Studies of Speech Intelligibility in Classrooms

By

Cheung, Man Lung Stanley

A thesis submitted in partial fulfillment of the requirements for the Degree of Master of Philosophy

Department of Mechanical Engineering

The Hong Kong Polytechnic University

February, 2004

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written nor material which has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________(Signed)

Cheung, Man Lung Stanley (Name of Student)

Abstract

The classroom is the place in which spoken language communication takes place to enable students to learn essential knowledge. Thus, the classroom can be defined as a communication channel for learning. My research is to focus on the special requirement and parameters placed on this communication channel by the listeners. The objective of my is to quantify the effects on the listeners in terms of their age, hearing status and background knowledge status. In addition, English, Cantonese and Mandarin proficiency on speech intelligibility in silent or noisy area is also examined to find out the signal-to-noise ratios that enable all students to obtain high intelligibility. The purpose of these studies is to apply different parameters such as "signal-to-noise ratio", "early decay time"(EDT), "speech transmission index"(STI) and "reverberation time"(RT) to predict the classroom characteristics. Besides, analysis of the younger generations for whom Cantonese, Mandarin and English are either their first or secondary language will be reported. Finally, various methods of estimating signal-to-noise ratios at different positions inside the classroom will be introduced, such as "ray tracing and image source method of computer simulation" and "Rapid Speech Transmission index"(RASTI) estimation in "Maximum Length Sequence System Analysis"(MLSSA), which predicts the intelligibility of different languages.

Acknowledgements

I would like to thank my supervisor, Dr. Kai Ming Li, for his invaluable guidance, continuous supervision, encouragement and advice throughout this work. His patience, corrective and insightful comments on the manuscript of this thesis are also gratefully appreciated. Special thanks to the group of the teachers from the Department of Chinese Language and Culture in the Hong Kong Polytechnic University for their kind support and participation of the speech intelligibility tests. I would like to acknowledge the Research Office of the Hong Kong Polytechnic University for offering me this studentship and opportunity of studying.

Finally, my most sincere thanks to my parents who brought me up. I would also like to thank my wife, Maggie Wong, for her love, encouragement and emotional support.

List of Figures and Captions

Chapter 1

Chapter 2

Chapter 3

List of tables

List of Symbols

r	Density of air
S_o	Radiating surface
S_B	Boundary surface of room
c	Velocity of sound
$G_{(r\ ,r0)}$	Green' s function for a rectangular room
$G_{E(r\ ,r0)}$	Representation of the Green' s function in terms of eigenfuctions
$G_{o(r\ ,r0)}$	Representation of the Green' s function in terms of acoustic images
n	Unit normal
n_j	Unit normal to the j-th surface element
N_o	Number of surface elements on radiating surface
$p_{(r)}$	Acoustic pressure field
P_j	Acoustic pressure on the j-th surface element
V	Particle velocity
P	Pressure
$\tilde{N}$	Gradient of
t	Time
a	Sabine absorption coefficient
f	Velocity potential
s_e	Specific flow resistance
W	Angular frequency
a_e	Porosity
f	Acoustic velocity potential
k	Wave number
f	Frequency
R	Distance from source point r_o to measurement point r
A	Constant that depends on the source strength
z_s	Source at a distance from a plane boundary
u_{x-}, u_{x+}	Acoustic reflection coefficient at $x = 0$ and $x = l_x$
e_n	Neumann symbol (= 1 if $n = 0$, = 2, otherwise)
V	Volume of room
V_{int}	Region inside S_0
V_{ext}	Region outside S_0 but inside S_B
b	Specific acoustic admittance ratio of the surface
q	Angle of incidence
d	Delta-function source
$J_0^{(1)}$	Bessel function of first kind of zero order
K	Horizontal wave number
$k_j(z)$	Medium wave number for layer
A, B	Arbitrary coefficient
N	Number of layers
k_z	Wave number in z direction
b_b	Admittance of the lower plane
f_D	Velocity potential includes the screen and the ground effect
¶$_n$	Normal derivative
f_n	Room eigenfunction
S	Screen for integration

R(q)	Plane-wave reflection coefficient
Q	Slow varying function with respect to horizontal distance
p	Sound field, a function of r and z
k_0	Reference wave number
N(r,z)	Index of refraction
TL	Transmission loss
TL	Insertion loss
CAN	Speech intelligibility scores of Cantonese
MAN	Speech intelligibility scores of Mandarin
ENG	Speech intelligibility scores of English
RT	Reverberation time
EDT	Early decay time
S/N	Signal-to-noise ratio
C_{50}	Clarity of 50ms
C_{80}	Clarity of 80ms
U_{80}	Useful and detrimental ratio 80ms
d_{ij}	Kroenecker delta (=1 if i = *j*, = *0*, otherwise)
STI	Speech transmission index
SI	Speech intelligibility
AI	Articulation index

Contents

Chapter1

Introduction and literature review

1.1 Identification of problems

It is controversial in Hong Kong whether the mother tongue, Cantonese, should be used as the medium of instruction in classrooms. Some believe that use of the mother tongue as the teaching medium facilitates the transfer of ideas. However, others maintain that using a second language such as English is much more important for the development of knowledge. Since English is the international language and Hong Kong is an international city, it is very important for students in Hong Kong to master the English language.

There is a continual debate about which language is the best medium for communication in classrooms. Should Cantonese, English or even Putonghua be used in classrooms for the transfer of ideas in Hong Kong? In this thesis, we do not intend to take part in this debate, but rather to explore the speech intelligibility of these three different languages in classrooms of various shapes, sizes and volumes, since teachers use different combinations of vowels and consonants in their verbal communication with students.

There have been many experimental studies on speech intelligibility in classrooms in which English was used as the medium of instruction [1, 3-5, 10-15]. Although Houtgast [65, 72-75] and Kang [31-32] compared speech intelligibility between different languages, there have been relatively few related studies focused on these language-specific effects in classrooms. There is also strong experimental evidence [35, 72] suggesting that different languages have different speech syntheses and intensity levels. Hence, the vowel and consonant tones of different languages demonstrate different levels of intelligibility inside the same environment. In an enclosed environment, the variations of such acoustical parameters as sound pressure levels (SPL), signal-to-noise ratios (S/N), early decay times (EDT), and speech transmission index (STI) are inter-correlated and complicated to analyze. However, these acoustical parameters can be used as indicators for the speech intelligibility of different languages in the classroom. The main theme of this thesis is to explore the use of SPL, S/N, EDT, STI and other acoustical parameters to assess the acoustic performance of three different languages in classrooms.

1.2 Parameters for determining the acoustic characteristics in a room

1.2.1 Reverberation time

The sound in a room is composed of that coming directly from the source and sound reflected by the walls and objects in the room. The overall array of these multiple reflected rays is called the reverberant sound field. It decreases with the amount of sound absorption in the room, and is affected by the room shape and absorbent contents. The reverberation affects speech intelligibility by affecting the arrival of early and late sound energies [29]. The early energy is defined as the total energy radiated by the speech source which arrives at a receiver less than 50 ms after the arrival of the direct sound. The late energy is usually defined as the total energy radiated by the speech source which arrives at a receiver position more than 50ms after the arrival of the direct sound. The clarity and intelligibility of speech is directly related to the ratio of early-arriving to late-arriving sound energy [29, 33]. An excessive reverberation in a classroom leads to a reduced ratio of early-to-late energy, and therefore to lower intelligibility of speech. Reverberation can be quantified by a number of measures, such as the early-to-late energy fraction, reverberation time (RT) and early decay time (EDT). It is interesting to note that the early-to-late energy fraction, and hence speech intelligibility, can be improved without changing the overall RT or EDT. These parameters, which are normally measured in octave bands, tend to be highly correlated in the case of real classrooms [33].

The present investigation aims to study the effects of the signal-to-noise ratio and reverberation on the design of classrooms. A number of measures, such as the speech transmission index (STI) and speech intelligibility (SI) [13-15, 19, 26, 29, 33-36, 48-60], have been proposed to quantify the combined effect of these two factors on verbal communication. The most important octave-band frequencies for speech intelligibility are 2 kHz, followed by 1 and 4 kHz [4, 11, 68]. A different weighting factor is applied at each octave band in the case of calculation of the STI, by summing all contributions. These weighting factors indicate their relative contribution to overall speech intelligibility.

Researchers [3-5, 48-60] have suggested that in order to obtain high speech intelligibility for adults with normal hearing working in their first language, the signal-to-noise ratio must exceed 15 dB and the reverberation time must be within the range of 0.5 to 0.7 s. The requirements are even more stringent for people who are acoustically more challenged, such as children, people with impaired hearing and people who use a second language for communication. An optimal reverberation time at a mid-frequency range is recommended at a range between 0.4 and 0.5 s. A minimum signal-to-noise ratio of 20 dB, and a maximum background-noise level of 25 to 30 dB have been suggested in addition to the recommended reverberation time. An important issue, seldom discussed, is whether these criteria should be applied to occupied or unoccupied classrooms. Although most experiments have been conducted in occupied classrooms, architects are more likely to have specified their design criteria for unoccupied classrooms.

The reverberation times may be measured by charting the decay of sound following a gunshot or balloon burst, or it may be estimated from:

$$R = \frac{0.16V}{S\bar{a}} \tag{1.1}$$

where V = room volume (m^3)

R = reverberation time (s)

S = surface area of the room (m^2)

$\bar{a}$ = average room absorption coefficient

This concept was first proposed by W. C. Sabine [67] in the 19th century. Eyring [22] provided a modification of the Eyring's formula as follows:

$$RT_{60} = \frac{55.2V}{-Sc\ln(1-\bar{a}) + 4hV} \qquad (1.2)$$

where c is the sound speed measured in m s^{-1} and S is the total surface area of the space.

According to the comparison of Dance and Shield, [16-18], the average absorption coefficient of the space is usually greater than 0.2. They also showed that the Eyring formula is not applicable in some situations. The average surface area should be multiplied independently by the respective absorption coefficient before the summation in the denominator shown in Equation (1.3). Millington [46] modified the equation further particularly at the term multiplication of the absorption coefficient. Each term of the surface area and absorption coefficient was considered independently rather than simply taking the average value.

The modified formula for the reverberation time, RT, is as follows:

$$RT = \frac{0.161V}{\sum_i S_i a_i} \qquad (1.3)$$

where S_i is the room surface area for each individual surface and α_i is the corresponding absorption coefficient.

The Sabine, Eyring and Millington formulas all give reasonable prediction of the reverberation time in an enclosed space. However, there are uncertainties which one of these three models gives the most accurate prediction of the reverberation time. Dance and Shield [16-18] introduced the “Millington absorption coefficients” to represent the most accurate prediction of RT in enclosed spaces with absorbent surfaces.

In addition to these analytical formulations for calculating the reverberation time, there are different types of computer models for predicting the acoustic performance in a room. Essentially there are four types of computer models that have been developed in recent years for prediction of the reverberation time. These are the image-source method, the ray-tracing technique, the beam-tracing approach and the sound particle tracing model.

Dance and Shield [16-18] compared the accuracy of several types of absorption coefficients, such as the Sabine, Eyring and Millington formulas used in different prediction schemes. The four different approaches: “classical theory”, “numerical solutions”, “empirical expressions”, and “physical scale and mathematical models” were studied. The authors discovered that the “image source method” and the “ray tracing method” were more appropriate than others. Some typical examples of the introduction of the ray-tracing model and the image-source method to predict the RT in a room were given by Dance and Shield. They used these numerical models to estimate and compare the overall absorption coefficients in a factory space containing a barrier.

According to Dance and Shield, use of the Millington formula at various receiver positions and frequencies gave a 10.9% prediction error in the reverberation times. This compared to respective errors of 12.8% and 20.1% for the Sabine and Eyring formulas. Their studies established the validity of the use of the Millington absorption coefficient for the ray-tracing model or the image source method in predicting reverberation time.

1.2.2 Signal-to-noise ratio

The signal-to-noise ratio is equal to the level of speech minus the level of background noise, both at the listener's position. The speech level depends mainly on the speaker's voice level, the distance between the speaker and the listener, and on the acoustical characteristics of the classroom. The background noise comes from noise sources like ventilation systems, classroom equipment such as the projector and computers, student activities in the classroom, and sources outside the classroom. The levels are dependent on the acoustical characteristics of the classroom. The average speech level, background noise level and signal-to-noise ratio are normally expressed in terms of A-weighted values.

1.2.3 Early decay time

The early decay time (EDT) is sometimes favored over the reverberation time to assess the 'absorptiveness' of a classroom. In an unoccupied classroom, most empirical models for predicting the early decay time are usually based on the classic diffused-field theory [20, 33, 56]. The EDT can be expressed as a function of the room volume [57], the total surface area, and the amount of sound absorption in the room as follows:

$$EDT = \frac{0.16}{VA_u} \tag{1.4}$$

where EDT is specified as a function of room volume, total surface area and the amount of sound absorption in the room;

$A_u = a \cdot S + 4mV$ is the total unoccupied classroom absorption in m^2;

α is the average classroom-surface absorption coefficient;

S is the total surface area of the classroom, in m^2;

m is the air-absorption exponent in Np/m at a specific frequency; and

V is the volume of the classroom, in m^3.

1.2.4 Clarity index

The clarity index, C_{50}, is a commonly-used criterion for interpreting perception of the speech intelligibility of the signal received. It is defined as the ratio of the early- to late-arriving sound energy in the first 50 ms [65] as follows:

$$C_{50} = 10\log\left[\frac{\int_0^{50ms} p^2(t)dt}{\int_{50ms}^{\infty} p^2(t)dt}\right] \text{ dB,} \qquad (1.5)$$

where the function $p(t)$ is the impulse response of the sound pressure at the receiver location. An alternative index for speech clarity may be defined as C_{80}. The division between early and late arriving sound for speech sounds is taken as 80 ms after the arrival of the direct sound and the early-to-late arriving sound ratio. Similarly, C_{80} can be expressed as

$$C_{80} = 10\log\left[\frac{\int_0^{80ms} p^2(t)dt}{\int_{80ms}^{\infty} p^2(t)dt}\right] \text{ dB} \qquad (1.6)$$

The clarity index C_{80} has gained considerable acceptance as a subjective parameter of musical clarity [65], but it has now also been used for speech intelligibility.

1.2.5 Useful-to-Detrimental ratio

The Useful-to-Detrimental ratio is another essential parameter that has been developed to take into account the effects of room reflections and ambient noise on speech intelligibility. This terminology originates from two criteria known as 'Deutlichkeit' and 'Definition'. They are used as subjective parameters for the perception of loudness. The 'Definition-50' (D_{50}) of the sound energy is defined as the 'useful' sound energy arriving within the first 50ms to the total sound energy. It can be calculated from the impulse response *p*(t) at the observed point as follows:

$$D_{50} = \frac{\int_0^{50ms} p^2(t)dt}{\int_0^{\infty} p^2(t)dt} \times 100\% \qquad (1.7)$$

1.2.6 Speech Transmission Index (STI)

The Speech Transmission Index (STI) is an objective criterion that characterizes speech intelligibility. The STI is an indicator that takes into account all the possible causes of speech intelligibility alterations, except the non-linear effect [73]. Any modifications of the room characteristics can result in an effective reduction of the signal modulation with a delay. The Modulation Transfer Function (MTF) may be obtained from the impulse response of the room *p*(*t*) by calculating the modulation index. MTF firstly need to be evaluated from the modulated sound intensity, which emitted by sound source and sound intensity at receiver in the following equations:

$$I_i(t) = \overline{I_i}\left[1 + m_i \cos(2pFt)\right] \qquad (1.8)$$

$$I_0(t) = \overline{I_0}\left[1 + m_0 \cos(2pF(t - t))\right] \qquad (1.9)$$

where $\overline{I_i}$ and $\overline{I_0}$, m_i and m_0 are average intensities and modulation indices at sound sources and receiver, respectively. F is the modulation frequency. t is the time constant dependent on the positions of sound source and receiver. Then, the modulation index, which is a function of frequency *F*, is defined by the ratio as follows:

$$m(F) = \frac{m_0}{m_i} \quad , \tag{1.10}$$

Using the modulation index [68], we can compute the signal-to-noise ratio per frequency band by:

$$S/N_k = 10\log\frac{m(F)}{1-m(F)} \quad , \tag{1.11}$$

where *k* is the corresponding octave band. The overall mean is found by combining the seven octave band specific means after each S/N_K value has been weighted by a factor W_k. The weightings which will be derived from a standard speech spectrum. These weightings have been developed from the results, which obtained from subjective tests such as using the phonetically balanced (PB) words or articulation index (AI) as usually described in chapter 2.

Truncating the effective range of the S/N_K ratio at each octave band to ±15 dB, we can calculate the average value of the signal-to-noise ratio as

$$\overline{S/N} = \sum_k W_k (S/N)_k \tag{1.12}$$

The value of W_k, which is refer to the seven octave band with center frequencies from 125Hz to 8kHz, are 0.13, 0.14, 0.11, 0.12, 0.19, and 0.17 respectively. [71-72]

Finally, the mean value of signal to noise ratio is converted to a STI value as follows:

$$STI = \frac{[\overline{S/N} + 15]}{30} \quad . \tag{1.13}$$

The equations shows that the STI value is basing on the range from 0 to 1.

However, the assumption of diffuse field should be clearly made [50-51], since the diffuse field theory should be implacable and inherent in the above equations.

1.3 Theoretical prediction models

There are three useful theories for predicting sound fields in enclosed spaces. The first is the diffuse field theory. The other two are the wave acoustics theory and the geometrical acoustics theory. The later two theories are useful for the prediction of room impulse responses.

The geometrical acoustics model is more popular for predicting the acoustical characteristics of an enclosed space. The basic principle of geometrical acoustics is that the sound is considered to propagate just like light rays, in geometrical optics. The effects of diffraction and refraction are not taken into account in this model. Only the laws of reflection, absorption by the walls, and absorption in the air are considered.

Two basic algorithms have been developed over the centuries. The first is known as the ray tracing method, and the second as the image source algorithm. They are described in the following sections.

1.3.1 The Ray-Tracing Algorithm (RTA)

Similar to the image-source model, the ray tracing method is based on geometrical room acoustics [48-59]. However, the ray tracing method allows the inclusion of the effect of sound scattering as it occurs when sound is reflected from rough surfaces. The scattering effect is not negligible if we do not want to restrict ourselves to obtain the prediction of sound fields at low frequencies.

One assumes that the energy emitted by the sound source is distributed into a discrete number of sound rays. Each ray has an initial energy equal to the total energy of the source, which is divided by a finite number of rays. Each ray travels at the speed of sound and collides with the walls, floor and ceiling, where it is reflected in accordance with the law of specula reflection. Its energy level decreases at reflections by means of the absorption of surfaces. The energy level is also reduced as the sound rays travel inside the medium due to the absorption of sound in air.

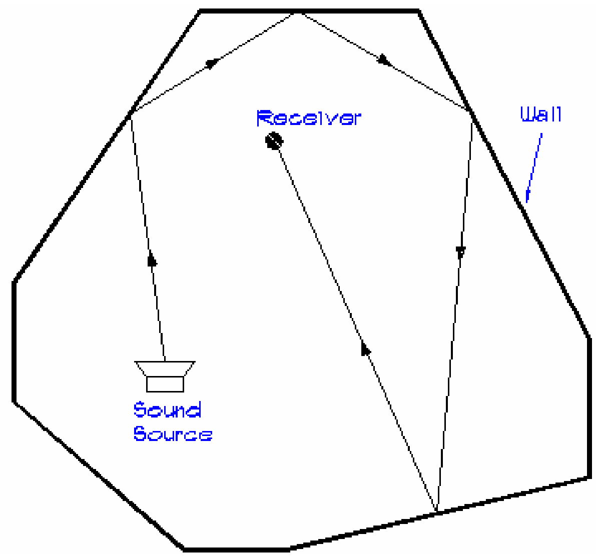

Fig. 1.1 Tracing a sound particle from the sound source S to receiver R.

Figure 1.1 shows a typical ray emanating from a sound source interacting with boundaries before it reaches the receiver. In order to calculate the sound energy at different points inside a room, receiver cells with finite volumes are defined. Each ray is checked to see whether it crosses the receiver volume. The number of rays crossing a receiver volume and the energy contributions of those rays give a measure of the sound pressure level. Losses due to spherical divergence are included as a result of the increasing separation between the rays as they spread out from the source with increasing travel time. The effect of absorption of the sound in air can be incorporated and its effect becomes more important at high frequencies.

1.3.2 The Image Source Algorithm (ISA)

The Image Source Algorithm [45], ISA, is based on the idea that the sound source can be thought of as the origin of the image source, which is the image of the original sound source in the mirror formed continuously by the boundary walls. The contribution from each image source is calculated by considering the total distance of the reflected sound ray before its arrival at the receiver position. The total sound field can then be evaluated by summing the contributions from all image sources.

There are some limitations imposed on the use of the image source algorithm, because the computational time for determining the higher-order image sources is relatively higher. Hence, the computational times required for the evaluation increase exponentially with the increasing order of image sources. More importantly, the effects of diffuse reflections cannot be modeled in the ISA. Also, some of the image sources may not be visible to the receptor and a test of 'visibility' must be performed. The computational time required for calculation of the impulse response due to a point source increases exponentially with the number of reflecting surfaces in a room. As a result, it is rather impractical to use solely the ISA in most practical situations.

Concerning the use of the geometrical acoustics in the image source algorithm, the incorporation of the diffusion coefficients of all boundary surfaces is one of the greatest difficulties. The predicted results can vary significantly as there is neither a database nor a standardized method for determining the diffusion coefficients of room surfaces. Some particular rules for choosing these coefficients have been proposed by Vorländer [60], but doubts have also been cast [50-51, 60] on their validity in determining the magnitudes.

Both the ray-tracing method and the image-source algorithm (ISA) are based on geometrical acoustics. In an enclosure, sound rays can be traced from a receiver point to the sound source. For this purpose, virtual image sources must correspond to the room shape. We start by constructing the first-order mirror images with respect to all walls. We then proceed in the same way to generate the second-order image sources. This process can be repeated up to a prescribed order of the image sources. For instance, an image source of the fourth order represents a sound ray that has suffered four-wall reflections.

Figure 1.2 illustrates a visibility test for an image source of the second order. The actual receiving point, R, is connected with the image source (S_{12}). Hence, the visibility is established. For image source S_{12}, the last index indicates that wall 2 is involved in the last reflection of the sound path in the mirroring algorithm. If the intersection point of this connecting line is situated within the polygon that defines the actual wall portion on this plane, the image source is considered visible. We proceed in the same way for image source S_{12} by considering the intersection point for the ray path connecting receiver point R and image source S_1 in the chain of image sources. This procedure is repeated until it ends at the original sound source.

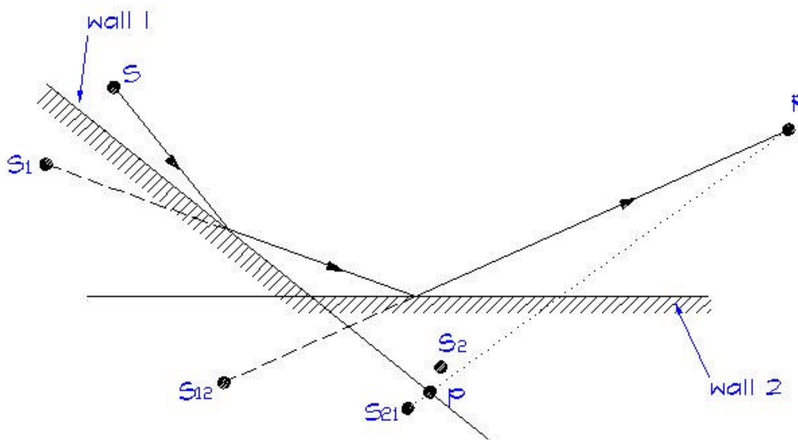

Fig. 1.2 Diagram of Image-Source Algorithm.

In irregular rooms, visibility tests are needed to confirm the presence of a particular image source. The example in the Figure 1.3 shows that receiver microphone R_1 can be reached by a first-order image, while receiver microphone R_2 cannot. In other words, R_1 is 'visible' from S_1 but R_2 is not. Consequently, it is necessary to check all calculated intersection points to confirm whether they are situated within the real physical boundaries.

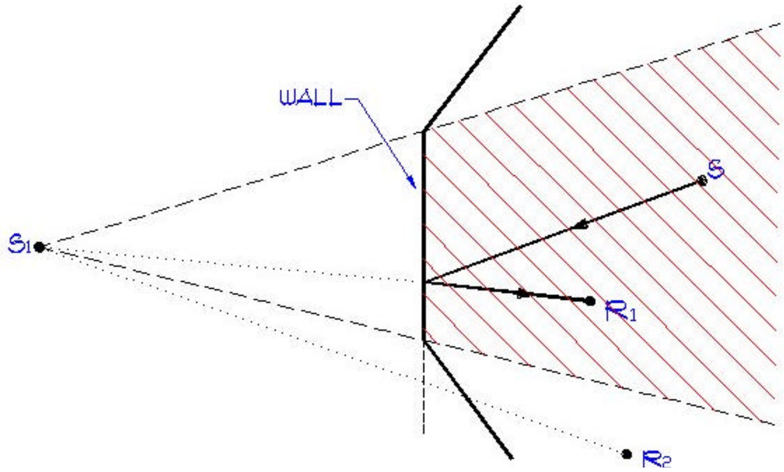

Fig. 1.3 Diagram of Visibility Test.

Owing to these visibility tests, the ISA suffers from long computation times. The problem is particularly acute when a large number of surfaces is involved and when the mean absorption of the room is small. Both conditions lead to the necessity for more images in the computation in order to achieve the required accuracy.

1.3.3 Image method with formulated Green's function

The approximate image-formulated Green's function for a rectangular room is given by [28]:

$$G_1(r,r_0) = \frac{e^{ikR}}{4\pi R} + \sum_{\substack{j=-\infty \\ \neq 0}}^{\infty} f_l(\mu_x) f_m(\mu_y) f_n(\mu_z) \frac{e^{ikR_j}}{4\pi R_j}, \quad (1.14)$$

where

$$R_j = |r - r_j|, \quad (1.15)$$

$$r_j = [l l_x + (-1)^l x_0, m l_y + (-1)^m Y_0, n l_z + (-1)^n z_0], \tag{1.16}$$

$$f_l = \begin{cases} m_{x-}^{|l/2|} - m_{x+}^{|l/2|} & l \text{ is even} \\ m_{x-}^{|l-1|/2} - m_{x+}^{|l-1|/2} & l \text{ is odd} \end{cases}, \tag{1.17}$$

and μ_{x-} and μ_{x+} are the reflection coefficients of the walls at x = 0 and x = l_x respectively. Also, *Rj* is the distance from source point r_j to measurement point *r* with respect to the *j*-th image of the reflective wall in the enclosure.

The reflection coefficients can be related to the Sabine energy absorption coefficient α by:

$$\alpha = 1 - \mu^2. \tag{1.18}$$

The Green's function can also be expressed as a sum over the eigenfunctions (normal modes) of the room. Since the eigenfunctions Φ_n of a rectangular room with rigid walls form a complete set, the approximate Green's function with β small and constant over each wall can also be written in the form:

$$G_E(r,r_0) = \sum_{n=0}^{\infty} \frac{\Phi_n(r)\Phi_n(r_0)}{V\Gamma_n(k_n^2 - k^2 - it_n)}, \tag{1.19}$$

where

$$\Phi_n(r) = (\cos k_{nx}x) \cdot (\cos k_{ny}) \cdot (\cos k_{nz}z), \tag{1.20}$$

$$k_{nx} = \frac{pn_x}{l_x},\ k_{ny} = \frac{pn_y}{l_y},\ k_{nz} = \frac{pn_z}{l_z}. \tag{1.21}$$

The damping coefficient of the nth mode τ_n is given by:

$$t_n = k \cdot \left[e_{nx}\left(\frac{b_{x0} + b_{xl}}{l_x} \right) + e_{ny}\left(\frac{b_{y0} + b_{yl}}{l_y} \right) + e_{nz}\left(\frac{b_{z0} + b_{zl}}{l_z} \right) \right], \quad (1.22)$$

Hence, $$G_n = 1/e_{nx}e_{ny}e_{nz} \quad (1.23a)$$

where the Neumann symbols e_{nx}, e_{ny} and e_{nz} are given by

$$e_{nx}, e_{ny}, e_{nz} = \begin{cases} 1 & \text{when } n \text{ is zero} \\ 2 & \text{otherwise} \end{cases} \quad (1.23b)$$

1.3.4 Choice of Green's function

The Helmholtz integral equation formulation and the finite element techniques are two approaches for calculating the radiation from a finite body in an enclosed space. The finite-element method is a comparatively more versatile technique. However, the Helmholtz surface integral techniques involve only surface elements of the same characteristic length scale, which leads to a significant reduction in the use of the computer memory. Both of these two methods require a prohibitively large number of surface elements, making realistic problems impossible to solve with today' s computers. In the 1990s, Lam and Hodgson [83-84] used the Helmholtz integral equation formulation to predict the sound fields from a vibrating sphere in a rectangular enclosure. They modified a Combined Helmholtz Integral Equation Formulation (CHIEF) [25] to deal with the radiation from an arbitrary body inside an enclosure. Their improved formulation gave a much greater accuracy in predicting noise levels in an enclosed space.

