OPTIMIZING ACOUSTICAL CONDITIONS FOR SPEECH INTELLIGIBILITY IN CLASSROOMS

by

WONYOUNG YANG

B.Sc., Hanyang University, Korea, 1998
M.Sc., The Pennsylvania State University, USA, 2002

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Occupational and Environmental Hygiene)

THE UNIVERSITY OF BRITISH COLUMBIA

October 2006

ABSTRACT

High speech intelligibility is imperative in classrooms where verbal communication is critical. However, the optimal acoustical conditions to achieve a high degree of speech intelligibility have previously been investigated with inconsistent results, and practical room-acoustical solutions to optimize the acoustical conditions for speech intelligibility have not been developed. This experimental study validated auralization for speech-intelligibility testing, investigated the optimal reverberation for speech intelligibility for both normal and hearing-impaired listeners using more realistic room-acoustical models, and proposed an optimal sound-control design for speech intelligibility based on the findings.

The auralization technique was used to perform subjective speech-intelligibility tests. The validation study, comparing auralization results with those of real classroom speech-intelligibility tests, found that if the room to be auralized is not very absorptive or noisy, speech-intelligibility tests using auralization are valid. The speech-intelligibility tests were done in two different auralized sound fields — approximately diffuse and non-diffuse — using the Modified Rhyme Test and both normal and hearing-impaired listeners. A hybrid room-acoustical prediction program was used throughout the work, and it and a 1/8 scale-model classroom were used to evaluate the effects of ceiling barriers and reflectors.

For both subject groups, in approximately diffuse sound fields, when the speech source was closer to the listener than the noise source, the optimal reverberation time was zero. When the noise source was closer to the listener than the speech source, the optimal reverberation time was 0.4 s (with another peak at 0.0 s) with relative output power levels of the speech and noise sources $SNS = 5$ dB, and 0.8 s with $SNS = 0$ dB. In non-diffuse sound fields, when the noise source was between the speaker and the listener, the optimal reverberation time was 0.6 s with $SNS = 4$ dB and increased to 0.8 and 1.2 s with decreased $SNS = 0$ dB, for both normal and hearing-impaired listeners. Hearing-impaired listeners required more early energy than normal-hearing listeners. Reflective ceiling barriers and ceiling reflectors — in particular, parallel front-back rows of semi-circular reflectors — achieved the goal of decreasing reverberation with the least speech-level reduction.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

)

LIST OF SYMBOLS

α	Sound absorption coefficient
$\bar{\alpha}$	Room average sound absorption coefficient
A	Absorptive ceiling barrier
C	Early-to-late sound energy ratio (clarity factor)
CC	Semicircular ceiling reflector with curved side down
CF	Semicircular ceiling reflector with flat side down
d	Sound diffusion coefficient
$\bar{d}$	Room average sound diffusion coefficient
E_d	Direct energy of sound signal
E_e	Early-arriving reflected energy of sound signal
E_l	Late-arriving energy of sound signal
E_n	Noise energy
HoH	Hearing impaired
L	Listener
L_{eq}	Equivalent sound pressure level
L_{nfl}	Long-term anechoic levels at 1 m directly in front of the noise sources in a free field
L_{sfl}	Long-term anechoic levels at 1 m directly in front of the speech in a free field
L_w	Sound power level
N	Noise source
q	Sound directivity
R	Reflective ceiling barrier
r1C	Receiver at front in centre line
r1L	Receiver at front in left-side line
r1R	Receiver at front in right-side line
R^2	Coefficients of determination
r2C	Receiver at middle in centre line
r2L	Receiver at middle in left-side line
r2R	Receiver at middle in right-side line
r3C	Receiver at back in centre line
r3L	Receiver at back in left-side line
r3R	Receiver at back in right-side line
r_n	Distance from the noise source to the listener

r_s	Distance from the speech source to the listener
r_{h_t}	Reverberation radius (critical distance)
S	Speech source
SNR	Speech-to-noise level difference at receiver
SNS	Relative output power levels of the speech and noise sources
U	Useful-to-detrimental sound energy ratio

LIST OF ABBREVIATIONS

%ALcons	Articulation Loss for Consonants
AI	Articulation Index
ANOVA	Analysis of Variance
BRIR	Binaural Room Impulse Responses
CWH	Central West Health
EDT	Early Decay Time
FRT	Fairbanks Rhyme Test
FS	Full-Scale value
HL	Hearing Loss
HRTF	Head Related Transfer Function
JND	Just Noticeable Differences
KEMAR	Knowles Electronics Mannequin for Acoustics Research
MLS	Maximum Length Sequences
MLSSA	Maximum Length Sequence System Analyser
MRT	Modified Rhyme Test
mSTI	modified Speech Transmission Index
MTF	Modulation Transfer Function
NH	Normal Hearing
NLA	A-weighted Noise Level
RASTI	Rapid Speech Transmission Index
RT	Reverberation Time
SAI	Speech Audibility Index
SD	Standard Deviation
SI	Speech Intelligibility
SII	Speech Intelligibility Index
SLA	A-weighted Speech Level
SLA_N	A-weighted Speech Level for a 'Normal' voice
SL_N	unweighted Speech Level for a 'Normal' voice
STI	Speech Transmission Index
UBC	University of British Columbia
FEM	Finite Element Method
BEM	Boundary Element Method

FDTD	Finite-Difference Time-Domain method
SEA	Statistical Energy Analysis
RTC	Randomized Tail-corrected Cone-tracing

ACKNOWLEDGMENTS

It has been four years since I came to Vancouver to study. At the moment I looked at the aerial photos of Vancouver and UBC, which my supervisor, Murray, showed me at an acoustical conference in Pittsburgh, I realized that I was going to fall in love with this beautiful city. Last but not least, my deep gratitude goes to this wonderful city, Vancouver, which inspired the beauty of nature and the preciousness of life into me.

October 2006

Wonyoung Yang

DEDICATION

To Moon Jung in the memory of your sacrifice and love.

1 INTRODUCTION

1.1 Background

People now spend the majority of their time in indoor environments, so appropriate room acoustics are imperative for human health, productivity, and comfort. It is particularly important in acoustically sensitive spaces such as classrooms, where verbal communication is critical. The acoustics of such rooms should achieve a high degree of speech intelligibility for listeners. Speech signals should be transmitted without distortion and be clearly audible.

Since subjective speech-intelligibility tests were proposed for evaluation of telephonic intelligibility [1,2,3], speech intelligibility has been considered to be important [4] in a wide variety of different fields. In speech science, speech intelligibility is used for clinical purposes [5]; in audiology, speech intelligibility is important in developing and evaluating hearing-aids [6]; in classroom acoustics, speech intelligibility is critical to improve conditions for teaching and learning ability [7].

In this thesis, speech intelligibility will be studied in relation to room-acoustical parameters. Speech intelligibility can be defined as the percentage of speech material correctly recognized by a listener, and it can be measured by specially designed speech-intelligibility tests or predicted by way of speech-intelligibility metrics. Speech-intelligibility metrics have been developed based on fundamental room-acoustical parameters. The room-acoustical parameters affecting speech intelligibility are known to be speech-to-noise level difference (often referred to in the literature as the signal-to-noise ratio) and reverberation. In general, speech intelligibility tends to increase with increased speech-to-noise level difference and to decrease with increased reverberation. The signal-to-noise level difference is considered to be the more dominant factor for controlling speech intelligibility in rooms [8]. However, in rooms the situation is complicated by the fact that reverberation and steady-state levels interact. Increased reverberation increases speech and noise levels by increasing the

reverberant sound energy; this increases or decreases the speech-to-noise level difference at a listener position depending on the listener's relative distances to the speech and noise sources [9]. In order to achieve a high degree of speech intelligibility, the room-acoustical parameters – speech-to-noise level difference and reverberation – should not be maximized or minimized but be optimized. How to do this is the topic of this thesis.

1.2   Literature Review

This section introduces fundamental concepts related to speech-intelligibility measures, which were used in this thesis (Section 1.2.1), reviews current issues related to optimal acoustical conditions for speech intelligibility and points out problems with the current experimental methods on speech-intelligibility studies (Section 1.2.2). This section also reviews feasibility and fundamentals of auralization for speech-intelligibility test (Section 1.2.3), and provides methods chosen to evaluate for this work (Section 1.2.4).

1.2.1   Speech-intelligibility measures

Subjective speech-intelligibility tests

Subjective speech-intelligibility tests measure human ability to recognize speech signals. These tests present test speech materials to appropriate listeners who record what they hear, the accuracy of their perceptions is scored. In general, the choice of the test is related to the purpose of the study. There are good overviews focused on the assessment of subjective speech-intelligibility tests [10,11]. In this section, the birth of speech-intelligibility tests and their application in room acoustics are discussed.

Speech-intelligibility tests frequently used in room acoustics are phoneme-level rhyme tests. Black [12] developed the first closed-set discrimination tests. Fairbanks [13] developed the Fairbanks' Rhyme Test (FRT), using the lists of rhyming monosyllabic words guided by the data of Thorndike and Lorge [14]. Modifications of Fairbanks' Rhyme Test

were made by House *et al.* [15] and became known as the Modified Rhyme Test (MRT). Kreul *et al.* [16] altered the MRT to make it more clinically useful. Adaptations were made in the areas of timing consistency, carrier phrase, test instructions, and forms used, with both male and female talkers. Further revision of the MRT was made by Griffiths [17], who added new items to examine phonemic confusion in the responses. The Fairbanks' Rhyme Test and the Modified Rhyme Test have been widely used: for example, Latham [18] used the FRT to validate the useful-to-detrimental sound-energy ratio; Bradley [19 20 8] has used the FRT in his studies of speech intelligibility; Nabelek *et al.* [21,22,23] have conducted the MRT in studies of speech perception.

Objective speech-intelligibility metrics

Subjective speech-intelligibility tests are by far the most accurate and reliable methods for intelligibility testing regarding subjective perception [24]. However, subjective speech-intelligibility tests are complicated to set up, time-consuming to conduct, and require extensive statistical analysis to interpret. Hence, various objective metrics have been developed to predict intelligibility scores.

One of the most prominent early efforts in speech intelligibility prediction was the development of the Articulation Index (AI) by French and Steinberg [25]; it can be computed from the intensities of speech and noise received by the ear, both as a function of frequency. The AI was developed as an acoustical metric that could be used to predict the speech recognition ability of speech-transmission systems. The method was reconsidered by Kryter [26,27] who increased its accessibility by the introduction of a calculation scheme, work sheets, and tables. Later, Peutz [28] published a new method for predicting speech intelligibility – the articulation loss for consonants (%ALcons) – which is computed from measurements of the direct-to-reverberant ratio, signal-to-noise ratio and the reverberation time. Later, the Speech Transmission Index (STI) was developed using the Modulation Transfer Function (MTF) of a transmission channel [29,30,31,32,33]. The STI combines two major phenomena that affect speech intelligibility, reverberation and noise, to extract a single index that gives good correlation with subjective perception. The Rapid Speech Transmission Index (RASTI) was developed as a simplified version of the STI [34]. The RASTI measures

in only two octave bands centered at 500 Hz and 2 kHz. The Speech Intelligibility Index (SII) [35] is a modified version of the AI and includes reverberation, noise and distortion, all of which are accounted for in the modulation transfer function. Recently, Speech Audibility Index (SAI) [36] was defined as the proportion of the useful speech signal (direct speech and early reverberation) that is above the level of the effective noise (actual noise and late reverberation). The SAI is similar to the STI in that it accounts for both noise and reverberation in terms of changes in the amplitude envelope of speech.

Speech intelligibility can be predicted from impulse responses. Early-arriving sound reflections (reflections arriving within some early-energy time of the direct sound) make the direct sound louder, so that the integrated early-arriving reflections increase the intelligibility of speech [37]. Late-arriving reflections are not integrated with the direct sound and degrade speech intelligibility by causing one speech sound to blur into the next. The ratio of early-to-late arriving sound has been proposed as an indicator of the effects of room acoustics (*i.e.* reverberation) on the clarity and intelligibility of speech [38]. Lochner and Burger [39] developed their Signal-to-Noise Ratio which provided a measure of useful and detrimental reflected speech energy according to the integration and masking characteristics of hearing by adding a background-noise component to the early-to-late arriving sound-ratio concept. The useful-to-detrimental ratio was extended by Latham [18] to account for the effect of fluctuating ambient background noise on speech intelligibility. The useful-to-detrimental ratio is considered to be a suitable intelligibility metric for predicting speech intelligibility for normal-hearing listeners [19,20]. Typically, the time limit for the early-arriving reflections has been taken to be 50 ms for speech sounds [8,20].

1.2.2   Effects of noise and reverberation on speech intelligibility

Experimental studies

Experimental studies have been used to examine speech intelligibility in different acoustical conditions which are combinations of various signal-to-noise ratios and reverberation times. Nabelek and Robinson [23] found that speech intelligibility was inversely proportional to the reverberation time in the absence of noise, but their results did not explain

4

what happens when noise sources are introduced. Nabelek and Pickett [21,22] found that speech intelligibility decreases with increased reverberation time. They only had two different conditions, with reverberation times of 0.3 s and 0.6 s. Finitzo-Hieber and Tillman [40] developed Nabelek and Pickett's tests, including a reverberation time of zero, and found that reverberation degraded speech discrimination for both normal and hearing-impaired children. Yacullo and Hawkins [41] also examined the effects of noise and reverberation time on monaural speech recognition using sentence materials, and concluded that speech intelligibility was adversely affected by noise and reverberation. Because of the fragmentary character of each study, it is difficult to draw systematic conclusions from these studies. However, these early experimental studies consistently found an optimal reverberation time of zero. They used fixed signal-to-noise ratios with different reverberation times at the listener's position, and put the speech and the noise at the same distance from the listener.

Experimental studies on speech intelligibility for hearing-impaired people have been done using subjective speech tests. Listeners with even mild sensorineural hearing loss may have greater difficulty when listening in noisy environments than do normal-hearing listeners [42,43,44,45]. Reverberation has a particularly detrimental effect on speech intelligibility for hearing-impaired listeners [46,47]. Nabelek and Dagenais [48] examined both noise and reverberation. Although the mean speech-test scores for the noise and reverberant conditions were not significantly different, the patterns of errors for these two conditions were different. They concluded that temporal smearing in reverberant condition should be considered. Duquesnoy and Plomp [49] investigated the applicability of the Speech Transmission Index (STI) to hearing-impaired subjects with added reverberation and noise. They proposed reduced reverberation times to improve communication for elderly people in rooms [50]. Humes *et al.* [51] found that the Articulation Index (AI) and the Speech Transmission Index (STI) had significant shortcomings for hearing-impaired listeners, and proposed the modified Speech Transmission Index (mSTI) as an alternative speech-intelligibility index for hearing-impaired listeners. Attempts to predict more accurate speech-recognition performance for sensorineural hearing-impaired listeners, several correction factors have been applied to the AI [52]. Payton, Uchanski, and Braida [53] used the AI and the STI; they found the AI was unable to represent the reduction in intelligibility scores due to reverberation for the hearing-impaired listeners. Neither Articulation Index (AI) nor Speech Intelligibility Index (SII) can

accurately predict the speech intelligibility observed in many hearing-impaired listeners [54,55,56].

In all of these experimental studies, noise was incorporated in an unrealistic manner. In reality, reverberation results in increased speech levels as well as noise levels [57]. In real rooms, the speech level at the listener's position is higher with reverberation, as is the noise level. In the presence of noise, it is important to recognize the spatial relationship between speaker and noise source, since where the noise sources are located differs their sound levels in reverberant rooms. Therefore, using the fixed signal-to-noise ratio at the listener in these experiments eliminated the accurate effect of reverberation on speech intelligibility.

Theoretical studies

Theoretical studies generally have predicted non-zero reverberation times. Bradley [20], using both subjective test results and the objective intelligibility measures, suggested optimum reverberation times for classrooms of 0.4 to 0.5 s with a background noise level of 30 dBA. The effects of reverberation on signal-to-noise ratio were taken into account by calculating an overall A-weighted signal-to-noise ratio for each speech level at each listener. Plomp and Mimpen [58] found non-zero optimum reverberation times with a variety of room sizes and signal-to-noise ratios by applying the image method. In their study, the effects of reverberation on noise were incorporated by considering the audience as a collection of individual noise sources. Bistafa and Bradley [59] predicted nonzero optimal reverberation times using a number of metrics. They found that increased reverberation increased early energy and intelligibility. However, too much reverberation decreased speech intelligibility. Increases of speech levels with increasing reverberation were considered in their study, but the effects of reverberation on noise levels were ignored. Hodgson and Nosal [9] addressed how to incorporate noise in a realistic manner. Optimum reverberation times depended on the signal-to-noise level difference at the listener's position, the positions and orientations of the speaker and the noise source, and the number of noise sources. They found that if the speech source was farther from the listener than the noise source, then speech levels increased more with reverberation than did noise levels, and the level difference increased with reverberation, tending to increase intelligibility. If, on the other hand, the noise source was farther from the

listener than the speech source, then noise levels increased more with reverberation than did speech levels, tending to decrease intelligibility. Thus, the effect of reverberation on intelligibility depended on the distances from the listener of both the speech source and the noise source.

1.2.3 Speech intelligibility in virtual environments

Subjective speech-intelligibility tests give more accurate results than objective metrics. However, subjective speech-intelligibility tests in real rooms have the limitation that they are difficult to perform with a large number of subjects. Auralization brings a solution to the limitations of the subjective speech tests, having unlimited capability of reproduction of the realistic listening environments and making it possible for subjective speech intelligibility to be predicted in a room before it is actually built.

Auralization or acoustical virtual reality is "the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modelled space" [60]. Historical reviews of auralization are presented in Kleiner *et al.* (1993) [60] and Lokki (2002) [61].

To perform auralization, room impulse responses are recorded or calculated, then are convolved with source signals recorded in an anechoic chamber. Figure 1.1 shows the auralization procedure. The image-source model and ray-tracing methods are well known algorithms for room-acoustic prediction used in auralization [62,63,64,65]. The influence of the absorption, scattering, and diffraction of surfaces has been incorporated in room-acoustic prediction [66,67]. Currently, calculated room impulse responses are widely used in auralization which, in that case, is called 'fully-computed' auralization [65,68,69,70].
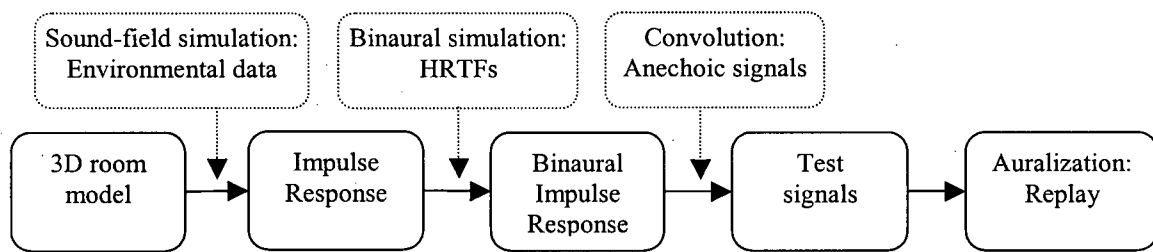
Figure 1. 1. Flow chart of the fully-computed auralization procedure

Auditory perception is sensitive with respect to listening environment [71]. Due to the auditory characteristics of humans, headphone playback systems need to compensate for the sound localization, sound transmission from a free field to a point in the ear canal of a human, and to consider the frequency response of the headphone used. The head-related transfer functions (HRTFs) are modeled as a filter which accounts for the effect of the reflections from the pinnae and shoulder, as well as the shadowing effect of the head itself. The HRTFs not only vary in a complex way with azimuth, elevation, range, and frequency, but also vary significantly from person to person [72,73]. Wightman and Kistler [72,74] developed and validated synthesized stimuli over headphones using the head-related transfer functions (HRTFs) measured from each subject. Judgments of azimuth were quite accurate, and there were no obvious differences between judgments made with free-field signals and those made with the synthesized ones. Judgments of elevation, however, tended to be more variable with the synthesized signals [72]. Wenzel *et al.* [75] extended Wightman and Kistler's work to sixteen untrained subjects with non-individual HRFTs. The results suggested that most listeners could obtain useful directional information of azimuth dimension from an auditory display without requiring the use of individual HRTFs. However, they found the high rate of front-to-back confusions to be asymmetric around the interaural axis with non-individual HRTFs. Virtual stimuli processed using non-individual HRTFs have been cited in the literature as degrading localizaion accuracy, decreasing externalization, and increasing reversal errors [76,77].

Although the non-individual HRTFs have drawbacks, measurement of individual HRTFs is impractical in an application developed for a large number of people [78,79].

Bronkhorst [80] found no significant effect of using individualized HRTFs on reversals with two different noise spectra. When virtual speech stimuli were incorporated, no significant main effects were found for head tracking or use of individual HRTFs [81]. Whether or not individualized cues are required may depend upon the nature of the task [79].

Subjective speech-intelligibility tests using auralization techniques have been used for various purposes. Nordlund, Kihlman, and Lindblad [82] developed a method for speech-intelligibility test using an artificial speaker having human voice directivity and an artificial head constructed for stereophonic recording in an auditorium. Kleiner [83] used more advanced auralization techniques for speech-intelligibility tests including calculated echograms and measured echograms to synthesize sound fields. Ricard and Meirs [84] incorporated head-related transfer functions (HRTFs) [72,73] into their Modified Rhyme Test (MRT) phrases for both recognition and localization tasks. Besing and Koehnke [85,86] used source-to-eardrum transfer functions to develop speech-intelligibility tests in noise. Peng [87] found that auralization with simulated binaural room impulse responses (BRIRs) was more accurate to test speech intelligibility than that with monaural room impulse responses. Detailed review on speech intelligibility in virtual environments will be presented in Chapter 2.

It is important to know how reliable auralization is, and to what extent acoustical details are actually simulated [88]. Although a number of auralization systems have been developed, few studies to validate the quality of the systems have been reported. This might be a consequence of the fact that such evaluation is laborious and that absolute quality is hard to define. Comparison of recordings and auralized test material has been used to evaluate the quality of auralization. Kleiner [83] compared direct listening and three different auralization techniques -- binaural recordings made in the theatre, convolution with calculated impulse response, and convolution with measured impulse response -- using speech-intelligibility tests. Lokki and Järveläinen [89] used a recording made in real room to evaluate auralization quality. Rindel and Christensen [88] compared three auralization techniques to validate the quality of their auralization. Another way of comparing real and auralized sound fields is to compare the just-noticeable differences (JNDs) for various parts of the binaural impulse response. Prodi and Velecka [90] used the JNDs to evaluate the binaural playback systems for virtual sound

fields. They concluded that all required quality features should be tested with a given specific application and it could be defined whether the performance of auralization is plausible with the specific application.

1.2.4 Room-acoustical modeling methods

For this work it was desired to predict the acoustical conditions in rooms and to obtain auralized sound fields using room-acoustical modeling methods. There are two different methods used in the modeling of room acoustics in this work: computational modeling and physical scale-modeling.

Computational modeling

There are three approaches in computational modeling in room acoustics: wave-based modeling, ray-based modeling, and statistical modeling [91]. The wave-based modeling methods (methods that account for wave phase) such as the finite-element method (FEM), boundary-element method (BEM), and finite-difference time-domain method (FDTD), are suitable only for small enclosures and low frequencies due to the heavy computational requirements [92]. The statistical modeling methods such as the statistical energy analysis (SEA), are used for prediction of noise levels in coupled rooms where sound transmission by structures is important [93].

Another statistical approach is diffuse-field theory, in which the theoretical assumption is that the reverberant sound field is perfectly diffuse [94]. In this thesis, it was used to calculate absorption coefficients and reverberation times from measured data in scale-model rooms. Since the Sabine formula was derived, diffuse-field theory has been developed [95]. It is applicable to fairly reverberant rooms with uniform surface absorption. Eyring improved upon Sabine formula to make it applicable to less reverberant spaces, by treating the waves as being absorbed only at the surfaces [96]. Fitzroy improved upon Sabine by allowing the absorbent material to be distributed unevenly [97]. Although a diffuse sound field is an idealization in rooms, it can be applicable in real rooms when the room absorption is

uniformly distributed and the room shape is quasi-cubic, especially if the room surfaces are diffusely reflecting [98].

The ray-based modeling methods, such as the ray-tracing and the image-source methods, are most commonly used for room acoustical modeling [99,100,101,62]. There are also hybrid models, which are combined ray-tracing and image-source method together [64,102]. In hybrid models, early reflections are calculated with the image-source method due to its accuracy in finding reflection paths, and later reflections are calculated with ray-tracing method, due to its efficiency in computational requirement.

One of the hybrid acoustical prediction and auralization software systems, CATT-Acoustic v8.0 [103], was used in this work. For early sound reflections, the image-source method was utilized, with added first-order diffuse reflection. The direct sound, first-order diffuse and specular reflections and second-order specular reflections are handled by the image-source method. For fully-detailed calculation, a randomized tail-corrected cone-tracing (RTC) method was used. The RTC combines features of both specular cone-tracing, standard ray-tracing and the image-source method [104].

Physical scale-modeling

Physical scale-modeling methods are now well proven and there is considerable knowledge regarding their ability to represent the acoustical conditions of real rooms [105, 106, 107, 108,109]. Acoustic scale modeling was first undertaken by Spandöck in 1934 [110]. The principle behind acoustic scale modeling is simple. In order to create a 1/n scale model, all dimensions are scaled by 1/n. If the wavelength is similarly reduced then wavelength-to-dimension ratios remain unchanged. If the wavelength is reduced, the frequency must be increased. However the propagation medium (generally air), in the model is the same as in the real room, so for a 1/n scale model, frequency should be increased by n times (possibly involving ultrasonic model frequencies). At these higher model frequencies, the absorption and diffusion properties of the model surfaces and the air should be the same as in the real room at normal frequencies.

A major constraint to the choice of scale factor relates to the transducers, which tend to have a high frequency limit around 100 kHz, both in the case of loudspeakers and microphones. Air absorption is also a major problem in scale-model measurement, air absorption increases approximately with the square of the frequency [111]. Since it should increase in proportion to frequency, the air-absorption effect is excessive in a scale model and cannot be neglected. The air absorption can be calculated at the model test frequencies for each temperature and relative humidity measured in the scale-model room [111].

1.2.5 Summary

Previous studies of optimal acoustical conditions for speech intelligibility were based on two major approaches. A series of experimental and theoretical studies have been carried out to understand the acoustical conditions important for speech intelligibility. Different types of studies, with different ways of incorporating noise, resulted in different apparent optimal reverberation times for speech. In the experimental studies, the use of fixed signal-to-noise ratios at the listener eliminated the accurate inclusion of the effect of reverberation on speech intelligibility. In the theoretical predictions, it was found that the effect of reverberation on intelligibility was dependent on the distances from the listener of both the speech source and the noise source. In order to prove the theoretical prediction, more accurate experimental study is imperative.

Subjective speech-intelligibility tests widely used in room acoustics, objective metrics to predict speech intelligibility, and the possibility of using the auralization technique were reviewed for the experimental study on speech intelligibility. Applying auralization to speech-intelligibility tests was deemed plausible in terms of its repeatability of test signals, as long as auralization was first validated for speech-intelligibility testing. Two room modeling methods chosen for this work, CATT-Acoustic and physical scale-modeling, were introduced.

1.3 Objectives of the Thesis

The primary objective of this thesis is to find the optimal reverberation for speech intelligibility by an experimental method using more realistic room-acoustical models. The interaction between reverberation and steady-state level will be incorporated into room-acoustical models. The effect of doing so is to increase the reality and accuracy of speech-intelligibility tests, resulting in more valuable data for the acoustic design of rooms for speech.

The second objective is to find the optimal reverberation for speech intelligibility for both normal and hearing-impaired listeners. Optimal room-acoustical conditions for hearing-impaired people on speech intelligibility are not well known, despite there being are a number of audiological studies on speech intelligibility for hearing-impaired listeners. Considering that the hearing-impaired population requires acoustically more elaborate listening environments, knowledge of the optimal room-acoustical conditions is imperative.

The third objective is to propose an optimal architectural design for speech intelligibility based on the findings of this thesis. Ceiling barriers and reflectors will be used as a design variable in classrooms. The results provide a design concept for high quality classroom acoustics.

1.4 Approach

The auralization technique was employed as a fundamental method for reproducing sound fields, and its application was validated for speech intelligibility in this thesis.

The first step in finding the optimal reverberation for speech intelligibility using auralization was to create idealized virtual rooms with an approximately diffuse sound field in order to have results to be tested in the same conditions of the previous experimental studies. For more accurate modeling of the interaction between reverberation and speech-to-noise ratio, the concepts of the speech-to-noise level difference at a listener's position (SNR) and the

relative output-power levels of the speech and noise sources were chosen to give differences (*SNS*) that are clearly defined and used in this thesis.

The next step consisted of applying the results from the first step into more realistic virtual rooms with non-diffuse sound fields, modeled based on existing classrooms. This step repeats the first step, and confirms and extends the results of the previous study.

Finally, a novel practicable method for achieving high speech intelligibility in classrooms is proposed based on the results of the previous steps. Ceiling barriers and reflectors for optimizing classroom acoustics for speech intelligibility were evaluated by using a physical scale-model and computer prediction.

1.5   Thesis Outline

This thesis is organized as follows. Chapter 2 discusses validation of speech-intelligibility testing in virtual classrooms in comparison with real classrooms. Chapter 3 investigates the optimum reverberation for speech intelligibility when noise is incorporated in a realistic manner in a simplified virtual room for both normal and hearing-impaired listeners. Chapter 4 develops Chapter 3 in the case of realistic virtual rooms. Chapter 5 reports tests of ceiling barriers and reflectors as an architectural solution to improve speech intelligibility based on the optimum acoustical conditions found in Chapter 3 and Chapter 4. Chapter 6 concludes the thesis and discusses the future work.

References

[1]   G. A. Campbell, "Telephonic intelligibility," Phil. Mag. **19**, 152-159 (1910).

[2]   H. Fletcher, "The nature of speech and its interpretation," J. Franklin Inst. **193** (6), 729-747 (1922).

[3]   H. Fletcher and J. C. Steinberg, "Articulation testing methods," J. Acoust. Soc. Am. **1** (1), 1-48 (1930).

[4]   P. Pratt, in *Sound & Vibration Control* (1996), pp. 49-54.

[5]   A. A. Tyler and L. C. Tolbert, "Speech-language assessment in the clinical setting," Am. J. Speech-Lang. Pat. **11** (4), 215-220 (2002).

[6]   D. Noffsinger, G. B. Haskell, V. D. Larson et al., "Quality rating test of hearing aid benefit in the NIDCD/VA Clinical Trial," Ear Hear. **23** (4), 291-300 (2002).

[7]   A. C. Neuman and I. Hochberg, "Children's perception of speech in reverberation," J. Acoust. Soc. Am. **73** (6), 2145-2149 (1983).

[8]   J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. **106** (4), 1820-1828 (1999).

[9]   M. Hodgson and E.-M. Nosal, "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms," J. Acoust. Soc. Am. **111** (2), 931-939 (2002).

[10]   H. J. M. Steeneken,"Quality evaluation of speech processing systems," in *Digital speech processing: speech coding, synthesis, and recognition*, edited by I. A. Nejat (Kluwer Academic Publishers, Boston, 1992), pp. 127-160.

[11]   J. P. Penrod,"Speech threshold and word recognition/discrimination testing," in *Handbook of cliniclal audiology*, edited by J. Katz (Williams & Wilkins, Baltimore, 1994), pp. 147-164.

[12]   J. W. Black, "Multiple-choice intelligibility tests," J. Speech Hear. Dis. **22** (2), 213-235 (1957).

[13]   G. Fairbanks, "Test of phonemic differentiation: The rhyme test," J. Acoust. Soc. Am. **30** (7), 596-600 (1958).

[14]   E. L. Thorndike and I. Lorge, *The teacher's word book of 30,000 words*. (Teachers college Columbia university, New York, 1944).

[15]   A. S. House, C. E. Williams, M. H. L. Hecker et al., "Articulation-testing methods: Consonantal differentiation with a closed-response set," J. Acoust. Soc. Am. **37** (1), 158-166 (1965).

[16]   E. J. Kreul, J. C. Nixon, K. D. Kryter et al., "A proposed clinical test of speech discrimination," J. Speech Hear. Res. **11** (3), 536-552 (1968).

[17]   J. D. Griffiths, "Rhyming minimal contrasts: A simplified diagnostic articulation test," J. Acoust. Soc. Am. **42** (1), 236-241 (1967).

[18]   H. G. Latham, "The signal-to-noise ratio for speech intelligibility--An auditorium acoustics design index," Appl. Acoust. **12** (4), 253-320 (1979).

[19]   J. S. Bradley, "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am. **80** (3), 837-845 (1986).

[20]   J. S. Bradley, "Speech intelligibility studies in classrooms," J. Acoust. Soc. Am. **80** (3), 846-854 (1986).

[21]   A. K. Nabelek and J. M. Pickett, "Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing-aids," J. Acoust. Soc. Am. **56** (2), 628-639 (1974).

[22]   A. K. Nabelek and J. M. Pickett, "Monaural and binaural speech-perception through hearing-aids under noise and reverberation with normal and hearing-impaired listeners," J. Speech Hear. Res. **17** (4), 724-739 (1974).

