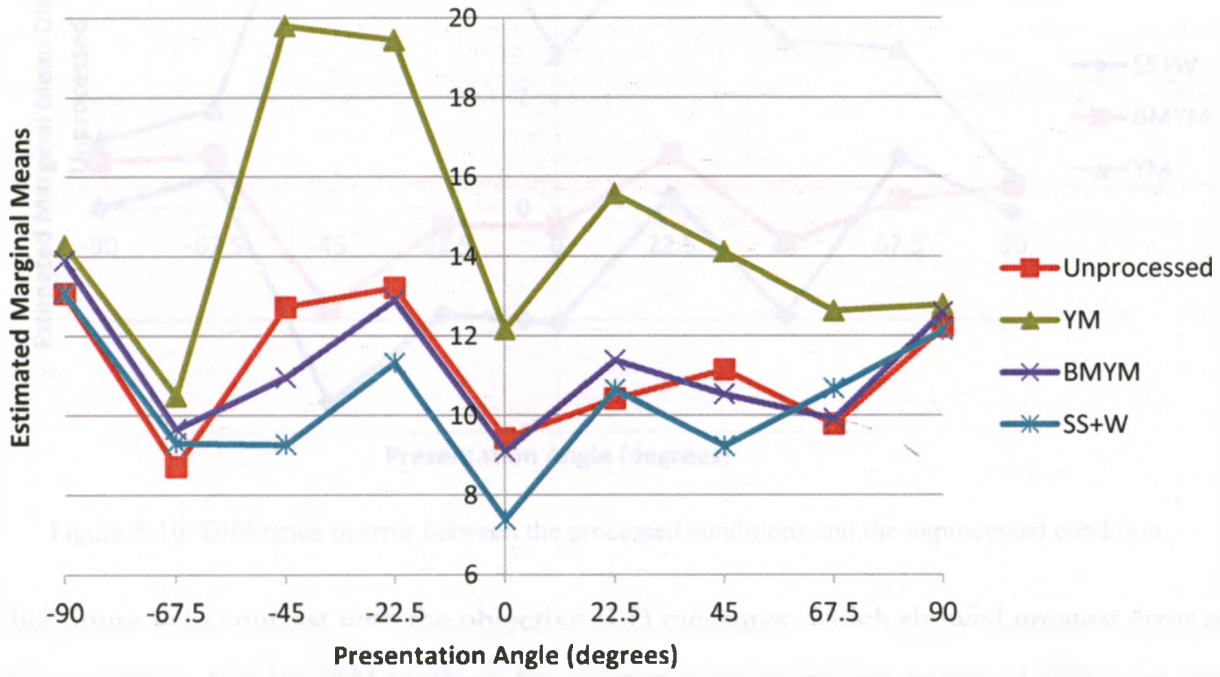


Figure 5-9: Mean azimuth estimation error from white noise linear fit.

The interaction of the algorithm and presentation angle is significant. The YM method experiences considerably greater error than the other conditions, and is most pronounced at ±22.5 and ±45 degrees. The error appears asymmetric. Stimuli presented from -22.5 and -45 degrees exhibited greater error than from +22.5 and +45 degrees. This result is believed to have occurred due to asymmetry in the room setup. This result would seem to not agree with the measured Δ ILD means and variances, which show greatest error at ±67.5 degrees. However, the error in localization is dependent on many factors other than the ILD. It is well known that it becomes more difficult to discriminate between locations

as azimuth increases [61]. As well, the listener also uses the ITD to aid in azimuth localization. To mitigate these effects, the effect of modifying the ILD is isolated by looking at the difference between the processed and unprocessed errors. The result is shown in Figure 5-10. The error difference at the outer angles is now considerably less than the error difference between -45 and +67.5 degrees.

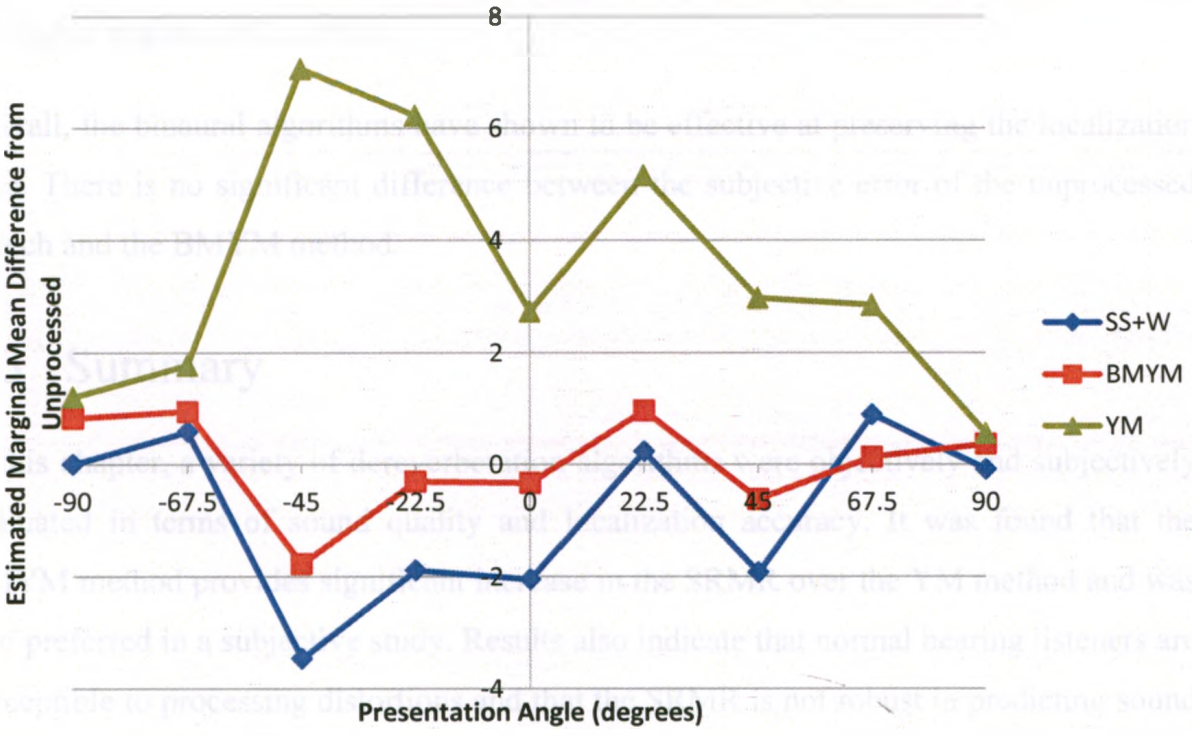


Figure 5-10: Difference in error between the processed conditions and the unprocessed condition.

This figure is in contrast with the objective ILD measures, which showed greatest error at ±67.5 degrees. For the YM method, the listener is experiencing expected ITD cues and biased ILD cues at any given angle. Macpherson and Middlebrooks [11] showed that the ILD can be biased in opposition of the ITD to shift a listener's perception of the location and developed dimensionless weights to explain the relative effects of biasing the ILD and the ITD. As an example, if the natural ILD is 5 dB and was biased by 10 dB and the subject indicated that the location occurred at the natural position, the ILD weight was said to be zero. Likewise, if the subject had indicated that the location was at a position for which the ILD is 15 dB, the ILD weight was said to be 1. For wideband noise, it was shown that the ILD bias weight is approximately 0.5 [11]. In other words, biasing the ILD has moderate influence on a listener's perceived sound source location for a

wideband sound. The subjective results can be explained in this way. The error is high in regions where the ILD has been biased within normal ILD levels. However, Macpherson and Middlebrooks [11] did not investigate the effects of biasing the ILD outside of naturally occurring values. It is suspected that when the ILD is raised outside the range of natural ILDs, there is no added bias for the subject in the corresponding direction of the ILD, and hence a decrease in error with respect to the reverberant case is experienced at the higher angles (±67.5, ±90).