The key element of their formulation was to use the Green's function, which satisfies the boundary condition on the walls. There are two approaches for calculating the Green's function. First, it can be computed by summing an infinite series of eigenfunctions. Secondly, it can also be calculated by summing all image sources. Each method has its advantages and disadvantages. The eigenfunction of the Green's function converges most rapidly for a small separation between source and receiver. However, the image source method is most suitable for a relatively large source/receiver separation. It is important to choose the appropriate Green's function and check the rate of convergence as discussed. [83-84]

1.3.5 Hybrid models

A new hybrid method for the prediction of room impulse responses is recommended [45], based on two well-known computer algorithms: the ray-tracing and the image-source models. This hybrid method combines the advantages of the ray-tracing process with the accuracy inherent in the image-source model, which is sufficient to calculate the sound field in the frequency domain. This is important if we intend to perform convolution of the impulse responses with reverberation-free signals. The sampling rate of the signal must be at least twice the required time resolution.

1.4 Experimental techniques for classroom acoustics

Controlled experiments should be conducted in unoccupied classrooms in order to analyze the characteristics of the university's studying facilities. It is not uncommon to find that speech sounds received at the back of some classrooms may not be strong enough, but is comparatively clearer than in other classrooms where there is excessive reflection from the back walls. Hodgson initiated some studies [50-51, 60] to investigate speech intelligibility in classrooms. He statistically analyzed the acoustic properties of the absorbent classrooms in various shapes and sizes. His studies provide the preliminary elements for good acoustical design of classrooms in the future. However, the major concern of this thesis is to identify the optimal design for high intelligibility in most positions in a classroom.

1.4.1 Speech intelligibility measurement methodology for experiment

Several different types of acoustical parameters have been used as predictors of speech intelligibility in rooms of different sizes and acoustical conditions. These include the signal-to-noise ratio, speech and background levels, the speech transmission index derived from modulation transfer functions, and the useful-to-detrimental sound ratio obtained from the early/late sound ratio. The different parameters have similar prediction accuracy, but the useful-to-detrimental ratios based on a 0.08 s early time interval are the most reliable. The physical measures, although based on very different calculation procedures, are strongly related to each other. [29]

Bradley [33] conducted speech intelligibility tests and acoustical measurements in 10 occupied classrooms. He also measured the octave-band background noise levels, early decay times, reverberation times, and various early or late sound ratios. Various octave-band useful-to-detrimental ratios were calculated, along with the speech transmission index. He studied the interrelationships of these measures in order to evaluate the appropriate parameters for assessing speech intelligibility in classrooms. The best predictors for speech intelligibility scores were also identified. From these results, ideal design goals for the acoustical conditions of classrooms have been determined either in terms of the 50 ms useful-to-detrimental ratio or a combination of the reverberation times and background noise levels.

Steeneken and Houtgast [68] conducted pilot studies to evaluate the use of the STI for testing digital speech-transmission channels. They developed a physical method for measuring the quality of speech transmission channels. Essentially, the method represents an extension of the concept of the Articulation Index (AI), which has been developed mainly to account for distortions in the frequency domain. The underlying concept of this approach, which is based on the Modulation Transfer Function (MTF) of a transmission channel, has been adapted to account for nonlinear distortions such as the clipping effect at the peak amplitudes and for distortions in the time domain such as the effects of reverberation and echoes. The resulting parameter, the Speech Transmission Index (STI), has been correlated with the subjective intelligibility scores obtained on 167 different transmission channels with a wide variety of disturbances. The relative predictive power of the STI, expressed in Phonetically Balanced (PB) word scores, appeared to be 5%. This accuracy is comparable with results obtained from subjective measurements with the participation of four listeners. Expressed in terms of the signal-to-noise ratio, the accuracy was about 1dB.

Currently, listening conditions can be accurately estimated in terms of the Speech Intelligibility Index (SII) [4, 11]. The input variables for these prediction schemes include the equivalent speech spectrum level, the equivalent noise spectrum level, and the equivalent hearing threshold level. This includes conditions where either speech or noise may not exist as directly measurable physical quantities (e.g., conditions where speech-correlated noise is present, such as a reverberated speech) but where an equivalent speech spectrum level, equivalent noise spectrum level, and equivalent hearing threshold level can nevertheless be calculated. The predictions made using these schemes are correct only on the average, that is, across a group of talkers and a group of listeners of both genders. The scope of the prediction schemes is limited to natural speech, normal listeners, and communication conditions excluding multiple, sharply filtered bands of speech or sharply filtered noise of speech or sharply filtered noise. In addition, the listeners should have no linguistic or cognitive deficiencies with respect to the language used.

These two prediction schemes [4, 11] define the method for computing SII, which is highly correlated with the intelligibility of speech under a variety of adverse listening conditions, such as noise masking, filtering, and reverberation. The SII is computed from acoustical measurements or estimates of speech spectrum level, from noise spectrum level, and from psycho-acoustical measurements or estimates of hearing threshold level. Various frequencies contribute different amounts of speech intelligibility and, within a certain range, a higher speech-to-noise ratio contributes to intelligibility. The intelligibility of a speech communication system can be predicted by measuring the speech-to-noise ratio in each contributing frequency band and adding the results.

1.4.2 Measurement of the speech transmission index

Due to the recent development of computer software for measurements and simulations of acoustical impulse responses, several numerical methods have been developed to predict sound fields accurately. Commercially available software, such as RAYNOISE [62] using in some application shown as the Figure 1.4 and the Figure 1.5, and hardware, such as the Maximum Length Sequence System Analyzer, are available for predicting and measuring the STI in a room.

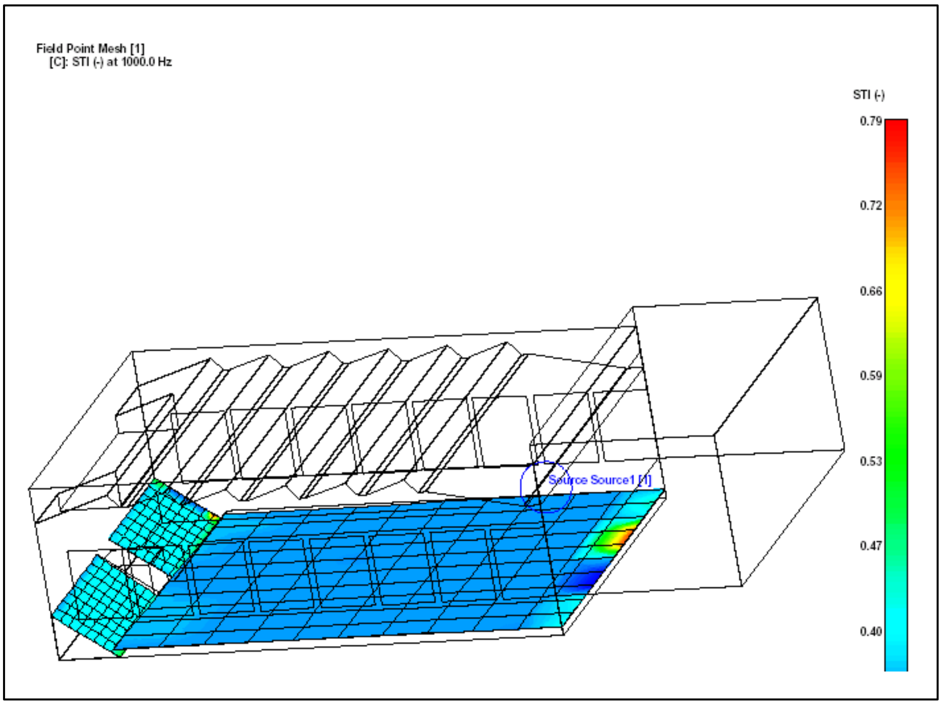

Fig. 1.4 An examples of auditorium application by RAYNOISE

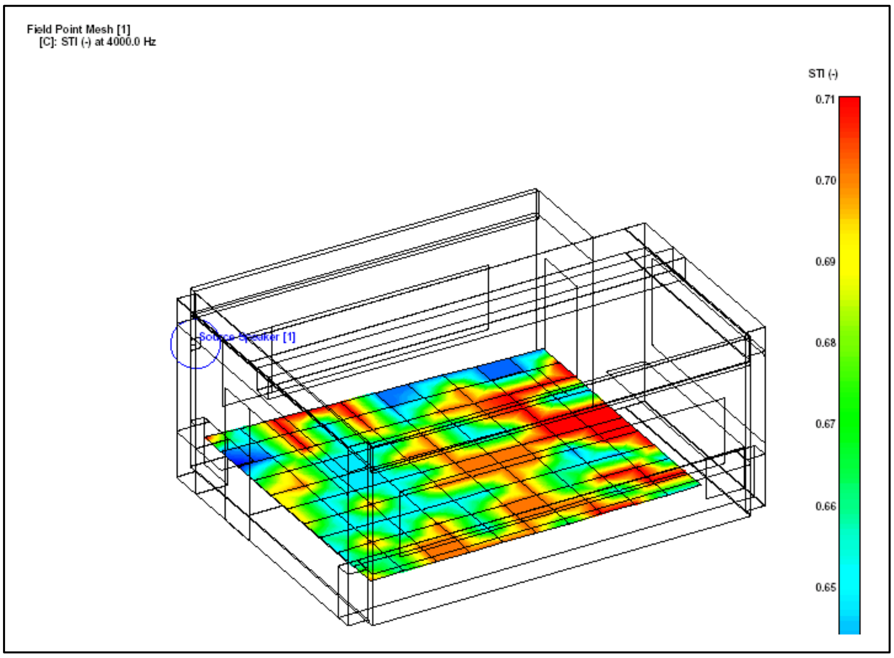

Fig. 1.5 An examples of classroom application by RAYNOISE.

There is no doubt that all the important parameters can be calculated from the impulse response between a pair of source-receivers in the room. However, not only is it hard to estimate the diffuse sound field inside a room, but the measurement is hardly repeatable. These effects can cause a nonlinear measurement and thus lower the dynamic range of the signal-to-noise ratio.

The technique of Maximum Length Sequence (MLS) has been developed and its use encouraged because a pseudo-random signal has been used which has an advantage in its auto-correlation function. Also, the MLS can show the cross-correlation of the driving signal. The system's response is the impulse response of the system. The non-correlated noise is eliminated because the driving signal is known, and only one of the signal receiver channels is required in the measurement device. This result leads to a higher signal-to-noise ratio, which is one of the main characteristics of the technique.

1.5 Summary

This chapter introduces the theoretical background and experimental work before we conduct the analysis of the tests for speech intelligibility in the following chapters. The literature survey provides the direction and the rationale of our current study. More of the literature will be discussed in detail in the following chapters. Technical terms and definitions will be clarified further in subsequent chapters.

Chapter 2

Speech intelligibility tests and overall intelligibility trends

The study of speech intelligibility (SI) is an interesting topic in an enclosed space. Generally speaking, speech intelligibility can be measured by means of an objective measure that is known as the Articulation Index (AI). Different countries use different languages for communication, and there are obvious discrepancies in the speech intelligibility languages that are due to different formations and combinations of vowels and consonants. For instance, some English consonants, such as affricates, fricatives, nasals, and plosives, may be easily confused in reverberant conditions, whereas the tones in Mandarin are more helpful for word intelligibility in such conditions [31].

Although Cantonese is the language that is spoken by the majority of people in Hong Kong, Mandarin and English are also widely used for communication in many circumstances. We are primarily interested in investigating the speech intelligibility of these three spoken languages under similar environmental conditions in the classroom.

The situation in Hong Kong is rather distinct in that students are frequently required to use Cantonese (their first language) and two second languages (English and Mandarin) to communicate in the classroom. Typically, teachers use Cantonese to explain their ideas but may also use English to supplement or elaborate the concept and content of the lessons. In this study, a group of university students were invited to participate in our experimental studies by undertaking speech intelligibility tests in the three languages. The experiments were conducted in a classroom at the Hong Kong Polytechnic University. We use the experimental results to compare the discrepancies in the speech intelligibility of Cantonese, English, and Mandarin for a typical undergraduate student in Hong Kong.

In addition to the use of the Articulation Index (AI), such physical parameters as the speech transmission index (STI) and early decay time (EDT) are also employed to assess speech intelligibility in the classroom [31, 33-34, 56]. We are interested in exploring the discrepancies in the speech intelligibility of the three languages and the way in which the STI, EDT, reverberation time (RT), and other pertinent parameters influence this intelligibility.

In this chapter, the main objective is to present the overall trend of the data in this study.

The data are very detailed, and include many assumptions and results from different experiments. It is therefore our aim in this chapter to offer a complete whole picture by detailing the overall results. In Chapter 3, the results are divided into different categories of listeners and speakers for each language, and the statistical variance is also discussed in detail. It is also necessary to further examine the effect of the different languages among the different categories of listeners and speakers, which is undertaken in Chapter 4.

2.1 Essential parameters for speech intelligibility

The concept of an architect designing a room with "good acoustics" is an interesting one. In general, there is no single encompassing set of criteria that ensures good acoustics for all rooms for all purposes. Small classrooms, large lecture theaters, auditoria, music rooms, cafeterias, and gymnasia all have different acoustical requirements, but whether these different spaces can be used for different purposes and how they should be designed to meet the right acoustic criteria is a matter of some debate.

2.1.1 Sound transmission method

It is important to understand the acoustical requirements of different types of rooms, and also to be familiar with a few basic properties of the propagation of sound in an enclosed space. Sound waves radiate in all directions from a source until they encounter obstacles such as walls or ceilings. In architectural acoustics, the two characteristics of sound waves that are of particular interest are *intensity* and *frequency.* Intensity is a physical measurement of a sound wave that relates to how loud the sound is perceived to be by the receiver. The frequency of a sound wave is perceived as pitch. A sound with just one frequency is called a pure tone, but most everyday sounds, such as speech, music, and noise, are complex sounds that are composed of a mix of different frequencies. The importance of frequency is revealed when a sound wave encounters a surface, as the absorption of the sound by the surface is different for different frequencies. The sensitivity of human ears to incoming sounds also varies with frequency.

Sound can be imagined as a beam, or ray of light, that passes through a medium and encounters an obstacle. When sound waves strike a surface, a number of things can happen, as follows.

- **Transmission –** The sound passes through the surface into the space beyond it, like light passing through a window.
- **Absorption –** The surface absorbs the sound like a sponge absorbs water.
- **Reflection –** The sound strikes a surface and changes its direction like a ball bouncing off a wall.

- **Diffusion –** The sound strikes a surface and is scattered in many different directions, like skittles being hit by a bowling ball.

However, several actions can occur simultaneously when a sound wave interacts with a surface. For instance, a sound wave can be both reflected and partially absorbed by a wall at the same time. This results in the reflected sound being softer than the initial sound. This process is dependent on the frequency of the impinging sound waves. Some surfaces absorb the energy of a high frequency sound but reflect the energy of a low frequency sound. The **absorption coefficient** is used to specify the ability of a material to absorb sound. The "sound absorption coefficient" (α) and "noise reduction coefficient" (NRC) are used to specify the ability of a material to absorb sound.

There are many parameters, such as the reverberation time (RT), speech transmission index (STI), and early decay time (EDT), that are used to assess the absorptiveness of the boundary walls in an enclosed space. These parameters may have different definitions according to different studies [12, 27, 31, 33, 57], but are discussed in general terms in the following sections.

2.1.2 Speech intelligibility (SI)

The ability to predict the acoustical conditions in a classroom accurately allows an architect to optimize the acoustical design construction or renovation of such rooms. The main acoustical feature in a classroom is verbal communication. The associated measurable quantity, speech intelligibility (*SI*), is the percentage of speech material that is presented to the average listener that is correctly heard. It can be evaluated in an existing classroom by using a word list. To generate an articulation index, several tests are performed, wherein one person recites words from a standard list and listeners write down what they hear. The percentage of words that the listeners correctly identify is a measure of speech intelligibility in the room.

2.2 Experimental setup

Experiments were carried out in a lecture theater with a maximum capacity of 120 and a classroom (Room BC512) with a maximum capacity of 55 at the Hong Kong Polytechnic University. For all of the measurements, the source was a Brüel & Kjær 4296 omni-directional speaker. A BSWA TECH MK224 half-inch condenser microphone was used as the receiver. The omni-directional speaker was connected via a Brüel & Kjær 2716 power amplifier to a maximum length sequence system analyzer (MLSSA) with an MLS card that was installed in a PC computer [45]. The analyzer was also connected to a half-inch condenser microphone via a BSWA TECH MA201 preamplifier. The MLSSA system was used both as the signal generator for the source and as the signal processing analyzer. To establish the noise levels that were emitted from the omni-directional speaker in the classrooms, the measurement system was initially calibrated in an anechoic chamber by placing the sound source 1 m away from the microphone.

Before the speech intelligibility tests, rooms FJ303 and BC512 were calibrated by measuring the *EDT*, *RT30*, *STI*, *RASTI*, *S/N* ratio, C_{50}, C_{80}, and U_{50} at various seating positions that were distributed uniformly throughout the classrooms. As the positions of the ears of all of the participants were assumed to be around 1.0 to 1.3 m above the floor level of the classrooms, a nominal height of 1.2 m was chosen for the microphone placement for all of the measurements.

Figure 2.1 shows the trace of a ray of sound in a typical lecture theater with a nominally rectangular shape. Sound rays may be reflected once or many times before they arrive at the receiver, and thus it is essential to obtain not only the signal-to-noise (S/N) ratio, but also the reverberation and early decay times for further analysis.

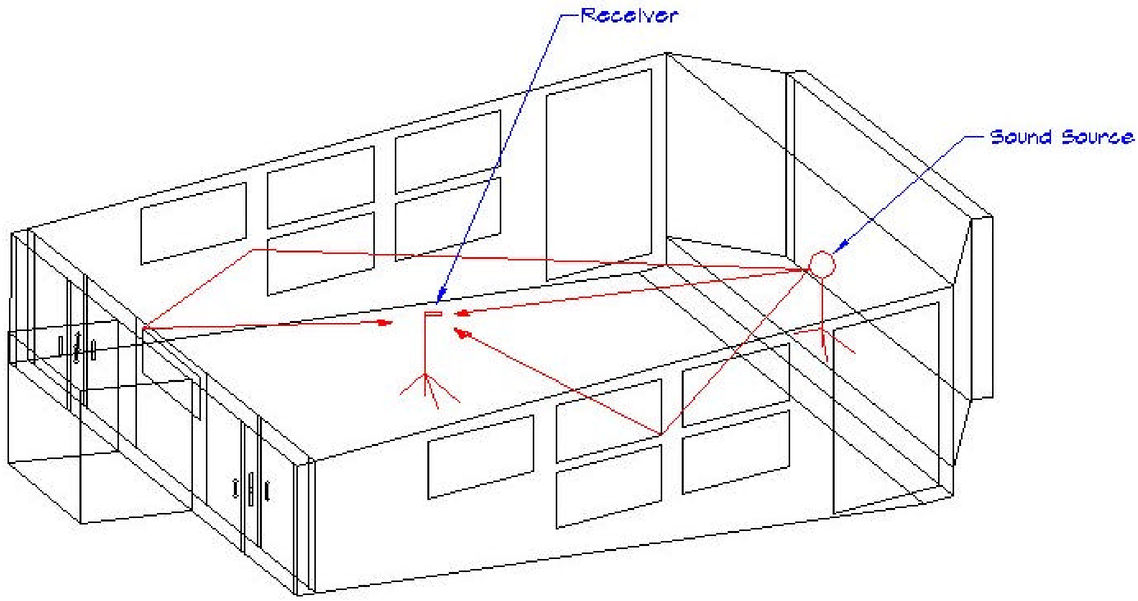

Fig. 2.1 The ray model in a typical lecture theater.

As the experimental study was focused on speech intelligibility and verbal communication in the classroom, it was necessary to use human subjects. A group of 15 individuals was invited to participate in each set of measurements. A total of over 300 participants were involved in the experiments. One of the principal objectives of the study was to explore the variation in the speech intelligibility of different languages (the mother tongue and two second languages) in the classroom. In our experiments, native trained speakers and listeners were invited to record phonetically balanced (PB) words as test materials according to the established International Standards [4-5, 11].

2.2.1 L_1 and L_2 signal

As classrooms in the Hong Kong Polytechnic University were chosen as the venue for our experiments, it was deemed appropriate to choose undergraduate students from local universities to participate in the tests. All of the equipment in the lecture theater, such as the projector, lights, and air-conditioner setting, was maintained in the normal operational conditions that would prevail during normal lecture periods. To eliminate the possibility of lip reading by the listeners, a Brüel & Kjær (B&K) 4296 omni-directional speaker was used to broadcast the test materials in the experiments. Two different signal settings were employed – soft and loud – which are marked either "L_1" or "L_2" in our presentation of the experimental results in Table 2.1. Each set of signals was tested and calibrated carefully inside an anechoic chamber, using a set of testing signals that comprised one vowel and consonant combination. Background noise was also introduced in the test to create either a silent or a noisy condition. This background noise was provided by loudspeakers inside the venue.

Soft signal in Silent condition								
Position	1	2	3	4	5	6	7	8
STI	0.681	0.681	0.646	0.637	0.618	0.646	0.601	0.651
S/N ratio at 1kHz (dB)	29.4	28.3	32.3	28.3	33.9	34.1	24.8	24.4
Noise Background (dB)	30.9	33.6	26.9	29.9	25.6	27.6	30.4	30.5
Position	9	10	11	12	13	14	15	
STI	0.606	0.619	0.601	0.589	0.58	0.58	0.583	
S/N ratio at 1kHz (dB)	26.5	25.8	22.7	21.7	25.9	30.2	25	
Noise Background (dB)	28.3	29.9	30.4	31.3	26.6	26	27.1	

Soft signal in Noisy condition								
Position	1	2	3	4	5	6	7	8
STI	0.374	0.448	0.343	0.363	0.334	0.315	0.287	0.25
S/N ratio at 1kHz (dB)	33.2	30.9	28.7	29.3	27.2	29.2	27.6	27.8
Noise Background (dB)	35.8	36.5	35	33	32.1	33.1	32.4	32.3
Position	9	10	11	12	13	14	15	
STI	0.298	0.299	0.296	0.194	0.284	0.261	0.253	
S/N ratio at 1kHz (dB)	27.8	28.9	28.4	27.6	28.7	28.5	26.2	
Noise Background (dB)	32.3	33.7	33.1	32.8	33.6	33.3	31.7	

Table 2.1 STI as measured in an anechoic chamber for the L_1 and L_2 settings.

MTF Matrix (Calibrated)

Frequency-Hz	125	250	500	1000	2000	4000	8000
level dB-SPL	47.8	55.3	56.0	54.7	58.6	57.4	55.8
m-correction	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.63	0.192	0.929	0.981	0.984	0.992	0.991	0.991
0.80	0.163	0.925	0.977	0.981	0.989	0.988	0.988
1.00	0.142	0.917	0.972	0.975	0.985	0.983	0.984
1.25	0.176	0.906	0.965	0.967	0.978	0.976	0.977
1.60	0.297	0.888	0.954	0.954	0.968	0.964	0.966
2.00	0.359	0.862	0.938	0.936	0.953	0.948	0.952
2.50	0.211	0.828	0.917	0.911	0.933	0.925	0.930
3.15	0.163	0.781	0.887	0.876	0.904	0.891	0.899
4.00	0.174	0.715	0.846	0.827	0.866	0.844	0.856
5.00	0.165	0.639	0.799	0.767	0.821	0.788	0.806
6.30	0.109	0.549	0.750	0.694	0.769	0.717	0.744
8.00	0.077	0.462	0.696	0.609	0.708	0.632	0.673
10.00	0.186	0.394	0.625	0.518	0.642	0.542	0.599
12.50	0.103	0.321	0.522	0.409	0.557	0.443	0.511
octave MTI	0.270	0.674	0.805	0.784	0.825	0.800	0.814

STI value= 0.719 (0.797 modified)   ALcons= 3.5%   Rating= GOOD

ESC to exit, F1 to print, Shift-F1 to dump.   MLSSA: STI

Table 2.2 The L_1 soft signal in the silent condition at position 7A in classroom FJ303.

MTF Matrix (Calibrated)

Frequency-Hz	125	250	500	1000	2000	4000	8000
level dB-SPL	71.9	70.0	40.6	36.5	35.8	34.9	33.0
m-correction	1.000	1.000	0.787	0.999	1.000	1.000	1.000
0.63	0.119	0.109	0.099	0.191	0.572	0.586	0.725
0.80	0.158	0.157	0.095	0.192	0.572	0.583	0.723
1.00	0.193	0.200	0.090	0.195	0.572	0.578	0.720
1.25	0.171	0.177	0.090	0.207	0.569	0.572	0.715
1.60	0.076	0.069	0.095	0.217	0.561	0.562	0.708
2.00	0.080	0.076	0.091	0.184	0.557	0.549	0.695
2.50	0.080	0.092	0.070	0.194	0.541	0.538	0.676
3.15	0.050	0.053	0.082	0.179	0.513	0.520	0.653
4.00	0.180	0.046	0.098	0.169	0.506	0.484	0.612
5.00	0.110	0.051	0.045	0.132	0.472	0.452	0.574
6.30	0.073	0.022	0.096	0.123	0.429	0.405	0.521
8.00	0.107	0.086	0.052	0.123	0.395	0.350	0.457
10.00	0.089	0.077	0.089	0.078	0.371	0.291	0.382
12.50	0.062	0.104	0.069	0.092	0.309	0.225	0.300
octave MTI	0.186	0.155	0.148	0.257	0.497	0.485	0.564

STI value= 0.349 (0.369 modified)   ALcons= 25.7%   Rating= POOR

ESC to exit, F1 to print, Shift-F1 to dump.   MLSSA: STI

Table 2.3 The L_2 loud signal in the noisy condition at position 7D in classroom FJ303.

For example, the data in Table 2.2 & 2.3 was obtained using the L_1 setting in the silent condition in the following manner. An impulse signal from the MLSSA was introduced, and white noise was generated at the -39dB level with the power amplifier set to -10dB at position 7. These settings were not decided upon by guess work. The 15 positions were evenly distributed inside the classroom, and the impulse signal at each position was tested and analyzed carefully before the MLSSA software was set and the power amplifier nailed down. To widen the spread of the STI range, repeated measurements were taken at all of the positions.

Although the impulse response that was received might be distorted or altered by the background noise, it was expected that the MLSSA would recognize or correlate the signal up to a certain acceptable level of noise. With the MLSSA, if the impulse response becomes totally distorted or cannot be recognized under severely noisy conditions, then an error message appears, the STI value will not be shown, and the MTI will seem unreasonable.

Through trial and error with repeated tests inside a single classroom, the objective articulation index was finalized ready for the subjective test of speech intelligibility among the teachers and students.

In the experimental study, 2,000 similar sets of signals were used and calibrated. It was assumed that the diffused field theory was applicable in each test in the experiment. Each signal was checked to verify that it covered the main STI spread from 0.05 to 0.980 to ensure that enough data would be collected for analysis. In the MLSSA, the STI rating is automatically classified as "poor," "fair," "good," or "excellent." However, it was also necessary to carefully check that the percentage of Articulation Loss for Consonants (AL_{cons}) was consistent with the two weighting factors of the STI and modulation transmission index (MTI) for each signal.

As has been mentioned, the background noise (dB) level, or environmental condition, was also controlled in each test. The background noise was provided by a pair of loudspeakers that were connected to the B&K2716 power amplifier, and was maintained at a much higher level than the overall SPL of 110(dB) that was generated inside the anechoic chamber. It was assumed that this would provide a suitable noisy condition inside the classroom when the test was taken. As the STI was controlled and calibrated for each position inside the classroom, adjustments could be made when the receiving signal was deteriorated by the background noise. For example, when the receiving S/N ratio was out the calibrated range of -30 dB and +30 dB, the weighting coefficient index triggered an “error” message inside the system, and the calculated result of the transmission index was rejected.

2.2.2 Phonetically balance (PB) word list

This experiment focuses on a comparison of the difference between first and second languages, and several standardized tests [1, 3-5, 10-11] have been developed by researchers for this purpose. Professor D. Y. Maa [19] carried out this type of research on Mandarin in mainland China using over a thousand randomly chosen statements transmitted over a public address (PA) system. In Hong Kong, no research has been carried out to compare the intelligibility of Cantonese, English, and Mandarin.