[23]   A. K. Nabelek and P. K. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71** (5), 1242-1248 (1982).

[24]   A. K. Nabelek and I. V. Nabelek,"Room acoustics and speech perception," in *Handbook of clinical audiology*, edited by J. Katz (Williams & Wilkins, Baltimore, 1994), pp. 624-637.

[25]   N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19** (1), 90-119 (1947).

[26]   D. K. Karl, "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34** (11), 1689-1697 (1962).

[27]   D. K. Karl, "Validation of the articulation index," J. Acoust. Soc. Am. **34** (11), 1698-1702 (1962).

[28]   V. M. A. Peutz, "Articulation loss of consonants as a criterion for speech transmission in a room," J. Audio Eng. Soc. **19** (11), 915-919 (1971).

[29]   T. Houtgast and H. J. M. Steeneken, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," Acustica **28**, 66-73 (1973).

[30]   T. Houtgast and H. J. M. Steeneken, "Evaluation of speech transmission channels by using artificial signals," Acustica **25**, 355-367 (1971).

[31]   T. Houtgast, H. J. M. Steeneken, and R. Plomp, "Predicting speech intellgibility in rooms from the modulation transfer function. I. General room acoustics," Acustica **46** (1), 60-72 (1980).

[32]   T. Houtgast, H. J. M. Steeneken, and R. Plomp, "Predicting speech intellgibility in rooms from the modulation transfer function. II. Mirror image computer model applied to rectangular rooms," Acustica **46** (1), 73-81 (1980).

[33]   H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67** (1), 318-326 (1980).

[34]   T. Houtgast and H. J. M. Steeneken, "A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria," Acustica **54** (4), 185-199 (1984).

[35]   American National Standards Institute, "ANSI S3.5-1997 Methods for caculation of the speech intelligibility index," 1997.

[36]   A. Boothroyd, "Room acoustics and speech perception," Semin. Hear. **25** (2), 155-166 (2004).

[37]   H. Haas, "The influence of a single reflection on the audibility of speech," J. Audio. Eng. Soc. **20**, 145159 (1972).

[38]   W. Reichardt, O. Abdelalim, and W. Schmidt, "Definition and basis of making an objective evaluation to distinguish between useful and useless clarity defining musical performances," Acustica **32** (3), 126-137 (1975).

[39]   J. P. A. Lochner and J. F. Burger, "The influence of reflections on auditorium acoustics," J. Sound Vib. **1** (4), 426-448 (1964).

[40]   T. Finitzo-Hieber and T. W. Tillman, "Room acoustics effects on mono-syllabic word discrimination ability for normal and hearing-impaired children," J. Speech Hear. Res. **21** (3), 440-458 (1978).

[41]   W. S. Yacullo and D. B Hawkins, "Speech recognition in noise and reverberation by school-age children," Audiology **26**, 235-246 (1987).

[42]   R. Plomp and A. M. Mimpen, "Speech-reception threshold for sentences as a function of age and noise level," J. Acoust. Soc. Am. **66** (5), 1333-1342 (1979).

[43]   J. R. Dubno, D. D. Dirks, and D. E. Morgan, "Effects of age and mild hearing loss on speech recognition in noise," J. Acoust. Soc. Am. **76** (1), 87-96 (1984).

[44]   R. Plomp, "A signal-to-noise ratio model for the speech-reception threshold of the hearing-impaired," J.Speech Hear. Res. **29** (2), 146-154 (1986).

[45]   C. C. Crandell, "Speech recognition in noise by children with minimal degrees of sensorineural hearing loss," Ear Hear. **14** (3), 210-216 (1993).

[46]   A. K. Nabelek and L. Robinette, "Reverberation as a parameter in clinical testing," Audiology **17** (3), 239-259 (1978).

[47]   A. K. Nabelek and T. R. Letowski, "Vowel confusions of hearing-impaired listeners under reverberant and nonreverberant conditions," **50** (2), 126-131 (1985).

[48]   A. K. Nabelek and P. A. Dagenais, "Vowel errors in noise and in reverberation by hearing-impaired listeners," J. Acoust. Soc. Am. **80** (3), 741-748 (1986).

[49]   A. J. Duquesnoy and R. Plomp, "Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis," J. Acoust. Soc. Am. **68** (2), 537-544 (1980).

[50]   R. Plomp and A. J. Duquesnoy, "Room acoustics for the aged," J. Acoust. Soc. Am. **68** (6), 1616-1621 (1980).

[51]   L. E. Humes, D. S. D. Dirks, T. Bell et al., "Application of the articulation index and the speech transmission index to the recognition of speech by normal-hearing and hearing-impaired listeners.," J. Speech Hear. Res. **29** (4), 447-462 (1986).

[52]   C. V. Pavlovic, G. A. Studebaker, and R. L. Sherbecoe, "An Articulation Index based procedure for predicting the speech recognition performance of hearing-impaired individuals," J. Acoust. Soc. Am. **80** (1), 50-57 (1986).

[53]   K. L. Payton, R. M. Uchanski, and L. D. Braida, "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95** (3), 1581-1592 (1994).

[54]   C. A. Kamm, D. D. Dirks, and T. S. Bell, "Speech recognition and the articulation index for normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77** (1), 281-288 (1985).

[55]   L. E. Humes, "Factors underlying the speech-recognition performance of elderly hearing-aid wearers," J. Acoust. Soc. Am. **112** (3), 1112-1132 (2002).

[56]   D. J. Schum, L. J. Matthews, and F. S. Lee, "Actual and predicted word-recognition performance of elderly hearing-impaired listeners," J. Speech Hear. Res. **34** (3), 636-642 (1991).

[57]   W. Yang and M. Hodgson, "Acoustical evaluation of preschool classrooms," Noise Control Eng. J. **53** (2), 43-52 (2005).

[58]   R. Plomp and A. M. Mimpen, "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18** (1), 43-52 (1979).

[59]   Sylvio R. Bistafa and John S. Bradley, "Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics," J. Acoust. Soc. Am. **107** (2), 861-875 (2000).

[60]   M. Kleiner, B. I. Dalenback, and P. Svensson, "Auralization - An overview," **41** (11), 861-875 (1993).

[61]   T. Lokki, *Physically-based auralization: Design, implementation, and evaluation*, D.Tech., Helsinki University of Technology, 2002.

[62]   J. Borish, "Extension of the image model to arbitrary polyhedra," J. Acoust. Soc. Am. **75** (6), 1827-1836 (1984).

[63]   D. R. Begault, "Challenges to the successful implementation of 3-D sound," **39** (11), 864-870 (1991).

[64]   G. M. Naylor, "ODEON--Another hybrid room acoustical model," **38** (2-4), 131-143 (1993).

[65]   K. H. Kuttruff, "Auralization of impulse responses modeled on the basis of ray-tracing results," J. Audio Eng. Soc. **41** (11), 876-880 (1993).

[66] J. H. Rindel, "Modelling the angle-dependent pressure reflection factor," Appl. Acoust. **38** (2-4), 223-234 (1993).

[67] R. R. Torres, U. P. Svensson, and M. Kleiner, "Computation of edge diffraction for more accurate room acoustics auralization," J. Acoust. Soc. Am. **109** (2), 600-610 (2001).

[68] W. Ahnert and R. Feistel, "Ears auralization software," J. Audio Eng. Soc. **41** (11), 894-904 (1993).

[69] B. I. Dalenback, M. Kleiner, and P. Svensson, "Audibility of changes in geometric shape, source directivity, and absorptive treatment - Experiments in auralization," **41** (11), 905-913 (1993).

[70] A. Mochimaru, "A study of the practicality and accuracy of impulse-response calculations for the auralization of sound-system design," J. Audio Eng. Soc. **41** (11), 881-893 (1993).

[71] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*, Rev. ed. (MIT Press, Cambridge, Mass., 1997), pp.xiii, 494.

[72] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85** (2), 868-878 (1989).

[73] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91** (3), 1637-1647 (1992).

[74] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am. **85** (2), 858-867 (1989).

[75] E. M. Wenzel, M. Arruda, D. J. Kistler et al., "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94** (1), 111-123 (1993).

[76] H. Moller, M. F. Sorensen, C. B. Jensen et al., "Binaural technique: Do we need individual recordings?" J. Audio Eng. Soc. **44** (6), 451-469 (1996).

[77] J. Sodnik, R. Susnik, M. Stular et al., "Spatial sound resolution of an interpolated HRIR library," Appl. Acoust. **66** (11), 1219-1234 (2005).

[78] E. M. Wenzel, P. K. Stone, S. S. Fisher et al., "A system for three-dimensional acoustic 'visualization' in a virtual environment workstation," Proc. 1st IEEE Conference on Visualization '90, San Francisco, CA, USA (1990).

[79] E. M. Wenzel and D. R. Begault, "Are individualized head-related transfer functions required for auditory information displays?" J. Acoust. Soc. Am. **105** (2), 1035 (1999).

[80] A. W. Bronkhorst, "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **98** (5), 2542-2553 (1995).

[81] D. R. Begault and E. M. Wenzel, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," J. Audio Eng. Soc. **49** (10), 904-916 (2001).

[82] B. Nordlund, T. Kihlman, and S. Lindblad, "Use of articulation tests in auditorium studies," J. Acoust. Soc. Am. **44** (1), 148-156 (1968).

[83] M. Kleiner, "Speech-intelligibility In real and simulated sound fields," Acustica **47** (2), 55-71 (1981).

[84] G. L. Ricard and S. L. Meirs, "Intelligibility and localization of speech from virtual directions," Human Factors **36** (1), 120-128 (1994).

[85] J. Besing and J. M. Koehnke, "A test of virtual auditory localization," Ear Hear. **16** (2), 220-229 (1995).

[86]   J. Koehnke and J. M. Besing, "A procedure for testing speech intelligibility in a virtual listening environment.," Ear Hear. **17** (3), 211-217 (1996).

[87]   J. Peng, "Feasibility of subjective speech intelligibility assessment based on auralization," Appl. Acoust. **66**, 591-601 (2005).

[88]   J. H. Rindel and C. L. Christensen, "Room acoustic simulation and auralization - How close can we get to the real room?" Proc. 8th WESPAC, Melbourne (2003).

[89]   T. Lokki and H. Jarvelainen, "Subjective evaluation of auralization of physics-based room acoustics modeling," Proc. 7th International Conference on Auditory Display, Espoo, Finland (2001).

[90]   N. Prodi and S. Velecka, "The evaluation of binaural playback systems for virtual sound fields," Appl. Acoust. **64** (2), 147 (2003).

[91]   L. Savioja, *Modeling techniques for virtual acoustics*, D.Tech., Helsinki University of Technology, 2000.

[92]   D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," J. Acoust. Soc. Am. **98** (6), 3302-3308 (1995).

[93]   C. B. Burroughs, R. W. Fischer, and F. R. Kern, "An introduction to statistical energy analysis," J. Acoust. Soc. Am. **101** (4), 1779-1789 (1997).

[94]   H. Kuttruff, *Room acoustics*, 4th ed. (Spoon Press, London New York, 2000), pp.116-122.

[95]   W. C. Sabine, *Collected papers on acoustics*. (Dover Publications, New York, 1964).

[96]   C. F. Eyring, "Reverberation Time in "Dead" Rooms," J. Acoust. Soc. Am. **1** (2A), 168 (1930).

[97]   D. Fitzroy, "Reverberation formula which seems to be more accurate with nonuniform distribution of absorption," J. Acoust. Soc. Am. **31** (7), 893-897 (1959).

[98]   M. Hodgson, "When is diffuse-field theory applicable," Appl. Acoust. **49** (3), 197-207 (1996).

[99]   A. Kulowski, "Algorithmic representation of the ray tracing technique," Appl. Acoust. **18** (6), 449-469 (1985).

[100] K. H. Kuttruff, "Auralization of impulse responses modeled on the basis of ray-tracing results," J. Audio Eng. Soc. **41** (11), 876-880 (1993).

[101] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65** (4), 943-950 (1979).

[102] M. Vorlander, "Simulation of the transient and steady-state sound-propagation in rooms using a new combined ray-tracing image-source algorithm," J. Acoust. Soc. Am. **86** (1), 172-178 (1989).

[103] B. -I. Dalenbäck, CATT-Acoustic v8.0 (Gothenburg, 2004).

[104] CATT, in *User's Manual* (Gothenburg, 2002), pp. 2-69-84.

[105] P. S. Veneklasen, "Model techniques in architectural acoustics," J. Acoust. Soc. Am. **47** (2A), 419-423 (1970).

[106] H. D. Harwood and A. N. Burd, "Acoustic scaling of studios and concert halls," Acustica **28** (6), 330-340 (1973).

[107] M. Barron and C. B. Chinoy, "1:50 scale acoustic models for objective testing of auditoria," Appl. Acoust. **12** (5), 361-375 (1979).

[108] M. Barron, "Auditorium acoustic modelling now," Appl. Acoust. **16** (4), 279-290 (1983).

[109] M. Hodgson and R. J. Orlowski, "Acoustic scale modeling of factories, Part I: Principles, instrumentation and techniques," J. Sound Vib. **113** (1), 29-46 (1987).

[110]   F. Spandöck, "Akustische Modellversuche," Annalen der Physik **20**, 345-360 (1934).

[111]   International Organisation for Standardisation (ISO), "ISO 9613-1 Acoustics - Attenuation of Sound During Propagation Outdoors - Part 1: Calculation of the Absorption of Sound by the Atmosphere First Edition," 1993.

2 VALIDATION OF THE AURALIZATION TECHNIQUE: SPEECH INTELLIGIBILITY TESTS IN VIRTUAL AND REAL CLASSROOMS *

2.1 Introduction

Subjective speech-intelligibility tests give more realistic results than do the measurement of prediction of objective metrics. However, performing such tests in real rooms has limitations – for example, they are difficult to perform with a large number of subjects. Auralization offers a solution to these limitations, having an unlimited capability to reproduce realistic listening environments, and making it possible for speech intelligibility to be assessed in a room before it is built. However, before it can be used with confidence, the auralization technique must be validated in comparison with direct listening in real rooms.

There are several ways to perform subjective speech intelligibility tests, as follows:

(1) on-site listening tests:

- a person speaks speech material in the room;

- a loudspeaker plays speech material, recorded in an anechoic chamber, into the room.

(2) off-site listening tests:

- speech material recorded in the room are reproduced by loudspeakers in an anechoic chamber or via headphones;

- speech material recorded in an anechoic chamber are convolved with room impulse responses measured in the room, and are reproduced by loudspeakers in an anechoic chamber or via headphones.

- speech material recorded in an anechoic chamber are convolved with calculated room impulse responses and are reproduced by loudspeakers in an anechoic chamber or via headphones ('fully-computed' auralization).

For off-site listening methods, auralization techniques are required. The quality of speech-intelligibility tests depends on how well the speech source is modeled and how speech sound is transmitted to listeners – that is, the accurate modeling of the room and the listener. This is what auralization seeks to achieve.

Speech-intelligibility tests using auralization methods have been used for acoustical evaluation of rooms, for validation of the auralization technique, and for clinical purposes. Nordlund, Kihlman, and Lindblad [1] developed a method for studying the correlation between the acoustical features and speech intelligibility of the room. They made recordings of nonsense monosyllables using an artificial speaker having human voice directivity and an artificial head constructed for stereophonic recording in an auditorium, and played the recordings through headphones. This method laid the groundwork for an auralization study on speech intelligibility. Kleiner [2] used more advanced auralization techniques for speech intelligibility testing. He used calculated and measured echograms to synthesize sound fields. Sound fields simulated in an anechoic chamber were recorded using a dummy-head and presented to listeners using headphones. The speech-intelligibility scores in on-site listening tests were compared with off-site listening-test results using dummy recordings, measured echograms, and calculated echograms. In general, the three different off-site speech-intelligibility scores were correlated with the on-site speech-intelligibility scores; however, the work revealed errors using the auralization technique for speech intelligibility. The speech-intelligibility scores in on-site listening tests in a theatre with a reverberation time os 0.8 s in the 1-kHz octave band were higher than those in off-site listening tests using auralization techniques.

Ricard and Meirs [3] studied speech intelligibility and listener ability to localize speech in a virtual sound field. They employed head-related transfer functions (HRTFs) to condition Modified Rhyme Test (MRT) phrases for both recognition and localization tasks. Localization of azimuth was accurate, but front-back confusions in the range of other

localization studies that used HRTFs were found. Besing and Koehnke [4,5] used source-to-eardrum transfer functions to develop speech intelligibility tests for virtual auditory localization. The source-to-eardrum transfer functions were measured using a broadband stimulus in actual sound fields for each source location in each environment using the KEMAR [6] manikin. They concluded that the virtual localization test enabled potential problems inherent in free-field localization tests to be avoided. Peng [7] conducted subjective speech-intelligibility tests using auralization with binaural room impulse responses (BRIRs). He found that auralization with simulated BRIRs is a better means of testing speech intelligibility than that with monaural room impulse responses.

In summary, research has been done to develop and validate the auralization technique, but there is lack of literature to support the validity of subjective speech-intelligibility testing in auralized sound fields for speech-intelligibility studies. The objective of the present work was to validate the fully-computed auralization technique for use in a subsequent speech-intelligibility study. The results of the real-classroom measurements are compared with results of tests in virtual classrooms. The similarities and differences between the results for the real classrooms and virtual classrooms are explained. The process of verification demonstrates the reliability of the auralization technique and, in particular, the fidelity of a virtual classroom model as a representation of a realistic listening environment.

2.2   Methods

2.2.1   Subjects and test materials

Twelve university students (eight females and four males) completed the listening tests in both the real classrooms and the virtual classrooms. All of the subjects had normal hearing (thresholds of 20 dB HL or better) at octave frequencies from 250 to 8000 Hz.

The Modified Rhyme Test [8] was used for the speech-intelligibility test. A test list consists of 50 words, which are common one-syllable words for consonant identification - 25 words tested initial consonants and 25 tested final consonants. The 300 words of the MRT were recorded in an anechoic chamber by a male talker who spoke standard Canadian English.

Each word was presented in the carrier phrase, "Say (test word)", at an average rate of two syllables per second. The word order of each list was randomized to obtain the 12 lists of the MRT, each containing 50 words. The r.m.s. amplitude of each word list was normalized tot the same value. As a noise signal, four-talker babble by AudiTec of St. Louis [9] was presented with the MRT lists.

2.2.2 Classrooms and acoustical measurements

Speech-intelligibility tests were performed in two existing, medium-sized university classrooms in order to evaluate the fidelity of the 'fully-computed' auralized sound fields. The classrooms were architecturally identical, with volumes of 400 m^3, but had different acoustical characteristics. One (Room A) had acoustical treatment to reduce reverberation, the other (Room B) had no acoustical treatment. Three listening positions (r1, r2, and r3) were chosen along the centre line of the classrooms. The speech source was positioned at the front, where the instructor might typically stand. The noise source was at the back of the classroom between r2 and r3 (see Figure 2.1).

Acoustical measurements were made in both classrooms. Impulse responses and speech levels were measured using the Maximum Length Sequence System Analyser (MLSSA) at each listening position. Reverberation Time (RT), Early Decay Time (EDT), early-to-late energy ratio (C_{50}), and useful-to-detrimental energy ratio (U_{50}) were calculated from the measured and predicted impulse responses. Mid-frequency (500, 1000, and 2000 Hz octave-band average) values were used for the RT and the EDT. C_{50} and U_{50} were calculated from the measured and predicted unfiltered impulse responses by Eqs. (1) and (2).

$$C_{50} = 10\log\left\{\frac{E_d + E_{d-50ms}}{E_{50ms-\infty}}\right\}, dB \qquad (1)$$

$$U_{50} = 10\log\left\{\frac{E_d + E_{d-50ms}}{E_{50ms-\infty} + E_n}\right\}, dB \qquad (2)$$
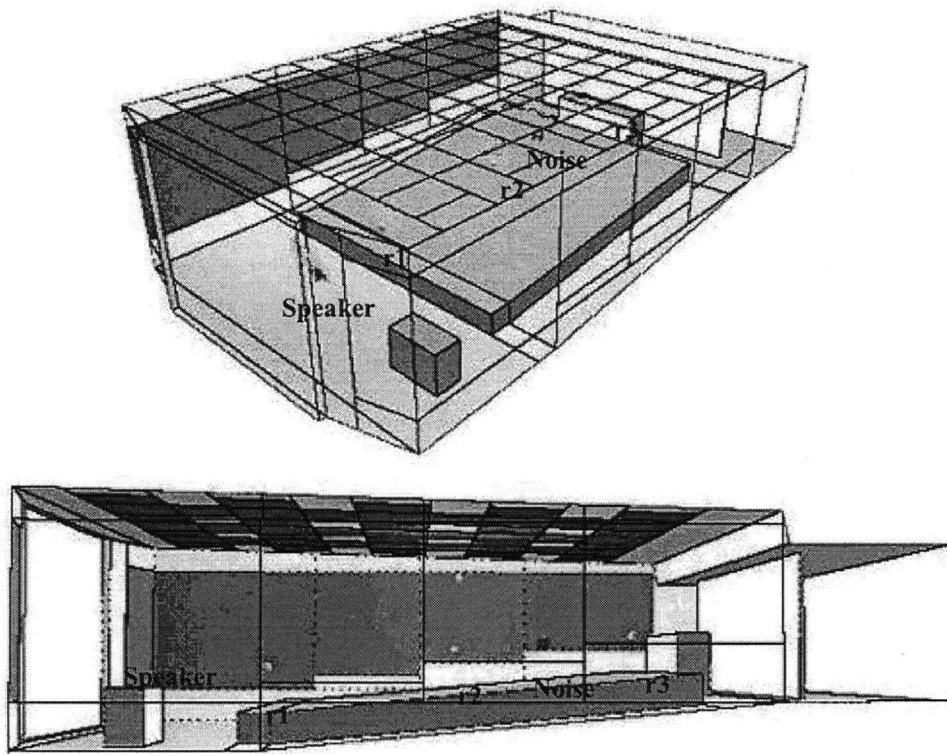
Figure 2. 1. 3D model of Room A showing sources (Speaker and Noise) and receiver positions (r1, r2 and r3).

Here, E_d is the direct energy from the source, E_{d-50ms} is the total early speech energy arriving up to 50 ms after the direct sound, $E_{50ms-\infty}$ is the total late speech energy arriving after 50 ms, and the noise energy is E_n.

2.2.3   Auralization and listening tests

The classrooms were modeled, and acoustical prediction and auralization were performed, using CATT-Acoustic v8.0 [10]. Each classroom had 153 surfaces; sources emitted 13508 rays. The classroom surfaces were discretized according to the distribution of materials in Room A, which was acoustically treated. For predicting Room B, the 3D model for Room A was also used, to maintain the same surface numbers and discretization, though

with different materials. Figure 2.1 shows the speech source, noise source and receiver positions for Room A.

The absorption coefficients and diffusion coefficients of the surface materials in the classrooms were the only acoustical quantities that could be varied to model the virtual classrooms. Table 2.1 shows absorption and diffusion coefficients of each surface material, and the average absorption and diffusion coefficients used in the room modeling. The absorption coefficients were chosen based on published absorption coefficients of the materials [11]. Room B consisted of a floor made of polished concrete, walls made of painted plywood, a ceiling made of textured plaster, and a blackboard. Room A had acoustical panels on the walls and the ceiling, with absorption coefficients 0.20, 0.57, 0.90, 0.98, 0.98, 0.97 in the 125 to 4000 Hz octave bands. The diffusion coefficients were chosen based on previous work [12], then adjustments were made to achieve a best fit to the measured RTs.

Table 2. 1. Individual and average absorption coefficients $(\alpha, \bar{\alpha})$ and diffusion coefficients $(d, \bar{d})$.

Material	Area (m^2)	Hz	125	250	500	1k	2k	4k
Glazed concrete	78.08	α	0.02	0.02	0.02	0.01	0.01	0.01
		d	0.05	0.05	0.05	0.05	0.05	0.05
Painted plywood	Room A: 72.39 Room B:132.57	α	0.28	0.07	0.06	0.07	0.07	0.06
		d	0.03	0.03	0.02	0.01	0.01	0.01
Textured plaster	Room A: 66.37 Room B: 124.24	α	0.28	0.07	0.06	0.07	0.07	0.06
		d	0.03	0.03	0.02	0.01	0.01	0.01
Wood	44.14	α	0.14	0.10	0.06	0.08	0.10	0.10
		d	0.60	0.45	0.32	0.38	0.30	0.30
Blackboard	19.32	α	0.15	0.25	0.20	0.12	0.10	0.05
		d	0.10	0.05	0.04	0.04	0.03	0.03
Seat area	60.77	α	0.13	0.16	0.15	0.13	0.18	0.25
		d	0.62	0.72	0.80	0.80	0.85	0.85
Absorption Room A only	118.05	α	0.20	0.57	0.90	0.98	0.98	0.97
		d	0.39	0.42	0.90	0.98	0.95	0.95
Room A	459.12	$\bar{\alpha}$	0.18	0.21	0.29	0.31	0.31	0.31
		$\bar{d}$	0.26	0.27	0.39	0.41	0.40	0.40
Room B	459.12	$\bar{\alpha}$	0.20	0.08	0.07	0.07	0.08	0.08
		$\bar{d}$	0.17	0.16	0.16	0.16	0.16	0.16

26

Table 2. 2. Freefield sound levels (L_{eq}) at 1m from source and power levels (L_w).

Sources (dB)		125	250	500	1k	2k	4k	unweighted
MLS	L_{eq}	43.7	46.5	50.4	57.8	60.6	61.7	65.3
	L_w	53.5	55.5	57.9	64.1	64.9	64.8	69.9
MRT	L_{eq}	51.5	57.6	59.2	56.3	54.4	48.8	63.7
	L_w	61.3	66.6	66.7	62.6	58.7	51.9	71.2
Noise	L_{eq}	50.4	59.4	62.5	51.5	43.4	33.4	64.7
	L_w	61.3	70.3	73.4	62.4	54.3	44.3	75.6

The impulse responses predicted by CATT-Acoustic were convolved with the MRT lists recorded in the anechoic chamber and with the babble noise signals. Since the Maximum Length Sequences (MLS) were used to measure the room impulse responses, and a speech signal (MRT) was used to conduct the listening tests, two sets of predictions were made to represent each case. In order to compare predictions with the results measured using the MLSSA system, the MLS spectrum was used as the speech signal in the prediction of the acoustical parameters. However, for auralization, the MRT spectrum was used as the speech signal. Free-field sound levels at 1 m from the source were measured in the anechoic chamber using both MLS and MRT sources, and are presented in Table 2.2.

For both real and virtual classroom tests, loudspeakers were used to reproduce the speech and noise. In order to account for a human speaker's directivity pattern, a loudspeaker designed for human voice testing, which had directional characteristics similar to that of the human voice, was used as the speech source. Figure 2.2 shows the measured directivities. An omnidirectional loudspeaker was used as the babble-noise source. The presentation levels of the speech-intelligibility tests were calibrated using the one-minute-average sound level ($L_{eq,1min}$) of the MRT list 1, based on Table 2.2. Subjects completed the tests at all three receiver positions (r1, r2, r3) for each configuration.

For the virtual classroom test, head related-transfer functions (HRTFs) [13] and headphone transfer functions were included in the simulations to provide more realistic sound reproduction. In order to take into account the effects of head, shoulder, and pinna, the
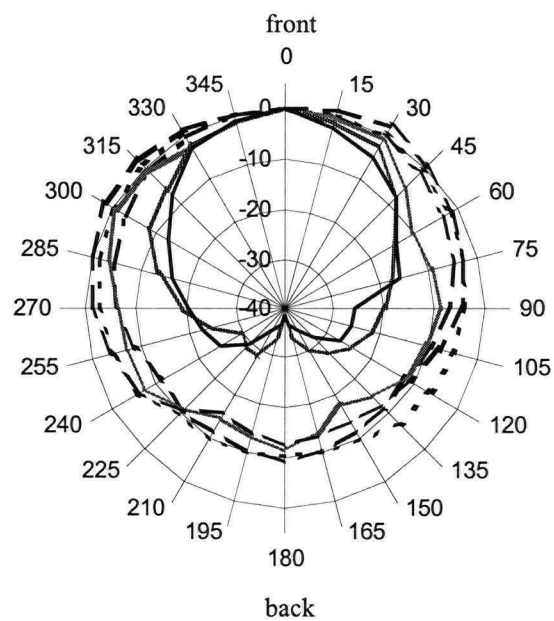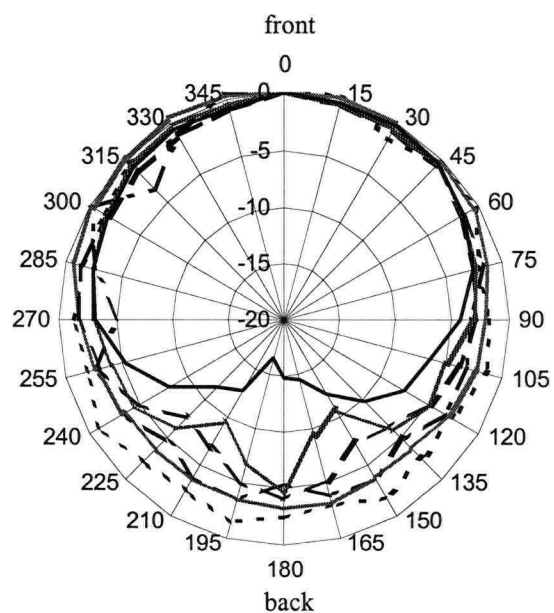
Figure 2. 2. Directivity patterns of the speech sources (a: loudspeaker radiating an MRT signal for speech intelligibility test; b: loudspeaker radiating am MLS signal for acoustical measurements and prediction. ... 125 Hz, -·- 250 Hz, — 500 Hz, --- 1000 Hz, ⚊ 2000 Hz, — 4000 Hz).

average HRTF provided by CATT-Acoustic was used, along with headphone transfer function compensation for Beyer DT 990 Pro headphones [beyerdynamic, Heilbronn]. The computed binaural impulse responses were convolved with the MRT lists and the babble noise using CATT-Acoustic. CATT-Acoustic does not provide absolute calibration of auralization levels [14]. Thus, the sound levels at the three receiver positions for the convolved MRT and the convolved babble signals were calibrated relative to one another using the calibration function in CATT-Acoustic. The speech signal and the babble noise for each combination were combined; all of the sound processing was conducted using CATT-Acoustic. Since the difference between the 1-m free-field sound levels of the MRT, and of the babble noise was 1.0 dB (see Table 2.2), which is perceptually insignificant [15], the calibrated MRT and babble noise were added without sound-level scaling. The final auralization test materials were transferred to a compact disc for presentation via a CD player. Each virtual test was performed individually in a soundproof booth using the Beyer DT 990 Pro headphones. The headphone output levels were set by measuring the equivalent sound-pressure levels of the MRT signal at the drive unit.

2.3   Results and Analysis

2.3.1   Room acoustical parameters

Figures 2.3a and b show the measured and predicted RTs and EDTs; Table 2.3 shows their differences. RT and EDT showed good agreement between measurement and prediction in both Room A and Room B. The differences were less than 10 %, except at r2 and r3 in Room A (see Table 2.3). A 10 % accuracy can be considered to be an indicator of the minimum practically significant difference [16]. Although the actual difference was not great in the high absorption classroom (Room A), the difference in percentage was greater because the measured RT and EDT were very low compared to Room B.