Overall, the binaural algorithms have shown to be effective at preserving the localization cues. There is no significant difference between the subjective error of the unprocessed speech and the BMYM method.

5.3 Summary

In this chapter, a variety of dereverberation algorithms were objectively and subjectively evaluated in terms of sound quality and localization accuracy. It was found that the BMYM method provides significant increase in the SRMR over the YM method and was also preferred in a subjective study. Results also indicate that normal hearing listeners are susceptible to processing distortions and that the SRMR is not robust in predicting sound quality results when audible processing distortions are present.

The binaural algorithms preserve the ILD to within 0.25 dB, well within the minimum audible difference. The YM method exhibited significant increase in both objective measures of the ILD and in azimuth estimation accuracy from a subjective study. However, the ILD measures do not agree with the subjective error.

Chapter 6

Conclusion

Recent advancements in wireless technology have created a new class of digital hearing aid. These binaural hearing aids communicate and share information over a wireless link, opening up possibilities for synchronization of processing for the devices on each ear. Classical dereverberation techniques are designed bilaterally and little work has been done on investigating how binaural dereverberation may offer improvements in speech quality and localization. Objective speech quality measures are an important tool for the creation and verification of speech enhancement algorithms. Objective and subjective measures are used to evaluate these effects. Objective measures allow for on-demand feedback, which can be especially useful for procedures that would otherwise require an infeasible number of trials for parameter optimization. Measures for predicting the quality of speech in reverberant environments are still relatively new and have not undergone extensive testing. As well, few objective measures exist that aim to predict the accuracy of a localization task. This chapter summarizes the work done and provides direction for future work.

6.1 Contributions

The main contributions of this thesis are outlined as follows:

- Objective and subjective testing of speech processing algorithms require a wide variety of conditions for accurate assessment. Room simulation models suffer from large computational time and in some cases do not reproduce realistic

reverberation conditions. Therefore, a binaural room impulse response database consisting of four reverberant conditions and one anechoic condition was developed. Each condition is accurately characterized by the RT_{60}. Measurements were taken at 16 angles, allowing for use in generating stimuli for localization tasks or investigating the effects of azimuth on speech perception. Using this database, generation of real room sounds complete with head shadowing and inter-aural time delays becomes computationally insignificant. The flexibility that simulation methods provide can be achieved by combining BRIR databases.

- The YM method, a single-channel dereverberation algorithm was implemented and examined. It was demonstrated that when implemented bilaterally, the YM method disturbs the ILD cue used in azimuth estimation. Several ways of implementing the YM in a binaural sense were investigated. Using a delay-and-sum beamformer to generate a single reference channel from which weights are calculated proved to be the most effective binaural implementation.
- Investigation of the YM method led to the conclusion that the suggested parameters did not function well for a wide variety of conditions. To improve the robustness of the YM method, an adaptive weighting parameter was implemented based on the moving average of the smoothed entropy function. This parameter is shown to improve the effectiveness of the YM method for a much wider range of speech stimuli.
- Speech quality was objectively and subjectively evaluated using the SRMR metric and a speech quality study. All tested algorithms show significant increase in SRMR over the YM method. For a RT_{60} of less than 0.55 seconds, the BMYM method offers greater improvement in the SRMR than the Two-Stage method. The subjective study demonstrated that normal hearing listeners are influenced highly by processing distortions, ranking the unprocessed condition as the most similar to an anechoic sample. The Δ SRMR was found to have moderate correlation ($r = 0.46$) with the Δ MUSHRA score.
- Localization performance was objective and subjectively evaluated by calculating the mean Δ ILD and Δ ILD variance and a localization task. The subjective study showed no difference in localization accuracy between the unprocessed and

binaural conditions, indicating that they effectively preserve the binaural cues used in azimuth estimation. The YM demonstrated significantly more error at angles around 0 degrees (±22.5° and ±45°). This is thought to be the result of conflicting location estimates from the disturbed ILD with those of the undisturbed ITD.

6.2 Future work

This thesis accomplished the development of a BRIR database, a binaural dereverberation algorithm and the objective and subjective evaluation of sound quality and localization in reverberation. The results yield insight into how evaluation of binaural speech enhancement systems can be improved in the future. As well, future improvements to the BRIR database and BMYM method are discussed.

The BRIR database developed in this thesis added RIRs at higher reverberation times (0.9s – 1.3s) to the available pool of impulse responses. However, reverberation cannot be completely characterized by the RT_{60}. Differently shaped and different wall materials can change the perceived nature of the reverberation while maintaining a constant RT_{60}. It is therefore desirable to extend the NCA database into lower reverberation times comparable with the AIR database. This way, differences in the colouration, and its effect on sound quality, between two equally reverberant rooms can be analyzed. As well, the NCA database can be enhanced through greater angular discrimination.

The BMYM method shows significant improvement over the YM method in terms of sound quality and preservation of localization cues by modifying the gross weight function. However, the idea of having a gross/fine weight function can be extended easily using the variable mapping parameter. By increasing/decreasing the length of the moving average of the entropy function, multiple weighting functions can be implemented to form a composite function for overall improved performance. As well, the Two-Stage method demonstrates the effectiveness of using dereverberation blocks to target early and late reverberation. This idea also follows the natural extension of including additional blocks to target specific characteristics of the offending reverberation. Composite

methods will likely prove popular and more effective at robust dereverberation in the near future.

Since the BMYM method is blind, it is theoretically possible to implement into a portable system. However, the binaural nature of the BMYM method may be hindered by bandwidth limitations on portable devices. Further work is necessary to determine these limitations and how the binaural transfer can be implemented efficiently.

Sound quality was investigated for normal hearing listeners who were found to be largely influenced by processing distortions. A contributing factor to this is that normal hearing listeners likely do not have difficulty hearing in most reverberant conditions. However, hearing impaired subjects may be more robust to processing distortions in exchange for the improved intelligibility dereverberation should offer. Thus, this study could be extended to include hearing-impaired subjects and evaluate the differences between their scores and those of normal hearing listeners. As well, an extension to this study with speech processed in different reverberant conditions could provide insight as to the relationship between the SRMR and subjective scores based on processing distortions.

The localization task demonstrated that the ILD measures used are not good indicators of azimuth estimation for bilateral dereverberation. However, it may be possible to develop a more accurate measure using the change in ILD and ITD and the biasing weights reported in [11].

Works Cited

[1] A.K. Nabelek and JM Pickett, "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners," *Journal of Speech, Language and Hearing Research*, vol. 17, no. 4, 1974.

[2] William A. Yost, *Fundamentals of Hearing: An Introduction*, 5th ed.: Elsevier, 2007.

[3] A.K. Nabelek and D. Mason, "Effect of noise and reverberation on binaural and monoaural word identification by subjects with various audiograms.," *Journal of Speech, Language, and Hearing Research*, vol. 24, pp. 375-383, 1981.

[4] A.K. Nabelek, "Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss," *J. Acoust. Soc. Am.*, pp. 476-484, 1988.

[5] A.K. Nabelek and P. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages.," *J. Acoust. Soc. Am.*, vol. 71, no. 5, pp. 1242-1248, 1982.

[6] Jennifer J Lister, Dashielle M Febo, Joan M Besing, Harvey B Abrams Catherine L Rogers, "Effects of bilingualism , noise , and reverberation on speech perception by listeners with normal hearing," *Applied PsychoLinguistics*, vol. 27, no. 3, pp. 465-485, 2006.

[7] C. Giguere and S.M. Abel, "Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay," *J. Acoust. Soc. Am.*, vol. 94, no. 2, pp. 769-776, 1993.

[8] Brian F. G. Katz, "Acoustic absorption measurement of human hair and skin within the audible frequency range," *J. Acoust. Soc. Am.*, vol. 108, no. 5, pp. 2238-2242, 2000.