We now introduce the testing methodology and provide some general results. In the methodology, phonetically balanced (PB) word lists were generated for the different languages, as shown in Table 2.4 to 2.8, by random selection from about a thousand words based on the general standards for each language. As the meanings of words are similar in Cantonese and Mandarin, over two hundred pairs of words from the PB word list were chosen for the tests that involved these languages.

	List 1	List 2	List 3	List 4
1	are	awe	ache	bath
2	bad	bait	air	beast
3	bar	bean	bald	bee
4	bask	blush	barb	blonde
5	box	bought	bead	budge
6	cane	bounce	cape	bus
7	cleanse	bud	cast	bush
8	clove	charge	check	cloak
9	crash	cloud	class	course
10	creed	corpse	crave	court
11	death	dab	crime	dodge
12	deed	earl	deck	dupe
13	dike	else	dig	earn
14	dish	fate	dill	eel
15	end	five	drop	fin
16	feast	frog	fame	float
17	fern	gill	far	frown
18	folk	gloss	fig	hatch
19	ford	hire	flush	heed
20	fraud	hit	gnaw	hiss
21	fuss	hock	hurl	hot
22	grove	job	jam	how
23	heap	log	law	kite
24	hid	moose	leave	merge
25	hive	mute	lush	move
26	hunt	nab	muck	neat
27	is	need	neck	new
28	mange	niece	nest	oils
29	no	nut	oak	or
30	nook	our	path	peck
31	not	perk	please	pert
32	pan	pick	pulse	pinch
33	pants	pit	rate	pod
34	pest	quart	rouse	race
35	pile	rap	shout	rack
36	plush	rib	sit	rave
37	rag	scythe	size	raw
38	rat	shoe	sob	rut
39	ride	sludge	sped	sage
40	rise	snuff	stag	scab
41	rub	start	take	shed
42	slip	suck	thrash	shin
43	smile	tan	toil	sketch
44	strife	tang	trip	slap
45	such	them	turf	sour
46	then	trash	vow	starve
47	there	vamp	wedge	strap
48	toe	vast	wharf	test
49	use	ways	who	tick
50	wheat	wish	why	touch

Table 2.4 English PB word list.

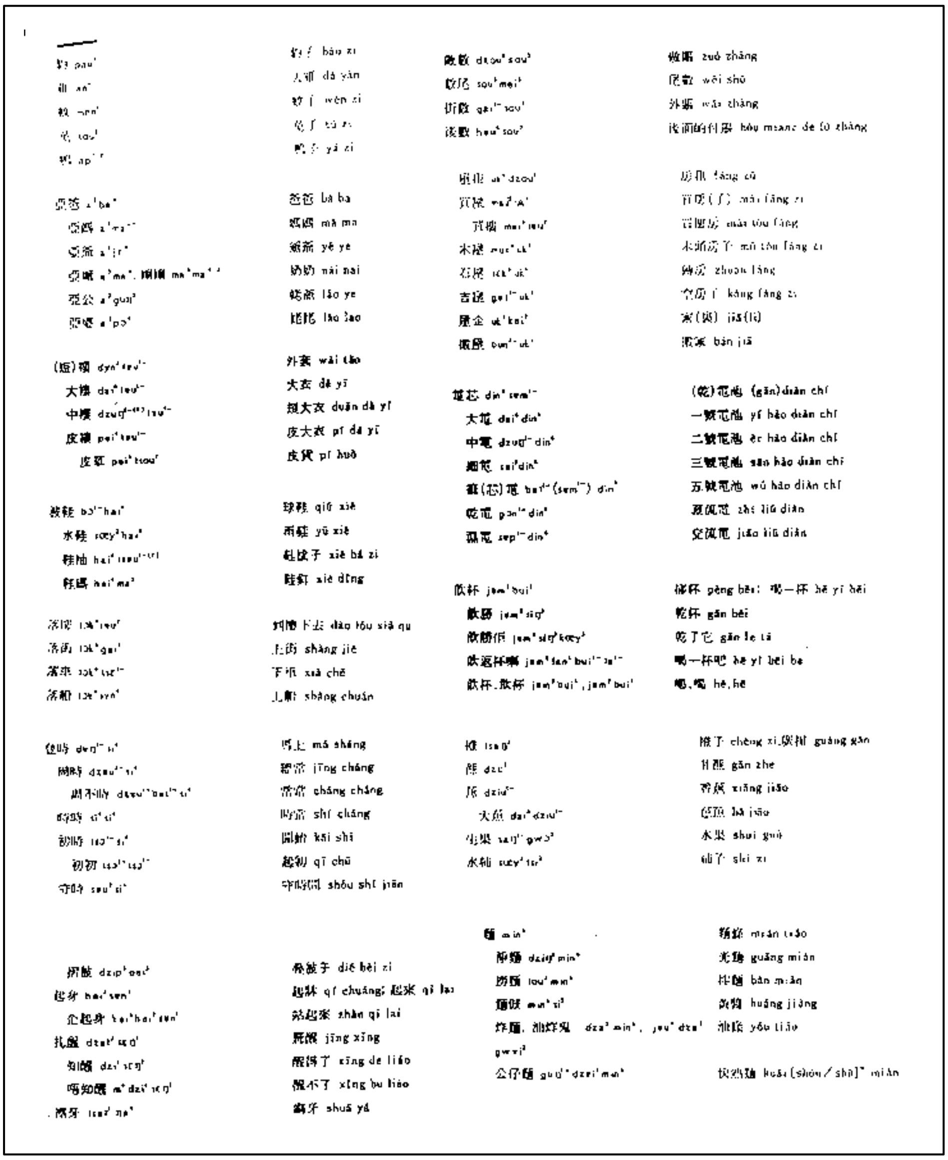

Table 2.5 Cantonese and Mandarin PB word list.

The test materials consisted of spoken words in Cantonese, Mandarin and English. For Cantonese and Mandarin, a list of 2,000 pairs of words from a selected word list. Tables 2.6 and 2.7 display the respective lists of Mandarin and Cantonese words that were used. For the English tests, 2,500 PB words were chosen, as suggested by the International Standards. Table 2.8 shows a typical list of the English words that were chosen. The selected words were recorded by trained native speakers of the respective languages on a tape recorder. A selection of 50 of the recorded words for each language was chosen on a random basis for use in the tests. In the tests, the recordings of the words were played back through the loudspeaker.

語音測試材料（普通話詞彙）			
豹子	bao4	zi0	
大雁	da4	yan4	
蚊子	wen2	zi0	
兔子	tu4	zi0	
鴨子	ya1	zi0	
外套	wai4	tao4	
大衣	da4	yi1	
短大衣	duan3	da4	yi1
皮大衣	pi2	da4	yi1
皮貨	pi2	huo4	

Table 2.6 Sample list of selected PB words in Mandarin

語音測試材料（廣州話詞彙）

豹	paau3	
雁	ngaan6	
蚊	man1	
兔	tou3	
鴨	aap3	
短褸	dyun2	lau1
大褸	daai6	lau1
中褸	zung1	lau1
皮褸	pei4	lau1
皮草	pei4	cou2

Table 2.7 Sample list of selected PB words in Cantonese

	List 1		List 2		List 3		List 4
1	are	1	awe	1	ache	1	bath
2	bad	2	bait	2	air	2	beast
3	bar	3	bean	3	bald	3	bee
4	bask	4	blush	4	barb	4	blonde
5	box	5	bought	5	bead	5	budge
6	cane	6	bounce	6	cape	6	bus
7	cleanse	7	bud	7	cast	7	bush
8	clove	8	charge	8	check	8	cloak
9	crash	9	cloud	9	class	9	course
10	creed	10	corpse	10	crave	10	court
11	death	11	dab	11	crime	11	dodge
12	deed	12	earl	12	deck	12	dupe
13	dike	13	else	13	dig	13	earn
14	dish	14	fate	14	dill	14	eel
15	end	15	five	15	drop	15	fin
16	feast	16	frog	16	fame	16	float
17	fern	17	gill	17	far	17	frown
18	folk	18	gloss	18	fig	18	hatch
19	ford	19	hire	19	flush	19	heed
20	fraud	20	hit	20	gnaw	20	hiss

Table 2.8 Sample list of selected PB words in English

To ensure comparable results, all of the participants were asked to confirm that they suffered no hearing or speech defects, although no hearing and speech tests were conducted to confirm this. All of the participants were assumed to be able to distinguish between the vowels and consonants in the PB words under normal circumstances.

2.2.3 Calibration of the experimental venue

The scope of the experimental study was to evaluate the speech intelligibility of the three languages among undergraduate students in selected in classrooms. In general, the intelligibility of speech in classrooms is usually measured by comparing the monosyllabic words that listeners receive with words that are delivered by the speaker or speech coder [3-5, 10-11]. The International Standards [3-5, 10-11, 12] also require that all measurements be conducted using speakers and listeners who have no speech or hearing defects. In addition, both speakers and listeners are also required to be native speakers of the respective languages. However, in this study, English and Mandarin were normally not the first language of most of the participants in the tests.

The experiments were conducted in a medium-sized lecture theater at the Hong Kong Polytechnic University (Room FJ303). As is shown in Figure 2.2, the room was divided into 15 zones in the front, middle, and back rows for analysis. The positions were so chosen to attain good coverage of the different areas with the measurements. The room is of the parallel-piped type, with sloped seating and a length that is greater than its width. The classroom has moderate absorption due to a suspended acoustical ceiling and upholstered seating for the students. The floor of the classroom is covered with carpet that has a fairly high rate of sound absorption.

In the preliminary measurements, the test signals were generated by the MLS system and subsequently amplified by the B&K 2716 power amplifier. The test signals were then fed to the omni-directional source, which is marked *S* in Figure 2.2. As has been mentioned, the test signals were divided into two different categories. Soft test signals were used to model soft consonants such as fricatives and nasals, and loud signals were used to represent strengthened vowels such as shouts and plosives. Both the soft and loud signals were transmitted in similar environmental conditions throughout the experiments. The effect of variations in temperature, humidity, air absorption, and scattering was negligible for the different sets of measurements. The SPL of the soft signals (105 dB at 1 m from the source) was the same as that of L_1 in the anechoic chamber, and the SPL of the loud signals was analogous to L_2, which had a level of 102 dB at 1 m from the source in the anechoic chamber.

Earlier experimental results have suggested that there is a direct correlation between speech intelligibility and the Speech Transmission Index (STI) in a room. To broaden the spread of the STI in room FJ303, the background noise level was augmented by placing two loudspeakers at the end of the room, as shown in Figure 2.2. The loudspeakers were fed with white noise that had a nominal SPL of 108 dB measured at 1 m from the source in the anechoic chamber. These two loudspeakers were turned on for the measurements taken under noisy conditions.

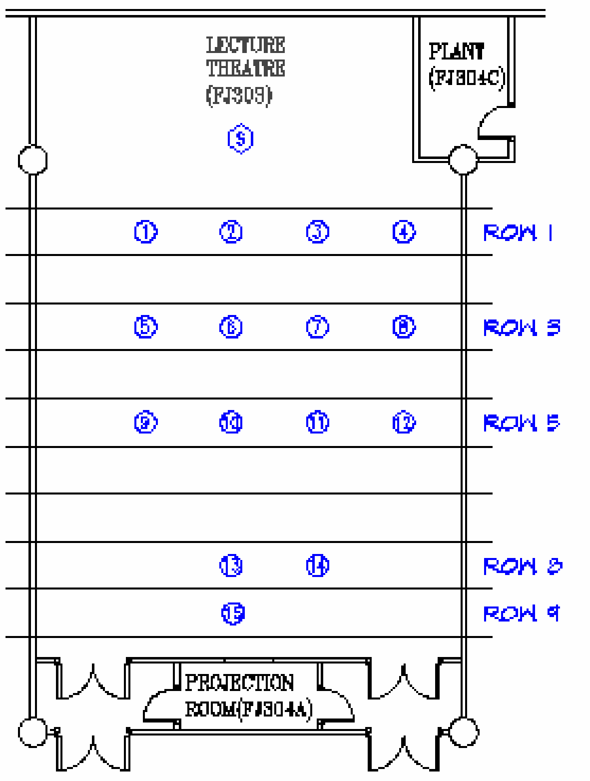

Fig. 2.2 Layout of the experimental venue (room FJ303 of the Hong Kong Polytechnic University).

Before the speech intelligibility tests were conducted, the classrooms were calibrated at the different positions with various acoustical conditions. The seating plan for the experiments is shown in Figure 2.2. Fifteen positions were chosen that were distributed evenly around the room, some of which were located directly under the exhaust of the air-conditioning system. To extend the range of the STI in our experiments, preliminary tests were conducted to determine the STI at different positions in the classroom. It was found that the STI was about 0.7 in the front row (closest to the speaker), and about 0.4 in the back row (nearest the door of the classroom). The two loudspeakers that generated artificial noise increased the STI range to give a highest STI of about 0.9 and a lowest STI of less than 0.1 at frequency bands of between 250 Hz and 1250 Hz. This ensured that the STI spread was wide enough and that the venue was suitable for the speech intelligibility tests. The variation in the STI at four positions in classroom FJ303 is shown in Figure 2.3.

It can be argued that human speech has some directive dependency. In this experimental study the omni-directional speaker type was used to generate the sound signals, rather than a human voice that directly emitted the sound rays to the listeners. The "dummy head" instrument is highly recommended for use with the MLSSA, as it not only models the speaker, but also correlates the sound ray that is emitted from the "speaker's" mouth in the model. However, as this piece of apparatus was not available, the omni-directional speaker was considered adequate to deal with the analysis. As all of the tests involved the comparison of a subjective intelligibility score and the objective measurement of the STI at each position, it can be assumed that the influence of directivity would have been negligible. Furthermore, the data from all of the positions included a measure of directivity error. Nonetheless, the test results would have been more accurate had the directivity issue been more explicitly considered through the use of the dummy head instrument.

2.3 Overall trend in the speech intelligibility tests

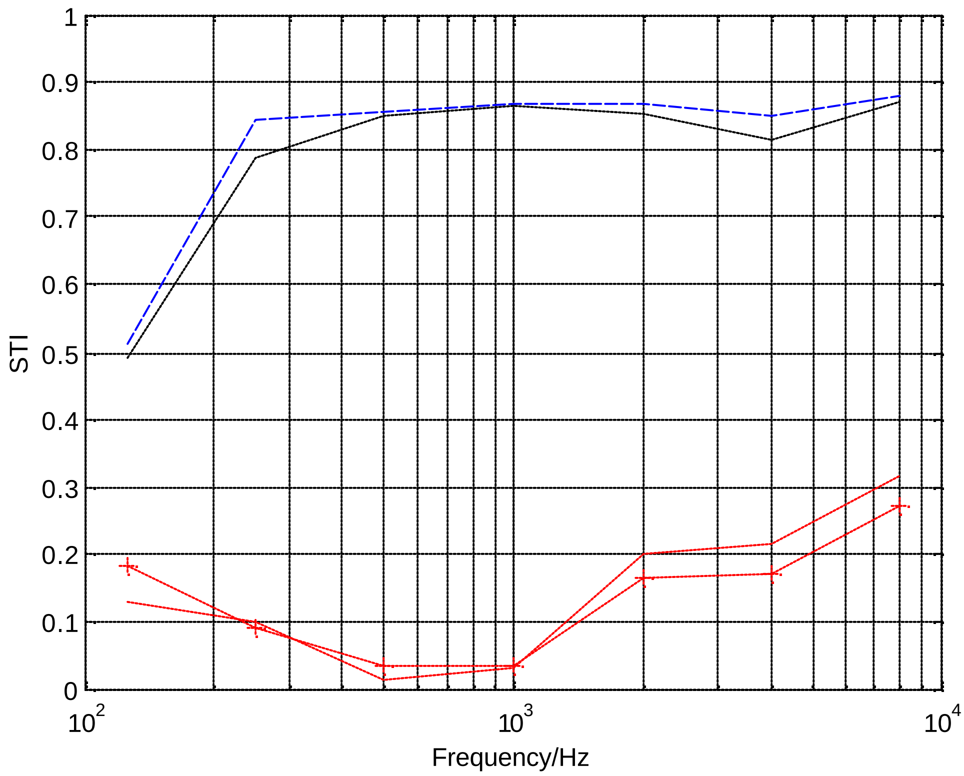

Fig. 2.3 STI versus frequency inside room FJ303.

The straight line represents the STI at position 1a, the dashed thick line the STI at position 2a, the dotted thin line the STI at position 13d, and the line with crosses the STI at position 14d.

As is shown in Figure 2.3, some of the positions were located directly in front of the sound source and some directly under the exhaust or supply outlets of the air-conditioning system in the classroom. However, the most important consideration of the study was to investigate the STI inside the classroom (see Figure 2.4), and although it would be useful to analyze the effect on speech intelligibility of background noise that is caused by the ventilation systems, we restricted our attention only to investigating the speech intelligibility of different languages among the students in the classroom.

Using the methods that are detailed in the foregoing, it was verified that the setup was sufficient in preparation for the subjective SI test, as the full spread of the STI covered the 0 to 0.88 range. The overall trends of the results are presented in the following.

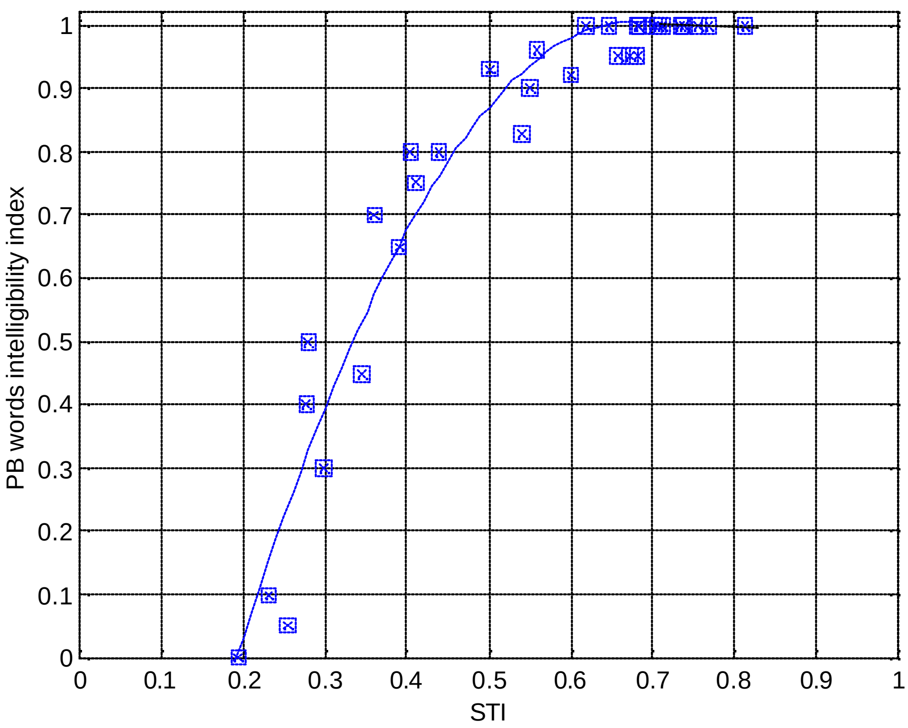

Fig. 2.4 Speech intelligibility of Cantonese among Hong Kong university students versus the STI.

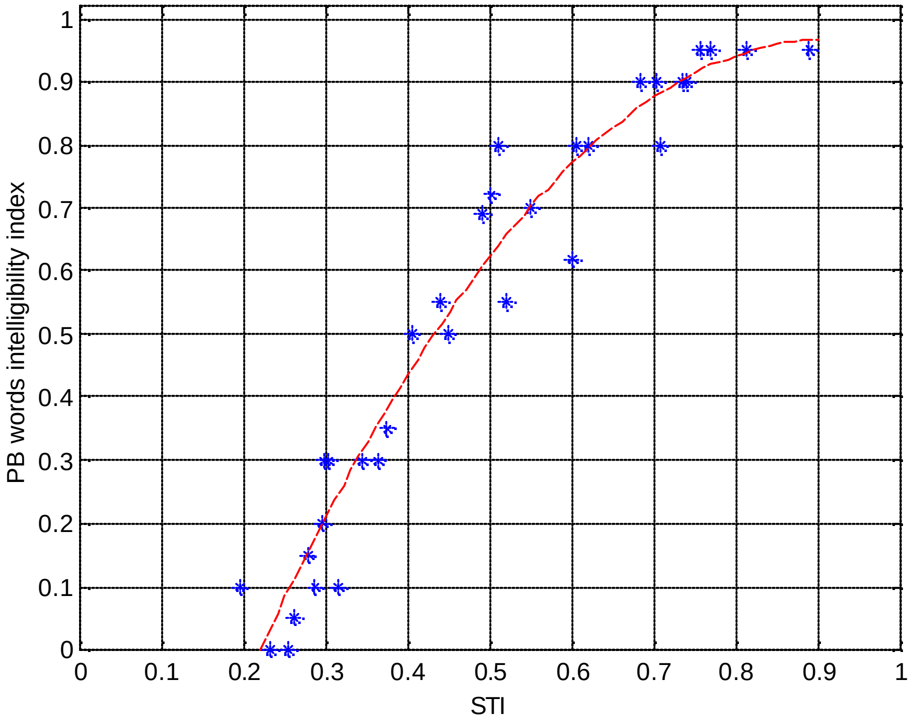

Fig. 2.5 Speech intelligibility of English among Hong Kong university students versus the STI.

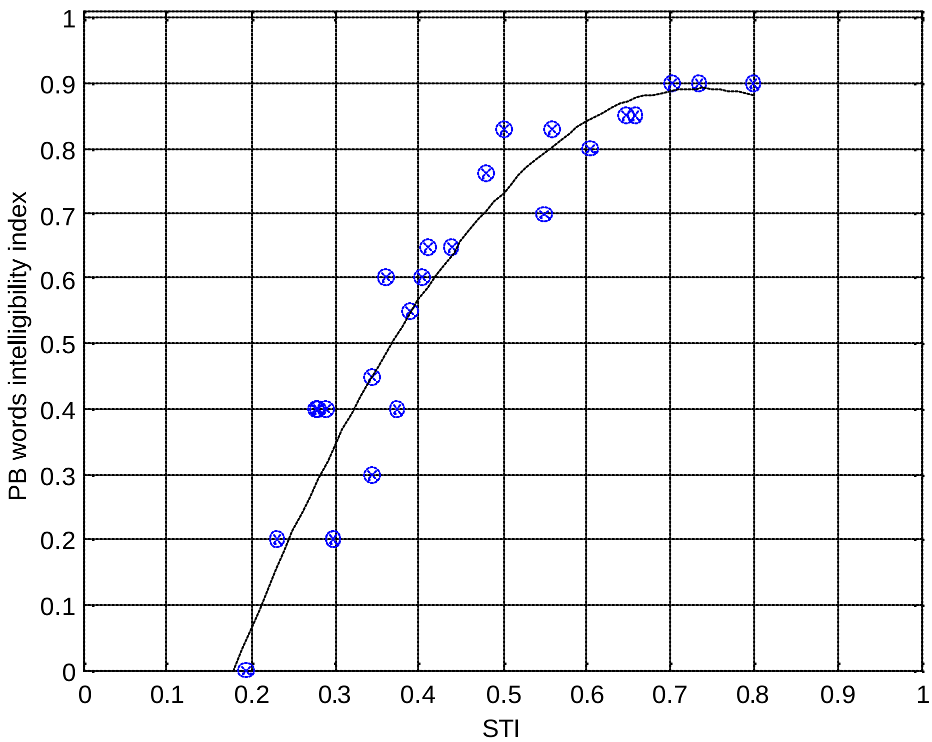

Fig. 2.6 Speech intelligibility of Mandarin among Hong Kong university students versus the STI.

Figures 2.4, 2.5, and 2.6 show the relationship between the speech intelligibility of the three different languages and the STI. They show that intelligibility among the students follows the same pattern as the distribution spread of the STI. However, for Cantonese, the slope of the curve is larger, which implies that it has a tendency to be dramatically saturated. The curve almost reaches a speech intelligibility level of 100% between the STI range of 0.6 to 0.7, and then stabilizes. This implies that when a teacher uses the mother language as the verbal communication channel for the lesson, more than 90% of students are capable of hearing the teacher clearly and intelligibly when sitting at any position in the studied classroom.

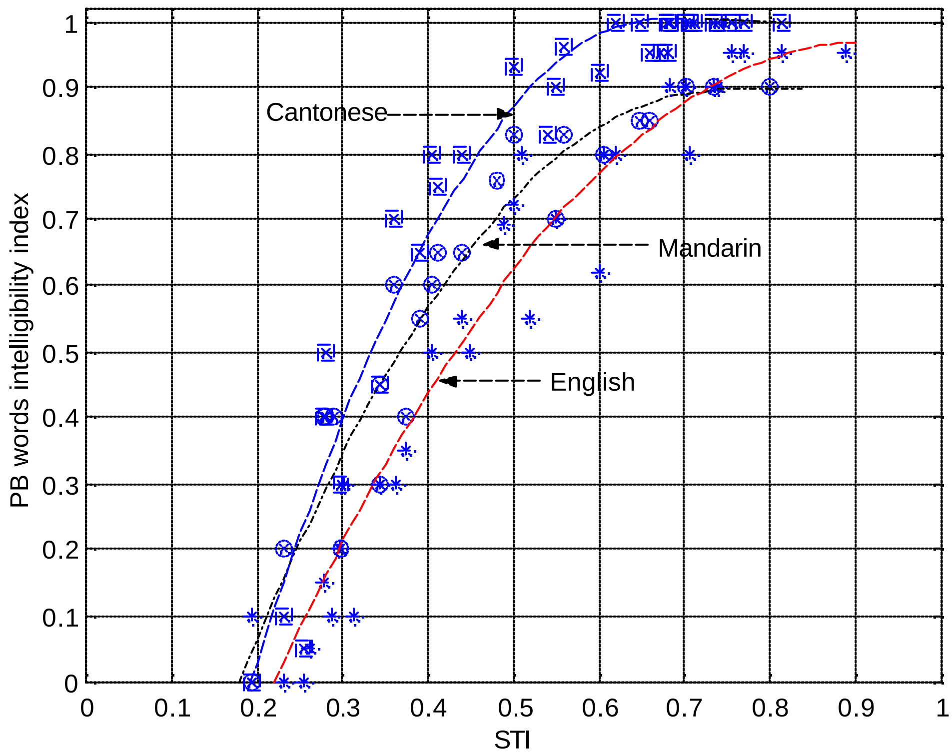

Fig. 2.7 Speech intelligibility of all three languages among Hong Kong university students.

In contrast, the three overlapping curves in Figure 2.7 show that the two second languages could not attain a 100% speech intelligibility. For Mandarin the maximum figure is 89%, and for English it is 97%. Thus, even if the classroom had undergone some acoustical treatment that forced the STI to 0.8 or more, the intelligibility of Mandarin would only reach the 0.89 level and that of English only the 0.97 level. This result implies that students' understanding of a second language is constrained at certain levels, even in a classroom with good acoustics. It may be that the knowledge level, educational background, and level of language acceptance among Hong Kong students was a factor in determining the intelligibility of the languages, which throws up some interesting topics for future development. For example, it would be interesting to see whether, under the same circumstances, this result could be replicated for other foreign languages, such as Japanese or Italian. The student participants found the English speech to be more intelligible than the Mandarin speech, which indicates that Hong Kong students may more receptive if English is used as the medium of verbal communication in the classroom, rather than Mandarin. However, these issues are beyond the scope of this study.