C_{50} showed disagreements between measurement and prediction for both Room A and Room B (see Figure 2.3c). The differences in the C_{50}s varied from -3.7 dB to 5.6 dB, as shown in Table 2.3. The just noticeable difference (JND) in C_{50} values is known to be 1.1 dB
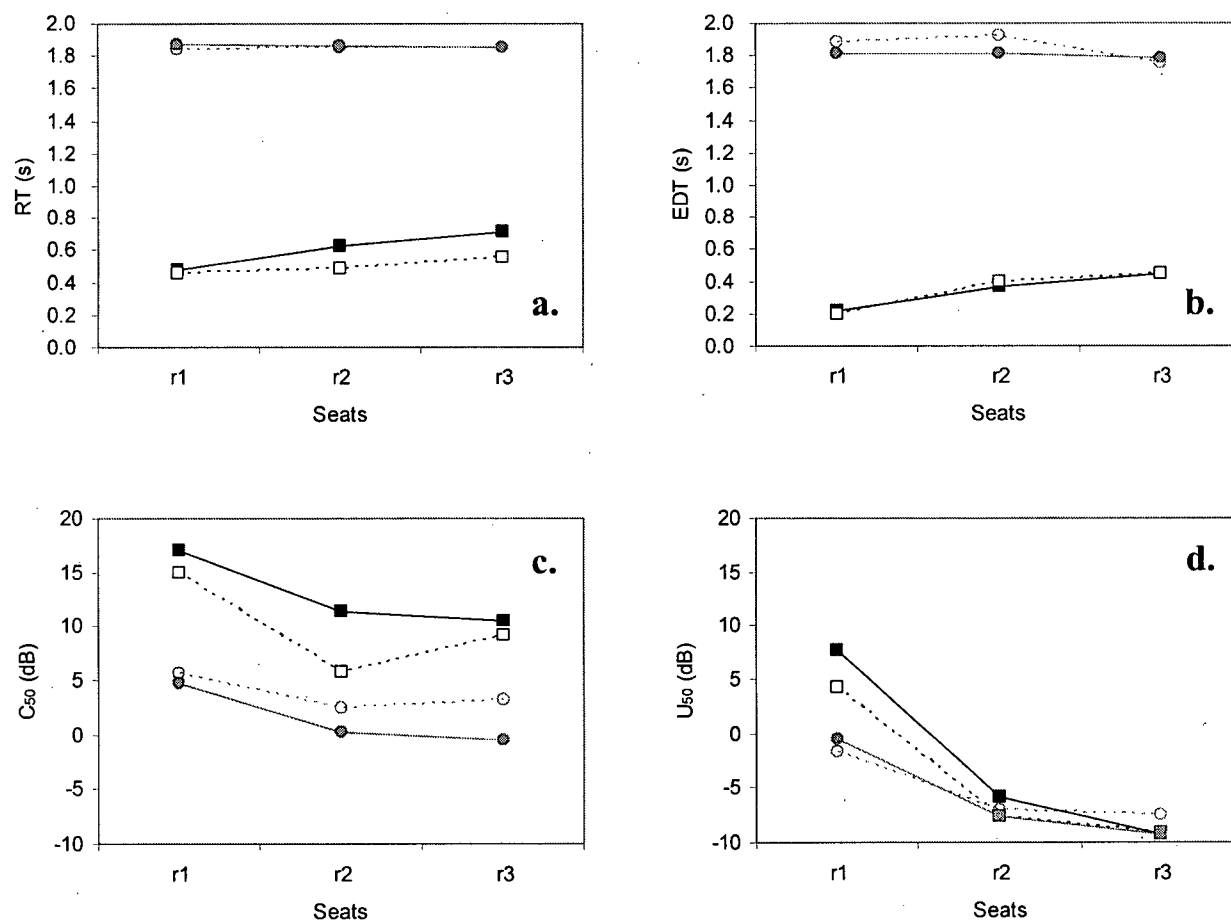
Figure 2. 3. Measured and predicted reverberation metrics (a: Reverberation times; b: Early decay times; c: C_{50} without babble; d: U_{50} with babble. ■: Predicted Room A; □: Measured Room A; ●: Predicted Room B; ○: Measured Room B).
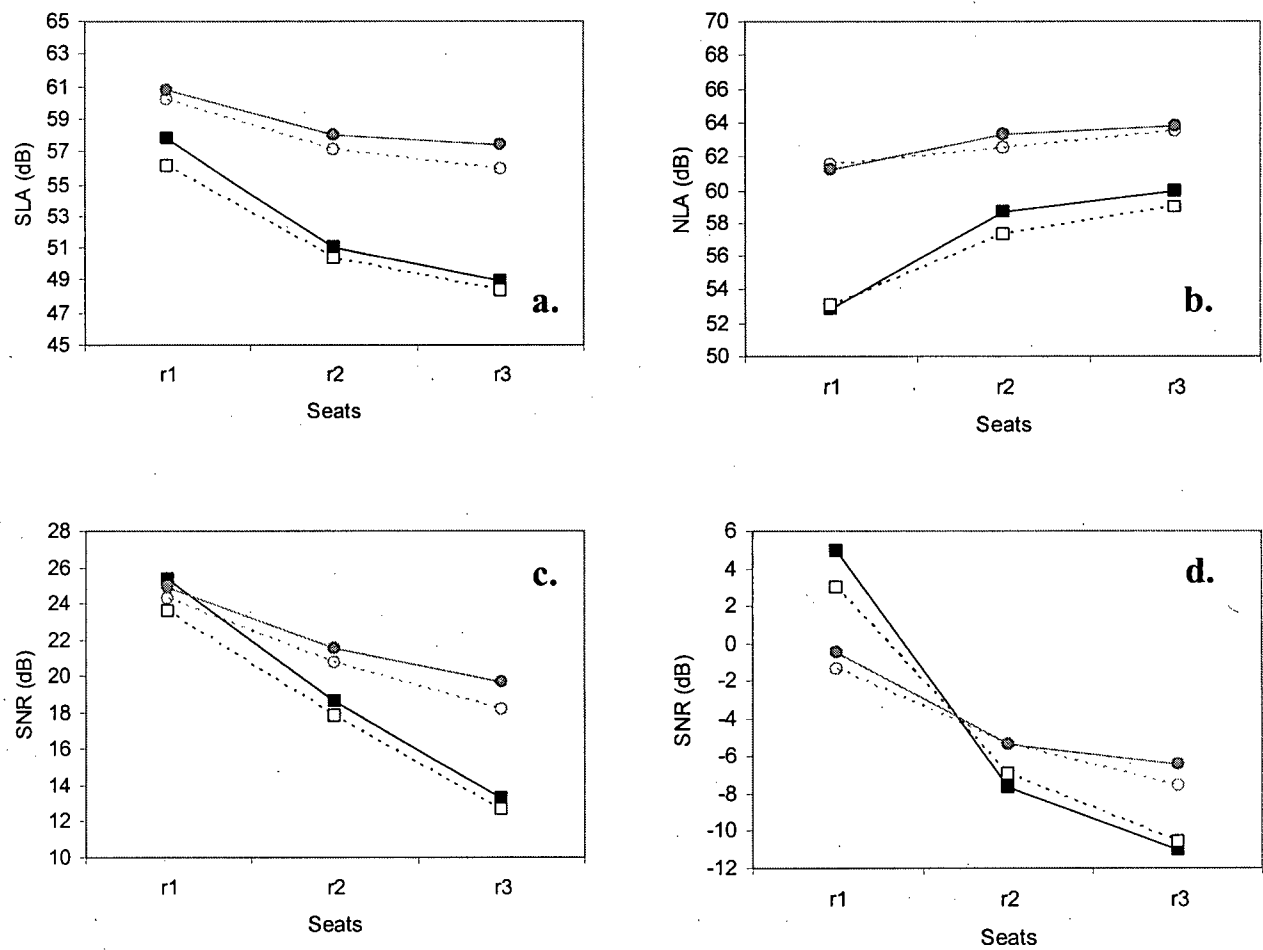
Figure 2. 4. Measured and predicted speech and babble noise levels and speech-to-noise level differences (a: speech levels; b: babble levels; c: speech-to-noise ratio with background noise, d: speech-to-noise ratio with babble noise. ■: Virtual Room A; □: Real Room A; ●: Virtual Room B; ○: Real Room B).

Table 2. 3. Difference between prediction and measurement parameters.

Predicted - Measured	Room	r1	r2	r3
RT (s)	A	0.02 (4.4 %)	0.13 (26.5 %)	0.16 (29.1 %)
	B	0.03 (1.6 %)	0.01 (0.5 %)	0.00 (0.0 %)
EDT (s)	A	0.02 (10.0 %)	-0.03 (-7.5 %)	0.00 (0.0 %)
	B	-0.07 (-3.7 %)	-0.11 (-5.7 %)	0.03 (1.7 %)
C_{50} (dB)	A	2.1	5.6	1.3
	B	-0.9	-2.2	-3.7
SPEECH (dBA)	A	1.7	0.8	0.6
	B	0.6	0.9	1.5
U_{50} (dB)	A	3.4	1.8	-0.2
	B	1.2	-0.7	-1.8
BABBLE (dBA)	A	-0.2	1.4	1.0
	B	-0.3	0.8	0.3

[17]. The measured C_{50} at r2 was lower than at r3 in both classrooms. In the high-absorption classroom (Room A), the predicted C_{50} values were higher than the measured C_{50}; however, in the low-absorption classroom (Room B), the measured C_{50} values were in general higher than the predicted C_{50}s. In Room A, C_{50} was higher if EDT was lower; however, in Room B, the relationship between C_{50} and EDT was not clear. It may be due to the large amount of late sound energy compared to the direct sound and early sound energies.

U_{50} showed better agreement than C_{50}. The differences in the U_{50}s varied from -1.8 dB to 3.4 dB, as shown in Table 2.3. The JND for U_{50} might be higher than that for C_{50} (no reference was found for the JND for U_{50}), because U_{50} includes a noise component, and the noise might desensitize the hearing ability.

Speech levels and babble noise levels were both measured and predicted. Figure 2.4 shows the speech levels, noise levels and speech-to-noise level difference, with background noise and with babble noise. Their differences are shown in Table 2.3. In most cases, the predicted sound levels were slightly higher than the measured levels.

2.3.2 Speech intelligibility tests

The mean percentage speech-intelligibility scores and 95 % confidence intervals are shown in Figure 2.5. In the high absorption classroom (Room A) without babble noise, the difference between the virtual and real classrooms was greater than in the low-absorption classroom (Room B). Contrastingly, in Room B with the babble noise, the difference between the virtual and real classrooms was much greater than in Room A. The difference in the MRT scores between the virtual and real classrooms for Room B with the babble was more than 15 %, with the virtual classroom having lower speech intelligibility relative to the real classroom.

Table 2.4 presents the mean speech-intelligibility scores and standard deviations. As expected, the standard deviations were generally greater when the mean speech-intelligibility scores were lower with babble noise and vice versa. When the babble noise was not present, the standard deviations in the real classrooms, both Room A and Room B, were greater than in the virtual classrooms, except at r3 in Room A and at r1 in Room B. This suggests that the virtual classrooms are reliable for speech-intelligibility tests without significant noise signals. Contrastingly, when the babble noise was present, the standard deviations in the virtual classrooms were greater than in the real classrooms. This suggests that auralization is sensitive to babble noise.

Table 2. 4. Mean speech-intelligibility scores and standard deviations.

		Room A				Room B			
		Without babble		With babble		Without babble		With babble	
		Virtual	Real	Virtual	Real	Virtual	Real	Virtual	Real
r1	Mean	99.17	91.83	92.93	92.33	94.67	97.00	80.83	95.50
	SD	1.34	3.35	4.63	5.25	3.55	3.46	5.49	4.04
r2	Mean	99.33	91.00	84.83	85.5	93.17	95.17	67.00	87.50
	SD	1.30	4.05	7.26	4.52	2.76	4.71	7.21	6.88
r3	Mean	96.33	94.00	77.17	80.33	91.67	94.33	59.33	80.00
	SD	2.67	2.09	8.96	5.90	2.93	4.66	6.17	6.27
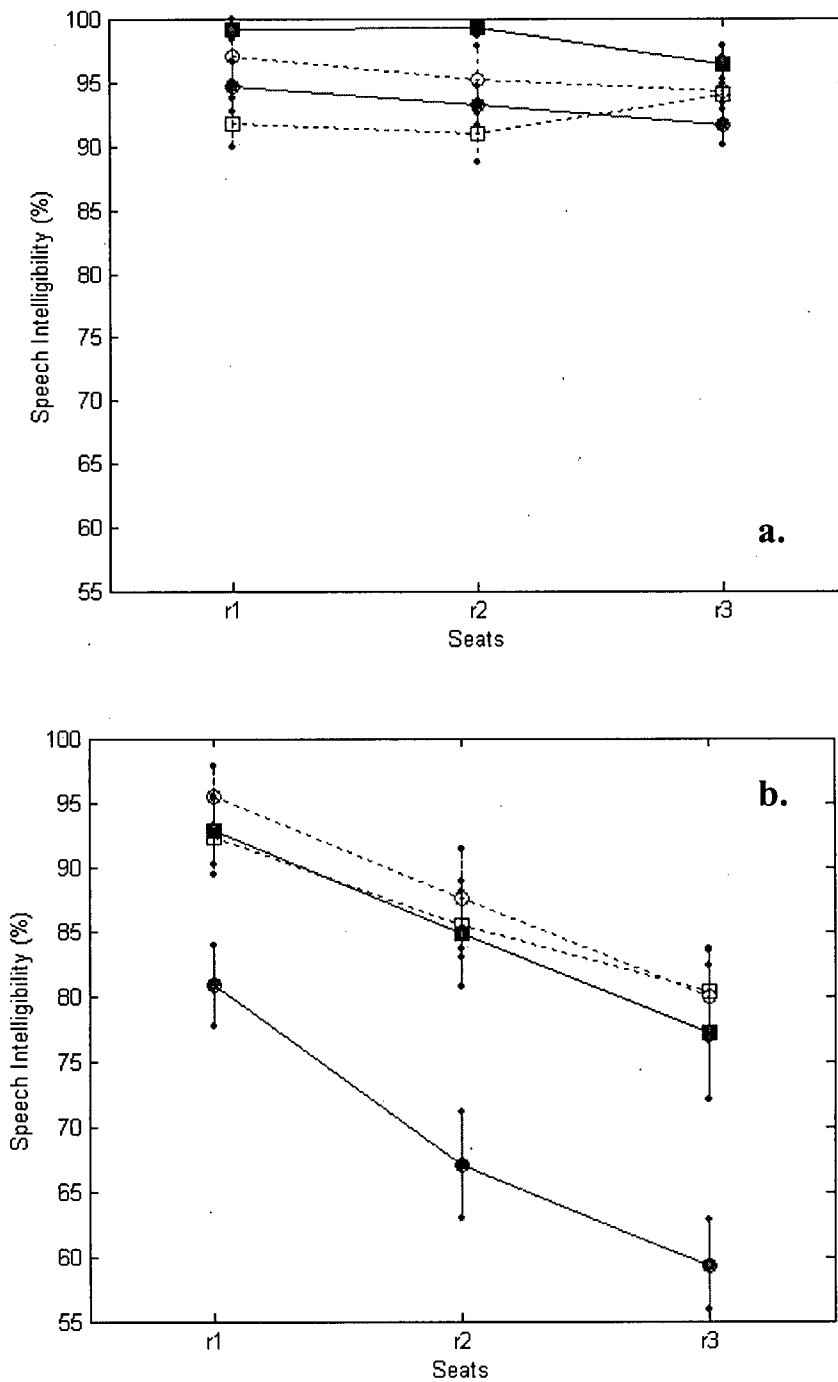
Figure 2. 5. Measured and predicted speech-intelligibility test scores and 95% confidence interval (a: without babble; b: with babble. ■: Virtual Room A; □: Real Room A; ●: Virtual Room B; ○: Real Room B).

Table 2. 5. Difference between predicted and measured MRT results.

MRT	Room	Without Babble			With Babble		
		r1	r2	r3	r1	r2	r3
Difference (%)	A	7.34 (7.99)	8.33 (9.15)	2.33 (2.48)	0.50 (0.54)	-0.67 (-0.78)	-3.16 (-3.93)
	B	-2.33 (-2.40)	-2.00 (-2.10)	-2.66 (-2.82)	-14.67 (-15.36)	-20.50 (-23.43)	-20.67 (-25.84)
Paired *t*-test (*p*-value)	A	8.16 (0.000)	6.30 (0.000)	2.55 (0.027)	0.25 (0.810)	-0.21 (0.836)	-0.98 (0.347)
	B	-2.38 (0.036)	-1.54 (0.153)	-2.4 (0.035)	-10.32 (0.000)	-6.54 (0.000)	-11.60 (0.000)

These results were interpreted by statistical analysis using the paired *t*-test, as shown in Table 2.5. Without babble, there was a highly significant difference in the speech intelligibility scores between the real and virtual classrooms in Room A at r1 and r2, but no statistically significant difference in Room B. With babble, there was no statistical difference between the real and virtual classrooms in Room A, but in Room B there was a highly statistical difference.

The mean speech-intelligibility scores were compared with C_{50}, U_{50}, and speech-to-noise level differences in Figure 2.6. Without babble noise, the speech intelligibility was highly correlated with C_{50} and speech-to-noise level difference in Room B (see Figure 2.6a and b). However, there was no clear relationship between speech intelligibility and C_{50} or speech-to-noise level difference in the high absorption classroom (Room A). With babble noise, the speech intelligibility depends on U_{50} and speech-to-noise level difference in both the high and the low absorption classrooms (Room A and Room B) (see Figures 2.6c and d). The equations and coefficients of determination (R^2) associated with the linear regression lines are presented in Table 2.6. Thus, in general, the mean speech-intelligibility scores are highly correlated with the speech-intelligibility metrics in both the low and high absorption classrooms, and in both real and virtual classrooms, with babble noise. In quite absorptive classrooms, the difference in the mean speech-intelligibility scores between the virtual classroom and the real classroom were large, and the relationship between the mean speech-intelligibility scores and speech-intelligibility metrics were not clearly seen.
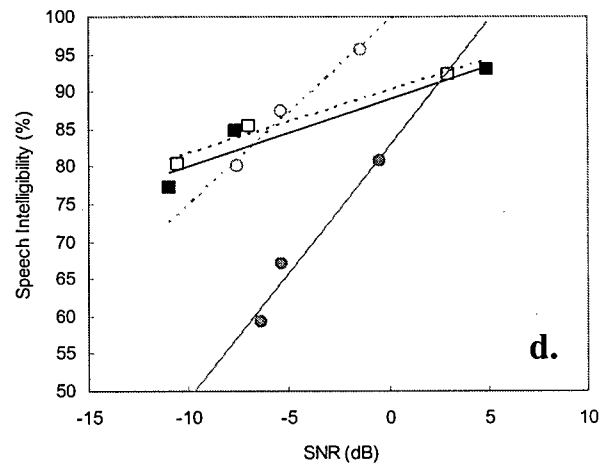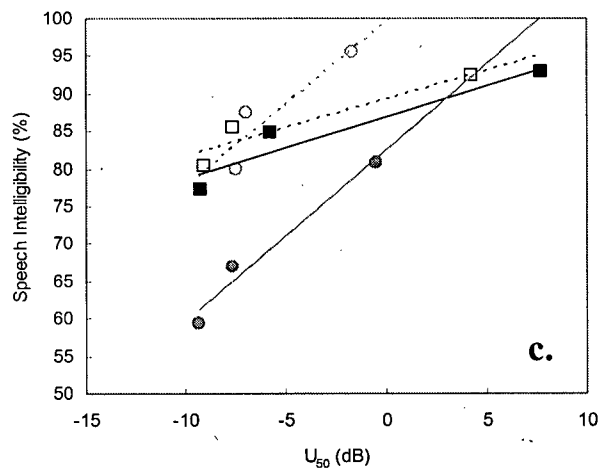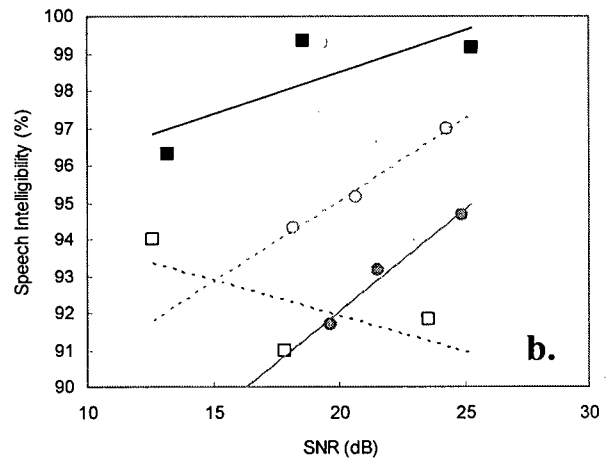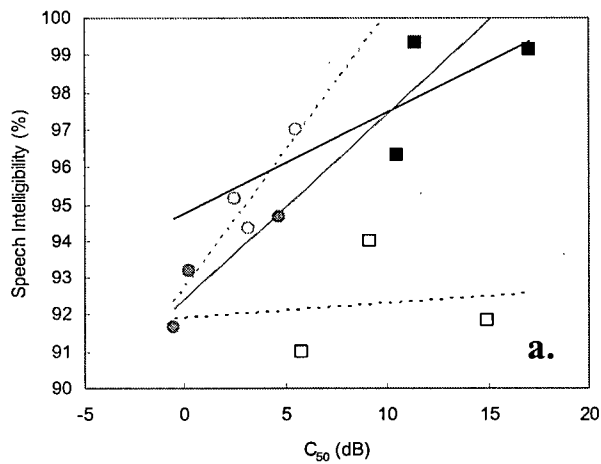
Figure 2. 6. Variation of speech intelligibility with acoustical parameters (a: C_{50}; b: Speech-to-noise level difference without babble; c: U_{50}; d: Speech-to-noise level difference with babble. ■: Virtual Room A; □: Real Room A; ●: Virtual Room B; ○: Real Room B).

Table 2. 6. Equations and coefficients of determination (R^2) associated with linear regression of C_{50}, U_{50}, and speech-to-noise level difference with mean speech-intelligibility score.

		Without Babble		With Babble	
		SI-C_{50}	SI-SNR	SI-U_{50}	SI-SNR
Virtual Room A	Eq	SI=0.2697•C_{50} +94.777	SI=0.2248•SNR +93.996	SI=0.8324•U_{50} +86.976	SI=0.8868•SNR +88.978
	R^2 (%)	32.76	64.66	90.24	90.68
Real Room A	Eq	SI=0.0393•C_{50} +91.884	SI=-0.193•SNR +95.753	SI=0.7738•U_{50} +89.252	SI=0.8408•SNR +90.117
	R^2 (%)	1.39	46.22	88.57	96.63
Virtual Room B	Eq	SI=0.4971•C_{50} +92.448	SI=0.5551•SNR +80.931	SI=2.3001•U_{50} +82.477	SI=3.3649•SNR +82.708
	R^2 (%)	86.34	97.43	96.79	96.43
Real Room B	Eq	SI=0.7356•C_{50} +92.762	SI=0.4409•SNR +86.216	SI=2.1998•U_{50} +99.515	SI=2.4374•SNR +99.329
	R^2 (%)	74.18	98.80	83.15	97.87

2.4   Discussion

Differences between the speech intelligibility in the real and virtual classrooms can be explained by one or a combination of the potential reasons listed below:

- the early-to-late energy ratio was not correctly estimated;

- surface diffusion was not sufficiently taken into account;

- the localization of the speech and noise sources in the virtual classrooms was not accurately modeled;

- sound reproduction via headphones was not the same as natural listening;

- the calibration of the speech signals and babble signals after convolution was not correctly done;

- the distortion of the electroacoustic equipment was too high.

Each of these points will be discussed in detail in this section.

Information about the early-to-late energy ratio can be obtained by considering the room impulse responses. As shown in Figures 2.7a and b, in Room A the predicted impulse

response had a stronger direct sound compared to the measured impulse response. This made predicted C_{50} values higher than the measured C_{50} in this configuration. In contrast, the predicted impulse response had a relatively long reverberant tail in the reverberant room, Room B. This can also be understood by comparing the reverberation times and early decay times shown in Figures 2.3a and b. In Room B, the predicted RTs and EDTs did not change much with receiver position in the room; however, the measured RTs were lower than the measured EDTs. Thus, the measured impulse response had more early energy than the predicted one. Even though the measured and predicted reverberation times were similar, the ratios of early energy to late energy differed.
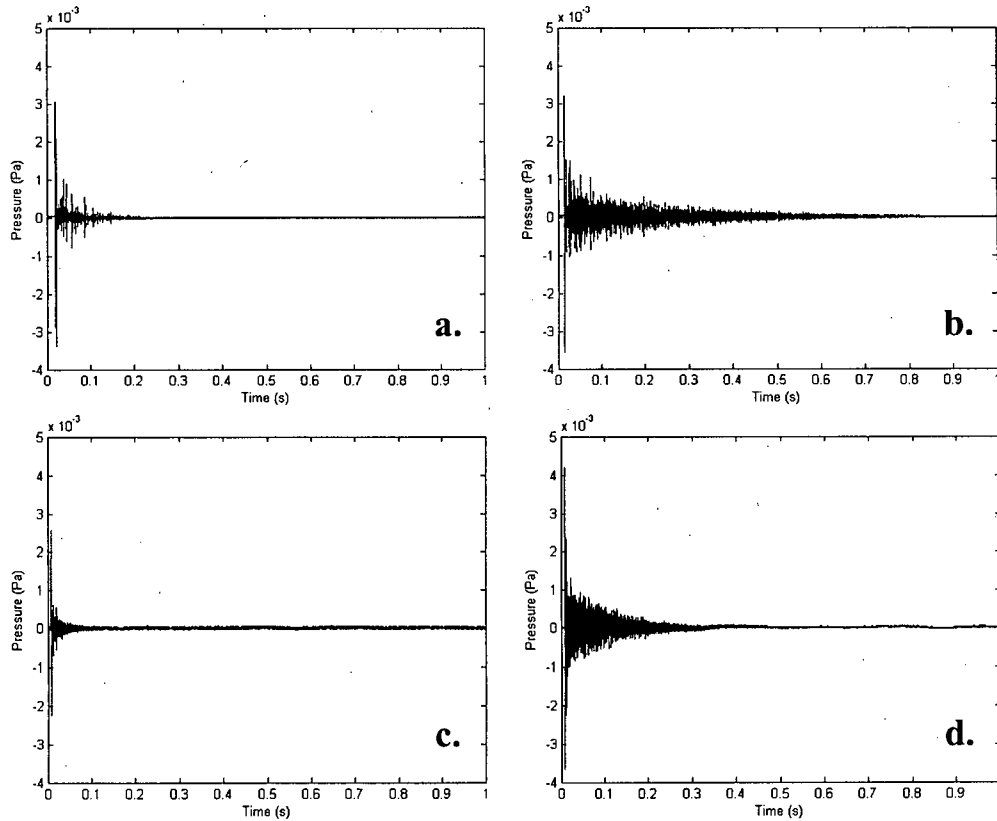
Figure 2. 7. Room impulse responses at r2 (a: Predicted Room A; b: Predicted Room B; c: Measured Room A; d: Measured Room B).

38

The room impulse responses (Figure 2.7) provide information about sound-field diffuseness as well as the early-to-late energy ratio. A number of strong individual peaks appear in the predicted room impulse responses in both Room A and Room B. Thus, the predicted rooms have more specular reflections than in the measured rooms. This suggests that the predicted rooms should have had more diffusion (i.e., higher diffusion coefficients).

For the reverberant classroom, Room B, the real classroom speech-intelligibility scores were higher than the virtual classroom test scores, as in Kleiner's results [2]. He found that speech-intelligibility results using simulated sound fields with RTs 0.8 to 1.0 s were worse than for direct listening in the theatre tested. The standard deviation of the mean speech-intelligibility scores also supported the conclusion that the auralization technique was adversely influenced by the reverberation and noise, as shown in Table 2.4. It is difficult to localize sound in a reverberant room or in noise. Sound localization is adversely affected by reverberation [18,19]. Since Good and Gilkey [20] found adverse effects of noise in localization, it has been studied by many researchers [21,22,23,24]. However, this issue remains unsolved. It may be also caused by experimental errors associated with on-site listening testing in real classrooms. For example, in the real-classroom tests, listeners can move their heads to change the signal levels they receive, whereas in the virtual-classroom tests using headphones, the test-signal levels were kept constant.

HRTFs are important to enhance spatial reproduction in an auralized sound field [25]. Wenzel *et al.* [26,27,28] compared individual HRTFs and non-individual HRTFs. The data showed that, while the interaural cues to horizontal location were robust, the spectral cues were distorted by a synthesis process that used non-individualized HRTFs. However, many listeners were able to obtain at least some useful directional information from an auditory display, without requiring the use of individualized HRTFs. The relatively small standard deviations of the mean speech-intelligibility scores in the virtual classrooms suggest that considering differences in the subjects' HRTFs may not be necessary, and that the effect of headphone distortion is negligible.

It was necessary to perform convolution twice for each configuration, since we had two different sound signals — the speech and the babble noise. Calibration of the convolved

sound levels relied on CATT-Acoustic in this study. Absolute calibration was difficult to achieve due to the various scaling factors involved in the calculation [14]. It should be possible to improve the accuracy of the sound-level calibration using the relative level difference between the speech and the babble noise.

In the high-absorption classroom, Room A, the differences between the real and virtual classrooms may not only be due to limitations in the auralization techniques. It may be caused by experimental errors in real-classroom measurements due to distortion added by the electro-acoustic equipment. As shown in Figures 2.6a and b, the speech intelligibility in Room A without noise did not correlate with the two objective speech-intelligibility metrics, C_{50} and speech-to-noise level difference. Contrastingly, in the configuration which had differences between the real and virtual classrooms (Room B with babble noise), the speech-intelligibility scores were well correlated with both C_{50} and speech-to-noise level difference.

2.5   Summary and Conclusions

This comparison study showed that, if the room to be auralized is not too absorptive (e.g., $RT = 0.4$ s at the center of the room) or noisy (e.g., $SNS = 0$ dB), speech-intelligibility tests using auralization are reliable.

The fundamental assumption of the auralization technique is that the room impulse response characterizes the room. The calculated impulse response is convolved with sound signals to obtain the sounds to be auralized. How close the predicted impulse response is to the measured impulse response is the key factor determining the accuracy of the auralization procedure. In this study, the results for the virtual classroom demonstrated limitations of current auralization techniques in certain conditions. Auralizing reverberant sound fields with noise is challenging. The low-absorption virtual classroom (Room B) did not agree well with the real classroom in terms of the speech-intelligibility test scores.

Mixing two different sound signals – the MRT signals and the babble noise – turned out to be another problem. In real sound fields, a listener can distinguish the sound information which he/she wants to listen to, from other noises. Sound directivity and spatial hearing can help this hearing procedure. However, in virtual classrooms, when the babble noise was presented in the low-absorption classroom (Room B), the MRT words were completely 'mixed up' with the babble noise.

In general, the prediction results showed good agreement for most room-acoustical parameters, except for C_{50}, even though the MRT test scores showed some disagreement.

References

[1]   B. Nordlund, T. Kihlman, and S. Lindblad, "Use of articulation tests in auditorium studies," J. Acoust. Soc. Am. **44** (1), 148-156 (1968).

[2]   M. Kleiner, "Speech-intelligibility In real and simulated sound fields," Acustica **47** (2), 55-71 (1981).

[3]   G. L. Ricard and S. L. Meirs, "Intelligibility and localization of speech from virtual directions," Human Factors **36** (1), 120-128 (1994).

[4]   J. Besing and J. Koehnke, "A Test of Virtual Auditory Localization," Ear Hear. **16** (2), 220-229 (1995).

[5]   J. Koehnke and J. M. Besing, "A procedure for testing speech intelligibility in a virtual listening environment.," Ear Hear. **17** (3), 211-217 (1996).

[6]   W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97** (6), 3907-3908 (1995).

[7]   J. Peng, "Feasibility of subjective speech intelligibility assessment based on auralization," Appl. Acoust. **66**, 591-601 (2005).

[8]   E. J. Kreul, J. C. Nixon, K. D. Kryter et al., "A proposed clinical test of speech discrimination," J. Speech Hear. Res. **11** (3), 536-552 (1968).

[9]   Auditec of St. Louis (St. Louis, MO, 2003).

[10]   B. -I. Dalenbäck, CATT-Acoustic v8.0 Users' Manual (Gothenburg, 2004).

[11]   M. D. Egan, *Architectural acoustics*. (McGraw-Hill, New York, 1988), pp.52-53.

[12]   M. Hodgson and E.-M. Nosal, "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms," J. Acoust. Soc. Am. **111** (2), 931-939 (2002).

[13]   F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85** (2), 868-878 (1989).

[14]   CATT, in *User's Manual* (Gothenburg, 2002), pp. 7-38-39.

[15] E. Zwicker and H. Fastl, *Psycho-acoustics: Facts and models*, 2nd ed. (Springer-Verlag, Heidelberg, 1999), pp.175-182.

[16] S. R. Bistafa and J. S. Bradley, "Predicting reverberation times in a simulated classroom," J. Acoust. Soc. Am. **108** (4), 1721-1731 (2000).

[17] J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. **106** (4), 1820-1828 (1999).

[18] C. Giguere and S. M. Abel, "Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay," J. Acoust. Soc. Am. **94** (2), 769-776 (1993).

[19] B. G. Shinn-Cunningham, N. Kopco, and T. J. Martin, "Localizing nearby sound sources in a classroom: Binaural room impulse response," J. Acoust. Soc. Am. **117** (5), 3100-3115 (2005).

[20] M. D. Good and R. H. Gilkey, "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99** (2), 1108-1117 (1996).

[21] K. S. Abouchacra, D. C. Emanuel, I. M. Blood et al., "Spatial perception of speech in various signal to noise ratios," Ear Hear. **19** (4), 298-309 (1998).

[22] C. Lorenzi, S. Gatehouse, and C. Lever, "Sound localization in noise in normal-hearing listeners," J. Acoust. Soc. Am. **105** (3), 1810-1820 (1999).

[23] C. Lorenzi, S. Gatehouse, and C. Lever, "Sound localization in noise in hearing-impaired listeners," J. Acoust. Soc. Am. **105** (6), 3454-3463 (1999).

[24] E. H. A. Langendijk, D. J. Kistler, and F. L. Wightman, "Sound localization in the presence of one or two distracters," J. Acoust. Soc. Am. **109** (5), 2123-2134 (2001).

[25] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*, Revised ed. (Hirzel Verlag, Stuttgart, 1996), pp.372-392.

[26] E. M. Wenzel, F. L. Wightman, and D. J. Kistler, "Localization with non-individualized virtual display cues," Proc. SIGCHI (1991).

[27] E. M. Wenzel, M. Arruda, D. J. Kistler et al., "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94** (1), 111-123 (1993).