[9] S. S. Stevens and E. B. Newman, "The localization of actual sources of sound," *The American Journal of Psychology*, vol. 48, no. 2, pp. 297-306, Apr. 1936.

[10] W. E. Feddersen, T. T. Sandel, D. C. Teas, and L. A. Jeffress, "Localization of High-Frequency Tones," *J. Acoust. Soc. Am.*, vol. 29, no. 9, pp. 988-991, Sept. 1957.

[11] Ewan A. Macpherson and John C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.*, vol. 111, no. 5, 2002.

[12] E.M. Wenzel, "The relative contribution of interaural time and magnitude cues to dynamic sound localization," in *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*, 1995, pp. 80-83.

[13] Hans Wallach, Edwin B. Newman, and Mark R. Rosenzweig, "The Precedence Effect in Sound Localization," *The American Journal of Psychology*, vol. 62, no. 3, pp. 315-336, July 1949.

[14] W.M. Hartmann, "Localization of sound in rooms," *J. Acoust. Soc. Am.*, vol. 74, no. 5, pp. 1380-1391, 1983.

[15] Schuyler R. Quackenbush, Thomas P. Barnwell III, and Mark A. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs: Prentice Hall, 1988.

[16] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proceedings of the 16th international conference on Digital Signal Processing*, 2009, pp. 550-554.

[17] B. Yegnanarayana and P.S. Murthy, "Enhancement of reverberant speech using LP residual signal," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 267-281, 2000.

[18] M. Jeub, M. Schafer, T. Esch, and P. Vary, "Model-Based Dereverberation Preserving Binaural Cues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1732-1745, Sept. 2010.

[19] N. Kitawaki, H. Nagabuchi, and K. Itoh, "Objective quality evaluation for low-bit-rate speech coding systems," *Selected Areas in Communications, IEEE Journal on*, vol. 6, no. 2, pp. 242-248, 1988.

[20] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 5, pp. 819-829, 1992.

[21] A.W. Rix, J.G. Beerends, M.P. Hollier, and A.P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 2-5, 2001.

[22] Yi and Loizou, P.C. Hu, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio Speech and Language Processing*, vol. 16, no. 1, pp. 229 - 238 , 2008.

[23] J.H.L. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *ICSLP*, Sydney, Australia, 1998, pp. 2819-2822.

[24] F. Itakura and S. Saito, "An analysis-synthesis telephony based on maximum likelihood method," in *Proc. Int. Congr. Acoust.*, Tokyo, Japan, 1968, pp. C-5-5.

[25] Dennis H. Klatt, "Prediction of perceived phonetic distance from critcal-band spectra: A first step," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1982, pp. 1278-1281.

[26] J.Y.C. Wen, N.D. Gaubitch, E.A.P. Habets, T Myatt, and P.A. Naylor, in *Proc. Int. Workshop Acoust. Echo Noise Control*, Paris, 2006, pp. 1-4.

[27] TH Falk and Chenxi Zheng, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, vol. 18, no. 7, pp. 1766-1774, 2010.

[28] Christof Faller and Juha Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Am.*, vol. 116, no. 5, pp. 3075-3089, 2004.

[29] Slaney M., Auditory toolbox, 1998.

[30] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice, "Spiral vos final report, part a: The auditory filterbank," Cambridge Electronic Design, Cambridge, U.K., Internal Report 1988.

[31] B. Glasberg and B. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, pp. 103-138, 1990.

[32] Y.A. Huang and J. Benesty, "Adaptive blind channel identification: Multi-channel least mean square and Newton algorithms," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2, 2002, pp. II-1637-II-1640.

[33] T Aboulnasr and K Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1421-1424, 1999.

[34] K Kokkinakis and P.C. Loizou, "Selective-Tap Blind Dereverberation for Two-Microphone Enhancement of Reverberant Speech," *IEEE signal processing letters*, vol. 16, no. 11, p. 961, 2009.

[35] Yonggang Zhang, J.A. Chambers, Saeid Sanei, Paul Kendrick, and T.J. Cox, "A new variable tap-length LMS algorithm to model an exponential decay impulse response," *Signal Processing Letters, IEEE*, vol. 14, no. 4, pp. 263-266, 2007.

[36] G. Tong Zhou, "A variable step size and variable tap length LMS algorithm for impulse responses with exponential power profile," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 3105-3108.

[37] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 2, pp. 145-152, 1988.

[38] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943-950, 1979.

[39] B Yegnanarayana, SR Prasanna, and K.S. Rao, "Speech enhancement using excitation source information," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on* , 2002, pp. I-542.

[40] B.W. Gillespie, H.S. Malvar, and D.A.F. Florencio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," , vol. 6, 2001.

[41] S Mosayyebpour, A Sayyadiyan, M Zareian, and A Shahbazi, "Single Channel Inverse Filtering of Room Impulse Response by Maximizing Skewness of LP Residual," in *International Conference on Signal Acquisition and Processing*, 2010, pp. 130-134.

[42] Mingyang Wu and D.L. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 3, pp. 774-784, 2006.

[43] Mingyang Wu and D.L. Wang, "A one-microphone algorithm for reverberant speech enhancement," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, 2003, pp. I-892-I-895.

[44] T. Nakatani and M. Miyoshi, "Blind dereverberation of single channel speech signal based on harmonic structure," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, 2003, pp. I-92-5.

[45] Rainer Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on* , 1988, pp. 2578-2581.

[46] I.a. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 709-716, 2003.

[47] R.K. Cook, RV Waterhouse, RD Berendt, S. Edelman, and MC Thompson Jr, "Measurement of correlation coefficients in reverberant sound fields," *The Journal of the Acoustical Society of America*, vol. 27, p. 1072, 1955.

[48] M. Jeub and P. Vary, "Binaural dereverberation based on a dual-channel Wiener filter with optimized noise field coherence," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* , 2010, pp. 4710-4713.

[49] T J Klasen, S Doclo, T Van Den Bogaert, M Moonen, and J Wouters, "Binaural Multi-Channel Wiener Filtering for Hearing Aids: Preserving Interaural Time and Level Differences," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on* , vol. 5, 2006, p. V.

[50] T V den Bogaert, J Wouters, S Doclo, and M Moonen, "Binaural Cue Preservation for Hearing Aids using an Interaural Transfer Function Multichannel Wiener Filter," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2007, pp. IV-565 - IV-568.

[51] Richard L. Freyman and UmaHelfer, Karen S. Balakrishnan, "Spatial release from informational masking in speech recognition," *The Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2112-2122, 2001.

[52] J. Rindel, "The use of computer modeling in room acoustics," *Journal of Vibroengineering*, vol. 3, no. 4, pp. 41–72, 2000.

[53] Jens Holger Rindel and Claus Lynge Christensen, "Room acoustic simulatio and auralization - How close are we to the real room?," in *The Eighth Western Pacific Acoustics Conference*, Melbourne, 2003.

[54] Edgar J. Berdahl and Julius O. Smith III, REALSIMPLE Project, 2008.

[55] M. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am*, vol. 24, no. 4, pp. 320-327, 1965.

[56] ANSI S1.1-1986 (ASA 65-1986): Specifications for Octave-Band and Fractional-Octave-Band Analog and Digital Filters, 1993.

[57] C. Knapp and G Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 4, pp. 320-327, 1976.

[58] John S. Garofolo et al., TIMIT Acoustic-Phonetic Continuous Speech Corpus, 1993.

[59] ITU-R recommendation BS.1534-1, 2003.

[60] Y Benjamini, D Drai, G Elmer, N Kafkafi, and I Golani, "Controlling the false discovery rate in behavior genetics research," *Behavioural brain research*, vol. 125, no. 1-2, pp. 279-284, 2001.

[61] J. Middlebrooks and D. Green, "Sound localization by human listeners," *Annual review of psychology*, vol. 42, pp. 135-159, 1991.

