INSTITUTE OF SOUND AND VIBRATION RESEARCH
FACULTY OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF SOUTHAMPTON

AUDITORIUM SPEECH QUALITY: EFFECT OF DELAYED SAME SPEECH MASKERS

By

Krissada Vivatvakin

A thesis submitted for the degree of

Doctor of Philosophy

September 2007

UNIVERSITY OF SOUTHAMPTON

ACKNOWLEDGEMENTS

My debts of gratitude for assistance in the execution of this thesis are multitudinous. In preparing the thesis, I have been assisted and supported by many dedicated and supportive people. Although I cannot possibly thank all of them here, I wish to express my appreciation to those whose help has been most valuable.

First of all, I would like to express my gratitude to my supervisor, Dr. Ian Flindell who has broadened my knowledge and helped me shape my research interests. His overwhelming support, guidance, considerable patience, and comments have enabled me to produce such a thesis, which could not have been produced without his tremendous assistance.

Secondly, many suggestions form Professor Mark Lutman to improve my work are gratefully acknowledged. Their broad knowledge and openness to discussion was very influential to me and to this thesis.

Thirdly, I am deeply indebted to the Institute of Sound and Vibration Research which have granted me a scholarship to study this PhD at the University of Southampton for three years. I also would like to thank Tim Griffin, who has given the needed financial support during my time in Southampton.

Last, but not least, there were many people who encouraged me during my time of study. Mr. Athikom Thakulchaovalit, Dr. Weerachai Variyart, my friends, have a great influence on my decision to continue study after many years off. Finally, I would like to thank my parents for their patience and support throughout the many years of my studies.

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE

INSTITUTE OF SOUND AND VIBRATION RESEARCH

Doctor of Philosophy

AUDITORIUM SPEECH QUALITY: EFFECT OF DELAYED SAME SPEECH MASKER.

In standard practice, reverberation time is always used in acoustical room design, but the method can not always discriminate between rooms of different acoustical quality. Even though there is extensive literature on alternative methods, none of these alternatives have come into general use because all alternative methods are always developed based on subjective preference. To avoid the uncertainties associated with subjective preference, this study investigates the masking effect of a simulated delayed same-signal reflection on the direct sound. The main objective is to contribute to understanding the phenomenon of speech masked by a delayed same-speech reflection.

Three experiments were designed. In the first experiment, the results show that the masking thresholds obtained are affected by the level difference between the signal and the reflection, the time difference between the signal and the reflection, and the direction of the masker relative to the direct sound. The second experiment suggests that the speech signal can be detected in the silence gap as well as in the low amplitude consonant part of the speech. In addition, it also suggests that low frequencies, such as vowels, may not always mask high frequencies as consonants in the context of overlap masking within the range of parameters tested. In the last experiment, the direction of reflection and the effects of head rotation affect the masking threshold.

Following the three experiments, it can be concluded that the sound quality in any room may be assumed to be good when the delay is short. Unfortunately, this statement cannot be applied when sound reinforcement system is used. Interestingly, when the sound reinforcement system with multi-loudspeakers are used, the short delay may not always provide the good sound quality. For further development of objective parameter determining speech quality based on masking effect in a room some existing objective parameters such as Lateral Fraction (LF) and Late Lateral Level (GLL) might be importance because the direction of the reflection is significant. However, from this study, the quality of speech may depend on auditory scene analysis when the initial time between direct sound and reflection is long. That means there are many factors, such as fundamental frequency, continuation, and localisation, involved. Therefore, to develop a precise indicator of speech quality, a method of measuring speech quality might require a computer-based system, which can be complicated. It may not depend on the geometry of a room and its finishes alone.

CONTENTS OF THESIS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Background to Research

The main objective of this study is to contribute to understanding the phenomenon that speech can be masked by delayed same speech reflection. The reverberation time provides a relatively crude indication of the amount of delayed reflections but it does not always discriminate very well between rooms of quite different acoustical quality. Many researchers have attempted to devise superior indicators of acoustical quality, but none of these has yet been fully successful. This project has addressed the problem by studying the effect of delayed same signal masking under controlled laboratory test conditions. The findings suggest that the direction of reflection as compared to the direct sound can be significant. Therefore some parameters such as Lateral Fraction (LF) and Late Lateral Level (GLL) may be importance in determining the speech quality based on masking effect. Please note that the good sound quality based on masking effect in this study refers to the speech clarity as the speech signal can be detected in the present of the reflection.

Standard practice suggests that the reverberation time should be as short as possible to avoid both self masking and overlap masking of wanted speech (see for example; Maekawa and Lord, 1994). However the reverberation time should not be too short, as this sounds 'un-natural' to most listeners and also there is no reinforcement of wanted sounds. There is some evidence that listeners use existing reverberation to orientate themselves within any enclosed space and they obviously will not be able to do this if the room is too heavily damped. Un-amplified performers tend to rely on existing reverberation to be able to hear themselves and other performers clearly. The optimum amount of reverberation in any interior space is often a compromise between the avoidance of masking effects without making the room sound too un-natural.

The literature on reverberation time dates back to 1895. The President of Harvard University, Charles W Eliot, invited Professor Wallace Clement Sabine to investigate the poor acoustical properties of a large lecture hall (called the 'Hunt Hall') in the University. Sabine had been investigating the effect of acoustical reflections for some time and eventually developed the 'man in the box' method of measuring reverberation time. This method required an observer – who was enclosed in a non-absorbent wooden box to avoid undue interference with the existing internal acoustics of the room – to time the decay of reverberation from suddenly ceased organ pipe sounds until he could no longer hear it. The time to inaudibility was originally defined as the reverberation time. This was later defined as the time to decay by 60 dB when appropriate measuring instruments became available many years later (Egan, 1972).

One of the main advantages of the reverberation time as measure of interior acoustical quality is that it is possible to calculate it from easily measured physical properties of the room and of the internal surfaces (Maekawa and Lord, 1994). The main disadvantage is that it does not provide a complete description of interior acoustical quality. In 1971, Beranek (1986) provided recommended reverberation times for different uses based on case study evidence of subjective preferences in different types of rooms. Acoustical designers usually find Beranek's recommended reverberation times to be a good place to start, but acoustics problem can still arise even when the recommended reverberation times have been met. For example, Heringa, Melkemijer, and Pentz (1988) found that some halls with short reverberation time were judged satisfactory, whereas others with exactly the same reverberation time were not. Jeffe, Scarbough, Copper, and Clement (1999) found that in the Kennedy Centre concert hall, which was fully compliant with Beranek's recommendations, the entire stage needed to be rebuilt because musicians had difficulties hearing on stage. Davis and Davis (1987) reported that in halls with long reverberation time such as in the Froy Savings Bank, where the reverberation time was more than 3 seconds, normal conversational speech could still be heard clearly from the stage at the back of the upper balcony.

It can be seen that designing to the recommended reverberation time does not guarantee success (Barwick, 1994). The reverberation time does not provide any information on the detailed pattern of reflections or on the directions from which those reflections arrive at the listeners ears. Indeed, there is an extensive literature on alternative methods, additional to reverberation time. To date, a number of alternative indicators have been developed such as the EDT, C50, C80, IACC, etc. Many of these alternative indicators have been shown to have high statistical correlations with subjective descriptors such as fullness of tone, clarity, intimacy, spaciousness, timbre, development, ensemble, dynamic range, etc. Unfortunately, while many of these alternative methods have been shown to provide significant advantages under certain specialised conditions, none of these alternatives have come into general use. This might be because most investigations have been conducted by using direct judgment method with questionnaires or interviews that could be opened to bias. In most questionnaires, subjective descriptors such as dull, bright, dry, muddy, warm, etc may be interpreted differently by different individuals and this can contribute to misunderstanding and errors. For this research, possible uncertainties associated with subjective preferences were avoided by studying objective masking phenomena. .

Delayed same signal masking effects in a room were first reported in 1929 by Knudsen (1929). Most investigators have attempted to explain same signal masking effects in a room in terms of the difference between the relative magnitude of the wanted signal and the masker e.g. Bolt and Macdonald, 1949; Kurtović, 1975; Nábělek and Dagenais, 1986; Nábělek et al., 1989. This model of delayed same signal masking is directly analogous to background noise masking. For this research, it is assumed that there is no background noise masking, that the adverse effects of any intrusive sounds entering a room from outside have already been controlled out.

Delayed same signal masking can be further categorised as overlap masking and self masking. These masking effects are presented only with reverberant or echo. Overlap masking is similar to the forward masking. It occurs when the reverberant decay from a previous sound masks a following sound. Self masking is similar to the backward masking. It takes place over a much shorter time scale when reverberant decay from the onset of a sound contributes to ongoing masking of that same sound. If the duration between the direct sound and reflection is very short, the effect of self masking will be reduced.

The separate effects of self masking and overlap masking cannot be distinguished in real rooms because there is no technical possibility of separating out the two types of masker signal from any measurement or recording of real running speech or continuous music programme with reverberation. Previous investigators have attempted to control reverberant content by editing recorded sounds to remove the direct sound or by using various combinations of anechoic and reverberant rooms to record the original sounds (see for example; Burgtorf and Wagener, 1968; Koenig et al., 1977; Zurek et al., 2004). However, none of these methods have overcome the fundamental problem that there is no known method for separating out direct sound from reverberant sounds when they overlap in time.

For this research, to measure the masking threshold of direct sound when it was masked by the reverberant seems to be impossible; because the direct sound and reflections cannot be separated when they overlap in time. To overcome the problem, the study was conducted by using a method of simulation where a single delayed reflection can be added to the direct sound under conditions of complete experimental control. To measure the masking threshold of the speech signal in the present of the delayed same speech masker, the level of the masker was fixed and the level of the speech signal was measured. That means the delayed same speech masker from different direction was always louder than the direct speech signal. It can be seen that the condition in which the reflection is louder than the direct sound has almost never been the case in real auditoria. Therefore the testing conditions those were conducted in this experiment can only be applied to any room when sound reinforcement system is used.

To use a method of simulation where a single delayed reflection can be added to the direct sound, five main variables were identified as follows;

- level difference between the signal and the masker
- time difference between the signal and the masker
- direction of the masker relative to the direct sound
- frequency spectrum of the direct and the delayed sounds
- type of programme material

Firstly, the level difference between the signal and the masker is a direct measure of the signal to noise ratio.

Secondly, the time difference between the signal and the masker is a fundamental property of the acoustic space within which the reverberation occurs.

Thirdly, the direction of the masker relative to the direct sound may be important because it will be affected by the acoustic space within which the reverberation occurs

but it is not taken into account by any measure of reverberant decay which is not directionally sensitive.
Fourthly, the different materials inside the room may absorb and reflect sound differently across the frequency range, and this may have an effect on perceived sound quality.

Finally, it is well known that the type of programme has a material effect on subjective preferences in different kinds of acoustic spaces.

Four out of the five variables mentioned above were directly investigated in this study. They are the level difference between the signal and the masker, the time difference between the signal and the masker, the direction of the masker, and the type of programme material. The frequency spectrum was investigated indirectly through the type of programme material because the type of programme material contains frequency spectrum as time variable. To represent the frequency spectrum on its own, the signal will be a pure tone or tone combination that can hardly be found in real life. Therefore, in the previous research, for example the works of Burgtorf and Wagener (1968), Koenig et al., (1977), and Zurek et al. (2004), the type of programme such as speech has been used to represent a sound source as in real situation instead of tone. Following their foot step, the type of programme, which is speech, is investigated together with its frequency spectrum.

Three experiments were carried out to measure the different effects of the previously described variables on masked threshold levels of the direct sound. In each experiment, the masked threshold levels were measured by varying the level of the direct sound in the presence of a range of delayed masker signals covering the key variables of interest. The method was similar to staircase or method of up-down but it was conducted without reducing the step size. The potential difficulty of being able to distinguish between a wanted speech signal and a delayed version of the same signal was overcome by careful training of the listeners – see chapter 3 for further details of this aspect of the work. All experiments used the same pre-recorded and edited speech as the programme material.

The following four main variables were investigated in the first experiment; level difference, time difference; masker direction; and type of programme (male versus female speech). The first experiment used separate loudspeakers to represent the signal and the masker in an anechoic listening room so that the relative angle between the two could be varied according to the experimental design. Listener head movement was not restricted, as in natural listening.

The second experiment was carried out to test for the separate effects of the 'silence gap' between utterances in running speech under a range of frequency bandwidth conditions for both male and female speech. Headphone listening was used because it is more convenient than loudspeaker listening when the relative angle between the masker and the direct sound is not an issue. In addition, headphone listening automatically controls for any effect of head movement while listening.

The third and final experiment compared head restrained and head movement allowed conditions over a small range of relative angles between the masker and direct sound to test for the effect of head movement under narrow angle listening conditions, and with a fixed time delay.

1.2 Delimitation of scope

The three experiments were designed to investigate self and overlap masking effects when speech is masked by a delayed same signal reflection. The kinds of multi-reflections which occur in real life were avoided because of the need to keep the experimental designs as simple as possible. It is accepted that single reflection simulations have lower construct validity than multi-reflections simulations in this context, but on the other hand, it should also be noted that each doubling of the number of reflections considered more than doubles overall size of the experiment.

1.3 Outline of the thesis

The outline of this thesis contains five chapters and is shown in Figure 1.2 below.

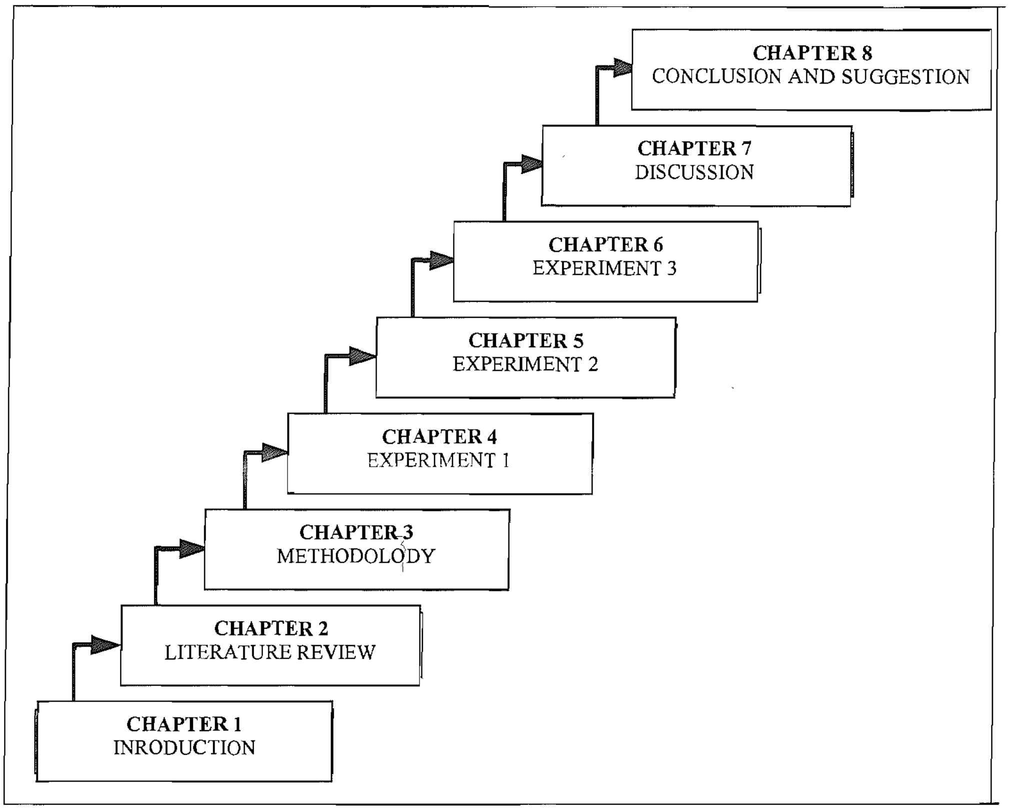


Figure 1.2: The outline of the Thesis' eight Chapters

As seen in Figure 1.2, Chapter 1 is the introduction of the thesis. This chapter provides the overview of the thesis. It describes general background, overall problems, delimitation of scope, outline of the thesis, and conclusion.

Chapter 2 refers to the literature review. This chapter begins with the auditorium acoustics, masking effect in the field of auditorium acoustics, and then it reviews the basic theories of masking effect, including the simultaneous masking and the non-simultaneous masking, together with the binaural masking release. The previous works concerning the masking effect on the speech masker under both headphones and free-field listening conditions are described. Research issues, including research problem, research questions, research objectives, and research hypotheses then are revealed.

Chapter 3 describes the research methodology. This chapter provides the justification for the methodology. It begins with the psychoacoustics method chosen and the method in determining the threshold. Then, the experimental design, including the experimental apparatus and the experimental procedures are shown. Also the data collection, the data analysis, and ethical considerations are depicted.

Chapter 4, Chapter 5, and Chapter 6 are the experiments. The masking of speech by delayed same speech reflection reinforcement is investigated in the first experiment. The second experiment is the investigation in the role of simultaneous and non-simultaneous masking on the speech signal when speech is masked by delayed same speech, under the headphone listening condition. Then the last experiment is the investigation on the role of head rotation on masking effect under the free-field listening condition. In each experiment, the introduction is described briefly, followed by the experimental procedure, and the results. These chapters also picture the analysis of the data for each experiment.

Chapter 7 is discussion. In this chapter, the results from those three experiments were discussed.

Chapter 8 draws the conclusion and recommendation of the thesis. In this chapter, the research hypotheses are concluded. Finally, the contributions of the thesis, limitations, and further research are discussed.

CHAPTER 2
LITERATURE REVIEW

This chapter begins by reviewing the development of descriptors for measuring sound quality in a room, followed by the broad problem caused by the direct judgment method. The masking effect is then introduced in this study. The theory of masking effect is then reviewed, including simultaneous masking, non simultaneous masking, binaural masking release, best ear, binaural masking level difference, spatial localisation, and effect of head rotation. Following the theory described, the current status of masking effect on speech is pointed out together with research gaps in the masking effect, which is the setting of this study. Finally the research issue is described.

2.1 Sound Quality

Many objective descriptors of sound quality that correlate with subjective impression have been developed in the past. These objective descriptors can mainly be separated into three groups based on their characteristics. The first group is the objective descriptors whose development is based on decay time in a room, known as *reverberation time* (RT). The second group is those developed by using the concept of *impulse response*. Finally, the objective descriptors were developed by using the concept of *binaural hearing* based on the direction of the reflections.

2.1.1 Reverberation time

The literature on reverberation time dates back to 1895. The President of Harvard University, Charles W Eliot, invited Professor Wallace Clement Sabine to investigate the poor acoustical properties of a large lecture hall (called the 'Hunt Hall') at the University. Sabine had been investigating the effects of acoustical reflections for some time and eventually developed the 'man in the box' method of measuring reverberation time. This method required an observer – who was enclosed in a non-absorbent wooden box to avoid undue interference with the existing internal acoustics of the room – to time the decay of reverberation of the sound from a suddenly stopped pipe organ until he could no longer hear it. The time to inaudibility was originally defined as the reverberation time, but was later defined as the time it takes a sound to decay by 60 dB when more sensitive measuring instruments became available many years later (Egan, 1972).

One of the main advantages of reverberation time as a measure of interior acoustical quality is that it is possible to calculate it from easily measured physical properties of the room and of the internal surfaces. The main disadvantage is that it does not provide a complete description of interior acoustical quality. Reverberation gives a rich sonority, meaning that it is preferable to have a fairly long reverberation time for the performance of music, while speech generally requires a shorter reverberation time. This is because too much reverberant sound tends to mask speech, making it less intelligible. Beranek (1986) noted that baroque music (1600-1750) tended to be played in a palace music room where reverberation time was typically between 1.6-2.0 seconds. The romantic period of music of the late nineteenth century and the early twentieth century was written for halls with longer reverberation times. Beranek also stated that Kuhl performed a series of experiments using artificially created reverberations and found that a reverberation time of approximately 1.5 seconds was perfect for classical music, while a reverberation time of 2 seconds was better suited for music of the romantic period. In addition, Beranek (1986) also provided recommended reverberation times for different uses based on case studies and evidence of subjective preferences in different types of rooms. Acoustical designers usually find Beranek's recommended reverberation times to be a good place to start, but acoustics problems can still arise even when the recommended reverberation times have been met. For example, Heringa, Melkemijer, and Pentz (1988) found that some halls with short reverberation time were judged satisfactory, whereas others with exactly the same reverberation time were not. Jeffe, Scarbough, Copper, and Clement (1999) found that in the Kennedy Centre concert hall, which was fully compliant with Beranek's recommendations, the entire stage needed to be rebuilt because musicians had difficulties hearing on stage. Davis and Davis (1987) reported that in halls with long reverberation time, such as in the Froy Savings Bank where the reverberation time was more than 3 seconds, normal conversational speech could still be heard clearly from the stage at the back of the upper balcony.

It can be seen that a superior, a good and a fair-to-good room can also have similar reverberation times, despite the dramatic difference in subjective preference (Beranek, 1999). Reverberation time does not provide any information on the detailed pattern of reflections or on the directions from which those reflections arrive at the listeners ears. Meanwhile, there exists extensive literature on alternative methods in addition to reverberation time. To date, a number of alternative indicators have been developed such as EDT, ITG, LF, GLL, IACC, etc.

2.1.2 Early Decay Time (EDT)

Early Decay Time (EDT) was developed when Schroeder and Sessler (1965) asked volunteers to subjectively compare linear (exponential) with non-linear decay of reverberation time, and discovered that there was a good correlation between the decay rates over the first 160 ms. Following these findings, Jordan (1969) suggests that since the sensation of reverberation seems to correlate well with the early slope of the decay curve, the time required for 10 dB decay multiplied by six might be more useful. Jordan named this objective descriptor as "Early Delay Time (EDT)". In addition to this early 10 dB decay, the time required for a 15 dB or 20 dB decay has also been proposed by other workers (Maekawa, 1994). Although early decay time is related to the subjective sensation of reverberation to a greater degree than reverberation time is, they are generally the same.

2.1.3. Initial Time-delay Gap (ITG)

Back in 1965, Schroeder was the first person who introduced the "integrated impulse" method. He demonstrated that the squared impulse response from time t to infinity equals the average sound intensity during decay. He has further shown that build up and decay of the reverberation is purely complementary with an integrated impulse response. Therefore, when the integrated impulse response is used, replication of reverberation time (RT) measurement is no longer necessary.

With the impulse response, Kulh reported in 1957 the importance of time delay between the direct sound and the first reflection (Meakawa, 1994). He noticed that the larger this Initial time-delay Gap, the larger spatial impression might be obtained, and suggested that an approximate 30 ms delay would be desirable, while Beranek stated that it would be better if it was less than 20 ms (Beranek, 1986).

Initial time-delay Gap seems to affect the sound quality in the room in some way, but the criteria has not yet been established. In addition, it is hardly used in design because it relates to only the room's size, not other factors.

2.1.4 Lateral Fraction (LF) and Late Lateral Level (GLL)

The lateral fraction a measurement based on the difference between the lateral sound energy in the first 80ms of an impulse response as measured with a figure-of-eight microphone, and the total sound pressure in the first 80 ms as measured with an omni directional microphone. The lateral fraction (LF) is determined for the time period of 0 to 80 milliseconds and is the average of the lateral fraction's in the four frequency bands, 125, 250, 500 and 1,000 Hz. It is equal to the ratio of the weighted energy in the sound that does not come from the direction of the source to that which comes from all directions including that of the source. Bradley & Soulodre (1995) indicated that the lateral fraction correlates with two spatial subjective effects. They are Source Broadening or Apparent Source Width (ASW) and Listener Envelopment (LEV). The effect previously known as "spatial impression" that is associated with early reflections is in fact related to source broadening and should be measured subjectively as apparent source width (ASW). For Listener Envelopment (LEV), there is evidence that the effect comes mostly from late lateral energy and late overall sound level. Bradley and Soulodre (1995) proposed the Late Lateral Level (GLL) as an objective measure for Listener Envelopment (LEV).

The Late Lateral Level (GLL) is a measure of the absolute strength of the lateral sound energy 80 ms and more after the direct sound. The listener envelopment (LEV) is found to be an important component of spatial perception that is distinct from the source broadening (ASW) predicted by Lateral Fraction. There is evidence that the listener envelopment (LEV) is strongly related to late energy, both overall and lateral. Unfortunately, the Late Lateral Level (GLL) is still not an internationally accepted standard measure.

2.1.5 Inter-Aural Cross Correlation (IACC)

Apart of the indicators mentioned previously, some indicators have been developed bases on binaural hearing in related to sound localisation. Keet (1968) suggested that the degree of the spatial effect was related to the short time cross-correlation between signals at the two ears. He discovered that the degree of the effect was a function of the level of the music. He was also the first to establish subjective quality as "apparent source width (ASW)". Instead of using the output of two microphones at the ear canals, he measured the cross correlation by using two microphones positioned with a directivity axes 90degree apart.

Following Keet's suggestion, Ando (1977) defined and adopted the "Inter-Aural Cross Correlation (IACC)" as one of the important factors making up subjective preferences.

Ando (1977) conducted subjective preference tests with a simulated single reflection in an anechoic room. The degree of preference in relation to a long auto correlation function (AFC) of sound and the Inter-Aural Cross Correlation was studied. The results indicated that the score was found to increase by decreasing the degree of IACC.

A few years later, Ando (1979) extended the 1977 signal reflection experiment to a field with multi-reflections. The preference test gave nearly the same results as the previous tests of sounds fields with single reflection.

Later in 1986, Ando established that IACC is one of four orthogonal acoustical parameters that explained the subjective preferences of the listeners in his experiments. The other three factors were listening level, the initial time delay gap and the reverberation time.

The IACC has also been applied to measure sound quality in the real hall by Hidaka and Beranek (1995). They reported that a (1-$IACC_{E3}$) of three octave bands (500Hz, 1 kHz, and 2 kHz) is a sensitive enough objective descriptor to represent apparent source width.

Since spatial impression was concerned, the investigations on the quality of sound in a room turned to binaural hearing conditions instead of monaural hearing conditions. However these objective descriptors have not been accepted as a good guidance in design because the investigations have been done in a complicated way and do not provide enough knowledge related to the structure of a room, especially its furnishings.

2.2 Difficulties in auditorium sound quality studying

As it has been mentioned previously over many decades, a number of indicators have been shown to have high statistical correlations with subjective preference. Unfortunately, while many of these alternative methods have been shown to provide significant advantages under certain specialised conditions, none of these alternatives have come into general use. One reason might be that most investigations have been conducted by using direct judgment methods with questionnaires or interviews that could be opened to bias.

2.2.1 Direct Judgement method

Direct judgement method is always used when studying the sound quality in a room. In most investigations, questionnaires and an interview after the performance are used to collect the data. These data will be analysed for a statistical correlation between subjective judgement and objective measurement. It can be seen that subjective judgement is an important parameter. However, the subjective judgement collected might be open to bias for two reasons:

First is short term memory. The direct judgement method is based on a subject's immediate impression of the performance in a room. Therefore, it is impossible to compare the sound quality between one room and the other because we cannot recall the exact elements of sound in the previous room visited (Moore, 1983). In addition, the quality of sound in a room also depends on the performance on the stage.

Second is that the questionnaires could contain ill-defined terminology. To collect data using the direct judgement method, the subject's task is normally to judge the sound quality of a room due to each subjective term provided in an interview or questionnaire. In most questionnaires, subjective descriptors, such as dull, bright, dry, muddy, warm, etc., are always used. Some of these subjective terms are ambiguous and may lead to misunderstandings and unfair ratings (Meakawa, 1994). In order to avoid these kinds of terminology, Ando (1977) used the term "subjective preference" instead of using specific term in questionnaires. However, an accurate judgement is still difficult because the subjective preference may very depending on the subject's parameters such as personal taste.

2.3 Researcher point of view

To date, many objective descriptors have been developed but there is yet no universal accepted better objective descriptor that correlates with overall subjective impression. This might be due to the subjective judgement method. Therefore, in this research, possible uncertainties associated with subjective preferences were avoided by studying objective masking phenomena.

2.4 The Masking effect

The masking effect presents one of the most basic phenomena of hearing. Moore (1982) states that masking is defined as "the process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound or the amount by which the threshold of audibility for one sound is raised by the presence of another (masking) sound (American Standard Association, 1960)".

The masking effect has been investigated for many decades. There are two types of masking effects – simultaneous masking and non-simultaneous masking.

2.4.1 Simultaneous masking

Simultaneous masking is the most basic masking phenomenon. This is masking in which the signal and masker are presented at the same time (Moore, 1982). Therefore this masking effect is a not time domain phenomenon but it rather a frequency domain phenomenon. That means a signal with a low level of sound pressure can be made inaudible by a simultaneously occurring masker with a higher level of sound pressure. Gelfand (1998) reported that the earliest work of simultaneous masking was done by Mayer in 1894. Mayer found low frequency tones effectively mask higher frequencies, whereas the high frequencies are not good maskers of low frequencies. These results were later demonstrated by Wegel and Lane (1924). They investigated the masking threshold of various frequency tone signals in the fixed frequency masker. However, their results had an error caused by a beat when the signal and masker were too close in frequency, allowing for detection of the signal. In order to avoid the beat in the later experiment, a narrow-band noise was used for either signal or masker instead of a tone.

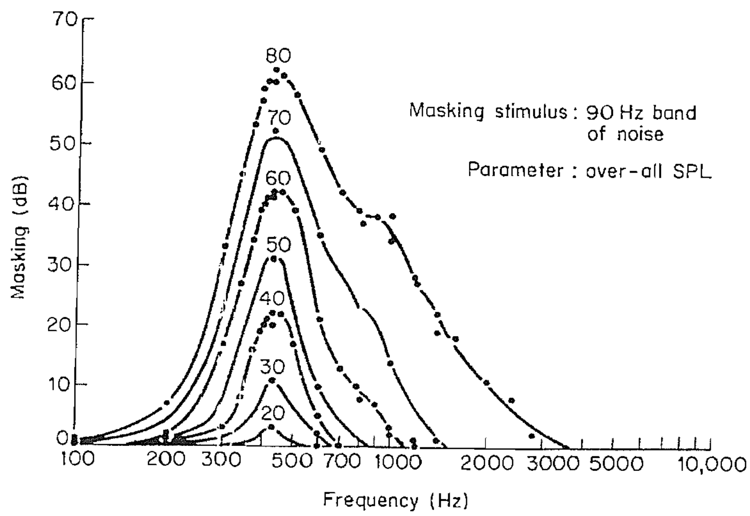


Figure 2.1: Threshold in quiet and masking threshold from Egan and Hake (1950). (Moore B.C.J, 1983)

In 1950, Egan and Hake published a masked audiogram for a narrow band of noise centred at 410Hz as shown in Figure 2.1. The audiogram showed an asymmetric increase in the masking threshold level, as previously published by Mayer and Wegel and Lane (Gelfend, 1998). It showed steep slopes on the low frequency side, with a gentler curve on the high frequency side. On high frequency side, the slopes of the curves tend to become shallower at high levels. That meant that the masking grows non-linearly when the masker level is increased. Their results also showed that the frequency of a signal would have a masking effect on the adjacent frequency component. This was also applied to speech. As the level of speech was increased, the low frequency became more and more effective in masking the higher frequency component and resulted in a loss of audibility of the important information carried in the mid and high frequency component (Moore, 1982).