[28] E. M. Wenzel and D. R. Begault, "Are individualized head-related transfer functions required for auditory information displays?" J. Acoust. Soc. Am. **105** (2), 1035 (1999).

3   OPTIMUM REVERBERATION TIMES FOR SPEECH INTELLIGIBILITY FOR NORMAL AND HEARING-IMPAIRED LISTENERS IN IDEALIZED CLASSROOMS WITH DIFFUSE SOUND FIELDS *

3.1   Introduction

Verbal communication is one of the most important acoustical activities in rooms, from small meeting rooms and classrooms to larger auditoria and conference rooms. The acoustical designs of such rooms should therefore achieve a high degree of speech intelligibility for listeners. Speech intelligibility is directly related to speech-to-noise level difference, and is inversely related to reverberation time [1]. However, in rooms the situation is complicated by the fact that reverberation and steady-state levels of speech and noise interact. Increased reverberation increases both speech and noise levels by increasing the reverberant sound energy; this increases or decreases the speech-to-noise level difference at a listener position depending on the listener's relative distances to the speech and noise sources. Here we consider rooms with approximately diffuse sound fields, for which reverberation can be accurately described by the reverberation time (RT), so that the results can be related to previous experimental research.

The literature reports a number of experimental and theoretical studies which investigated the relationship between the prevailing acoustical conditions and resulting speech intelligibility. These studies accounted for noise and the interaction between reverberation and speech-to-noise level difference, with varying degrees of realism. A brief overview of the literature is presented here - see Ref. 2 for a full review and discussion.

Nabelek and Robinson [3] showed that, in the absence of noise, speech intelligibility was inversely related to reverberation time - that is, the optimal reverberation time for speech intelligibility was zero. Nabelek and Pickett [4,5] and Finitzo-Hieber and Tillman [6] performed speech-intelligibility tests with normal-hearing and hearing-impaired subjects, for various fixed speech-to-noise ratios in rooms with various fixed reverberation times, again finding that speech intelligibility decreased with increased reverberation time. Hearing-impaired people were more sensitive to reverberation than normal-hearing people. However, these experimental studies were unrealistic in effectively assuming a diffuse sound field by involving exponential sound decays, and in not accounting for the interaction between reverberation and sound levels. The theoretical studies, on the other hand, were based on speech-intelligibility metrics which are considered to be good predictors of speech intelligibility, and which account for the interaction. Bradley [1], using both subjective test results and theoretical prediction of the U_{50} useful-to-detrimental energy metric in diffuse sound fields, found optimum reverberation times of 0.4 to 0.5 s for classrooms with a uniform background-noise level of 30 dBA. Houtgast et al. [7] used a numerical model and found non-zero optimum reverberation times for a variety of rooms with non-diffuse sound fields, and for various speech-to-noise level differences. The effects of reverberation on noise were incorporated by considering the audience as a collection of individual noise sources. Bistafa and Bradley [8] used a theoretical model and predicted non-zero optimal reverberation times in diffuse sound fields using a number of metrics. They found that increased reverberation increased early energy and intelligibility, but that too much reverberation decreased intelligibility. However, noise levels were again unrealistically assumed to be uniform throughout the room. Hodgson and Nosal [2] incorporated noise into a theoretical model in a realistic manner, by including noise sources in the rooms, which had diffuse sound fields. Predicted optimum reverberation times depended on the source directivities, the speech-to-noise level difference at the listener's position, the positions and orientations of the speaker and the noise source, and the number of noise sources. They found that if the speech source was farther from the listener than the noise source, then speech levels increased more with reverberation than did noise levels, and the level difference increased with reverberation, tending to increase intelligibility. If, on the other hand, the noise source was farther from the listener than the speech source, then noise levels increased

more with reverberation than did speech levels, tending to decrease intelligibility. Thus, the effect of reverberation on intelligibility depended on the relative distances from the listener to the speech and the noise sources. The effect increases with source/receiver distance and, therefore, is greater when source/receiver distances are larger than the reverberation radius (or critical distance).

In this paper, auralization techniques are used to identify optimal reverberation times in an idealized classroom with speech and babble-noise sources and an approximately diffuse sound field, in order to validate theoretical prediction. Considering individual speech and noise sources is more realistic. Subjective test has the potential to be more accurate than theoretical prediction, since it directly reflects listeners' perception. Optimal reverberation times are found by performing speech-intelligibility tests with normal-hearing and hearing-impaired adult subjects. Considering the U_t speech-intelligibility metric that has been shown to be well-suited to the prediction of speech intelligibility in classrooms [1], the early/late energy time t that best predicts speech intelligibility is found for both subject groups. Involving both normal- and hearing-impaired subjects allows similarities and differences between these two subject groups to be determined.

3.2   Theoretical Considerations

U_t is a metric based on the useful-to-detrimental energy-ratio concept. This concept divides acoustical energy received after the arrival of the direct sound into useful and detrimental parts. The useful part consists of the direct energy from the speaker, E_d, and the early-arriving, reflected energy from the speaker, E_e. The remaining reflected, or late-arriving, energy, E_l, is considered detrimental. In addition to the late-arriving reflected energy, noise energy, E_n, is detrimental. Thus, the measured useful-to-detrimental ratio calculated from measured data is defined as,

$$U_{t,m} = 10\log\left\{\frac{E_d + E_{e,t}}{E_{l,t} + E_n}\right\} \ \text{dB}, \tag{1}$$

The U_t useful-to-detrimental ratio can also be predicted based on diffuse-field theory [2],

$$U_{t,p} = 10\log\left\{\frac{(r_{h_s}^2/r_s^2)+1-e^{-kt}}{e^{-kt}+10^{(L_{nf1}-L_{sf1})/10}(r_{h_s}^2/r_n^2+q_s/q_n)}\right\} \text{ dB,} \qquad (2)$$

where the subscripts 'm' and 'p' refer to measured and predicted data, respectively. Here $k = \ln(10^6)/RT$, where RT is the reverberation time, r_s is the distance from the speech source to the listener, r_n is the distance from the noise source to the listener, and r_{h_s} is the reverberation radius (or critical distance) associated with the speech source. L_{sf1} and L_{nf1} are the long-term anechoic levels at 1 m directly in front of the speech and noise sources, in a free field, respectively; q_s and q_n are their directivity indices.

3.3   Experimental Methodology

3.3.1   Classroom and sound-field simulation procedures '

In this study, one of the objectives was to model an idealized room with an approximately diffuse sound field and exponential sound decay. Thus, the design was based on previous research into the factors that relate to a diffuse sound field [9]. The virtual classroom was based on a real 95-seat classroom of simple, rectangular geometry. It was 11-m long, 7-m wide and 5-m high (volume = 385 m^3). Predictions were made in octave bands from 125 to 4000 Hz. The same absorption coefficients were used for all octave bands and for all surfaces. The absorption coefficient was varied to achieve different reverberation times. Values of 1.0, 0.68, 0.40 and 0.21 were used, respectively, to obtain reverberation times of 0.0, 0.2, 0.4 and 0.8 s. The corresponding reverberation radii varied from 3 to 1 m. In order to avoid strong specular reflections from the surfaces, and to promote diffuse fields with exponential sound decays, all surfaces were defined to be 30 % diffusely reflecting [10]. CATT-Acoustic v8.0 [11] was used to predict and auralize the sound fields. The number of rays and the truncation time were 10,088 and 1.0 s, respectively, for both prediction and auralization. In order to verify the diffusiveness of the simulated sound fields, predicted EDT's and RT's were compared; these should be very similar in a diffuse sound field with an exponential sound decay. The differences were always less than 0.05 s at mid frequencies (500, 1000, and 2000 Hz).

The classroom contained a speech source, a noise source and a virtual listener, all located at least 2 m from surfaces. Two noise-source positions were considered – one between the speech source and the listener, such that the noise source was closer to the listener than the speech source (Noise 1), and one farther from the speech source than the listener (Noise 2). Figure 3.1 shows the floor-plan of the virtual room, with the relative positions of the listener, the speaker, and the noise sources. The speech source had human-like directivity, the noise source was omnidirectional. The listener and the speech source faced each other. The relative output power levels of the speech and noise sources were chosen to give differences (*SNS*) of 0 and 5 dB. Note that *SNS* is different from the difference in the speech and noise levels at a receiver location (*SNR*). Table 3.1 lists the *SNR* for both *SNS* values. The values of *SNS* and *RT* were selected on the basis of preliminary listening tests. These covered a wide range of *SNS* and *RT* values and allowed the more limited ranges, resulting in realistic *SNR* values and expected to contain the optimal *RT* values which were used here to be identified. The test *RT* values were additionally chosen to cover the range including zero and the optimal values specified in classroom standards [1].

Binaural impulse responses between the listener and the speech and noise sources were predicted. In order to take into account the effects of the head, shoulder and external auditory systems of the virtual listener, the head-related transfer functions (HRTFs) provided with the CATT-Acoustic system were used. Headphone playback without equalization, and diffuse-field HRTF data, were used in this work.

The resulting sound fields were auralized using the CATT-Acoustic software. The speech-signal level was chosen to approximate typical classroom levels. The noise-source levels were set relative to the input speech level to achieve the two test *SNS* values. A total of 16 different sound-field configurations were created, consisting of all combinations of the two speech- and noise-source relative output levels (*SNS* = 0 and 5 dB), the four reverberation times (*RT* = 0, 0.2, 0.4, and 0.8 s), and the two positions of the noise source (Noise 1 and Noise 2). As shown in Table 3.1, the speech-to-noise level differences received at the listener position (*SNR*), corresponding to the two *SNS* values, varied in the various configurations from -6 dB to +8.5 dB.

47

Table 3. 1. Received speech-to-noise level differences (*SNR* in dB) for all test sound-field configurations.

RT(s)	SNS = 0 dB		SNS = 5 dB	
	Noise 1	Noise 2	Noise 1	Noise 2
0.0	-6.0	3.5	-1.0	8.5
0.2	-4.5	2.1	0.4	7.2
0.4	-2.8	1.2	2.1	6.2
0.8	-1.5	0.6	3.3	5.7
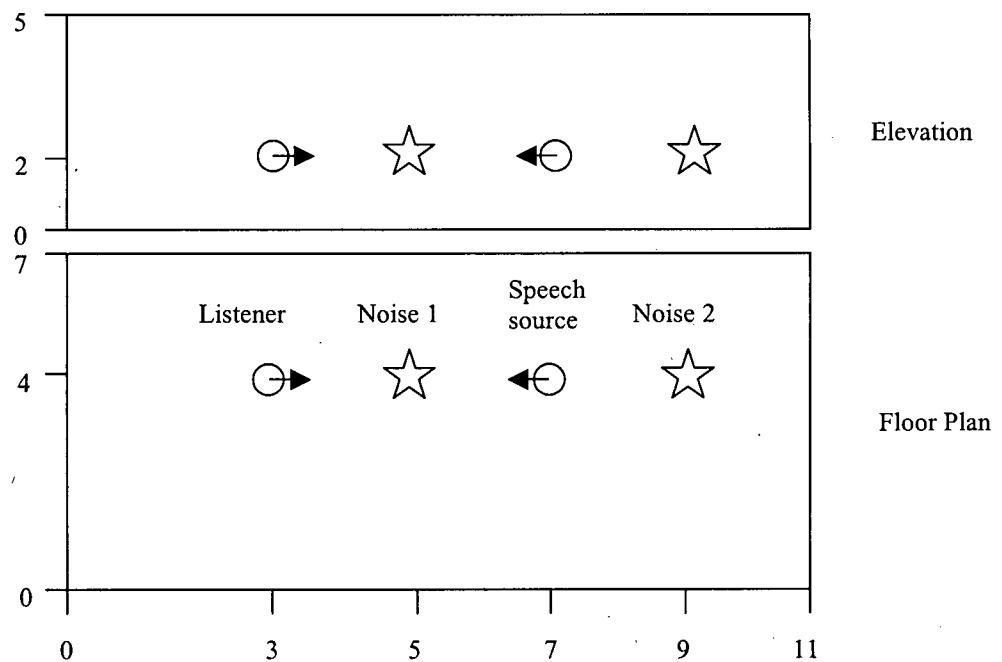
Figure 3. 1. Floor-plan and elevation of the virtual classroom, showing the speaker, listener and noise-source positions. All coordinates are in metres.

The Modified Rhyme Test (MRT) [12] was used as the speech-intelligibility test method. Twelve, fifty-word MRT word lists recorded by a male, native-Canadian talker in an anechoic chamber were combined through the CATT-Acoustic system with four-talker babble noise (available from AUDITECH [13]). The MRT-speech and babble-noise signals for each test configuration were mixed together using the Goldwave v5.1 sound-editing program [14], at levels corresponding to the predicted sound-pressure levels at the listener position. The resulting, final auralization test materials were transferred to a compact disc for presentation to subjects using a CD player. The test material was replayed through Sony MDR V600 headphones in a soundproof room. Each subject was tested individually. Each listened to a complete list of 50 words for each of the 16 different sound-field configurations. In order to avoid score inflation caused by the closed-set method used here, subjects were instructed not to guess the answer. The tests were presented in randomized order. The presentation levels were set by the predicted levels of the sound-field configurations, and used for both normal and hearing-impaired groups.

3.3.2 Subjects

Hearing-screening tests were done prior to the speech-intelligibility testing, to identify the hearing categories of the subjects. Subject groups for the study were normal-hearing adults, and hearing-impaired adults with a mild to moderate sensorineural hearing loss, whose first language was English. The hearing-loss criteria for the hearing-impaired subjects were lower than 25 dBHL (HL = Hearing Loss) between 250 Hz and 1 kHz, and between 30 and 55 dBHL from 2 to 8 kHz, with no more than 15 dB difference between the two ears at any two frequencies. This represents a typical frequency response for sensorineural hearing loss [15]. The hearing-impaired subjects in this work did not use hearing aids in their everyday lives.

Data collection was done at two different sites: the University of British Columbia, Vancouver, BC ('UBC') and Central West Health, Grand Falls-Windsor, NL ('CWH'). Forty-three normal-hearing and twenty-eight hearing-impaired subjects, with mean ages of 26 and 48 years, respectively, completed the tests. For the normal-hearing subjects, the difference between the UBC and CWH groups was not statistically significant. For the

hearing-impaired subjects, the difference was statistically significant ($p < 0.05$), with the CWH hearing-impaired group showing a lower average MRT score than the UBC group. However, the subject groups at the two locations showed similar variations of scores with the different test configuration (*i.e.* for different *SNS* and *RT*); thus the results for the two groups were combined. The exact reason for the difference in the results for the hearing-impaired subjects at the two test sites is not known.

3.4 Results

3.4.1 Speech Intelligibility

Mean speech-intelligibility scores for the sixteen sound-field configurations were calculated separately for the normal-hearing and hearing-impaired subject groups. Results are presented as the percentage of correct responses. Figure 3.2 shows the variations of the mean speech-intelligibility score (with 95 % confidence interval) with reverberation time and noise-source position.

For normal-hearing subjects, when the noise source was farther from the listener than the speaker (Noise 2), with either *SNS* = 5 dB or 0 dB, the mean speech-intelligibility scores exceeded 85 % for all reverberation times. Analysis of variance (ANOVA) was employed to compare the sensitivity of the scores to variations in the speech- and noise-source output-level difference (*SNS*) and reverberation time (*RT*). The mean scores for the four different *RT*'s were found to be statistically different ($p < 0.0005$). The mean scores at each *RT* were ranked using Tukey's paired comparison test [16] ($\alpha = 0.05$). The rank orders by the mean speech-intelligibility scores were varied in inverse relation the the *RT* except for two pairs for which scores were not statistically different: *RT* = 0.0 and 0.2 s with *SNS* = 5 dB, and *RT* = 0.4 and 0.8 s with *SNS* = 0 dB. With *SNS* = 5 dB, for *RT*'s of 0.2 and 0.4 s, the mean speech-intelligibility scores with *SNS* = 5 dB were significantly higher than those with *SNS* = 0 dB ($p < 0.001$). However, the difference at *RT* = 0.8 s was not significant, although the mean speech-intelligibility scores with *SNS* = 5 dB were lower than those with *SNS* = 0 dB. When *RT* was 0.0 s there was no statistical difference between *SNS* = 5 dB and *SNS* = 0 dB ($p = 0.554$).

For hearing-impaired subjects, with Noise 2, the mean speech-intelligibility scores were between 68.3 and 80.0 % with SNS = 5 dB, and between 64.3 and 72.5 % with SNS = 0 dB. The differences between the scores for the normal-hearing and hearing-impaired subjects were 13.3 to 19.5 % with SNS = 5 dB, increasing to 17.8 to 22.7 % with SNS = 0 dB. The ANOVA results for the difference in the mean speech-intelligibility test scores between the normal-hearing and hearing-impaired subjects indicated a statistically significant difference (p < 0.0005). For the hearing-impaired subjects in this case, the 95% confidence intervals overlapped at the different RTs; therefore, statistical confidence in the rank order was lower. The peak (*i.e.* the locally highest value, possibly not statistically significant) occurred at RT = 0.2 s with SNS = 5 dB, and its confidence interval was relatively narrow; with SNS = 0 dB, the highest mean speech-intelligibility score was at RT = 0.0 s. There were no significant differences between the scores with SNS = 5 dB and 0 dB, except in the case of RT = 0.2 s (p < 0.05).

For normal-hearing subjects, when the noise source was positioned between the listener and the speaker (Noise 1), with SNS = 5 dB, mean speech-intelligibility scores varied from 80.4 to 88.3 %. There were two peaks in the score, at RT = 0.0 and 0.4 s; the difference was not statistical significant according to their rank order by Tukey's pairwise comparison test (α = 0.05). For the hearing-impaired subjects in the same conditions, mean speech-intelligibility scores varied from 61.7 to 67.6 %. Of the values tested, the highest score occurred at RT = 0.8 s. The differences between the scores for the normal-hearing and hearing-impaired subjects varied from 12.8 to 22.3 % with SNS = 5 dB, and the differences were statistically significant.

In the case of Noise 1, with SNS = 0 dB, mean speech-intelligibility scores increased with increasing reverberation time for both normal- and hearing-impaired groups. Of the RTs tested, the mean speech-intelligibility scores had its highest value at RT = 0.8 s for both subject groups. The mean scores for the four RT's were again ranked using Tukey's paired comparison test (α = 0.05), and were found to vary directly with RT, with 99.9 % confidence.
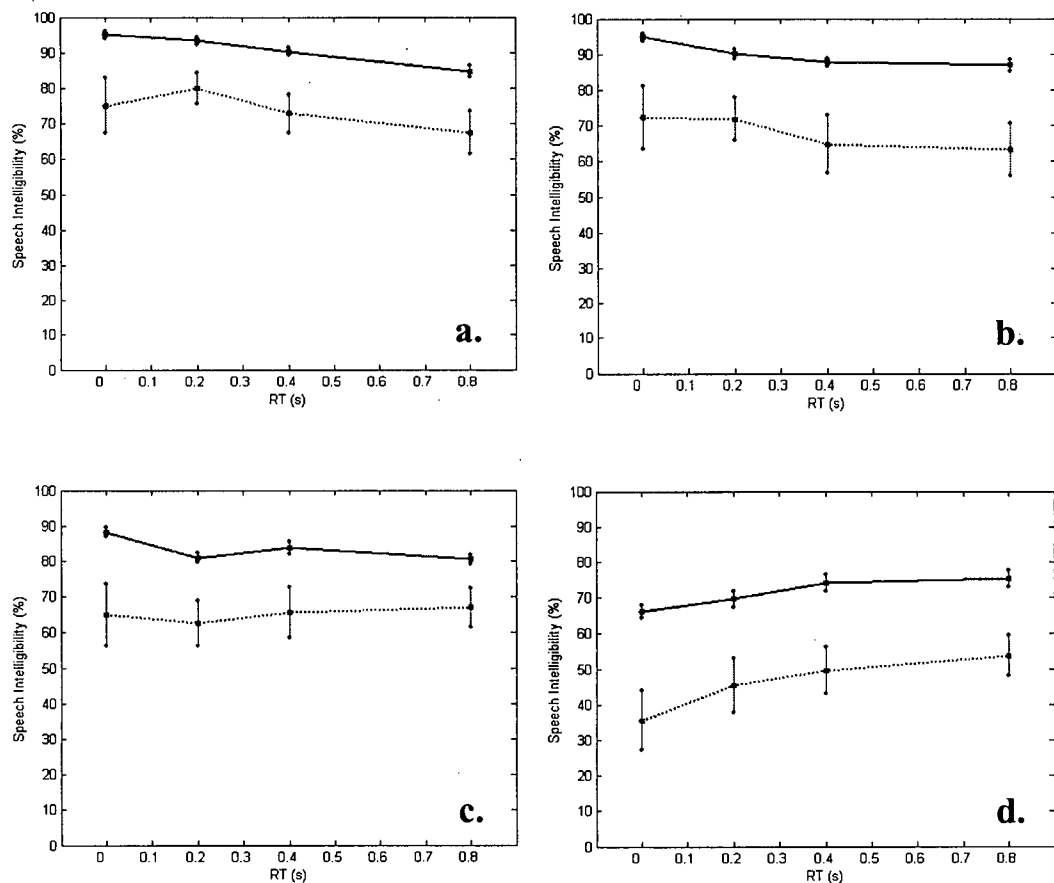
Figure 3. 2. Variation of mean speech-intelligibility score, and 95% confidence interval, with RT: a. *SNS* **= 5 dB, Noise 2; b.** *SNS* **= 0 dB, Noise 2; c.** *SNS* **= 5 dB, Noise 1; d.** *SNS* **= 0 dB, Noise 1). (━━━━) normal-hearing; (.............) hearing-impaired.**

The lowest mean speech-intelligibility scores measured in the various test configurations varied between 66.2 and 75.3 % for normal-hearing subjects, and between 36.7 and 55.1 % for hearing-impaired subjects. The differences between the scores for the two subject groups varied from 20.2 to 29.5 %, the biggest differences seen in the four cases. The difference between the results for *SNS* = 0 and 5 dB was statistically significant ($p < 0.0005$) for both subject groups.

52

3.4.2 Best predicting early-time limit

For each sound-field configuration, useful-to-detrimental ratios were calculated from the predicted impulse responses and the applicable speech and noise levels, according to Eq. (1). They were also predicted using Eq. (2) for comparison with theory. Early-time limits of $t = 20, 30,..., 120$ ms were used. In order to identify the early-time limits which best predicted the measured speech intelligibility, regression analyses were performed on the mean speech-intelligibility scores for each sound-field configuration. Since the relationships were clearly not linear or quadratic, and following Bradley [17], third-order polynomials were fit.

Table 3.2 shows the strengths of the relationships – quantified by the goodness-of-fit measure, R^2 – between each measure and speech intelligibility, for both the normal- and hearing-impaired results.

Table 3. 2. Coefficients of determination (R^2) associated with third-order-polynomial regression fits for each U_t value, for both normal-hearing and hearing-impaired subjects. The highest values are in bold.

U_t	Normal hearing		Hearing-impaired	
	Measured	Predicted	Measured	Predicted
U_{20}	74.7	72.6	70.2	66.4
U_{30}	81.1	80.0	81.4	78.6
U_{40}	83.8	84.3	88.6	86.6
U_{50}	**85.2**	86.0	92.6	91.0
U_{60}	84.7	**86.4**	93.9	93.3
U_{70}	83.9	86.2	**94.5**	94.5
U_{80}	83.2	85.7	**94.5**	95.0
U_{90}	82.4	85.1	**94.5**	**95.3**
U_{100}	81.9	84.4	94.4	**95.3**
U_{110}	81.3	83.8	94.3	95.2
U_{120}	80.8	83.3	94.1	95.1

Since the form of the trendline and the number of data points was the same in every case, the success of each measure can be compared by comparing the corresponding R^2 values. In both calculation and prediction results, the trends were similar. $U_{50,m}$ and $U_{60,p}$ were most accurate at predicting the speech-intelligibility results for the normal-hearing subjects. However, for the hearing-impaired subjects, $U_{70,m}$, $U_{80,m}$ and $U_{90,m}$ and $U_{90,p}$ and $U_{100,p}$ predicted the results best; that is, with early-time limits 20 - 40 ms higher than for the normal-hearing group. That the limit was higher in prediction than in measurement may be due to the fact that the sound fields in the virtual rooms were not perfectly diffuse as assumed in prediction. The $U_{50,m}$ (for normal hearing, 'NH') and $U_{80,m}$ (for hearing-impaired, 'HoH') regression curves are shown in Figure 3.3. The corresponding equations are as follows:

Normal hearing:

$$SI_{NH,m} = 85.2 + 1.94 \cdot U_{50,m} - 0.167 \cdot U_{50,m}^2 + 0.00911 \cdot U_{50,m}^3 \qquad R^2 = 85.2\,\%$$

$$SI_{NH,p} = 85.6 + 1.77 \cdot U_{60,p} - 0.176 \cdot U_{60,p}^2 - 0.012 \cdot U_{60,p}^3 \qquad R^2 = 86.4\,\%$$

Hearing-impaired:

$$SI_{HoH,m} = 64.1 + 2.86 \cdot U_{80,m} - 0.232 \cdot U_{80,m}^2 + 0.010 \cdot U_{80,m}^3 \qquad R^2 = 94.5\,\%$$

$$SI_{HoH,p} = 64.7 + 2.66 \cdot U_{100,p} - 0.245 \cdot U_{100,p}^2 - 0.013 \cdot U_{100,p}^3 \qquad R^2 = 95.3\,\%$$
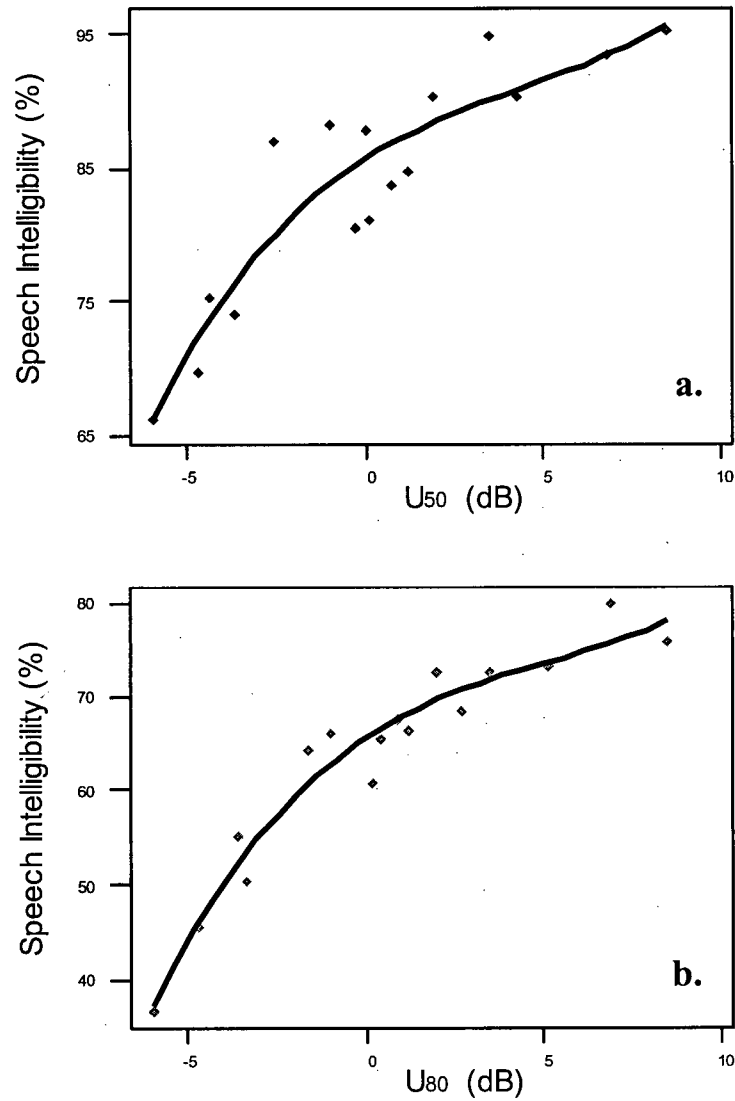
Figure 3. 3. Variation of mean speech-intelligibility score with useful-to-detrimental ratio using the best-fit early-time limit for all sound-field configurations and the best-fit third-order polynomial regression curves: a. normal-hearing, $U_{50,m}$; b. hearing-impaired, $U_{80,m}$.

3.5 Discussion

The speech-intelligibility results for the normal-hearing and hearing-impaired subjects obtained with the Modified Rhyme Tests revealed some basic differences in perception by the two groups, and also some similarities.

Normal-hearing subjects in our study showed decreased speech intelligibility with increased reverberation time when the speech source was closer to the listener than the noise source (Noise 2). When the noise source was between the speech source and the listener (Noise 1), of the RTs tested, the optimal reverberation time varied from 0.4 to 0.8 s when the SNS varied from 5 to 0 dB, except for normal-hearing subjects with $SNS = 5$ dB, for which the highest score was obtained with $RT = 0$ s. Except for this result, the results are in good agreement with those of Hodgson and Nosal [2] (who also proposed a detailed explanation for their results).

In general, hearing-impaired subjects in our study showed similar trends to the normal-hearing subjects. This is consistent with recent work by Bradley, Sato and Picard [18]. Of course, the hearing-impaired listeners were more adversely affected by reduced speech-to-noise level difference. Increased reverberation time generally resulted in increased speech intelligibility when the noise source was closer than the speech source to the listener (Noise 1). Increased reverberation time above $RT = 0.2$ s decreased speech intelligibility when the speech source was closer to the listener than the noise source (Noise 2). When the noise source was farther from the listener than the speech source, the optimal reverberation time included zero and low, non-zero values. The results for the hearing-impaired subjects had relatively large standard deviations on their mean speech-intelligibility scores. The greatest standard deviations always occurred at $RT = 0$ s among the four reverberation times. The smallest standard deviations always occurred at the RTs resulting in the peak mean speech intelligibility scores – *i.e.* at the optimal reverberation times.

When the noise source was farther from the listener than the speech source, the difference in score resulting from the two different speech- and noise-source output level

differences (*SNS*) was smaller than those when the noise was between the speech source and the listener, for both the normal- and hearing-impaired subject groups.

3.6   Conclusions

The results of this work generally support previous theoretical predictions [2]. With the noise source incorporated in a realistic manner, the optimal reverberation times were dependent on its positions relative to the speaker and the listener in a room. The optimal reverberation time was zero or near zero when the noise source was farther than the speaker; zero and non-zero reverberation times were found to be optimal when the noise source was between the listener and the speaker. If the speech-to-noise level difference is adverse for a subject group, some reverberation is required to increase the speech signal. The best early-time limit in the useful-to-detrimental energy ratio was 50 - 60 ms for normal-hearing subjects. U_{70}, U_{80} or U_{90} were the most accurate predictors of the mean speech-intelligibility score. Hearing-impaired subjects apparently require more early energy than normal-hearing subjects with this range of the speech-to-noise level difference.

In this study, a simple, idealized classroom with approximately diffuse sound field and exponential sound decay was studied, as was the case in previous experimental work reported in the literature. Thus, the results depended on the overall reverberation in the room. They also depend on the source and receiver locations involved; Ref. 2 contains further discussion of the effect of varying these parameters. Ignored in this study is the influence of detailed room-acoustical factors such as individual reflections from the wall, floor or ceiling; these relate to the exact room geometry and surface-absorption distribution, and exist in realistic rooms. It would be interesting to repeat the study using a more realistic model of a classroom to improve the current work. The optimal reverberation might be found to vary from room to room depending, for example, on details of the arrival of reflections at the receiver. It would also be interesting to include replay-headphone equalization and angularly-varying HRTF data in the simulations, though the relatively small standard deviations of the mean speech-intelligibility scores among the normal-hearing subjects suggests that considering differences in the subjects' HRTFs may not be necessary, and that the effect of headphone distortion is negligible.

References

[1]   J. S. Bradley, "Speech intelligibility studies in classrooms," J. Acoust. Soc. Am. **80**(3), 846-854 (1986).

[2]   M. R. Hodgson and E.-M. Nosal, "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms," J. Acoust. Soc. Am. **111**(2), 931-939 (2002).

[3]   A.K. Nabelek and P.K. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**(5), 1242-1248 (1982).

[4]   A.K. Nabeleck and J.M. Pickett, " Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners," J. Speech Hear. Res. **17**(4), 724-739 (1974).

[5]   A.K. Nabelek and J.M. Pickett, "Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids," J. Acoust. Soc. Am. **56**(2), 628-639 (1974).

[6]   T. Finitzo-Hieber and T.W. Tillman, "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," J. Speech Hear. Res. **21**(3), 441-458 (1978).

[7]   T. Houtgast, H. J. M. Steeneken and R. Plomp, "Predicting speech intelligibility in rooms from the modulation transfer function. II. Mirror image computer model applied to rectangular rooms," Acustica **46**(1), 73-81 (1980).

[8]   S.R. Bistafa and J.S. Bradley, "Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics," J. Acoust. Soc. Am. **107**(2), 861-875 (2000).

[9]   M. R. Hodgson, "When is diffuse-field theory applicable?," Appl. Acoust. **49**(3), 197-207 (1996).

[10]   M. Hodgson, "Evidence of diffuse surface reflections in rooms," J. Acoust. Soc. Am. **89**(2), 765-771 (1991).