Appendix A: Reverberation and Anechoic Chamber Setup

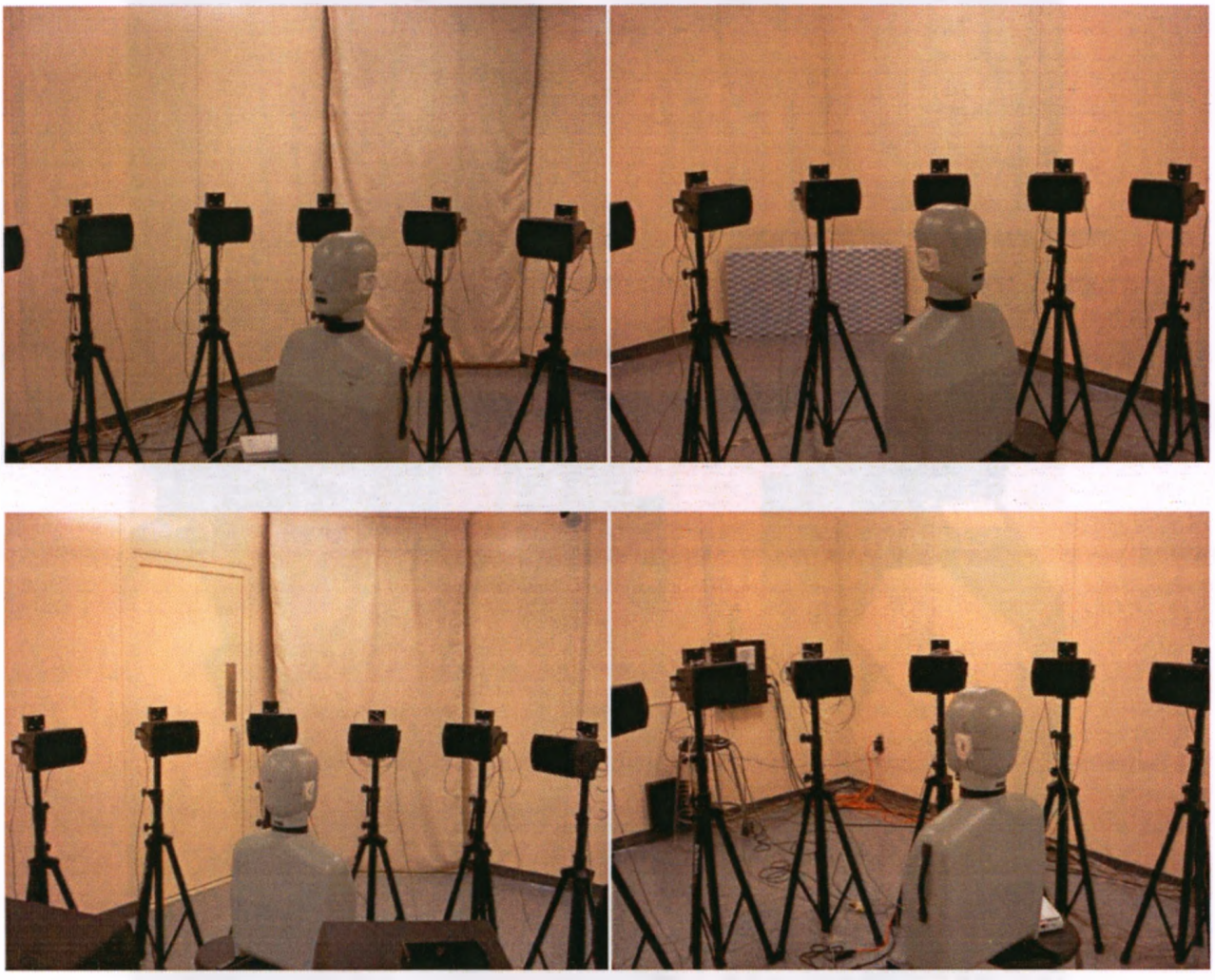


Figure A-1: The HATS positioned on a stool in the centre of the reverberation chamber, surrounded by the speaker array.

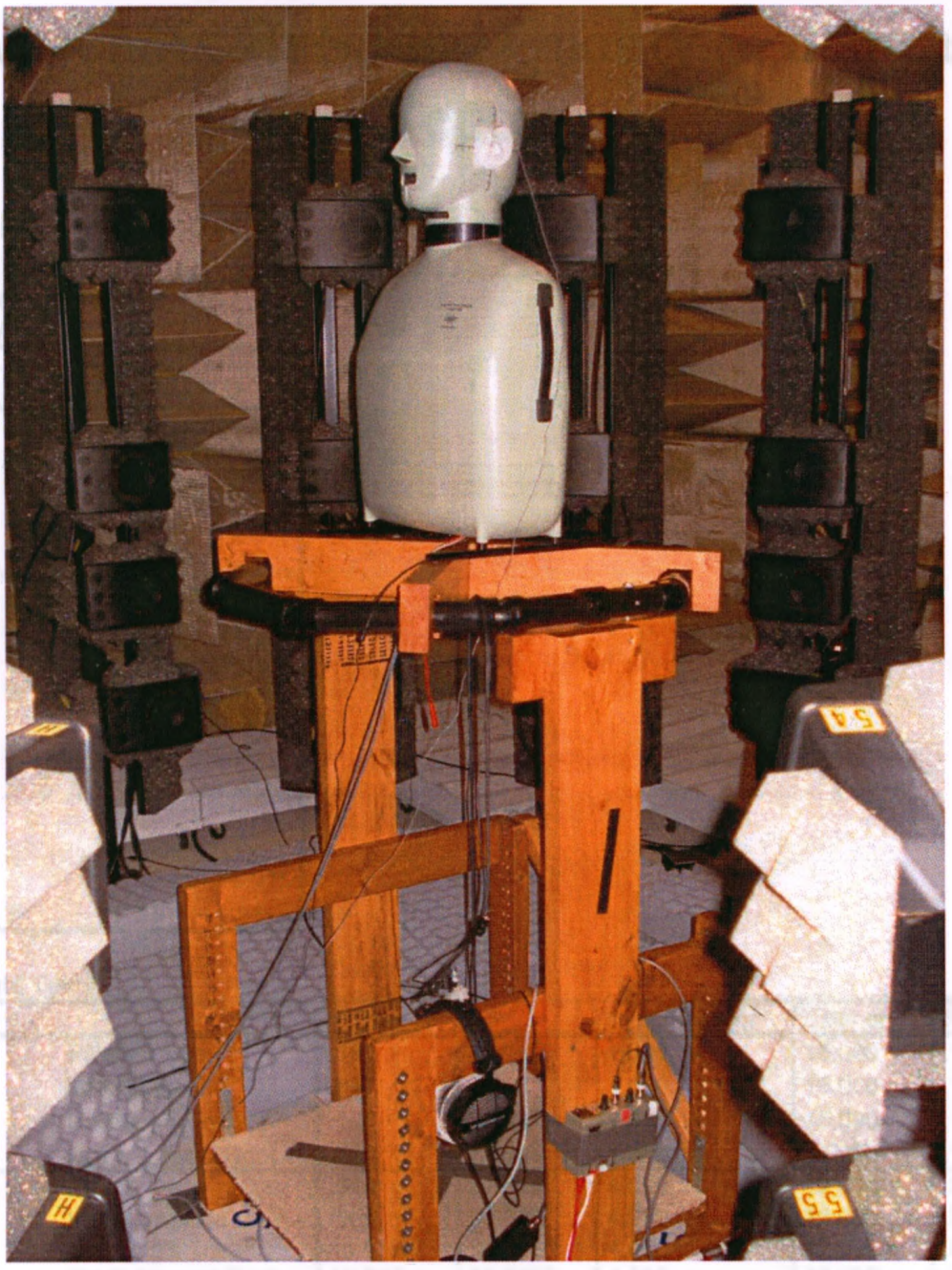


Figure A-2: The HATS in the anechoic chamber, shown with digital hearing aids.