Apart from the characteristics of simultaneous masking as mentioned above, it was found that masking had more of an effect if the signal and masker were close enough to each other in frequency (Hamilton, 1957). Moore's (1997) report on the work of Fletcher (1940) found that increases in noise bandwidth beyond a certain critical value have little effect on the threshold of a tone. The critical value that was mentioned by Fletcher (1940) had been referred to as the concept of a critical band.

The critical band concept was defined by Fletcher in 1940. He introduced the concept of an auditory filter in order to explain the phenomenon of masking. He suggested that the peripheral auditory system behaves as if it contained a bank of band pass filters. When a listener is trying to detect a signal in a noise, one will attend to the filter with a centre frequency close to that of the signal. Only the components in the noise that pass through the filter will have an effect on masking the signal. In other words, the components of noise filtered outside the filter band do not contribute to masking (Fletcher, 1940). That means man has the ability to separate an element of complex sound into frequency ranges due to the frequency-resolving power of the basilar membrane (Moore, 1982). Fletcher (1940) also suggested that different frequencies produce their maximal effects at different locations along the basilar membrane, so that each location responds only to a limited range of frequencies.

According to this concept, Fletcher (1940) conducted an experiment to measure the masked threshold for a tone produced by various bandwidths of noise centred on the test tones. He found that the masked threshold of tone increased as the bandwidth of the masking noise was widened. However, once the noise band reached a certain bandwidth, further widening of the band did not result in any more masking of the tone. Thus he demonstrated that only a certain width of bandwidth contributes to the masking of a tone at the centre of a band, and called this certain width of bandwidth "critical bandwidth". Later on, the same findings were repeatedly confirmed by Hamilton (1957), and Boer (1966).

Consequently, it seems as though sounds and frequencies that are separated by critical bandwidth will not interact. But in fact, one tone may mask another over frequency separations greater than the critical band, especially when the masker is low frequency and the signal is high frequency. In addition, critical bands have been reported in a range of frequencies with upper and lower limits (Scharf, 1970), But Moore (1982) reports that the critical bands are continuous rather than discrete, as there is no evidence for any discontinuity. He stated that the critical band can be approximately any centre frequency in the audible range which we choose.

In summary, simultaneous masking is a basic phenomenon wherein the signal threshold is raised in the presence of a masker. This phenomenon can easily be explained by using the concept of critical bands. This means that the components of a speech masker are effective in masking the signal when the signal and the component of a masker have their frequency components close together. Also, when the overall level of the masker is increased, the upward spread masking can be observed.

2.4.2 Non-Simultaneous masking

In addition to simultaneous masking, non-simultaneous masking also plays an important role in human auditory perception. Non-simultaneous masking is a term describing masking situations where the signal and the masker do not overlap in time as shown in Figure 2.2. The signal is presented, and then the masker is presented after a brief time delay following the signal offset. This masking is called "backward masking" because the masking effect occurs backward in time. On the other hand, the masker is presented first and then the signal. Masking that occurs forward in time is called "forward masking".

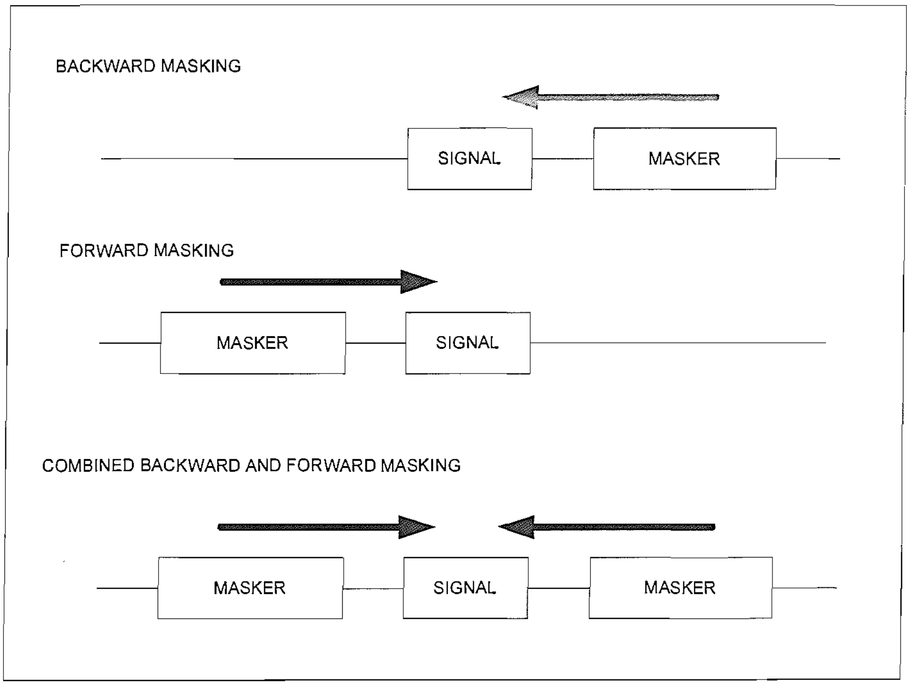


Figure 2.2: Non-simultaneous masking diagrams (Gelfand, 1998)

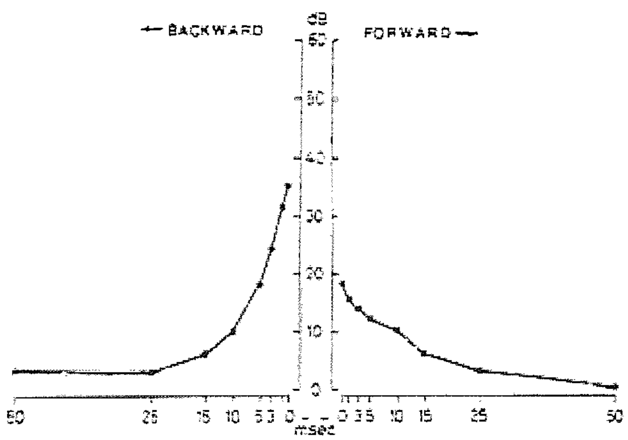


Figure 2.3: Characteristics of non-simultaneous masking: timing effects. Modified from Gelfand (1998)

The non-simultaneous masking, including forward and backward masking was first studied by Pickett (1959), and followed by the work of Elliott (1962). They conducted the experiments following the schemes in Figure 2.2. The masking threshold of the signal is determined under various conditions of both signal and masker, including time interval between signal and masker, frequency, masker level and masker duration. One of the results taken from Elliott (1962) is shown in Figure 2.3. This Figure is a typical pattern for non-simultaneous masking that shows the effect of non-simultaneous masking relative to the time interval between the signal and masker. The main features of non-simultaneous masking have been summarised by Moore (1982) as follows:

1. The nearer the time interval between the masker and the signal, the more the threshold elevated.
2. More backward masking than forward masking occurs for very short time intervals between the signal and the masker.
3. Most backward and forward masking occurs within about ±100 ms of the masker onset or termination.

4. Non-simultaneous masking is more pronounced when the signal and masker are presented to the same ear, however, it can be observed under dichotic listening condition.
5. The amount of masking increases with increasing masker duration between 1-20 ms, but the masking threshold is independent of masker duration when the masker duration is more than 20 ms.
6. Non-simultaneous masking has more effect if the frequency of the signal and masker are in the same critical bands (just as in simultaneous masking).
7. Increments in masker intensity do not produce corresponding increments in the amount of forward and backward masking.
8. The additional backward and forward masking can occur when a probe signal is presented in the gap between two maskers.

As described previously, non-simultaneous masking depends on the interval time between the signal and masker, the frequency relationship, and level. Apart from these parameters, the size of non-simultaneous masking also depends on the duration of the masker. It has been observed that more forward masking is produced by a masker 200 ms in duration than by a 25 ms masker. However, this is not a case for backward masking (Elliott, 1971).

As far as forward and backward masking is concerned, its basis is not fully clear. Moore (1982) reported that forward masking could be explained in terms of persistence in the pattern of neural activity evoked by the masker. But backward masking is difficult to explain because it is the interference of a later masker with a signal that is already completed. In addition, backward masking occurs over time intervals that are longer than those required for the decay of a travelling wave pattern along the basilar membrane.

In order to understand the basis of forward and backward masking, the basic curve of forward and backward masking in Figure 2.3 is carefully investigated. This curve shows that the amount of masking decreases dramatically as a function of the time gap between signal and masker as it increases from 0 to 15 ms, while the amount of backward masking decreases very little as the interval is lengthened further. The similar pattern is also found for the forward masking curve, but the reduction is of a lesser degree. Thus, there should be two active mechanisms (Duifhuis, 1973). Firstly, it would suggest that at short intervals, approximately 15-20 ms, and the dominant component in both forward and backward masking results from a temporal overlap in cochlear response (Duifhuis, 1971). In other words, the vibrations evoked by the first signal on the basilar membrane have not completely died down before the second signal arrives. Secondly, in the long interval, the neural events are responsible for the shallower non-simultaneous masking curve.

So far the empirical phenomena of forward and backward masking are well known, even though the masking functions obtained depend on the extent of the parameters used in the specific experiment. However, the effect of forward and backward masking seen in an experiment is different from everyday situations, because neither forward nor backward masking occurs alone in daily situations. They always occur together.

2.4.2.1 Additive Forward and Backward masking

The combined forward and backward masking is a result that is obtained by placing the signal between the two maskers. This phenomenon was studied by Pollack in 1964 (Moore, 1982) and Elliott (1971). Both studies show that forward and backward masking cumulate under some circumstances. Unfortunately they did not concurrently obtain separate function of forward and backward masking, so it is not possible to judge what addition the two types of masking exhibited. Later, Wilson and Carhart (1971) conducted another experiment on the interactions and addition of forward and backward masking. Their experiment examined the effect of backward and forward masking alone and then examined the effect of combined forward and backward masking. From their results, as shown in Figure 2.4, it shows that more masking occurs when backward and forward masking are combined than would result for the individual contribution if backward and forward masking were simply added together (Wilson and Carhart, 1971).

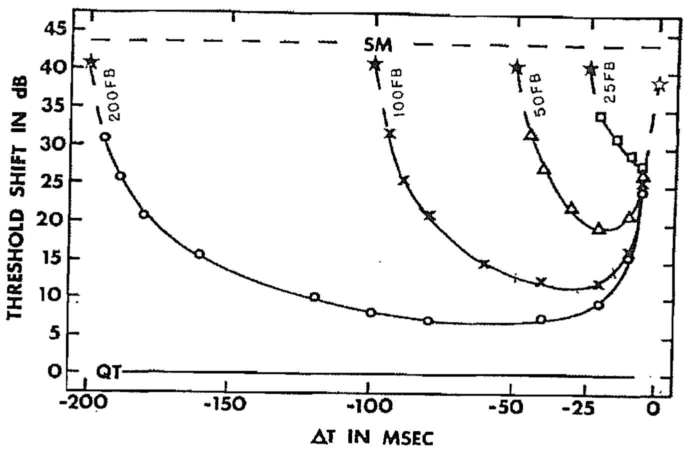


Figure 2.4: Addition of backward and forward masking effect. (Wilson and Carhart, 1971).

From Figure 2.4, the masking threshold of combined forward and backward masking is affected by gap duration between those maskers. The bigger gap duration provides a greater release from masking. The masked threshold is even closed to absolute threshold at one point when the gap duration is larger than 300 ms. However, this effect of gap duration did not affect the masking threshold level when gap between the signal and the masker was small, as the masked threshold was only influenced by the backward or forward masking effect.

Because of the masking effect mentioned above, some sounds in our daily lives may disappear in the presence of the other sounds. Luckily, humans can still hear the sound because masking can be released under certain conditions, such as the phenomenon known as "comodulation masking release (CMR)". In addition, the masking can also be released when a person listens binaurally or under spatial hearing conditions. In spatial hearing, masking effects can be released in some amount because of various phenomena, including the best ear signal to noise ratio, binaural masking level difference (BMLD), spatial localisation, and motion theory.

2.4.3 Comodulation Masking Release (CMR)

As has been described previously, masking can be released under some conditions when an off-frequency band is added to the masker. This masking release is known as "Comodulation Masking Release (CMR)".

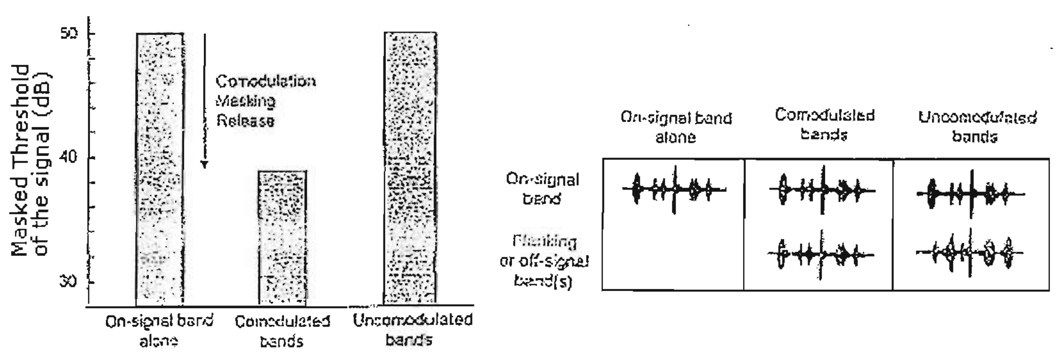


Figure 2.5: Comodulation Masking Release (Hall, Haggard, and Fernandes, 1984)

As it has been mentioned previously, only a certain critical bandwidth of noise around a signal tone is involved in the masking of that tone. The masked threshold of signal will not be changed by widening the noise bandwidth beyond the critical bandwidth or adding one or more other bands outside of the critical band. However, a different situation occurs when the masking noise is amplitude modulated, as illustrated in Figure 2.5 (Hall, Haggard, and Fernandes, 1984). Assume that the noise band centred on the test tone to be called the on-signal band and for any other bands of noise to be called off-signal band. Firstly, it shall begin by masking a pure tone signal by an on-signal band of noise that is being amplitude modulated ("on-signal band alone" in Figure 2.5). The graph on the left shows that the masked threshold is 50dB in present of this amplitude modulated noise on the signal. Then add another band of noise that is outside of the critical bandwidth of the test tone. The off-signal band will be amplitude modulated in exactly the same ways as the on-signal band ("comodulated bands" in Figure 2.5). These two noise bands are said to be comodulated bands because the envelops of their modulated waveform follow the same pattern over time even though they contain different frequencies. Under this condition, it has been found that the masked threshold of the signal becomes better (lower) by up to 11 dB for the comodulated bands. This improvement is called "Comodulation Masking Release (CMR)". It has also been noticed that the masked threshold does not improved (right bar in Figure 2.5) if the on-signal and off-signal band are not comodulated. Therefore, the comodulation masking release (CMR) may be observed if a normal speech signal was used as a masker and the filtered signal was a signal. Skewed frequencies in the masker would cause some degree of masking release due to this phenomenon. In addition, comodulatation masking release (CMR) is observed with both monaural and binaural hearing.

2.5 Spatial Hearing conditions

Spatial hearing should be examined in the context of biological evolution. "The most important higher cortical functions of an animal brain are environmental representation and prediction, and the planning of behavioural response with the goal of maximising the chances of survival and perpetuation of the species (Roederer, 1995)." Our sensory systems, necessary for environmental representation, evolved based on their usefulness in picking out the information that is most useful to our survival from the sea of energy around us (Gray, 1991).

McEachern (1992) argues that signal detection, identification, and location (localisation) were the most critical signal processing tasks for the eyes and ears of early humans. To hunt and avoid being hunted, man had to detect nearby movement, identify it as prey, predator, or friend, and determine its location so they could run towards or away from it as necessary.

All of these views point to the evolutionary advantage of environmental representation through perception of spatial object-person relationships. Spatial hearing exists because it is advantageous to humankind's survival.

2.5.1 Early Theory

The study of spatial hearing began in the area of psychoacoustics and has since been embraced by other disciplines, including acoustics. One of the oldest psychoacoustic theories of spatial hearing is Rayleigh's duplex theory of sound localisation (Kendall, 1995). Based on experiments with sine wave stimuli, Rayleigh found that differences in the signals reaching the listener's ears strongly affected spatial perception. Inter-aural intensity differences (IID) and inter-aural time differences (ITD) were found to have a significant impact in specific frequency ranges. Inter-aural intensity differences also are called inter-aural level differences (ILD) in the literature.

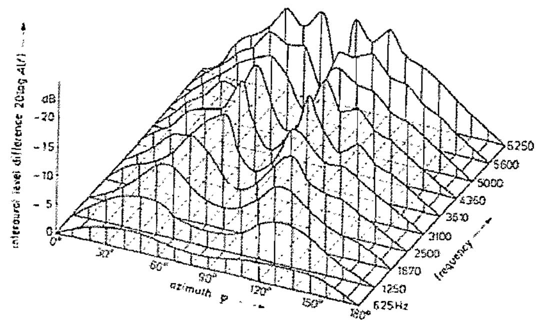


Figure 2.6: Inter-aural intensity differences as a function of sound incidence angle (Blauert, 1997)

The inter-aural intensity differences (IID) are not equally effective at all frequencies because of the physical properties of sound. Due to these traits, sound can diffract around an object; in this case, the subject's head. At low frequency, sound has a long wavelength compared to the size of the head, thus the sound bends around the head. The result is that there is no incidence of a 'head shadow'. On the other hand, high frequency signals have short wavelengths when compared to the size of the head, thus creating a 'head shadow'. With low frequency waves, there will be no inter-aural intensity difference found; but with high frequency waves, it's possible to find inter-aural intensity as large as 20 dB (see Figure 2.6). That means the inter-aural intensity differences (IID) were found to dominate localisation for frequencies above about 1.5 kHz because the head is much larger than wavelengths of sound. Unlike the inter-aural intensity differences (IID), the inter-aural time differences (ITD) are not as important in this frequency range because it is more difficult to judge time delay based on phase differences of higher frequency stimuli. Therefore, the inter-aural time differences (ITD) dominate localisation for frequencies below 1.5 kHz, where sound waves more easily diffract around the head, causing less of an intensity difference between the ears. The inter-aural time difference (ITD) is shown in Figure 2.7 (Wendt, 1963; Feddersen et al, 1957).

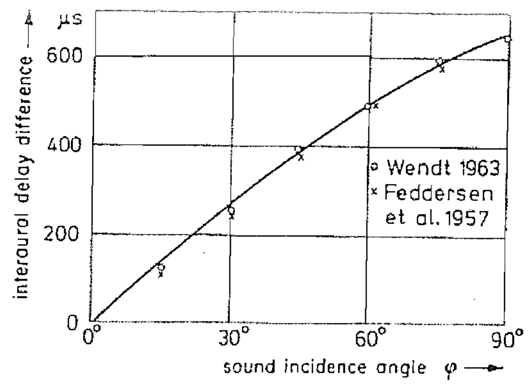


Figure 2.7: Inter-aural time difference as a function of sound incidence angle based on calculation (Blauert, 1997)

2.5.2 Localisation Blur

"Localisation" is the law by which the location of an auditory event is related to specific attributes of a sound event or other stimulus correlated with the auditory event (Blauert, 1997). Basically, the typical question regarding localisation is, "where does the auditory event appear, given a specific position of the sound source?"

"Localisation blur," is the smallest change in specific attributes of a sound event or other event correlated to an auditory event that is sufficient to produce a change in the location of the auditory event (Blauert, 1997). The typical question regarding localisation blur is "what is the smallest possible change of position of the sound source that produces a just-noticeable change of position of the auditory event?"

Sound Localisation blur has been investigated by many researchers since 1966. In an experiment conducted by Preibisch-Effenberger and Haustein and Schirmer (Blauert, 1997), localisation blur was measured in the horizontal plane as shown in Figure 2.8.

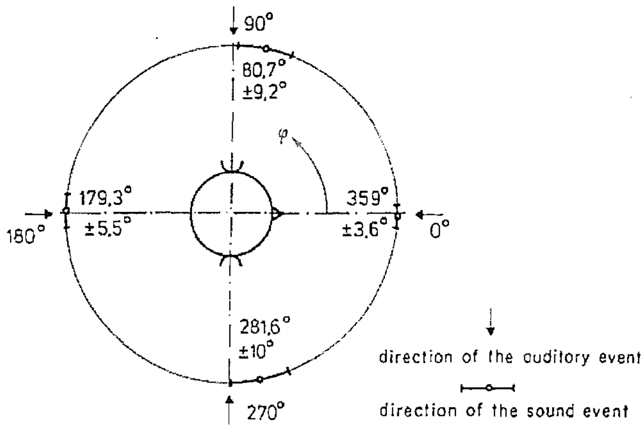


Figure 2.8: Sound localisations blur (Blauert, 1997)

Localisation blur in the horizontal plane for a sound source at 0° azimuth has been measured between about ±3.6°, Localisation blur increases to about ± 9.2° and ± 10° at the sides (90° or 270°) and decreases again to about ± 5.5° in the rear (180°) (Blauert, 1997).

2.5.3 Cone of Confusion

Later investigations of spatial hearing discovered that the inter-aural intensity difference (IID) and the inter-aural time difference (ITD) do not explain localisation sufficiently. In fact, they only affect the lateralisation of the sound source, where lateralisation is the perception of position along the inter-aural axis on the frontal plane. When signals presented to the ears possess only the inter-aural intensity difference (IID) or the inter-aural time difference (ITD), listeners can describe the extent to which the signals are to their left or right, but not whether they are in front of, behind, above, or below them. Kendall (1995) stated that Woodworth called this ambiguity of location at a given degree of lateralisation the "cone of confusion," as shown in Figure 2.9. It was given its name because all points that occur at the same distance from the left and right ears form a cone opening outward from each ear.

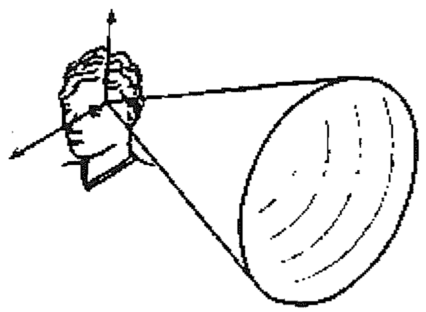


Figure 2.9: Cone of confusion (Handel, 1989)

It can be shown that the cone of confusion forms a hyperbola on the horizontal plane. For experiments concerned only with horizontal azimuth, this "hyperbola of confusion" results in confusions between auditory event locations in front of or behind the frontal plane. These confusions are called front-back reversals (or front-back confusion).

2.5.4 Head Related Transfer functions (HRTF)

Rayleigh's duplex theory was insufficient to describe spatial listening because it did not describe how sound waves reaching the ears are affected acoustically by the listener's body, how these effects change under head motion, and other temporal cues to localisation.

Head-related transfer functions (HRTF) describe the acoustic interactions that a sound wave has with the listener's torso, head, pinnae (outer ears), and ear canals. The complexity of these interactions makes the HRTF at each ear strongly dependent on the direction of the sound (Kendall, 1995).

Head-related transfer functions (HRTF) may be examined in the time domain or the frequency domain. In the time domain, the original impulses are spread over 1 to 3 ms by acoustic interactions with the listener's body (Kendall, 1995). Differences between the HRTF time domain representations are expressed as the inter-aural intensity difference (IID) and the inter-aural time difference (ITD). When a sound source is directly to the listener's side, the inter-aural time difference (ITD) reaches its maximum value between 0.6 to 0.8 ms depending on the signal. Alternatively, the inter-aural intensity difference (IID) must be between 15 and 20 dB for localisation to be completely to one side of the listener (Blauert, 1997).

More subtleties are seen in head-related transfer functions (HRTF) when they are examined in the frequency domain. Comparing HRTF from two ears, we see that magnitude spectra are more similar for frequencies below about 1,500 Hz. Thus magnitude spectra differences are more evident for higher frequencies. Phase spectra may be interpreted as either phase delay or group delay. Phase delay differences are greatest for sound waves at low frequencies because their diffraction around the head slows them relative to those at high frequencies. The transition between these low and high frequency regions exists from 500 to 2,500 Hz and is centred around 1,500 Hz.

Obviously the frequency domain perspective on head-related transfer functions (HRTF) corresponds well with classical duplex theories. However, the inter-aural intensity differences (IID) and the inter-aural time difference (ITD) vary in complex ways across frequencies because of the constructive and destructive interference of the direct wave with sound reflected off the body. Sound above 4 kHz is reflected mainly by the pinnae, while sound below 2 kHz is reflected mainly from the torso. In between, there is a region of overlapping influence. The pinnae are especially important to spatial hearing and may be considered as an acoustic, linear filter. "By distorting incident sound signals linearly and differently depending on their direction and distance, the pinnae codes spatial attributes of the sound field into temporal and spectral attributes (Blauert, 1997)." The frequency dependence of the inter-aural intensity difference (IID) and the inter-aural time difference (ITD) are important to the resolution of front-back reversals and are not captured by Rayleigh's duplex theory.

2.5.5 Motional Theories

Because head-related transfer functions (HRTF) are such strong functions of the direction of the sound source relative to the position of the head, head movement plays an important role in spatial hearing.

Head movement is a natural behaviour. Man always moves his head in order to locate the sound source (Thurlow, Mangel and Runge, 1967). Thurlow, Mangel, and Runge (1967) observed more than 50 subjects whom were permitted to move their head freely, while they were attempting to determine the position of a sound source in an anechoic chamber in order to study the natural movement of the head on sound localisation. They observed that every subject moved their heads in order to locate the sound source. The movements were always obvious and are a combination of classes of movements involving rotating and tipping. It was also observed that the subjects often moved their heads more than once to allow the localisation to become more precise. This experiment also shows that the natural movement of the head can be distinguished into two classes. Firstly, the movement of the head is toward the position of the auditory event and then toward the most probable position of the sound source. In this case they require a more or less precisely located auditory event before the movement begins. Secondly, the movement of head occurred when the position of the auditory event is unclear. Searching and orienting movement is then undertaken. Their goal is clearly to assemble more information in order to establish a final position for an auditory event that is at first not sharply located. Generally, the subjects' knowledge of the location of the auditory event becomes more precise during the head movement.

With the head movement, the acoustic interference of the sound wave with the head changes and the head-related transfer functions (HRTF) change accordingly. Whereas the duplex theory alone led to front-back reversals, the changing HRTF resulting from head movement considerably reduced their occurrence (Wightman et al, 1994; Kendall, 1995). Assuming the duration of a sound event is long enough, exploratory head movements almost always increase localisation precision (Blauert, 1997).

2.5.6 Auditory Scene Analysis

Not only Rayleigh's duplex theory of sound localisation and head-related transfer functions (HRTF) have been investigated in the last few decades; the auditory perception and the formation of auditory images have also been a subject of interest. Bregman (1990) suggests that our auditory perceptual faculties evolved as a means of allowing us to construct a useful representation of reality. It provides us with the what, when, and where of the events around us. The primary task of the auditory system is to arrange the cacophony of frequency wisps into meaningful clumps that correspond to various real-world activities. In other words, the act of hearing may be likened to the work of a cartographer constantly drafting maps of the auditory scene.

Some sounds, such as the slamming of a door, mark the occurrence of unique events. But the world of sound is not merely a succession of momentary incidents. Even discrete sounds, such as a series of footsteps, are often caused by an on-going coherent activity. Most sounds have a lineage or history. The mental images that form from such "lines of sound" that Bregman (1990) dubbed 'auditory streams' and the study of the behaviour of these images is the study of auditory streaming.

Auditory streaming entails two complementary domains of study. Firstly, how sounds cohere to form a sense of continuation is the subject of perceptual fusion. The other is how concurrent activities retain their independent identities. That is the subject of stream segregation.

2.5.7 Auditory Stream Segregation

Auditory stream segregation is (1) the perceptual process in which relationships are formed between different sensations and (2) the effect of relationships on what is included and excluded from our perceptual descriptions of distinct auditory events (Bregman, 1990). It is the method by which our brains group together or "fuse" multiple sensations of acoustic stimuli because they are understood to have come from a common source. Spatial location of a sound source is one of the many perceptual cues to auditory stream segregation; others include timbre, melodic direction, and even position or movement perceived visually. Localisation of auditory events at different positions is really only important in this project as it relates to their segregation into different streams.

Spatial cues affect stream segregation for sensations resulting from non-simultaneous or simultaneous events. Segregation by localisation does occur for the case of non-simultaneous events. However, it seems weak unless supported by other bases for segregation. For simultaneous sound events, or ones that run in and out of simultaneity such as two people talking, localisation effects on stream segregation are much stronger.

2.5.8 Cocktail Party effect

The "cocktail party effect" describes the ability to focus one's listening attention on a single talker among a mixture of conversations and background noises, ignoring other conversations. The cocktail party effect is the most famous example of the localisation effects on stream segregation. It arises from the fact that a desired signal S with a certain direction of incidence is less effectively masked by an undesired noise N from a different direction when subjects listen binaurally (with two functioning ears) than when they listen monaurally (Cherry, 1953; Moore, 1982). The cocktail party effect falls into the more general case of sound events of lower subjective loudness not being masked by those of higher loudness because of different localisations. This is also known as "binaural masking release".