[11]   B. -I. Dalenback, CATT-Acoustic v8.0 Users' Manual (Gothenburg, 2004).

[12]   E. J. Kreul, J. C. Nixon, K. D. Kryter et al., "A proposed clinical test of speech discrimination," J. Speech Hear. Res. **11**(3), 536-552 (1968).

[13]   Auditec of St. Louis, (St. Louis, MO, 2003).

[14]   Goldwave Inc., GoldWwave v5.0 (St. John's, NL, 2004).

[15]   D. Henderson, R. J. Salvi, F. A. Boettcher, and A. E. Clock, "Neurophysiologic correlates of sensory-neural hearing loss," J. Katz, Handbook of Clinical Audiology, Chapter 4, 4th Ed. Williams & Wilkins (1994).

[16]   W. Mendenhall, R. J. Beaver, and B. M. Beaver, *Introduction to Probability and Statistics*, 10th ed. (Duxbury Press, Belmont, CA, 1999).

[17]   J. S. Bradley, "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am. **80**(3), 837-845 (1986).

[18]   J. S. Bradley, H. Sato and M. Picard, "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**(6), 3233-3244 (2003).

4   OPTIMUM REVERBERATION FOR SPEECH INTELLIGIBILITY FOR NORMAL AND HEARING-IMPAIRED LISTENERS IN REALISTIC VIRTUAL CLASSROOMS USING AURALIZATION *

4.1   Introduction

Conceptually, speech intelligibility is directly related to signal-to-noise level difference and inversely related to the amount of reverberation [1]. However, situations involving rooms become complicated, since reverberation and steady-state levels interact so that they are not independent. Increased reverberation increases speech and noise levels by increasing the reverberant sound energy. Hodgson and Nosal [2] reviewed the effect of reverberation on speech and noise levels, and using theoretical prediction found that the effect of reverberation on speech intelligibility depends on the relative distances of the listener to the speech and the noise sources.

Regarding the optimization of reverberation for speech intelligibility, early reflections have been considered since Lochner and Burger [3] applied the concept of the useful-to-detrimental energy ratio. The useful-to-detrimental ratio was extended to account for the effect of fluctuating ambient background noise on speech intelligibility by Latham [4] for more accurate prediction. The useful-to-detrimental ratio is considered to be a suitable intelligibility metric for predicting speech intelligibility for normal-hearing listeners [1,5,6]. Typically, the time limit for the early-arriving reflections has been taken to be 50 ms for speech sounds [1,7]. The useful early-time limit for hearing-impaired listeners for speech intelligibility was unknown. Bradley, Sato and Picard [8] studied the effect of early reflection

---

for speech intelligibility for hearing-impaired listeners. They found that increased early-reflection energy has the same effect on speech-intelligibility scores as does an equal increase in the direct-sound energy for both normal and hearing-impaired groups.

In a previous paper [9], the authors used auralization to experimentally confirm Hodgson and Nosal's [2] prediction results, incorporating realistic noise sources into rooms with approximately diffuse sound fields. Subjective speech intelligibility testing with human subjects showed that hearing-impaired listeners require more early sound energy to achieve the benefit of increased speech level by reverberation than do normal-hearing listeners.

The objective of the present work was to expand and validate the previous work on speech intelligibility with diffuse sound fields into more realistic rooms with non-diffuse sound fields. This was done using speech-intelligibility tests with both normal and hearing-impaired listeners in auralized sound fields of existing classrooms and their variations. The test reverberation times were also extended to correspond to those of the existing classrooms, and to confirm the optimal reverberation times for speech intelligibility for both normal and hearing-impaired listeners in the case when the noise source was positioned between the talker and the listener; this was not clearly elucidated in the previous work, due to the limited values of the test reverberation times used. The second purpose of this work was to examine and compare widely used speech-intelligibility metrics. Useful-to-detrimental ratios and Speech Transmission Index (*STI*) were correlated with the speech-intelligibility test scores. The best predicting early time limits for useful sound energy were identified for both normal and hearing-impaired listeners.

4.2 Methods

4.2.1 Subjects

Subjects for the study were normal-hearing adults, and hearing-impaired adults with mild to moderate sensorineural hearing loss, whose first language was English. Twenty-five normal-hearing and thirteen hearing-impaired subjects participated in the tests; they varied in age from 18 to 64 years old, with mean ages of 24 and 41 years, respectively. Hearing-

screening tests were done prior to the speech-intelligibility testing, to categorize the hearing of the subjects. The subjects in the normal-hearing groups had pure-tone auditory thresholds less than 25 dBHL (HL = Hearing Loss) at octave frequencies from 250 to 8000 Hz [10]. The hearing-loss criteria for the hearing-impaired subjects were less than 25 dBHL between 250 Hz and 1 kHz, and between 30 and 55 dBHL from 2 to 8 kHz, in either ear. This represents a typical frequency response for sensorineural hearing loss [11]. The hearing-impaired subjects in this study did not wear hearing aids in their everyday lives.

4.2.2 Classroom configurations

In this study, realistic virtual classrooms were created with various reverberation times. Thus, the design was based on previous research and the acoustical conditions of the selected existing classrooms [9,12]. A configuration with zero reverberation time was excluded, since it is unrealistic. Six classroom configurations were defined to have reverberation times varying from 0.3 s to 1.9 s, by changing the amount of surface absorption.

Two typical medium-sized university classrooms were selected as models for the virtual classrooms. The two classrooms are architecturally identical, but have different acoustical characteristics. One has acoustical treatment to reduce the reverberation (RT_{mid} = 0.6 s at the listener position, L), and the other has no acoustical treatment (RT_{mid} = 1.9 s at L). Figure 4.1 shows a three-dimensional model of the virtual classroom. It has 96 seats, length = 15 m, width = 8.5 m, and height = 4 m (volume = 400 m^3, total surface area = 464 m^2, volume to surface-area ratio = 0.86 m) with a sloped floor in the seating area. There was a speech source, a noise source and a virtual listener in the classroom. The speech source (S) was positioned at the front of the classroom where the instructor might typically stand. The listener was in the middle of the classroom (source/receiver distance = 5.0 m). Speech-intelligibility tests were auralized in six classroom sound-field configurations having different reverberation times from 0.3 to 1.9 s. The corresponding reverberation radii of the speech source varied from approximately 4 to 1 m. The noise source (N) was positioned between the

speech source and the listener at 2.5 m from both. The speech source, the noise source, and the listener were on the center line of the classroom (see Figure 4.1).

Figure 4.1 shows six classroom configurations with different applications of absorbing material to vary the reverberation time. The sound-absorbing material used corresponded to 60-mm-thick glass fibre, with the following sound-absorption coefficients in the six octave frequency bands from 125 Hz to 4 kHz: 0.20, 0.57, 0.90, 0.98, 0.98, and 0.97. The absorption and diffusion coefficients of the surface materials are shown in Table 4.1. In Figure 4.2, configurations (0.6) and (1.9) correspond to the two existing classrooms, configuration (0.3) has more absorbing material on the ceiling and the walls, configuration (0.8) has no ceiling absorption, but the same amount of wall absorption as configuration (0.6), configurations (1.2) and (1.5) have the absorbing materials only on the side walls. Each configuration name indicates the predicted mid-frequency reverberation time at the listener position in the unoccupied classroom. The surface-diffusion coefficients were set based on the work of Hodgson [13]. Table 4.2 lists the octave-band average surface-absorption coefficients ($\bar{\alpha}$) and diffusion coefficients ($\bar{d}$) of the six classroom configurations.

Two different speech- and noise-source output-level differences (*SNS*) were tested: *SNS* of 0 and 4 dB, selected on the basis of preliminary predictions. The test values of *SNS* and *RT* covered a wider range of *RT* values than those in the previous study [9], and allowed the more limited ranges, resulting in realistic *SNR* values and expected to contain the optimal *RT* values which were used here, to be identified.
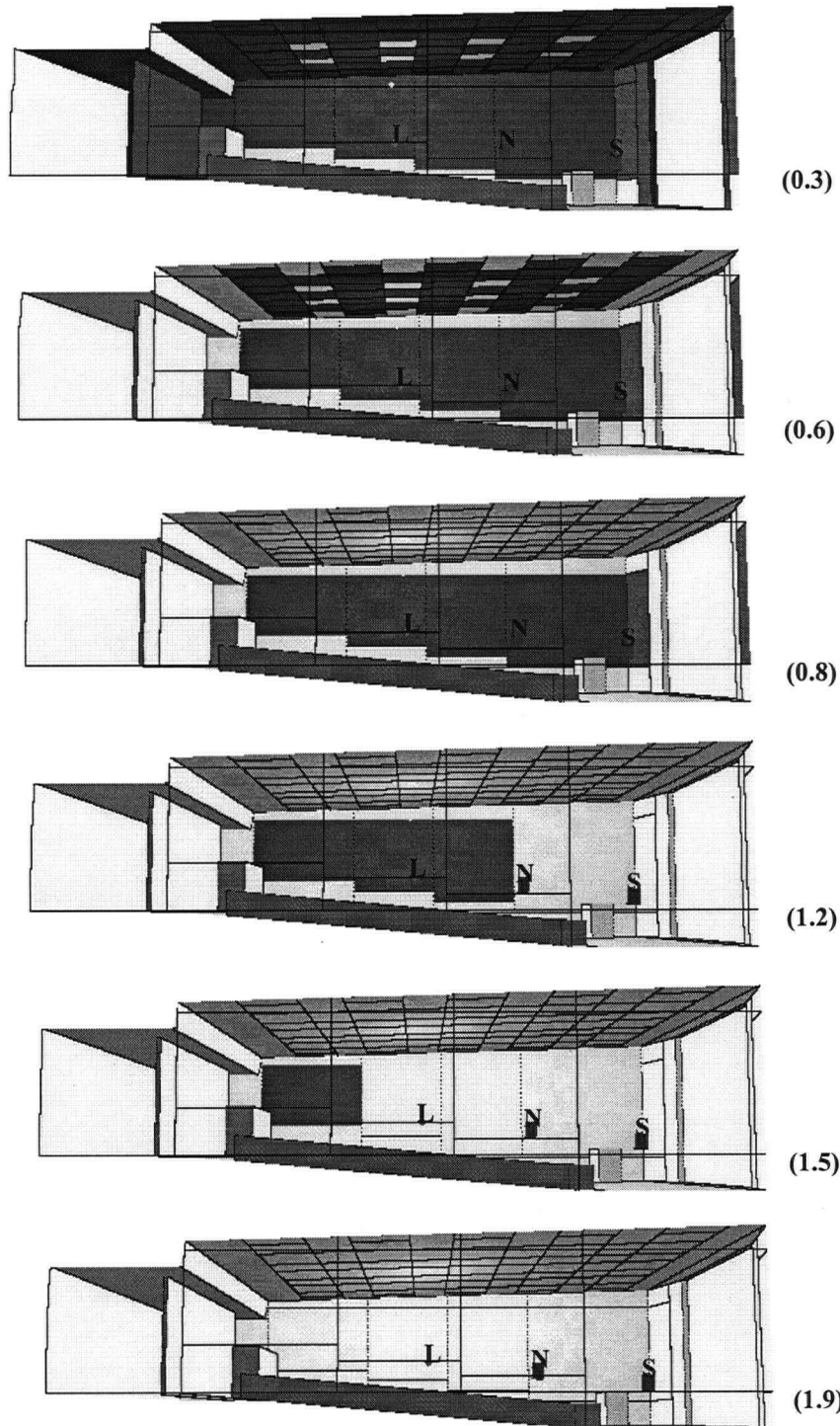
Figure 4. 1. Classroom configurations with sound-absorbing areas shown in grey (dark grey for sound absorbing material).

Table 4. 1. Absorption (α) and diffusion (d) coefficients of the surface materials.

material		125	250	500	1000	2000	4000
Glazed	α	0.02	0.02	0.02	0.01	0.01	0.01
concrete	d	0.05	0.05	0.05	0.05	0.05	0.05
Carpet	α	0.03	0.09	0.25	0.31	0.33	0.44
	d	0.06	0.07	0.10	0.15	0.15	0.20
Audience	α	0.13	0.16	0.15	0.13	0.18	0.25
	d	0.62	0.72	0.80	0.80	0.85	0.85
Painted	α	0.28	0.07	0.06	0.06	0.06	0.05
Plywood	d	0.03	0.03	0.02	0.01	0.01	0.01
Wood	α	0.14	0.10	0.06	0.08	0.10	0.10
	d	0.60	0.45	0.32	0.38	0.30	0.30
Blackboard	α	0.15	0.25	0.20	0.12	0.10	0.05
	d	0.10	0.05	0.04	0.04	0.03	0.03
Fibreglass	α	0.20	0.57	0.90	0.98	0.98	0.97
	d	0.39	0.42	0.90	0.98	0.95	0.95

Table 4. 2. Octave-band average surface-absorption coefficients ($\bar{\alpha}$) and diffusion coefficients ($\bar{d}$) of the six classroom configurations.

RT (s)		125	250	500	1000	2000	4000
0.3	$\bar{\alpha}$	0.17	0.35	0.54	0.58	0.59	0.61
	$\bar{d}$	0.32	0.34	0.60	0.65	0.64	0.65
0.6	$\bar{\alpha}$	0.18	0.22	0.29	0.31	0.32	0.32
	$\bar{d}$	0.26	0.27	0.39	0.42	0.41	0.41
0.8	$\bar{\alpha}$	0.19	0.15	0.18	0.19	0.20	0.20
	$\bar{d}$	0.22	0.22	0.28	0.29	0.28	0.28
1.2	$\bar{\alpha}$	0.19	0.12	0.13	0.13	0.14	0.14
	$\bar{d}$	0.19	0.19	0.22	0.23	0.22	0.22
1.5	$\bar{\alpha}$	0.20	0.10	0.09	0.09	0.10	0.10
	$\bar{d}$	0.18	0.17	0.18	0.18	0.18	0.18
1.9	$\bar{\alpha}$	0.20	0.08	0.07	0.07	0.08	0.08
	$\bar{d}$	0.17	0.16	0.16	0.16	0.16	0.16

4.2.3 Sound-field simulation and speech-intelligibility test procedure

CATT-Acoustic v8.0 [14] was used to predict and auralize the test sound fields. The number of rays involved in the predictions was 22,854 in each configuration and the truncation times varied from 1.0 s to 2.2 s depending on the target reverberation times in the classrooms.

Human speech directivity was associated with both the speech source and the noise source. The listener and the speech source faced each other, as in a real classroom situation. The noise source faced the front in the room. Predictions were made in six octave bands from 125 to 4000 Hz. Binaural impulse responses between the listener and both the speech and the noise sources were predicted. The head-related transfer functions (HRTFs) provided with the CATT-Acoustic system were used; theses take into account the effects of reflections from the pinnae and shoulders, as well as the shadowing effect of the head itself. Playback, with headphone transfer function compensation, with Beyer DT 990 Pro headphones, was used in this work.
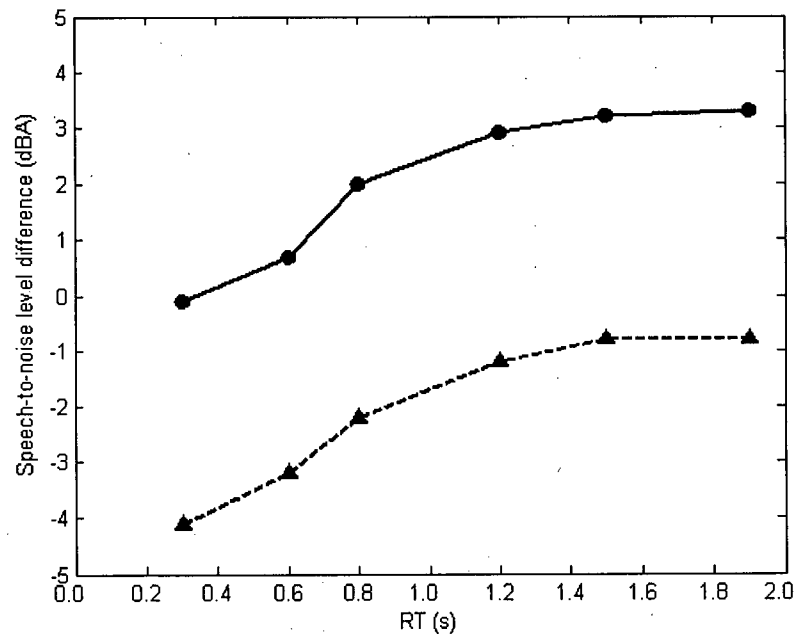
Figure 4. 2. Speech-to-noise ratio (SNR) at listener's location: (—●—) *SNS* = 4 dB; (---▲---) *SNS* = 0 dB.

The predicted sound fields were auralized using the CATT-Acoustic software. The speech-signal level was chosen to approximate typical classroom levels in the six octave frequency bands from 125 Hz to 4 kHz: 54.5, 60.6, 62.2, 59.3, 57.4, and 51.8 dB [15]. The noise-source levels were set relative to the input speech level to achieve the two test SNS values. A total of 12 different sound-field configurations were created, consisting of all combinations of the two speech- and noise-source relative output levels ($SNS = 0$ and 4 dB), and of the six reverberation times ($RT = 0.3, 0.6, 0.8, 1.2, 1.5$ and 1.9 s). As shown in Figure 4.2, the speech-to-noise level differences received at the listener position (SNR), corresponding to the two SNS values, varied in the various configurations from -4.1 dB to +3.3 dB.

Modified Rhyme Test (MRT) [16] word lists were used as the speech stimuli. The MRT consists of 300 words embedded in a carrier phrase, "Say /word/.", that are arranged in six lists each having fifty words: twenty five differ by the initial consonant and twenty five by the final consonant. The six lists were recorded by a male, native-Canadian talker in an anechoic chamber, and the 50-word orders in each list were randomized to obtain twelve lists for the twelve classroom configurations. The r.m.s. amplitudes of the lists were normalized to the same value. The MRT lists were combined through the CATT-Acoustic system with four-talker babble noise (available from AudiTech, St. Louis, MO [17]). The MRT-speech and babble-noise signals for each test configuration were mixed together using the Goldwave v5.1 [18] sound-editing program, at levels corresponding to the predicted sound-pressure levels at the listener position. The final auralization test materials were transferred to a compact disc for presentation to subjects using a CD player. The test material was replayed through Beyer DT 990 Pro headphones in a soundproof room. Each subject was tested individually. Subjects listened to a complete list of 50 words for each of the 12 different sound-field configurations. In order to avoid score inflation caused by the closed-set method used here, subjects were instructed not to guess the answer.

4.2.4 Data analysis

Descriptive statistics for the speech-intelligibility test scores for each of the twelve sound-field configurations were calculated both for the normal-hearing and the hearing-impaired listeners. Results are presented as the percentage of correct responses. Analysis of variance (ANOVA) was employed to compare the sensitivity of the scores to variations in the speech- and noise-source output-level difference (*SNS*) and reverberation time (*RT*). Tukey's multiple-comparison test was used for *post hoc* measures of individual mean differences [19]. A significance level of 0.05 was used to evaluate the statistical outcomes of the speech-intelligibility tests.

For each sound-field configuration, useful-to-detrimental ratios were calculated from the predicted impulse responses and the applicable speech and noise levels, as described in Ref. 9. Early-time limits of 10 ms intervals from 20 to 120 ms were used.

4.3 Results

4.3.1 Speech intelligibility

Figure 4.3 shows the variations of the mean speech-intelligibility score (with 95 % confidence interval) with reverberation time. For normal-hearing subjects, with *SNS* = 4 dB, the mean speech-intelligibility scores were between 77.1 % and 90.1 %. The mean scores for the six different *RT's* were statistically different ($p < 0.0005$). The highest mean speech-intelligibility score occurred at *RT* = 0.6 s, which was statistically significant for the six different *RT's* by the Tukey's multiple-comparison test ($\alpha = 0.05$). For hearing-impaired subjects, with *SNS* = 4 dB, the mean speech-intelligibility scores were between 72.9 % and 83.5 %, and were also statistically different ($p < 0.024$). The highest mean speech-intelligibility score also occurred at *RT* = 0.6 s; however it was not statistically significant by Tukey's multiple-comparison test ($\alpha = 0.05$).

For normal-hearing subjects with $SNS = 0$ dB, mean speech-intelligibility scores varied from 69.2 % to 78.5 %. There was a significant difference in the mean scores of the six RTs tested ($p < 0.0005$). With $RT = 0.8$ s and 1.2 s, the highest mean speech-intelligibility scores occurred with the same value of 78.5 %; the difference was not statistically significant according to their rank order by Tukey's multiple-comparison test ($\alpha = 0.05$). For the hearing-impaired subjects in the same conditions, mean speech-intelligibility scores varied from 64.0 to 71.1 %. Of the values tested, the highest score occurred at $RT = 0.8$ s.

The differences between the scores for the normal-hearing and hearing-impaired subjects were statistically significant, in general. When each configuration was examined individually, more detailed results were obtained. The mean speech-intelligibility scores for the normal-hearing and hearing-impaired subjects had statistically significant differences, except for the case of $RT = 1.5$ s, for both $SNS = 0$ dB and 4 dB.
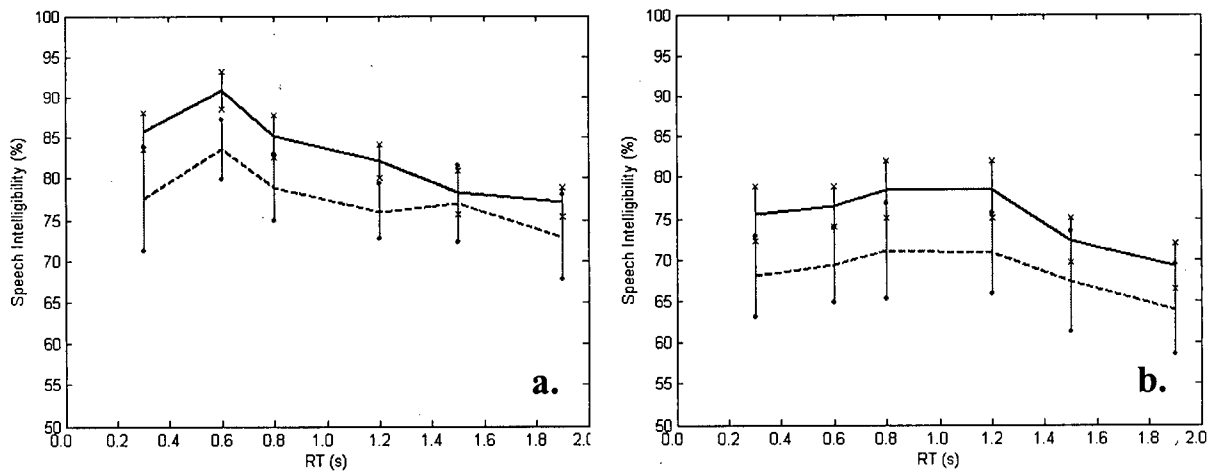
Figure 4. 3. Variation of mean speech-intelligibility score, and 95% confidence interval, with RT: a. $SNS = 4$ dB; b. $SNS = 0$ dB. (————) normal-hearing; (- - - - -) hearing-impaired.

4.3.2 Useful-to-detrimental ratio and Speech Transmission Index

Figure 4.4 shows U_t values with $SNS = 0$ dB and 4 dB using both methods. For $t = 30$ and 40 ms, the highest U_t values occurred at $RT = 0.6$ s with both SNS values. With $SNS = 4$ dB, the highest U_t values occurred at $RT = 0.8$ s for $t = 50$ ms and higher. When SNS decreased to 0 dB, the RT giving the highest U_t value increased to 1.2 s. Considering that the RT having the highest mean speech-intelligibility scores was 0.6 s for $SNS = 4$ dB and 0.8 s to 1.2 s for $SNS = 0$ dB, the variation of U_t prediction with RT showed fairly good agreement with the mean speech-intelligibility results.

In order to identify the early-time limits for which speech intelligibility was highest, regression analyses were performed on the mean speech-intelligibility scores for each sound-field configuration [20]. Table 4.2 shows the strengths of the relationships – quantified by the coefficient of determination, R^2 – between each measure and speech intelligibility, for both the normal- and hearing-impaired results. Since the form of the regression curves and the number of data points were the same in every case, the success of each measure can be compared by comparing the corresponding R^2 values. U_{30} was most accurate at predicting the speech-intelligibility results for the normal-hearing subjects. For the hearing-impaired subjects, U_{50} predicted the results best – that is, with early-time limits 20 ms higher than for the normal-hearing group. The limit using CATT-Acoustic was lower in this study than in the study of rooms with approximately diffuse sound fields [9]. This may be due to the fact that the sound fields in the virtual classrooms had relatively large numbers of specular reflections, so that the useful early sound time limits were decreased. The U_{30} (for normal hearing, 'NH') and U_{50} (for hearing-impaired, 'HoH') regression curves are shown in Figure 4.5a and b. The resulting best-fit regression equations are as follows:

Normal hearing:

$$SI_{NH} = 89.1 + 2.73685 \cdot U_{30} - 0.01553 \cdot U_{30,m}{}^2 \qquad R^2 = 91.3\ \%$$

Hearing-impaired:

$$SI_{HoH} = 79.5 + 1.63113 \cdot U_{50} - 0.26769 \cdot U_{50}{}^2 \qquad R^2 = 91.2\ \%$$

69

Table 4. 3. Coefficients of determination (R^2) associated with third-order-polynomial regression fits for various U_t values, for both normal-hearing (NH) and hearing-impaired (HoH) subjects.

R^2	U_{20}	U_{30}	U_{40}	U_{50}	U_{60}	U_{70}	U_{80}	U_{90}	U_{100}	U_{110}	U_{120}
NH	76.5	**91.3**	90.0	84.1	81.8	76.1	71.2	64.2	59.5	55.2	49.5
HoH	59.5	84.2	90.7	**91.2**	90.7	89.1	87.0	82.5	79.3	76.1	72.0
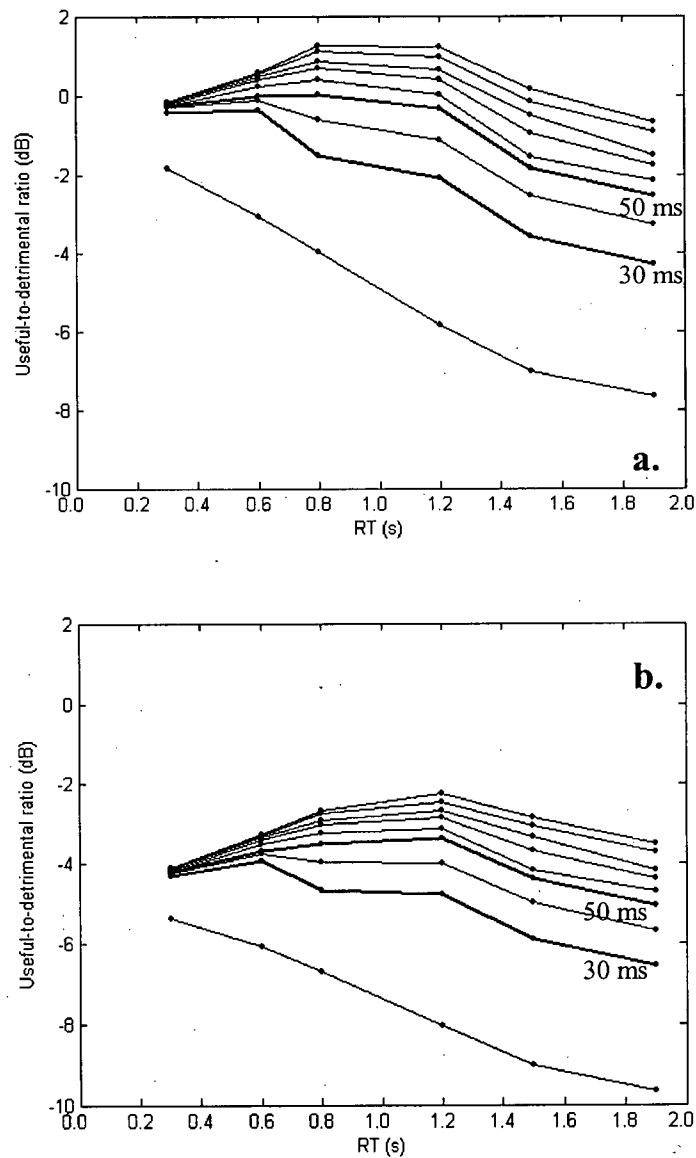
Figure 4. 4. Variation of useful-to-detrimental ratios with reverberation time for different early-time limits (from 20 ms to 100 ms): a. *SNS* = 4 dB; b. *SNS* = 0 dB.
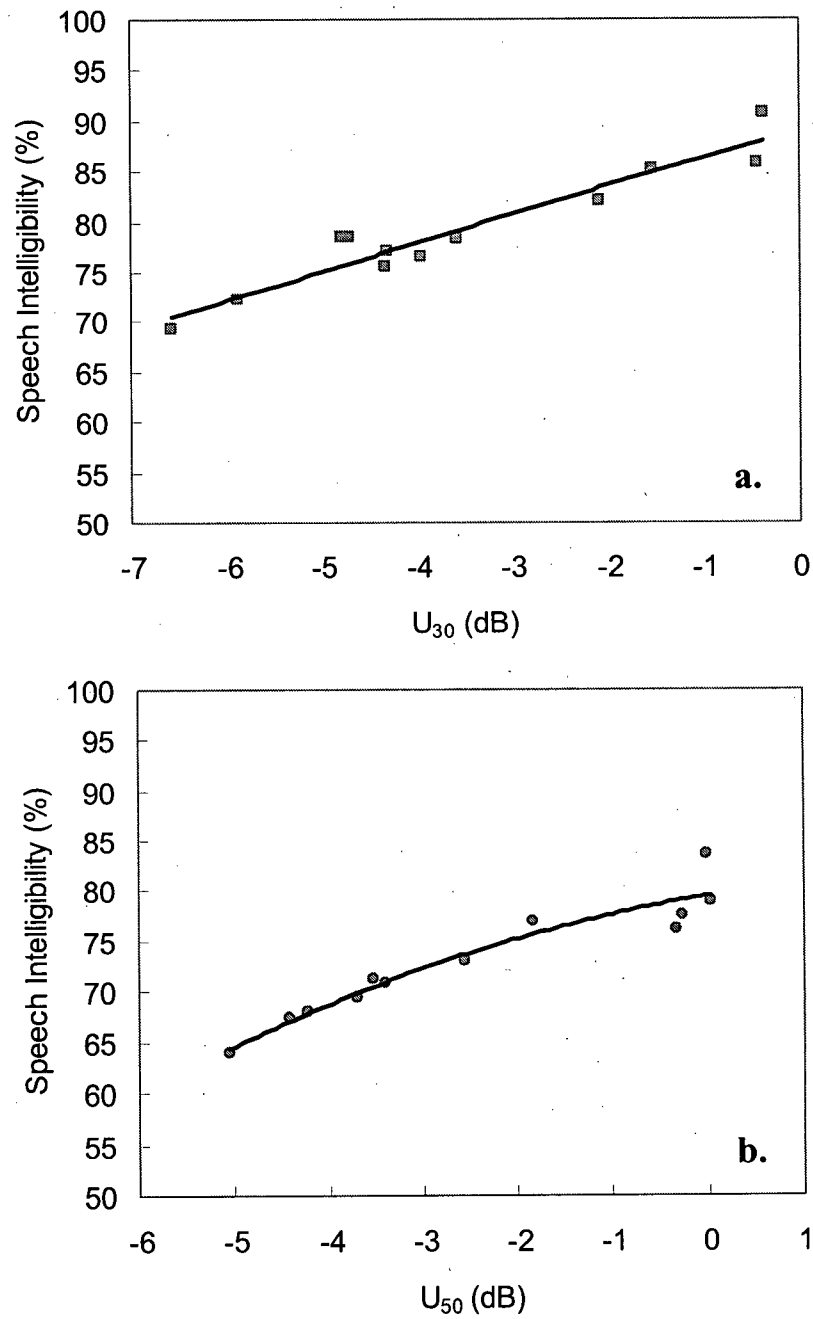
Figure 4. 5. Variation of mean speech-intelligibility score with useful-to-detrimental ratio using the best-fit early-time limit for all sound-field configurations and the best-fit quadratic regression curves: a. normal-hearing, U_{30}; b. hearing-impaired, U_{50}.