Appendix B: Sound Quality Statistical Report

General Linear Model

Within-Subjects Factors
Measure:MEASURE_1

gender	alg	Dependent Variable
1	1	m_reverb
	2	m_bilatLP
	3	m_binaurLP
	4	m_SS
	5	m_SSW
2	1	f_reverb
	2	f_bilatLP
	3	f_binaurLP
	4	f_SS
	5	f_SSW

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
gender	Pillai's Trace	.153	2.708[a]	1.000	15.000	.121
	Wilks' Lambda	.847	2.708[a]	1.000	15.000	.121
	Hotelling's Trace	.181	2.708[a]	1.000	15.000	.121
	Roy's Largest Root	.181	2.708[a]	1.000	15.000	.121
alg	Pillai's Trace	.914	31.835[a]	4.000	12.000	.000
	Wilks' Lambda	.086	31.835[a]	4.000	12.000	.000
	Hotelling's Trace	10.612	31.835[a]	4.000	12.000	.000
	Roy's Largest Root	10.612	31.835[a]	4.000	12.000	.000
gender * alg	Pillai's Trace	.470	2.658[a]	4.000	12.000	.085
	Wilks' Lambda	.530	2.658[a]	4.000	12.000	.085
	Hotelling's Trace	.886	2.658[a]	4.000	12.000	.085
	Roy's Largest Root	.886	2.658[a]	4.000	12.000	.085

a. Exact statistic
b. Design: Intercept
Within Subjects Design: gender + alg + gender * alg

Mauchly's Test of Sphericity[b]
Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
gender	1.000	.000	0	.
alg	.095	31.648	9	.000
gender * alg	.255	18.342	9	.033

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

b. Design: Intercept
Within Subjects Design: gender + alg + gender * alg

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Epsilon[a]		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
gender	1.000	1.000	1.000
alg	.499	.575	.250
gender * alg	.577	.687	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: gender + alg + gender * alg

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square
gender	Sphericity Assumed	81.796	1	81.796
	Greenhouse-Geisser	81.796	1.000	81.796
	Huynh-Feldt	81.796	1.000	81.796
	Lower-bound	81.796	1.000	81.796
Error(gender)	Sphericity Assumed	453.135	15	30.209
	Greenhouse-Geisser	453.135	15.000	30.209
	Huynh-Feldt	453.135	15.000	30.209
	Lower-bound	453.135	15.000	30.209
alg	Sphericity Assumed	25418.357	4	6354.589
	Greenhouse-Geisser	25418.357	1.995	12739.428
	Huynh-Feldt	25418.357	2.301	11046.624
	Lower-bound	25418.357	1.000	25418.357
Error(alg)	Sphericity Assumed	14170.354	60	236.173
	Greenhouse-Geisser	14170.354	29.929	473.469
	Huynh-Feldt	14170.354	34.515	410.555
	Lower-bound	14170.354	15.000	944.690
gender * alg	Sphericity Assumed	144.131	4	36.033
	Greenhouse-Geisser	144.131	2.306	62.492
	Huynh-Feldt	144.131	2.750	52.419
	Lower-bound	144.131	1.000	144.131
Error(gender*alg)	Sphericity Assumed	1058.506	60	17.642
	Greenhouse-Geisser	1058.506	34.596	30.596
	Huynh-Feldt	1058.506	41.244	25.665
	Lower-bound	1058.506	15.000	70.567

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source	F	Sig.

gender	Sphericity Assumed	2.708	.121
	Greenhouse-Geisser	2.708	.121
	Huynh-Feldt	2.708	.121
	Lower-bound	2.708	.121
alg	Sphericity Assumed	26.907	.000
	Greenhouse-Geisser	26.907	.000
	Huynh-Feldt	26.907	.000
	Lower-bound	26.907	.000
gender * alg	Sphericity Assumed	2.042	.100
	Greenhouse-Geisser	2.042	.139
	Huynh-Feldt	2.042	.127
	Lower-bound	2.042	.173

Tests of Within-Subjects Contrasts
Measure:MEASURE_1

Source		alg	Type III Sum of Squares	df	Mean Square
gender	Level 2 vs. Level 1		32.718	1	32.718
Error(gender)	Level 2 vs. Level 1		181.254	15	12.084
alg		Level 2 vs. Level 1	19701.631	1	19701.631
		Level 3 vs. Level 1	9312.250	1	9312.250
		Level 4 vs. Level 1	1557.782	1	1557.782
		Level 5 vs. Level 1	11718.063	1	11718.063
Error(alg)		Level 2 vs. Level 1	4486.977	15	299.132
		Level 3 vs. Level 1	3759.719	15	250.648
		Level 4 vs. Level 1	5864.452	15	390.963
		Level 5 vs. Level 1	8561.094	15	570.740
gender * alg	Level 2 vs. Level 1	Level 2 vs. Level 1	44.223	1	44.223
		Level 3 vs. Level 1	121.000	1	121.000
		Level 4 vs. Level 1	114.223	1	114.223
		Level 5 vs. Level 1	11.391	1	11.391
Error(gender*alg)	Level 2 vs. Level 1	Level 2 vs. Level 1	286.012	15	19.067
		Level 3 vs. Level 1	479.125	15	31.942
		Level 4 vs. Level 1	615.715	15	41.048
		Level 5 vs. Level 1	1919.734	15	127.982

Tests of Within-Subjects Contrasts
Measure:MEASURE_1

Source		alg	F	Sig.
gender	Level 2 vs. Level 1		2.708	.121
alg		Level 2 vs. Level 1	65.863	.000
		Level 3 vs. Level 1	37.153	.000
		Level 4 vs. Level 1	3.984	.064
		Level 5 vs. Level 1	20.531	.000
gender * alg	Level 2 vs. Level 1	Level 2 vs. Level 1	2.319	.149
		Level 3 vs. Level 1	3.788	.071

	Level 4 vs. Level 1	2.783	.116
	Level 5 vs. Level 1	.089	.770

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	54956.253	1	54956.253	1217.790	.000
Error	676.918	15	45.128		

Appendix C: Adjusted Azimuth Estimation Error Statistical Report

General Linear Model

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.
angle	Pillai's Trace	.792	3.817[a]	8.000	8.000	.038
	Wilks' Lambda	.208	3.817[a]	8.000	8.000	.038
	Hotelling's Trace	3.817	3.817[a]	8.000	8.000	.038
	Roy's Largest Root	3.817	3.817[a]	8.000	8.000	.038
alg	Pillai's Trace	.849	24.385[a]	3.000	13.000	.000
	Wilks' Lambda	.151	24.385[a]	3.000	13.000	.000
	Hotelling's Trace	5.627	24.385[a]	3.000	13.000	.000
	Roy's Largest Root	5.627	24.385[a]	3.000	13.000	.000
angle * alg	Pillai's Trace	.[b]	.	.	.	.
	Wilks' Lambda	.[b]	.	.	.	.
	Hotelling's Trace	.[b]	.	.	.	.
	Roy's Largest Root	.[b]	.	.	.	.

a. Exact statistic
b. Cannot produce multivariate test statistics because of insufficient residual degrees of freedom.
c. Design: Intercept
Within Subjects Design: angle + alg + angle * alg

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
angle	.011	54.377	35	.027
alg	.706	4.777	5	.445
angle * alg	.000	.	299	.