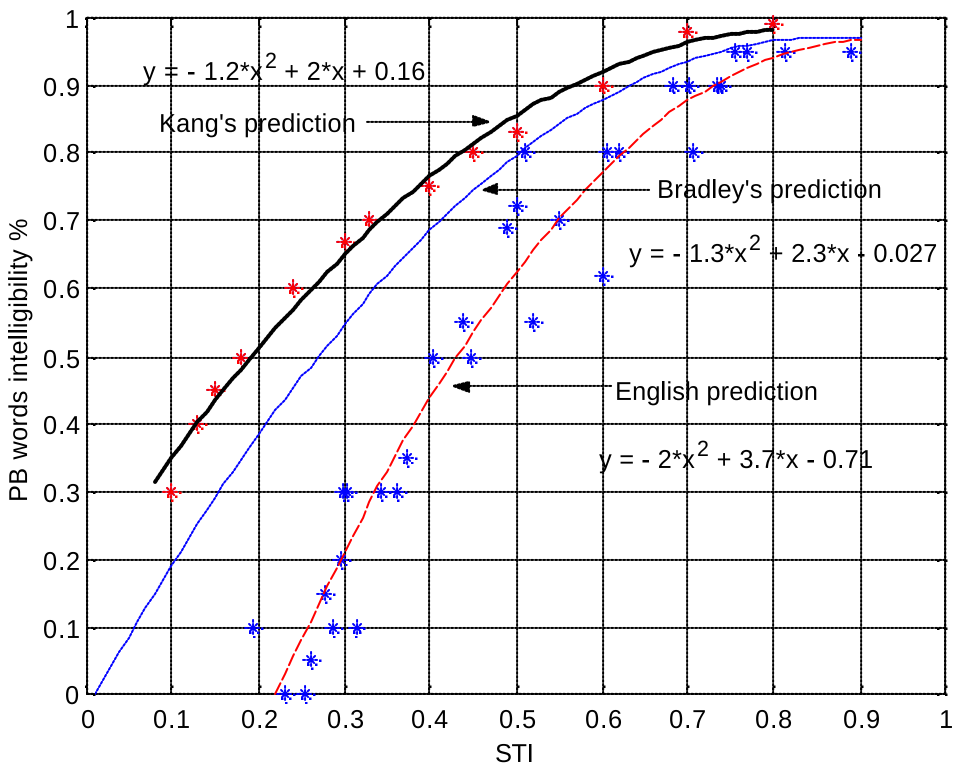

Fig. 2.8 Comparison of the predictions of Kang, Bradley, and this study of the speech intelligibility of English PB words.

Having compared the results for the intelligibility of the three languages among Hong Kong university students, it would be interesting to compare the intelligibility levels with those that have been found in other studies. Figure 2.8 shows the level of intelligibility of English among trained native speakers according to the predictions of Kang and Bradley. It can be seen that the levels of intelligibility in their results are greater than those among the Hong Kong university students in our study. This is understandable as, unlike the Hong Kong students, the mother language of the participants in the studies of Kang and Bradley was English.

A comparison of the results for the speech intelligibility of Cantonese among the Hong Kong students and the speech intelligibility of English according to the predictions of Kang and Bradley shows that the results are similar across the full range of the articulation index. However, it is necessary to point out that the results of the three studies were obtained under different conditions, using different measurements and levels of training, and among groups of speakers with different educational backgrounds.

These differences notwithstanding, it is interesting to study the discrepancy in the intelligibility of Mandarin among native Mandarin speakers and Hong Kong students. Once again, owing to the effect of the mother language, the performance of the native speakers in the studies of Kang and Bradley is better than that of the second language speakers over the full STI range.

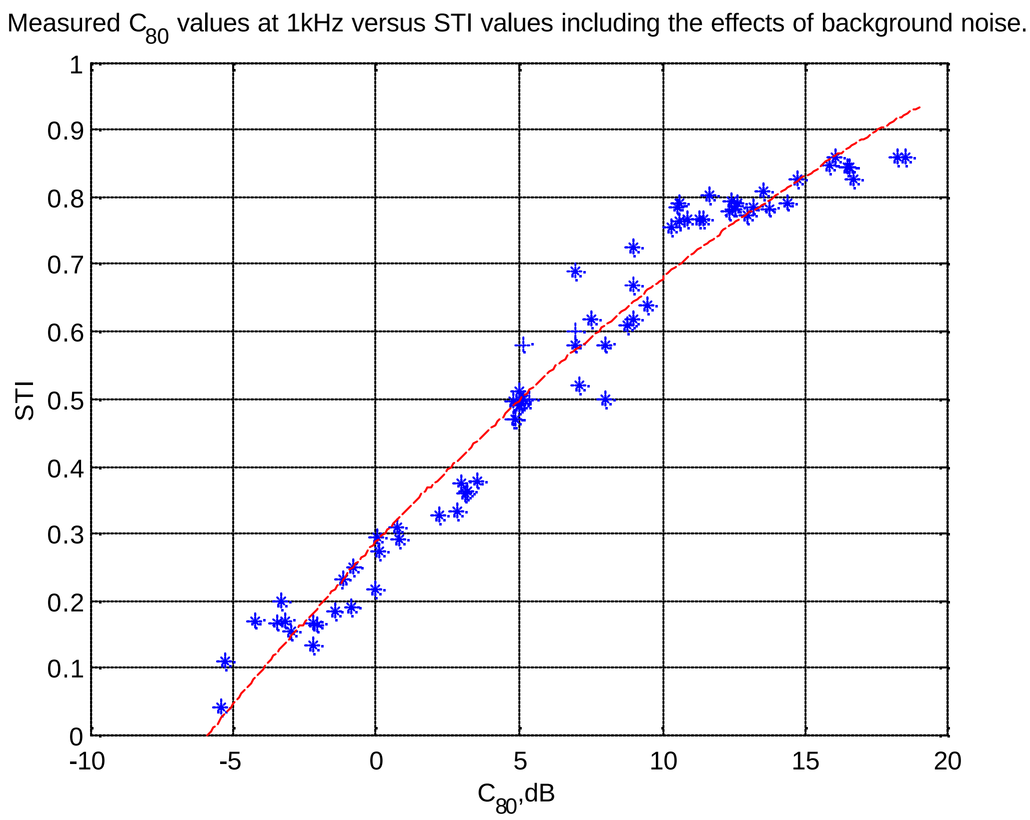

Fig. 2.9 Relationship between the STI and clarity C_{80}

Having considered the intelligibility of the three languages and compared it with the intelligibility of these languages among native speakers under the same STI conditions, the empirical equation of the relationship between STI and other acoustic parameters is further analyzed. Figure 2.9 shows the relationship between the STI and clarity C80 at a frequency of 1 kHz. This empirical relationship is expressed in Equations (2.1) to (2.3).

Empirical equation for the relationship between the speech intelligibility of Cantonese and the STI:

$$CAN = -4.26(STI)^2 + 5.78(STI) - 0.955 \qquad (2.1)$$

Empirical equation for the relationship between the speech intelligibility of Mandarin and the STI:

$$MAN = -2.88(STI)^2 + 4.24(STI) - 0.688 \qquad (2.2)$$

Empirical equation for the relationship between the speech intelligibility of English and the STI:

$$ENG = -2.01(STI)^2 + 3.68(STI) - 0.711 \qquad (2.3)$$

These three empirical equations are based on the results from the tests that were conducted among the Hong Kong university students.

The relationship between the three parameters is explored in more detail in Chapter 4. By fitting the second-order polynomials of the predictor variables, the best-fit curves for speech intelligibility can be obtained from the foregoing figures to give the relevant multiple correlation coefficients. This gives rise to the following empirical equations (2.4) to (2.6).

$$CAN = -1.43\times10^{-4}\left(C_{80}\right)^4 + 2.22\times10^{-4}\left(C_{80}\right)^3 - 1.06\times10^{-2}\left(C_{80}\right)^2 + 0.15\left(C_{80}\right) + 0.37 \quad (2.4)$$

$$MAN = -9.68\times10^{-5}\left(C_{80}\right)^4 + 1.5\times10^{-4}\left(C_{80}\right)^3 - 7.34\times10^{-3}\left(C_{80}\right)^2 + 0.12\left(C_{80}\right) + 0.32 \quad (2.5)$$

$$ENG = -6.75\times10^{-5}\left(C_{80}\right)^4 + 1.5\times10^{-4}\left(C_{80}\right)^3 - 5.53\times10^{-3}\left(C_{80}\right)^2 + 0.12\left(C_{80}\right) + 0.19 \quad (2.6)$$

2.4 Summary

This chapter describes the testing of the speech intelligibility of English, Cantonese, and Mandarin among Hong Kong university students. The speech intelligibility test materials comprised a list of phonetically balanced (PB) words and a recording of a selection of these words for broadcast to the listeners in the tests. The achievement of reasonable results in both tests required considerable effort, because the test materials had to cover a broad combination of the different vowels and consonants in each of the languages. In the Cantonese language alone, several hundred thousand PB words were needed for comprehensive testing.

Results were obtained on the speech intelligibility of the different languages among university students in a calibrated university classroom in both noisy and silent conditions. Some discrepancy was found between the intelligibility of the first and second languages among the students. However, a larger sample of listeners is needed to generate fully generalizable results. This chapter also introduces a methodology for the prediction of the various acoustic parameters that affect intelligibility. By comparing the test results for the three languages with the predictions of other studies, the empirical equations for the relationship between the intelligibility of the languages and the STI are obtained. This is the first study on speech intelligibility to compare Cantonese with the other second languages in Hong Kong.

Chapter 3

Testing and statistical analysis of speech intelligibility among different categories of listeners and speakers

An important aspect of disturbance that is caused by noise is its effect on speech intelligibility. Houtgast *et al.* [71-73] conducted speech intelligibility tests and revealed that the ambient noise at different locations in a classroom had certain effects on speech intelligibility. Information on the relationship between speech intelligibility and the ambient noise level is pertinent to the formulation of appropriate noise criteria for the planning and design of classrooms in school buildings.

In classrooms, two environmental factors are important for speech intelligibility: the level of reverberation inside the room and the interfering ambient noise. These two factors are interrelated. Ignoring the reverberant nature of a classroom, for example, will lead to an unrealistic estimate of the effect of ambient noise.

Theoretical and experimental studies have been carried out on speech intelligibility in classrooms that contained twenty teachers and about 500 pupils [34, 56], and a critical noise level for speech interference in classrooms has been suggested. Intelligibility tests among teachers and pupils have shown that satisfactory intelligibility in the classroom is noticeably compromised when the ambient noise level exceeds a critical value, and that ambient noise should be less than 15 dB for the long-term (reverberant) speech level of the teacher to be intelligible.

This chapter describes how further statistical and variance analysis was carried out in different classrooms to compare the speech intelligibility indices of Cantonese, which is the mother tongue of most people in Hong Kong, and English and Mandarin (which are the two most commonly used second languages in Hong Kong). The performance of trained speakers of Mandarin and Cantonese was compared with that of "average" listeners, who in this case were typically students at the Hong Kong Polytechnic University [31, 74]. Native speakers who were fluent in both Cantonese and Mandarin were also recruited for the speech intelligibility tests, and their test scores were compared with those of typical college students who were fluent in Cantonese but had learnt Mandarin as a second language. We use the results of tests that were carried out among native speakers of English [31] and compare them with the intelligibility scores of typical Hong Kong college students who had learned English as a second language.

In this chapter, the test results are employed to illustrate the speech intelligibility of the three languages. Each group of “trained” or “native” speakers is categorized, and the results for each group are plotted on graphs to show the objective intelligibility of each language. Finally, from the perspective of statistical and variance analysis, the discrepancy between the first language and second language for each category of listener and speakers is illustrated.

3.1 Experimental Setup

The objective of this chapter is to compare and analyze the speech intelligibility of first and second languages. Before the speech intelligibility test, the speech transmission index (STI) at each of the listener positions was determined. The test material included a list of phonetically balanced (PB) words.

3.1.1 Initialization and Calibration of the STI in the Classroom

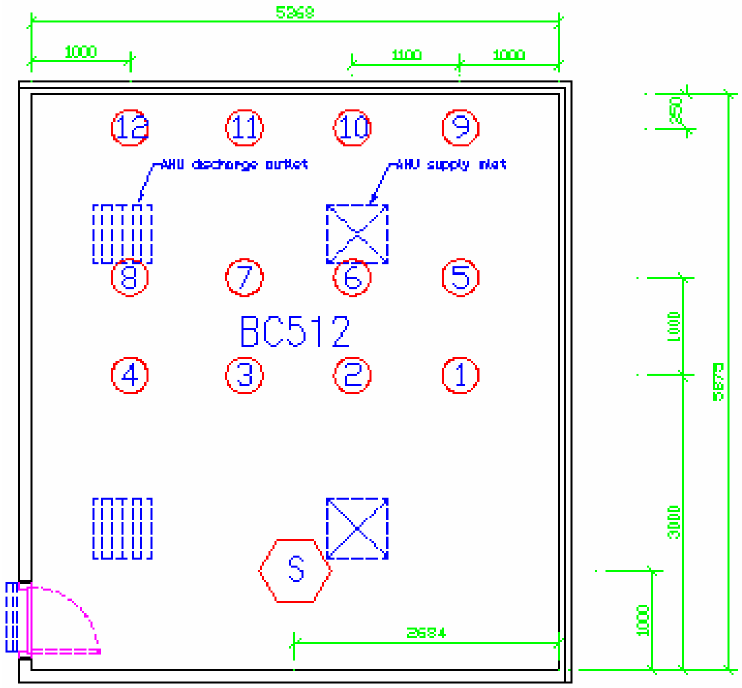

Fig. 3.1 Plan layout of the acoustic measurement in classroom BC512.

Room BC512 of the Hong Kong Polytechnic University was chosen as the main venue for the initial measurements for the speech intelligibility tests among the different speaker and listener categories, as shown in Figure 3.1. Room BC512 is a relatively small classroom in the university, and has a capacity of 12 people. The room is rectangular in shape and measures about 5.8 m × 5.3 m. The layout of the classroom is shown in Figure 3.2.

The room was divided into twelve different listener positions that were evenly distributed among the front, middle, and back rows. The main purpose of taking the initial measurements was to provide benchmark results for future reference.

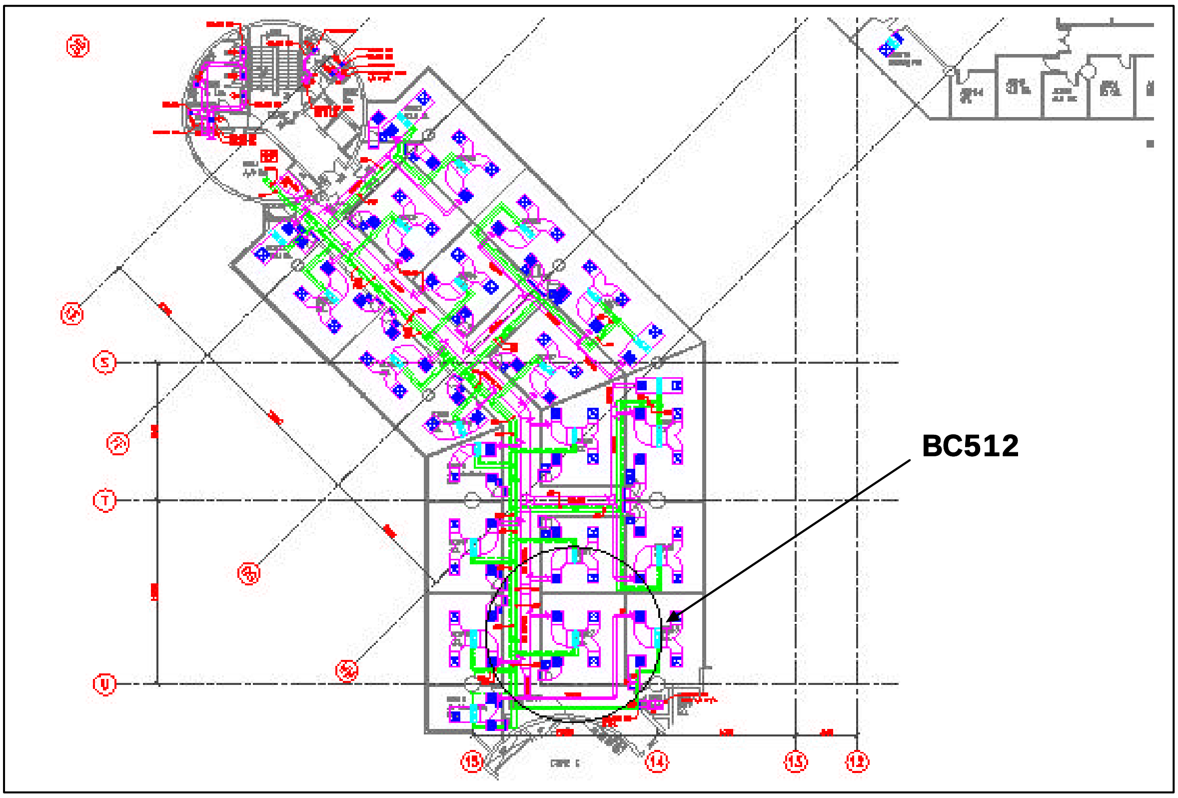

Fig. 3.2 Plan of the listener positions in classroom BC512.

We note that many of the classrooms at the Hong Kong Polytechnic University are not rectangular in shape (see Figure 3.2). It is more difficult to obtain a set of "standard" data for these rooms because of the complexity of their geometries. Due to its small size and rectangular shape, BC512 was the deemed to be the best location for the preliminary design of the speech intelligibility tests and the calibration of the experimental equipment.

The ceiling of room BC512 is lined with a sound absorbent material, the floor is carpeted, and all of the surrounding walls are concrete surfaces with no windows.

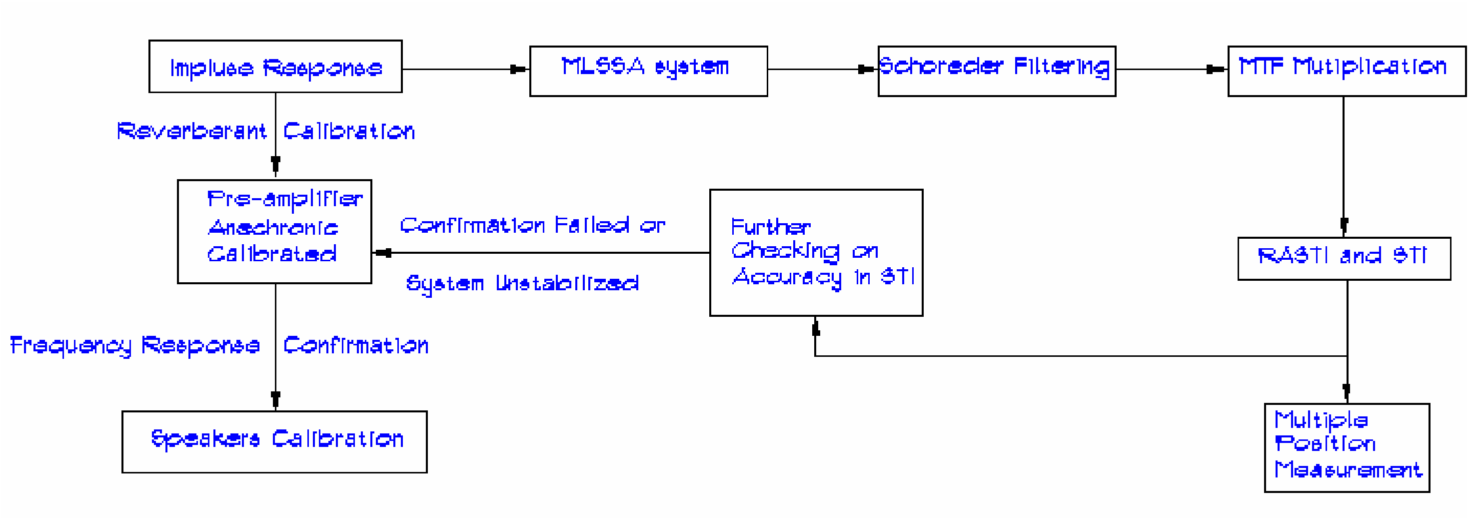

Fig. 3.3 Schematic diagram of the STI and RASTI experimental setup and MLS system calibration.

In this experiment, a BSWA TECH MK2 24 half-inch condenser microphone and a BSWA TECH MA201 pre-polarized pre-amplifier were used as the receivers. This measuring system was used to measure the reverberation time (RT), early decay time (EDT), and STI at the various listener positions in the classrooms. A Brüel & Kjær 4296 omni-directional speaker was used as the sound source, and was connected to a Brüel & Kjær 2716 power amplifier and subsequently to a maximum length sequence system analyzer (MLSSA) with an MLS card that was installed in a PC. The MLSSA system was used both as the signal generator and the signal-processing analyzer. The measuring system and the omni-directional speakers were initially calibrated inside an anechoic chamber before the field measurements were taken (see the schematic diagram in Figure 3.3. for the flow chart of the calibration procedures).

3.1.2 Calibration of room BC512

For the preliminary measurements in room BC512, the MLSSA was set to generate an impulse response with an overall sound level that was maintained at about 100 dB. The EDT and RT30 were measured at various receiver positions with the MLS system by recording the sound spectrum under Schroeder filtering of the linear data [67]. As has been described, room BC512 was divided into 12 listener positions for the initial measurements. The height of the positions was 1.2 m, which is approximately the ear level of an average person sitting on a chair. To measure the acoustical parameters such as the RT, EDT, and STI, the receiver positions were set according to international standards [1, 10] as far as was practicable. However, due to the size of Room BC512, the receivers in the last row were placed at a distance of only about 300 mm from the wall.

3.1.3 The effect of furniture

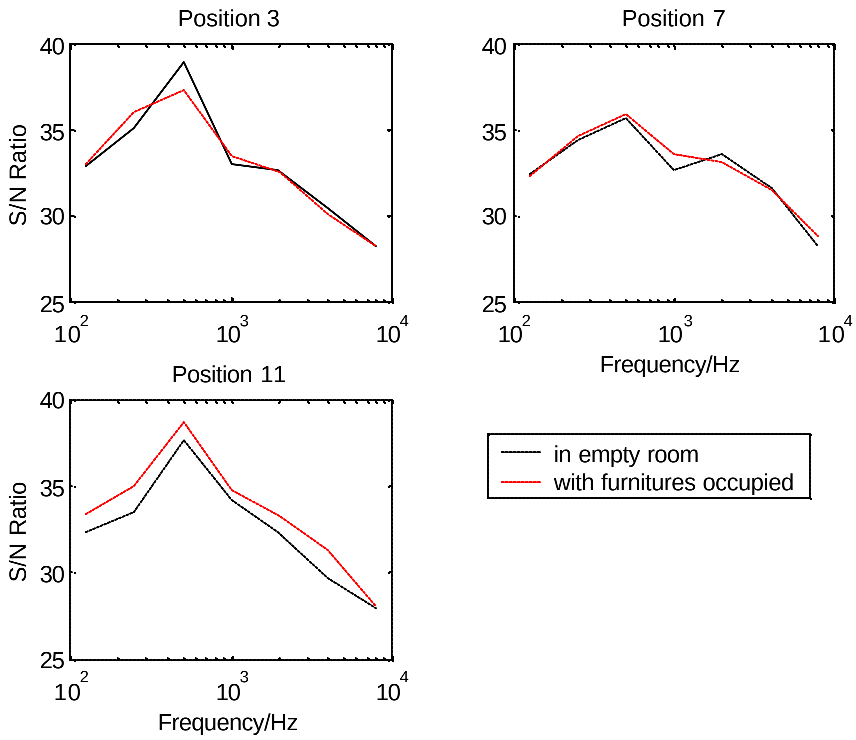

Fig. 3.4 S/N ratio analysis of the front, middle, and back positions in the presence of furniture. The solid line represents the data from an empty room and the dotted line represents t he measurements in the presence of furniture.

Several researchers have undertaken studies on the influence of furniture in university classrooms [48, 50]. Although it is likely that different universities use different furniture in their classrooms, it is of interest to study the effect of furniture on the acoustical characteristics of room BC512. Figure 3.4 illustrates the discrepancies in the signal-to-noise ratio (S/N ratio) at three receiver locations in the room. The solid line denotes the measured data for the S/N ratio in the absence of furniture and the dotted line represents the measured S/N ratio in the presence of furniture. We note that the effect of the furniture is rather small, with the difference in the S/N ratio not exceeding 3 dB in any of the cases. The receivers in the middle row suffered less measurement "error" because they were far from the reflecting walls. A difference of more than 2 dB was measured in the S/N ratio in the last row, but this is understandable, because the reflections from the surface of the back wall would have caused a higher reverberant field in this row.

The EDT is an essential parameter that is directly related to the S/N ratio and the RT. Generally, a room is more sound absorbent if the measured EDT is shorter. Figure 3.5 shows plots of the EDT spectra at the different positions in room BC512. It can be clearly seen that there are noticeable differences in the EDT at all frequencies in all of the positions, except for position 11 in the last row, where the effect of furniture on EDT was minimal for frequencies below 800 Hz. The measured data suggest that the furniture inside the classroom absorbed sound energy. It might be possible to improve the EDT from 1 s to 4 s by using absorbent materials in the furniture, especially for the receivers that were located in the front and middle rows.

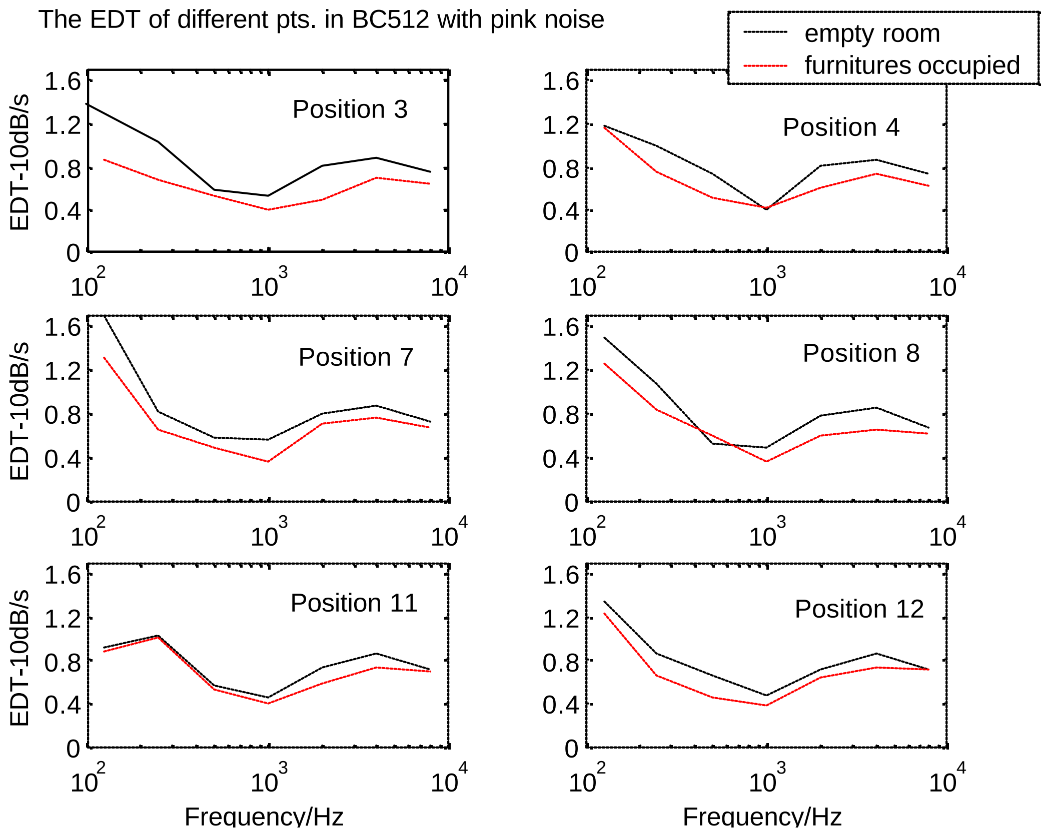

Fig. 3.5 Measured EDT at different positions with and without the presence of furniture. The solid line represents data in an empty room and the dotted line represents the measurements in the presence of furniture.

3.1.4 Comparison of the parameters in different rooms

It was deemed highly desirable to investigate the essential acoustical parameters in different rooms before conducting the speech intelligibility tests for the different languages. We conducted three additional sets of measurements in three classrooms, the details of which are shown in Table 3.1. The same measurements were repeated in the classrooms at various receiver locations to ensure the repeatability and reproducibility of the experimental results. Figure 3.6 shows a typical data set that includes the plots of the

STI at different frequencies and different positions in Room CF302. This figure is not intended to show a comparison of the variation in STI at the different positions in CF302, but to show the trend of the STI spectra for the different receiver locations.

Classroom	Volume m^3	No. of seats	No. of tables	Acoustic absorbent treatment
BC512	137	18	10	Absorbent ceiling tiles, concrete surface walls, fire-rated timber door, and carpeted floor.
DE307	243	55	28	Absorbent ceiling tiles, concrete surface walls, fire-rated timber door, and carpeted floor.
CF302	200	55	28	Absorbent ceiling tiles, timber partition wall surface, fire-rated timber door, and carpeted floor.
FJ303	360	144	48	Absorbent ceiling tiles, acoustic absorbent fabric walls, upholstered seating, and carpeted floor. Parallel pipe shaped.

Table 3.1 The four classrooms in the university that were selected for measurement.