2.6 Binaural masking release

Binaural masking release can be regarded as having four components. One is an advantage arisen directly from improvements in signal-to-noise ratio at the 'best ear', which are caused by a 'head shadow' (Bronkhorst and Plomp, 1988). The second is an advantage arisen from binaural masking level differences (BMLD), which are largely facilitated by differences in inter-aural time delay (ITD) between competing sources (Bronkhorst and Plomp, 1988; Zurek, 1993; Culling and Summerfield, 1995; Breebaart et al., 2001a,b,c;). Thirdly, the masking can be influenced by the perceived spatial locations of signal and masker (Kopčo, 2003). And finally, the masking released due to head rotation according to "motional theory" (Koenig, 1950; Burger, 1958; Thurlow and Mergener, 1971; Lambert, 1974; Rodgers, 1981; Wightman et al., 1994; Blauert, 1997; Yost and Mapes-Riordan, 1997).

2.6.1 Best ear

In general, signal detection could be in either ear separately or binaurally. When signal and noise are generated in both ears under diotic listening, the signal to noise ration in both ears is equal. Therefore, theoretically, the signal thresholds of both ears are the same. There should be no masking release. On the other hand, when signal and noise are presented in sound field under binaural listening, the signal to noise ratio in the left ear and the right ear are different because of the inter-aural intensity differences (IID) provided by the head shadow effect. For instance, if the signal is presented from the front and the noise is presented from the left hand side; the signal level in both ears will be the same, but the noise level in the left ear will be higher than the right ear. Therefore the signal to noise ratio in the right ear is better than the left ear. According to these differences in signal to noise ratio, the signal threshold for binaural hearing trends to be similar to the best signal to noise ratio ear (Bronkhorst and Plomp, 1988). However, the result from Bronkhorst and Plomp (1988) indicated that the threshold for actual binaural hearing is somehow lower than the threshold in the best ear. Therefore, they argued the additional masking release might be provided by the other binaural cue that is inter-aural time differences (ITD). The binaural release from masking using inter-aural time differences (ITD) is known as "binaural masking level difference (BMLD)".

2.6.2 Binaural masking level difference (BMLD)

The binaural masking level difference (BMLD) was first introduced in 1948 by Licklider (1948). Later in the same year, Hirsh fully described this phenomena in detail. He reported that differences in the binaural masking level are a great way to detect signals in the presence of a masker. This phenomenon was discovered when speech signals were masked by a broadband noise. Later, it has been extensively investigated in a much simpler form: a single low-frequency tone masked by noise. When the noise is the same at both ears, a tone presented at 180° out-of-phase at both ears is 10-15 dB more detectable than when the tone is identical at the ears, as shown in Figure 2.10.

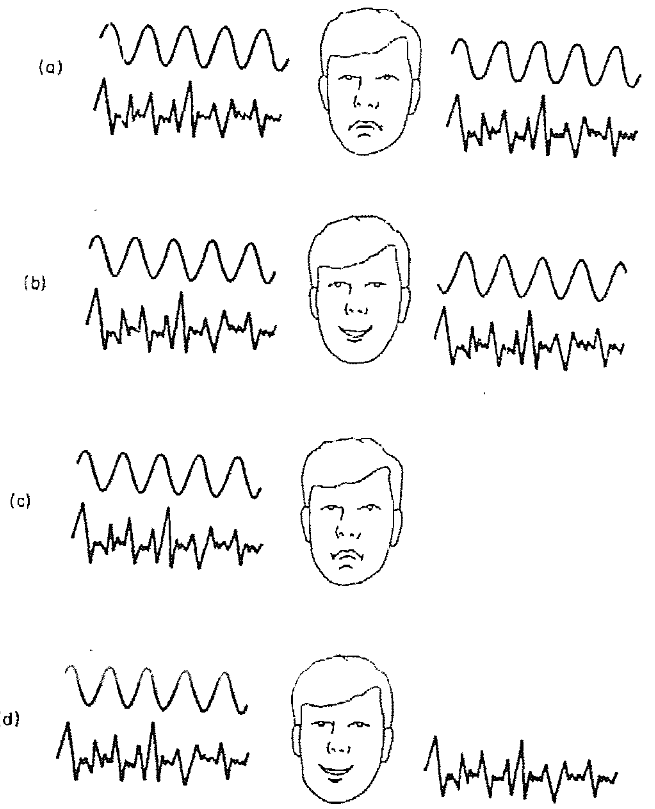


Figure 2.10: Illustrating the situation of binaural masking level difference (BMLD) (a) and (c) detectability is poor, while in (b) and (d), the inter-aural relations of the signal and masker are different, the detectability is good (Moore, 1997).

Binaural masking level differences have not only been studied under conditions shown in Figure 2.10, but also under various conditions since 1948. A large amount of work has been done on the behaviour of the binaural masking level difference under various conditions, including the works of Hirsh (1948), Licklider (1948), Carhart et al. (1969), Koenig et al. (1977), Yost and Walton (1977), and of Zwicker and Henning (1984). They surveyed and published the literature as the basic behaviour of BMLD that can be summarised as shown in Figure 2.11 (Gelfand, 1998). From the Figures, the type of sound presentation will be indicated, as also in most of the literature, by the following symbols:

S for the (desired) signal

N for (interfering) noise

m	for monotic presentation
0	for diotic presentation with the Inter-aural phase term 0
φ, π	for dichotic presentation with the Inter-aural phase term φ or π
τ	for dichotic presentation with the Inter-aural time difference τ
u	for dichotic presentation with Inter-aurally uncorrelated signal

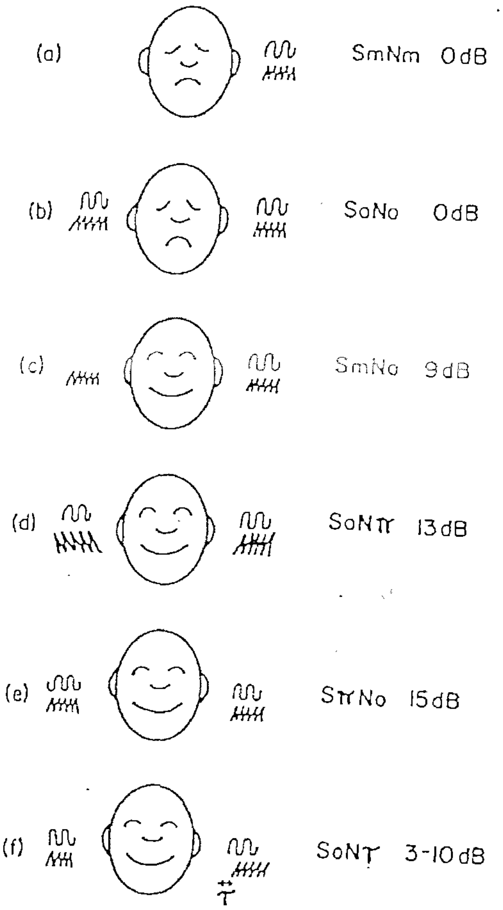


Figure 2.11: Masking level difference (MLDs) for various conditions. (a) and (b) detectability is poor, while in (c), (d), (e), and (f), the smiles mean that the detectability is good (Gelfand, 1998)

From Figure 2.11, a signal S is barely masked by a noise N when the signal and noise are presented in one ear (monotically), as in Figure (a), or when the signal and noise are presented in both ears (diotically), as in Figure (b). From Figure (a), suppose the signal is still monotic but the noise is now diotic, as in Figure (c), the previous masked signal now becomes audible. This is not the only situation in which the masked signal becomes audible. The masked signal in Figure (b) can be made audible again by reversing the phase of the noise between the ears, as in Figure (d), or by reversing the phase of the signal between the ears, as in Figure (e). Finally the signal becomes audible when the noise in one ear is delayed in respect to the other. As far as the release from masking under these conditions is concerned, the binaural advantage occurs only when the stimuli are in some way different at the two ears. However the levels of release from masking in each condition shown in Figure 2.11 are not the same.

From Figure 2.11, Hirsh (1948) studied the hierarchy for the BMLD that occurs with a combination of signals and noises that are easily generated using headphones. The results of his experiment are shown below:

N_0S_0: BMLD = 0 dB (reference condition)

N_0N_m: BMLD = 6-9 dB

$N_\pi S_0$: BMLD = 9-12 dB

N_0S_π: BMLD = 12-15 dB

It can be seen that the size of binaural masking difference varies from as large as about 15 dB to as little as 0 dB. The largest BMLD are obtained when either the signal or the noise are opposite in phase at the two ears ($N_\pi S_0$, N_0S_π). The large BMLD obtained under anti-phase conditions have been repeated confirmed by Jeffress et al. (1952). A few years later, Green and Henning (1969) suggested that the large BMLD associated with anti-phase conditions might be related to the firing patterns of auditory nerve fibres, particularly at low frequencies. According to this suggestion, the Binaural masking level difference will be greatest at low frequency and then decrease as frequency becomes higher.

The reduction in Binaural masking level difference at high frequency has been observed in many instances of research (Hirsh, 1948; Webster, 1951; Hirsh and Burgeat, 1958; Durlach, 1963, Schenkel, 1964; Rabiner et al., 1966). The relationship between the size of a binaural masking level difference and stimulus frequency of a sinusoidal signal when the interfering signal is broad band noise for a number of studies (Hirsh, 1948; Webster, 1951; Hirsh and Burgeat, 1958; Durlach, 1963, Schenkel, 1964; Rabiner et al., 1966) is summarised by Durlach (1972), can be seen in Figure 2.12.

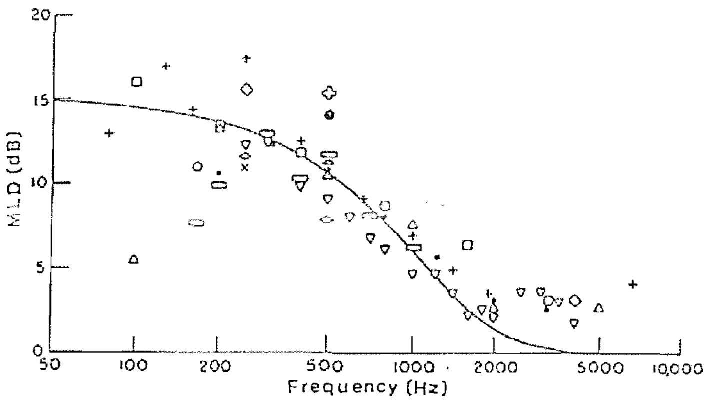


Figure 2.12: Magnitude of the MLD as a function of frequency from many studies. (Blauert, 1997)

From Figure 2.12, it can be seen that the BMLD reaches a maximum of 12-15 dB between 200-300Hz, and then decreases for higher frequencies until a constant of about 3 dB is maintained at about 1500-2000Hz. This reduction of BMLD at high frequencies is often explained by saying that the phase difference between two ears is the determining variable for this BMLD.

The wide range of measured BMLD at low frequencies is apparently due to the differing signal level used. Researcher has shown that an increase of the non-linear BMLD masker at low frequency gives a value of approximately 15 dB when the noise level reaches 50 dB or more (Schenkel, 1964; Dolan, 1968; McFadden, 1968). This non-linearity increase in BMLD lead to the level dependence that McFadden (1968) explained term of internal noise generated inside the auditory system. He said that the internal noise dominates over the external noise when the level of the external noise is very low.

As has been described previously, the binaural masking release affected by frequency (Hirsh, 1948; Webster, 1951; Hirsh and Burgeat, 1958; Durlach, 1963, Schenkel, 1964; Rabiner et al., 1966) was summarised by Durlach (1972) as the phase and time differences between two ears.

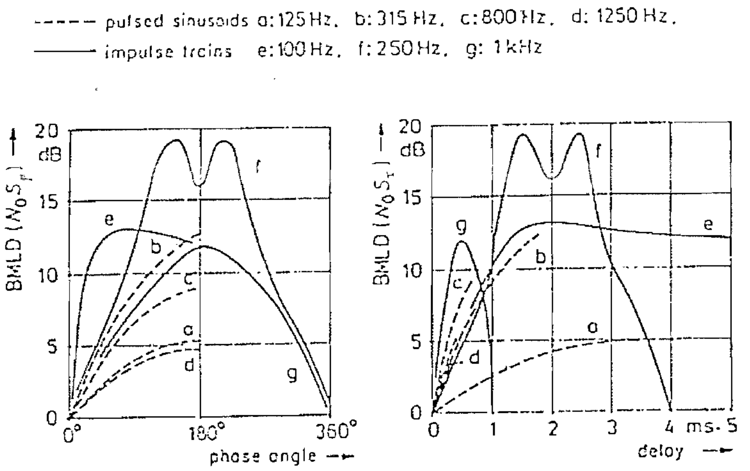


Figure 2.13: BMLD as a function of Inter-aural phase difference or time difference of signal. (Blauert, 1997)

The BMLD as a function of Inter-aural phase and time difference for pulse signals and pulse trains that was studied by Schenkel (1964) and Flanagan and Watson (1966) is summarised by Blauert (1997), as shown in Figure 2.13. It can be seen that the maximum BMLD occurred with an Inter-aural time difference of 1.5-2 ms. This maximum BMLD was presented for that the signal or impulse trains at frequency of 250Hz which explains the position of the maximum in Figure 2.13. From this Figure, it also shows that the BMLD as a function of Inter-aural time difference change relative to the function of Inter-aural phase difference. This relation was also shown in a work of Langford and Jeffress (1964). They carried out an experiment to investigate the dependence of BMLD on Inter-aural time difference. They examined the BMLD of 500Hz sinusoidal signal in two conditions that were $N_\tau S_0$, and $N_\tau S_\pi$. The result indicated that the BMLD reaches a maximum whenever the Inter-aural phase relationship of the signal is opposite that of the noise.

So far, it can be seen that the effects of BMLD are partly caused by the difference in signal or noise in two ears. If the noises at the two ears are derived from the same source, the noise waveforms will be exactly the same in both ears. In other words, the noises are perfectly correlated. However, when the noises are generated from difference sources, the noises are uncorrelated. These uncorrelated noises have been reported to affect the size of BMLD; as the degree of correlation decreases the BMLD results become larger (Robinson and Jeffress, 1963).

This resulting of BMLD from uncorrelated noise may contribute to understanding the effects of reverberation. The relationship between BMLD and reverberation was first demonstrated by Koenig et al. in 1977. He determined the masking level difference in a reverberant environment by conducting two type of masking situations (NR_uS_0, and NR_0S_0) in an experiment. The band-limited noises were generated and recorded in a reverberant field. It was used as masker, while the speech was used as signal. The threshold measure was obtained by using headphones. The experiment showed that the values of BMLD depended on the degree of noise correlation, with the binaural processing providing approximately a 3 dB masking level difference in a reverberant field (Koeing et al., 1977).

Apart from all work on BMLD mentioned above, there is precious little work that has been done to investigate BMLD when the signal and noise are not presented simultaneously. Almost all work in this area was done around 1969-1970 by Deatherage and Evan (1969), and Dolan and Trahiotis (1970) but the most thorough work was done by Yost and Walton in 1977. These two conducted an experiment to measure the masking threshold under four binaural conditions (S_0N_0, S_mN_0, S_0N_π, and $S_\pi N_0$), and in four temporal masking conditions (simultaneous masking, forward masking, backward masking, and combination forward-backward masking). The results indicated that BMLD can also be observed in connection with both forward and backward masking, as shown in Table 2.1.

Table 2.1: The average BMLD in dB (refer to S_0N_0) Under each binaural and temporal conditions (Yost and Walton, 1977)

Binaural	Temporal Condition			
Condition	Simultaneous	Forward	Backward	Forward & Backward
S_mN_0	10 dB	4 dB	5 dB	9 dB
S_0N_π	14 dB	7 dB	7 dB	13 dB
$S_\pi N_0$	17 dB	9 dB	9 dB	16 dB

All of the previously described investigations of BMLD were carried out using headphones. Free-field BMLD was first investigated in the 1960s, but subsequent major investigations were done by Burtorf (1963), Ebata et al. (1968), and Damaske (1970). These works were concerned with the masked threshold of reflections, which is relevant to this context. Their results showed that if the incident direction of signal and noise differ, the threshold generally becomes smaller than when the signal and noise arrived from the same direction. Therefore the phenomenon of binaural signal detection in free field may be in conjunction with the direction of incidence or spatial localisation that will be discussed in the next section.

2.6.3 Spatial location

Spatial localisation plays an important role in masking release because it has been found that binaural release from masking is the improvement in detection obtained when a signal is separated in space from a masker; this detection is probably involved in listening in noisy environments. Ebata et al. (1968) even hypothesised that the

BMLD can be especially large if the subject knows the direction of incidence of the desired signal (Ebata et al, 1968). This spatial release from masking occurs because of two factors. They are the head shadow effect, which appears to influence thresholds for high-frequency stimuli, and binaural processing, which takes advantage of inter-aural time differences (ITDs) at low frequencies (Good et al., 1997). However, the head shadow effect and the effect on binaural processing has always been mentioned together in literature (Santon, 1987; Saberi et al., 1991; Gilkey and Good, 1995).

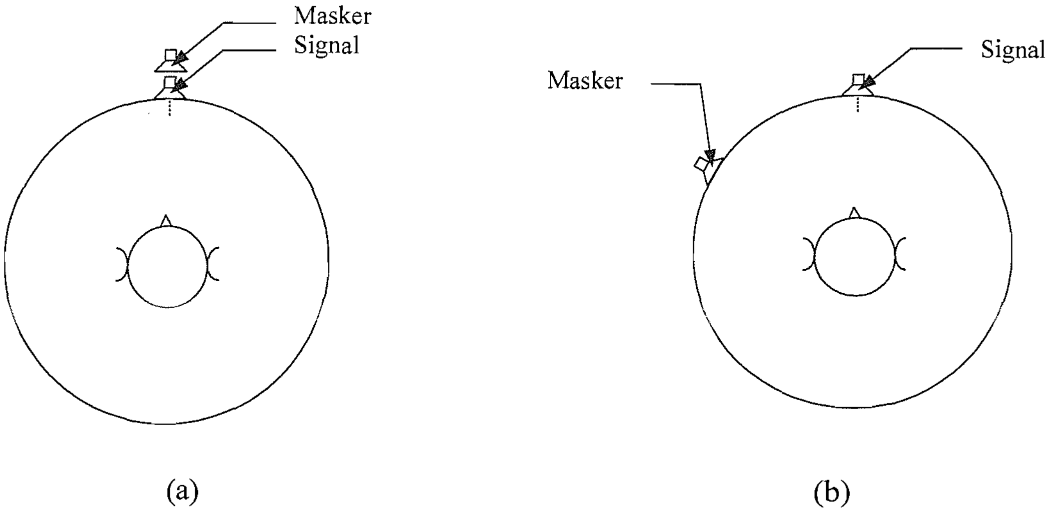


Figure 2.14: Direction of signal and masker (a) from the same azimuth angle of incidence (b) from the difference azimuth angles of incidence.

In 1987, Santon studied the BMLD under free field conditions. He presented a tone at frequencies between 250 and 4000Hz in a background pink noise and measured the threshold levels for various positions of masking. He found that the poorest detection is not when the signal and masker are spatially coincident. He explained his results in term of a model of the diffraction of sound waves around the head. But, later, Saberi et al. (1991) explained that this effect failed with broad band signal by reporting on the work of Kurozumi and Ohigushi, and conducted an experiment to examine this effect. They investigated the detection of broad band click train in background Gaussian noise for various position of the signal. Their results confirmed that the poorest detection occurs when the signal and masker are spatially coincident, with a reduction of about 18 dB found when the signal and the masker were spatially separated (Saberi et al, 1991). However, this experiment is unclear that this observed effect would be expected in all frequencies because it did only investigate on broad band stimuli. Therefore, in 1995, another experiment was conducted by Gilkey and Good (1995), further investigating the influence of the frequency content. This experiment was similar to Saberi et al.'s experiment, apart from the fact that the frequency of the stimuli was separated into low frequency, mid frequency, and high frequency. The results showed that the reduction of the threshold occurred when the signal and masker are spatially separated, which was the same as the previous experiment. However it also showed that the great reduction in masking observed in high frequency than the reduction in masking in low frequency, or mid frequency.

2.6.4 Head rotation

As has been described previously, many studies have investigated the effect of head movement on localisation, but none of these studies has investigated the effect of head movement on binaural release from masking. The head movement changes inter-aural time differences (ITD) and inter-aural intensity differences (IID) that contribute to binaural masking release. Therefore, head movement might affect the masking effect under dichotic listening conditions. Consequently the effect of head rotation on binaural masking will be investigated in chapter 6.

2.7 Masking of reverberation

Masking of reverberation has been a subject of interest for many decades. In the literature, the masking of reverberation has been investigated from two points of view. One is that the masking of reverberation has been studied in term of energy. The other is that the masking of reverberation has been investigated based on masked threshold level.

2.7.1 Based on energy

Masking effect in auditorium acoustics has been a subject of interest for several decades. It is known that a masking effect is an important factor in evaluation the overall sound quality in a room (Marshell, 1967; Marshell 1968a, b; Kuttruff, 1979; Beranek, 1986; Maekawa and Lord, 1994; Beranek, 1996), and is also an important factor in the perception of speech in a room (Knudsen, 1929; Bolt and Macdonald, 1949; Haas, 1972; Nábělek and Dagenais, 1986; Nábělek et al., 1989).

Lawrence and Lord (1989) state that reverberation always masks the direct sound and causes direct sound to become inaudible, with no impression of overall sound quality. This masking effect has been observed in speech as well as in music (Beranek, 1996). Therefore, to avoid speech degradation due to the masking effect, a short-reverberant room is always preferred (Maekawa and Lord, 1994).

It is known that reverberation appears to act as noise masker in reducing speech intelligibility; however this is actually an oversimplification. The reflected energy of reverberation overlaps the direct (original) speech signal, so that perceptual cues are masked.

Delayed same-signal masking effects in a room were first reported in 1929 by Knudsen (1929). He theorised that masking effects arise when the energy of a preceding sound overlaps onto the following sound. This is called "overlap masking". He observed that overlap masking causes the errors to occur among final consonants in the perception of non-sense syllables. Since that time, the masking effect has been investigated by number of studies, such as Bolt and Macdonald (1949), Nábělek and Dagenais (1986), Nábělek et al. (1989).

In 1949, Bolt and Macdonald reported that the masking effect in a room can be classified into two types – overlap masking and self masking. They found that overlap masking does not usually appear alone in an auditorium, but often appears together with self-masking, which they defined as an internal temporal smearing of the energy within each sound. It takes place over a much shorter time scale when reverberant decay from the onset of a sound contributes to ongoing masking of that same sound However, they assumed that self-masking must be relatively unimportant because the effect of this masking on speech intelligibility can be reduced by decreasing the duration between direct sound and reflection (Bolt & Macdonald, 1949). Using this reasoning, self-masking has never been considered a vital element in determining the masking effect in a room (Kurtović, 1975).

Until 1986, Nábělek and Dagenais (1986) reported that self-masking must be important in reverberations because self-masking can cause some vowel confusion. They observed that vowels that were preceded by a consonant /b/ were confused when presented as a reverberant. They surmised that it was unlikely that any significant overlap masking occurred because the intensity of the /b/ sound was much lower than the intensity of the following vowels. A similar effect was also found in an experiment that was conducted a few years later by Nábělek et al. (Nábělek et al., 1989). Nábělek et al. reported that most errors with synthetic syllables in reverberant tests were caused by a combination of overlap masking and self masking. Therefore, the effect of both types of masking must be important in perception of speech in a room.

In summary, there are two types of masking in a room. Overlap masking occurs when the reverberant decay from a previous sound masks the following sound. Self masking takes place over a much shorter time scale when reverberant decay from the onset of a sound contributes to ongoing masking of that same sound. The separate effects of self masking and overlap masking cannot be distinguished in real room because there is no technical possibility of separating out the two types of masker signal from any measurement or recording of real running speech or a continuous music programme with reverberation.

2.7.2 Based on threshold measurement

The investigations of masking effect in the presence of signal and reverberation have been studied since the 1960s. Experiments have attempted to control reverberant content by using various methods as in the experiments of Burgtorf and Wagener (1968), Koenig et al. (1977), Zurek et al. (2004).

Burgtorf and Wagener (1968) investigated the masking effect by means of subjectively diffuse sound fields. The sound source was generated in a reverberant room. The reverberation was picked up by a set of microphones and was then played back in anechoic room through a set of loudspeakers. The delay of reverberant sound and direction of reverberation were controlled in the experiment.

In 1977, the difference in masking level between two ears under a reverberant environment was studied by Koenig et al. Band-limited noises, which were generated and recorded in a reverberant field, was used as masker while speech was used as signal. Thresholds were obtained by using headphone listening conditions. Two types of masking situations were conducted. They were the signal between two ears were the same and the band limited noise between two ears were difference (NR_uS_o), and the signal between two ears were the same and the band limited noise between two ears were same (NR_oS_o).

Apart from the study of Koenig et al. (1977), the most recent work found is the paper published by Zurek et al. (2004). They studied auditory target detection in reverberation. The signal and masker were passed through a software simulation of a reverberant room and then generated to the subject through a headphone. The experiment was designed under headphone conditions because this made it possible to control the properties of reverberation (masker) (same as in Koenig et al.). This experiment was conducted to develop a detection model to predict speech intelligibility based on listening mode (monaural or binaural) and source location.

Consequently, none of these methods have managed to overcome the fundamental problem of there being no known method for separating out direct from reverberant sounds when they overlap in time. For this research, the problem was overcome by using a method of simulation where a single delayed reflection can be added to the direct sound under conditions of complete experimental control.

2.8 Variable for single reflection

In a simulation where there is a single delayed reflection, there are five possible variables. They are: level difference between the signal and the masker; time difference between the signal and the masker; direction of the masker relative to the direct sound; frequency spectrum of the direct and delayed sounds; and type of programme material.

2.8.1 Level difference between the signal and the masker

Firstly, the level difference between the signal and the masker is a direct measure of the signal to noise ratio. The threshold of the signal will increase when the masker level is increased (see simultaneous masking). In addition, Henning and Zwicker (1984) reported that, under binaural conditions, the threshold is increased with the increasing masker level but at a rate that slower than in the monaural condition, as the inter-aural level difference cues are used in binaural conditions. They also observed that the binaural masking level difference was slightly smaller at the lower masker level, when masker levels of 40 dB and 60 dB were examined.

2.8.2 Time between the signal and the masker

Secondly, the time difference between the signal and the masker is a fundamental property of the acoustic space within which the reverberation occurs. It is known that when two signals differ from one another by more than 1 ms at the position of the listener, then the position of the auditory event in most cases is determined only by the position of the sound source that arrived first. The signal that arrives at the ear first is taken into consideration, while the latter ones are suppressed in the interpretation process. This effect is called "the law of first wave front". It is also referred to under a variety of other names, like the "Haas effect", or "precedence effect" (Aoki, 1992).

The precedence effect in sound localisation and room acoustics has received a considerable amount of attention in literature. The sound quality will be satisfactory in a room with a reflection arriving no later than 50 ms. If the delay between two signals exceeds this upper limit, the two auditory events appear one after the other; One connected with each sound source. The second auditory event is called the echo. It may be perceived and is highly irritating for values large than 100 ms (Beranek, 1986; Maekawa and Lord, 1994).

In terms of masking threshold, Wilson and Carhart (1971) showed that the threshold depended on the time duration between signal and masker in the gap between maskers when they studied the addition of forward and backward masking. The masking threshold shrank when the time interval between signal and masker was increased.

2.8.3 Direction of masker relative to the direct sound

Thirdly, the direction of the masker relative to the direct sound may be important because it will be affected by the acoustic space within which the reverberation occurs. It is not taken into account by any measure of reverberant decay which is not directionally sensitive.

The direction of masker was found to affect the masking threshold. Saberi et al (1991) reported that the thresholds were found to be poorest when the signal and masker are spatially coincident in most studies using speech or a broadband signal. A reduction in masking of 9 dB was observed when the signal and masker were separated within the horizontal plane. The results were also confirmed by Gilkey (1995) for a high frequency sound with an 18 dB reduction of masking threshold.

2.8.4 Frequency spectrum and type of programme material

The last two variables are frequency spectrum and type of programme material. The frequency of masker and signal affect the masking threshold according to both simultaneous and non-simultaneous masking effect (Moore, 1987). The type of programme material affects subjective preferences in different kinds of acoustic spaces. However the type of programme material is coordinated with the frequency spectrum in some senses. The type of programme material contains a frequency spectrum as a time variable. To represent the frequency spectrum on its own, the signal will be a pure tone or tone combination that can hardly be found in real life. Therefore, in the previous research, for example the works of Burgtorf and Wagener (1968), Koenig et al., (1977), and Zurek et al. (2004), speech was used to represent a sound source as in a real situation instead of a tone. Following their footsteps, the speech was investigated together with its frequency spectrum for this study.

2.9 Speech

Speech can be investigated in three dimensions: frequency domain; time domain; and a combination of time-frequency as a spectrogram.

To investigate characteristics of speech in the frequency domain, the fundamental frequency (F0) is a robust feature of the speech signal. This fundamental frequency identifies the difference between adult male and female/prepubescent speech. Before puberty, the fundamental frequency for normal speech ranges between 150-400 Hz for both males and females. After puberty, the vocal cords of males undergo a physical change, which has the effect of lowering their pitch frequency to the range 80-160 Hz. This difference is also readily apparent in the speech signal itself.

To investigate speech in the time domain, speech is not represented in a fixed acoustic pattern. The waveform is quasi-periodic. It can be observed that separated words are uttered at around 2.5 words per second in normal speech. However, it is not easy to analyse the characteristics of continuous speech in this domain alone. In other words, continuous speech is composed of a particularly complex and variable waveform because it contains a large number of words that may have different acoustical characteristics. To make it easier, speech is always analysed in the time-frequency domain or spectrogram. In addition, speech is often broken down into smaller units than whole words, as in vowels and consonants.