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

b. Design: Intercept
Within Subjects Design: angle + alg + angle * alg

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Epsilon[a]		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
angle	.557	.821	.125
alg	.822	.996	.333
angle * alg	.281	.534	.042

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.
a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Epsilon[a]		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
angle	.557	.821	.125
alg	.822	.996	.333
angle * alg	.281	.534	.042

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: angle + alg + angle * alg

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square
angle	Sphericity Assumed	1477.808	8	184.726
	Greenhouse-Geisser	1477.808	4.454	331.790
	Huynh-Feldt	1477.808	6.568	225.005
	Lower-bound	1477.808	1.000	1477.808
Error(angle)	Sphericity Assumed	8686.468	120	72.387
	Greenhouse-Geisser	8686.468	66.811	130.016
	Huynh-Feldt	8686.468	98.518	88.171
	Lower-bound	8686.468	15.000	579.098
alg	Sphericity Assumed	1529.339	3	509.780
	Greenhouse-Geisser	1529.339	2.467	620.032
	Huynh-Feldt	1529.339	2.989	511.624
	Lower-bound	1529.339	1.000	1529.339
Error(alg)	Sphericity Assumed	545.055	45	12.112
	Greenhouse-Geisser	545.055	36.998	14.732
	Huynh-Feldt	545.055	44.838	12.156
	Lower-bound	545.055	15.000	36.337
angle * alg	Sphericity Assumed	919.039	24	38.293
	Greenhouse-Geisser	919.039	6.741	136.341
	Huynh-Feldt	919.039	12.816	71.710
	Lower-bound	919.039	1.000	919.039
Error(angle*alg)	Sphericity Assumed	3329.516	360	9.249
	Greenhouse-Geisser	3329.516	101.111	32.929
	Huynh-Feldt	3329.516	192.242	17.319

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square
angle	Sphericity Assumed	1477.808	8	184.726
	Greenhouse-Geisser	1477.808	4.454	331.790
	Huynh-Feldt	1477.808	6.568	225.005
	Lower-bound	1477.808	1.000	1477.808
Error(angle)	Sphericity Assumed	8686.468	120	72.387
	Greenhouse-Geisser	8686.468	66.811	130.016
	Huynh-Feldt	8686.468	98.518	88.171
	Lower-bound	8686.468	15.000	579.098
alg	Sphericity Assumed	1529.339	3	509.780
	Greenhouse-Geisser	1529.339	2.467	620.032
	Huynh-Feldt	1529.339	2.989	511.624
	Lower-bound	1529.339	1.000	1529.339
Error(alg)	Sphericity Assumed	545.055	45	12.112
	Greenhouse-Geisser	545.055	36.998	14.732
	Huynh-Feldt	545.055	44.838	12.156
	Lower-bound	545.055	15.000	36.337
angle * alg	Sphericity Assumed	919.039	24	38.293
	Greenhouse-Geisser	919.039	6.741	136.341
	Huynh-Feldt	919.039	12.816	71.710
	Lower-bound	919.039	1.000	919.039
Error(angle*alg)	Sphericity Assumed	3329.516	360	9.249
	Greenhouse-Geisser	3329.516	101.111	32.929
	Huynh-Feldt	3329.516	192.242	17.319
	Lower-bound	3329.516	15.000	221.968

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		F	Sig.
angle	Sphericity Assumed	2.552	.013
	Greenhouse-Geisser	2.552	.041
	Huynh-Feldt	2.552	.021
	Lower-bound	2.552	.131
alg	Sphericity Assumed	42.088	.000
	Greenhouse-Geisser	42.088	.000
	Huynh-Feldt	42.088	.000
	Lower-bound	42.088	.000
angle * alg	Sphericity Assumed	4.140	.000
	Greenhouse-Geisser	4.140	.001
	Huynh-Feldt	4.140	.000
	Lower-bound	4.140	.060

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	alg		Type III Sum of Squares	df	Mean Square	F	Sig.
angle	Linear		13.144	1	13.144	.882	.362
	Quadratic		8.590	1	8.590	.353	.561
	Cubic		6.731	1	6.731	.187	.672
	Order 4		46.845	1	46.845	1.549	.232
	Order 5		70.796	1	70.796	5.837	.029
	Order 6		142.486	1	142.486	11.378	.004
	Order 7		.377	1	.377	.062	.807
	Order 8		80.484	1	80.484	9.516	.008
Error(angle)	Linear		223.490	15	14.899		
	Quadratic		365.203	15	24.347		
	Cubic		541.359	15	36.091		
	Order 4		453.519	15	30.235		
	Order 5		181.941	15	12.129		
	Order 6		187.838	15	12.523		
	Order 7		91.400	15	6.093		
	Order 8		126.867	15	8.458		
alg		Level 1 vs. Level 4	104.968	1	104.968	4.664	.047
		Level 2 vs. Level 4	2602.268	1	2602.268	76.174	.000
		Level 3 vs. Level 4	111.458	1	111.458	7.314	.016
Error(alg)		Level 1 vs. Level 4	337.606	15	22.507		
		Level 2 vs. Level 4	512.435	15	34.162		
		Level 3 vs. Level 4	228.575	15	15.238		
angle * alg	Linear	Level 1 vs. Level 4	8.754	1	8.754	.447	.514
		Level 2 vs. Level 4	55.290	1	55.290	1.713	.210
		Level 3 vs. Level 4	11.007	1	11.007	.741	.403
	Quadratic	Level 1 vs. Level 4	83.391	1	83.391	6.412	.023
		Level 2 vs. Level 4	777.442	1	777.442	27.935	.000
		Level 3 vs. Level 4	24.154	1	24.154	1.220	.287
	Cubic	Level 1 vs. Level 4	29.091	1	29.091	1.594	.226
		Level 2 vs. Level 4	127.336	1	127.336	4.237	.057
		Level 3 vs. Level 4	3.699	1	3.699	.295	.595

	Order 4	Level 1 vs. Level 4	4.912	1	4.912	.331	.574
		Level 2 vs. Level 4	.308	1	.308	.007	.935
		Level 3 vs. Level 4	16.244	1	16.244	1.877	.191
	Order 5	Level 1 vs. Level 4	15.810	1	15.810	1.066	.318
		Level 2 vs. Level 4	131.100	1	131.100	9.931	.007
		Level 3 vs. Level 4	.093	1	.093	.006	.939
	Order 6	Level 1 vs. Level 4	88.502	1	88.502	6.553	.022
		Level 2 vs. Level 4	358.362	1	358.362	27.491	.000
		Level 3 vs. Level 4	16.871	1	16.871	3.438	.083
	Order 7	Level 1 vs. Level 4	1.894	1	1.894	.120	.734
		Level 2 vs. Level 4	28.845	1	28.845	1.140	.303
		Level 3 vs. Level 4	3.076	1	3.076	.148	.706
	Order 8	Level 1 vs. Level 4	56.590	1	56.590	2.459	.138
		Level 2 vs. Level 4	.212	1	.212	.008	.929
		Level 3 vs. Level 4	7.750	1	7.750	.306	.588
Error(angle*alg)	Linear	Level 1 vs. Level 4	293.602	15	19.573		
		Level 2 vs. Level 4	484.077	15	32.272		
		Level 3 vs. Level 4	222.907	15	14.860		
	Quadratic	Level 1 vs. Level 4	195.092	15	13.006		
		Level 2 vs. Level 4	417.451	15	27.830		
		Level 3 vs. Level 4	296.968	15	19.798		
	Cubic	Level 1 vs. Level 4	273.776	15	18.252		
		Level 2 vs. Level 4	450.782	15	30.052		
		Level 3 vs. Level 4	188.218	15	12.548		
	Order 4	Level 1 vs. Level 4	222.852	15	14.857		
		Level 2 vs. Level 4	674.024	15	44.935		
		Level 3 vs. Level 4	129.832	15	8.655		
	Order 5	Level 1 vs. Level 4	222.487	15	14.832		
		Level 2 vs. Level 4	198.012	15	13.201		