Speech Transmission Index was calculated based on the RT at 1k Hz and the A-weighted signal-to-noise ratios for each classroom configurations using a simplified version of the procedure developed by Steeneken and Houtgast based on the modulation transfer function, as described in Ref. 21. Figure 4.6 shows the mean speech-intelligibility scores versus STI. The resulting best-fit regression equations are:

Normal hearing:

$$SI_{NH} = 47.456 + 89.616 \cdot STI - 0.4005 \cdot STI^2 \qquad R^2 = 87.0\ \%$$

Hearing-impaired:

$$SI_{HoH} = 38.82 + 121.67 \cdot STI - 88.535 \cdot STI^2 \qquad R^2 = 54.5\ \%$$

STI predicted speech-intelligibility with $R^2 = 87.0\ \%$ for the normal-hearing subjects. The speech-intelligibility converged asymptotically to 100 % as STI was greater than 0.6. For the hearing-impaired subjects, STI could not describe the speech-intelligibility with high accuracy; that is, STI was unable to predict speech intelligibility for hearing-impaired listeners. This is consistent with previous results reported in the literature [22,23,24].
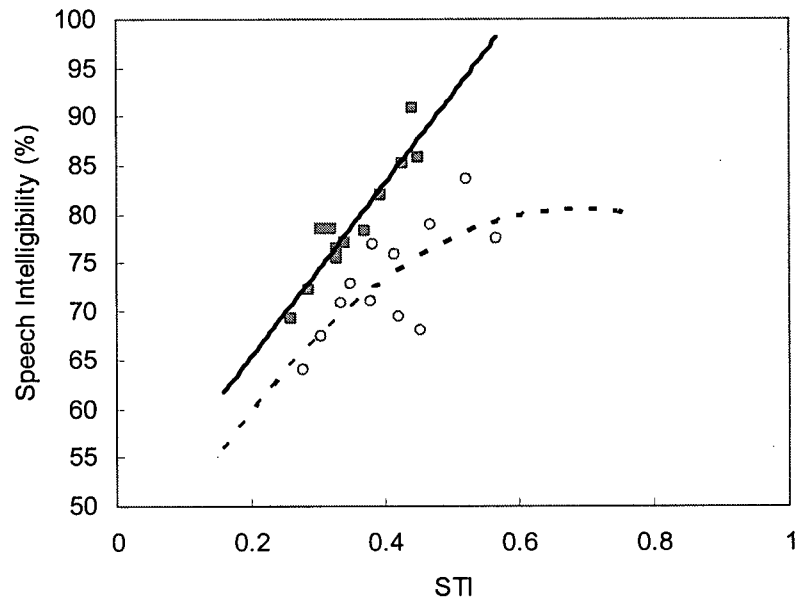
Figure 4. 6. Variation of mean speech-intelligibility score with Speech Transmission Index using the best-fit early-time limit, for all sound-field configurations, and the best-fit quadratic regression curves (eqns in text). (■,————) normal-hearing; (○,- - - - -) hearing-impaired.

4.4 Discussion

The results for the normal-hearing and hearing-impaired subjects, obtained with the MRT, revealed some basic differences in perception by the two groups, and also some similarities.

In this study, the range of RTs examined was broadened in order to investigate the speech intelligibility for the case of a noise source between the talker and the listener, which was expected to have non-zero optimal reverberation times [2], over a higher range of RTs than in the previous study [9]. When the noise source was between the speech source and the listener, of the RTs tested, the optimal reverberation time varied from 0.6 s to 0.8 s. The results are in good agreement with those of the previous study [9].
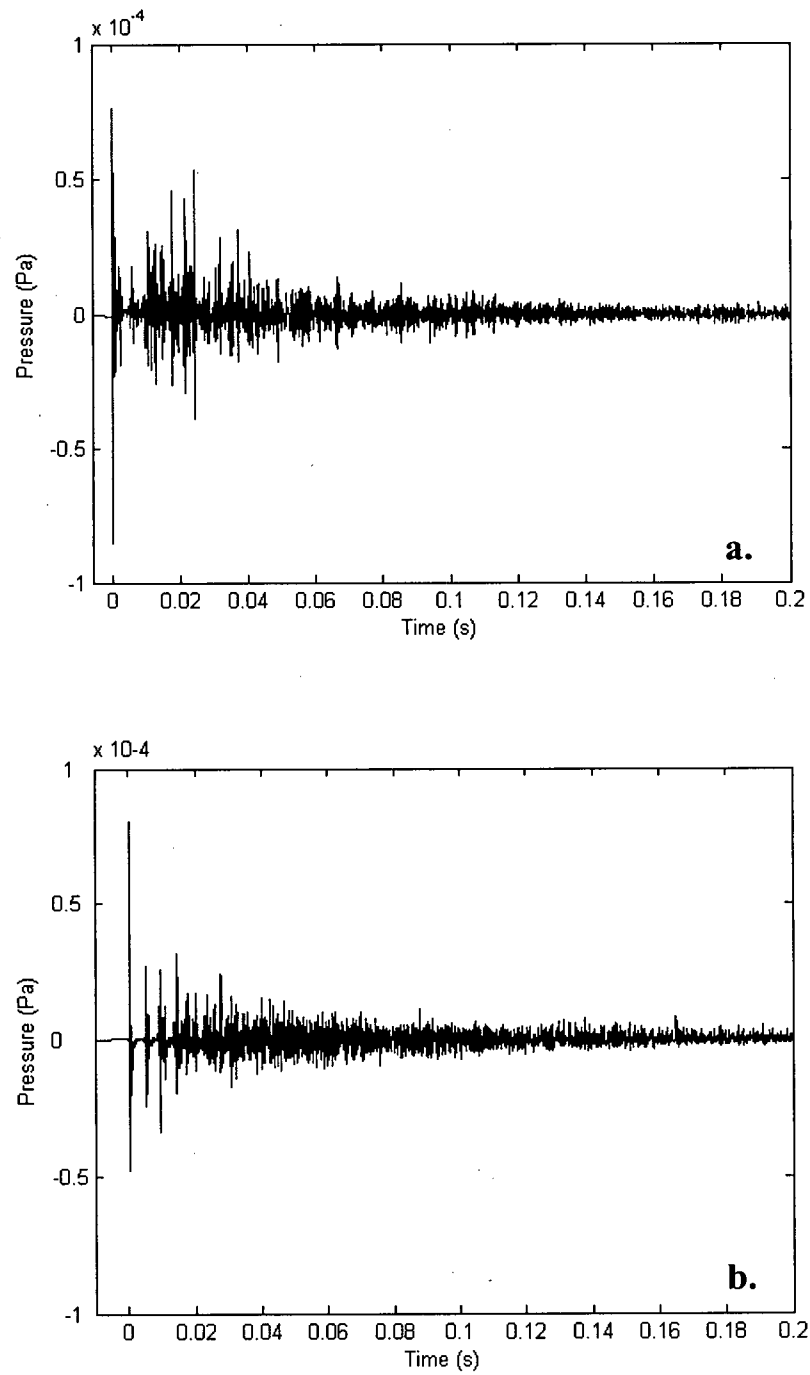
Figure 4. 7. The initial 0.2 s of the impulse response. a. RT = 0.8 s in the realistic classroom; b. RT = 0.8 s in the idealized classroom in Ref 9.

In general, hearing-impaired subjects in our study showed similar trends to the normal-hearing subjects in varied room-acoustical conditions. This is consistent with recent work by Bradley, Sato and Picard [8]. As expected, the hearing-impaired listeners were more adversely affected by reduced speech-to-noise level difference. Increased reverberation time generally resulted in increased speech intelligibility when the noise source was closer than the speech source to the listener; at times greater than the optimal reverberation time, increased reverberation time decreased speech intelligibility.

The useful early-time limits found here using CATT-Acoustic were shorter in the realistic virtual classrooms than in the approximately diffuse sound fields of the previous study [9]. This can be explained by examining the impulse responses. Figure 4.7 shows the initial 0.2 s of the impulse responses for the realistic classroom and for the idealized classroom with an approximately diffuse sound field, both with $RT = 0.8$ s. Figure 4.7a clearly shows strong specular reflections in the realistic virtual classroom. This shows that optimum reverberation times may vary from room to room according to the details of the reflection arrival in the impulse response.

4.5 Conclusions

Extending previous work in classrooms with approximately diffuse sound fields, this study used auralization and speech-intelligibility tests to find the optimum reverberation time for speech intelligibility in more realistic rooms with non-diffuse sound fields. For both normal and hearing-impaired subjects, the results of this study agreed with previous results [9]. The optimal reverberation time was not zero when the noise source was between the listener and the speaker, with the received speech-to-noise level difference varying from -4 to +3 dB. The optimal reverberation time increased with decreasing the speech-to-noise level difference. If the speech-to-noise level difference is adverse for a subject group, some reverberation is required to increase the speech signal. Hearing-impaired subjects required more early energy than normal-hearing subjects for this range of speech-to-noise level differences.

In this study, a typical medium-sized classroom was the model for the virtual classroom. It would be interesting to repeat the study using various types of classroom to improve the current work. While the hearing-impaired listeners in this study benefited from early reflections more than the normal listeners, it would be also interesting to verify the result for elderly listeners or more severely impaired listeners who use hearing aids in their everyday lives to investigate optimal reverberation for hearing aids.

References

[1]   J. S. Bradley, "Speech intelligibility studies in classrooms," J. Acoust. Soc. Am. **80** (3), 846-854 (1986).

[2]   M. Hodgson and E.-M. Nosal, "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms," J. Acoust. Soc. Am. **111** (2), 931-939 (2002).

[3]   J. P. A. Lochner and J. F. Burger, "The influence of reflections on auditorium acoustics," J. Sound Vib. **1** (4), 426-448 (1964).

[4]   H. G. Latham, "The signal-to-noise ratio for speech intelligibility--An auditorium acoustics design index," Appl. Acoust. **12** (4), 253-320 (1979).

[5]   J. S. Bradley, "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am. **80** (3), 837-845 (1986).

[6]   J. S. Bradley, "Relationships among measures of speech intelligibility in rooms," J. Audio Eng. Soc. **46** (5), 396-405 (1998).

[7]   J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. **106** (4), 1820-1828 (1999).

[8]   J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113** (6), 3233-3244 (2003).

[9]   W. Yang and M. Hodgson, "Auralization study of optimum reverberation times for speech intelligibility for normal and hearing-impaired listeners in classrooms with diffuse sound fields," J. Acoust. Soc. Am. **120**(2), 801-807 (2006).

[10]   J. G. Clark, "Uses and abuses of hearing loss classification," ASHA **23** (7), 493-500 (1981).

[11]   D. Henderson, R. J. Salvi, F. A. Boettcher et al.,"Neurophysiologic correlates of sensory-neural hearing loss," in *Handbook of cliniclal audiology*, edited by J. Katz (Williams & Wilkins, Baltimore, 1994), pp. 37-55.

[12]   W. Yang and M. Hodgson, "Speech intelligibility tests in virtual and real classrooms," Proc. Forum Acusticum 2005, Budapest (2005).

[13]   M. Hodgson, "Evidence of diffuse surface reflections in rooms," **89** (2), 765-771 (1991).

[14]   B. -I. Dalenbäck, CATT-Acoustic v8.0 Users' Manual (Gothenburg, 2004).

[15]   American National Standards Institute, "ANSI S3.5-1997 Methods for caculation of the speech intelligibility index," 1997.

[16]   E. J. Kreul, J. C. Nixon, K. D. Kryter et al., "A proposed clinical test of speech discrimination," J. Speech Hear. Res. **11** (3), 536-552 (1968).

[17]   Auditec of St. Louis (St. Louis, MO, 2003).

[18]   GoldWave Inc., GoldWave v5.1 (St. John's, NL, 2005).

[19]   F. J. Gravetter and L. B. Wallnau, *Essentials of statistics for the behavioral sciences*, 4th ed. (Wadsworth, Pacific Grove, CA, 2002), pp.327-330.

[20]   J. S. Bradley and S. R. Bistafa, "Relating speech intelligibility to useful-to-detrimental sound ratios (L)," J. Acoust. Soc. Am. **112** (1), 27-29 (2002).

[21]   M. Hodgson, "Rating, ranking, and understanding acoustical quality in university classrooms," J. Acoust. Soc. Am. **112** (2), 568-575 (2002).

[22]   K. L. Payton, R. M. Uchanski, and L. D. Braida, "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95** (3), 1581-1592 (1994).

[23]   L. E. Humes, S. Boney, and F. Loven, "Further validation of the Speech Transmission Index (STI)," J Speech Hear Res **30** (3), 403-410 (1987).

[24]   L. E. Humes, D. S. D. Dirks, T. Bell et al., "Application of the articulation index and the speech transmission index to the recognition of speech by normal-hearing and hearing-impaired listeners.," J. Speech Hear. Res. **29** (4), 447-462 (1986).

5 CEILING BARRIERS AND REFLECTORS TO OPTIMIZE LECTURE-ROOM SOUND FOR SPEECH INTELLIGIBILITY *

5.1 Introduction

The importance of classroom acoustics and of speech intelligibility is well recognized. The room-acoustical parameters affecting speech intelligibility are known; generally, speech intelligibility tends to increase with increased speech-to-noise level difference and to decrease with increased reverberation. Reverberation for speech intelligibility is best quantified by the clarity factor C_{50}, based on the early-to-late energy fraction. C_{50} is usually highly correlated with early-decay time and reverberation time in classrooms [1]. The effect of speech-to-noise level difference on speech intelligibility dominates that of reverberation [2]. Other studies [1,3,4] have confirmed that the values of these parameters are often non-optimal. Thus, the question remains as to how to achieve the optimal values of the parameters in real classrooms in a practical, cost-effective way. Many newly designed or renovated classrooms use absorptive materials to reduce reverberation and late-arriving energy. However, these absorptive materials also decrease speech levels and may reduce beneficial early reflections.

The purpose of the study reported here was to find an effective way to design lecture rooms – *i.e.* larger classrooms with an instructor at the front of the room, speaking to a group of students in front of him/her – and control sound to achieve optimum reverberation and adequate speech levels, especially at the back of the room, for speech intelligibility. The effectiveness of a novel system of ceiling barriers and reflectors for optimizing speech intelligibility is investigated, using a room-prediction model and physical scale-modeling.

A room consists of a floor, walls and a ceiling. Among these three room components, the ceiling is chosen to be modified because it has a large flexibility compared to the walls and floor, and it can help reflect a teacher's voice toward the back of the room. Various ceiling-barrier and ceiling-reflector configurations were designed. Each design was incorporated into computer and scale-model lecture-room models, and the effects of the barriers and reflectors on the sound field were predicted, to optimize the designs.
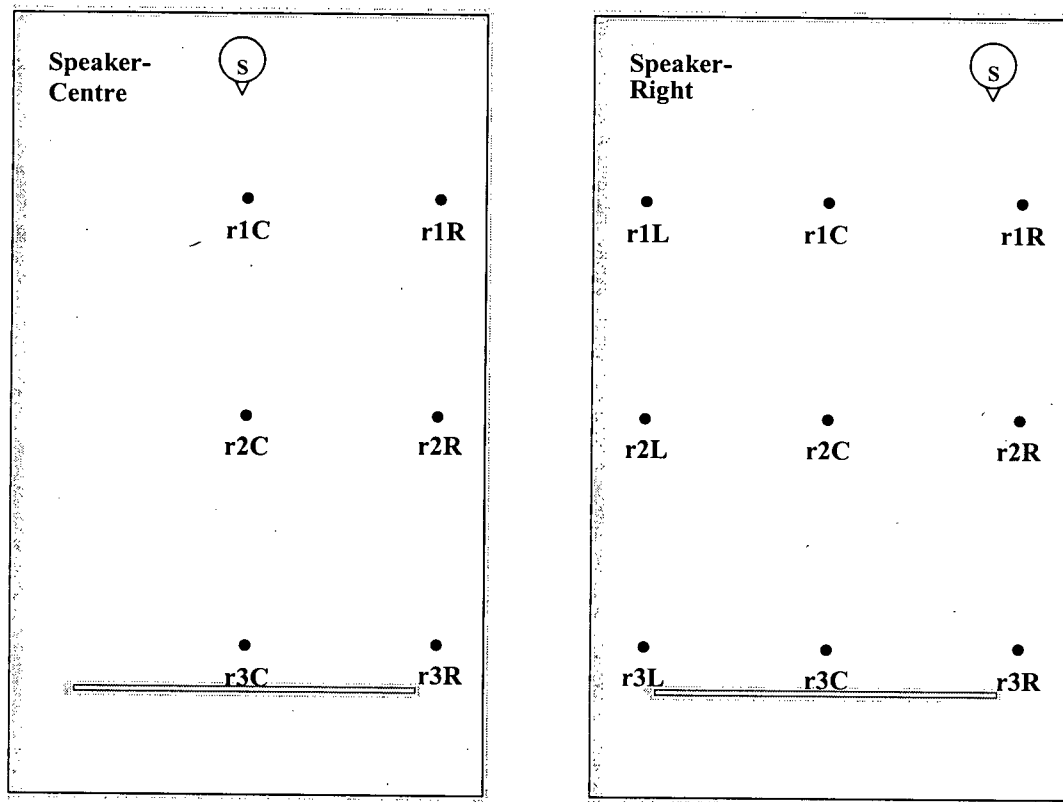
Figure 5. 1. Lecture-room floor-plan showing the speech-source (S) and receiver (r1L, r2L, r3L, r1C, r2C, r3C, r1R, r2R, r3R) positions.

5.2 Methods

5.2.1 Lecture-room configurations

A typical medium-sized university lecture room was selected as the model for tests of the effects of the ceiling barriers and reflectors. Figure 1 shows the floor plan. The lecture-room has 96 seats, length = 15 m, width = 8.5 m, and height = 4 m (volume = 510 m^3, total surface area = 443 m^2, volume to surface-area ratio = 1.15 m). The room surfaces and the seats with writing tablets, are sound-reflective. The mid-frequency reverberation time measured in the unoccupied classroom was about 2 s. Average octave-band surface-absorption coefficients calculated using diffuse-field theory varied from 125 to 2000 Hz as follows: 0.22, 0.12, 0.08, 0.08, 0.08. A speech source was positioned either at the front-center or at the front-right of the lecture room, and three listening positions were positioned in front (r1C – source/receiver distance = 2.0 m), middle (r2C = 5.5 m), and back (r3C = 9.0 m) seats on the center line of the lecture room. Six additional listening positions were used when predicting the acoustical conditions in side areas of the lecture room (see Figure 5.1).

5.2.2 Ceiling barrier and reflector configurations

Two basic types of ceiling barriers and reflectors were studied. The first involved parallel ceiling barriers, projecting down from the ceiling and running front-to-back in the classroom. They were expected to absorb reverberant sound incident on the ceiling from many angles, reducing late-arriving energy, while leaving speech signals, reflecting from the (reflective) ceiling between the barriers to the back of the room, unaffected. Both sound-absorptive and sound-reflective barriers were considered. Different shapes, materials, spacings and depths of the ceiling-barrier configurations were considered in pilot studies. Six configurations were selected for detailed study using computer prediction and scale-model measurement. These were reflective (configuration R) and absorptive (A) barriers, 0.6-m deep, separated by either 0.3 m (R1/A1), 0.6 m (R2/A2) or 1.2 m (R4/A4).

Based on preliminary barrier results, ceiling reflectors involving lengths of obstacles of semicircular cross-sectional shape (configuration C), suspended from the ceiling with either the flat (F) or curved (C) side down, were evaluated in the scale model. These were expected to reflect and scatter speech sounds like 'fittings' in an industrial workshop (see below). The shapes were based on common suspended light fixtures. The diameter of the semicircular reflectors was 0.3 m, the distance from the ceiling to the bottom of the reflectors was 0.6 m. They were 7-m long, and ran front-back in the lecture room, separated by 1.2 m, and with either the flat side down without absorption on the curved side (CF), the flat side down with absorption on the upper curved side (CFA), or the curved side down (CC) without absorbtion. Alternatively, either thirty (configuration CF30) or sixty (CF60) reflectors, 0.8-m long, were hung randomly from the ceiling with flat side down without absorption.

5.2.3 Physical scale-modeling

The lecture room was studied without and with ceiling barriers and reflectors using a 1:8-scale model. According to the fundamental principle of the scale-modeling technique [5], according to which all dimensions are scaled down by the scaling factor, the lecture room with length = 15 mFS (FS=full-scale value), width = 8.5 mFS, and height = 4 mFS, was a 1.88 m x 1.06 m x 0.50 m 1/8-scale model. The floor of the model was of polished concrete. The walls, ceiling, the partition at the back of the room, and the rows of seats, were of varnished plywood. Figure 5.2a is a photograph of the model showing the rows of seats and the rear partition. In order to investigate how the effects of ceiling barriers and reflectors vary with room occupancy in the lecture-room, two different occupancies (unoccupied and 37 % occupied/26 students) were used in the scale model.

Air absorption is a major problem in scale-model measurement. Air absorption increases approximately with the square of the frequency [6]. In a scale model, wavelength-to-dimension ratios are maintained, so wavelengths are scaled down by the scale factor, resulting in scaled-up model test frequencies. Since the test frequencies are high, air absorption is excessive in a scale model and cannot be neglected. For prediction, air-absorption exponents were calculated at the model test frequencies for the temperature and relative humidity measured in the scale model, as described in Ref. 6.
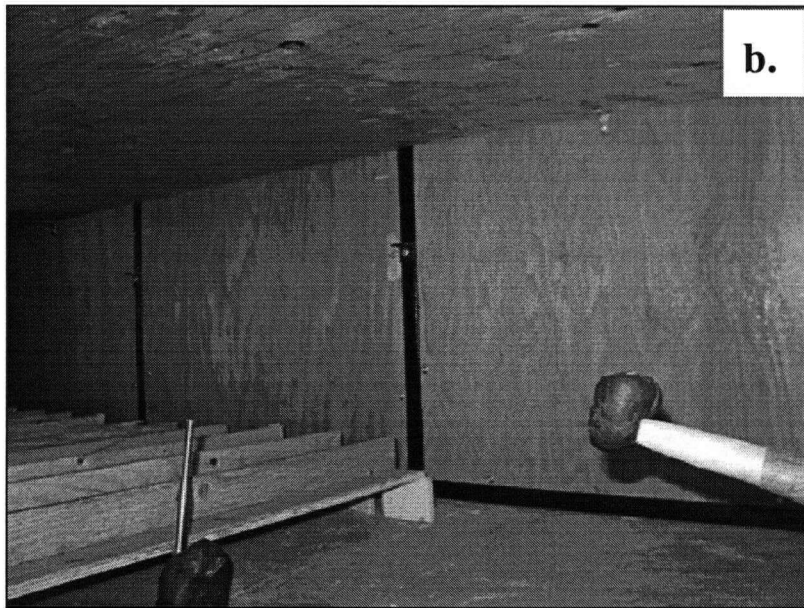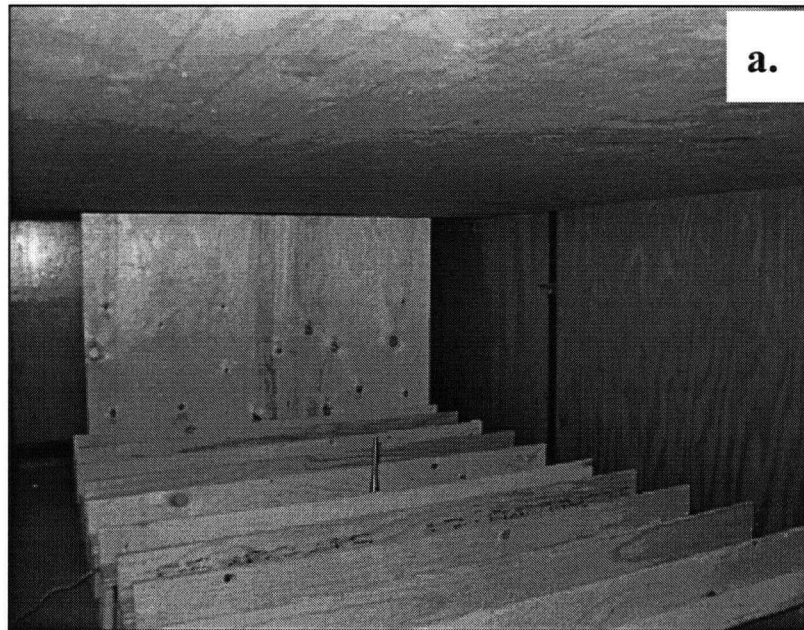
Figure 5. 2. Photographs of the 1/8-scale-model without ceiling barriers or reflectors: a. showing the seats and rear partition; b. showing the model speech source and microphone.

Speech sources in lecture rooms are mainly human talkers. Source directivity can strongly influence speech levels in lecture rooms. For accurate scale modeling of the lecture room, a model speech source is required which radiates with the directional characteristics of human speech. Such a source was created using a 1:8-scale head made of modeling clay, and formed around the end of a hollow cone driven by a 'tweeter' loudspeaker, which narrowed down to a 3-mm-diameter opening as the mouth, to represent human speech directivity in the scale model (see Figure 5.2b). The power levels and directivities of the model speech source were measured in an anechoic chamber. Figure 5.3 shows the measured horizontal-plane directivity.
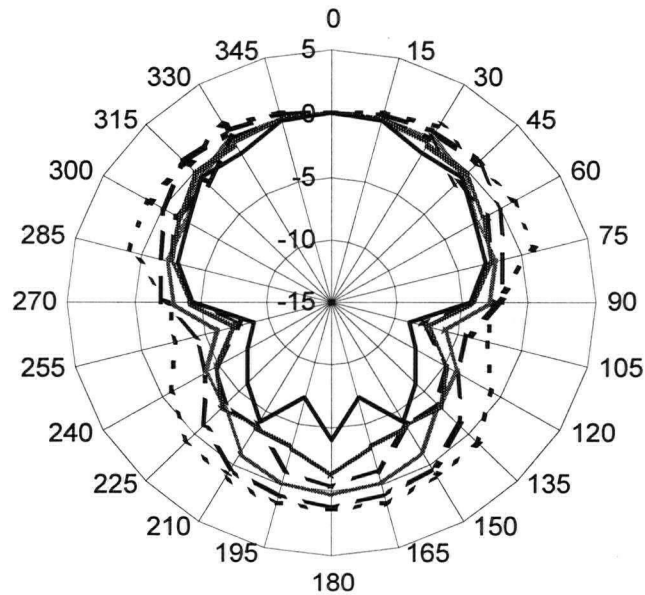
Figure 5. 3. Measured octave-band, horizontal-plane directivity factors of the 1:8-scale-model speech source; levels are normalized to 0 dB at 0°: (.........) 125 HzFS; (— · —) 250 HzFS; (⌇⌇⌇⌇⌇⌇) 500 HzFS; (— — —) 1000 HzFS; (⎯⎯⎯) 2000 HzFS; (⎯⎯⎯) 4000 HzFS.

The ceiling barriers were made of varnished plywood, and could be covered with thin industrial carpet to make them absorbent. The absorption of the carpet was estimated from the change in reverberation time that occurred when a sample of it was introduced into the empty model. The absorption was similar to that of 50-mm-thick glass fibre at full-scale frequencies. The semicircular ceiling reflectors were made of painted wood; the same carpet was used to make them absorbent. The carpet was also used to cover the seats to simulate occupied absorption. The corresponding occupancy was estimated from the reduction in *EDT* at 1kHz [7]. Figure 5.4 shows photographs of some of the scale-model ceiling barriers and reflectors.

Acoustical measurements were made using the Maximum Length Sequence System Analyzer (MLSSA), which measured the impulse response between the model speech source and a Bruel & Kjaer 4135, 1/4"-diameter microphone used to receive the sound signals. All measurements were made after the pre-calibration of the equipment. Early-decay times (*EDT* in s) and steady-state levels (with the speech-source output levels always kept constant) were measured. Measurements were made in octave bands from 1 to 16 kHz (125 to 2000 HzFS) at all three receiver positions. Average mid-frequency·EDT_{mid} values most relevant to speech intelligibility were calculated by averaging the octave-band *EDT*'s at 500, 1000 and 2000 HzFS. In the case of the model without ceiling barriers or reflectors, average surface-absorption coefficients were calculated from the measured octave-band *EDT*'s using diffuse-field theory; values increased with frequency from 0.06 to 0.1, close to those in the full-scale room. The octave-band steady-state levels were converted to total A-weighted 'speech' levels SLA_N corresponding to a typical adult talking in a normal voice level, using the relative output power levels of such a talker [8] and of the model speech source.

5.2.4 Computer simulation

The ceiling barriers were also studied using CATT-Acoustic v8.0 [9] computer simulation. The lecture room and ceiling-barrier configurations were modeled, octave-band *EDT*'s and speech levels predicted, and corresponding values of EDT_{mid} and SLA_N calculated. The lecture-room configuration was exactly the same as in the scale-model measurements. A sound source and nine receivers were positioned as shown in Figure 5.1. The output level
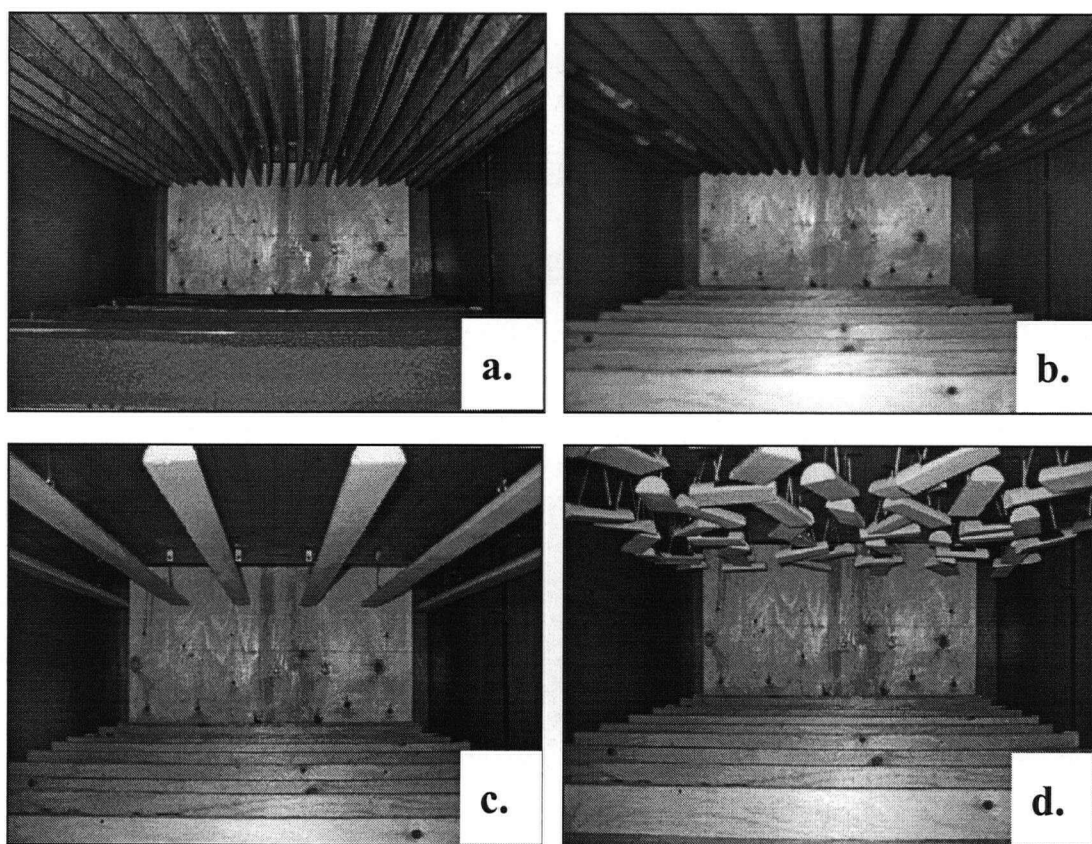
Figure 5. 4. Scale-model ceiling barrier and reflector configurations in the unoccupied lecture-room: a. R1o; b. A1u; c. CFu; d. CF60u.

and the directivity of the sound source were identical to the values used in the scale-model measurements. Unoccupied seats were modeled as one large 1-m-deep seat block (see Figure 5.5). Figure 5.5 shows computer models of the virtual lecture room without and with ceiling barriers. Ceiling reflectors were not studied by computer prediction as it was not clear exactly how to model them. The absorption coefficients of the room surfaces used in the simulation were the average values measured in the scale model without ceiling barriers or reflectors. Diffuse-reflection coefficients of the surfaces were set to increase with frequency from 0.1 to 0.3, based on previous research [10].
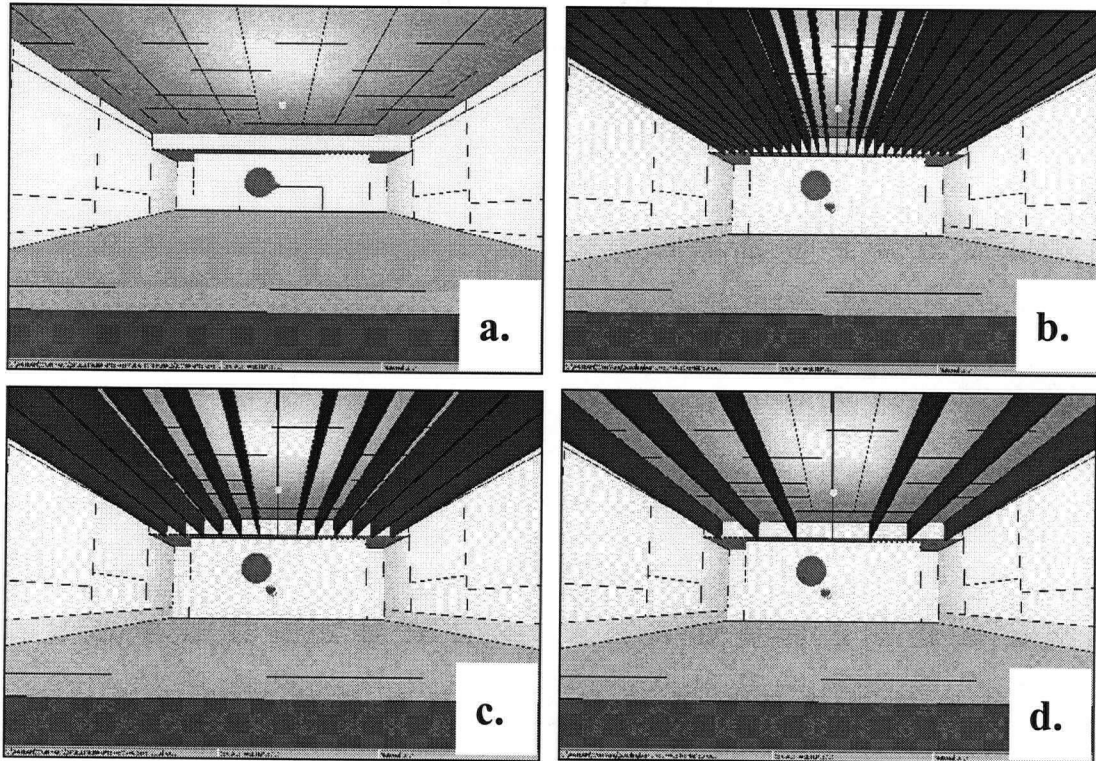
Figure 5. 5. Computer models of the lecture room with: a. no barriers; b. R1/A1 barriers; c. R2/A2 barriers; d. R4/A4 barriers.