2.9.1 Vowels

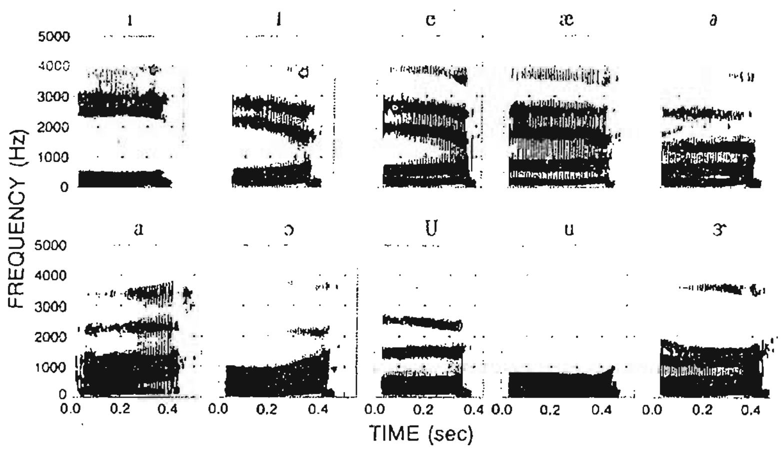


Figure 2.15: The spectrogram of vowel sounds. (Rabiner and Juang, 1993)

Vowels are the simplest sounds to analyse and describe acoustically. They are associated with a steady state articulatory configuration and a steady state acoustic pattern. They can be indefinitely prolonged as an articulatory or acoustic phenomenon. Therefore it is not necessary to consider the time dimension beyond choosing an instant that is taken as representative of the vowels production. In addition, vowels have often been characterised with a very simple set of acoustics descriptors, namely, “formants”. The formant frequencies are generally determined by the shape of the vocal tract that behaves as a filter modifying the source. They are partly a consequence of the frequency and time trade-off. They appear on a wide band spectrogram as thick bars, as shown in the Figure above. The formants are numbered, the one with the lowest frequency is called the first formant, the next is the second, and so on; as many as six formants can be observed in speech spectrogram, but the first two formants are the most important for the purpose of identifying the vowel, as shown in Figure 2.15.

Rabiner and Juang (1993) report on the work of Peterson and Barney that there are frequency overlaps between the formant frequencies for different vowel sounds in different talkers as shown in Figure 2.16. The ellipses drawn in this figure present gross characterisations of the regions in which most of the tokens of the different vowels lie. It has been clearly seen that it is not simple to measure formant frequencies or spectral peaks accurately to precisely classify vowel sounds, as the formant frequency of vowel sounds depend on the type of talker. The differences between male and female speech will produce different formant frequencies for each vowel (Table 2.2).

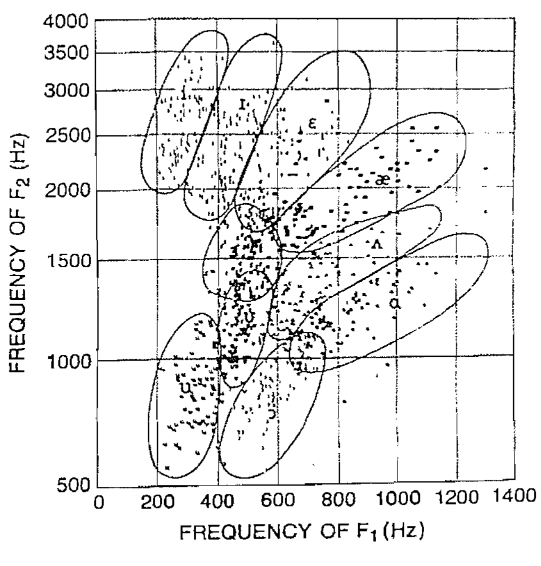


Figure 2.16: Measured frequencies of first and second formant for a wide range of talkers for several vowels (Rabine and Juang, 1993 after Peterson and Barney, 1952).

Table 2.2: Average frequencies of the first three formants (F1, F2, F3) of the vowels of men, women and children (from Appleton and Perera, eds., The Development and Practice of Electronic Music, Prentice-Hall, 1975, p.42; after Peterson and Barney, Journal of the Acoustical Society of America, vol. 24, 1952, pp. 175-84).

	Formant	heed	head	had	hod	haw'd	who'd
Male	F1	270	530	660	730	570	300
	F2	2290	1840	1720	1090	840	870
	F3	3010	2480	2410	2440	2410	2240
Female	F1	310	610	860	850	590	370
	F2	2790	2330	2050	1220	920	950
	F3	3310	2990	2850	2810	2710	2670
Children	F1	370	690	1010	1030	680	430
	F2	3200	2610	2320	1370	1060	1170
	F3	3730	3570	3320	3170	3180	3260

As has been described previously, the formant frequencies of vowels vary with each speaker. In order to describe the formant frequency in a common way, the formant frequencies of vowels are represented by an average behaviour and are not represented variably across talkers. If a plot of a vowel sound is made against variation in the frequencies of formant 1 & 2 (F1 and F2), they are seen to form a roughly triangular fashion, as in Figure 2.17. This Figure shows the classic vowel triangle. The common frequencies of formant can also be represented in terms of formant positions by the data given in Table 2.3.

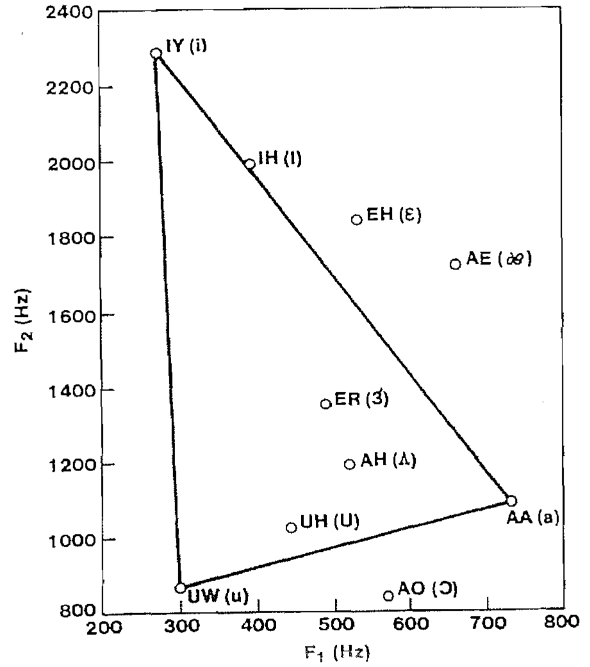


Figure 2.17: The vowel triangle of the common vowels (Rabiner and Juang, 1993).

From the Figure, the vowels triangle represents the extremes of formant locations in the F1-F2 plane. It represents /i/ as low F1 and high F2, /u/ as low F1 and low F2, and so on. In the other words, the Table demonstrates the first and second formant frequencies of each vowels sound.

Table 2.3: Formant frequencies for typical vowels

Vowel	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
F1	280	370	405	600	860	830	560	430	400	330	680
F2	2230	2090	2080	1930	1550	1170	820	980	1100	1260	1310

According to Figure 2.17 and Table 2.3, the vowels can be identified into formant frequencies. The concept of formant frequencies in analysis vowel in a speech has been one of the most frequently used to identify the characteristics of vowels. The most important formants used are always the first formant and the second formant (Kent and Read, 1992).

2.9.2 Consonants

Consonants are sounds that are non-periodic and are weak components of speech. They are produced by a partial or complete obstruction of the flowing air sound along the vocal tract. This obstruction causes the sound to become quasi-periodic or non-periodic and noise-like. They may be classified according to their manner of articulation, such as stop consonants, fricative consonants, affricate consonants, and nasal consonants.

The physical nature of consonants means they will not be as static as vowels. The bandwidth of most consonants is more than an octave, and together their energy can range from 100 Hz to 10 kHz. However, their spectrum will change as a function of time. This change is known as 'formant transitions' as shown in Figure 2.18.

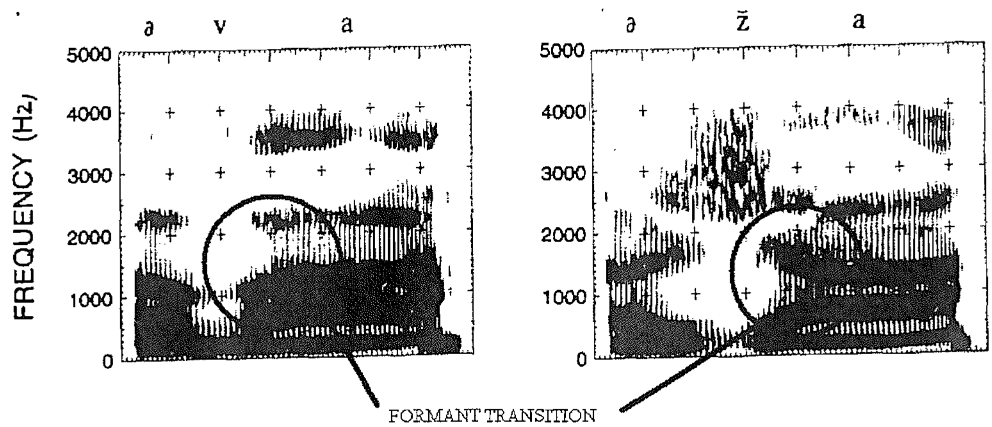


Figure 2.18: The formant transition between the frequency of a consonant and formant frequency of a vowel in /ə-v-a/ and /ə-zh-a/.

Formant transition is the only main characteristic of consonants. It can only be observed in a word. In a word that contains consonant and a vowel, there is always a rapid change in the frequency of a given formant or set of formant between formant frequency of vowel and frequency of consonant or formant transitions. Although this formant transition has been observed to be an important acoustic cue for the perception of consonants in speech, it unfortunately does not fully explain the characteristics of consonants.

2.9.3 Silence interval or silence gap

Speech may show a number of time intervals where there is little or no spectral energy as silent interval. These silent intervals do not always correspond to spaces between the words, and often occur during the speech sounds representing a particular word. Thus, in consideration of speech characteristics, the silent interval is also important. It controls the speed of the speech sound. A long silent interval can slow the speech down. On the other hand, speech speeds up when the silent interval is short.

To summarise, speeches do not represent a fixed acoustic pattern because they contain vowels, consonants, and silent intervals that all have different acoustical characteristics. In order to study the masking effects of speech, these characteristics should be taken into consideration.

2.10 Speech masking

The classical masking effects of noise on speech signals were carried out by Miller (1947), Steven et al. (1946). They observed that the masking effect of broad-band noise on speech is similar to the masking effect of pure tone on pure tone. That is, the masking effect increases linearly with its level, spreading upward in frequency. Therefore, when a noise masker reaches a certain level, an increase in the noise level will result in an equivalent increase in the speech threshold. The low frequency band of a noise masker is more effective in masking the speech signal than one whose energy is concentrated in the higher frequencies (Miller, 1947).

It can be seen that the effect of speech masked by noise is similar to the masking effect of a pure tone on pure tone, or of noise on a pure tone. However this pattern of masking effect might not be exactly the same as the masking effect of speech on speech because the speech masker is not represented in a fixed acoustical pattern as noise, but is composed of variable waveforms as vowels, consonants and silent intervals.

According to the characteristics of speech, the speech masker effect is different from that of broad band noise. Speech can provide an important cue in signal detection. If the speech signal is presented as a masker, the signal will be presented as either vowels or consonants or even a silent interval of the speech masker; therefore, the signal can be detected in each part of the speech masker. However, the silent interval, or silent gap in the speech masker, should play an important role in this signal detection. This is illustrated by the results of many investigators (Wilson and Carhart, 1971; Zwicker and Henning, 1984; Akeroyd and Summerfield, 1999). Wilson and Carhart (1971) investigated the interactions and addition of forward and backward masking effects in the gap between two maskers, with varied duration. They reported that the masking threshold level depends mainly on the gap's duration. A smaller release from masking is found in a smaller gap. On the other hand, a great release from masking is found in a bigger gap. The masking threshold may drop as the absolute threshold when the gap is bigger than 400 ms. However, speech is unlike noise. Speech not only has a silent gap to provide a cue in signal detection, but low amplitude consonants can also provide a cue in signal detection because they do not act like a silent gap (Spiegel, 1987).

In 1987, Spiegel (1987) stated that the investigations using restricted speech stimuli could not explain the masking effects of natural speech. Therefore Spiegel (1987) conducted an experiment to study the masking effect of a normal speech masker. He investigated the effects of simultaneous and non-simultaneous masking within stop /d/ and flap /f/ closures. The results indicated that the amount of combined simultaneous and non-simultaneous masking can be 10-15 dB more than simultaneous masking alone. It has been also found that combination between these masking occurs in a low energy regions when is adjacent to a high energies regions. Whenever the energy of speech between low and high regions differed by 20 dB or more, within an interval of about 20-30 ms, in any frequency band, non-simultaneous masking contributed to the overall masking effect. Consequently, Spiegel (1987) concludes that the threshold of masking is generally the strongest masking influence, as non-simultaneous masking always occurs near the formant peak because natural speech contains rapid amplitude fluctuations. Finally he suggested that it is possible that other natural listening conditions will show greater effects of non-simultaneous masking.

To summarise, with speech masking under monaural conditions, the signal can be detected in the silent intervals or the low amplitude consonants of the speech masker. The following section describes research issues in this study

2.11 Research issues

In this study, three experiments were carried out to measure the different effects of the previously described variables on masked threshold levels of the direct sound. In each experiment, the masked threshold levels were measured by varying the level of the direct sound in the presence of a range of delayed masker signals covering the key variables of interest. The method used was staircase or up-down. The potential difficulty of being able to distinguish between a wanted speech signal and a delayed version of the same signal was overcome by careful training of the listeners – see chapter 3 for further details of this aspect of the work. All experiments used the same pre-recorded and edited speech as programme material.

The following four main variables were investigated in the first experiment: level difference; time difference; masker direction; and type of programme (male vs. female speech). The first experiment used separate loudspeakers to represent the signal and the masker in an anechoic listening room so that the relative angle between the two could be varied according to the experimental design. Listener head movement was not restricted, as in natural listening.

The second experiment was carried out to test for the separate effects of the 'silence gap' between utterances in running speech under a range of frequency bandwidth conditions for both male and female speech. Headphone listening was used because it is more convenient than loudspeaker listening when the relative angle between the masker and the direct sound is not an issue. In addition, headphone listening automatically compensates for any effect of head movement while listening.

The third and final experiment used head restrained and head movement testing, which allowed conditions over a small range of relative angles between the masker and direct sound to test for the effect of head movement under narrow angle listening conditions, and with a fixed time delay.

CHAPTER 3
METHODOLOGY

In this chapter, the research methodology used in this thesis is reviewed. The contents of this chapter consist of several parts. General research tools are first reviewed, and then research methods are discussed. The choice between the threshold measurement method and experimental design is also described and defended. Finally, limitations of research and ethical considerations were drawn.

3.1 Justification for the Methodology

Research on speech masking effects in a room can be conducted through either the study on speech intelligibility or threshold measurement, depending upon the researcher's purpose. Although researchers often use speech intelligibility in the investigation of the masking effect in a room (Knudsen, 1929; Bolt and Macdonald, 1949; Kurtović, 1975; Nábělek and Dagenais, 1986; Nábělek et al., 1989), this thesis was dedicated to using threshold measurement on the basis that the definition of masking is the threshold for one sound raised by the presence of another sound.

Measurement of masking thresholds explains the masking effect in the better way. It helps to understand a hearing phenomenon, such as the factors to address when the threshold changed.

To measure the threshold of speech signal in the presence of a delayed same speech masker and determine its effect, the basic requirement for a measurement system in this study may be summarised as follows:

- The method must give a reliable threshold level.
- It must be repeatable.
- The method must be sufficiently sensitive.
- It should take a short time to arrive at a result.

There are several psychoacoustics methods which satisfy the conditions above. The most commonly used are listed below.

1. Method of Adjustment
2. Two-Interval Forced Choice, Three-Interval Forced Choice
3. The Up-Down method or Staircase

3.2 Psychoacoustic method

3.2.1 Principle method of measurement

As was mentioned above, there are some common methods used in psychoacoustics that satisfy the requirements of this study. In this section, each method mentioned above is described in detail, including its advantages and disadvantages.

3.2.1.1 Method of adjustment

Method of Adjustment is one of the classical methods and the most basic psychoacoustic method, in which the subject has control over the stimulus. One can vary the level of stimulus continuously until it is barely audible. In practice, the stimulus control must be unlabeled; otherwise, it may provide a cue that could bias the result. Furthermore it is common practice to insert a second control between the subject's dial and the instrument, allowing the experimenter to vary the starting point of a test series.

The Method of Adjustment test gives the threshold level in a short time, but it is difficult for the experimenter to control the procedure (Zwicker and Fastl, 1990). In addition the results measured by this method might have errors because of a phenomenon known as "persistence of the stimulus" or "preservation of the response". This phenomenon is when a subject continues to turn the level down below threshold level as though the sound was still audible, causing a lower threshold level. On the other hand, the subject keeps turning the level up until the true threshold is passed.

3.2.1.2 Two- (or more) Interval Forced Choice

Two- (or more) Interval Forced Choice is an adaptive procedure which is combined with various approaches to presenting stimuli and obtaining responses. In the Two-Interval Forced Choice Procedure, the subject is presented with two intervals, and has to decide whether the signal occurs in the first or second interval. Sometimes three or four intervals may be used and the task of the subject is to decide in which interval the sound is different with respect to some quality of the stimulus. With this procedure, feedback is frequently given. This means that after each trial the subject is informed of the right answer, usually by a light indicating the interval that contained the signal.

Two- (or more) Interval Forced Choice has an advantage of reasonable precision. It does, of course, take a long time to arrive at a relevant result. Under many listening conditions studied, the test must be divided into many sessions in order to avoid fatigue.

3.2.1.3 The Up-Down method or Staircase

This method is one of the adaptive procedures. Galfand (1998) reported that this method was first introduced in 1948. It involves increasing the stimulus when the subject does not respond to the previous stimulus presentation, and decreasing the intensity when there was a response to the prior stimulus. This procedure shows some similarity to the method of tracking; however, in contrast to the method of tracking, the stimulus level is controlled by the experimenter.

The threshold is measured by starting with some level of stimulus, and then the level is lowered until it can be taken for granted that the subject does not hear the stimulus. The level is then increased until the subject clearly hears the stimulus, and after that it is again decreased. The step size is reduced with the number of reversals. When a predetermined same step size is reached, a value can be calculated with an accuracy of half the final step size by averaging the last few reversals.

This method has several advantages and limitations. It quickly converges up to the 50% point so that most trials are efficiently placed close to the point of interest. It also has an advantage of giving the tester the ability to follow changes in the subject's responses. On the other hand, the subject may bias his responses if one realised that the stimuli are presented according to a sequential rule. The other limitation is that if the step size is too small, a large number of trials are wasted; if the step size is too large, they are badly placed for estimating the 50% point.

3.2.2 Choice of a measurement method for this study

Of the measurement methods described above, only the stimulus from the method of Up-Down is controlled by the experimenter since, in the adjustment method, the stimulus is controlled by the subject. When the stimulus is controlled by the experimenter, the experimenter can control the procedure, and can follow changes in the subject's responses. In addition, the results measured by the Up-Down method should be more reliable compared to the Adjustment method because it determines the threshold level by tacking. Therefore, errors caused by "persistence of the stimulus" or "preservation of the response" can be resolved.

The major limitation of the Up-Down method is its step size, which was considered to be a minor disadvantage. In order to overcome this limitation, a predetermined small step size must be carefully designed with the pilot test.

As far as the choice of a measurement method is concerned, the Up-Down method was used in the study for measuring the threshold of the continuous signal.

3.2.3 Determination of threshold

In order to investigate the threshold level, the subject may be presented with a number of sounds of varying intensity and asked to respond in some way to those sounds. In this manner, the minimum intensity required to produce the sensation of hearing can be determined. The probability that a given auditory stimulus will produce a response depends primarily on the magnitude of the stimulus. It also depends on the disposition of the subject, namely, his alertness, his willingness and ability to listen, his motivation, and so on. Therefore, the threshold is not fixed for all time, but depends on the circumstances. In order to measure the threshold level, only an estimate made in terms of when a certain stimulus level will lead to just-audible hearing sensation. As shown in figure 3.1, both the stimulus magnitude and the sensation magnitude increase in the vertical direction. However, below threshold the stimulus does not lead to a sensation. The threshold is chosen to correspond to the probability 0.5. This means that in 50% of the trials, the 'threshold' stimulus leads to a sensation, whereas in the other 50% of the trials, no sensation is produced.

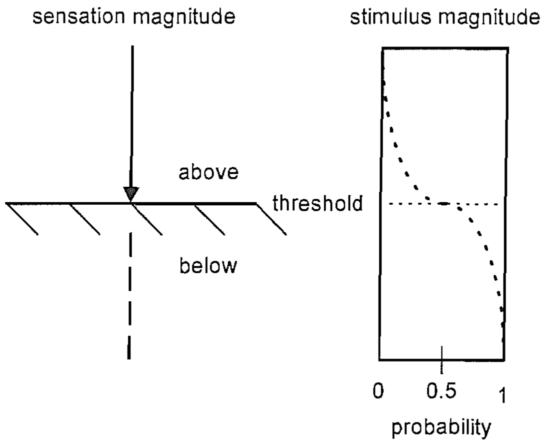


Figure 3.1: Determination of threshold for which the corresponding sensation or sensation increment is audible with a 50% probability (Zwicker and Fastl, 1990)

The Up-Down method converges on the 50% point of the psychometric function because each response leads to a decrease in the stimulus level and each non-response leads to an intensity increase. The chance of a response well below the 50% point is very small. Similarly, it is likely that stimuli presented at level above the 50% point will frequently be heard. As the intensity corresponding to 50% is approached, the chance of a response and a non-response become closer and closer. At the 50% point, the probability of a response is the same as that of a non-response. In other words, the up-down rule forces the intensity to the point on the psychometric function where the probabilities of response and non-response are equal (0.5 each).

3.2.4 Operation Procedure

In the previous section, the Up-Down method was chosen to be used in the measurement of hearing threshold. In the following section, the apparatus used for the

Up-Down method is described, including a schematic diagram. Then stimulus used is also explained, as is the calibration method and operational procedures.

3.2.4.1 Experimental Apparatus

In the present study, the investigation was conducted under two conditions - headphone and free field. Under both conditions, the main apparatus was a Kamplex AD27Audiometer.

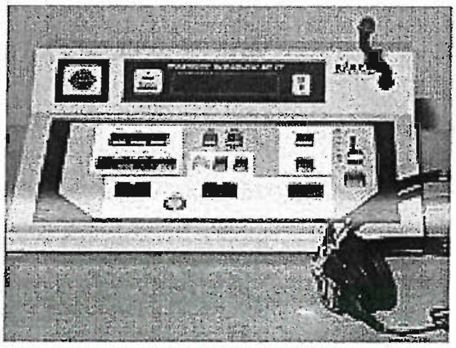


Figure 3.2: Kamplex AD27

The Kamplex AD27 is a portable audiometer mounted in a carrying case, which conveniently accommodates all its accessories, including a response push button. It was used to measure the threshold level of the signal in this experiments. In other words, the audiometer was used to increase and decrease the stimulus intensity according to a predetermined 5dB step size. In addition, this audiometer has feature that allows external input such as Minidisc or DAT to plug in, and allows external output. Therefore the audiometer output can be either headphone or external output which, when used in conjunction with external mixer and amplifiers, allows free field presentation of speech. This audiometer also comes with a response push button. When it is pushed, the red light on the audiometer is lit, indicating the response of the subject to stimulus. Therefore the experimenter can observe throughout the procedure.

In this experiment, the audiometer has been set into two different conditions. One is an experiment for listening under free field conditions, while the other is an experiment for listening under headphone conditions.

In the free field listening condition, the schematic diagrams for the apparatus setting is shown below.

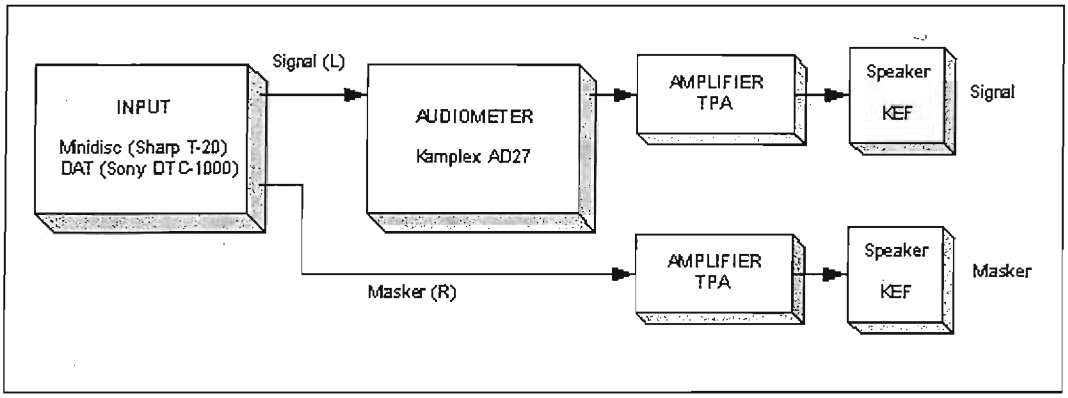


Figure 3.3: Schematic diagrams for apparatus setting for free field listening.

In figure 3.3, the source of stimulus is an input. In this study there are two input sources used. One is a Sharp T-20 Minidisc that was used as input source in experiment one, while the other a Sony DTC-1000 DAT, as shown in Figure 3.4, used in experiment 2 and 3. Using different input sources was not considered a factor that would affect the results because, with either source, the signal and the masker should have the same qualities.

Figure 3.4: DAT Sony DTC-1000

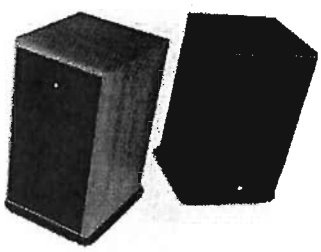


Figure 3.5: Speaker KEF

The input source was generated into two channels (see detail of stimulus in section 3.3.4.3). One is represented as the signal and the other is represented as the masker. The signal was sent to the audiometer, then to a TPA amplifier and a KEF C35 speaker. On the other hand, the masker was sent directly to a TPA amplifier and a KEF C35 speaker.

Similar to the settings of free field listening conditions, under headphone listening, the input source channel was connected to the audiometer as shown in the schematic diagram shown below.

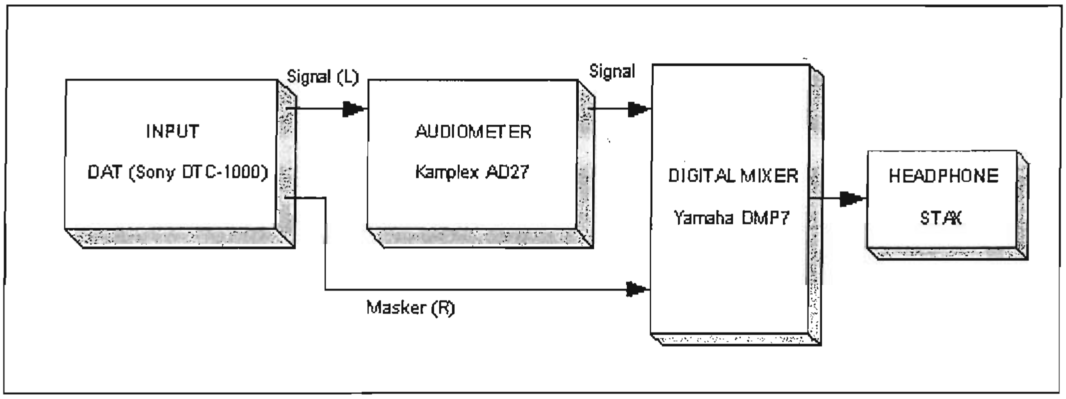


Figure 3.6: Schematic diagrams for apparatus setting under headphone listening conditions.

The signal input channel was sent to the audiometer in order to control the up-down level, and then out off the audiometer to a Yamaha DMP-7digital mixer. The masker input channel was sent directly to a digital mixer, shown in Figure 3.7.

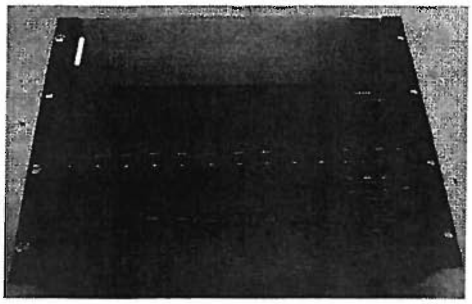


Figure 3.7: Yamaha DMP7

Figure 3.8: STAX headphones

The signal through the audiometer and masker directly from the source was mixed by a digital mixer before being presented to the subjects via the Stax headphones shown in Figure 3.8 in order to present both the signal and masker together in both ears. Not only was the mixer used in mixing the signal and masker, but it was also used in controlling the masker level.

3.2.4.2 Stimulus

Stimuli used in the present study were continuous male speech and continuous female speech. The speech chosen had to be clear speech without reverberation. Therefore running male and female speech was selected from an audio CD recorded by the BBC, as the voices of both speakers is very clear and lacks any noticeable reverberation. One track from the CD containing both male and female speech was converted into 'wave' format by using 'CDEX' software, provided by www.softseek.com, then saved on a personal computer as a *.wav file. From the wave file, both male speech and female speech were picked up and saved. It was decided to use only 2 minutes from each type of speech after it was determined to be sufficient for running the psychoacoustic method chosen. For mono speech stimulus, the left channel was always presented as the signal. On the other hand, the right channel was always presented as the masker. In order to investigate the masking effect, the signal was detected in various masker conditions as described in section 3.4.4. Both signal and masker were edited digitally by personal computer using either 'Goldwave' software or 'Wave studio' software according to their conditions. After they were edited, they were recorded on either Minidisc (Sharp T-20) or DAT (Sony DTC-1000ES) for the procedure as described in section 3.4.3.

3.2.4.2.1 Male speech and female speech

As has been stated previously, the stimuli used in this study were running male speech and running female speech, chosen for clarity and lack of reverberation. For further investigation on these speeches, it can be observed that, over time, female speech tends to leave a slightly longer silent gap between utterances than male speech does. Thereby, the female speech contained higher peaks in the long term rms (root mean square) ratio than male speech. In terms of frequency, the frequency responses for long term average of both male speech and female speeches were measured using a Hewlett-Packard 3566A frequency analyzer, with the results being shown below.

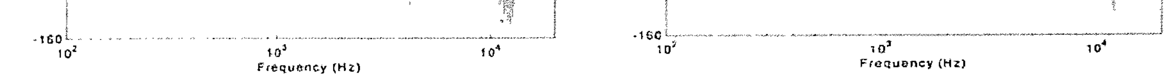


Figure 3.9: Frequency responses for male speech and for female speech

In the figures above, the x-axis represents the frequency in Hz and the y-axis represents the power spectrum in dB (ref. 1Vp-p). It can be seen that the female speech contains higher energy in almost all frequency bands. This might be due to the characteristics of female speech, which generally contains higher peaks and amplitude fluctuations.