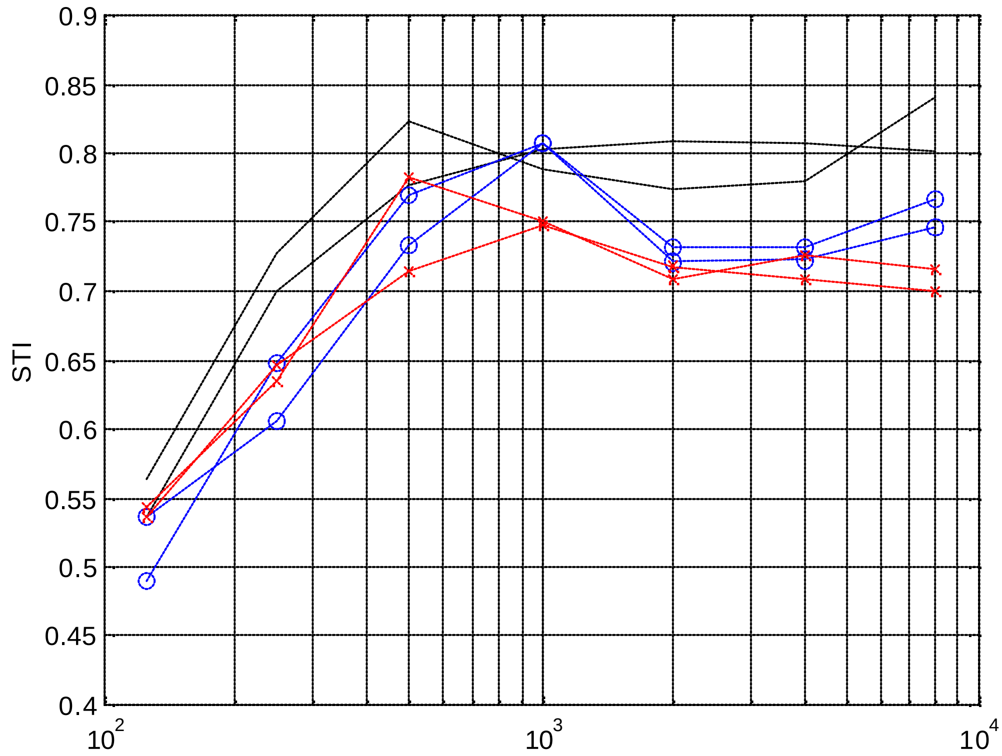

Fig. 3.6 The spread of the STI versus the frequency at different positions. The straight line denotes the measurements at position 1, the dashed line those at position 2, the straight line with circles those at position 5, the dashed line with circles those at position 6, the straight line with crosses those at position 9, and the dashed line with crosses those at position 10.

It is necessary to categorize the different classrooms in detail to allow further investigation. As is shown in Table 3.1, four classrooms were chosen for system calibration and testing, each of which had a different shape, size, sound-absorbent area, and volume to give appropriate variability in the investigation of the reliability of the method. In the following paragraphs, we compare the acoustical parameters of the four classrooms.

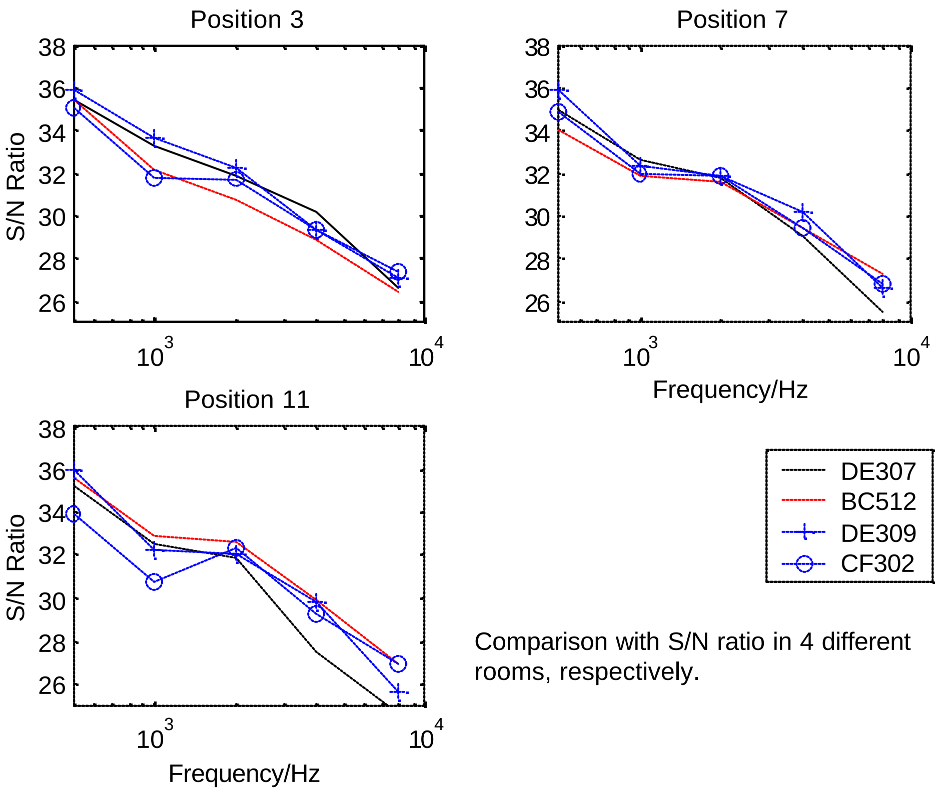

Fig. 3.7 S/N ratio measurement at different positions in the four classrooms.

The S/N ratio trend versus the frequency is plotted in Figure 3.7 for the three main streams of positions – the front, middle, and back rows – of the different classrooms. Each plot shows the S/N ratios of all four classrooms. All of the S/N ratios show a good agreement, especially for the front and middle rows. It is evident that the results for the back row positions were affected by reflections off the back wall, which led to greater reverberation.

Figures 3.8 and 3.9 show the frequency spectra for the RT30 and EDT at various positions in the four classrooms. Again, the RT30 and EDT have similar trends for the front, middle, and back rows for the different classrooms, and it thus appears that the classrooms all have similar acoustical characteristics for lecturing purposes. We chose the two most typical classrooms – FJ303 and BC512 – for our speech intelligibility tests.

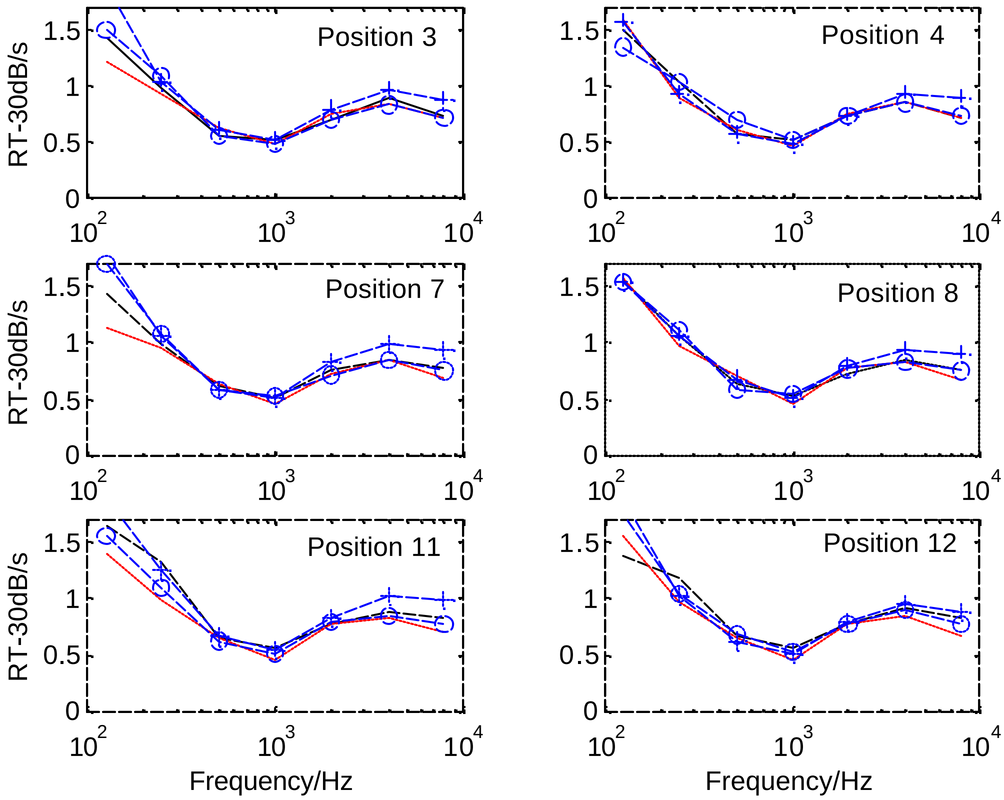

Fig. 3.8 RT-30 measurements at different positions in the four classrooms. The straight line denotes the measurements from room DE307, the dashed line those from room BC512, the straight line with crosses those from room DE309, and the straight line with circles those from room CF302.

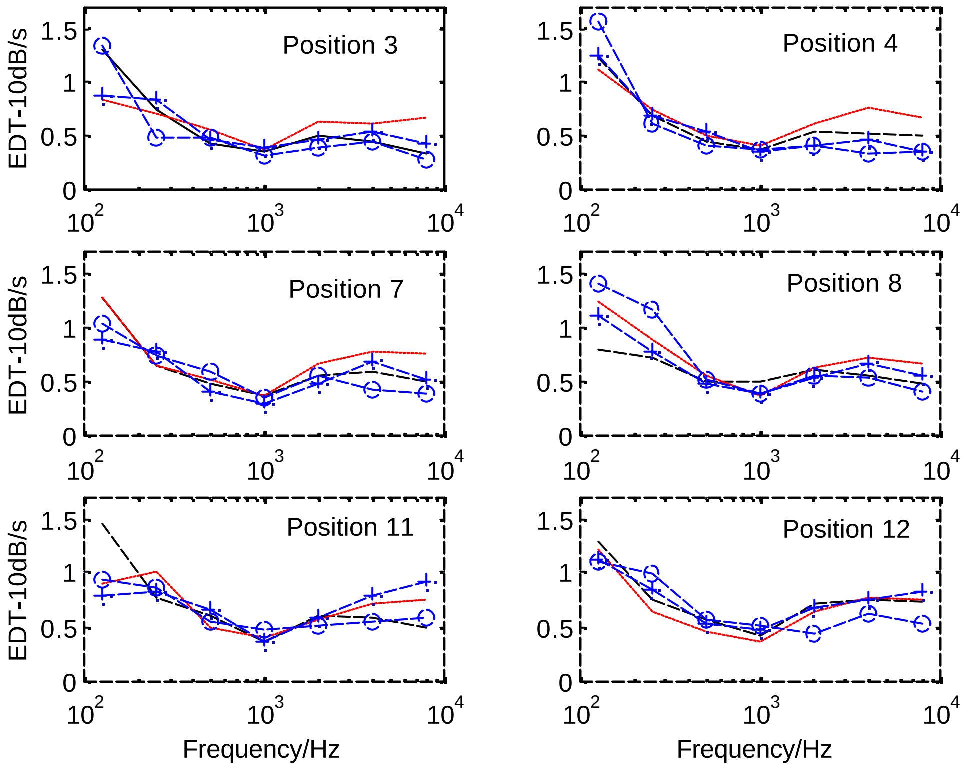

Fig. 3.9 EDT measurements at different positions in the four classrooms. The straight line denotes the measurements from room DE307, the dashed line those from room BC512, the straight line with crosses those from room DE309, and the straight line with circles those from room CF302.

Because of the symmetrical effect of the parallel-piped shape of room FJ303, we were able to choose any side for our measurements in this room. The locations that were chosen for analysis were Positions 1, 2, 5, 7, 9, 11, 14, and 15. Figure 3.10 shows the variation in the STI at various modulation transfer frequencies for these eight positions.

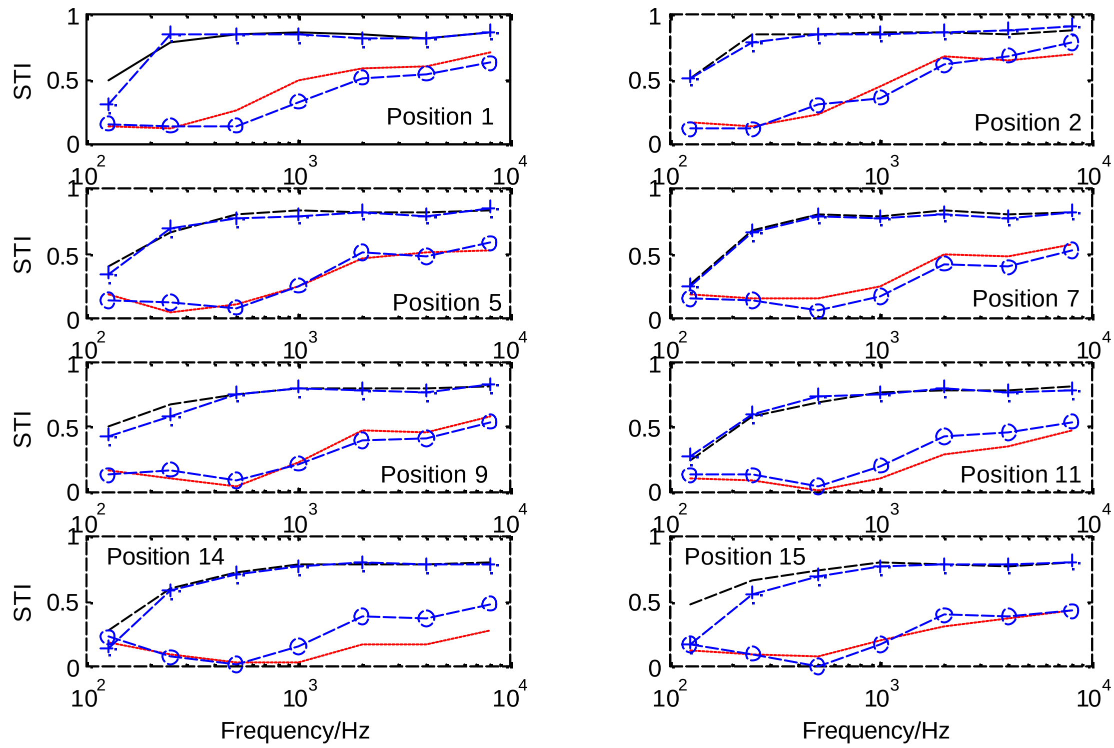

Fig. 3.10 Comparison of the two sets of signal categories in terms of STI against frequency response at the eight positions. The straight line denotes a loud signal without background noise, the dotted line a soft signal with a loud background noise, the line with crosses a loud signal with a loud background noise, and the line with circles a soft signal without background noise.

It can be seen from the figure that the maximum STI at positions 1 and 2 (the first row in the seating plan) is about 0.9, and thus these two positions are among the best in the classroom for speech intelligibility. In contrast, positions 14 and 15 (rows 8 and 9 of the seating plan) have the worst STI values at close to 0.01 from 750 Hz to 800 Hz. This is because of the combination of a soft sound signal and a high level of background noise that occurs at this position.

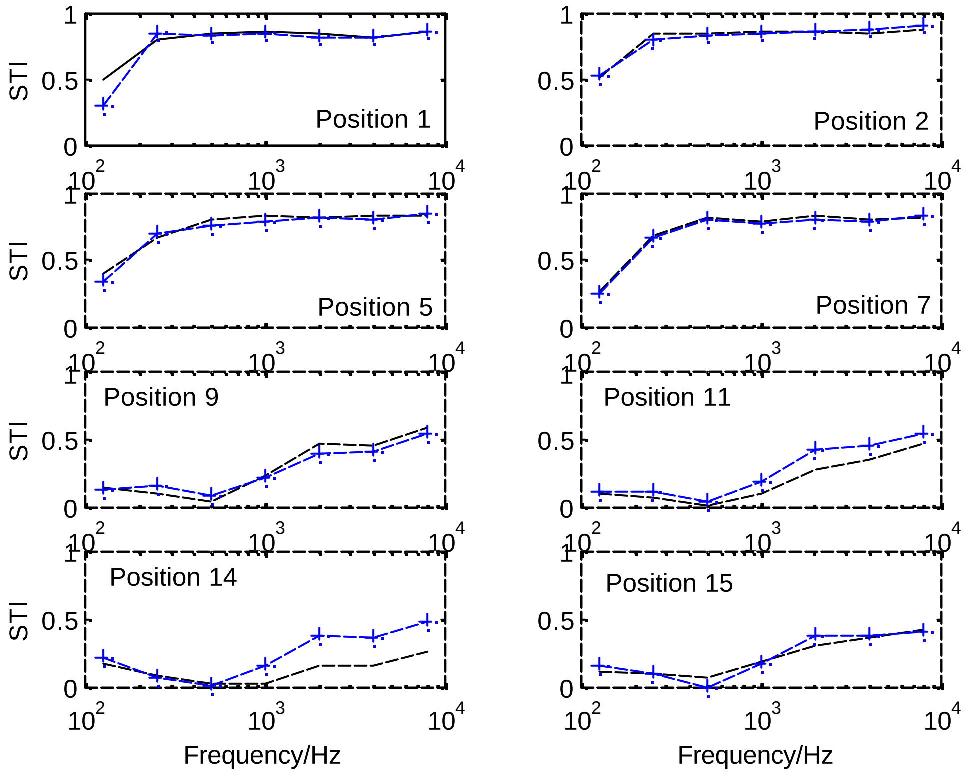

Fig. 3.11 The best fitting results for STI against frequency response in room FJ303. The straight line denotes a loud signal with a loud background noise and the line with crosses a loud signal without background noise.

In addition to the influence of the STI, as shown in Figure 3.11, there are other factors that affect speech intelligibility in the classroom. For instance, when the volume of a room is large, the resulting RT may be too long and the speech level too weak, which leads to a low S/N ratio. This means that listeners who are located between the middle and the back of the classroom will find it difficult to hear the speaker. Such a classroom would not be suitable for lecturing because of to the low level of speech intelligibility.

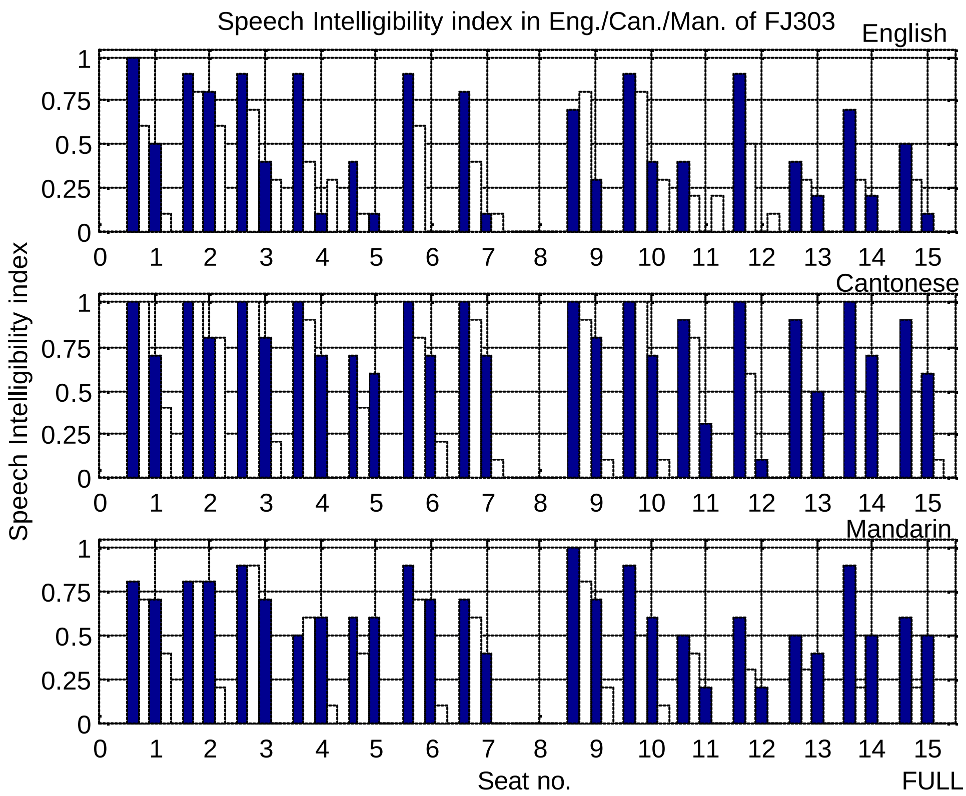

Fig. 3.12 Speech intelligibility as measured among Hong Kong university students. The first black bar denotes a loud signal measured in silent conditions, the second white bar a loud signal measured in noisy conditions, the third black bar a soft signal measured in silent conditions, and the fourth white bar a soft signal measured in noisy conditions.

Approximately 3,000 sets of data were collected from the two groups of experiments, and the results are shown in Figure 3.12. Although attendance is often uncertain in lectures, almost full seating was assumed in the experiment, except for Position 8. The symmetrical effect of the parallel-pipe shape of room FJ303 may mean that the derivation at that position was negligible. The speech intelligibility of Cantonese was found to be better than that of the other second languages, which shows that beyond doubt the mother language is more intelligible than other languages, even under the most difficult acoustic conditions.

3.1.5 Experimental setup for the speech intelligibility tests

As it is difficult to prevent bias in the language of an ordinary person, language instructors from the Hong Kong Polytechnic University who spoke Cantonese and Mandarin fluently were invited to participate in the speech intelligibility tests (these instructors are referred to as "trained listeners" hereafter). The use of experienced language instructors enabled the reduction of first-language bias error, which means that the results of the tests can be used to compare the intelligibility scores of native and non-native listeners.

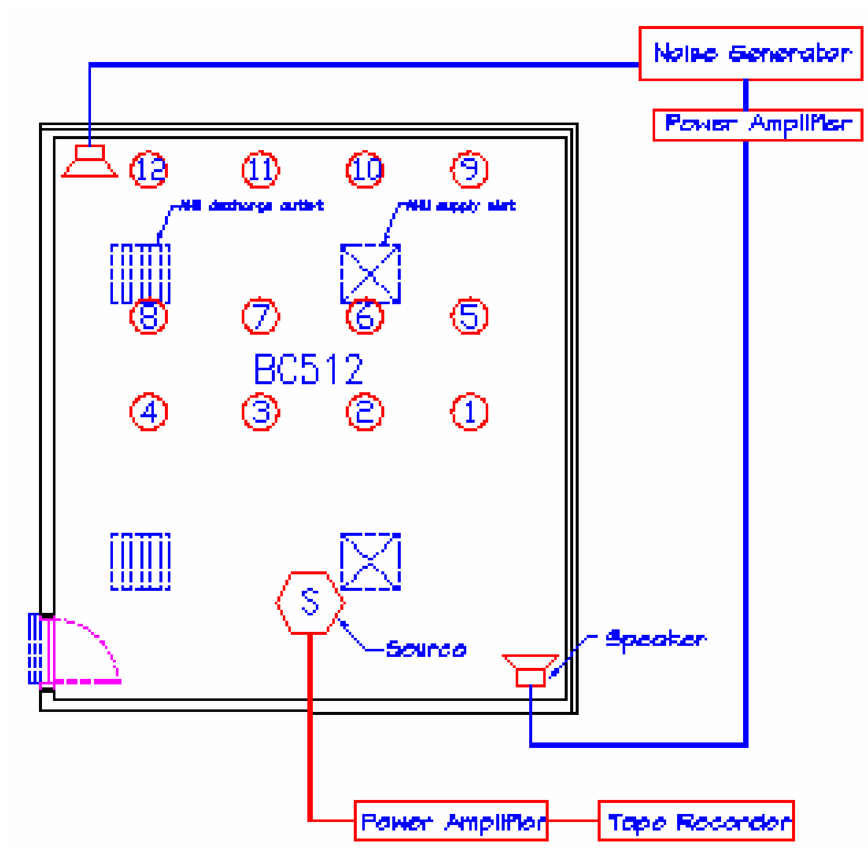

Fig. 3.13 Layout of the speech intelligibility testing equipment in classroom BC512.

Two tests were performed in two calibrated venues, rooms BC512 and FJ303, at the Hong Kong Polytechnic University. Both rooms were calibrated and the seating plan of Room FJ303 systematically controlled, as shown in Figures 3.13 and 3.14. The experimental conditions were controlled to meet two basic requirements: consideration of reverberation and ambient noise and coverage of a broad range of STI values from 0.1 to 0.9. To satisfy these two requirements, two extra loudspeakers were used to deliberately create noisy conditions inside the classroom.

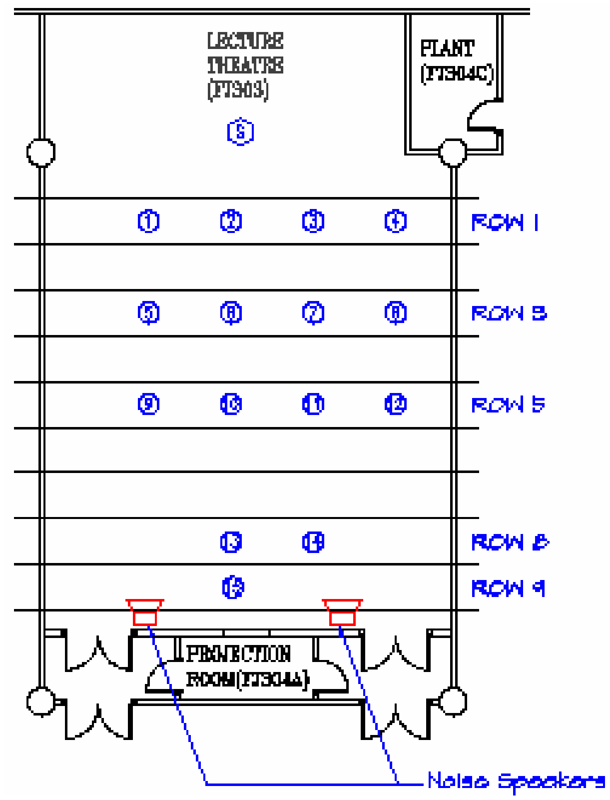

Fig. 3.14 Layout of the speech intelligibility testing equipment in room FJ303.

A list of words from each language was chosen randomly from the speech test materials. During the articulation tests to determine the level of speech intelligibility, a vocabulary was introduced, the pronunciation of which involved a combination of vowels and consonants, including affricates, fricatives, nasals, and plosives.

3.1.6 Definition of the listener and speaker categories

It is necessary to define each category of listener and speaker before presenting the results for each language.

Native listeners/speakers

We invited native speakers to record the sound samples for each language. For example, for the English language test, we invited native English speaking teachers to make recordings of the PB word list. Native speakers of the languages were used to record the samples to minimize pronunciation error.

Non-native listener/speakers

We recruited some listeners who had two mother tongues. For example, in one of the tests we recruited bilingual teachers from the language department of the university as participants, some of whom had both Cantonese and Mandarin as their mother tongue. In addition, some of the students had a father whose mother tongue was Cantonese and a mother whose mother tongue was Mandarin, and thus had grown up with both languages. These native listeners were categorized as native listeners in both the Mandarin and Cantonese tests. However, most of the university students who participated had only one mother tongue (Cantonese), and were defined as non-native listeners for the English and Mandarin tests.

Trained listeners

"Trained" listeners refer to the participants who announced the instructions at the start of the test before the test. All of the test materials were released to the trained listeners before the test, and they were given 5 minutes to go through the word list in the three languages before the test started. It was assumed that this group of listeners would be able to distinguish the vocabulary if the clarity of the signal at their position was sufficient.

Untrained listeners

This group of listeners comprised those who would be likely to recognize the words, vocabulary, or statements that they heard during the test, but took the test instantaneously and were not given the word list before taking the test. It was assumed that if the STI at their position was adequate, then they would be likely to identify a word or vocabulary, although they might confuse it with a word with similar pronunciation but a different meaning. The results from this group are the most pertinent to this study, as they should help us to identify the positions inside a classroom at which students can hear the teacher's voice, but due to reverberation or other root cause may mistake a word uttered one with a similar pronunciation but a different meaning.

3.2 Results and discussion

3.2.1 Statistical results

The experiment was carried out using two groups of participants: 40 trained listeners and 150 untrained listeners. Each group underwent at least 100 pairs of language tests, which gave a total of approximately 120,000 samples for the experiment. The articulation tests were conducted in the two classrooms, BC512 and FJ303. The experimental results for the tests are shown in Figure 3.15, and a marked difference in the performance of the trained and untrained listeners can be seen.