5.3   Results

5.3.1   Comparison of measurement and prediction

In order to confirm that the scale-model and virtual lecture rooms were reasonably similar, comparisons were made between measured and predicted speech levels and early-decay times for the three central receiver positions. Figure 5.6 shows the octave-band results. Predicted speech levels were somewhat lower than those measured – by about 4 dB at low frequency, decreasing with frequency to about 1.5 dB at high frequency. Predicted early-decay times varied negligibly with position; measured times showed much more variation. Predicted *EDT*'s tended to be lower than those measured, by up to about 0.7 s (25 %). The imperfect agreement between measurement and prediction is interesting, given that the average absorption coefficients involved in the virtual and scale models were very similar to
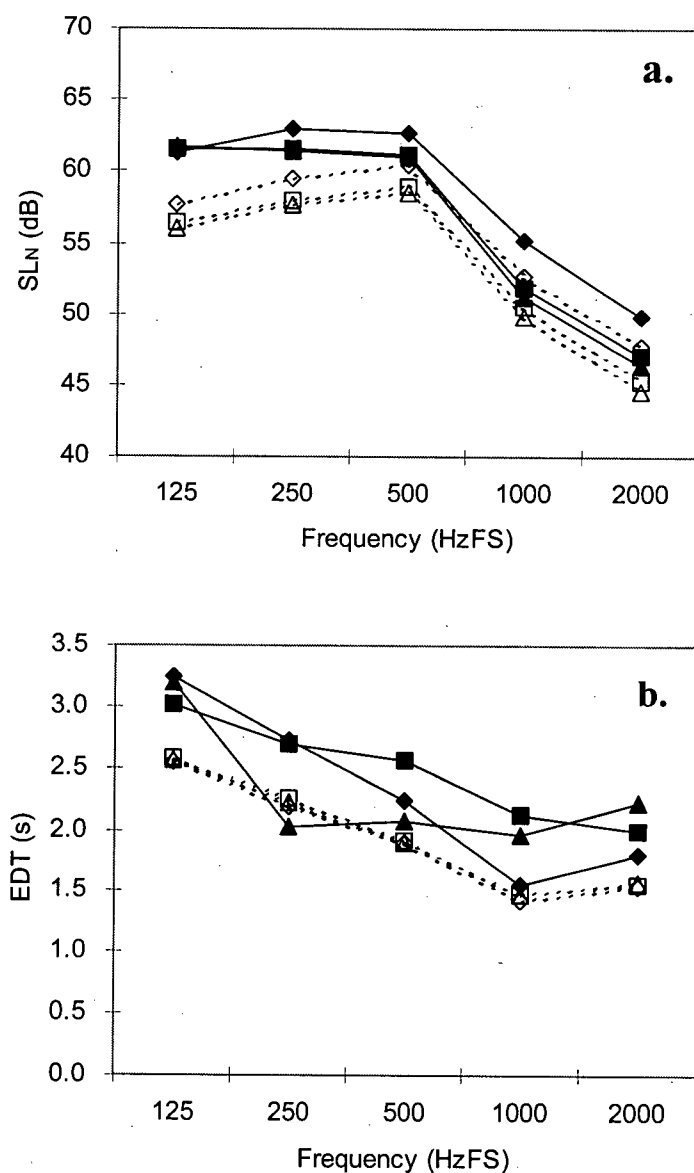
Figure 5. 6. Variation with frequency of (a.) speech levels SL_N and (b.) early-decay times *EDT* at three central positions with the centre speaker in an unoccupied lecture room without ceiling barriers or reflectors, as measured in a scale-model and as predicted by CATT-Acoustic: (—♦—) r1C-measured, (—■—) r2C-measured, (—▲—) r3C-measured, (····◊····) r1C-predicted, (····□····) r2C-predicted, (····△····) r3C-predicted.

one another. It can partly be explained by uncertainties in the scale-model measurements, differences in the values of important room parameters (e.g. the diffuse-reflection coefficients), and possible limitations of the computer simulation (for example, the seat block). In any case, it was concluded that the scale-model and virtual lecture-room models, while not identical, are sufficiently similar that both can be used to study the effects of ceiling barriers and reflectors. The two techniques have their advantages and disadvantages. The scale model has the advantage of physical realism (for example, including modal effects), while prediction has the advantage that the input data defining the virtual room are precisely known.

5.3.2   Ceiling barriers

Figure 5.7 compares the speech-level and early-decay-time results for the six ceiling-barrier configurations with the results for no barriers for the three positions along the center line in the unoccupied room. The measured and predicted levels are highest with no barriers, for which speech levels decrease by about 2 dBA from the front to the back of the lecture room. With the reflective ceiling barriers, as the number of barriers increased, predicted speech levels remained virtually unchanged at the front of the room, but decreased by up to about 1.5 dBA at the back. Measured levels showed more variability, but similar trends, especially at r3. With the absorptive ceiling barriers, and increasingly with the number of barriers, levels decreased more rapidly with distance relative to levels with the reflective barriers; speech levels decreased by up to between 4 and 6 dBA at the front and back of the room, respectively.

Ceiling barriers decreased the early-decay times in all cases. With the reflective ceiling barriers, the predicted EDT_{mid} progressively decreased with increased number of barriers, from about 1.6 s to about 0.8 s; the decreases were similar at the three receiver positions. With absorptive barriers, EDT_{mid} varied little with barrier spacing and was very low (around 0.5 s). Again, measured results were similar, but showed less clear trends, and barriers resulted in smaller decreases in EDT_{mid}.

The absorptive ceiling barriers do not achieve the objective, since they decrease speech levels significantly, as well as reducing early-decay times. Reflective barriers, on the other hand, do achieve the objective, reducing early-decay times significantly, while reducing speech levels little. Therefore, the study was focused on the reflective ceiling barriers. Effects of the occupancy in the lecture-room, and effects of the speaker-position, on the performance of the reflective ceiling barriers were investigated.
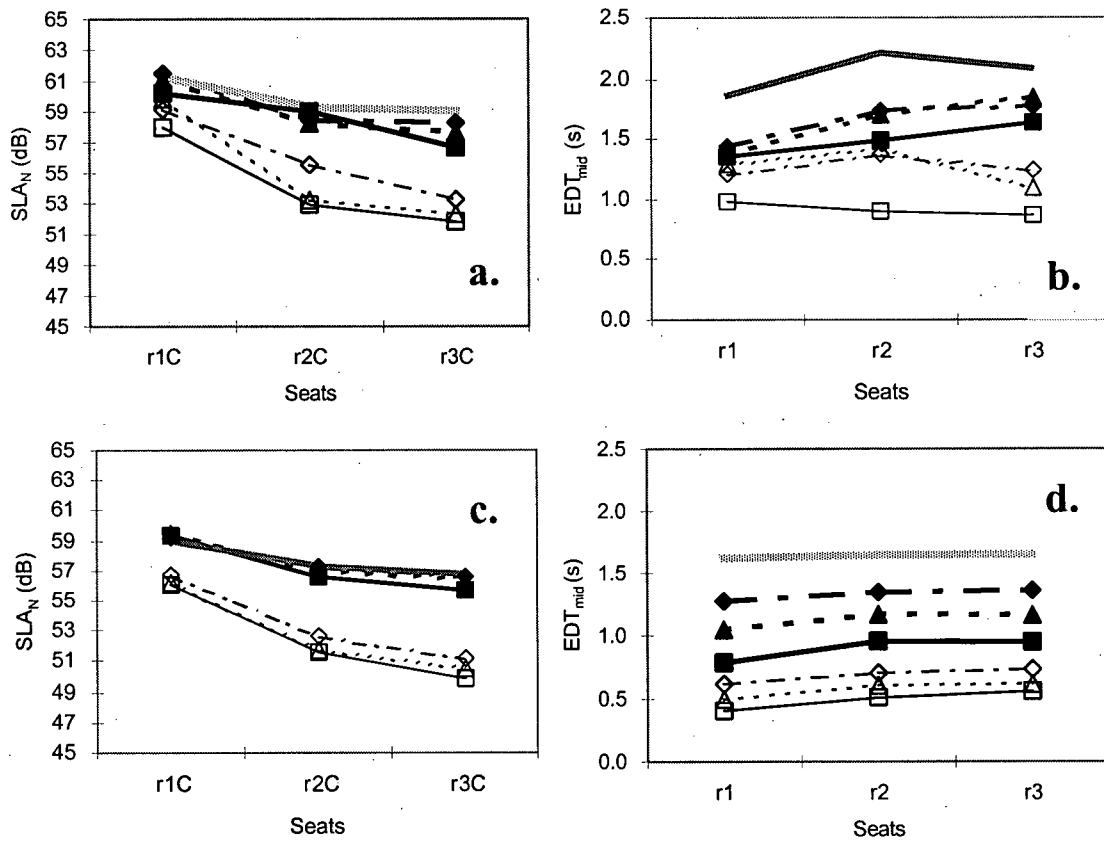
Figure 5. 7. Variation with position of speech levels SLA_N and early-decay times EDT_{mid} in an unoccupied lecture room without and with reflective and absorptive ceiling barriers, as measured in a scale model and as predicted by CATT-Acoustic along the center line with the centre speaker: a. SLA_N, measured; b. EDT_{mid}, measured; c. SLA_N, predicted; d. EDT_{mid}, predicted. (⎯⎯⎯) no barriers, (—■—) R1, (- -▲- -) R2, (—·—♦—·—) R4, (—□—) A1, (- - △ - -) A2, (—·—◇—·—) A4.

SLA$_N$ (dB)	(S)			EDT$_{mid}$ (s)	(S)	
	a.				**b.**	
NOo 60.4	NOo 58.9			NOo 1.08	NOo 1.33	
NOu 63.4	NOu 63.8			NOu 2.21	NOu 2.00	
R1o 60.3	R1o 57.8			R1o 0.95	R1o 1.05	
R2o 60.2	R2o 59.1			R2o 0.86	R2o 0.97	
R4o 60.8	R4o 59.1			R4o 1.07	R4o 0.97	
R1u 61.2	R1u 60.7			R1u 1.37	R1u 1.34	
R2u 62.1	R2u 60.2			R2u 1.53	R2u 1.73	
R4u 62.7	R4u 60.8			R4u 1.56	R4u 1.75	
NOo 56.0	NOo 57.1			NOo 1.76	NOo 1.56	
NOu 60.5	NOu 61.5			NOu 2.74	NOu 2.24	
R1o 56.5	R1o 55.7			R1o 1.05	R1o 1.10	
R2o 56.9	R2o 56.5			R2o 1.22	R2o 1.05	
R4o 57.1	R4o 56.8			R4o 1.06	R4o 1.17	
R1u 58.9	R1u 59.6			R1u 1.56	R1u 1.48	
R2u 58.8	R2u 58.9			R2u 1.94	R2u 1.89	
R4u 60.3	R4u 59.7			R4u 1.65	R4u 1.79	
NOo 57.7	NOo 55.5			NOo 1.55	NOo 1.55	
NOu 59.3	NOu 60.0			NOu 3.06	NOu 2.85	
R1o 57.3	R1o 54.9			R1o 1.07	R1o 1.15	
R2o 56.9	R2o 56.5			R2o 1.15	R2o 1.28	
R4o 56.7	R4o 55.9			R4o 1.23	R4o 1.15	
R1u 56.2	R1u 57.9			R1u 1.59	R1u 1.63	
R2u 59.1	R2u 58.3			R2u 1.66	R2u 1.83	
R4u 59.1	R4u 60.0			R4u 1.78	R4u 1.55	

Figure 5. 8. Variation with position of speech levels SLA_N **and early-decay times** EDT_{mid} **with the centre speaker without and with reflective ceiling barriers, as measured in a scale model: a.** SLA_N**; b.** EDT_{mid}**.**

Figure 5.8 shows the measured A-weighted speech levels for a 'Normal' voice (SLA_N) and EDT_{mid} values with reflective ceiling barriers for the six receiver positions with the centre speaker in the scale model. In both the occupied and unoccupied rooms, speech levels did not decrease significantly and speech levels in the side seats were slightly lower than in the center-line seats. In the occupied room, speech levels remained constant or increased slightly with barriers. Early decay times decreased with barriers in both the occupied and unoccupied rooms.

SNA$_N$ (dB) a. (s)

NOo 56.5	NOo 58.3	NOo 61.3
NOu 62.9	NOu 62.9	NOu 63.0
R1o 57.5	R1o 56.8	R1o 59.8
R2o 58.7	R2o 58.6	R2o 61.0
R4o 58.1	R4o 59.6	R4o 60.9
R1u 60.3	R1u 59.1	R1u 59.7
R2u 60.3	R2u 60.5	R2u 62.2
R4u 61.0	R4u 61.3	R4u 62.5
NOo 54.7	NOo 54.9	NOo 57.6
NOu 61.2	NOu 60.5	NOu 61.0
R1o 55.0	R1o 52.8	R1o 55.0
R2o 56.9	R2o 56.6	R2o 56.7
R4o 56.3	R4o 56.6	R4o 57.6
R1u 57.5	R1u 56.5	R1u 56.5
R2u 58.7	R2u 58.8	R2u 58.8
R4u 59.0	R4u 59.7	R4u 59.7
NOo 54.7	NOo 56.1	NOo 57.1
NOu 59.2	NOu 60.8	NOu 60.7
R1o 54.7	R1o 53.7	R1o 55.8
R2o 55.4	R2o 55.6	R2o 55.3
R4o 55.5	R4o 56.8	R4o 56.6
R1u 57.2	R1u 55.6	R1u 57.6
R2u 58.1	R2u 58.4	R2u 58.0
R4u 58.6	R4u 59.7	R4u 57.9

EDT$_{mid}$ (s) b. (s)

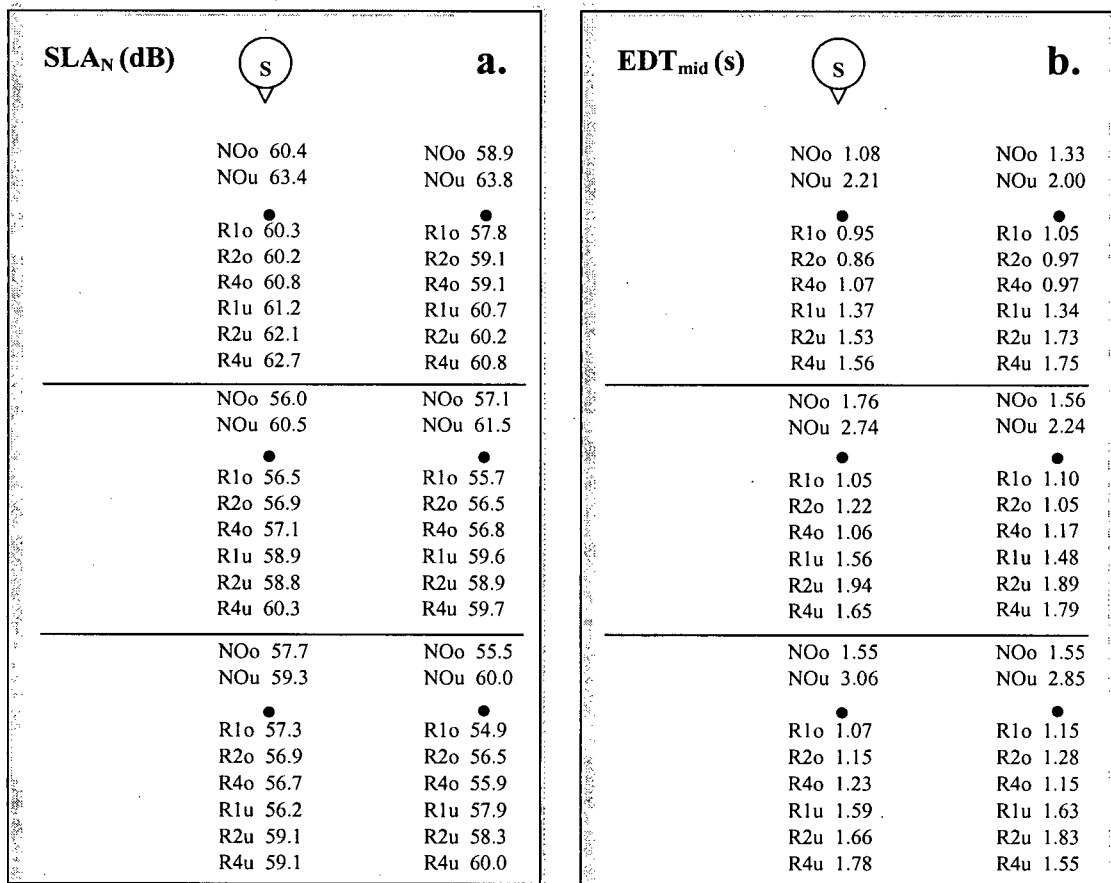|		
NOo 1.67	NOo 1.36	NOo 1.23
NOu 2.08	NOu 2.17	NOu 2.23
R1o 1.24	R1o 1.29	R1o 0.95
R2o 1.15	R2o 1.06	R2o 0.93
R4o 1.05	R4o 0.95	R4o 0.90
R1u 1.37	R1u 1.35	R1u 1.35
R2u 1.51	R2u 1.38	R2u 1.58
R4u 1.53	R4u 1.51	R4u 1.64
NOo 1.76	NOo 2.04	NOo 1.64
NOu 2.49	NOu 2.94	NOu 1.64
R1o 1.46	R1o 1.27	R1o 1.40
R2o 1.38	R2o 1.36	R2o 1.12
R4o 1.34	R4o 1.23	R4o 1.20
R1u 1.61	R1u 1.60	R1u 1.51
R2u 175	R2u 1.96	R2u 1.76
R4u 2.01	R4u 1.96	R4u 1.96
NOo 1.72	NOo 1.42	NOo 1.42
NOu 2.34	NOu 3.08	NOu 2.57
R1o 1.55	R1o 1.03	R1o 1.30
R2o 1.41	R2o 1.30	R2o 1.30
R4o 1.32	R4o 1.06	R4o 1.18
R1u 1.59	R1u 1.42	R1u 1.73
R2u 1.87	R2u 1.73	R2u 1.69
R4u 1.95	R4u 1.61	R4u 1.93

Figure 5. 9. Variation with position of speech levels SLA_N and early-decay times EDT_{mid} with the right speaker without and with reflective ceiling barriers, as measured in a scale model: a. SLA_N; b. EDT_{mid}; c. $C_{50,mid}$.

Figure 5.9 shows the measured A-weighted speech levels for 'Normal' voice (SLA_N) and EDT_{mid} values with reflective ceiling barriers for the nine receiver positions with the right speaker in the scale model. As shown in Figure 5.8, in both the occupied and unoccupied rooms, speech levels with reflective ceiling barriers did not decrease significantly, and speech levels in the side seats were slightly lower than in the center-line seats. In an occupied room, speech levels increased slightly with barriers at the left side seats (r1L, r2L, r3L). Reflective ceiling barriers decreased early decay times by between 0.2 and 0.8 s in the occupied room, and by between 0.1 and 1.0 s in the unoccupied room.

The ceiling-barrier results are reminiscent of those that occur when reflective scattering obstacles ('fittings') are introduced into an industrial workshop [11]. Reverberation times decrease sharply; steady-state levels from a sound source increase slightly close to the source due to back-scattering, and decrease farther from the source. Therefore, the alternative ceiling-reflector concept, consisting of reflective scattering obstacles suspended from the ceiling, was tested.

5.3.3 Ceiling reflectors

Figure 5.10 shows the speech levels and early-decay times for the semicircular ceiling reflectors at the three central positions with the centre speaker in the unoccupied room. Also shown are the results without barriers or reflectors, and for the R4 ceiling-barrier configuration which had the same spacing as the CF and CC reflectors. The CF reflectors decreased early-decay times by between 0.1 and 0.6 s, with the largest decreases occurring at r2C. The early-decay times were very close to those for barrier configuration R4. In this configuration, speech levels decreased by about 1 dBA at r1C, and increased slightly at r2C and r3C. The CC reflectors also decreased early-decay times, though much less than the CF reflectors or R4 barriers. The short, randomly-distributed ceiling reflectors (configurations CF30 and CF60) showed somewhat different results from those for the longer reflectors. Both CF30 and CF60 had little effect at r1C and r3C, but increased early-decay times at r3C. The CF30 reflectors had little effect on speech levels; the CF60 reflectors increased levels slightly at r1C, decreased them slightly at r2C, and left them unchanged at r3C.

Clearly, the CF reflectors were effective at decreasing early-decay time with the least reduction of speech levels. Further measurements were made with semi-circular ceiling reflectors which had sound-absorptive materials on the upper curved surfaces (configuration CFA). Figures 5.11 and 12 show the measured speech levels and early-decay times in both the occupied and unoccupied rooms. When the speaker was at front-centre, speech levels at the side seats were lower than those without reflectors; however when the speaker was at right-centre speech levels at the left side seats (r1L, r2L, r3L) were slightly higher than those without reflectors. In the occupied classroom, speech

levels decreased by between 2.4 and 4.5 dBA due to the added absorption. The CFA reflectors decreased both speech levels and early-decay times more than the CF reflectors did.
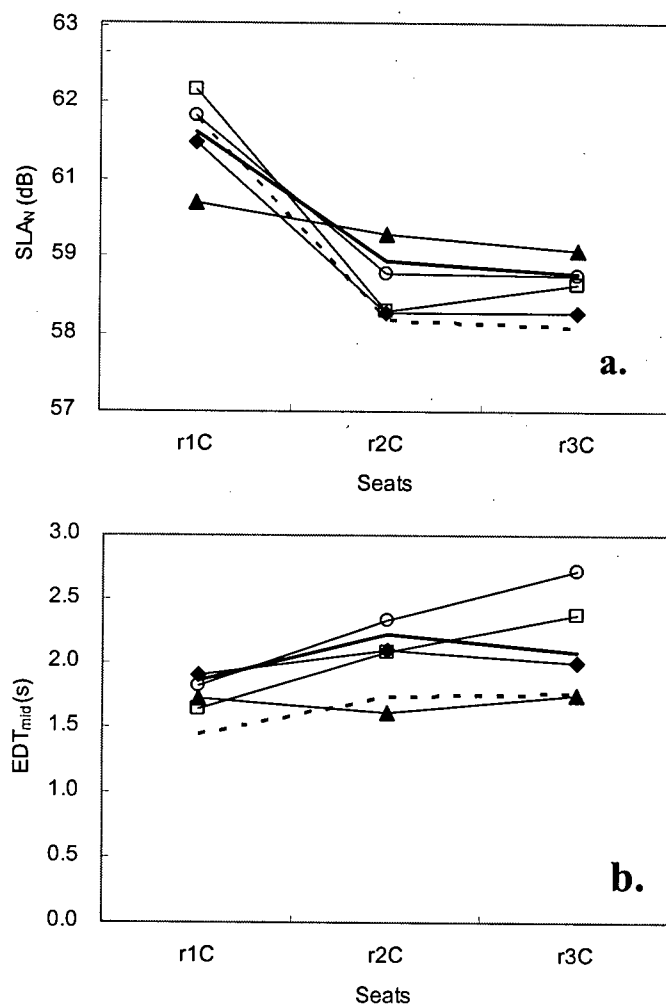
Figure 5. 10. Variation with position of (a.) speech levels, SLA_N and (b.) early-decay times EDT_{mid} in the scale-model lecture room at central positions without and with ceiling barriers and reflectors: (————) no reflectors, (.........) R4, (—▲—) CF, (—♦—) CR, (—□—) C60, (—O—) C30.

SNA_N (s)	(s)	**a.**

NOo 60.4   NOo 58.9
NOu 63.4   NOu 63.8
   ●   ●
CFo 60.8   CFo 60.4
CFAo 57.5   CFAo 58.8
CFu 63.8   CFu 63.3
CFAu 60.8   CFAu 56.3

NOo 56.0   NOo 57.1
NOu 60.5   NOu 61.5
   ●   ●
CFo 56.6   CFo 56.9
CFAo 54.1   CFAo 53.7
CFu 60.6   CFu 61.6
CFAu 55.9   CFAu 54.1

NOo 57.7   NOo 55.5
NOu 59.3   NOu 60.0
   ●   ●
CFo 56.6   CFo 56.7
CFAo 54.0   CFAo 52.2
CFu 62.4   CFu 60.6
CFAu 53.7   CFAu 54.3

EDT_mid (s)	(s)	**b.**

NOo 1.08   NOo 1.33
NOu 2.21   NOu 2.00
   ●   ●
CFo 1.16   CFo 1.35
CFAo 0.94   CFAo 1.10
CFu 1.82   CFu 2.23
CFAu 1.21   CFAu 1.40

NOo 1.76   NOo 1.56
NOu 2.74   NOu 2.24
   ●   ●
CFo 1.52   CFo 1.56
CFAo 1.46   CFAo 1.16
CFu 2.47   CFu 2.50
CFAu 1.52   CFAu 2.11

NOo 1.55   NOo 1.55
NOu 3.06   NOu 2.85
   ●   ●
CFo 1.50   CFo 1.42
CFAo 1.13   CFAo 1.53
CFu 2.23   CFu 2.46
CFAu 1.80   CFAu 1.65

Figure 5. 11. Variation with position of speech levels SLA_N and early-decay times EDT_{mid} with the centre speaker without and with ceiling reflectors, as measured in a scale model: a. SLA_N; b. EDT_{mid}.

SLA$_N$ (s)   **a.** (s)

NOo 56.5	NOo 58.3	NOo 61.3
NOu 62.9	NOu 62.9	NOu 63.0
•	•	•
CFo 60.3	CFo 60.3	CFo 60.3
CFAo 53.7	CFAo 54.9	CFAo 58.6
CFu 62.3	CFu 63.2	CFu 63.6
CFAu 58.3	CFAu 58.3	CFAu 58.7

NOo 54.7	NOo 54.9	NOo 57.6
NOu 61.2	NOu 60.5	NOu 61.0
•	•	•
CFo 58.4	CFo 58.5	CFo 57.3
CFAo 50.6	CFAo 50.7	CFAo 54.4
CFu 61.5	CFu 61.4	CFu 61.4
CFAu 55.4	CFAu 56.0	CFAu 55.7

NOo 54.7	NOo 56.1	NOo 57.1
NOu 59.2	NOu 60.8	NOu 60.7
•	•	•
CFo 59.2	CFo 59.7	CFo 57.9
CFAo 49.4	CFAo 50.6	CFAo 53.1
CFu 58.2	CFu 60.4	CFu 61.0
CFAu 54.8	CFAu 55.2	CFAu 54.8

EDT$_{mid}$ (s)   **b.** (s)

NOo 1.67	NOo 1.36	NOo 1.23
NOu 2.08	NOu 2.17	NOu 2.23
•	•	•
CFo 1.42	CFo 1.02	CFo 1.16
CFAo 0.82	CFAo 0.87	CFAo 0.85
CFu 1.96	CFu 1.97	CFu 2.08
CFAu 1.31	CFAu 1.19	CFAu 1.41

NOo 1.76	NOo 2.04	NOo 1.64
NOu 2.49	NOu 2.94	NOu 1.64
•	•	•
CFo 1.37	CFo 1.51	CFo 1.28
CFAo 1.23	CFAo 1.11	CFAo 1.07
CFu 2.11	CFu 2.28	CFu 2.77
CFAu 1.71	CFAu 1.35	CFAu 1.66

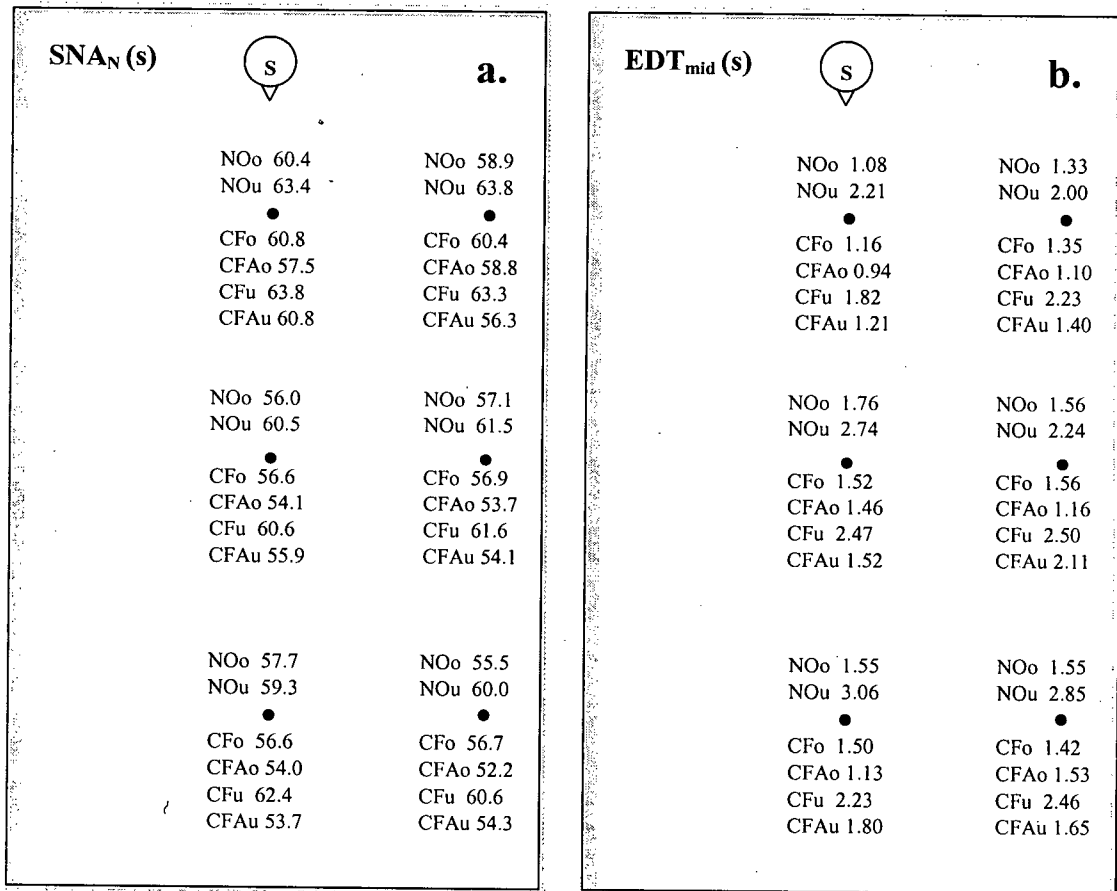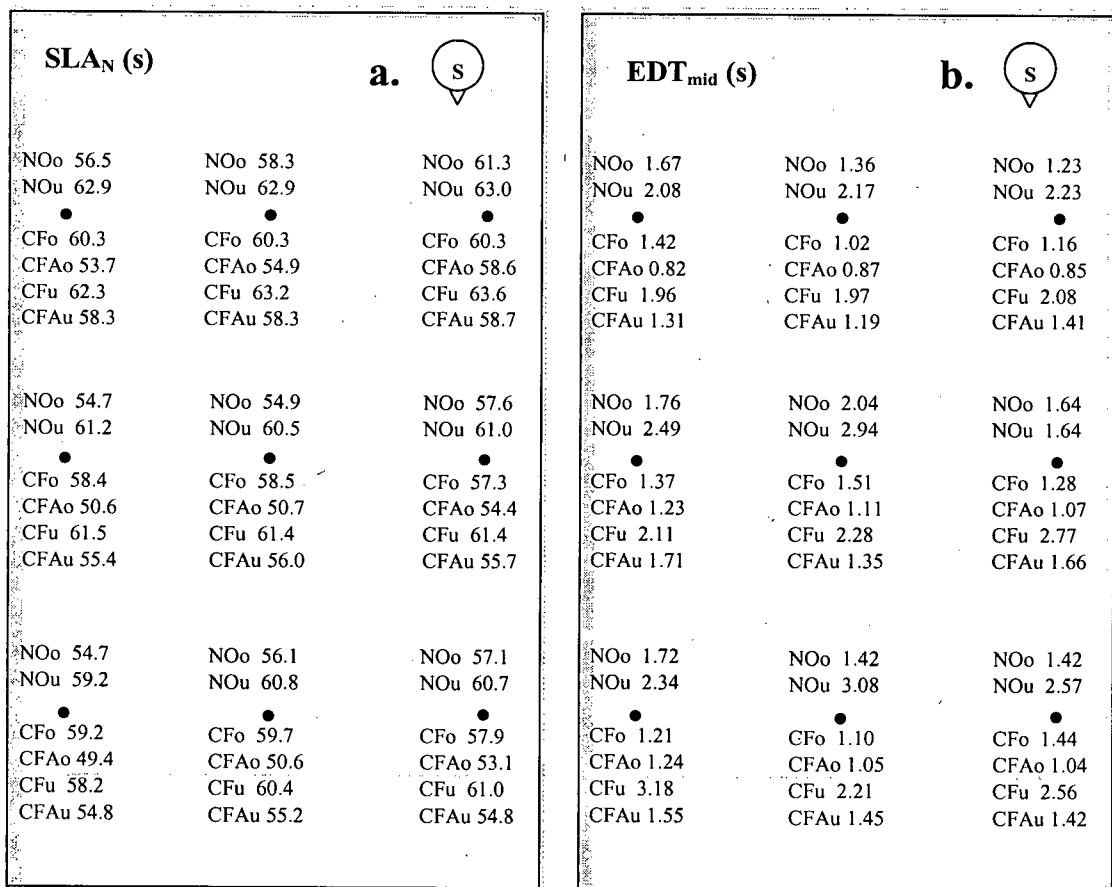| NOo 1.72 | NOo 1.42 | NOo 1.42 |
NOu 2.34	NOu 3.08	NOu 2.57
•	•	•
CFo 1.21	CFo 1.10	CFo 1.44
CFAo 1.24	CFAo 1.05	CFAo 1.04
CFu 3.18	CFu 2.21	CFu 2.56
CFAu 1.55	CFAu 1.45	CFAu 1.42

Figure 5. 12. Variation with position of speech levels SLA_N and early-decay times EDT_{mid} with the right speaker without and with ceiling reflectors, as measured in a scale model: a. SLA_N; b. EDT_{mid}.