In addition to normal male and female speech, filtered speech was also used in this study. In experiment 2, male and female speech was filtered into three frequency bands; below 800 Hz, between 800 Hz to 2,000 Hz, and above 2,000 Hz. Both male and female speech was filtered digitally using 'Wave Studio' software. In general, when any frequency filters are applied, distortion follows as a phase shift or a delay in the time domain. However, phase shift or delay may not affect the results of the study because the filtered stimuli would be used as a signal, then applied with some delays and used as a masker. However, other distortions can be minimized by selecting the correct type of filter. In this study, the Butterworth filter was chosen. This filter provided a flat response in the pass band with an adequate rate of roll off at 6dB/octave. The frequency response of Butterworth filter is shown in the figure below.

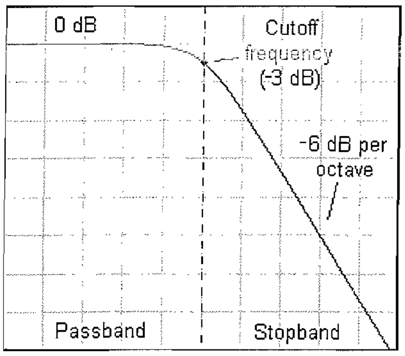
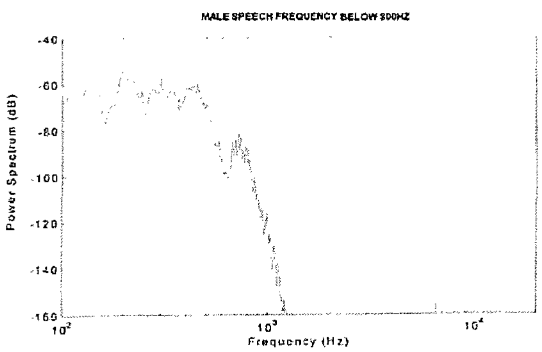
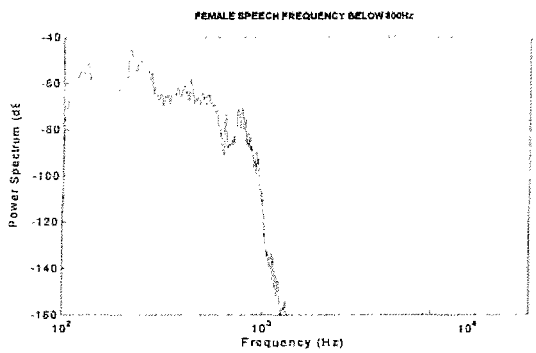


Figure 3.10: Frequency responses for Butterworth

After the male speech and female speech were filtered into frequency bands, the filter stimuli were measured for frequency response using a Hewlett-Packard 3566A frequency analyzer in order to ensure that the stimuli were filtered correctly. The filtered male and female speech is shown below.





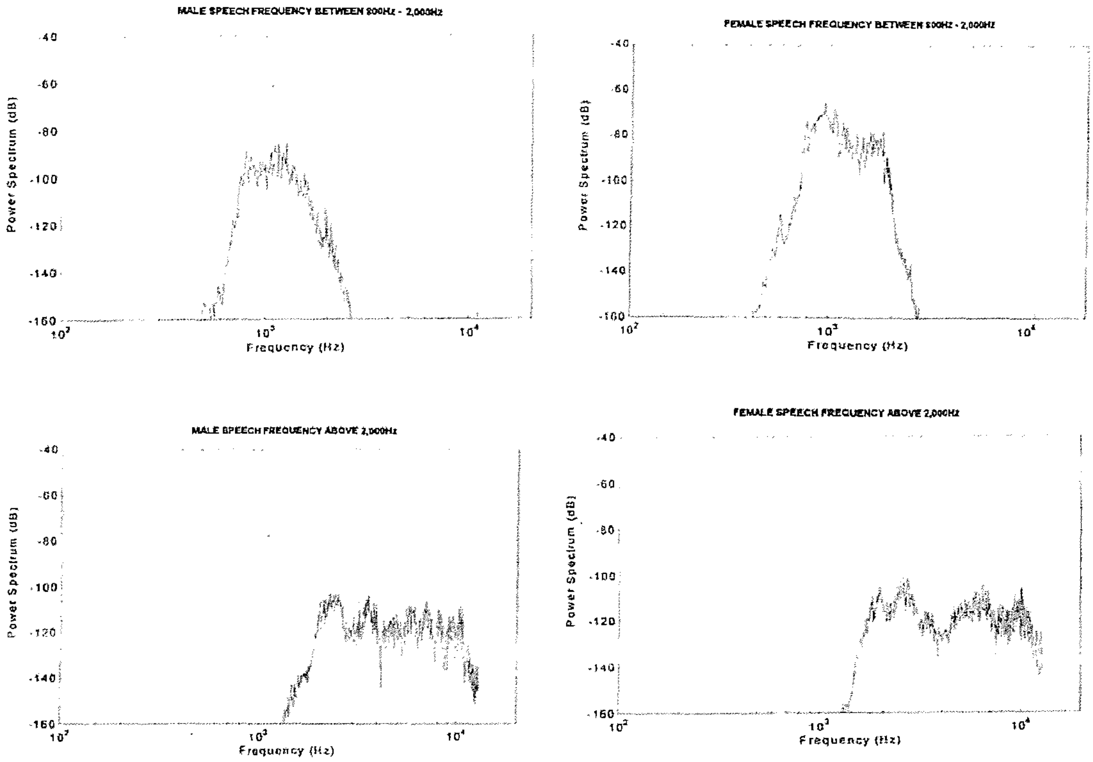


Figure 3.11: Frequency responses for filtered male and filtered female speeches

From the figure above, the x-axis represents the frequency in Hz and the y-axis represents the power spectrum in dB (ref. 1Vp-p).

3.2.4.2.2 Level

In this study, the level of male speech and female speech were studied in terms of 'equivalent continuous sound level, $L_{Aeq,T}$ (dBA)'. This continuous sound level is a notional sound level which, if maintained for a given length of time, would produce the same acoustic energy as a fluctuating noise over the same time period. It is noted that the equivalent continuous sound level should be accompanied by the time over which it was measured. It is widely used to measure any environmental sound which varies considerably with time. Therefore it was deemed suitable for measuring male speech and female speech which were time dependent. It was measured at a given length of time for one minute ($L_{Aeq,1min}$ (dBA). These measurements could be done with the customary sound level meters (sound level meter B&K type "2218" and type-2 sound level meter 'DAWE D-1421D').

In this study, the level was set at $L_{Aeq,1min}$ 40dBA, 60dBA, and 80dBA. A level as low as 40dBA was chosen because this level can be found in a normal listening room, while a level as high as 80dBA was chosen because this level can be found in a large concert hall. In addition to this, it is also considered a safe level for a listening test.

3.2.4.2.3 Time interval between signal and masker, or Delay

When the speech signal was presented together with a delayed same-speech masker, one thing that has to consider is phase. Phase is the time difference between two similar waveforms. When two signals are close in frequency and amplitude (level) but out of sync with each other, there is a phase difference. One cycle of sound is considered to have 360 degrees. When a second signal starts a half-cycle later, it is called 180 degrees "out-of-phase" with the first signal. When two signals are 180 degrees out-of-phase, the peaks of one signal are in time with the dips of the second signal and the result is a cancellation of the signals' energy. When the valleys and peaks of two signals start at the same time, the two signals are said to be "in-phase" and the energy of the two signals will double when the signals are combined.

A similar interference occurs in this study when a speech masker is delayed and played back into the original speech signal. These interference patterns are called 'Comb filters'. This is a filter that has a series of deep notches in its frequency response which correspond to multiples of the frequency of the lowest notch. When plotted on a linear frequency graph, the structure looks like a comb. The reinforcement and cancellation frequencies depend on the delay time (the time difference between the arrival time of the speech signal and the delayed speech masker). The frequency of the first cancellation occurs at $1/(2 \times t)$ Hz, where $t =$ the delay time in seconds. The cancellations are separated by $(1/t)$ Hz.

In this study, the delays chosen were 15 ms, 30 ms, 50 ms, and 100 ms. These delays were chosen to be long enough to assure incoherence of the two sounds with no coloration effects. The comb filters calculated for these delays are shown in the figures below.

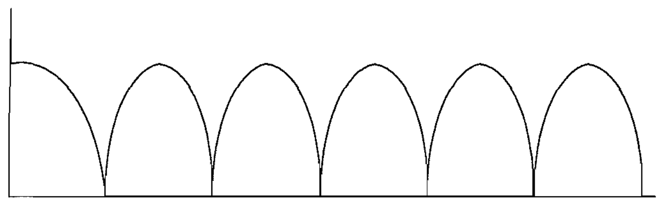


Figure 3.12: Comb filters for signal and masker with a delay of 15 ms

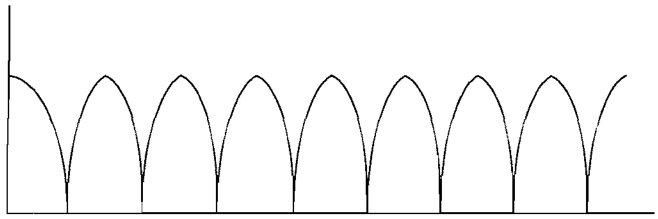


Figure 3.13: Comb filter for signal and masker with a delay of 30 ms

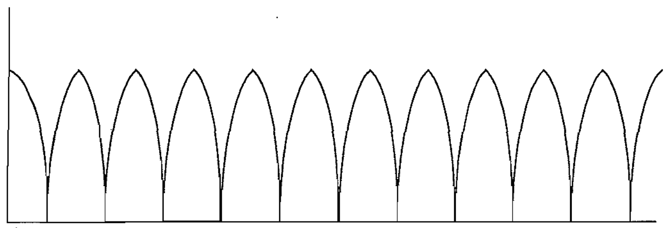


Figure 3.14: Comb filter for signal and masker with a delay of 50 ms

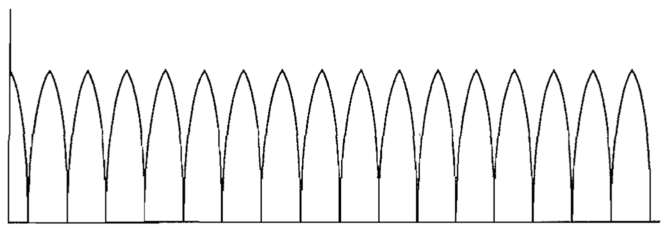


Figure 3.15: Comb filter for signal and masker with a delay of 100 ms

From the figure above, it can be seen whenever the speech signal is combined with delayed same-speech, there is always a comb filter. Fortunately, if the difference between the levels of the delayed speech masker and speech signal are kept dissimilar, there will be less cancellation (notches).

3.2.4.2.4 Azimuth angle of incidence

The azimuth angle of incidence is considered to be one of the factors that were studied under the free field listening tests. The signal was always presented to the subjects from the front, considering a 0 degrees azimuth angle, at ear level. The maskers or delayed speech was presented to the subjects at various azimuth angles on the left hand side of the subject due to the geometry of the anechoic room and the apparatus setting, also at ear level. The maskers' azimuth angles of incidence were set at 0 degrees, 30 degrees, 60 degrees, and 90 degrees in the first experiment. In the last experiment, because the experimenter would like to investigate the effect of head rotation on azimuth angle, the azimuth was chosen to be as fine as 0 degrees, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees. Consequently, the azimuth angles of incidence were set at 0 degrees, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, 60 degrees, and 90 degrees in this study.

3.2.4.3 Calibration

In this study, the level of both male speech and female speech were studied in terms of 'equivalent continuous sound level, $L_{Aeq,T}$ (dBA)' as it has been described in section 3.2.4.2.2. Three experiments have been conducted under both headphone listening conditions and free field listening conditions. In each experiment, both signal and masker were calibrated before they were presented to the subjects. There were 2 methods of calibration used in this study.

Figure 3.16: Artificial ear

Firstly, under headphone listening conditions as in experiment 2, the levels of stimuli were calibrated at 60dBA $L_{Aeq,1min}$. The main apparatus for calibration was a "B&K 4152" artificial ear. With this type of artificial ear, the socket for mounting the measuring microphone permits use of a 1" B&K type "4144" microphone condenser cartridge. The coupler supplied with artificial ear consists of a 6 cm^3 DB 0913 coupler which fulfils the requirements of the NBS 9A coupler (United States National Institute of Standards and Technology, formerly the National Bureau of Standards) and the ANSI S3.6–1969 and IEC 303 coupler for measurements on headphones. The frequency response of this artificial ear with a DB 0913 6 cm^3 coupler measured with a B&K type 4144 1" condenser microphone cartridge is shown in the figure below.

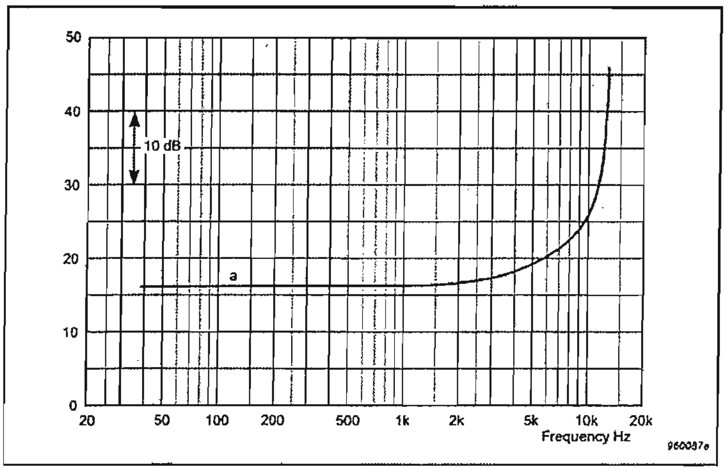


Figure 3.17: Typical frequency response curve for artificial ear type 4152 (source: B&K specification sheet)

The calibration was obtained by placing the test headphones on the artificial ear. The ear can be set up in a couple steps. First, the acoustic coupler of the artificial ear was unscrewed. Then the B&K type "4144" microphone cartridge was screwed into the microphone socket. The coupler was then replaced. Following this, a B&K type "2218" sound level meter was fitted into the socket of the main housing of the artificial ear. The ear was then ready for headphone calibration. However, the electrostatic headphones (Stax) are larger than the other headphones. Therefore it required a DB 0843 adaptor on top of the coupler.

To calibrate the levels of stimuli, their levels were adjusted at the digital mixer. Unfortunately, it had been observed that the level of each filtered stimulus were different from one another because of dynamic range of speech (see the frequency response of the filtered male speech and of the filtered female speech). In order to avoid bias from changing the level of the signal and masker in various conditions during the experiment by adjusting the gain at the mixer, the overall level (rms) of these filtered signal and masker were equalized digitally using "Gold Wave" software before re-recording on DAT (Sony DTC-1000ES). Finally, the level of each stimulus was calibrated to the reference level by adjusting at the digital mixer.

Free field listening conditions were quite dissimilar to using headphones. This method was used in experiment 1 and experiment 3 of this study. In the first experiment, the levels of stimuli were calibrated at 40dBA, 60dBA, and 80dBA. In the experiment 3, the level of stimuli were calibrated at 60dBA $L_{Aeq,1min}$. The levels of stimuli were calibrated by using a 'DAWE D-1421D' type-2 sound level meter. The sound level meter was placed at the seated subject's ear level, facing front toward the speaker. Then the level of each stimulus was calibrated to the reference level by adjusting at the digital mixer.

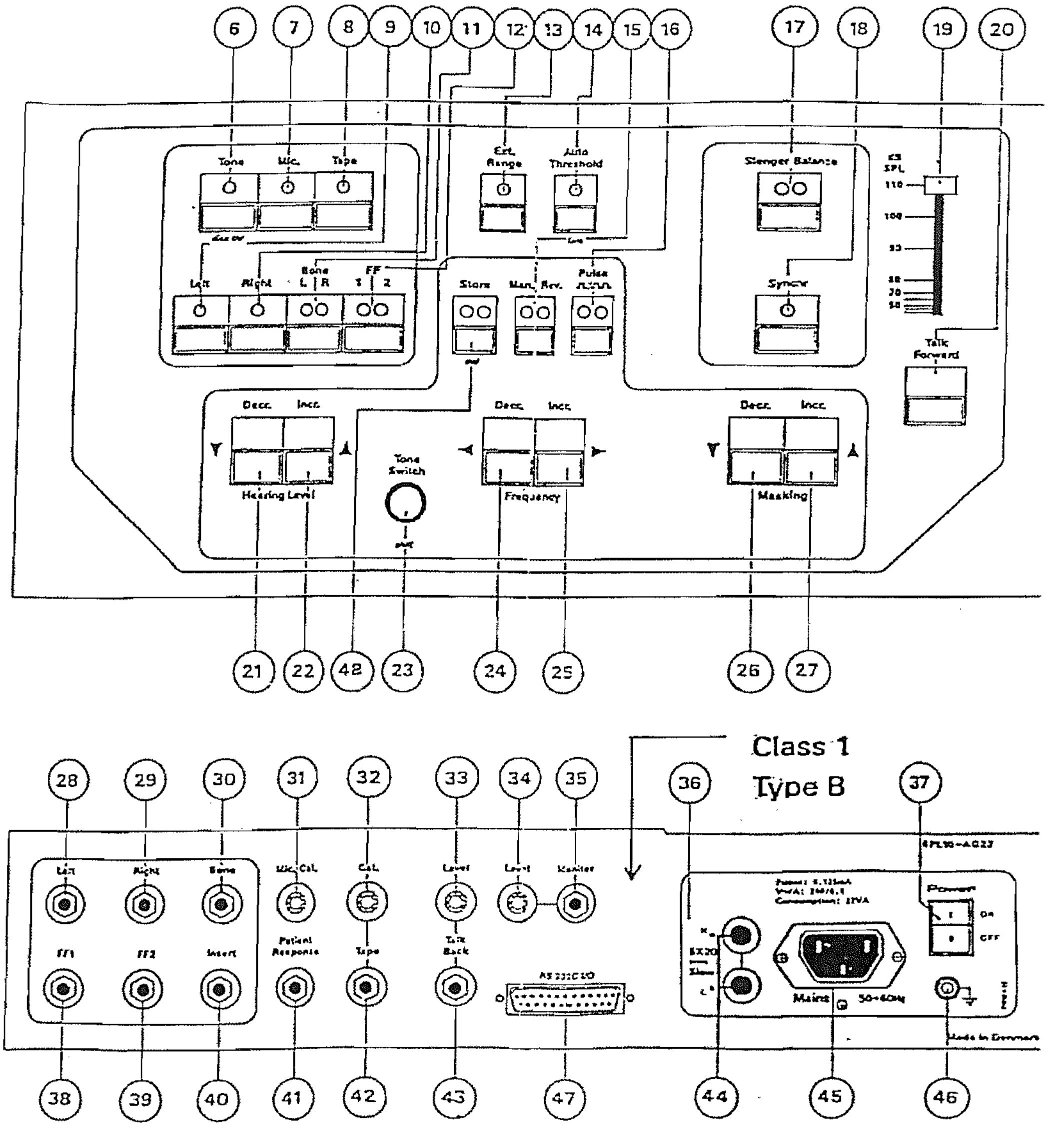


Figure 3.18: Kamplex Audiometer

3.2.4.4 Procedure

From figure 3.3 and figure 3.6, the masker stimuli were connected directly to either an amplifier and loudspeaker or a mixer and headphone. However, the signal stimuli were connected through the audiometer. The connecting of the audiometer to the other apparatus and method of operation of the procedure were as follows. The audiometer has only an input, which is 'TAPE'. It is the number '42' shown in figure 3.18. Unlike the single input, there are two outputs, labelled are 'FF1' and 'FF2'. They are the number 38 and 39 in figure 3.18. The signal stimulus was feed to the audiometer input 'TAPE'. The output from the audiometer, 'FF1', was connected to the amplifiers or the mixer.

At the beginning of the operation, the signal intensity was always presented through the audiometer at the same level as the masker intensity. Then the experimenter changed the signal intensity in order to make sure that the subject was aware of the existence of a signal. The task will be similar to a discrimination task because the subject has to identify the signal in the presence of same-speech masker. The 'response' light (number 1 in the figure 3.18) on the audiometer will light up when the subject presses the response button when they hear both signal and masker at the same time. The signal intensity was then decreased in 5dB increments. This step size was not further reduced following the actual Up-Down method because it was considered that the 5dB step size was a suitable size for the experiments from the pilot test. The decrease in signal intensity was done by the experimenter using the 'Drcr. Hearing Level' button (number 21 of figure 3.18) of the audiometer. The signal intensity was continued to drop at 5dB/second until the response light disappeared or there was no response from the subject. The level was then increased with the same step size by using the 'Incr Hearing Level' button (number 22 of figure 3.18) until the subject clearly heard the signal and responded. After that it was again decreased. Each threshold run consisted of 10 reversals. The level of the first 4 reversals was not recorded. Only the levels on the last 6 reversals were read and recorded. The mean of the levels of the last 6 reversals was taken as the threshold for that orientation.

3.3 Experimental Design

This section aims to detail research designs used in this study. The number of subjects is first described in the repeat measures design including the conditions of testing, followed by a power analysis section and how subjects were selected. Finally experimental procedure is explained.

3.3.1 Repeated measures design

In this study, the experiments were carried out using a repeated measures design, which is an experimental design that studies the same subjects at different times or under different conditions. This design was considered suitable for this study because each subject acts as their own control. In other words, this design eliminates variance caused by individual differences of subject characteristics that vary from one person to another. Furthermore, this design is very convenient, as the instructions were given to each subject only once, ensuring consistent comprehension among all subjects. Therefore the repeated measure design was used in three experiments of this study.

3.3.2 Testing conditions

As was described in the previous section, the same subjects repeatedly measured the masked threshold in various conditions. The conditions of all three experiments in this study were designed as factorial for three reasons (Shavelson, 1988). One is economy of time and resources. The second is that more information is gained from the factorial design than from the combination of one-way ANOVA designs. For example, if the study had been conducted as separate one-way ANOVA design, information about the main effects would be obtained, but information would not have been obtained about a certain combination where the largest differences between means arose. The final reason is that the error variance will be more precisely estimated in a factorial design than in one-way designs. Therefore the factorial design was used in these experiments. Each experiment was designed under different conditions as follows.

3.3.2.1 Experiment 1

The first experiment was a study on the effect of a masker's properties on a masked threshold. It has been hypothesised that each maskers' properties (type of speech, level, delay and azimuth angle) should affect the masked threshold. In order to test the hypothesis, the experiment was designed to be a 2 x 3 x 4 x 4 factorial experiment. The experiment tested the conditions of the masker as follows:

1. The type of speech masker was tested under 2 conditions – male and female speech.
2. The level of the masker was tested under 3 conditions – 40, 60, and 80 dB
3. Delay between signal and masker was tested under 4 conditions – 15, 30, 50, and 100 milliseconds.
4. The azimuth angle of masker was tested under 4 conditions – 0, 30, 60, and 90 degrees.

The testing condition for experiment 1 is shown in figure 3.19.

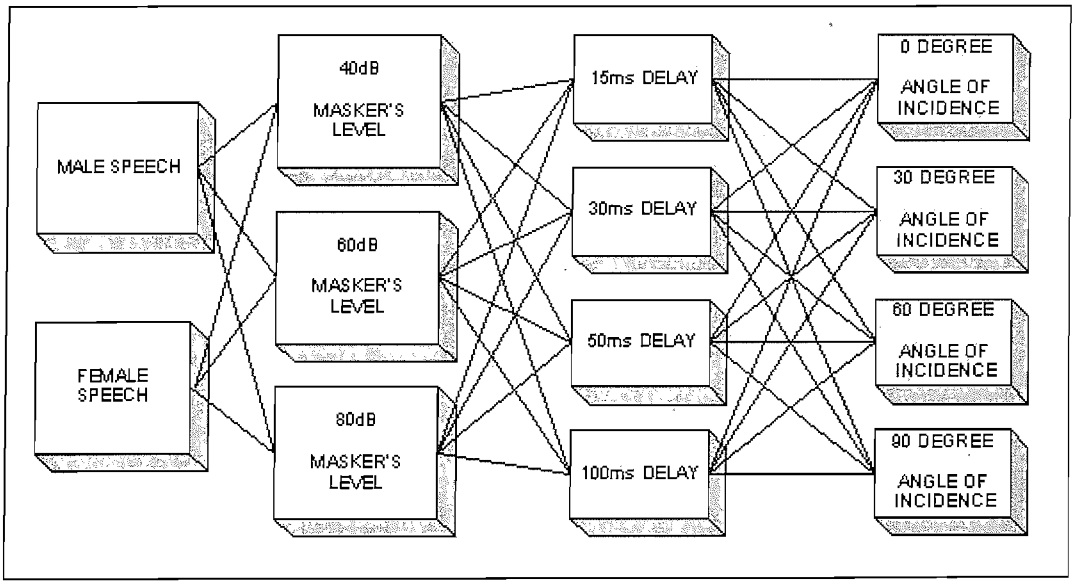


Figure 3.19: Testing condition for experiment 1

3.3.2.2 Experiment 2

The second experiment was a study on the role of simultaneous masking effect on speech threshold under headphone listening condition. It was hypothesised that simultaneous masking should play a dominant role based on the characteristics of speech under headphone listening conditions. In order to test this hypothesis, the experiment was designed in two parts.

3.4.2.2.1 Experiment 2 Part 1

In the first part, the experiment was designed to be a 2 x 2 x 3 factorial experiment. The experiment tested the conditions of the masker as follows:

1. Type of speech masker was tested under 2 conditions with male and female speech.
2. The characteristics of a speech signal and masker had been tested under 2 conditions – speech with a silent gap and speech without a silent gap.
3. Delay between signal and masker was tested under 3 conditions – 15, 30, and 50 milliseconds.

The testing conditions for experiment 2 part 1 are shown in figure 3.20.

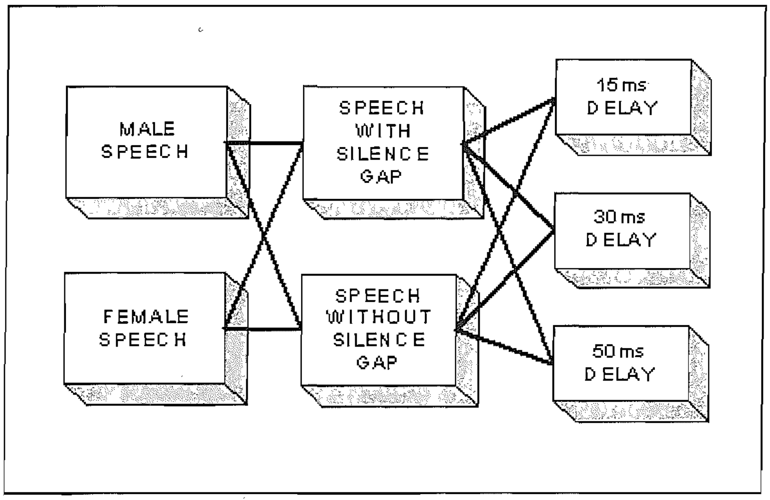


Figure 3.20: Testing conditions for experiment 2 part 1

3.4.2.2.2 Experiment 2 Part 2

In the second part, the experiment was designed to be a 2 x 9 x 3 factorial experiment. The experiment tested the conditions of the masker as follows:

1. Type of speech masker was tested under 2 conditions – male and female speech.
2. The characteristics of a speech signal and the characteristics of a speech masker were tested under 9 conditions – frequency of speech signal and frequency of speech masker that had been filtered into a *signal below 800 Hz - masker below 800 Hz,* signal below 800 Hz - masker between 800 Hz and 2000 Hz, *signal below 800 Hz - masker above 2000 Hz,* signal between 800 Hz and 2000 Hz - masker below 800 Hz, *signal between 800 Hz and 2000 Hz - masker between 800 Hz and 2000 Hz,* signal between 800 Hz and 2000 Hz - masker above 2000 Hz, *signal above 2000 Hz - masker below 800 Hz,* signal above 2000 Hz - masker between 800 Hz and 2000 Hz, and *signal above 2000 Hz - masker above 2000 Hz.*
3. The characteristics of a speech masker had been tested under 3 conditions – frequency of a speech masker that had been filtered into below 800 Hz, between 800 Hz and 2,000 Hz, and above 2,000 Hz.
4. Delay between signal and masker was tested under 4 conditions – 15, 30, and 50 milliseconds.

The testing condition for experiment 2 part 2 is shown in figure 3.21.

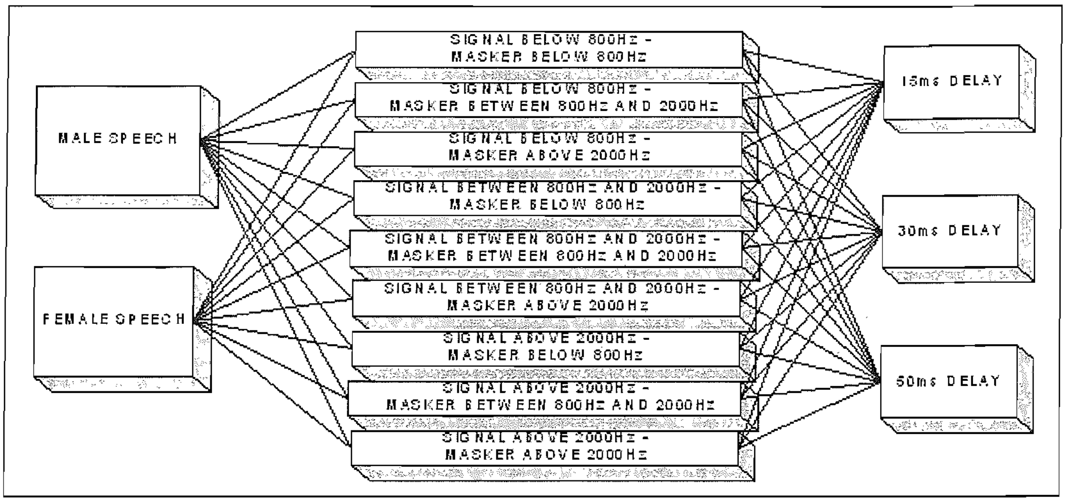


Figure 3.21: Testing conditions for experiment 2 part 2

3.3.2.3 Experiment 3

The last experiment was a study on the effect of head rotation on the speech masking threshold. It was hypothesised that head rotation should affect the masking release under free field listening conditions by changing BMLD conditions. In order to test the hypothesis, the experiment had been design to be a 2 x 9 factorial experiment. The experiment tested the conditions of the masker as follows:

1. Type of speech masker was tested under 1 condition – male speech.
2. Effects of head rotation were tested under 2 conditions – fixed and rotating.
3. Azimuth angle of masker was tested under 9 conditions – 0, 2, 4, 6, 8, 10, 15, 30, and 90 degrees.