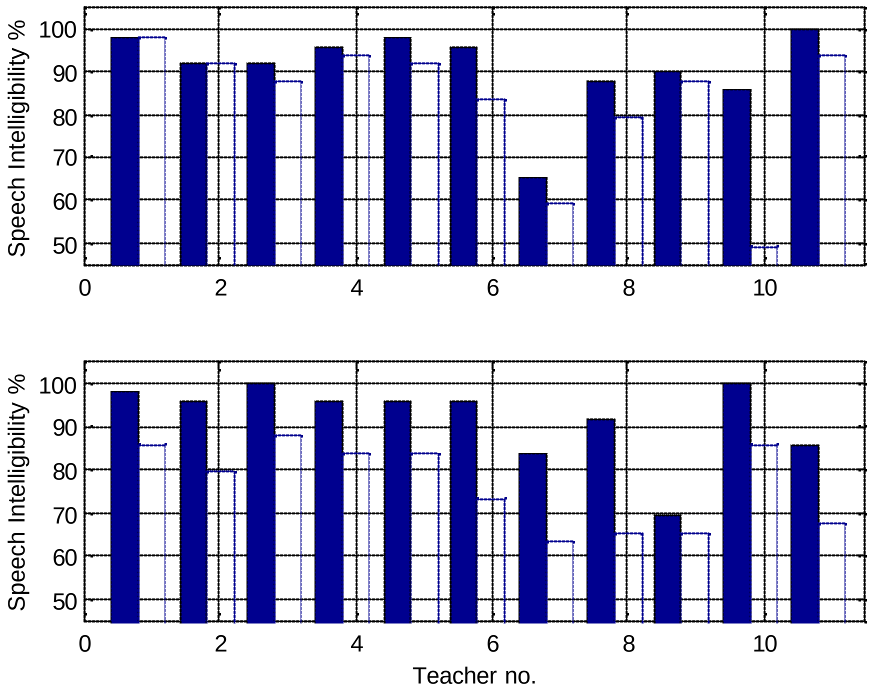

Fig. 3.15 Speech intelligibility scores for Cantonese and Mandarin in silent and noisy conditions. The black bar in the upper graph represents the intelligibility of Cantonese speech in silent conditions, the white bar in the upper graph represents the intelligibility of Cantonese speech in noisy conditions, the black bar in the lower graph represents the intelligibility of Mandarin speech in silent conditions, and the white bar in the lower graph represents the intelligibility of Mandarin speech in noisy conditions.

To explore other parameters that may have affected the speech intelligibility in the classrooms, such as the clarity and reverberation time, it is necessary to examine the correlation between these parameters. In Figure 3.16, we show the variation in the measured clarity (or useful-to-detrimental in 50 ms) with the measured RT at a frequency of 1kHz. The definition of the clarity or useful-detrimental ratio in 50 and 80 ms is discussed in Chapter 1, and is explored in greater detail in the next chapter.

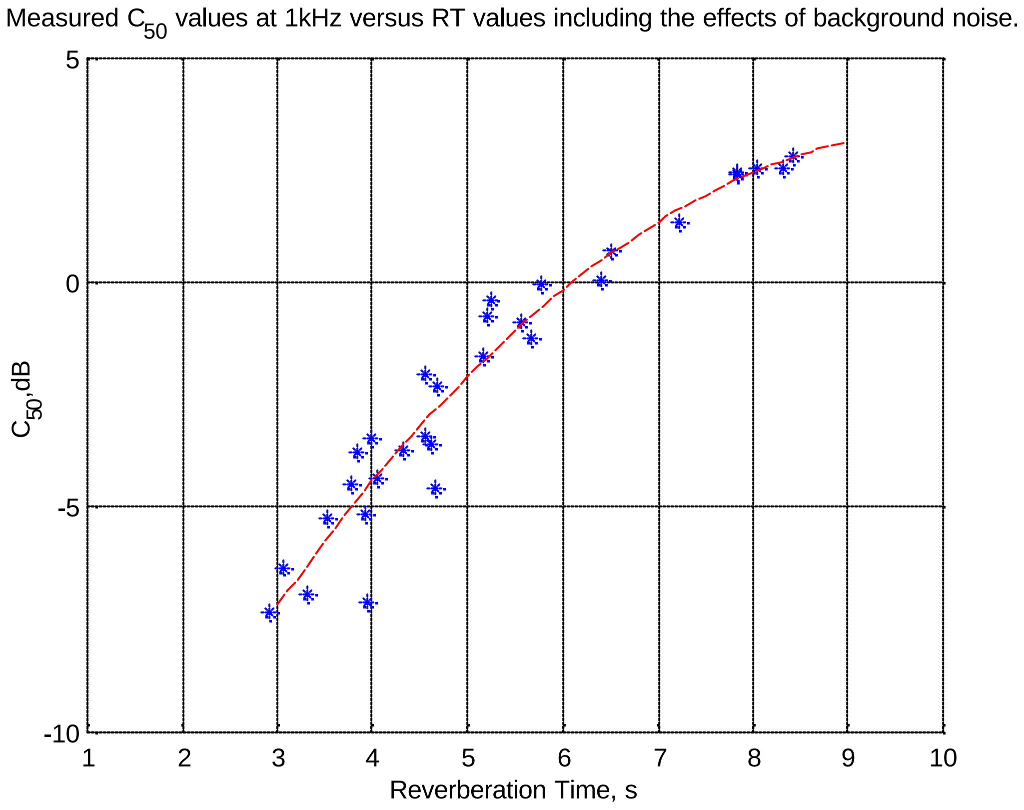

Fig. 3.16 Measured clarity 50 versus RT in the 1 kHz octave band. The effect of the ambient noise is included in the measurements.

According to the results, the variation in the clarity 50 was directly proportional to the RT in the 1 kHz octave band. We also obtained similar preliminary measurements for the EDT, D_{50}, and S/N ratio, which confirms the validity of the articulation tests and any subsequent analyses that are based on the tests.

The trained listeners in the Cantonese and Mandarin tests (teachers at the Chinese Department of the Hong Kong Polytechnic University) were located at Positions 1 to 6.

Their performance was very similar. The performance of the remaining listeners, who were either teachers or students from the engineering department, varied significantly. The results were separated into categories. All of the preliminary results were taken from room BC512, and further data were then collected from room FJ303. The trained and untrained listeners were invited to conduct articulation tests in room FJ303 under silent and noisy conditions, the results of which are summarized in Figure 3.17. The average intelligibility scores were also measured. Some of the teachers had lower scores in silent conditions than in noisy conditions, which is reasonable. Under certain specific conditions, reverberation or standing waves occurred at some positions. The statistical histogram in Figure 3.17 offers an adequate database for the final analysis of the speech intelligibility and its relation to the STI.

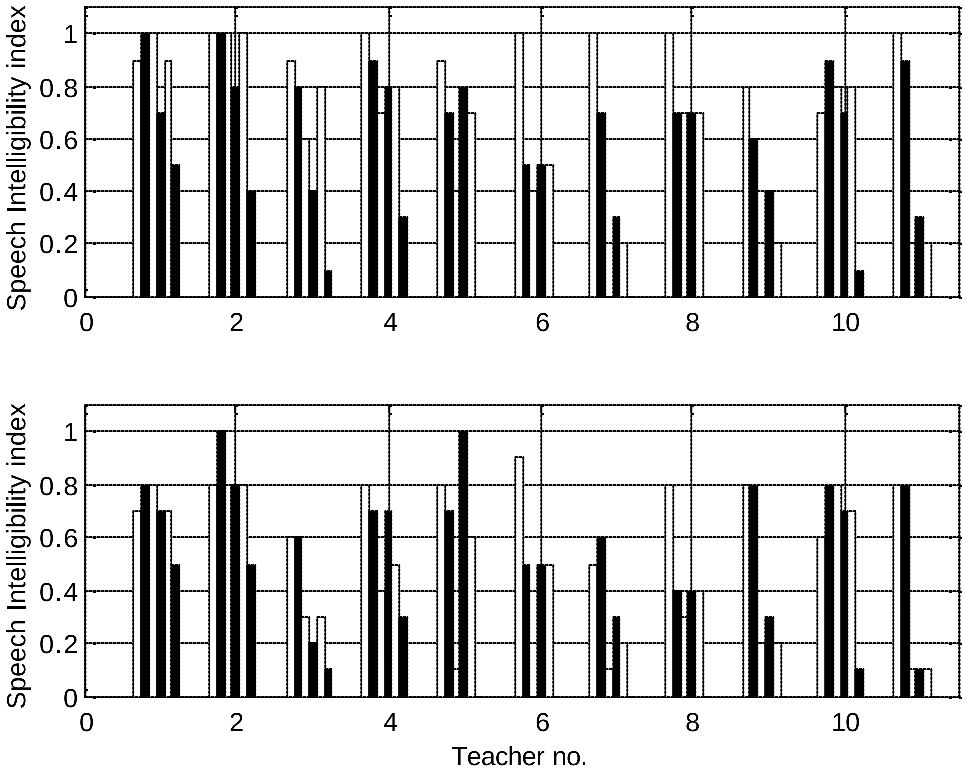

Fig. 3.17 Speech intelligibility scores for Cantonese, English, and Mandarin speech in silent and noisy conditions in FJ303. The upper graph represents speech intelligibility with a loud signal and the lower graph represents speech intelligibility with a soft signal. The first (white) bar denotes the intelligibility of Cantonese speech in silent conditions, the second (black) bar denotes the intelligibility of Cantonese speech in noisy conditions, the third (white) bar denotes the intelligibility of Mandarin speech in silent conditions, the fourth (black) bar denotes the intelligibility of Mandarin speech in noisy conditions, the fifth (white) bar denotes the intelligibility of English speech in silent conditions and the sixth (black) bar denotes the intelligibility of English speech in noisy conditions.

3.2.2 Comparison with other studies

To ensure that the variance in the experimental results is within an acceptable distribution, a comparison is made with acoustical data from other studies. This enables us to verify whether the results accurately show the discrepancy between the level of intelligibility under the various conditions among the native listeners and non-native listeners.

The students who participated in the two articulation tests showed a better understanding of the English PB words than the Mandarin words (see Figure 3.18), as is demonstrated by the fact that the intelligibility scores for the English words became “saturated” at a higher value of about 95 percent. Nevertheless, we can still conclude that a quieter classroom (more specifically, a room with a higher STI) is generally required for non-native speakers of a language to render speech in that language intelligible.

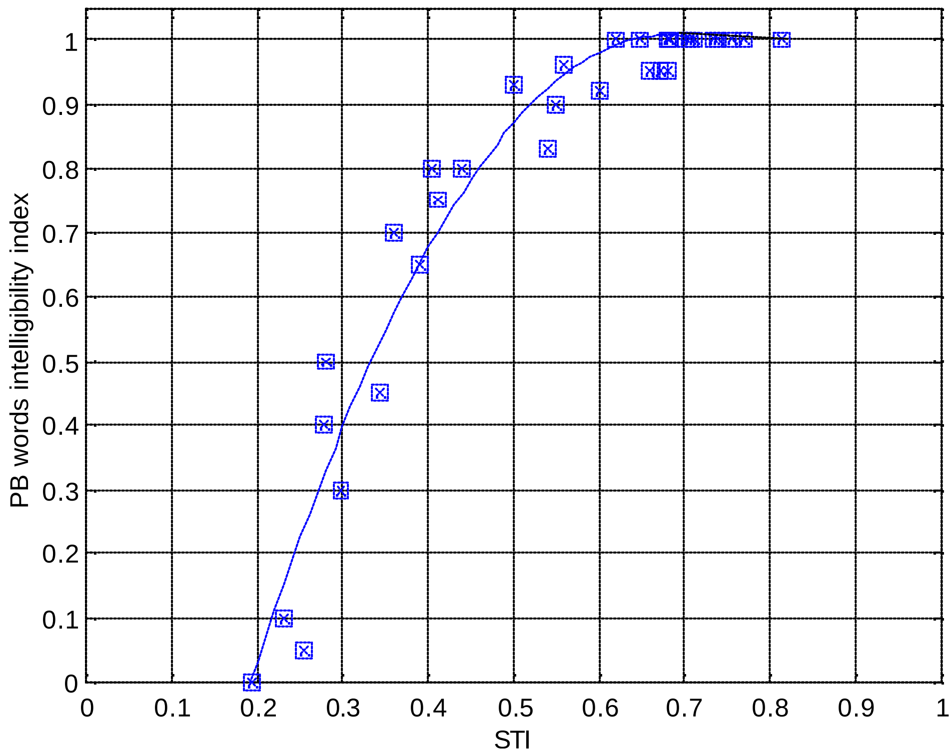

Fig. 3.18 Speech intelligibility scores for Cantonese versus STI in the experiment in room FJ303.

As Cantonese is the native language of Hong Kong, we further explored the variation in the PB word scores with the STI among the untrained listeners whose native language was Cantonese. The initial set of data came from the articulation tests that were conducted in Room FJ303. Figure 3.19 shows a plot of the speech intelligibility scores versus the STI for this test. As in Kang's experimental results for native speakers of English and Mandarin, the speech intelligibility scores reached a saturation point when the STI exceeded 0.6. The intelligibility scores for Cantonese and Mandarin are also compared in the figure. As expected, the Cantonese scores were higher because Cantonese was the first language of all of the participating students.

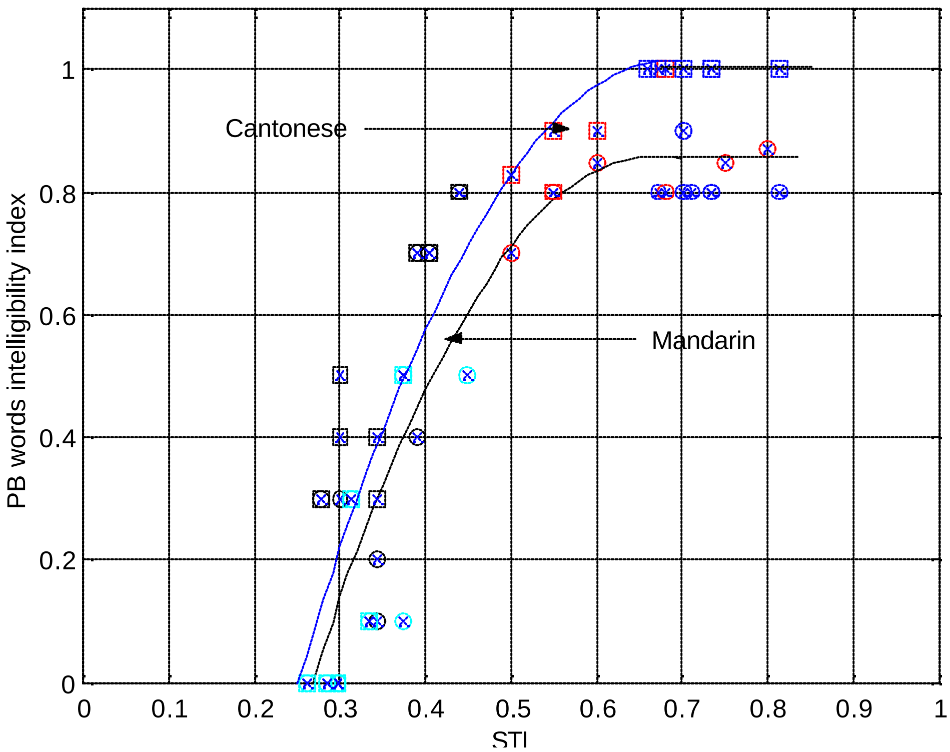

Fig. 3.19 Speech intelligibility for Cantonese and Mandarin versus STI in room FJ303.

Figure 3.19 also shows that the PB word scores for Mandarin among the untrained listeners saturated at an STI of around 0.83-0.85. This value would have been close to 1 had the listeners been native speakers of Mandarin. The discrepancy between the results is largely due to the personal knowledge of the listeners, their background, and their previous exposure to the language.

3.3 Prediction method for the empirical equations

The models of Kang, Hodgson, and Bradley were investigated and compared with the experimental results for Cantonese speech. These models are analyzed and discussed in detail in the next chapter. We used the derived empirical formulations from the three models to fit the experimental data for the three languages, the results of which are shown in Figures 3.20-3.22. Linear regression was used to obtain the best fit lines for the experimental data.

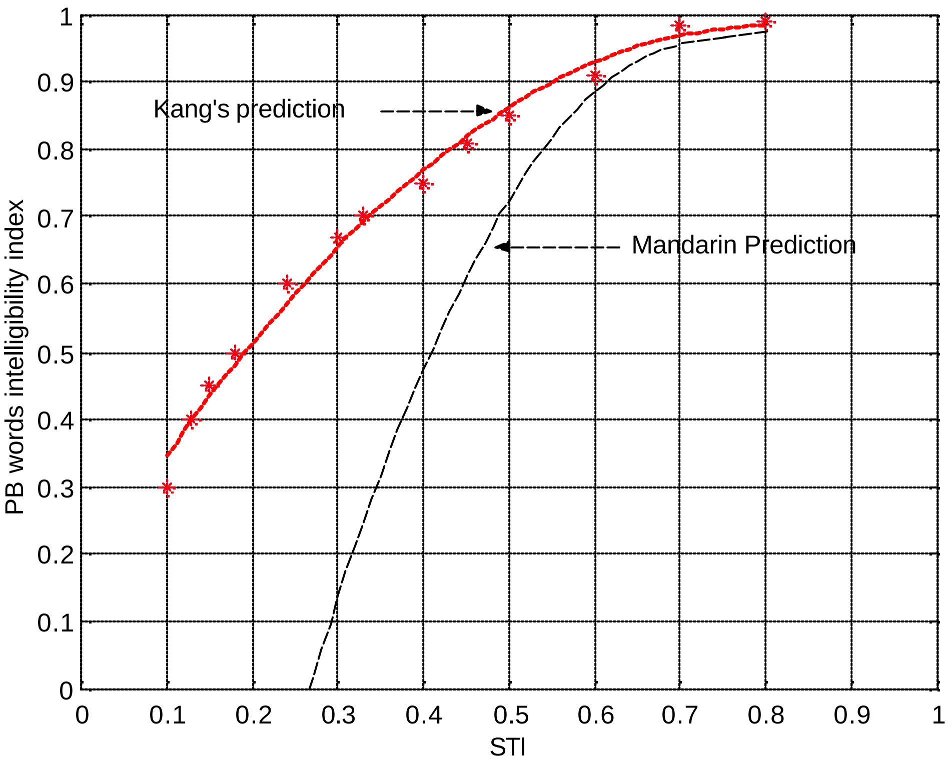

Fig. 3.20 Empirical equation for Mandarin speech: comparison of the experimental results and those of Kang.

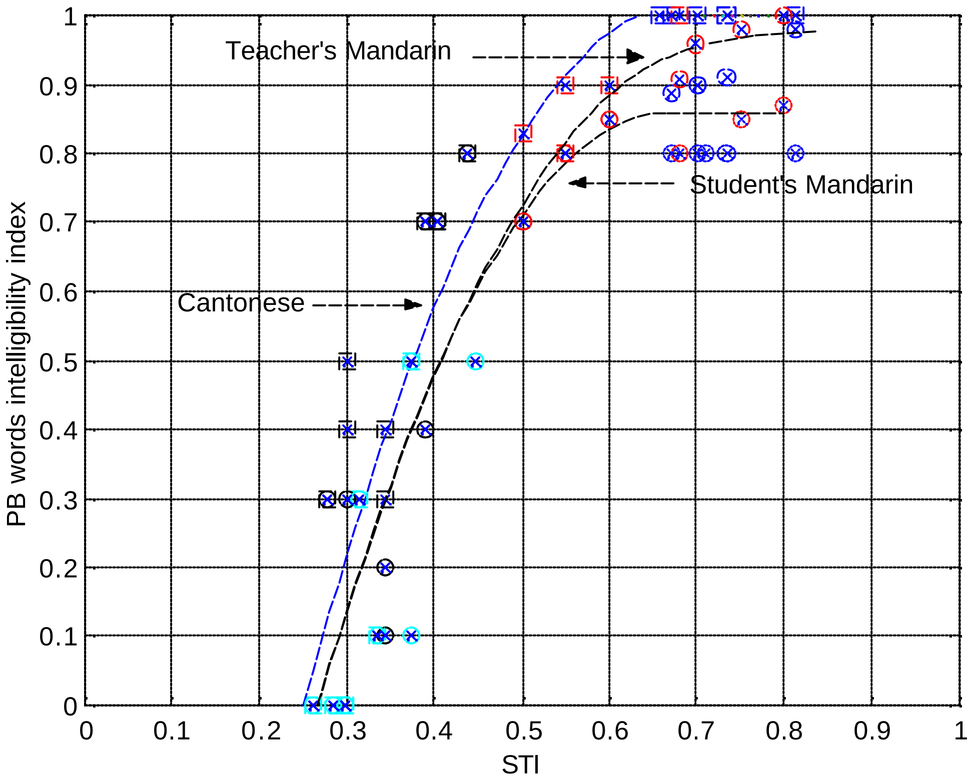

Fig. 3.21 Empirical equation for Cantonese and Mandarin speech: teacher and student model.

Figure 3.21 clearly shows a discrepancy in the performance of the native and non-native listeners. The plotted graph shows the Cantonese and Mandarin results for the group of bilingual teachers from the language department who had both Cantonese and Mandarin as a mother tongue. The curve of the students' performance shows the results of a group of ordinary university students in the Mandarin test. The saturation point of 0.87-0.88 clearly shows that the STI could not determine whether a signal would be precisely heard by the non-native speakers inside the classroom.

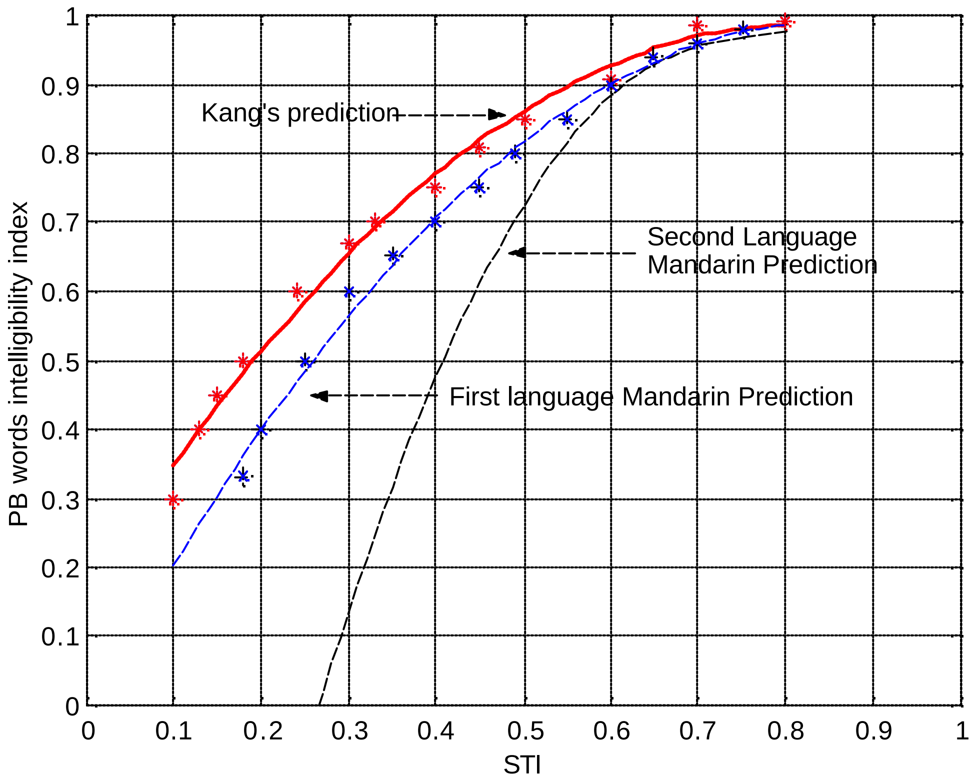

Fig. 3.22 Empirical equation for the experimental data for Mandarin speech: a comparison with Kang's model.

Figure 3.22 shows a comparison of Kang's [34] experimental results for native speakers of Mandarin with the experimental results of this study for the trained and untrained listeners. The PW word scores for the trained listeners in this study are similar to those that were obtained by Kang with native speakers of Mandarin. However, the untrained listeners achieved consistently lower PB word scores. The empirical formulae for the three separate experiments are shown.

3.4 Overall empirical equations for the test results of the untrained listeners in rooms FJ303 and BC512

To summarize, the empirical formulae for the intelligibility of the three languages among untrained listeners in the two rooms are as follows.

(1) Speech intelligibility of Cantonese (*CAN*):

$$CAN = -4.26(STI)^2 + 5.78(STI) - 0.955 \quad (3.1)$$

(2) Speech intelligibility of Mandarin (*MAN*):

$$MAN = -2.88(STI)^2 + 4.24(STI) - 0.668 \quad (3.2)$$

(3) Speech intelligibility of English (*ENG*):

$$ENG = -2.88(STI)^2 + 4.24(STI) - 0.668 \quad (3.3)$$

In contrast, the empirical formulae for Kang's data [31] on the speech intelligibility of English and Mandarin are as follows.

$$ENG = -1.2(STI)^2 + 2(STI) - 0.16 \quad (3.4)$$

and

$$MAN = -1.3(STI)^2 + 2(STI) + 0.15 \quad (3.5)$$

Bradley [33-34, 36] gave the empirical formula for the speech intelligibility of English as follows.

$$ENG = -1.3(STI)^2 + 7.3(STI) - 1.5 \quad (3.6)$$

3.5 Summary

This chapter reviews the data that were collected in previous chapter studies and introduces the definitions of the various categories of listeners and speakers, namely, native and non-native speakers and trained and untrained listeners. To ensure that the data are consistent, statistical and variance analysis is applied to show the differences between the data for each category.

The objective of this study was to compare the variation in the speech intelligibility of Cantonese, Mandarin, and English with the STI at various positions in different classrooms. A methodology has been successfully developed to measure the speech intelligibility of the three languages against the STI, and empirical formulae have been determined to predict the speech intelligibility for the different languages for a given STI.

Chapter 4

The effect of different languages among different categories of listeners and speakers

In Hong Kong, Cantonese, Mandarin, and English are all used for teaching activities in the classroom. It is therefore essential to determine the characteristics of these three languages – which are often used interchangeably in verbal communication in the classroom – including the difference between their intelligibility as first and second languages. Many Hong Kong residents frequently speak English, Cantonese, and Mandarin to communicate with each other in the street, at work, and in the classroom, and teachers frequently use the different languages to express their ideas to students during classroom activities. However, it may be the case that there is some discrepancy between the intelligibility of the languages that speakers use to communicate with each other in certain conditions.

The intention of this chapter is first to study the main physical parameters that affect speech intelligibility in the different languages, such as room dimensions, volume, RT, and S/N ratio. Second, it aims to reveal that the speech intelligibility of first and second languages is not the same at different positions inside the classroom, even when the same testing materials and trained listeners with a similar educational background are tested. As the statistical analysis of variance was completed in Chapter 3, this chapter focuses on the intelligibility of each language for the various categories of listeners and speakers.

4.1 Measurement of the speech intelligibility of the different languages

This chapter investigates the pronunciation of English, Cantonese, and Mandarin through the development of speech intelligibility indices for each language. Words are a natural unit for modeling the speech intelligibility of a language using an articulation index, as continuous sentences and statements involve combinations of words that are composed of different sets of sounds with specific contexts. In English, this is referred to as a list of phonetically balanced (PB) words, and is one of the methods that are used to identify a combination of different words. Vowels and consonants are the basic components that control the pronunciation of words.

4.1.1 Auralization

The International Phonetic Alphabet (or IPA) [8] is the recognized standard for pronunciation, and comprises a base of approximately 75 consonants and 25 vowels, which covers most of the languages that are spoken in the world. Many speech acousticians have characterized the similarities and differences between the various phones and phonemes in the IPA. In the following sections, we describe how the human vocal tract controls its various parts, which are shown in Figure 4.1, to produce different sounds.