5.4 Discussion

To reduce reverberation while minimizing decreases of the sound levels in lecture rooms, reflective ceiling-barriers can be used. Making the barriers sound-absorptive reduces reverberation, but also reduces speech levels, which is detrimental to speech intelligibility. Long, parallel obstacles of semicircular cross-section, with their flat sides facing downward, running front-to-back in the lecture room, can also be effective. They provide early reflections and increase early energy, increasing speech intelligibility. The benefit would be expected to be increase as the number of reflectors increases (i.e. as their spacing decreases).

As for the ceiling barriers, Figure 5.13 shows the predicted and measured percentage decreases of the early-decay times plotted against the decibel decreases of the speech levels for the six ceiling-barrier configurations at the center line seats, r1C, r2C, and r3C. Each set of data has been best-fit with a linear-regression trendline. For speech intelligibility, the goal is to achieve the minimum speech-level decrease with a significant *EDT* decrease. In Figure 5.13, this corresponds to the upper, left areas of the graphs. Both in the physical scale model (Figures 5.13a, c and e) and the computer prediction (Figures 5.13b, d, and f), the trendlines for the reflective ceiling barriers are always positioned above the trendlines for the absorptive ceiling barriers for small decreases of the speech levels. This means that even if the speech levels decrease a small amount, the decreases in the early-decay time would be greater with the reflective ceiling barriers than with the absorptive ceiling barriers in the center line. Thus, the early-decay times can be reduced effectively by the reflective ceiling barriers. In the computer simulation, the source-receiver distance clearly affected the results. When the receiver is farther from the source, the predicted effect of the reflective ceiling barriers is less than that measured in the scale model. This may result from the way CATT-Acoustic models the direct sound and diffuse surface reflection. Based on the results here, the sound field predicted by the computer simulation seems to be less diffuse than that measured in the physical scale model.

Figure 5.14 show the predicted percentage decreases of the early-decay times plotted against the decibel decreases of the speech levels for the six ceiling-barrier configurations at

the side seats, r1L, r2L, and r3L. The percentage decreases of the early-decay times remained nearly same as the speech level changes at each receiver position for the absorptive ceiling barriers. Thus, for the side seats some absorption is required to reduce the specular reflections from the walls. The absorptive and reflective ceiling barriers could be used in combination to optimize lecture-room acoustics for speech intelligibility upon the room layout.

In practice, it is the occupied room that is of most interest, with the occupants contributing significant absorption in the lower floor/seating region of the room. This absorption would reduce reverberation times and causes speech levels to decrease, especially at larger source/receiver distances [1]. Sound paths carrying energy from a talker at the front of the room to receivers, which involve reflection from the ceiling, would contribute more to the sound field at the receivers. Thus, it was expected that the potential effects of ceiling barriers and reflectors to increase. Figures 5.13 and 14 show the effects of the reflective ceiling barriers. The trendlines for the reflective ceiling barriers in the occupied room are always positioned above the trendlines for the reflective and absorptive ceiling barriers in the unoccupied room for small decreases or increases of the speech levels. This means that the reflective ceiling barriers can be effective even in an occupied classroom.

Figure 5. 13. Variation of percentage decrease of early-decay time with sound-level decrease at three central positions with the centre speaker with (...◊...) reflective and (...□...) absorptive ceiling barriers in the unoccupied room, and with (—♦—) reflective and (—■—) absorptive ceiling barriers in the occupied room, as measured in a scale model and as predicted, with linear trendlines: a. r1C-measured; b. r1C- predicted; c. r2C-measured; d. r2C-predicted; e. r3C-measured; f. r3C-predicted.

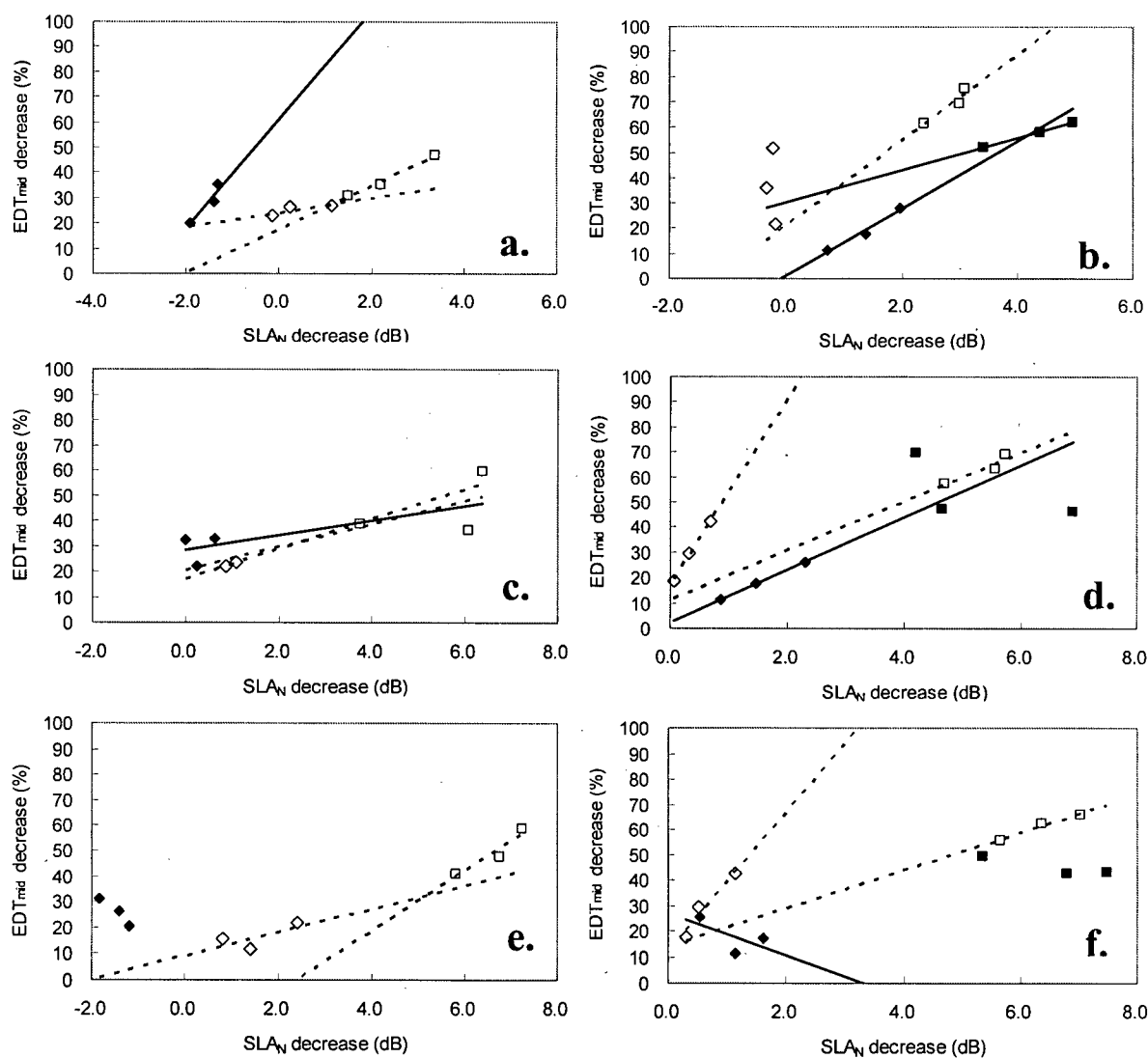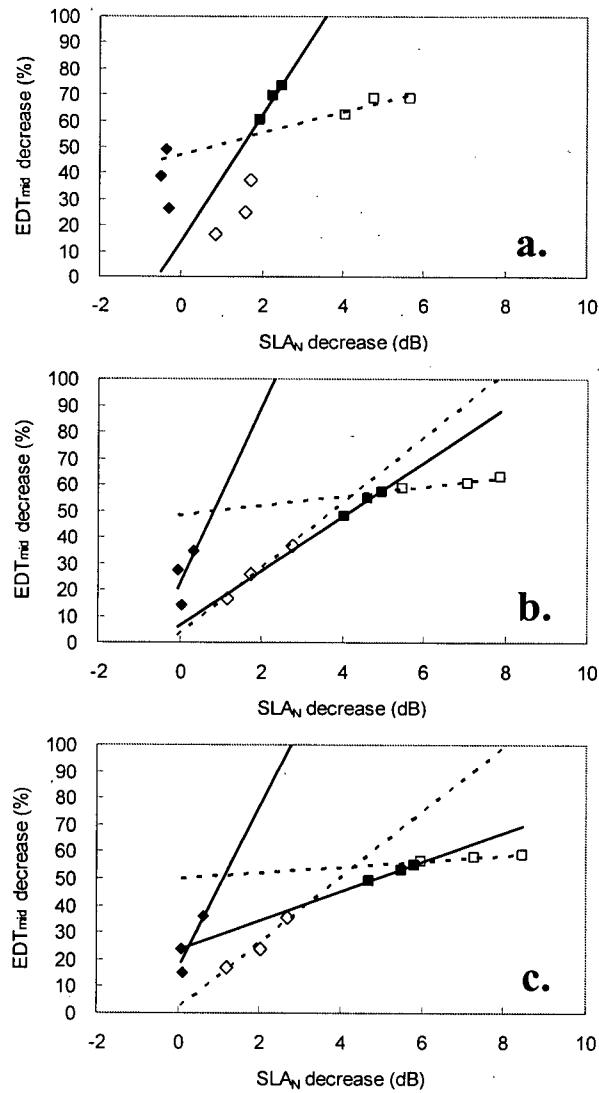**Figure 5. 14. Variation of percentage decrease of early-decay time with sound-level decrease at three side positions with the centre speaker in a lecture room with (...◊...) reflective and (...□...) absorptive ceiling barriers in the unoccupied room, and with (—♦—) reflective and (—■—) absorptive ceiling barriers in the occupied room, as predicted, with linear trendlines: a. r1L; b. r2L; c. r3L.**

5.5 Conclusions

Reflective ceiling barriers achieved the goal of decreasing reverberation with the least speech-level reduction; the effect increases with barrier density. Ceiling reflectors, in the form of long obstacles of semicircular cross-section, suspended below the ceiling in parallel, front-to-back lines with flat side down, were also effective, though a closer spacing would be desirable. However, some amount of absorption was necessary to prevent specular reflections from the walls. The shape of the semicircular ceiling reflectors was inspired by typical lighting fixtures. The results suggest that lighting fixtures could be effective at controlling lecture-room sound if they were made with flat, sound-reflecting (and, of course, optically transparent) bottoms, and arranged in long, parallel, front-to-back lines.

References

[1]   M. R. Hodgson, "Experimental investigation of the acoustical characteristics of university classrooms", J. Acoust. Soc. Am. 106(4), 1810-1819 (1999).

[2]   J. S. Bradley, R. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. 106(4), 1820-1828 (1999).

[3]   W. Yang, M. Hodgson, "Auralization study of optimum reverberation times for speech intelligibility for normal and hearing-impaired listeners in classrooms with diffuse sound fields," J. Acoust. Soc. Am. 120(2), 801-807 (2006).

[4]   M. Picard and J. S. Bradley "Revisiting speech interference in classrooms", Audiology 40(5), 221-244 (2001).

[5]   M. Hodgson, R. J. Orlowski, "Acoustic scale modeling of factories, Part I: Principles, instrumentation and techniques", J. Sound Vib. 113(1), 29-46 (1987).

[6]   H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Atmospheric absorption of sound: further developments", J. Acoust. Soc. Am. 97(1), 680-683 (1995).

[7]   M. Hodgson, "Empirical prediction of speech levels and reverberation in classrooms," J. Build. Acoust. 8(1), 1-14 (2001).

[8]   American National Standards Institute, "ANSI S3.5-1997 Methods for calculation of the speech intelligibility index," 1997.

[9]   B. -I. Dalenback, CATT-Acoustic v8.0 (Gothenburg, 2004)

[10]   M. Hodgson, "Evidence of diffuse surface reflections in rooms", J. Acoust. Soc. Am. 89(2), 765-771 (1991).

[11]   M. Hodgson, "Measurements of the influence of fittings and roof pitch on the sound field in panel-roof factories", Appl. Acoust. 16(5), 369-391 (1983).

6　CONCLUSION

6.1　Contributions

This work comprised three main themes; the auralization technique was validated for use in speech-intelligibility testing; the optimum reverberation for speech intelligibility for normal-hearing and hearing-impaired listeners was investigated using auralization; and novel architectural acoustical designs for optimizing the classroom acoustics were developed.

Following the validation of auralization for speech-intelligibility testing, this work then expanded the application of auralization in speech-intelligibility research. Tests showed that if the room to be auralized is not very sound-absorptive, or reverberant and noisy, speech-intelligibility tests using auralization are valid and reliable. In general, the prediction results showed good agreement for most room-acoustical parameters. Success in applying the auralization technique to speech-intelligibility testing can yield more facilitative [1] and reliable subjective speech-intelligibility tests than on-site speech-intelligibility tests. These were presented in Chapter 2. The fully-computed auralization procedure used in this study was done using the CATT-Acoustic — hybrid room-acoustical prediction and auralization program.

This work was motivated by discrepancies between the results of experimental studies and theoretical predictions of the optimal reverberation time for speech intelligibility. In this work, the traditional concept of the signal-to-noise ratio, which is the output sound-level difference of the signal and noise sources, was not used. Instead the concepts of the signal-to-noise level difference at listeners and the relative output power levels of the speech and noise sources were separated in the experimental setting: this had previously only been done in theoretical predictions. This made it possible to incorporate more realistic noise sources into the room models, and to eliminate the discrepancies between the results of experimental studies and theoretical predictions of the optimal reverberation time for speech

intelligibility. In particular it generally confirmed the results of theoretical predictions. This work was presented in Chapter 3 and Chapter 4.

The effects of the location of the speaker, receiver, and noise source on speech intelligibility in a room were confirmed by comparing two cases: when the noise source was closer than the speech source to the listener, and when the speech source was closer to the listener than the noise source (Chapter 3 and Chapter 4). Generally, the optimal reverberation time was zero or near zero when the noise source was farther than the speaker; non-zero reverberation times were found to be optimal when the noise source was between the listener and the speaker both, for normal and hearing-impaired listeners. This finding agrees with Hodgson and Nosal [2]'s theoretical prediction results. The optimum reverberation time in a room cannot be identified as a single defining value, because it depends on the relationship between the speakers, receivers, and noise sources, as well as on the room.

When the speech-to-noise level difference is adverse for both normal-hearing and hearing-impaired listeners, some reverberation is required to increase the speech signal and increase intelligibility. With the tested range of speech-to-noise level differences (-6 dB to 8.5 dB), hearing-impaired listeners needed more early-sound energy than normal-hearing listeners in the cases of both approximately diffuse sound fields and non-diffuse sound fields. This was shown in Chapter 3 and Chapter 4. Classroom noise levels are too high to achieve the ideal 15 dB speech-to-noise level difference [3] in many schools [4,5,6,7]. If it is hard to decrease existing noise levels (over 70 dBA) significantly, particularly in preschool classrooms where noise is frequently attributable to student activities [5,7,8], enhancing of early reflections rather than minimizing reverberation can improve speech intelligibility in a room. This can also be applied to classrooms for hearing-impaired students who do not use hearing-aids in their everyday lives. However, the priority in controlling sound for speech intelligibility is reducing noise levels.

Various types of ceiling barriers and reflectors proposed and evaluated in this work were effective at optimizing acoustical conditions for speech intelligibility. This novel system contains sound-reflective materials, in contrast with traditional sound-absorptive systems used to minimize classroom reverberance. Reflective ceiling barriers and ceiling reflectors —

in particular, parallel front-back rows of semi-circular reflectors — achieved the goal of decreasing reverberation with the least speech-level reduction. However, a small amount of absorption is necessary to prevent specular reflections from the walls. This was presented in Chapter 5. The novel system of the ceiling barriers and reflectors can be applied to classrooms to optimize acoustical conditions for speech intelligibility. Specifically, the ceiling reflectors inspired by typical lighting fixtures currently used in many classrooms can be directly installed, simply considering light-fixture arrays in classrooms.

6.2   Limitations

Auralized sound fields were validated in Chapter 2 and used in Chapter 3 and Chapter 4 for the speech-intelligibility tests in this work. Although some limitations were found on the auralization technique in sound-absorptive ($RT = 0.4$ s in the 1kHz octave band) or noisy, reverberant rooms ($SNS = 0$ dB and $RT = 2.0$ s), as shown in Chapter 2, in general the speech-intelligibility scores in the auralized classroom showed agreement with the scores in the real classroom.

The discrepancy between the on-site speech-intelligibility tests and the virtual speech-intelligibility tests in very sound-absorptive or noisy, reverberant classrooms is a weakness of the work. In very absorptive rooms the speech-intelligibility scores with auralized sound fields were higher than the scores in actual rooms; in noisy, reverberant rooms the speech-intelligibility scores with auralized sound fields were lower than the score in actual rooms. Even considering these facts, the overall findings may not change, since the decreased speech-intelligibility scores in real, absorptive rooms, and the increased speech-intelligibility scores in real noisy, reverberant rooms, support theoretical predictions in the literature that the optimum reverberation time is not zero with noise. Therefore, auralized sound fields can be used for speech-intelligibility studies, even if they do not perfectly represent the room sound field to be modeled.

The other limitation of this work is the relatively low statistical power associated with the number of hearing-impaired subjects in the speech-intelligibility tests. The tests were advertised using posters and email distribution and, specifically for recruiting hearing-

impaired subjects, local networks helped to disseminate the advertisement: the Disability Resource Centre of the University of British Columbia, local audiology clinics, and the Workers' Compensation Board of BC. For the speech-intelligibility tests in Chapter 3 and Chapter 4, the target numbers of subjects were thirty for each hearing group; however it proved very challenging to find hearing-impaired subjects to volunteer for the tests. In Chapter 3, forty-three normal-hearing and twenty-eight hearing-impaired subjects completed the speech-intelligibility tests. In Chapter 4, twenty-five normal-hearing and thirteen hearing-impaired subjects participated in the tests.

For the validation tests in Chapter 2, twelve subjects participated in the speech-intelligibility tests. The small number of subjects can decrease confidence in the current results. However, only three subjects could participate in each real classroom test, since there were three receiver positions (r1, r2, and r3) designed for the tests, as shown in Figure 2.1. In order to have more subjects in the real classroom listening tests, enormous amounts of time and classroom scheduling would have been required, which was unrealistic.

6.3 Future Work

The optimal reverberation for speech intelligibility has been explored experimentally in this work. However, this study opens some directions for future work in this area. In this study, a typical medium-sized classroom was used as the environment for the speech intelligibility test. There is a need for an experimental study using various types of classroom to improve the current work; e.g. small-sized classrooms and large-sized auditoria. As described in Chapter 3 and Chapter 4, the talker, listener, and noise sources were positioned on the centre line of the room for the speech-intelligibility tests. It would be interesting to observe whether other talker-listener-noise combinations can affect the optimal reverberation for speech intelligibility, especially when the azimuthal angle is incorporated into the talker-listener-noise combination. This can be studied by considering types of classroom activities; e.g. group activities, round table discussion, lectures, etc. Classroom activities are important factors affecting talker-listener-noise combinations. The investigation of more noise sources would be another improvement to understand speech intelligibility accurately. Only one noise signal (babble noise as a student noise) was used in this work. To obtain more realistic

classroom sound fields, different types of classroom noise sources could be studied - for example, a ventilation outlet, computer fan noise, a laser projector and other visual aids, as well as student-activity noise.

While the hearing-impaired listeners in this study benefited from early reflections more than the normal listeners, it would also be interesting to verify the result for elderly listeners or more severely impaired listeners who use hearing aids in their everyday lives to investigate optimum reverberation for hearing aids.

Reference

[1]   J. Besing and J. M. Koehnke, "A test of virtual auditory localization," Ear Hear. 16 (2), 220-229 (1995).

[2]   M. Hodgson and E.-M. Nosal, "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms," J. Acoust. Soc. Am. 111 (2), 931-939 (2002).

[3]   America National Standards Institute, "ANSI S12.60 Acoustical performance criteria, design requirements and guidelines for schools," 2002.

[4]   M. Hodgson, "Experimental investigation of the acoustical characteristics of university classrooms," J. Acoust. Soc. Am. 106 (4), 1810-1819 (1999).

[5]   M. Picard and J. S. Bradley, "Revisiting speech interference in classrooms," 40 (5), 221-244 (2001).

[6]   H. A. Knecht, P. B. Nelson, G. M. Whitelaw et al., "Background noise levels and reverberation times in unoccupied classrooms: Predictions and measurements," Am. J. Audiol. 11 (2), 65-71 (2002).

[7]   B. Shield and J. E. Dockrell, "External and internal noise surveys of London primary schools," J. Acoust. Soc. Am. 115 (2), 730-738 (2004).

[8]   W. Yang and M. Hodgson, "Acoustical evaluation of preschool classrooms," Noise Control Eng. J. 53 (2), 43-52 (2005).

APPENDIX A – Ethical Certificate

The University of British Columbia
Office of Research Services and Administration
Behavioural Research Ethics Board

Certificate of Approval

PRINCIPAL INVESTIGATOR	DEPARTMENT	NUMBER
Hodgson, M.R.	Occupational & Env Hygiene	B03-0256

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT
UBC Campus ,

CO-INVESTIGATORS
Yang, Wonyoung, Occupational & Env Hygiene

SPONSORING AGENCIES
Natural Science Engineering Research Council

TITLE
Optimum Acoustical Conditions in Classrooms

APPROVAL RENEWED DATE	TERM (YEARS)	
SEP 2 0 2005	1	

CERTIFICATION:
The protocol describing the above-named project has been reviewed by the Committee and the experimental procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval of the Behavioural Research Ethics Board by one of the following:
Dr. Peter Suedfeld, Chair,
Dr. Susan Rowley, Associate Chair

This Certificate of Approval is valid for the above term provided there is no change in the experimental procedures

APPENDIX B – Consent Forms

Chapter 2

Faculty of Graduate Studies

UBC
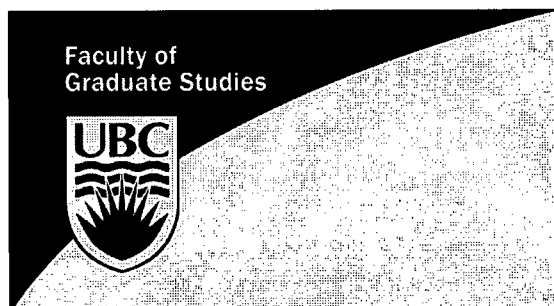
THE UNIVERSITY OF BRITISH COLUMBIA

School of Occupational & Environmental Hygiene
3rd Floor – 2206 East Mall
Vancouver, BC Canada V6T 1Z3
www.soeh.ubc.ca

(604) 822-9595 tel
(604) 822-9588 fax

INFORMED CONSENT FORM

Speech Intelligibility in Real and Virtual Classrooms

Principal Investigator: Murray Hodgson, Ph.D., Professor,
Occupational & Environmental Hygiene, UBC, 822-3073
Co-Investigators: Wonyoung Yang, MSc, Ph.D. Candidate,
Occupational & Environmental Hygiene, UBC, 822-9590

This study will form the basis of a part of Wonyoung Yang's Ph.D. research project.

Purpose:
The purpose of this study is to validate the acoustical virtual reality technique for speech intelligibility in classrooms. The project could result in improved designs for classrooms using acoustical virtual reality technique, auralization.

This study poses no significant risks to subjects.

Study Procedures:
For the purposes of this study we are recruiting normal hearing participants. Your participation in the study involves consenting to a hearing screening (comparable to those conducted in audiology clinics for hearing loss). This study consists of two parts; a live test in real classrooms and a virtual test in a laboratory. For a live test, you will be seated in a real classroom and word samples will be presented by a loudspeaker. For the first part of a virtual test, like that for a conventional hearing screening, you will be asked to raise your hand when you hear a tone presented over the headset. For the second part of the virtual test, you will be seated at a separate listening station and word samples will be presented over the same type of headset. These word samples will be presented at the same level of the live test. During the testing, you will be asked to choose what you hear on a response sheet.

You will receive your hearing screening results immediately after the listening test.
The total amount of your time needed for participating is 60 minutes per each test.

Confidentiality:
Your confidentiality will be respected. No information that discloses your identity will be released or published without your specific consent to the disclosure. No records which identify you by name or initials will be allowed to leave the Investigators' offices. Your rights to privacy are also protected by the 'Freedom of Information and Protection of Privacy Act' of British Columbia. This Act lays down rules for the collection, protection, and retention of your personal information, by public bodies, such as the University of British Columbia. Further details about this Act are available upon request. Your results will be securely stored for possible future follow-up analyses.

Compensation:
You will be compensated $10.^{00} for participating in this research study.

Contact for information about the study:
If you have any questions or desire further information with respect to this study, please contact: Prof. Murray Hodgson at 604-822-3073, or Won Yang at 604-822-9590.

Contact for information about the rights of research subjects:
If you have any concerns about your treatment or rights as a research subject, please contact the Research Subject Information Line in the University of British Columbia Office of Research Services at 604-822-8598.

Consent:
Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without adversely affecting your relationship with the investigators, any of your health care providers, or the University of B.C.
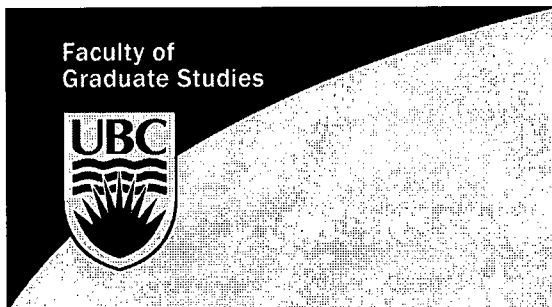
Your signature below indicates that you have received a copy of this consent form for your own records.

Your signature indicates that you consent to participate in this study.

Printed Name: _____

Signature of Participant: _____ Date: _____

Chapter 3, Chapter 4

INFORMED CONSENT FORM

Optimum acoustical conditions for normal and hard of hearing people in classrooms

Principal Investigator: Murray Hodgson, Ph.D., Professor,
Occupational & Environmental Hygiene, UBC, 822-3073
Co-Investigators: Wonyoung Yang, MSc, Ph.D. Candidate,
Occupational & Environmental Hygiene, UBC, 822-9575

This study will form a part of Wonyoung Yang's Ph.D. research project.

Purpose:
The purpose of this study is to determine the optimal reverberation for normal and hard of hearing people in classrooms. Reverberation, or the process of how speech is perceived, is regarded as an important issue with regards to speech intelligibility. The project could result in improved designs for classrooms for both normal and hard of hearing people.

This study poses no significant risks to subjects.

Study Procedures:
For the purposes of this study we are recruiting both normal and hard of hearing participants. Testing for normal and hard of hearing participants is identical. Your participation in the study involves consenting to a hearing screening (comparable to those conducted in audiology clinics for hearing loss). For the first part of the survey, like that for a conventional hearing screening, you will be asked to raise your hand when you hear a tone presented over the headset. For the second part of the survey, you will be seated at a separate listening station and word samples will be presented over the same type of headset. These word samples will be presented at a moderate listening level. During this portion of the testing, you will be asked to select what you hear on a response sheet. The word samples will be presented in such a way as to simulate a classroom setting. Various reverberations will be incorporated into the test using a specialized computer program.

You will receive your hearing screening results immediately after the listening test. The total amount of your time needed for participating is 50 minutes on one occasion only.

Confidentiality:
Your confidentiality will be respected. No information that discloses your identity will be released or published without your specific consent to the disclosure. No records which identify you by name or initials will be allowed to leave the Investigators' offices. Your rights to privacy are also protected by the 'Freedom of Information and Protection of Privacy Act' of British Columbia. This Act lays down rules for the collection, protection, and retention of your personal information by public bodies, such as the University of British Columbia. Further details about this Act are available upon request. Your results will be securely stored for possible future follow-up analyses.

Compensation:
You will be compensated $10.^{00} for participating in this research study.

Contact for information about the study:
If you have any questions or desire further information with respect to this study, please contact: Prof. Murray Hodgson at 604-822-3073, or, Won Yang at 604-822-9575.

Contact for information about the rights of research subjects:
If you have any concerns about your treatment or rights as a research subject, please contact the Research Subject Information Line in the University of British Columbia Office of Research Services at 604-822-8598.

Consent:
Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time without adversely affecting your relationship with the investigators, any of your health care providers, or the University of B.C.

Your signature below indicates that you have received a copy of this consent form for your own records.

Your signature indicates that you consent to participate in this study.

Printed Name: _____

Signature of Participant: _____ Date: _____

APPENDIX C – 300 MRT Words in a Response Sheet

RESPONSE SHEETS

Subject NO. _____

Please, circle the word you hear.

LIST 1

Number	a	b	c	d	e	f
1	led	bed	red	wed	shed	fed
2	shop	mop	cop	top	hop	pop
3	wick	lick	pick	tick	sick	kick
4	paw	jaw	saw	thaw	law	raw
5	pen	then	den	ten	hen	men
6	teach	tear	tease	teal	teak	team
7	tan	tack	tam	tab	tang	tap
8	sung	sun	sud	sub	sup	sum
9	cold	gold	hold	told	fold	sold
10	win	tin	din	fin	pin	sin
11	sip	sing	sit	sin	sill	sick
12	seethe	seek	seed	seep	seen	seem
13	sake	safe	same	save	sale	sane
14	sat	sass	sack	sad	sap	sag
15	sip	hip	lip	rip	dip	tip
16	nest	vest	test	rest	best	west
17	ray	raze	rate	rave	rake	race
18	puff	puck	pun	pub	pus	pup
19	not	tot	got	pot	hot	lot
20	pin	pip	pick	pit	pig	pill
21	peas	peak	peach	peat	peal	peace
22	pay	pale	page	pane	pave	pace
23	pat	pad	path	pack	pass	pan
24	fun	bun	run	nun	sun	gun
25	heat	seat	beat	neat	feat	meat

Number	a	b	c	d	e	f
26	way	day	gay	may	say	pay
27	math	map	mass	mad	mat	man
28	cook	shook	book	look	hook	took
29	mark	bark	park	lark	hark	dark
30	lane	lay	late	lake	lame	lace
31	till	bill	fill	kill	will	hill
32	kill	kin	king	kid	kit	kick
33	reel	feel	eel	keel	heel	peel
34	heat	heave	heal	heap	heath	hear
35	sang	fang	hang	gang	bang	rang
36	name	tame	fame	game	came	same
37	tale	pale	male	gale	bale	sale
38	wit	sit	kit	fit	bit	hit
39	fib	fill	fig	fin	fizz	fit
40	dug	dung	dud	dub	dun	duck
41	dim	din	dig	dip	did	dill
42	fig	pig	big	dig	wig	rig
43	went	sent	bent	dent	tent	rent
44	cup	cud	cut	cub	cuff	cuss
45	cake	came	cave	cane	case	cape
46	bus	buck	buff	but	bug	bun
47	rust	dust	just	bust	must	gust
48	oil	soil	toil	boil	foil	coil
49	beach	beat	beak	bead	beam	bean
50	bat	ban	bad	bass	back	bath