The testing conditions for experiment 3 are shown in figure 3.22.

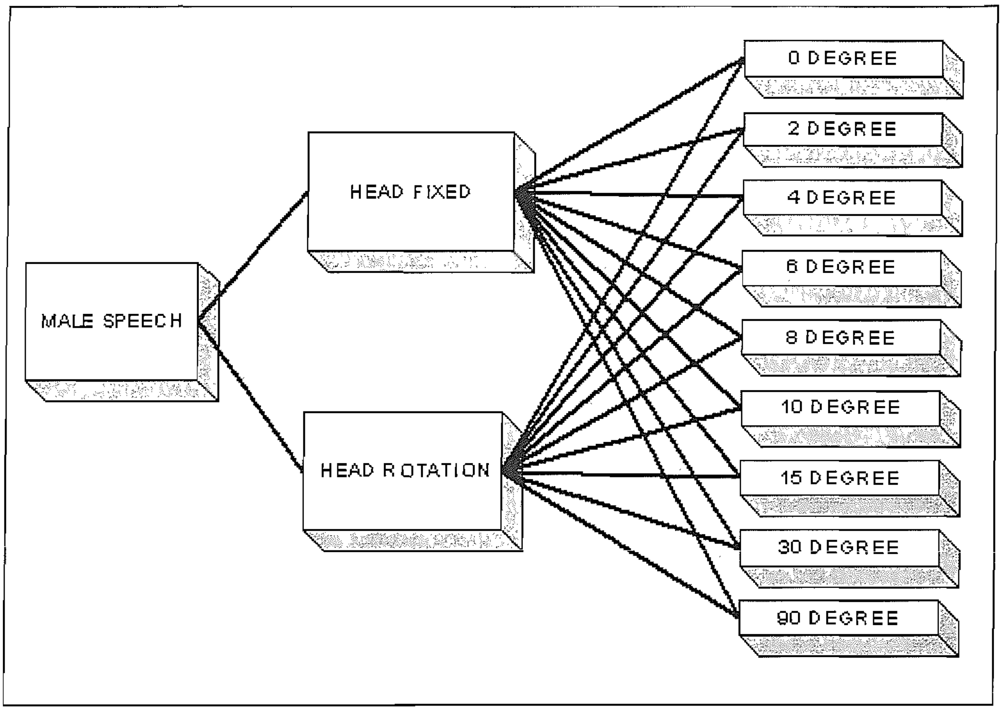


Figure 3.22: Testing condition for experiment 3

3.3.3 Sample estimation

Determining the appropriate sample size for an investigation is an essential step in the statistical design in this study. A technique for determining the sample size is statistical power analysis. Performing power analysis is important in experimental design, because without these calculations, sample size may be too high or too low. If sample size is too low, the experiment will lack the precision to provide reliable answers to the hypotheses it is testing. If sample size is too large, time and resources will be wasted, often for minimal gain.

After the experiment analysis method (as repeated measures ANOVA), was designed and planned, there remained three parameters which defined the power of the study. These are the sample size, the effect size and alpha. In considering these three parameters and the power of the study together, fixing any three will allow determination of the fourth.

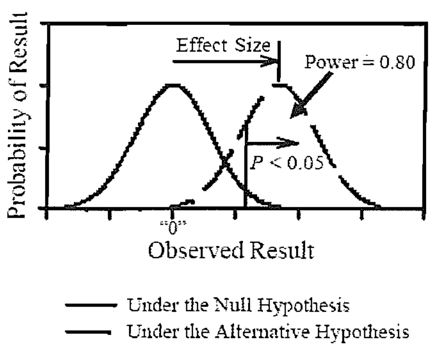


Figure 3.23: Graphic representation of power analysis (source: Lewis R., 2000)

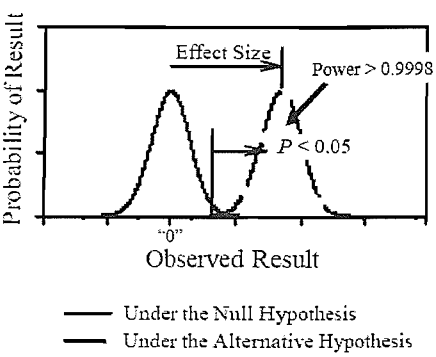


Figure 3.24: Large sample size: large power (source: Lewis R., 2000)

In the figure 3.23, the left curve shows the distribution of results expected under the null hypothesis of no difference, while the right curve shows the distribution of results expected under the alternative hypothesis. The difference in the location of the peak of these two curves is the ***effect size***. The vertical line shows the observed result which corresponds to a p-value of exactly ***alpha*** or 0.05. Any result at that location or to the right of that location will result in rejection of the null hypothesis and the conclusion that the alternative hypothesis is true. The ***power*** of the study is the area under the right hand curve which lies to the right of the vertical line, 80% under the curve, representing a power of 0.80.

If the small value for alpha is chosen while the sample size and effect size remain constant, the vertical line is moved to the right. Therefore the smaller areas under the right have curve lies to the right of the vertical line, representing the lower power. If the effect size is increased while the values for alpha and sample size remain constant, the right-hand curve is moved to the right, enlarging the area underneath the curve and right of the vertical line. Thus a larger power is achieved with a larger effect size. Unfortunately, the effect size and value of alpha are usually fixed because the effect size is usually dictated by scientific considerations, and the value of alpha is usually set at 0.05. Consequently, the only way to increase power is to increase the sample size.

In figure 3.24, the sample size is increased so that the curves representing the expected results under the null and alternative hypotheses are narrower. The criteria for statistical significance are moved to the left because of the narrower curve for outcomes under the null hypothesis. After this, the entire area under the right hand curve falls to the right of the vertical line. This means that the power of the study has been increased.

Power calculations can be done using the tables or charts provided in articles and texts (Lipsey, 1990). However these often require some hand calculations before they can be used, including interpolation between tabled values that is complicated and can give inaccurate results in some situations (Thomas and Krebs, 1997). In order to make power analysis more accurate, interactive and easy to perform, in this study, statistical power analysis was performed using statistical software called "PASS". The power packages PASS was recommended by Thomas and Krebs (1997) for beginner to intermediate level because of the accuracy and the ease of use.

In order to calculate the power using PASS software; first of all, between-subject factors and within-subject factors need to be determined. The between-subject factor is a factor that divides the sample into discrete subgroups, such as male and female. The within-subjects factor is defined for the group with the number of levels equal to the number of repetitions. After the factors are determined, the approximate mean threshold level of each group need to be filled in, and followed by their standard deviation (SD) and a value of autocorrelation. The value of autocorrelation of 0.7 was used in all experiments as it is considered to be appropriate (Machin and Campbell, 1998).

To determine the variables studied in each experiment as between subject factors or within subject factors, in general, the interested variables will be considered as between subject factors. Therefore, in fact, every variables studied shall be considered as between subject factors. Unfortunately, all variables cannot be considered as between subject factors at the same time. Therefore, when one variable is studied, the rest will be considered as within subject factor. That means if there are four variables studied, each of them will be considered as between subject factor once when the rest were considered as within subject factors. Consequently, in power calculation, only one condition was assumed when some of them were determined as between subject factors and the rest shall be considered as within subjects. In case that more than one variable is considered as between subject factors, the interaction effect of those variables can also be examined in analysis (ANOVA). However, in this study, the interaction between each variable was not interested.

To approximate mean threshold level of each group and their standard deviation, a pilot test was conducted for each experiment. The pilot test was done in the same way as the actual experiment except there are only two subjects attended the pilot test. Then the results from the pilot test were used in calculation of the approximate mean threshold level of each group and their standard deviation (SD).

After all information required was collected, the power calculation was conducted using PASS software.

3.3.3.1 Experiment 1

In the first experiment, there are four variables studied. They are speech type, masker level, delay, and the azimuth angle. The interested variable will be determined as between-subject factor and the rest will be determined as within-subjects. That means all of them shall be between-subject factors because, in this experiment, all variables are interested. Unfortunately, for PASS only two between subject factors can be considered. Therefore the design calls for two between-subject factors:

- speech type – male and female speech (A)

- masker levels of 40 dBA, 60 dBA, and 80 dBA (G)

Consequently, the other two variables studied were considered as within-subject factors:

- a delay of 15 ms, 30 ms, 50 ms, and 100 ms (D)
- an azimuth angle of 0 degrees, 30 degrees, 60 degrees, and 90 degrees (M).

A pilot test was conducted according the test condition described in section 3.3.2.1. Following the results of pilot test (See the result in appendix 4); when the speech type was considered, the desired means thresholds of speech type groups (male and female) were 40dB and 31dB. When the masker level was considered, the desired means thresholds of masker level groups (40dB, 60dB, and 80dB) were 53dB, 38dB, and 27dB. When the delay was considered, the desired means thresholds of delay groups (15ms, 30ms, 50ms, and 100ms) were 50dB, 43dB, 37dB, and 28dB. And when the azimuth angle of incidence was considered, the desired means thresholds of azimuth angle groups (0degree, 30degrees, 60degrees, and 90degrees) were 50dB, 40dB, 36dB, and 35dB.

To specify the covariance matrix it was found from the pilot test that a value of 15 is appropriate for Standard deviation (SD). An autocorrelation of 0.7 was used as it is considered to be appropriate (Machin and Campbell, 1998).

Finally, the power was calculated using the Greenhouse-Geisser F test (GG F) at the following sample sizes: 2, 3, 4, and 5. The power calculations with plot are shown in table 3.1 and figure 3.25.

Table 3.1: Table of power calculations for experiment 1

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
A(2)	GG F	0.5789	3	18	1.00	4.50	8.13	0.55	0.0500	0.4211
G(3)	GG F	0.9945	3	18	1.00	10.66	8.13	1.31	0.0500	0.0055
D(4)	GG F	0.9986	3	18	1.00	8.08	5.37	1.50	0.0500	0.0014
M(4)	GG F	1.0000	3	18	1.00	5.93	2.17	2.74	0.0500	0.0000
A(2)	GG F	0.8288	5	30	1.00	4.50	8.13	0.55	0.0500	0.1712
G(3)	GG F	1.0000	5	30	1.00	10.66	8.13	1.31	0.0500	0.0000
D(4)	GG F	1.0000	5	30	1.00	8.08	5.37	1.50	0.0500	0.0000
M(4)	GG F	1.0000	5	30	1.00	5.93	2.17	2.74	0.0500	0.0000

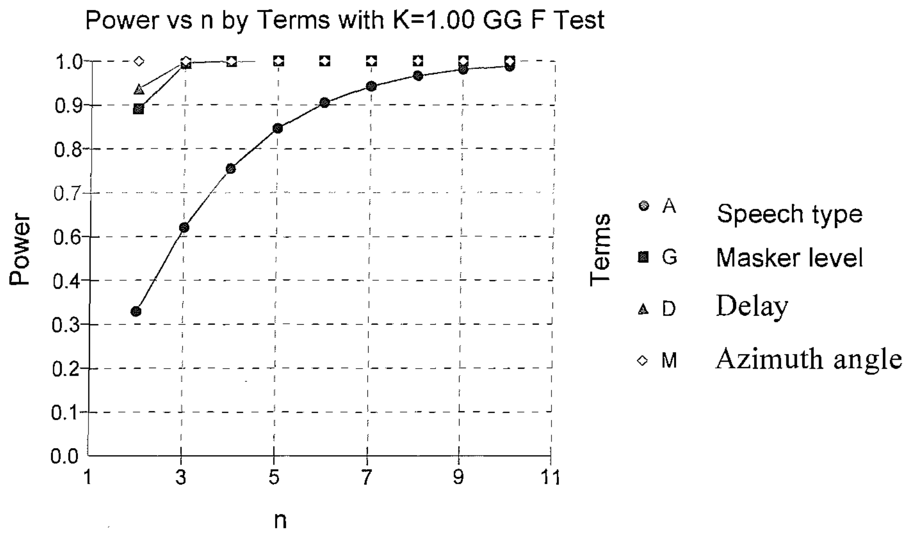


Figure 3.25: Results of power analysis for experiment 1

From the results of power calculation, repeated measured designs with 2 between-factors and 2 within factor had 6 groups with 5 subjects each, for a total of 30 subjects. Each subject is measured 16 times. This design achieves 83% power to test factor A (speech type) if a Greenhouse-Geisser corrected F test is used, with a 5% significance level and the actual effect standard deviation is 8.13 (an effect size of 3.55), achieves 100% power to test factor G (masker level) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 8.13 (an effect size of 1.31), achieves 100% power to test factor D (delay) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 5.37 (an effect size of 1.50) and achieves 100% power to test factor M (azimuth angle of incidence) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 2.17 (an effect size of 2.74).

3.3.3.2 Experiment 2

In experiment 2, there are two parts. Power calculations have been conducted in both parts.

In the first part of this experiment, there are three variables studied. They are speech type, delay, and silence gap. The interested variable will be determined as between-subject factor and the rest will be determined as within-subjects. That means all of them shall be between-subject factors because, in this experiment, all variables are interested. Unfortunately, for PASS only two between subject factors can be considered. Therefore the design calls for two between-subject factors:

- speech type – male and female speech (A)
- a delay of 15 ms, 30 ms, and 50 ms (G)

Consequently, the other one variable studied was considered as within-subject factors:

- with silence gap and without silence gap (D)

A pilot test was conducted according the test condition described in section 3.3.2.2. Following the results of pilot test (See the result in appendix 4); when the speech type was considered, the desired means thresholds of speech type groups (male and female) were 45dB and 40.5dB. When the delay was considered, the desired means thresholds of delay groups (15ms, 30ms, and 50ms) were 49dB, 44dB, and 38dB. And when condition of silence gap was considered, the desired means thresholds between with silence gap and without silence gap were 42dB and 40dB.

To specify the covariance matrix it was found from the pilot test that a value of 5 is appropriate for Standard deviation (SD). An autocorrelation of 0.7 was used as it is considered to be appropriate (Machin and Campbell, 1998).

Finally, the power was calculated using the Greenhouse-Geisser F test (GG F) at the following sample sizes: 2, 3, 4, 5, 6, and 7. The power calculations with plot are shown in table 3.2 and figure 3.26.

Table 3.2: Table of Power calculation for experiment 2 part 1

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
A(2)	GG F	0.4779	3	18	1.00	2.25	4.61	0.49	0.0500	0.5221
G(3)	GG F	0.9111	3	18	1.00	4.50	4.61	0.98	0.0500	0.0889
D(2)	GG F	0.5217	3	18	1.00	1.00	1.94	0.52	0.0500	0.4783
A(2)	GG F	0.7275	5	30	1.00	2.25	4.61	0.49	0.0500	0.2725
G(3)	GG F	0.9965	5	30	1.00	4.50	4.61	0.98	0.0500	0.0035
D(2)	GG F	0.7745	5	30	1.00	1.00	1.94	0.52	0.0500	0.2255
A(2)	GG F	0.8089	6	36	1.00	2.25	4.61	0.49	0.0500	0.1911
G(3)	GG F	0.9994	6	36	1.00	4.50	4.61	0.98	0.0500	0.0006
D(2)	GG F	0.8503	6	36	1.00	1.00	1.94	0.52	0.0500	0.1497

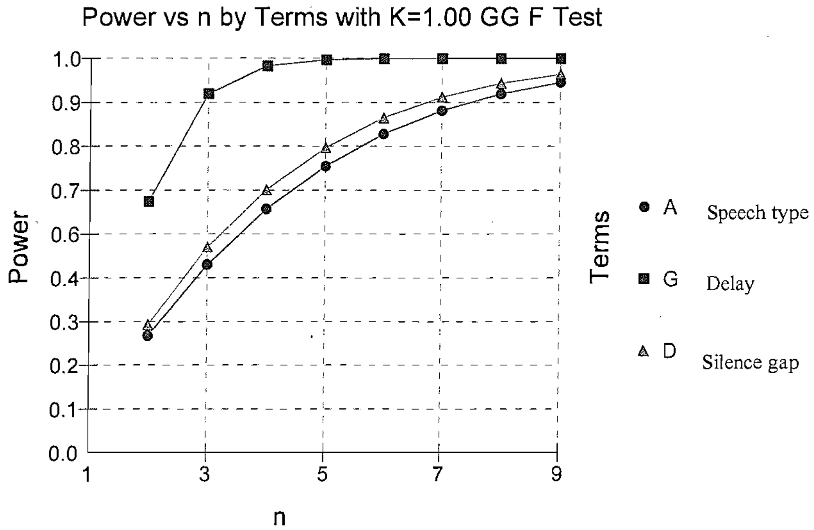


Figure 3.26: Results of power analysis for experiment 2 part 1

From the results of power calculation, a repeated measure design with 2 between-factors and 1 within-factor has 6 groups with 6 subjects each, for a total of 36 subjects. Each subject was measured 2 times. This design achieves 81% power to test factor A (speech type) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 4.61 (an effect size of 0.49), achieves 100% power to test factor G (delay) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 4.61 (an effect size of 0.98), and achieves 85% power to test factor D (silence gap) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 1.94 (an effect size of 0.52).

In the second part of this experiment, there are three variables studied. They are speech type, frequency of signal and frequency of masker, and delay. The interested variable will be determined as between-subject factor and the rest will be determined as within-subjects. That means all of them shall be between-subject factors because, in this experiment, all variables are interested. Unfortunately, for PASS only two between subject factors can be considered. Therefore the design calls for two between-subject factors:

- speech type – male and female speech (A)
- frequency of signal and frequency of masker - low-low, low-mid, low-high, mid-low, mid-mid, mid-high, high-low, high-mid, high-high (G)

Consequently, the other one variable studied was considered as within-subject factors:

- a delay of 15 ms, 30 ms, and 50 ms (D)

A pilot test was conducted according the test condition described in section 3.3.2.2. Following the results of pilot test (See the result in appendix 4); when the speech type was considered, the desired means thresholds of speech type groups (male and female) were 29dB and 24dB. When the frequency of signal and frequency of masker was considered, the desired means thresholds of this groups (low-low, low-mid, low-high, mid-low, mid-mid, mid-high, high-low, high-mid, high-high) were 44dB, 41dB, 39dB, 24dB, 20dB, 18dB, 16dB, 15dB, and 14dB. And when condition of delay was considered, the desired means thresholds between delay groups (15ms, 30ms, and 50ms) were 28dB, 25.5dB, and 23dB.

To specify the covariance matrix, it was found from the pilot test that a value of 11 is appropriate for SD1. An autocorrelation of 0.7 is used as it is considered to be appropriate (Machin and Campbell, 1998).

Finally, the power was calculated using the Greenhouse-Geisser F test (GG F) at the following sample sizes: 2, 3, 4, 5, 6, and 7. The power calculations with plot are shown in table 3.3 and figure 3.27.

Table 3.3: Table of Power calculation for experiment 2 part 2

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
A(2)	GG F	0.4653	3	54	1.00	2.50	9.55	0.26	0.0500	0.5347
G(9)	GG F	1.0000	3	54	1.00	11.48	9.55	1.20	0.0500	0.0000
D(3)	GG F	0.9079	3	54	1.00	2.04	3.86	0.53	0.0500	0.0921
A(2)	GG F	0.6883	5	90	1.00	2.50	9.55	0.26	0.0500	0.3117
G(9)	GG F	1.0000	5	90	1.00	11.48	9.55	1.20	0.0500	0.0000
D(3)	GG F	0.9918	5	90	1.00	2.04	3.86	0.53	0.0500	0.0082
A(2)	GG F	0.8297	7	126	1.00	2.50	9.55	0.26	0.0500	0.1703
G(9)	GG F	1.0000	7	126	1.00	11.48	9.55	1.20	0.0500	0.0000
D(3)	GG F	0.9994	7	126	1.00	2.04	3.86	0.53	0.0500	0.0006

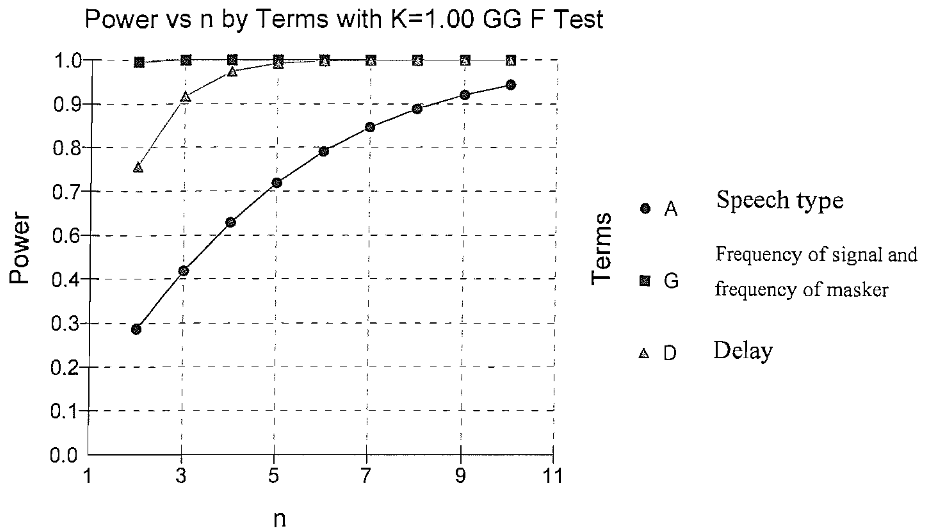


Figure 3.27: Results of power analysis for experiment 2 part 2

From the results of power calculation, repeated measured design with 2 between-factors and 1 within factor has 18 groups with 7 subjects each for a total of 126 subjects. Each subject is measured 3 times. This design achieves 83% power to test factor A (speech type) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 9.55 (an effect size of 0.26), achieves 100% power to test factor G (frequency of signal and frequency of masker) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 9.55 (an effect size of 1.20), and achieves 100% power to test factor D (delay) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 3.86 (an effect size of 0.53).

3.3.3.3 Experiment 3

In the last experiment, there are two variables studied. They are azimuth angle and head rotation. The interested variable will be determined as between-subject factor and the rest will be determined as within-subjects. Therefore the design calls for two between-subject factors:

- an azimuth angle of 0 degree, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees (A)

Consequently, the other variables studied were considered as within-subject factors:

- Head fixed and head rotated (D)

A pilot test was conducted according the test condition described in section 3.3.2.3. Following the results of pilot test (See the result in appendix 4); when the azimuth angle was considered, the desired means thresholds of azimuth angle groups (0 degree, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees) were 52dB, 52dB, 49dB, 46dB, 44dB, 42.5dB, 41dB, 39dB, and 37dB. And when head rotation was considered, the desired means thresholds between head fixed and head rotated were 47dB, and 42.5dB.

To specify the covariance matrix, from the pilot test, it was found that a value of 10 is appropriate for SD1. An autocorrelation of 0.7 is used as it is considered to be appropriate (Machin and Campbell, 1998).

Finally, the power was calculated using the Greenhouse-Geisser F test (GG F) at the following sample sizes: 2, 3, 4, 5, 6, and 7. The power calculations with plot are shown in table 3.4 and figure 3.28.

Table 3.4: Table of power calculations for experiment 3

Term	Test	Power	n	N	Multiply Means By	SD of Effects (Sm)	Standard Deviation (Sigma)	Effect Size	Alpha	Beta
A(9)	GG F	0.3216	3	27	1.00	4.99	9.22	0.54	0.0500	0.6784
D(2)	GG F	0.8145	3	27	1.00	2.25	3.87	0.58	0.0500	0.1855
A(9)	GG F	0.6258	5	45	1.00	4.99	9.22	0.54	0.0500	0.3742
D(2)	GG F	0.9665	5	45	1.00	2.25	3.87	0.58	0.0500	0.0335
A(9)	GG F	0.8306	7	63	1.00	4.99	9.22	0.54	0.0500	0.1694
D(2)	GG F	0.9949	7	63	1.00	2.25	3.87	0.58	0.0500	0.0051

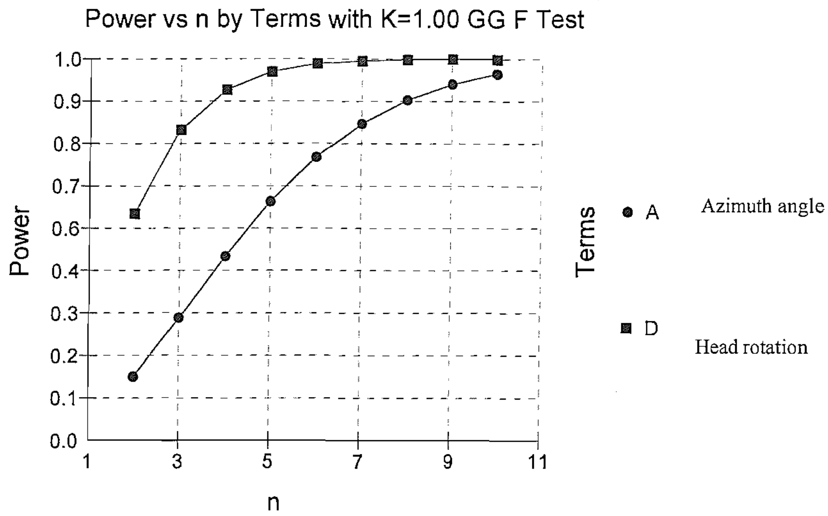


Figure 3.28: Results of power analysis for experiment 3

From the results of power calculation, repeated measured design with 1 between-factor and 1 within-factor had 2 groups with 7 subjects each, for a total of 63 subjects. Each subject was measured 9 times. This design achieves 83% power to test factor A (head rotation) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 9.22 (an effect size of 0.54), and achieves 99% power to test factor D (Azimuth angle of incidence) if a Greenhouse-Geisser corrected F test is used with a 5% significance level and the actual effect standard deviation is 3.87 (an effect size of 0.58).

Following the results of statistical power calculated shown in the graph, the summary of the results are shown in table 3.5. This table shows the number of subjects required.

Table 3.5: The total subjects used in this study

	Experiment 1	Experiment 2		Experiment 3
		Section 1	Section 2	
Number of subject from PASS	5	6	7	7
Available Number of subject in this study	5	7	7	10

In summary, judging from the power calculation, the subjects required for each experiment were:

- First experiment – 5
- Second experiment – Section one – 6
  Section two – 7
- Third experiment - 7

Consequently, the number of subjects who attended the test were:

- First experiment – 5
- Second experiment – Section one – 7
  Section two – 7
- Third experiment - 10

In the last experiment, more subjects attended the test simply because more were available.

3.3.4 Subjects selection

All subjects were students and staff at Southampton University. They were both male and female between the ages of 18 and 35. However, sex and age were not the main criteria for selection of the subjects in this study. For this study, the only criteria that was set was their hearing threshold. In other words, all subjects must have a pure-tone hearing level of 20dB or less at octave frequencies from 250 to 8000 Hz, which is considered as normal hearing, according to the classification of degrees of hearing loss by Jerger (Katz, 1985). The procedure for testing the hearing capabilities for each subject is described in the next section.

3.3.4.1 Hearing test

The hearing tests in this experiment were conducted in order to screen the subjects and to make sure that all of them have normal hearing. All subjects were tested inside an acoustical booth where the background noise level was as low as 8dB. The apparatus and hearing test procedure is shown in section 3.3.4.1.1 and section 3.3.4.1.2.

3.3.4.1.1 Apparatus

The main apparatus used for the hearing test was a Kamplex AD27 audiometer with TDH39 headphones and a response push button as described in section 3.2.4.1

3.3.4.1.2 Hearing test procedure

For the hearing test, the experimenter first gave the response task instructions to the subjects. The instructions stated that the subjects had to respond by pressing the response button as soon as the tone is heard, no matter how faint it may be. The subjects were told to hold the button down for as long as the tone lasts; when the tone stops, they were to release the button immediately.

After the instruction were read an understood, earphones were held in place by a headband with the earphone grid directly over the entrance to the ear canal of the subject. The earphone was put in place by the experimenter, not the subjects. For female subjects, long hair and other obstacles was cleared away.

When the subject had the earphones properly in place, the AD27 started performing threshold determination automatically. The test procedure is based on the Hughson-Westlake method (up 5dB, down 10dB) and conforms to ISO8253. The results of thresholds were obtained by the storage of audiometric data in an internal memory.

3.3.5 Experimental procedure

In this study, the experiments were conducted under both free field listening conditions and headphone listening conditions. Under the free field condition, the experiments were conducted in an anechoic room (4.33 m wide X 4.95 m long X 2.74 m high) with a background noise level of L_{Aeq} 8-12 dBA. The subject was seated on a chair at the centre of the loudspeaker array at a radian of 2 meters and was positioned in such a way that they could not observe the experimenter manipulating the controls, as shown in Figure 3.29.

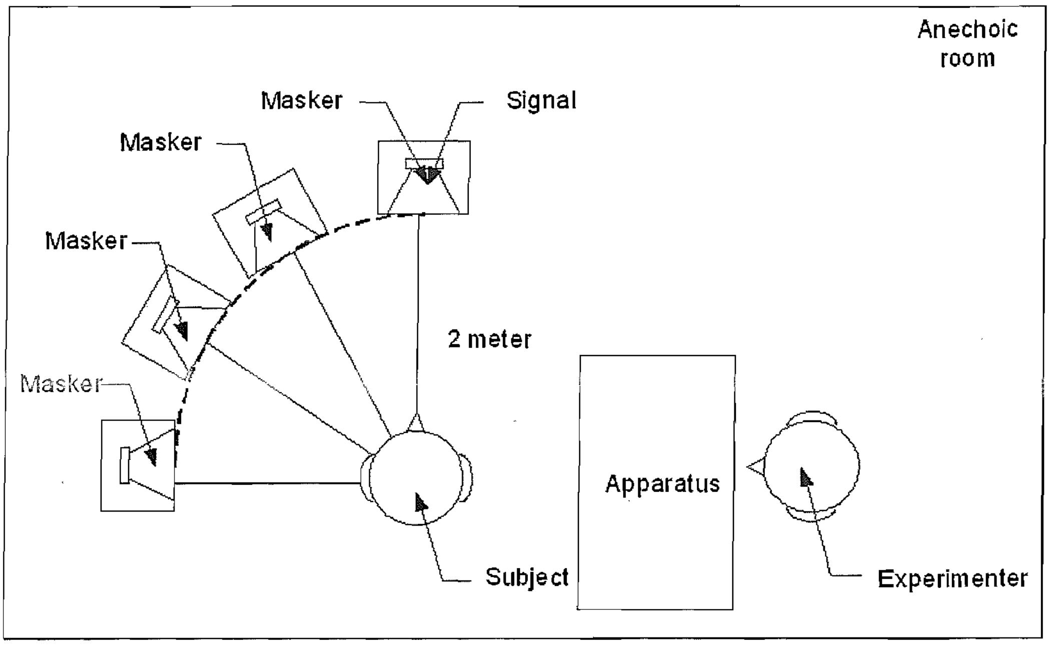


Figure 3.29: Diagram of apparatus set up in Anechoic room

Under headphone conditions, the experiment was conducted in a sound attenuating booth that had a background noise level of L_{Aeq} 10-12 dBA. The subject was also positioned in such a way that they could not observe the experimenter manipulating the controls, as shown in Figure 3.30.