		Level 3 vs. Level 4	232.163	15	15.478		
	Order 6	Level 1 vs. Level 4	202.576	15	13.505		
		Level 2 vs. Level 4	195.537	15	13.036		
		Level 3 vs. Level 4	73.615	15	4.908		
	Order 7	Level 1 vs. Level 4	236.715	15	15.781		
		Level 2 vs. Level 4	379.545	15	25.303		
		Level 3 vs. Level 4	312.107	15	20.807		
	Order 8	Level 1 vs. Level 4	345.132	15	23.009		
		Level 2 vs. Level 4	387.873	15	25.858		
		Level 3 vs. Level 4	379.483	15	25.299		

Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	20161.417	1	20161.417	248.248	.000
Error	1218.223	15	81.215		

Appendix D: Objective Localization Results

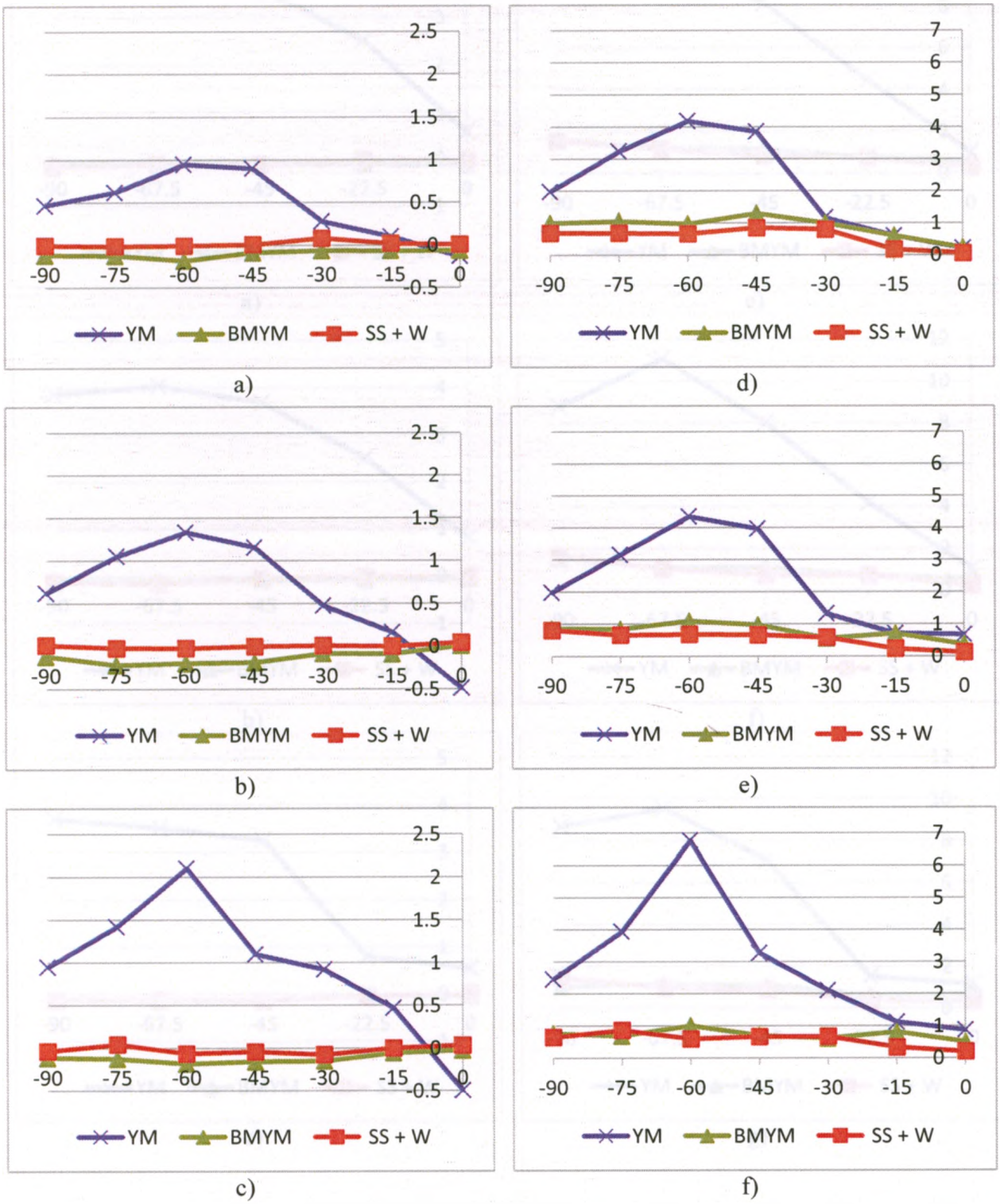
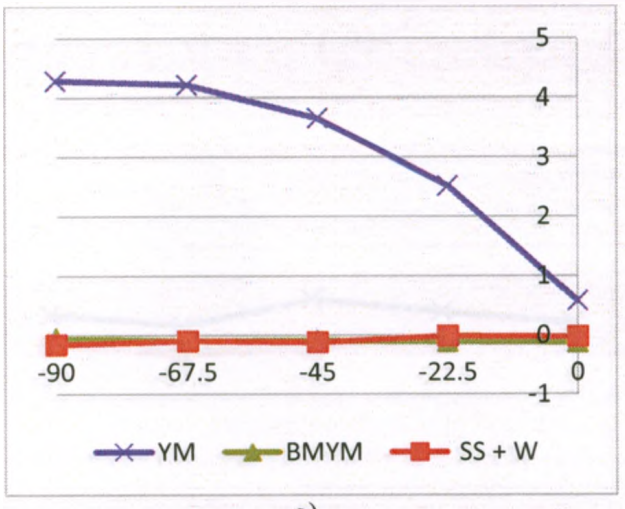
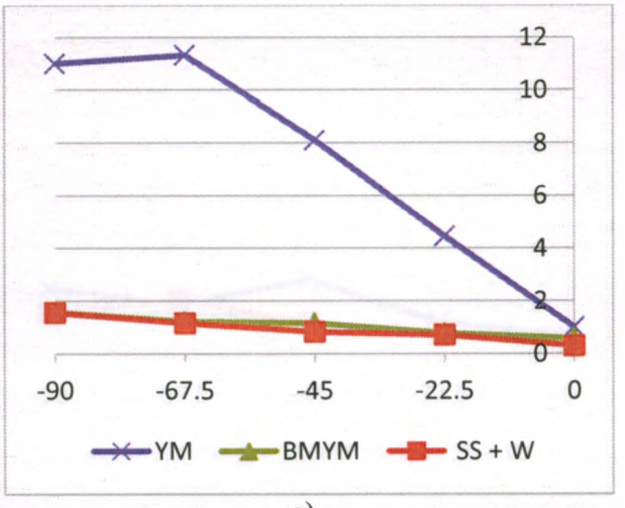


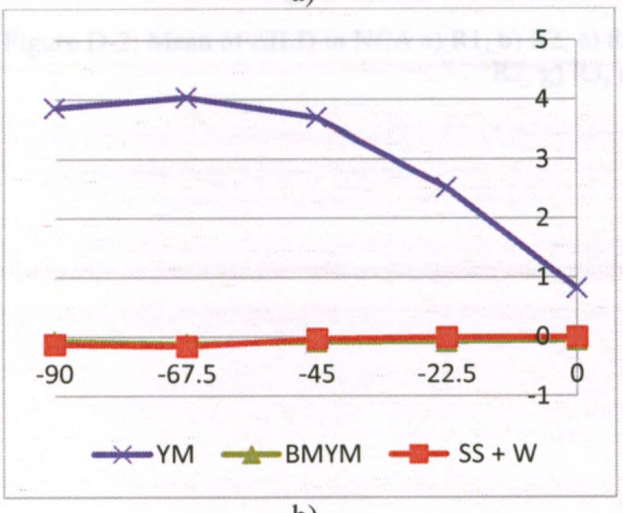
Figure D-1: Mean of ΔILD in AIR Stairwell a) 1, b) 2, and c) 3 and variance of ΔILD in AIR Stairwell d) 1, e) 2, f) 3.