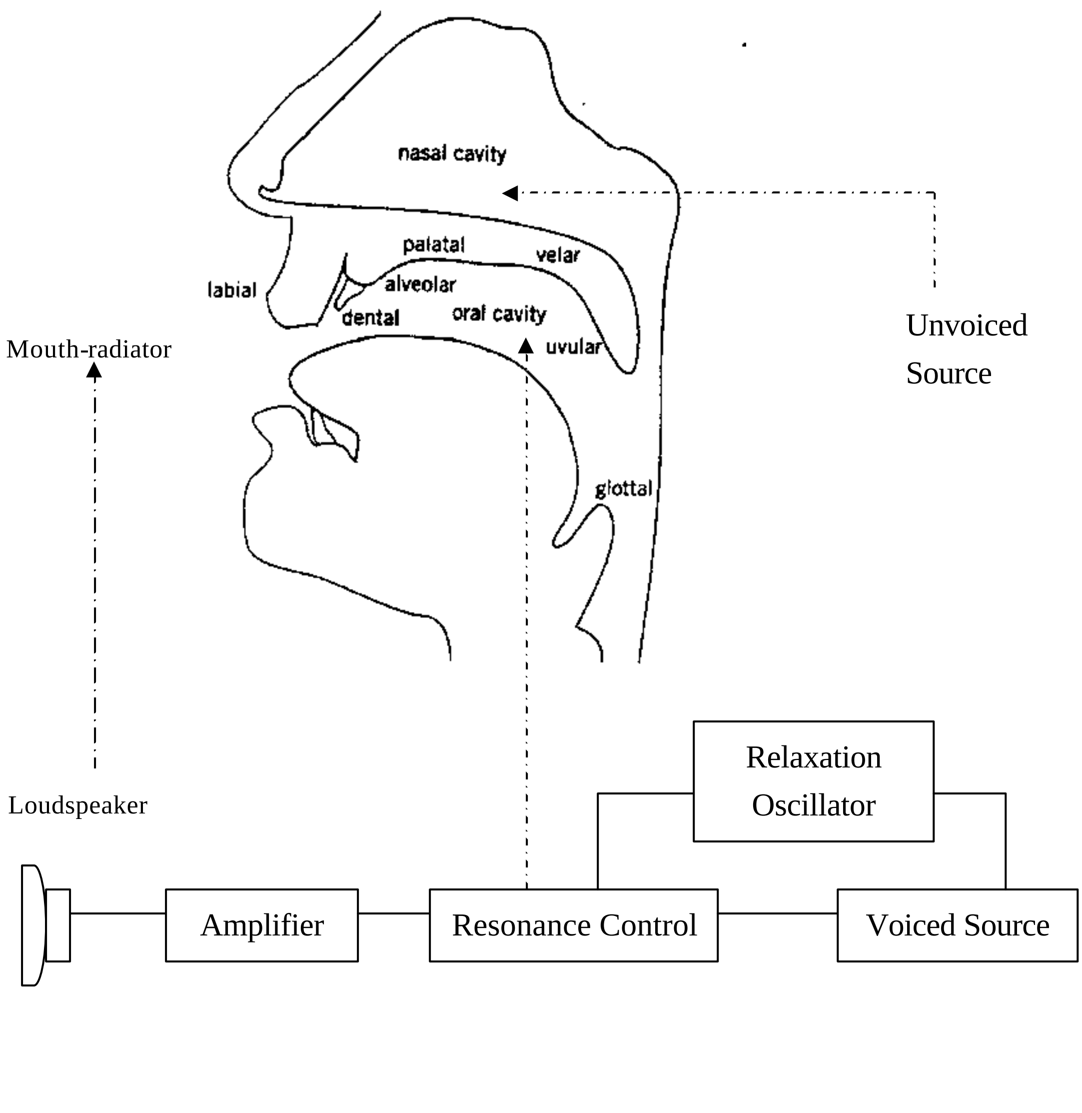

Fig 4.1 Vocal control of pronunciation and the parts of the oral cavity.

4.1.2 Articulation in English

In English, consonants are formed by constricting the tube of the vocal tract in various ways, usually with the tongue. The two main streams of sounds that are used to determine the similarities between phones are the place and manner of articulation. The place of articulation refers to the point of closest constriction in the oral cavity. The various manners and places of articulation that are found in English and some other languages are as follows.

Labiodental

Labiodental segments require speakers to place their lower lip against their upper teeth, as in the phones [f] and [v].

Interdental

For these segments, the tongue is placed between the teeth. The only sounds that fall into this category in English are those at the beginning of *thing* and *that*.

Palatal

The only English phone that is made by a constriction on the hard palate is that at the beginning of the word *judge*, although this sound can also be palatal-alveolar. In German, the sound of the letters "ch" after certain vowels such as "i" is made by a palatal constriction.

Alveolar

Behind the teeth sits the alveolar ridge. Constrictions here produce sound segments such as [t], [n], and [z].

Stops

These segments are produced by the complete blockage of the airstream. Stops often involve two acoustic events: the complete stoppage of the airstream (often called the closure), and then, after a short duration, the release of the closure. An example of this type of stop is [p]. Another type of stop is the flap or tap, the alveolar version of which is often heard in American English, replacing [t] or [d] in words like *butter*. Flaps are often shorter than full stops, with the tongue merely tapping the roof of the mouth for a short time before returning to a less constricting position.

Fricatives

Fricatives can be thought of as "almost stops," in that the tongue comes close enough to a complete closure to create turbulence in the airstream.

Affricates

An affricate is a combination of a stop and a fricative. Affricates start with a stop closure, but instead of a normal stop release, a fricative is produced. In the IPA, this is denoted by the combination of two symbols, such as in the English word *cherry*.

Nasals

Nasals are related to stops. Nasals such as [n] close off the oral cavity in the same way that stops do, but the nasal passage is opened at the velum, allowing air to escape through the nose, rather than building up behind the closure as in a stop.

4.1.3 Articulation in Cantonese and Mandarin

Several rules and criteria must be satisfied for the correct pronunciation of these languages. In Cantonese, there are 19 vowels and 53 consonants, and in Mandarin there are 23 vowels and 36 consonants. In Cantonese, there are up to six different lists of phonetically balanced words, and each phonetically balanced word has a number that is used to identify the method of pronunciation. The numbers are commonly captioned according to the movement of sound in the vocal cavity, for example, fu^{1}, sin^{5} and wa^{6}, which are shown in Figure 4.1. The pronunciation of the word lists in Mandarin is less complicated, and is based on four or five different combinations of movements to control the airstream inside the oral cavity.

4.2 Articulation loss for consonants (AL_{cons})

In Chapter 2, we discuss the calculation and measurement of the modulation transfer function (MTF) and Speech Transmission Index (STI). It is important to note that all of the measurements and analyses in this study are based on the assumption of a correlation between the Speech Transmission Index (STI) and the articulation loss for consonants (AL_{cons}). According to Houtgast and Steeneken [71, 74-75], STI is directly proportional to the AL_{cons}, as shown in Figure 4.2.

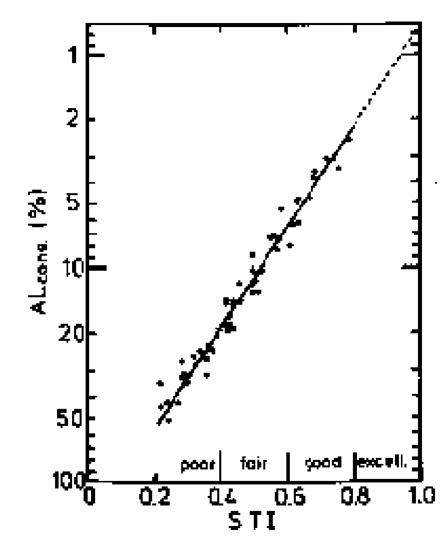

Fig 4.2 Relationship between the physical STI and articulation loss for consonants (AL_{cons}) [71, 73].

In general, the limit of the S/N ratio should be set at ±30 dB to maintain satisfactory speech intelligibility. Houtgast and Steeneken [71-75] suggested that the loss of consonants is directly correlated to the STI, and that the S/N ratio can therefore be used to evaluate the range of articulation loss for consonants (AL_{cons}).

Table 4.1 shows the AL_{con} and STI measurements that were obtained in Room FJ303. The first column shows the STI measured at various MTF values, and the bottom line shows that the overall STI was 0.756 with an AL_{con} of 2.8%. This set of measurements can be classed as excellent, as the scores for speech intelligibility are high.

MTF Matrix (Calibrated)

Frequency-Hz	125	250	500	1000	2000	4000	8000
level dB-SPL	51.2	59.7	59.5	58.8	61.3	58.8	60.0
m-correction	1.000	1.000	1.000	1.000	1.000	0.999	1.000
0.63	0.465	0.945	0.984	0.989	0.993	0.991	0.995
0.80	0.428	0.945	0.982	0.986	0.990	0.987	0.993
1.00	0.382	0.944	0.978	0.982	0.985	0.982	0.989
1.25	0.344	0.938	0.973	0.976	0.978	0.974	0.985
1.60	0.326	0.926	0.963	0.968	0.967	0.960	0.977
2.00	0.301	0.910	0.950	0.956	0.951	0.942	0.966
2.50	0.271	0.896	0.932	0.939	0.929	0.915	0.951
3.15	0.326	0.866	0.905	0.916	0.896	0.877	0.930
4.00	0.349	0.836	0.866	0.885	0.849	0.822	0.900
5.00	0.354	0.804	0.818	0.850	0.792	0.756	0.866
6.30	0.216	0.768	0.754	0.809	0.721	0.672	0.822
8.00	0.261	0.731	0.684	0.764	0.635	0.571	0.770
10.00	0.238	0.690	0.625	0.720	0.543	0.470	0.715
12.50	0.160	0.657	0.576	0.670	0.446	0.381	0.655
octave MTI	0.384	0.777	0.820	0.846	0.804	0.781	0.857

STI value= 0.756 (0.807 modified)   ALcons= 2.8%   Rating= EXCELLENT

ESC to exit, F1 to print, Shift-F1 to dump.   MLSSA: STI

Table 4.1 An example dataset of measurements from room FJ303.

From Table 4.1, it is possible to determine the critical level at which noise interferes with speech in the classroom. This critical level should lead to a criterion for determining a tolerable fluctuating noise level in terms of L_{10}, rather than L_{eq}. An experiment can be conducted using the measured AL_{cons} to verify this assumption. The measured data are unacceptable if the calculated AL_{con} according to the MLS is significantly different from the AL_{com} according to Figure 4.4 for the measured STI. For the example shown, when the measured STI is 0.756, the AL_{con} is found to be about 2.65%. This value is fairly close to the measured AL_{con} of 2.8%, and thus the set of measured data is deemed to be acceptable.

4.3 Comparison of prediction methods

The design of a classroom with good acoustics that allows a high level of speech intelligibility in such a highly populated and technological city as Hong Kong is a controversial topic. In university classrooms, the major acoustical concern is verbal communication. Non-optimal acoustical conditions that result in reduced verbal communication can lead to two main problems: they can reduce learning efficiency and can result in fatigue, stress, and health problems (headaches and sore throats) among lecturers who are forced to compensate for the poor acoustical conditions by raising their voices.

Many scientists and acousticians have tried to solve this puzzling and complex problem and have used different prediction methods to optimize the problem. Some of these prediction methods are detailed in the following sections.

4.3.1 Kang's prediction method

- Experimental condition and target:

  Kang compared the speech intelligibility of English and Mandarin [31] in a closed space by taking several measurements inside the long enclosure of a railway station beyond a corridor using the MIDAS system [31-32] and a tape recorder.

- Testing materials:

  Twenty-five words each were chosen from the PB word lists of English and Mandarin, and 25 sentences in each language that contained the words were recorded on a signal tape for use in the test. The sentences in each language were recorded by two male and two female native speakers. White noise was also provided to simulate the background noise of the trains.

- Conclusion:

  Kang's prediction model not only considered the reverberation, but also the ambient noise effect on the intelligibility of speech on the platforms when trains passed through the enclosed corridor. He discovered that Mandarin was more intelligible than English inside the enclosure. More precisely, he found that sentence intelligibility was better in English than in Mandarin, but that Mandarin was slightly more intelligible than English under reverberant conditions. However, the intelligibility of English was considerably better under noisy conditions.

4.3.2 Bradley's prediction method

- Experimental condition and target:

Bradley's studies [13-15] were based on numerical predictions of different acoustical parameters, such as the S/N ratio, clarity ratios C_{50} and C_{80}, and the useful-detrimental ratio U_{80} in 10 occupied classrooms. The target listeners in the test were junior students in grades seven and eight (12- to 13-year olds).

- Testing materials:

Four levels of loudspeakers that varied according to local conditions were used to produce a sound that was similar to the human voice. Four measurement groups were set up in each of the 10 classrooms, and four speech levels were tested. In total, 160 sets of data were obtained from the classrooms, which contained an average of 24.3 students.

➢ Conclusion:

The optimum reverberation time (RT) for the classrooms was determined to range between 0.4 and 0.5s. Bradley concluded that to ensure speech intelligibility for all age groups of listeners with normal hearing, the background noise level should be limited to approximately 30dB(A).

4.3.3 Hodgson's prediction method

➢ Experimental condition and target:

Hodgson undertook many studies to predict and compare the acoustical characteristics of different classrooms. As classrooms of a different shape, height, width, and volume behave differently, he aimed to find out the optimum size, absorbency, and shape of a classroom that would ensure the best speech intelligibility. Most importantly, he was also interested in finding out the difference acoustical characteristics of occupied and unoccupied classrooms. He randomly selected 10 occupied classrooms and 30 unoccupied classrooms at the University of British Columbia in which to carry out his measurements [51-52, 55].

➢ Testing materials:

A maximum length sequence system analyzer (MLSSA) was used for the predictions, which included one signal-generating source with between three and eight microphone positions. All of the predictions were made in both the occupied and unoccupied conditions. The STI, RT, and S/N ratio were used to represent the acoustical characteristics of the classrooms. For example, an STI of 0.0 to 0.3 denoted bad acoustical characteristics, an STI of 0.3 to 0.45 poor characteristics, and an STI of 0.75 to 1.0 excellent characteristics. No speech intelligibility scoring test was included in the experiments. Instead, the level of speech intelligibility was determined using the STI and global STI (GSTI). Finally, Hodgson also examined the variation in the STI with distance or doubling distance (dd).

- Conclusion:

Hodgson concluded that unoccupied classrooms have far from optimum acoustics, as many suffer from excessive reverberation. Using his results, Hodgson developed empirical models to predict the 1-kHz octave band early decay times and total A-weighted speech levels in classrooms in any state of occupancy.

4.3.4 Summary of the prediction methods

We can now compare the prediction method of this study with those of the three aforementioned studies. Kang's predictions mainly concentrated on the analysis of long enclosures, and the main contribution of his work is related to controlling the noise inside a railway station to create a better environment for passengers on platforms waiting for trains. The speech intelligibility of English and Mandarin in these conditions differs from the results that we obtained inside the classrooms. Moreover, the trained listeners in Kang's study were local Chinese people whose mother language was Mandarin, unlike Hong Kong citizens, whose mother language is Cantonese. Another difference is that Kang used the MIDAS system, whereas we used the MLSSA. Because of these differences, Kang's results can only be used as a general reference for classroom acoustics.

Bradley gave many empirical equations for the relationships between clarity, reverberation time, and the useful-detrimental ratio. However, because of the lack of use of he did not use the experimental data instrument MLS, the STI was not taken into account in determining these relationships, although the speech intelligibility was measured and matched with the predictions of the equations. Another difference between Bradley's work and this study is that all of the data in Bradley's study were based on English, rather than Cantonese or other languages.

Hodgson contributed a great deal to the study of classroom acoustics by taking measurements using the MLS and attempting to determine the relationship between different acoustical parameters. However, his target was to determine the influence of occupancy on classroom acoustics, and although. Although he empirically predicted the STI and global STI, the data were based on the shape, size, and height of the interior of the classrooms studied, whereas in this study language is the major concern. Nevertheless, Hodgson's work is useful for finding the optimum STI and speech intelligibility of different languages in any type of classroom.

4.4 Results and discussion

Different languages have different methods for the pronunciation of vowels and consonants, as is discussed earlier in this chapter. It is therefore useful to compare the speech intelligibility of the three main languages that are used in Hong Kong – English, Cantonese and Mandarin – in classrooms with similar acoustic environments.

In the earlier studies that are described in Chapter 3, we used English to compare the difference in relative intelligibility among local Hong Kong students who were native and non-native speakers of the language [31]. This chapter gives the experimental results of the experimental study of the intelligibility of English, Cantonese, and Mandarin under various classroom conditions, and compares them with the predictions of other researchers.

Five open testing experiments were conducted in classroom FJ303. Each experiment involved 15 people who were seated at evenly distributed positions in the room. The whole test took around 45 minutes, and was divided into four conditions with different S/N ratios and background noise levels. In each test, 200 PB words were randomly selected from the 2,000 words in the PB word list to ensure that at least 2,000 useful data items would be generated for analysis.

Fifty trained listeners, who comprised language instructors, native speakers, and teachers, and 300 untrained listeners were invited to participate in the tests. Before testing, it was necessary to confirm that all of the members of the testing audience had a healthy auditory system. The participants were required to complete a questionnaire confirming that their auditory system was healthy, and that they had not had surgery on their auditory system in the previous six months.

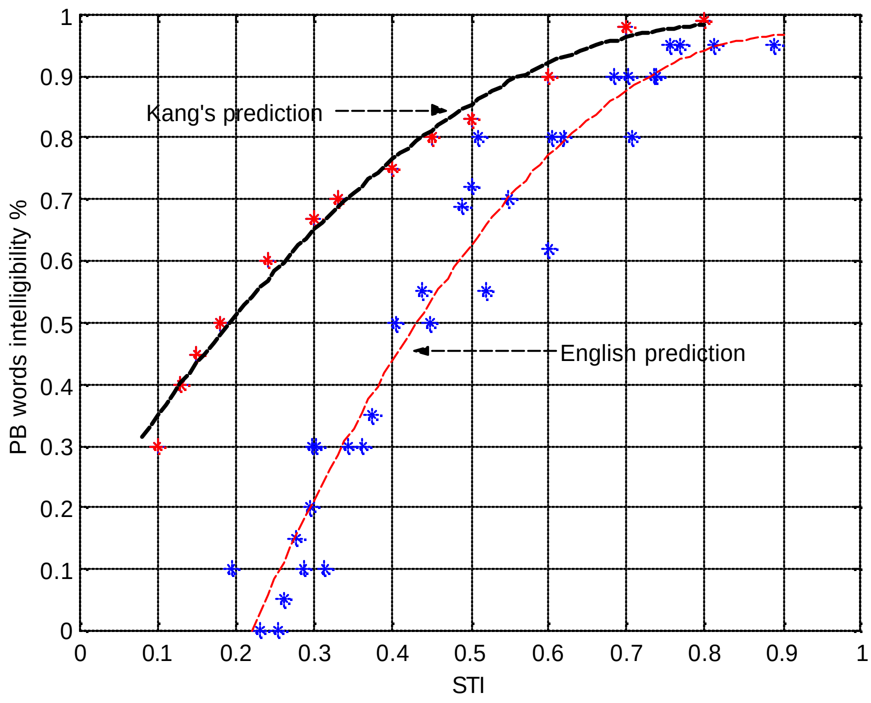

Fig 4.3 A comparison of the speech intelligibility of English PB words according to Kang's prediction and the results of the experimental study.

The experiment provided approximately 3,000 datasets for the various positions in the classroom, and so the next step was to identify the data that were essential and reasonable. This involved checking each dataset with a reasonable AL_{cons} with the correlated STI plot from previous studies [71, 75]. After checking, the preliminary experimental results in the same environment and conditions yielded approximately 2,000 useful data findings, which are shown in Figure 4.3.

Figure 4.3 shows a plot of the STI versus the PB word scores for English among a group of listeners whose native language was English and a group of typical Hong Kong listeners who used English as a second language. Kang's experimental results for the speech intelligibility of English are also shown in this figure for comparison.

The curve fitting in the graph is reasonable, although there is a degree of discrepancy between our results and those for the native speakers in Kang's study. This is largely because Kang's experiment was conducted in a corridor, rather than a classroom, and because the data for Kang's curve came from native speakers only, whereas the data from our study is based on both native and non-native speakers. It is interesting to note the differences in the level of intelligibility of English among the native speakers and non-native speakers who use three different languages to communicate inside the classroom, but it is important to further investigate the results for the other languages, Cantonese and Mandarin, which were the mother languages of all of the trained listeners and some of the untrained listeners.

To ensure a good speech intelligibility range for English, the university students in this test required a certain level of AL_{cons} with respect to the STI in the classroom. The different range of STI values implies that the data fall into different regions of AL_{cons} rating. These ratings are divided into four categories: poor, fair, good, and excellent, according to the clarity of the sound signal that is received by the listeners, which directly influences the articulation index and therefore speech intelligibility simultaneously.

Figure 4.4 shows a comparison of the intelligibility of Mandarin in the same test circumstances among Kang’ s native speakers and the Hong Kong university students in this study. It can be seen that there is a discrepancy in the lower range of the STI of between 0.1 and 0.5, which means that the lower the STI, the worse the intelligibility of Mandarin.

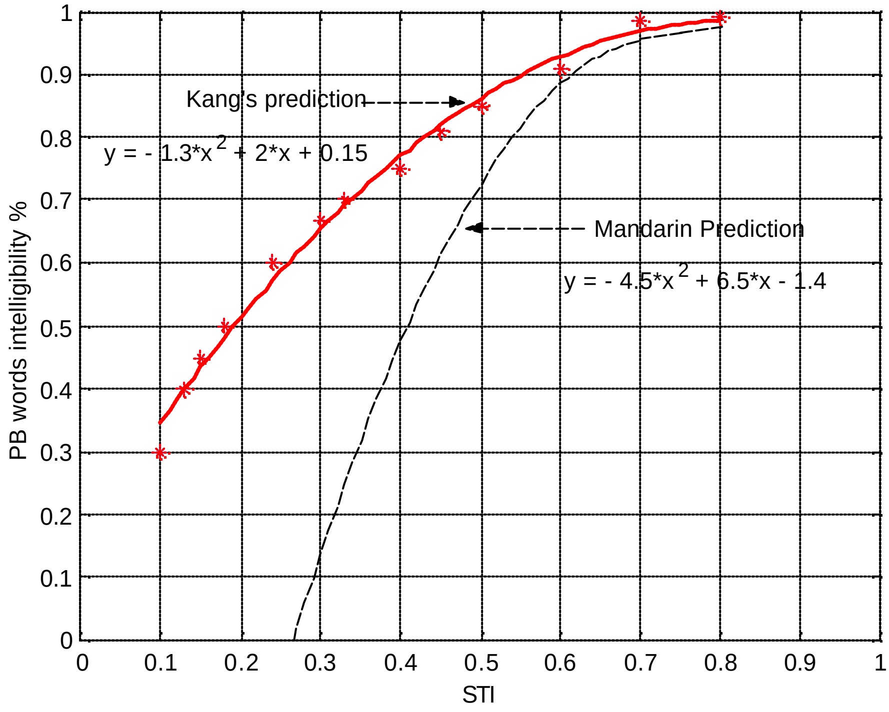

Fig 4.4 Comparison of the empirical equation for the intelligibility of Mandarin in Kang's results and the experimental results.

It is particularly interesting to observe that the rate of increase in the intelligibility of Mandarin between an STI of 0.3 and 0.6 is greater than the rate increase in the intelligibility of English. This implies that for the participants of this study to understand 50 percent of the English speech an STI of at least 0.42 would be required, whereas for Mandarin only an STI of 0.4 would be required. This shows that there is a difference in the intelligibility of the mother language and the second language under the acoustical conditions in this study.

Figure 4.4 also shows a plot of the PW word scores for Mandarin against STI. Again, Kang's experimental results for native speakers of Mandarin are shown as a comparison with our experimental results for typical listeners in Hong Kong who use Mandarin as a second language. The experimental results suggest that the intelligibility scores approach constant levels (equal to 100 percent for native speakers but about 90 percent for typical students in Hong Kong) when the STI is about 0.7 and above. The higher intelligibility scores for native speakers of Mandarin are understandable, and show that the minimum STI should be set at 0.7 to achieve excellent intelligibility in classrooms for these listeners. However, for the non-native speakers of Mandarin, the "saturation" level was only 90 percent. This is largely due to the limited ability of the students to master Mandarin. Due to this language barrier, the non-native speakers were able to perceive the speech but were unable to understand the meaning of some of the words, even when the STI was well over 0.7.

Among the non-native speakers, the speech intelligibility of Mandarin was around ± 0.3 when the STI was in the range of 0.2 to 0.4, and increased as the STI increased. This implies that the intelligibility of Mandarin can be increased in certain conditions through a better STI, even for secondary language listeners who are able to hear certain phonetically balanced words in Mandarin before saturation occurs. The native Mandarin speakers were capable of achieving a perfect SI score once the STI reached 0.8, but for the non-native speakers saturation occurred at 0.88 to 0.89.

The intelligibility of the Mandarin speech for the non-native speaking Hong Kong students appeared to be less satisfactory than the intelligibility of English under the conditions of less "clarity" but a sufficient AL_{cons}. It could be argued that the number of selected students was not sufficient to be representative of the majority in this experiment. However, all of the tests were taken in the same circumstances with the same environmental conditions and the same sample of average students.

Further confirmation of the results is given in Figure 4.5, which compares the experimental results with those of a similar articulation experiment that was conducted by Bradley [13-15]. Bradley measured the speech intelligibility levels of English among native speakers in ten occupied classrooms. His results show the interrelationship between the octave-band measurements of the useful-detrimental ratios, which he calculated along with the STI.

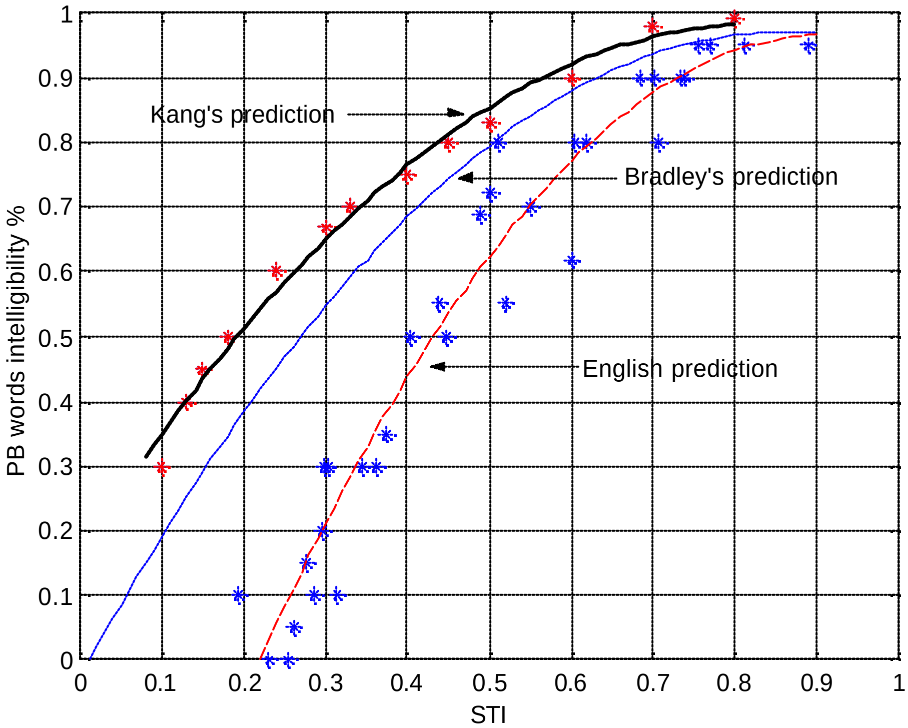

Fig 4.5 Comparison of the predictions of the speech intelligibility of English PB words of Kang, Bradley, and the experimental results of this study.

The plot shows that the findings of Kang and Bradley have a compatible intelligibility discrepancy of 0.5 to 1. The three separate measurements also approach 1 when the STI is increased from 0.7 to 0.95. This demonstrates the acceptability of the results for the intelligibility of English speech among non-native speakers in this study.

We next compare the intelligibility of the three languages among the Hong Kong university students independently. This is a major consideration in this study, and will provide a complete picture of the speech intelligibility of the three languages among trained and untrained speakers.

According to Figure 4.6, the intelligibility of the three languages among the students increased consistently from an STI of 0.2 until saturation occurred at an STI of approximately 0.7. It is clearly shown that the intelligibility of Cantonese was the best among the mother languages in the same environment. Given the increasing intelligibility of Cantonese between the STI values of 0.5 to 0.9, it can be concluded that the minimum STI requirement for Cantonese is only 0.35. This means that in the same environment, Cantonese speech should be understood by native speakers more than 50 percent of the time when the acoustical parameters of the listening and communication environment provide an STI of around 0.3 to 0.4.

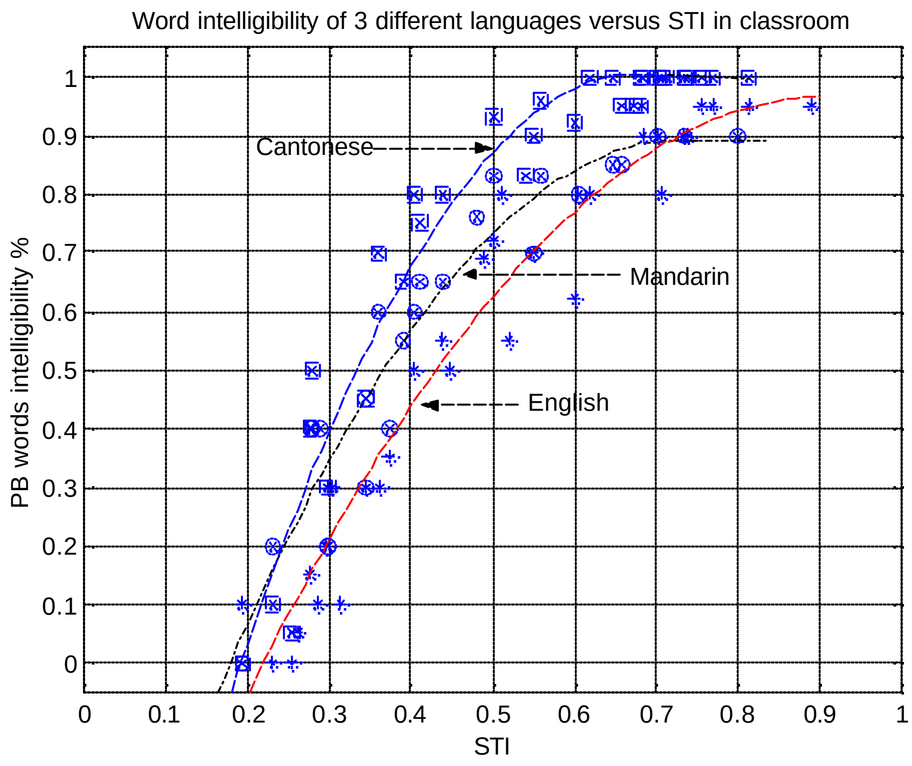

Fig 4.6 Speech intelligibility of the three languages among the participating students.