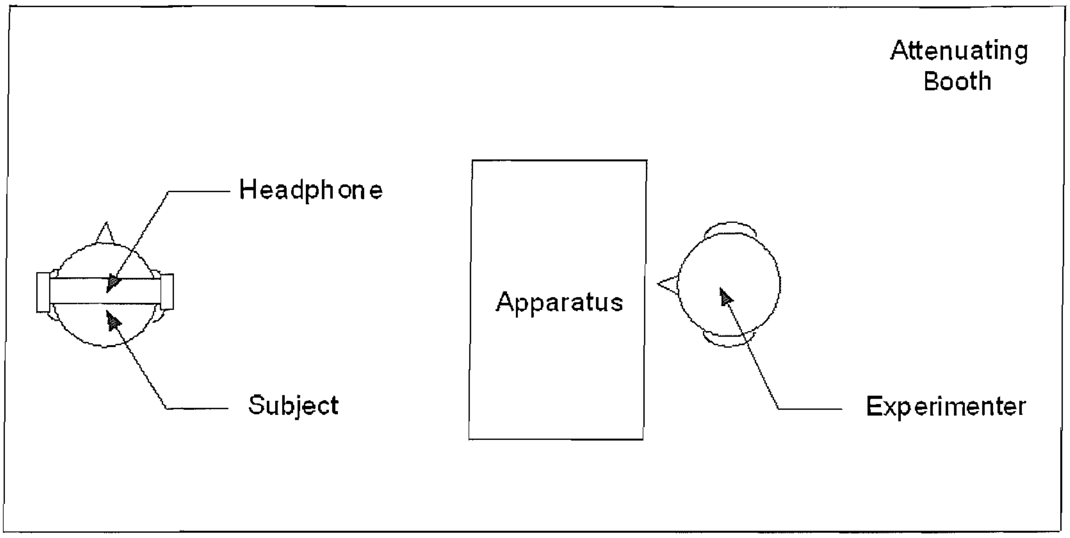


Figure 3.30: Diagram of apparatus set up in attenuating booth

After the subjects were seated, they were instructed to press the response button when they heard, or thought that they heard, the speech signal. In order to make it clearer, under free field listening, the subjects were instructed to detect the only speech signal that emanated from the front loudspeaker. They shall press the response button when then heard, or thought that they heard the speech from the front loudspeaker. For the experiment conducted under headphone listening, it was more complicated. The subjects were instructed to press the response button when they heard, or thought that they heard, two speech stimuli played together. In the other words, the subjects were instructed to press the response button or release the response button when they detect some changes in testing stimuli. For both listening conditions, prior to the actual tests, subjects were trained to make sure that they were familiar with the procedure. However, under headphone listening conditions, subjects were sometimes confused. Further instructions were often provided during the course of testing, based on each subject's responses to the stimuli.

In every experiment, the absolute threshold of both male a female speech stimuli, were always measured at the beginning of the first section.

For each experiment, if the duration of the testing was too long, the tests were divided into two sessions, including a short break, in order to avoid fatigue. In the first and the second experiment, the test took about 2 hours in total, depending on the speed of the subject. A 2-hour test was considered too long, therefore the tests were divided into two sessions of 1 hour each, and each test conducted on a different day. For the last experiment, the test took about 40 minutes, which was considered manageable. This test could be conducted in one day in one session, but a short break was allowed in order to avoid fatigue.

3.3.5.1 Experiment 1

In the first experiment, the examiner started testing with the masker condition for each subject according to the table shown in section 3.3.1. Although each subject started the test with a different level and type of speech, they all started of with a delay of 15 ms at an incidence angle of 90 degrees. Under each testing condition, the initial level of the signal was set at 0 dB. This level was adjusted upward in 5dB increments until the subject responded. As soon as the subject responded, the level of the signal was adjusted downward in 5 dB steps. As soon as the subject stopped responding, the level of the signal is adjusted upward again. This procedure is repeated. Each threshold run consisted of 10 reversals. During the first 4 reversals, the response levels were not recorded. Only the levels of the last 6 reversals was read and recorded.

3.3.5.2 Experiment 2

In experiment 2, each session contained 34 conditions across both sections of the experiment. In order to avoid a learning effect, the subjects were separated into 4 groups (2 subjects, 2 subjects, 2 subjects, and 1 subject). In the first group, the subject started the test with male speech in part one of the experiment, followed by male speech in part two of the experiment. In the second group, the subject started the test with female speech in part one of the experiment, followed by female speech in part two of the experiment. In the third group, the subject started the test with male speech in part two of the experiment, followed by male speech in part one of the experiments. The last subject started the test with female speech in part two of the experiment, followed by female speech in part one of the experiments.

3.3.5.3 Experiment 3

In last experiment, there were 18 conditions. In order to avoid a learning effect, subjects were separated into 2 groups (5 subjects in each group). In the first group, the subject started the test under head-fixed conditions followed by head-rotated conditions. In the other group, the subject started the test under head-rotated conditions followed by head-fixed conditions. The experiment was conducted under these circumstances in order to avoid the learning effect.

3.3.5.3.1 Head rotation

The subject had to attend two sessions in this study; head-fixed and head-rotated. In the head-fixed session, subjects were instructed to keep the head still by pushing their head against the head rest. Subjects were also instructed to report to the experimenter if they noticed that they might have moved their head even slightly. In the head-rotated session, subjects were always confused because they did not know how much they had to move their head. Therefore the subjects were instructed to move their head around the area between the sound source and the masker positions.

3.4 Data Collection

There were three types of results collected in this study – the threshold of hearing for each subject, the absolute (unmasked) threshold for speech, and the masking threshold.

The threshold of hearing and the masking thresholds for each subject were measured as explained in section 3.3.4.2.1 in order to screen the subject before starting the experiment. The data of threshold of hearing measured was noted on the results sheet (see appendix 3).

Each threshold run consisted of 10 reversals. The level of the first 4 reversals was not recorded. Only the levels of the last 6 reversals were read and recorded. The mean of the levels of the last 6 reversals was taken as the threshold for that orientation as explained in section 3.2.3.

3.5 Statistical Analysis

Data collected in each experiment had been statistically analyzed by SPSS software (Statistical Package for Social Science) version 11.5. that the resulting information was used for discussion and conclusions.

There are three statistics tests that were used in this study – test of normality, analysis of variance in repeated measure design, and compared means test (Shavelson, 1988).

3.5.1 Test of Normality

Test for normality are important because many data analysis methods that are parametric (repeated measure ANOVA and paired samples t-test) depend on the assumption that data was sampled from a normal distribution. If a parametric test, such as a simple pair t-test, is applied to non-normal distribution data, it will probably increase the risk of error. Therefore, the test for normality shall be applied first. There are 2 types of popular normality tests. They are Kolmogorov-Smirnov test and Shapiro-Wilk test. The Kolmogorov-Smirnov test compares the cumulative distribution of the data with the expected cumulative normal distribution, and bases its P value on the largest discrepancy. Unfortunately, Kolmogorov-Smirnov isn't very useful in practice because it requires a simple null hypothesis; that is, the distribution must be completely specified with all parameters known. When the null hypothesis does not specify the expected value and variance, it is necessary to apply the Lilliefors correction to the Kolmogorov-Smirnov test. Sometime it is called "Lilliefors test. The Kolmogorov-Smirnov Test with Lilliefor correction is always used when sample size is large, say greater than 2000. For the Shapiro-Wilk normality test, it was similar to the Kolmogorov-Smirnov Test in the sense that it was designed to test the normality, where P values will answer the question. However, the Shapiro-Wilk Test is suitable for small sample size, between 3 and 2000. Therefore, in this study, the test for normality used was the Shapiro-Wilk Test because the sample size in this study was smaller than 2000 in every experiment.

3.5.2 Analysis of variance in repeat measure (ANOVA)

The second statistical analysis used in this study is the "Repeated measure ANOVA". The Repeated measures procedure provides analysis of variance when the same measurement is made several times on each subject. If between-subject factors are specified, they divide the population into groups. Using this general linear model procedure, the null hypotheses can be tested on the effects of both the between-subject factors and the within-subject factors.

A repeated measures analysis can be approached in two ways. These are univariate and multivariate. The univariate approach considers the dependent variables as responses to the levels of within-subjects factors. The measurements on a subject should be a sample from a multivariate normal distribution, and the variance-covariance matrices are the same across the cells formed by the between-subjects effects. Certain assumptions are made on the variance-covariance matrix of the dependent variables. The validity of the F statistic used in the univariate approach can be assured if the variance-covariance matrix is circular in form.

To test this assumption, Mauchly's test of sphericity can be used. Mauchly's test will perform a test of sphericity on the variance-covariance matrix of an ortho-normalized transformed dependent variable. Mauchly's test is automatically displayed for repeated measures analysis in SPSS. For small sample sizes, this test is not very powerful. For large sample sizes, the test may be significant even when the impact of the departure on the results is small. If the significance of the test is large, the hypothesis of sphericity can be assumed. However, if the significance is small and the sphericity assumption appears to be violated, a correction is required in order to validate the univariate F statistic. The most well known corrections are those developed by Greenhouse and Geisser (the Greenhouse-Geisser correction) and Huynh and Feldt (the Huynh-Feldt correction). Each of these corrections works in roughly the same way. They attempt to adjust the degrees of freedom in the repeated measure ANOVA test in order to produce a more accurate significance (p) value. If sphericity is violated, the p values need to be adjusted upward. The first step in each test is to estimate something called "epsilon". The epsilon is considered to be a descriptive statistic indicating the degrees to which sphericity has been violated. If the sphericity is perfectly met, then epsilon will be 1. If epsilon is below 1 then sphericity is violated.

Three common corrections are available in the repeated measures procedure. They are lower bound, Greenhouse-Geisser, and Huynh-Feldt. The lower bound value of epsilon and correction could be used for the worst possible case. The Greenhouse-Geisser correction is a conservative correction. It tends to underestimate epsilon when epsilon is close to 1. The Huynh-Feldt correction could be used when the true value of epsilon is thought to be near or above 0.75.

In this study, the conservative correction Greenhouse-Geisser is always used, unless the epsilon is equal to 1. When the epsilon is close to 1, Greenhouse-Geisser correction tends to underestimate epsilon, then Huynh-Feldt correction is used. Using these corrections works well for relatively modest epsilon departures and when sample size is small. However, the sphericity is not always violated in this study; especially when factors have two levels. A repeat measures factor with only two levels is done in special cases where the sphericity assumption is always met.

3.5.3 Compare means test

The paired samples t-test had been used in the case of using a statistical method for testing the hypotheses of two means. Its procedure compares the means of two variables for a single group. It computes the differences between values of the two variables for each case and tests whether the average differs from 0. In this study the paired samples t-test was always used following the repeated measures ANOVA, in order to test the difference between means of any two variables.

3.6 Limitation of Research

Regarding the amount of data collected in the experiments, experimental procedure seemed to have more practical limitations than others did. Although no large limitations were found when subject attended the experiment, untrained subjects led to some limitations.

Although untrained subjects with normal hearing can be selected for the experiment, their limitation was that they did not understand the task clearly, as they could not identify between signal and masker since both were the same material. However this limitation can be overcome by careful training of the subjects and explanations of the task.

To train a subject, researchers would first allow the subjects to listen to various conditions of signal and masker. The researcher would vary the level of the signal up and down in order to make sure that the subject can observe the differences between the speech when the signal was present and when the signal was not present. The researcher would let the subject listen to the various conditions of signal and masker until the subject can observe the present of the signal. Then, before starting each experiment, the researcher again ran the procedure for one or two conditions in order to check the subject's response. The period of training may vary from one subject to another. After training, a short break was provided before the actual procedure was begun in order to avoid the learning process.

3.7 Ethical Considerations

Ethical considerations for the noise exposure experiment were considered a main issue for conducting the experiment on a subject. As general practice, after an experiment has been designed to be used with human subjects, the researcher has to send requests for approval of the ethical, safety and insurance aspects of the experiment to the committee.

All experiments in this thesis were designed for safety. The researcher ensured that the level of exposure was not higher than 85 dB at any time. The period of exposure is no more than 80 minutes. This had been designed in order to avoid fatigue, with the subjects being allowed to have a break at any time. In the very long tests, short breaks were always provided.

In order to prevent accidental over-exposure, all volume controls are fixed (unmoved) at a calibrated level. Therefore, there is no possibility of accidentally exposing the subject to a high sound level.

For experimental procedures, ethical considerations for experiments involving subjects are the same as for most other methods of social research. When selecting subjects, the researcher provided full disclosure regarding the purpose of the experiment and uses of subjects' contributions. Being honest, avoiding pressure and keeping subjects informed about the experiment was key.

The researcher also informed the subject about the experiment and times to be used in each session. Before the experiment, subjects were required to fact-check the details on the subject's consent form, with a signed form being provided from each individual.

CHAPTER 4

Experiment 1: *The masking of speech by single delayed same-speech reflection*

4.1 Introduction

The main objective of this experiment is to work toward understanding the roles of level difference between the signal and the masker; time difference between the signal and the masker; direction of the masker relative to the direct sound; and type of programme (male and female) on speech masking; when a speech is masked by a simulated delayed same-speech (as a single reflection). These four variables are the main properties of a reflection. They are always referred to as the variables that affect sound quality in a room. Unfortunately, their effect on masking in this context is still unclear.

Masking effects have been investigated for decades. There have been many variables reported affecting the masked threshold, including masker level, type of programme, time interval between signal and masker, and masker incidence angle.

The effect of masker level was first reported by Wegel and Lane (1924). Since then a number of investigations on the effects of masker levels have been studied under various conditions using the signal and masker in combination. For example, the signal is pure tone, the masker is a narrow band noise (Egan and Hake, 1950); the signal is speech and the masker is broad band noise (Miller, 1948), etc. They results show the effect of the masker level on the masked threshold. It can be seen that the masked threshold increases when the masker level is increased. We also expect the consistent results in this experiment when the speech signal is masked by delayed same-speech.

For certain types of programme, such as speech, the speech masker is described at times as acting like a noise. However it seems to be a bit of an oversimplification, as speech is more complicated than noise. It contains vowels, consonants, and silent intervals. Their masking effect is also different from that of broad band noise. Nábělek et al. (1989) reported that the speech masker causes some vowels to get confused and degrades speech intelligibility. However the vowel confusion can be reduced when the silent interval is increased. This means that the speech signal can be detected in the silent interval (Wilson and Carhart, 1971). Not only can the signal be detected in the silent interval part of the sentence, but the signal can also be detected in the low amplitude consonant parts of the speech (Spiegel, 1987). This suggests that the characteristics of speech itself may affect the masking threshold when speech is masked by a delayed same-speech signal. In other words, the different types of programme might affect the masked threshold in this experiment.

The effect of time interval between signal and masker, when speech is masked by reflections, was first mentioned in 1949 by Bolt & Macdonald. They suggested that speech intelligibility can be reduced by decreasing the time interval between direct sound and reflections. It is known that when signal and reflection differ from one another by more than 1 ms at the position of the listener, then the position of the auditory event is determined by the position of the sound source arriving first. This effect is known as "law of the first wave front" or "precedence effect" (Blauert, 1997). Regarding this theory, only the sound that arrived first is taken into consideration when two sounds arrived close together. On the other hand, when the interval between two sounds exceeds the limit, the two auditory events appear one after the other. That means when the interval between two sounds is long enough, both signal and reflection will be easily detected. This suggests that the time interval between direct sound and the delayed same-speech masker may affect the speech masking in this condition.

Finally, the relationship between the incidence angle of direct sound and reflections was reported to affect the masking threshold by Saberi et al (1991) and Gilkey and Good (1995). They reported that the threshold was found to be poorest when the incidence angle of a signal as direct sound and the incidence angle of a masker as a reflection are the same. On the other hand, when the incidence angle of a signal and masker are different, it has been observed that the masking threshold was reduced. This suggests that the azimuth angle of incidence might affect the masked threshold of the speech.

As mentioned previously, all four variables seem to affect the masked threshold under various signal and masker conditions. However, the roles of these variables on masked thresholds have not yet been fully understood when signal and masker were the same-speech.

In this experiment, the roles of those four variables, including different level, type of programme, time difference, and masker direction on masked thresholds were investigated. The study was carried out by first measuring the masked threshold levels of the direct sound in the presence of direct sound, and the simulated delayed same-speech through the loudspeakers in anechoic room. The masked threshold level of each condition was measured by varying the level of the direct sound in the presence of a range of delayed masker signals covering all four key variables of interest. The method used was the up-down method (see chapter 3 for details). Please note that during this study, the listener's head movement was not restricted, as in natural listening.

4.2 Experimental procedure

In order to investigate the relationship between the masker's properties and the masking threshold of the speech signal presented with a delay of single reflection reinforcement, the experiment design will be described in the following section.

4.2.1 Experimental design

This experiment was conducted to examine the effect of a masker's properties on speech signals. It was done under consistent conditions of direction originating from the front, while the originating point of the reflected sound was changed. Both signal and reflection were produced in an anechoic room using two loudspeakers, each at a distance of 2 meters from the centre of the listener's head as shown in Figure 3.29 (section 3.3.5).

The incidence angles of the single reflection were chosen as 0, 30, 60, and 90 degrees (left hand side only) (see section 3.2.4.2.4). The delay of the reflected sound with respect to the direct sound was chosen as 15, 30, 50, and 100 ms. These delay times were long enough to assure that the incoherence of the two sounds would provide minimal coloration effect occurring. In addition, in this experiment, the difference between the levels of the delayed speech masker and the speech signal are kept dissimilar, therefore there will be even less coloration effect occurring. (see section 3.2.4.2.3).

The masking threshold level of the direct sound was measured while the amplitude of reflected sound was chosen as 40, 60, and 80 dBA at the subject's head position, as a reflection level as low as 40 dBA can be found in a small listening room, while a high reflection level of 80 dBA can possibly be found in the large concert hall (see section 3.2.4.2.2). It is noted that refection level surely depends on level of direct sound.

4.2.2 Stimuli and Calibration

Stimuli were running speech of male and female (see section 3.2.4). The stimuli were edited digitally by using the apparatus described in section 3.2.4. With a monophonic speech signal, the left was unedited. The right channel was edited in order to give it a delay with respect to the direct sound of 15, 30, 50, and 100 ms. After they were edited, they were recorded on MiniDisc (SHARP MD-MT-20) through the sound card of the computer. The audio on a MiniDisc is compressed using the ATRAC format (**A**daptive **TR**ansform **A**coustic **C**oding). ATRAC is a psychoacoustic lossy audio compression scheme, so decompression of the compressed signal will not yield exactly the original signal. However, to the listener, the compressed signal from the MiniDisc would sound very similar to the original, and would not degrade the perceived quality of the sound (Yoshida, 1994). This compressed signal was unlikely to affect the result of this experiment because the sound quality of the direct speech and the delayed speech are the same. Also, the dynamic signal level of the MiniDisc is not relevant because the direct speech and delayed speech were recorded at the same level. The attenuation level of the direct sound was applied after the output of the MiniDisc using an audiometer. Therefore, the speech signal and speech masker were recorded for about 2 minutes per condition. The 2 minutes was considered to be long enough for running each condition due to the pilot listening test. To present to the subjects, both signal and masker were played back via loudspeakers (KEF C35) in an anechoic room. The level of masker held constant at 40, 60, and 80 dB $L_{Aeq,1min}$ at the subject's head position, using the calibration method as described in section 3.2.4.3. The masking threshold level of the speech signal was then controlled and measured by the experimenter using an audiometer (Kamplex AD-27).

4.2.3 Experimental session

Each subject attended two sessions. One was male speech, the other was female speech. In each session, the experiment lasted for about 60 minutes, include a short break in order to avoid fatigue. There were total 49 conditions as described in section 4.2.3.3, including absolute threshold of speech.

To avoid the possibility of a learning effect, the subject who attended this experiment attended a different order of masker conditions. They started with different levels and types of speech. The experimental session is shown in the Table 4.1.

Table 4.1: Experimental sessions

male speech level 40, 60, 80 dBA

no.	delay (ms)	azimuth (degrees)
1	15	0
2	15	30
3	15	60
4	15	90
5	30	0
6	30	30
7	30	60
8	30	90
9	50	0
10	50	30
11	50	60
12	50	90
13	100	0
14	100	30
15	100	60
16	100	90

female speech level 40, 60, 80 dBA

no.	Delay (ms)	azimuth (degrees)
1	15	0
2	15	30
3	15	60
4	15	90
5	30	0
6	30	30
7	30	60
8	30	90
9	50	0
10	50	30
11	50	60
12	50	90
13	100	0
14	100	30
15	100	60
16	100	90

4.2.4 Apparatus

The masking threshold was measured using the apparatus already described in section 3.2.4.1. The apparatus used to produce the stimuli have been described in detail in section 3.2.4.2 together with its calibration in section 3.2.4.3. The stimuli used in this experiment were male and female running speech that was defined in section 3.2.4.2.1.

4.2.5 Subject

According to power analysis calculation (see section 3.3.3.1), the number of subjects required in this experiment was five. They consisted of two male and two female university students, 23-30 years of age and one male 50 years of age. Some of them have had some experience in psychoacoustic experimentation. Two males had also attended the pilot test of this experiment.

All subjects were tested as having normal hearing before taking part in the experiment. The hearing test procedure was described in section 3.3.4.1. All subjects had a hearing threshold of less than 20 dB in both ears in all frequency tests.

4.2.6 Procedure

In the experiment, subject's task is to detect the signal from the front. The subjects were instructed to press the response button when they heard or thought that they heard speech from the front. Prior to the actual tests, subjects trained to make sure that they familiar with the procedure.

Details of operation procedure were described in section 3.2.4.4.

The threshold measured run consisted of 10 reversals. The level of the first 4 reversals was not recorded. Only the levels of the last 6 reversals were read and recorded. The mean of the last 6 levels were taken as the threshold for that orientation as explained in section 3.4.

4.3 Results and Analysis

4.3.1 Absolute threshold

At the beginning of the experiment, the thresholds of male and female speech were measured. The mean of the absolute threshold for male speech was 5.1 dBA, and female speech was 6.3 dBA. The distribution of data was tested for normality using the Shapiro-Wilk test (more reliable when subjects less than 50) as shown in table 4.2.

Table 4.2: The test of normality using Shapiro-Wilk test

	(mean±SD)	statistic	Sig.
Male	5.1±1.8	.872	.278
Female	6.3±1.1	.847	.185

* p <0.05, ** p<0.01

From table 4.2, it may be assumed that the data is adequate to perform a parametric test. With the first glance, it appears that the mean absolute threshold of male speech was a little bit lower than that of female speech. To investigate the difference between the mean absolute threshold of male speech and the mean absolute threshold of female speech, the hypothesis is that absolute thresholds obtained for male speech and female speech are different. The repeated factors measured in this case are male speech and female speech. The epsilon of 1.000 indicates perfect sphericity. That is, the sphericity assumption is met (see details in chapter 3). Of the multivariate test results given, Wilks' Lambda is used as shown in Table 4.3.

Table 4.3: Multivariate analysis of male and female data ((a) Exact statistic)

Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.
SPEECH TYPE	Pillai's Trace	.579	5.495(a)	1.000	1.000	.079
	Wilks' Lambda	**.421**	**5.495(a)**	**1.000**	**1.000**	**.079**
	Hotelling's Trace	1.374	5.495(a)	1.000	1.000	.079
	Roy's Largest Root	1.374	5.495(a)	1.000	1.000	.079

From the table 4.3, the multivariate analysis yields significance at the 0.079 level for the absolute threshold. That means no significant difference was found for the absolute threshold between male and female (Lambda = 5.495, p = .079). In other words, the absolute thresholds of male speech and female speech were at the same level.

4.3.2 Signal type

To investigate the effect of signal type, the hypothesis is that masking thresholds obtained for male speech and female speech are different. The repeated measures factored in this case there are only two – male speech and female speech. The epsilon of 1.000 indicates perfect sphericity. That is, the sphericity assumption is met (see details in chapter 3). Of the multivariate test results given, Wilks' Lambda is used as shown in Table 4.4.

Table 4.4: Multivariate analysis of male and female data ((a) Exact statistic)

Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.
SPEECH TYPE	Pillai's Trace	.976	10.178(a)	4.000	1.000	.230
	Wilks' Lambda	**.024**	**10.178(a)**	**4.000**	**1.000**	**.230**
	Hotelling's Trace	40.711	10.178(a)	4.000	1.000	.230
	Roy's Largest Root	40.711	10.178(a)	4.000	1.000	.230

From table 4.4, the multivariate analysis yields significance at the 0.230 level for speech types. That means no significant effect was found for speech type (Lambda = 10.178, p = .230). Therefore it would be concerned that threshold of male speech and of female speech were at the same level. The effect of speech types on masking threshold under various conditions of azimuth angle of incidence, delay, and masker level was plotted in figures 4.1, 4.2, 4.3, and 4.4.