5
4
3
2
1
0
-1
-90
-67.5
-45
-22.5
0
YM
BMYM
SS + W

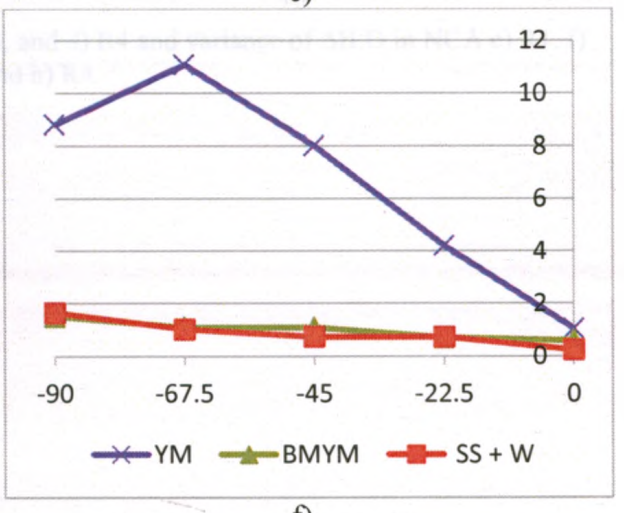
a)


12
10
8
6
4
2
0
-90
-67.5
-45
-22.5
0
YM
BMYM
SS + W

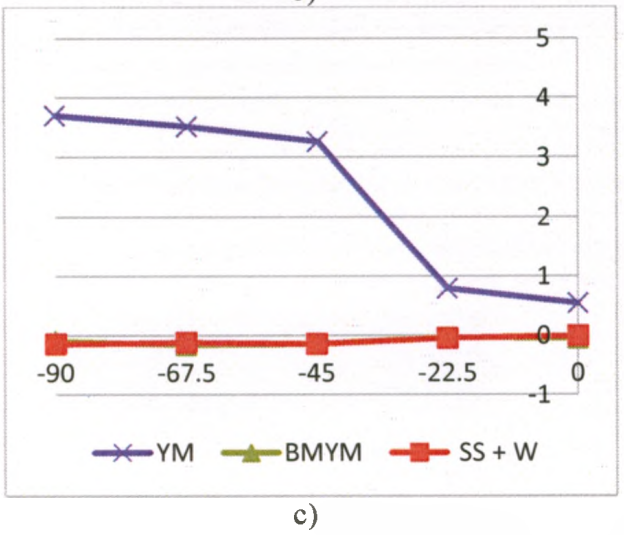
e)


5
4
3
2
1
0
-1
-90
-67.5
-45
-22.5
0
YM
BMYM
SS + W

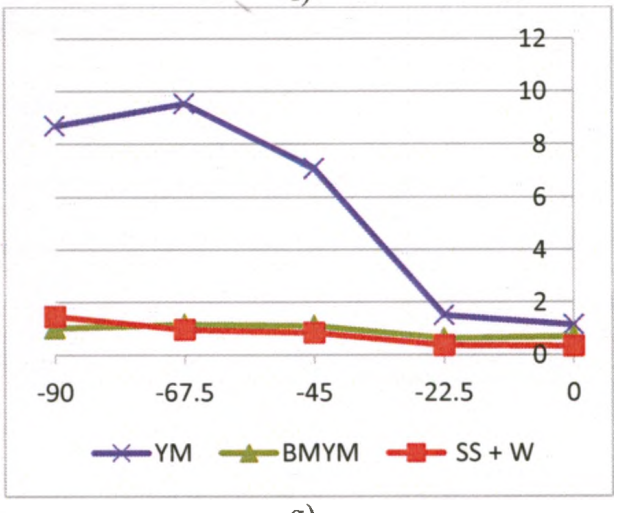
b)


12
10
8
6
4
2
0
-90
-67.5
-45
-22.5
0
YM
BMYM
SS + W

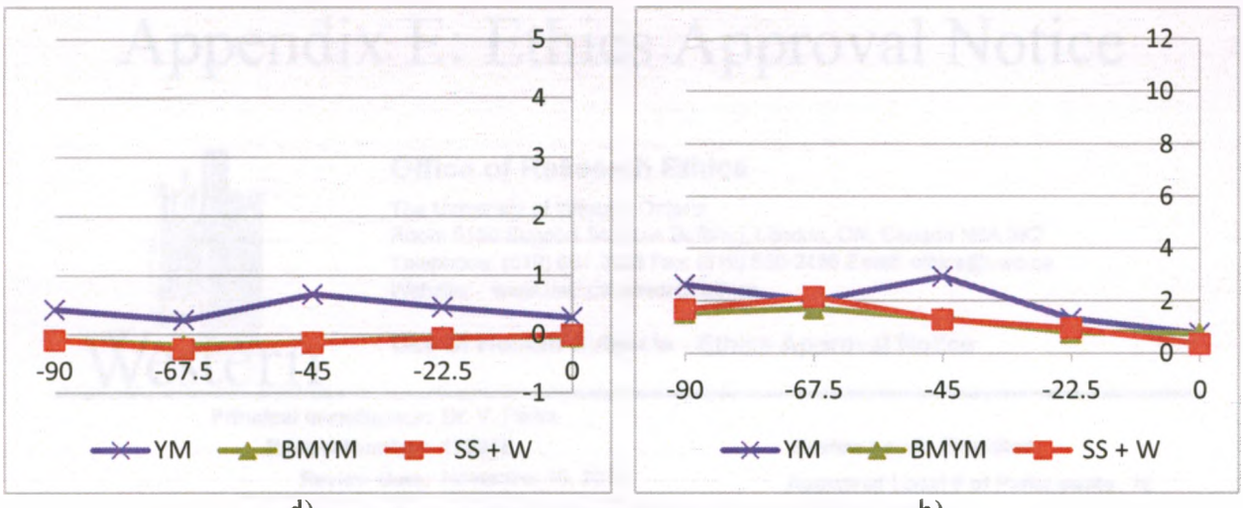
f)


5
4
3
2
1
0
-1
-90
-67.5
-45
-22.5
0
YM
BMYM
SS + W

c)


12
10
8
6
4
2
0
-90
-67.5
-45
-22.5
0
YM
BMYM
SS + W

g)



d) h)

Figure D-2: Mean of ΔILD in NCA a) R1, b) R2, c) R3, and d) R4 and variance of ΔILD in NCA e) R1, f) R2, g) R3, and h) R4.

Appendix E: Ethics Approval Notice

Office of Research Ethics
The University of Western Ontario
Room 5150 Support Services Building, London, ON, Canada N6A 3K7
Telephone: (519) 661-3036 Fax: (519) 850-2466 Email: ethics@uwo.ca
Website: www.uwo.ca/research/ethics

Western

Use of Human Subjects - Ethics Approval Notice

Principal Investigator: Dr. V. Parsa
Review Number: 17591E **Review Level:** Expedited
Review Date: November 10, 2010 **Approved Local # of Participants:** 16
Protocol Title: Subjective evaluation of binaural dereverberation algorithms
Department and Institution: Communication Sciences & Disorders, University of Western Ontario
Sponsor: NSERC-NATURAL SCIENCES ENGINEERING RSRCH COU
Ethics Approval Date: December 22, 2010 **Expiry Date:** September 30, 2011
Documents Reviewed and Approved: UWO Protocol, Letter of Information and Consent, Email
Documents Received for Information:

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced study on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:

a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

[redacted]

Chair of HSREB: Dr. Joseph Gilbert
FDA Reg. #: IRB 00000940

Ethics Officer to Contact for Further Information		
☐ Janice Sutherland (jsuther@uwo.ca)	☐ Elizabeth Wambolt (ewambolt@uwo.ca)	☒ Grace Kelly (grace.kelly@uwo.ca)

This is an official document. Please retain the original in your files.

cc: ORE File