In contrast, although the intelligibility of Mandarin was better than that of English in the middle STI region from 0.3 to 0.7, it became saturated at an intelligibility index of 0.9 at an STI of 0.7. However, the intelligibility of English continued to increase up to an intelligibility index of 0.98 or 1 after the saturation boundary of Mandarin at 0.7 STI. The preliminary conclusion can thus be drawn that the intelligibility of the two second languages in this study varies in the middle STI region of 0.35 to 0.7 region and the upper region of 0.7 and above.

It is useful to investigate the intelligibility of the mother language (Cantonese) among the students. The intelligibility of Cantonese became saturated at a lower index of approximately 0.65 STI. Once again, English was more intelligible than Mandarin in the higher STI range of 0.72 and above. Figure 4.6 provides the empirical equations for the speech intelligibility of the different languages among the students.

It is also of interest to compare the data from the trained listeners and the untrained listeners separately for each language to identify the differences between the two groups. Figure 4.7 shows the intelligibility of Cantonese and Mandarin among the trained listeners (teachers) and untrained listener (students). As expected, the trained listeners performed well in both tests, as they had a better understanding of the general pronunciation of words in both languages.

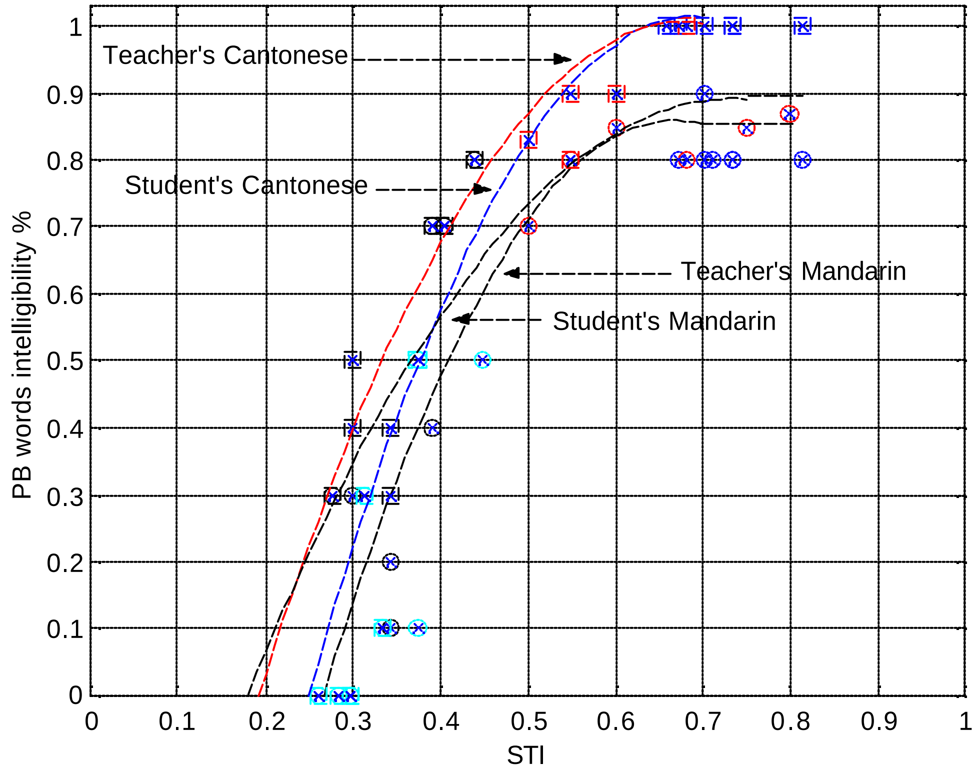

Fig 4.7 Intelligibility of Cantonese and Mandarin speech among the teachers and students.

Although their methods of prediction are different, the predictions of Kang and Bradley are similar, and the discrepancy error of approximately 0.05 in the different background acoustical parameters, such as EDT, LEF and C50, is acceptable. In Figure 4.8, the overall results of Kang's study are plotted against two sets of experimental results from this study. The first is the intelligibility of Mandarin among the trained listeners who were native speakers of Mandarin, and the other plots the intelligibility of Mandarin among the non-native speakers. This makes it easier to determine the difference between the two groups.

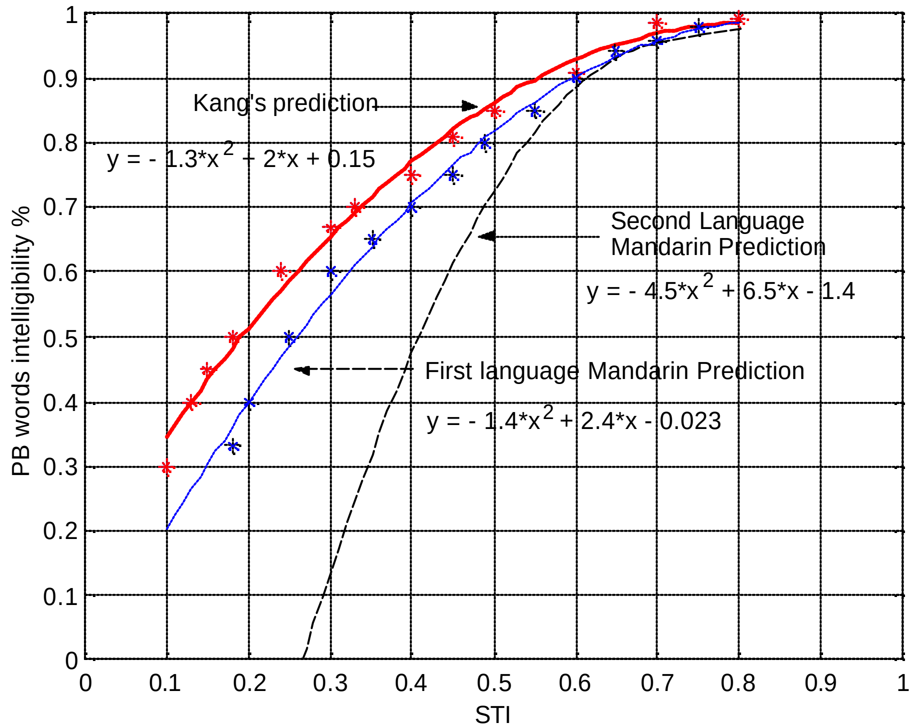

Fig 4.8 Empirical equation for the intelligibility of Mandarin speech among native and non-native speakers of Mandarin from the experimental results and the results of Kang.

Further analysis of the prediction methods for speech intelligibility shows that there is a certain discrepancy between the first and second languages as the STI increases. This discrepancy gradually narrows as the clarity or intelligibility of the consonants increases.

Although our results compare favorably with those from other studies, they would be more widely applicable if the experiment had been conducted among a greater number of listeners. It is suggested that at least 500 people take part in future tests to achieve more generalizable measurements.

4.5 Summary

Using the overall trends that are identified in Chapter 2, in this chapter we compare the speech intelligibility of the mother language (Cantonese) with that of the second languages that are most commonly used in university classrooms in Hong Kong (English and Mandarin) among native and non-native speakers.

This chapter also provides some empirical equations and intelligibility measurement methods using an articulation index, which will help acousticians to rationally analyze whether a university classroom is suitable for studying in multiple languages or not. The prediction methods of three previous studies and this study are evaluated.

The different effect of the languages can be seen for each category of speakers and listeners. Acceptable STI ranges are provided for each language as a reference for acousticians to objectively predict the speech intelligibility of these languages among trained and untrained listeners under specific environmental conditions.

Chapter5

Conclusions and suggestions of future work

In this chapter, not only are the experimental results and the theoretical predictions of the current study are summarized but further work and their implementations are suggested.

One of the main objectives for the present study is to identify the pertinent acoustical parameters for improving the speech intelligibility in classrooms. There are many earlier studies exploring different parameters such as the signal-to-noise ratio, the reverberant time, the articulation index for determining the acoustic environments of classrooms. In the present study, the Speech Transmission Index (STI) has been used to investigate the specific speech intelligibility at different positions in a classroom. Three languages - Mandarin, Cantonese and English have been used in the articulation tests. The interpretation of the characteristics of these languages has been described in Chapter 2. Each of these three languages has a specific combination of the vowels and consonants including the affricates, fricatives, nasals and plosives which may be confused easily in an environment with high background noise levels.

As mentioned in chapter 2, Hong Kong is one of the international cities where more than two languages have frequently been used in the verbal communication. Many earlier studies [17-18, 60] were to compare the speech intelligibility of English and Mandarin in an enclosed space. In Chapter 2, the overall trend of results have been shown, we examined the speech intelligibility of Cantonese, the first language used in Hong Kong. We also exploited the difference in the acoustic performance with the other two second languages, English and Mandarin.

In the current study, more than 2000 phonetically balanced words were selected randomly for the experiments to explore the speech intelligibility in classrooms. More than two hundred university students had been invited for the articulation tests for several times. In this circumstance, trained listeners provided valuable data for our experiments in the articulation tests. Moreover, the reverberation time, the early-decay-time, the signal-to-noise ratio and the other essential parameters have been combined in Chapter 3. The speech intelligibility of the first and second languages in classrooms has also been examined.

In the chapter 3, the speech intelligibility of different languages has been studied by applying the method of Maximum Length Sequence (MLS) to measure the Speech Transmission Index (STI), based on the statistic analysis, the speech intelligibility is predicted inside a classroom. More than 50 trained teachers, who used either Cantonese or Mandarin as their first language, were invited to participate in the articulation tests. More than twenty classrooms at the Hong Kong Polytechnic University were chosen to conduct the experiments.

Chapter 3 also described the articulation tests for Chinese language instructors at the Hong Kong Polytechnic University who were fluent both in spoken Cantonese and Mandarin. The speech intelligibility test materials were separated into the phonetically balanced word (PB) and public address statement (PA) analyses.

With the developed experimental techniques, the performance of the three different languages, Cantonese, English and Mandarin were analyzed. In Chapter 4, empirical formula was developed for the speech intelligibility of Cantonese that was dependent on such acoustical parameters as the useful-detrimental ratio, the clarity-50 (C50) and clarity-80 (C80) ratios.

In addition to the articulation tests of different languages, a comparison of the performance of different languages was also conducted in a typical classroom at the Hong Kong Polytechnic University. The comparisons provided a clear picture and guidelines for other acousticians, engineers and architects to consider the different acoustical parameters which affected the intelligible scores of different languages. In the analyses, the variations in STI and the different geometrical configurations of a university classroom are also considered

The main contribution of Chapter 4 is to implement the experimental techniques to explore the differences in the Speech Intelligibility scores for different languages for typical University students in Hong Kong. These students use Cantonese as their mother language and both English and Mandarin as their second languages.

The theoretical results and the experimental methods, which have been developed in the present study, can be extended in many ways. There are several suggestions for future related studies in the following paragraphs.

A list of over 2000 phonetically balance (PB) words were chosen in the present study. The words used in the intelligibility tests were chosen randomly in the list. However, different statements used in the Public Address (PA) System were commonly used in many public areas such as concourses and platforms of railway stations, hotel lobbies or the arrival and departure lounges of airports. The Investigation of the speech intelligibilities of different languages at these types of public areas remains to be one of the most challenging studies. For this study, it is required to develop a combination of complicated and complex PA statements. It is also expected that more related studies are required in this research area in order to develop an improved PA system that address the issues of different speech intelligibilities for different languages in this type of large indoor areas.

It is well-known that the shapes and the acoustic environments of a room affect the speech intelligibility. Different types of rooms may be used for different purposes. For instance, a classroom is used for studying, a meeting room is used for the communication of ideas and for negotiation, an office is used for the personal conversation, and the auditorium is used for drama and concerts. These different indoor environments have different acoustical parameters and intelligibility requirements. The empirical formulas developed in the present study may be used as a basis for optimal designs for different rooms. A hybrid method should be explored such that the use of computing software can be combined with the numerical model developed in the present study. This will provide a practical method of improving the speech intelligibility at different indoor environments.

In the present thesis, empirical formulas have been developed to estimate the speech intelligibility of different languages for a typical classroom. However, it will be most useful if the acoustic environments of the chosen classrooms can be controlled more closely. This can be achieved by using different absorbing materials and indoor acoustic diffusers for improving the reverberation time and Speech Transmission Index of the room. In addition, a more comprehensive list of PB words and many trained listeners should be used in future studies. With the improved control of the acoustic environments and the subjects for the articulation tests, a deeper understanding of the effect of the acoustic environments can be developed that help to design an indoor environment with improved speech intelligibility.

Appendices (Contributed papers):

Contribution arising from my research studies as shown below

1. M.L. Cheung, K.M. Li and C.L. Chan, “A computational approach for assessing the acoustic performance of barriers in a long industrial space.” *The 32nd International Congress and Exposition on Noise Control Engineering Inter-Noise Proceeding*, N822, pp 137 (2003).

2. Coriolanus C.L. Lam, Kai Ming Li, and Stanley M.L. Cheung “Articulation tests of the first and the second languages in classrooms,” *Journal of the Acoustical Society of America*, 116, p.2611 (2004).

References:

[1] "Acoustics – Measurement of sound absorption in a reverberation room." *BS EN* 20354:1993, ISO 354: (1985).

[2] Allan D. Pierce, "Diffraction of sound around corners and over wide barrier." *Journal of the Acoustical Society of America*, Vol. 55, pp.941-55 (1974).

[3] "American National Standard Method for the Calculation of the Speech Intelligibility Index", *ANSI* S3.5-1969 (1986).

[4] "American National Standard Method for the Calculation of the Speech Intelligibility Index", *ANSI* S3.79-1998 (1998).

[5] "American National Standard Method for the intelligibility of Speech Over Communication Systems", *ANSI* S3.2-1989 (1989).

[6] Arau-Puchades H., "An improved reverberation formula." *Acustica*, Vol.65, pp.163-79 (1988).

[7] Bart de Boer and Patricia K. Kuhl, "Investigating the role of infact-directed speech with a computer model." *Acoustics Research Letters Online, 2003 Acoustical Society of America*, Vol. 4 (4), pp.129-134 (2003).

[8] Ben Gold, Nelson Morgan, *Speech and audio signal processing — processing and perception of speech and music,* John Wiley & Sons, Inc., (2000)

[9] Bridget M. Shield and Julie E. Dockrell, "The Effects of Noise on Children at

School: A Review." *Journal of Building Acoustics,* Vol.10, (2), pp.97-116. (2003).

[10] "British Standard Method for Measurement of sound absorption in a reverberation room." *BS* 3638 (1987).

[11] "British Standard for the Objective rating of speech intelligibility by speech transmission index." *BSI BS EN* 60268-16, 1998 (1998).

[12] *Classroom acoustic – a resource for creating learning environments with desirable listening condition,* Architectural of the acoustical society of America, (2000)

[13] C.Wang and J.S. Bradley, "A mathematical model for a single screen barrier in open-plan offices." *Journal of Applied Acoustics,* Vol. 63, pp.849-66 (2002).

[14] C.Wang and J.S. Bradley, "Prediction of the speech intelligibility index behind a single screen in open-plan offices." *Journal of Applied Acoustics,* Vol. 63, pp.867-83 (2002).

[15] C.Wang and J.S. Bradley, "Sound propagation between two adjacent rectangular workstations in an open-plan office – part I: mathematical modeling." *Journal of Applied Acoustics,* Vol. 63, pp.1335-52 (2002).

[16] Dance S, Shield B. "Modelling of sound fields in enclosed spaces with absorbent room surfaces. Part I: Performances spaces." *Applied Acoustics* Vol. 58(1), pp.1-18 (1999).

[17] Dance S, Shield B. "Modelling of sound fields in enclosed spaces with absorbent room surfaces. Part II. Absorptive panels." *Applied Acoustics* Vol.61, pp.73-3847(2000).

[18] Dance S, Shield B. "Modelling of sound fields in enclosed spaces with absorbent room surfaces. Part III. Barriers." *Applied Acoustics*, Vol.61, pp.385-397 (2000).

[19] D.Y. Maa and H. Shen, Handbook of Acoustics, Science Press Beijing (in Chinese), (1987)

[20] D. Alarcão and J.L. Bento Coelho, "Lambertian Enclosures – A First Step Towards Fast Room Acoustics Simulation." *Journal of Building Acoustics*, Vol.10, (1), pp.33-54. (2003).

[21] Eswin M-L Yiu, Bruce Murdoch, Kathryn Hird and Polly Lau, "Perception of synthesized voice quality in connected speech by Cantonese speakers." *Journal of the Acoustical Society of America*, Vol. 112, (3) (2002).

[22] Eyring CF, "Reverberation time in 'dead' rooms." *Journal of the Acoustical Society of America*, Vol. 1, pp125-136 (1930).

[23] F.P. Mechel, *Formulas of acoustics*, Springer, (2002)

[24] Gao Ge, Li Ming and Hu Ruiming, "Research on pitch estimation algorithm." *Acta Acustica (in Chinese)*, Vol.28, (6), pp.540-544. (2003).

[25] H. A. Schenck, "A physical method for measuring speech-transmission quality."

Journal of the Acoustical Society of America, Vol. 44, pp.41 -58 (1968).

[26] Hammad R., "Simulation of noise distribution in rectangular rooms by means of computer modeling techniques.", *Journal of Applied Acoustics,* Vol. 24, pp.211-28 (1988).

[27] Henrich Kutruff, *Room Acoustics,* Elsevier Science Publishers Ltd., fourth edition, (2000)

[28] H.J.M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality." *Journal of the Acoustical Society of America,* Vol. 67, (1) (1980).

[29] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small room acoustics." *Journal of the Acoustical Society of America,* Vol.65, pp.943-950 (1979).

[30] Ji-qing Wang, "Diffusion and Auditorium Acoustics." *Journal of Building Acoustics,* Vol.10, (3), pp.211-219. (2003).

[31] J. Kang, "Comparison of speech intelligibility between English and Chinese." *Journal of the Acoustical Society of America,* Vol.103, (2) February 1998, pp.1213-1216 (1998).

[32] J. Kang, "Numerical modeling of the speech intelligibility in dining spaces." *Journal of Applied Acoustics,* Vol. 63, pp.1315-1333 (2002).

[33] J.S. Bradley, "Predictors of speech intelligibility in rooms." *Journal of the Acoustical Society of America,* Vol. 80, pp.837-845 (1986).

[34] J.S. Bradley, "Speech intelligibility studies in classrooms", *Journal of the Acoustical Society of America,* Vol. 86, pp.846-854 (1986).

[35] J.S. Bradley and R. Reich, "Optimizing Classroom Acoustics using Computer model studies." *Canadian Acoustics,* Vol. 26 (4), pp.15-21, (1998).

[36] J.S. Bradley and S.R. Bistafa, "Relating speech intelligibility to useful-to-detrimental sound ratios (L)", *Journal of the Acoustical Society of America,* Vol. 112, pp.27-29 (2002).

[37] Kohei Yamamoto and Koichi Takagi, "Expressions of Maekawa's chart for computation." *Journal of Applied Acoustics,* Vol. 37, pp.75-82 (1992).

[38] Kotarbinska E., "How to calculate the efficiency of an acoustic barrier in a flat room." *Journal of Applied Acoustics,* Vol. 23, pp.99-108 (1988).

[39] Kurze UJ. , "Scattering of sound in industrial spaces." *Journal of Sound and Vibration,* Vol. 98, pp.349-64 (1985).

[40] Lening Yang, *Computer Modelling of Speech Intelligibility in underground stations,* the thesis for the degree of PHD in the South Bank University, London, (1997)

[41] Lening Yang and Bridget M. Shield, "The prediction of speech intelligibility in

underground stations of rectangular cross section." *Journal of the Acoustical Society of America*, Vol.109, (1) (2001).

[42] Lindqvist E., "Sound attenuation in factory spaces." *Acustica*, Vol.50, pp.313-28 (1982).

[43] London: The Stationery Office, "Guidelines for Environmental Design in Schools. (Revision of Design Note 17)" *Building Bulletin 87*.

[44] Mark John Howarth, *The Application of Advanced Computer Models to the prediction of Sound in Enclosed spaces*, the thesis for the degree of PHD in the Department of Acoustics and Audio Engineering at the University of Salford, (1998)

[45] M.H.A. Gomes, S. N.Y. Gerges, R.A. Tenenbaum, "On the Accuracy if the Assessment of Room Acoustics Parameters: Using MLS Technique and Numerical Simulation." *Acustica*, Vol.86, pp.891-895 (2000).

[46] Millington G. "A modified formula for reverberation." *Journal of the Acoustical Society of America*, Vol.4, pp.69-81 (1932).

[47] M.L. Cheung, K.M. Li and C.L. Chan, "A computational approach for assessing the acoustic performance of barriers in a long industrial space." *The 32nd International Congress and Exposition on Noise Control Engineering Inter-Noise Proceeding*, N822, pp 137 (2003).

[48] Murray Hodgson, "On the prediction of sound fields in empty room." *Journal of the Acoustical Society of America,* Vol. 84, pp.253-61 (1988).

[49] Murray Hodgson, "Case History: Factory Noise Prediction using Ray Tracing – Experimental validation and the Effectiveness of Noise Control Measures." *Noise Control Engineering Journal ,* Vol. 33, pp.97-104 (1989).

[50] Murray Hodgson, "Evidence of diffuse surface reflections in rooms." *Journal of the Acoustical Society of America,* Vol. 89, pp.765-71 (1991).

[51] Murray Hodgson, "Experimental evaluation of the accuracy of the Sabine and Eyring theories in the case of non-low surface absorption." *Journal of the Acoustical Society of America,* Vol. 94, pp.835-40 (1993).

[52] Murray Hodgson, "UBC-Classroom Acoustical Survey." *Canadian Acoustics,* Vol. 22 (4), pp3-10, (1994).

[53] Murray Hodgson, "When is diffuse-field theory applicable?" *Journal of Applied Acoustics,* Vol. 49, pp.197-207 (1996).

[54] Murray Hodgson, "Scale-model evaluation of the effectiveness of novel absorber treatments for industrial noise control." *Institute of Noise Control Engineering,* Vol. 46 (2), pp.31-37 (1998).

[55] Murray Hodgson, Rod Rempel and Susan Kennedy, "Measurement and prediction of typical speech and background noise level in university classrooms during

lecture." *Journal of the Acoustical Society of America,* Vol.105, (1) (1999).

[56] Murray Hodgson, "Experimental investigation of the acoustical characteristics of university classrooms." *Journal of the Acoustical Society of America,* Vol. 106, (4) (1999).

[57] Murray Hodgson, "Empirical prediction of speech levels and reverberation in classrooms." *Journal of Building Acoustics,* Vol.8, (1), pp.1-14. (2001).

[58] Murray Hodgson, "Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms." *Journal of the Acoustical Society of America,* Vol. 111, (2) (2002).

[59] Murray Hodgson, "Rating, ranking and understanding acoustical quality in university classrooms." *Journal of the Acoustical Society of America,* Vol. 112, (2) (2002).

[60] M. Vorländer, "Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image source algorithm." *Journal of the Acoustical Society of America,* Vol. 86 (1989).

[61] Nicoleta Roman, Deliang Wang and Guy J. Brown, "Speech segregation based on sound localization." *Journal of the Acoustical Society of America,* Vol. 114 (4), pp.2236-2252 (2003).

[62] Ondet AM., Barbry JL"Modelling of sound propagation in fitted workshops using

ray tracing." *Journal of the Acoustical Society of America*, Vol. 85, pp.787-96 (1989).

[63] Rasmussen, K.B. "A note on the calculation of sound propagation over impedance jumps and screens." *Journal of Sound Vibration*, Vol.84, pp.598-602 (1982)

[64] *RAYNOISE User's Manual*, Version 2.1 LMS Numerical Technologies. (2001)

[65] Sabine WC, Collected papers on acoustics. Cambridge (MA): Harvard University Press, (1922).

[66] Schroeder MR. , "Modulation transfer functions: definition and measurement." *Acustica*, Vol. 49, pp.179-182 (1981).

[67] Schroeder MR. , "New method of measuring reverberation time." *Journal of the Acoustical Society of America*, Vol. 37, pp.35-72 (1990).

[68] Stephenson U., "Comparison of the mirror image source method and the sound particle simulation method." *Journal of Applied Acoustics*, Vol. 37, pp.35-72 (1990).

[69] Stephenson U., "Quantized pyramidal beam tracing — a new algorithm for room acoustics and noise emission prognosis." *Acustica*, Vol.82, pp.517-25 (1996).

[70] Tao Jianhua and Cai Lianhong, "Study of prosody model on Chinese speech synthesis based on the classification of syllabic prosody features." *Acta Acustica (in Chinese)*, Vol.28, (5), pp.358-402. (2003).

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

[72] T. Houtgast and H.J.M. Steeneken, "A physical method for measuring speech transmission quality." *Journal of the Acoustical Society of America,* Vol.67, pp.318-326 (1980).

[73] T. Houtgast, "The effect of ambient noise on speech intelligibility in classrooms." *Journal of applied acoustics,* Vol.14, pp.15-25 (1981).

[74] T. Houtgast, H.J.M. Steeneken, "A Multi-language Evaluation of the RASTI-method for Estimating Speech Intelligibility in Auditoria." *Acustica,* Vol.54, pp.186-99 (1984).

[75] T. Houtgast and H.J.M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria." *Journal of the Acoustical Society of America,* Vol. 77, pp.1069-77 (1985).

[76] T. Isei, T.F.W. Embleton and J.E. Piercy, "Noise reduction by barriers on finite impedance ground." *Journal of the Acoustical Society of America,* Vol. 67, pp.46-58 (1980).

[77] Trevor Cox and Peter D' Antonio, "Schroeder Diffusers: A Review[1]." *Journal of Building Acoustics,* Vol.10, (1), pp.1-32. (2003).

[78] U.J. Kurze and G.S. Anderson, "Sound attenuation by barriers." *Journal of Applied Acoustics,* Vol. 4, pp.35-53 (1988).

[79] Veimeir G., "Prediction of the sound field in industrial spaces." *Proceeding of Inter-Noise 92 Toronto, Canada,* pp.727-30 (1992).

[80] Wang Yunjia, Chu Min, He Lin and Feng Yongqiang, "The perception of disyllabic word stress of Chinese speech in utterance." *Acta Acustica (in Chinese),* Vol.28, (6), pp534-539. (2003).

[81] Wieland Weise, "Measurement Uncertainties for Sound Field Levels in Rooms." *Journal of Building Acoustics,* Vol.10, (4), pp.281-287. (2003).

[82] W. James Hadden, Jr. and Allan D. Pierce, "Sound diffraction around screens and wedges for arbitrary point source locations." *Journal of the Acoustical Society of America,* Vol. 69, pp.1266-76 (1981).

[83] Xi Xin, Zhang Wei, Zou Ling, Hong Mengdi and Shi Lei, "Real-ear measurement of ear canal acoustic resonance on Chinese children." *Acta Acustica (in Chinese),* Vol.28, (5), pp.453-456. (2003).

[84] Y.W. Lam, D.C. Hodgson, "The prediction of the singing noise of a drop hammer in a rectangular enclosure." *Journal of the Acoustical Society of America,* Vol. 93, (2) (1993).

[85] Y.W. Lam, D.C. Hodgson, "The prediction of the sound field due to an arbitrary

vibrating body in a rectangular enclosure." *Journal of the Acoustical Society of America*, Vol. 88, (4) (1999).

[86] Yap Sau Hee, "Classroom acoustics in Singapore." *The 8th International Congress on Sound and Vibration 2-6 July 2001, Hong Kong, China,* pp.887-894 (2001).

[87] Z. Maekawa, "Noise reduction by screens." *Mem. Fac. Eng., Kobe Univ.,* Vol. 11, pp.157-73 (1988).

The End