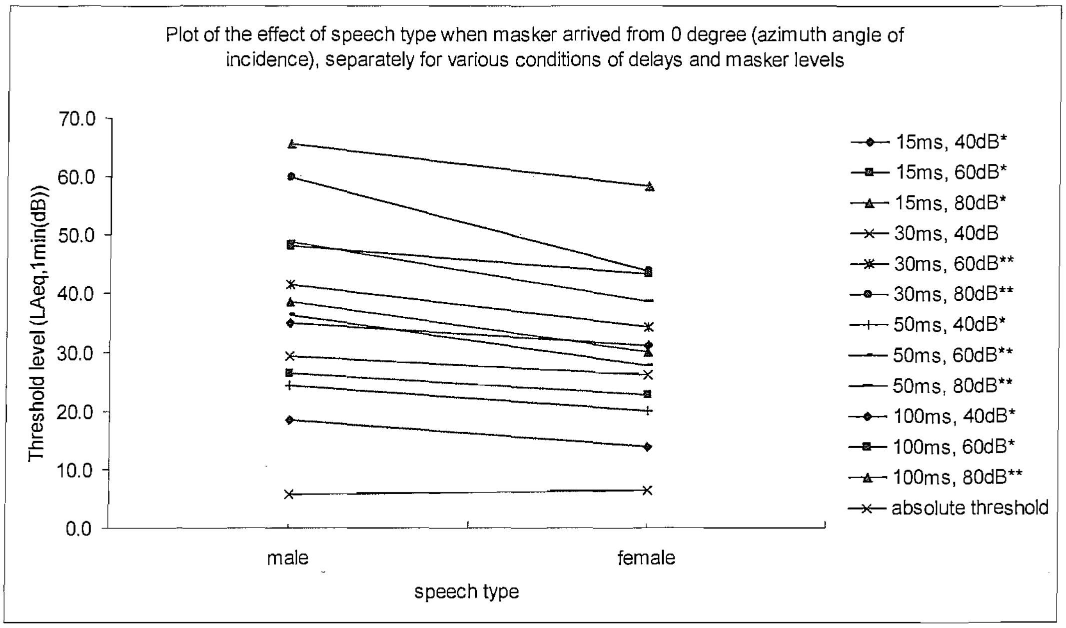


Figure 4.1: Effect of speech type on threshold when masker arrived from 0 degrees (azimuth angle), separated for various conditions of delays and masker levels. (* p <0.05, ** p<0.01)

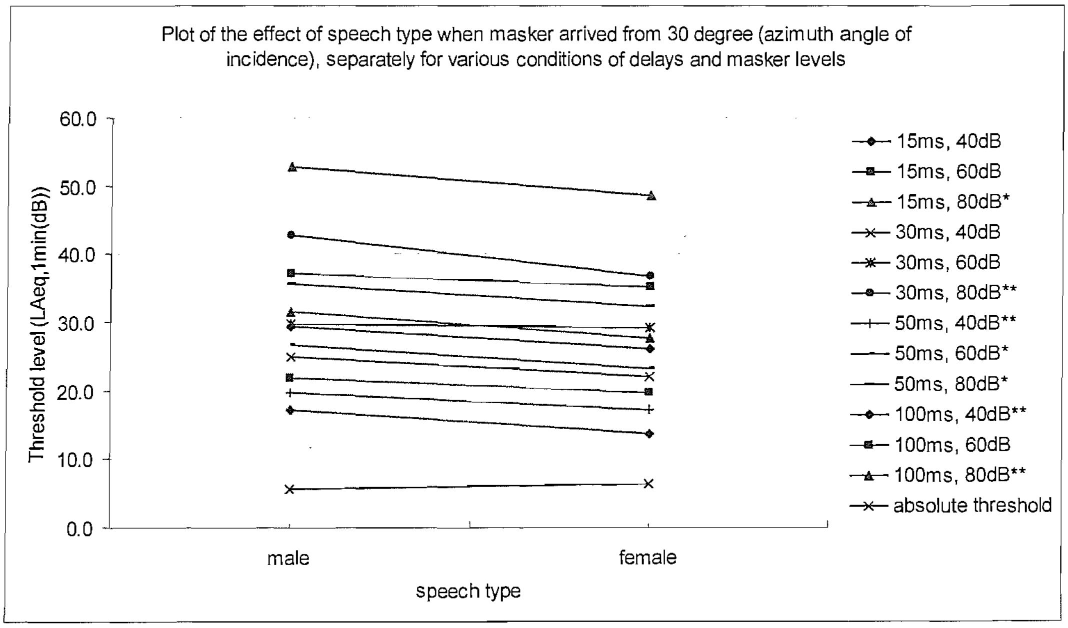


Figure 4.2: Effect of speech type on threshold when masker arrived from 30 degrees (azimuth angle), separated for various conditions of delays and masker levels. (* p <0.05, ** p<0.01)

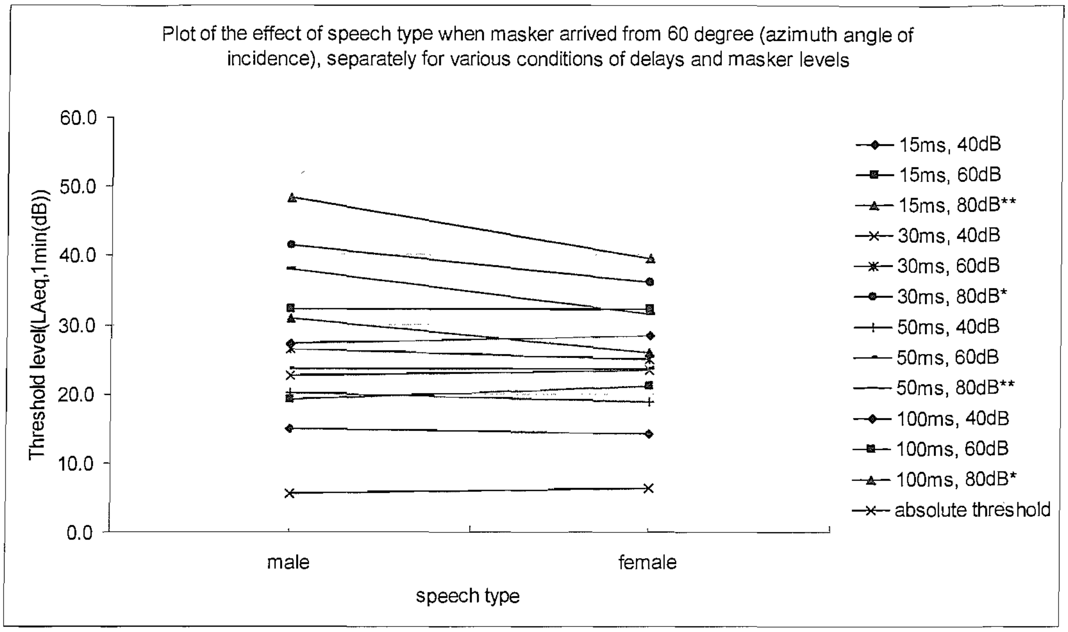


Figure 4.3: Effect of speech type on threshold when masker arrived from 60 degrees (azimuth angle), separated for various conditions of delays and masker levels. (* p <0.05, ** p<0.01)

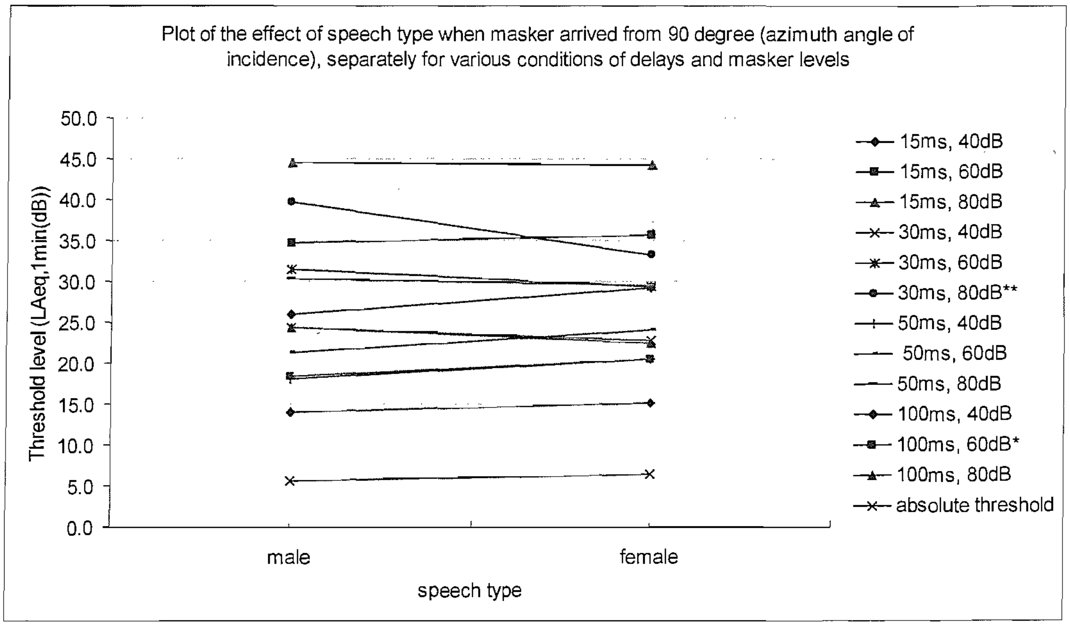


Figure 4.4: Effect of speech type on threshold when masker arrived from 90 degrees (azimuth angle), separated for various conditions of delays and masker levels.
(* p <0.05, ** p<0.01)

Although the multivariate indicated that the thresholds of male and female were the same, it can be observed from figures 4.1, 4.2, 4.3, and 4.4 that the threshold of male speech is slightly higher than the threshold of female speech under various conditions, especially when the masker arrived from 0 degrees azimuth angle of incidence and when the masker level is 80 dB. In order to have a closer look at the results, the univariate results and specific dependent variables (various conditions of azimuth angle of incidence, delays, and masker's levels) were measured. The table 4.5 gives results below.

Table 4.5: Threshold level comparing male speech to female speech under various conditions of azimuth angle of incidence, delays, and masker levels by using univariate ANOVA.

	Male (mean±SD)	female (mean±SD)	F-Value	Sig.
0 degrees, 15 ms, 40 dB	35.0±1.9	31.0±3.1	11.857	.026*
0 degrees, 15 ms, 60 dB	48.0±2.3	43.3±3.2	7.891	.048*
0 degrees, 15 ms, 80 dB	65.5±1.1	58.2±2.8	37.930	.004*
0 degrees, 30 ms, 40 dB	29.2±2.8	26.0±3.6	2.925	.162
0 degrees, 30 ms, 60 dB	41.5±2.8	34.2±3.1	82.426	.001**
0 degrees, 30 ms, 80 dB	59.7±0.9	43.7±2.1	238.942	.000**
0 degrees, 50 ms, 40 dB	24.3±1.5	20.0±2.6	13.926	.020*
0 degrees, 50 ms, 60 dB	36.2±2.5	27.7±2.6	26.403	.007**
0 degrees, 50 ms, 80 dB	48.7±3.2	38.5±2.5	44.609	.003**
0 degrees, 100 ms, 40 dB	18.3±2.0	13.8±3.3	13.485	.021*
0 degrees, 100 ms, 60 dB	26.2±1.7	22.7±2.3	10.152	.033*
0 degrees, 100 ms, 80 dB	38.5±2.2	30.0±3.7	77.476	.001**
30 degrees, 15 ms, 40 dB	29.5±0.7	26.2±3.6	3.995	.116
30 degrees, 15 ms, 60 dB	37.2±2.2	35.3±3.0	2.267	.207
30 degrees, 15 ms, 80 dB	52.8±2.7	48.5±1.8	14.730	.018*
30 degrees, 30 ms, 40 dB	25.0±3.5	22.0±1.3	4.537	.100
30 degrees, 30 ms, 60 dB	29.8±3.2	29.2±3.6	0.413	.555
30 degrees, 30 ms, 80 dB	42.7±3.2	36.8±2.9	48.966	.002**
30 degrees, 50 ms, 40 dB	19.8±3.1	17.2±2.5	30.157	.005**
30 degrees, 50 ms, 60 dB	26.7±2.4	23.2±3.6	13.185	.022*
30 degrees, 50 ms, 80 dB	35.7±3.1	32.3±2.3	8.000	.047*
30 degrees, 100 ms, 40 dB	17.2±1.7	13.8±1.3	159.418	.000**
30 degrees, 100 ms, 60 dB	21.8±2.2	19.7±3.4	3.280	.144
30 degrees, 100 ms, 80 dB	31.5±3.4	27.7±3.5	22.055	.009**
60 degrees, 15 ms, 40 dB	27.3±1.6	28.3±1.0	0.930	.389

60 degrees, 15 ms, 60 dB	32.3±1.6	32.3±2.5	0.000	.997
60 degrees, 15 ms, 80 dB	48.3±2.4	39.5±2.7	244.501	.000**
60 degrees, 30 ms, 40 dB	22.6±3.1	23.3±2.1	0.159	.711
60 degrees, 30 ms, 60 dB	26.5±3.3	25.0±2.6	1.103	.353
60 degrees, 30 ms, 80 dB	41.5±2.8	36.0±3.2	9.785	.035*
60 degrees, 50 ms, 40 dB	20.1±3.0	18.7±2.3	1.678	.265
60 degrees, 50 ms, 60 dB	23.6±2.7	23.5±2.7	0.016	.906
60 degrees, 50 ms, 80 dB	38.0±2.0	31.5±2.9	64.719	.001**
60 degrees, 100 ms, 40 dB	15.0±3.0	14.2±2.8	0.670	.459
60 degrees, 100 ms, 60 dB	19.1±3.5	21.0±3.0	1.232	.329
60 degrees, 100 ms, 80 dB	30.8±3.6	25.8±4.4	8.388	.044*
90 degrees, 15 ms, 40 dB	26.0±1.0	29.2±2.6	7.065	.057
90 degrees, 15 ms, 60 dB	34.6±3.2	35.7±3.0	1.538	.283
90 degrees, 15 ms, 80 dB	44.5±3.7	44.2±3.3	0.029	.874
90 degrees, 30 ms, 40 dB	24.3±1.8	22.7±3.2	0.754	.434
90 degrees, 30 ms, 60 dB	31.5±3.5	29.3±2.5	4.337	.106
90 degrees, 30 ms, 80 dB	39.6±1.8	33.3±2.8	45.915	.002**
90 degrees, 50 ms, 40 dB	18.1±2.9	20.5±2.8	2.348	.200
90 degrees, 50 ms, 60 dB	21.3±3.4	24.0±2.8	3.368	.140
90 degrees, 50 ms, 80 dB	30.3±1.5	29.5±3.2	0.400	.562
90 degrees, 100 ms, 40 dB	14.1±2.4	15.2±2.8	0.356	.583
90 degrees, 100 ms, 60 dB	18.4±2.8	20.5±2.8	8.478	.044*
90 degrees, 100 ms, 80 dB	24.3±2.1	22.5±2.8	6.532	.063

* $p<0.05$ ** $p<0.01$

Following the univariate ANOVAs showed that a statistically significant difference is observed between the threshold of male and of female speech at a significant level ($p < 0.05$) in almost every case when the azimuth angle of incidence is 0 degrees, except when the delay is 30 ms and the level is 40 dB. However when the azimuth angle of incidence increased toward 90 degrees, the difference between the threshold of male and of female speech and level were observed when azimuth angle of incidence, delay, and level were 30 degrees, 15 ms, 80 dB ($F=14.730$, $p < 0.05$); 30 degrees, 30 ms, 80 dB ($F=48.966$, $p < 0.01$); 30 degrees, 50 ms, 40 dB ($F=30.157$, $p < 0.01$); 30 degrees, 50 ms, 60 dB ($F=13.185$, $p < 0.05$); 30 degrees, 50 ms, 80 dB ($F=8.000$, $p < 0.05$); 30 degrees, 100 ms, 40 dB ($F=159.418$, $p < 0.01$); 30 degrees, 100 ms, 80 dB ($F=22.055$, $p < 0.01$); 60 degrees, 15 ms, 80 dB ($F=244.501$, $p < 0.01$); 60 degrees, 30 ms, 80 dB ($F=9.785$, $p < 0.05$); 60 degrees, 50 ms, 80 dB ($F=64.719$, $p < 0.01$); 60 degrees, 100 ms, 80 dB ($F=8.388$, $p < 0.05$); 90 degrees, 30 ms, 80 dB ($F=45.915$, $p < 0.01$); 90 degrees, 100 ms, 60 dB ($F=8.478$, $p < 0.05$).

It can be seen that, according to statistical analysis, the thresholds of male speech were higher than those of female speech in various conditions of azimuth angle, delay, and masker level as mentioned above. This finding suggested that the masking effect observed could have been due to the characteristics of speech itself. The fundamental frequency tends to be different, as female speech tends to have more energy at frequencies higher than male speech. In addition, the results could also have been due to several other factors such as the tendency for the speaker to leave slightly longer silent gaps between utterances, thereby changing the peak to long term rms (root mean square) ratio of the speech.

4.3.3 Time interval between signal and masker or delay

To investigate the effect of delay, the hypothesis is that masking thresholds obtained for time intervals between a signal and masker of 15 ms, 30 ms, 50 ms, and 100 ms are different. The repeated measures factored in this case are four – 15 ms, 30 ms, 50 ms, and 100 ms. From repeated measurement, the average epsilon of <0.75 indicates the sphericity is violated (see details in section 3.5.2). Therefore the results of the Greenhouse-Geisser corrections (Univariate test) are used. The univariate results and specific dependent variables (various conditions of speech type, azimuth angle of incidence, and masker's levels) were measured. The tabled results are provided below.

Table 4.6: Threshold level comparing delays of 15 ms, 30 ms, 50 ms, and 100 ms under various conditions of speech type, azimuth angle of incidence, and masker levels by using univariate ANOVA.

	15 ms (mean±SD)	30 ms (mean±SD)	50 ms (mean±SD)	100 ms (mean±SD)	F-Value	Sig.
male, 0 degrees, 40 dB	35.0 ± 1.9^{a}	29.2 ± 2.8^{b}	24.3 ± 1.5^{c}	18.3 ± 2.0^{d}	88.201	.000**
male, 0 degrees, 60 dB	48.0 ± 2.3^{a}	41.5 ± 2.8^{b}	36.2 ± 2.5^{c}	26.2 ± 1.7^{d}	372.870	.000**
male, 0 degrees, 80 dB	65.5 ± 1.1^{a}	59.7 ± 0.9^{b}	48.7 ± 3.2^{c}	38.5 ± 2.2^{d}	148.088	.000**

male, 30 degrees, 40 dB	29.5±0.7[a]	25.0±3.5[b]	19.8±3.1[c]	17.2±1.7[d]	52.519	.000**
male, 30 degrees, 60 dB	37.2±2.2[a]	29.8±3.2[b]	26.7±2.4[c]	21.8±2.2[d]	217.148	.000**
male, 30 degrees, 80 dB	52.8±2.7[a]	42.7±3.2[b]	35.7±3.1[c]	31.5±3.4[d]	260.189	.000**
male, 60 degrees, 40 dB	27.3±1.6[a]	22.6±3.1[b]	20.1±3.0[c]	15.0±3.0[d]	65.747	.000**
male, 60 degrees, 60 dB	32.3±1.6[a]	26.5±3.3[b]	23.6±2.7[b]	19.1±3.5[c]	46.964	.000**
male, 60 degrees, 80 dB	48.3±2.4[a]	41.5±2.8[b]	38.0±2.0[c]	30.8±3.6[d]	108.888	.000**
male, 90 degrees, 40 dB	26.0±1.0[a]	24.3±1.8[a,b]	18.1±2.9[b]	14.1±2.4[c]	55.748	.000**
male, 90 degrees, 60 dB	34.6±3.2[a]	31.5±3.5[b]	21.3±3.4[c]	18.4±2.8[d]	103.326	.000**
male, 90 degrees, 80 dB	44.5±3.7[a]	39.6±1.8[b]	30.3±1.5[c]	24.3±2.1[d]	114.575	.000**
female, 0 degrees, 40 dB	31.0±3.1[a]	26.0±3.6[b]	20.0±2.6[c]	13.8±3.3[d]	87.909	.000**
female, 0 degrees, 60 dB	43.3±3.2[a]	34.2±3.1[b]	27.7±2.6[c]	22.7±2.3[d]	150.301	.000**
female, 0 degrees, 80 dB	58.2±2.8[a]	43.7±2.1[b]	38.5±2.5[c]	30.0±3.7[d]	137.917	.000**
female, 30 degrees, 40 dB	26.2±3.6[a]	22.0±1.3[b]	17.2±2.5[c]	13.8±1.3[d]	46.661	.000**
female, 30 degrees, 60 dB	35.3±3.0[a]	29.2±3.6[b]	23.2±3.6[c]	19.7±3.4[c]	41.667	.000**
female, 30 degrees, 80 dB	48.5±1.8[a]	36.8±2.9[b]	32.3±2.3[c]	27.7±3.5[d]	177.535	.000**
female, 60 degrees, 40 dB	28.3±1.0[a]	23.3±2.1[b]	18.7±2.3[c]	14.2±2.8[d]	90.533	.000**
female, 60 degrees, 60 dB	32.3±2.5[a]	25.0±2.6[b]	23.5±2.7[b,c]	21.0±3.0[c]	59.874	.000**
female, 60 degrees, 80 dB	39.5±2.7[a]	36.0±3.2[b]	31.5±2.9[c]	25.8±4.4[d]	109.597	.000**
female, 90 degrees, 40 dB	29.2±2.6[a]	22.7±3.2[b]	20.5±2.8[b]	15.2±2.8[c]	114.589	.000**
female, 90 degrees, 60 dB	35.7±3.0[a]	29.3±2.5[b]	24.0±2.8[c]	20.5±2.8[d]	455.538	.000**
female, 90 degrees, 80 dB	44.2±3.3[a]	33.3±2.8[b]	29.5±3.2[c]	22.5±2.8[d]	418.082	.000**

* $p<0.05$ ** $p<0.01$

Following the univariate ANOVAs indicated that a statistically significant difference is observed among delays of 15 ms, 30 ms, 50 ms, and 100 ms at a significant level ($p<0.01$). Since the threshold is significantly affected by delays, what has to be determined is which delays differ significantly for thresholds under each condition. Therefore each pair of delays was tested using a paired samples t-Test. The results of the pair t-Test have been summarized in table 4.6 in terms of superscripts a, b, c, and d. The same superscripts indicate that there is no significant mean threshold difference between delays. On the other hand, the different superscripts indicate that mean thresholds between delays are different. The results can also be observed in figure 4.5, figure 4.6.

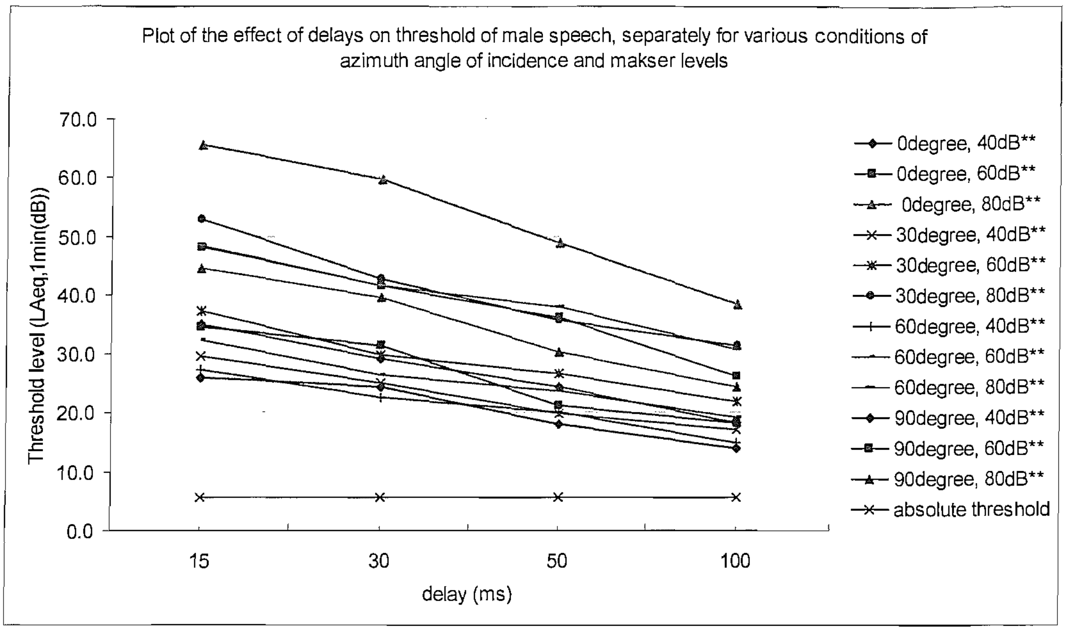


Figure 4.5: The effect of delay on threshold for male speech, separated for various conditions of azimuth angle of incidence and masker levels. (* p <0.05, ** p<0.01)

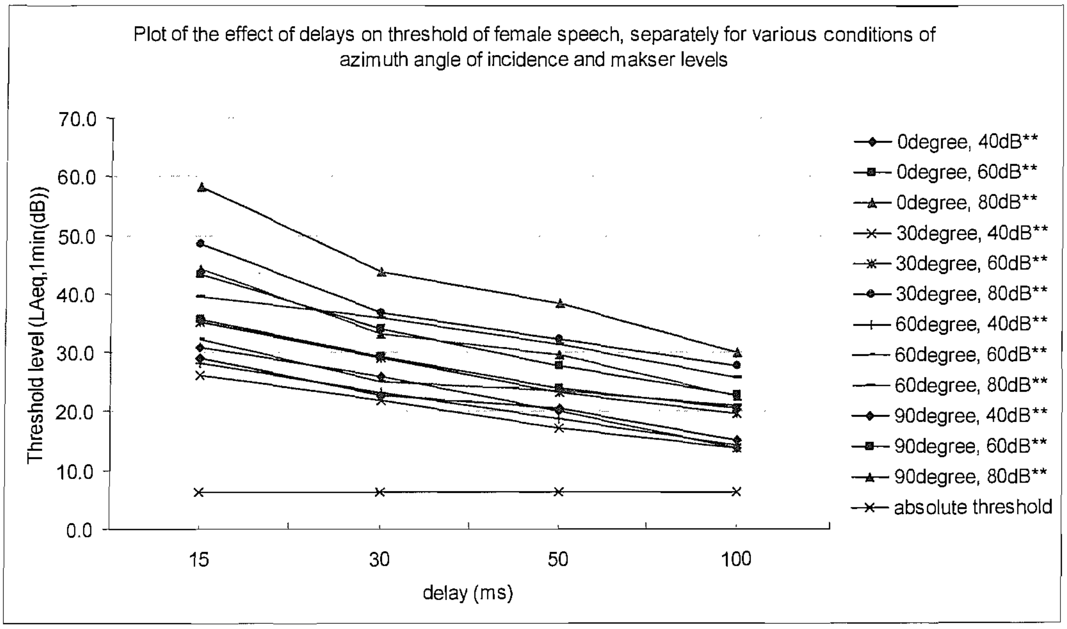


Figure 4.6: The effect of delay on threshold for female speech, separated for various conditions of azimuth angle of incidence and masker levels. (* p <0.05, ** p<0.01)

From figure 4.5 and 4.6, it can be seen clearly that the mean thresholds decline as the time interval between signal and masker increased toward 100 ms at a significant level (p <0.01) under all conditions investigated. Delay tends to be the most significant factor in speech signal detection. With a short delay, subjects have some difficulty identifying the speech signal when presented together with a speech masker because of "the precedence effect" as described in chapter 2 (Haas, 1972). However, when the delay was increased, the subjects can perceive the speech signal and the speech masker as two sound events separately. Therefore the threshold of speech signal was lower when the delay was increased toward 100 ms.

4.3.4 Azimuth angle of incidence

To investigate the effect of azimuth angle of incidence, the hypothesis is that masking thresholds obtained for azimuth angles of incidence of 0 degrees, 30 degrees, 60 degrees, and 90 degrees are different. The repeated measures factored in this case are four – 0 degrees, 30 degrees, 60 degrees, and 90 degrees. From the analysis, the average epsilon of <0.75 indicates that sphericity is violated (see details in chapter 3). Therefore the results of the Greenhouse-Geisser correction (Univariate test) are used. The univariate results and specific dependent variables (various conditions of speech type, delays, and masker's levels) were measured. The tabled results are provided below.

Table 4.7: Threshold level comparing azimuth angles of incidence of 0 degrees, 30 degrees, 60 degrees, and 90 degrees under various conditions of speech type, delays, and masker levels by using univariate ANOVA.

	0 degrees (mean±SD)	30 degrees (mean±SD)	60 degrees (mean±SD)	90 degrees (mean±SD)	F-Value	Sig.
male, 15 ms, 40 dB	35.0 ± 1.9^{a}	29.5 ± 0.7^{b}	27.3 ± 1.6^{c}	26.0 ± 1.0^{d}	52.173	.000**
male, 15 ms, 60 dB	48.0 ± 2.3^{a}	37.2 ± 2.2^{b}	32.3 ± 1.6^{c}	$34.6\pm3.2^{b,c}$	59.458	.000**
male, 15 ms, 80 dB	65.5 ± 1.1^{a}	52.8 ± 2.7^{b}	$48.3\pm2.4^{b,c}$	$44.5\pm3.7^{c,d}$	78.086	.000**
male, 30 ms, 40 dB	29.2 ± 2.8^{a}	$25.0\pm3.5^{a,b}$	$22.6\pm3.1^{b,c}$	$24.3\pm1.8^{a,c}$	5.878	.026*
male, 30 ms, 60 dB	41.5 ± 2.8^{a}	29.8 ± 3.2^{b}	26.5 ± 3.3^{b}	31.5 ± 3.5^{b}	34.930	.001**
male, 30 ms, 80 dB	59.7 ± 0.9^{a}	42.7 ± 3.2^{b}	41.5 ± 2.8^{b}	39.6 ± 1.8^{b}	108.054	.000**

male, 50 ms, 40 dB	24.3 ± 1.5^{a}	19.8 ± 3.1^{b}	20.1 ± 3.0^{b}	18.1 ± 2.9^{b}	14.610	.004**
male, 50 ms, 60 dB	36.2 ± 2.5^{a}	26.7 ± 2.4^{b}	23.6 ± 2.7^{b}	21.3 ± 3.4^{b}	44.249	.000**
male, 50 ms, 80 dB	48.7 ± 3.2^{a}	35.7 ± 3.1^{b}	38.0 ± 2.0^{c}	30.3 ± 1.5^{d}	48.521	.001**
male, 100 ms, 40 dB	18.3 ± 2.0^{a}	$17.2\pm1.7^{a,b}$	15.0 ± 3.0^{b}	14.1 ± 2.4^{b}	10.293	.010**
male, 100 ms, 60 dB	26.2 ± 1.7^{a}	$21.8\pm2.2^{a,b}$	$19.1\pm3.5^{a,b}$	18.4 ± 2.8^{b}	16.009	.005**
male, 100 ms, 80 dB	38.5 ± 2.2^{a}	31.5 ± 3.4^{b}	$30.8\pm3.6^{b,c}$	24.3 ± 2.1^{d}	30.319	.000**
female, 15 ms, 40 dB	31.0 ± 3.1^{a}	26.2 ± 3.6^{a}	28.3 ± 1.0^{a}	29.2 ± 2.6^{a}	4.020	.064
female, 15 ms, 60 dB	43.3 ± 3.2^{a}	35.3 ± 3.0^{b}	32.3 ± 2.5^{b}	35.7 ± 3.0^{b}	73.571	.000**
female, 15 ms, 80 dB	58.2 ± 2.8^{a}	48.5 ± 1.8^{b}	39.5 ± 2.7^{c}	44.2 ± 3.3^{b}	50.855	.001**
female, 30 ms, 40 dB	26.0 ± 3.6^{a}	22.0 ± 1.3^{a}	23.3 ± 2.1^{a}	22.7 ± 3.2^{a}	4.140	.080
female, 30 ms, 60 dB	34.2 ± 3.1^{a}	$29.2\pm3.6^{a,b,c}$	25.0 ± 2.6^{b}	29.3 ± 2.5^{c}	20.529	.008**
female, 30 ms, 80 dB	43.7 ± 2.1^{a}	36.8 ± 2.9^{b}	$36.0\pm3.2^{b,c}$	33.3 ± 2.8^{d}	72.506	.000**
female, 50 ms, 40 dB	20.0 ± 2.6^{a}	17.2 ± 2.5^{a}	18.7 ± 2.3^{a}	20.5 ± 2.8^{a}	5.176	.051
female, 50 ms, 60 dB	27.7 ± 2.6^{a}	$23.2\pm3.6^{a,b}$	$23.5\pm2.7^{a,b}$	24.0 ± 2.8^{b}	8.899	.009**
female, 50 ms, 80 dB	38.5 ± 2.5^{a}	32.3 ± 2.3^{b}	31.5 ± 2.9^{b}	29.5 ± 3.2^{b}	48.359	.000**
female, 100 ms, 40 dB	13.8 ± 3.3^{a}	13.8 ± 1.3^{a}	14.2 ± 2.8^{a}	15.2 ± 2.8^{a}	0.741	.488
female, 100 ms, 60 dB	22.7 ± 2.3^{a}	19.7 ± 3.4^{a}	21.0 ± 3.0^{a}	20.5 ± 2.8^{a}	2.598	.157
female, 100 ms, 80 dB	30.0 ± 3.7^{a}	27.7 ± 3.5^{a}	$25.8\pm4.4^{a,b}$	22.5 ± 2.8^{b}	30.319	.001**

* p<0.05 ** p<0.01

Following the univariate ANOVAs indicated that a statistically significant difference is observed among azimuth angles of incidence of 0 degrees, 30 degrees, 60 degrees, and 90 degrees are different at a significant level (p<0.01) for almost all male speech conditions, except with the delay of 30 ms at 40 dB, a statistically significant difference is observed among azimuth angles that are different at a significant level (p<0.05). For female speech, the statistically significant difference is observed among azimuth angles of incidence that are different at a significant level (p<0.01) when delayed, the levels being 15 ms, 60 dB (F=73.571, p < 0.01); 15 ms, 80 dB (F=50.855, p < 0.01); 30 ms, 60 dB (F=20.529, p < 0.01); 30 ms, 80 dB (F=72.506, p < 0.01); 50 ms, 60 dB (F=8.899, p < 0.01); 50 ms, 80 dB (F=48.359, p < 0.01); 100 ms, 80 dB (F=30.319, p < 0.01). Since the threshold is significantly affected by azimuth angle of incidence, what has to be determined is which azimuth angle differs significantly for the threshold under each condition. Therefore, each pair of azimuth angles were tested using paired a samples t-Test. The results of the pair t-Test have been summarized in table 4.7 in terms of superscripts a, b, c, and d. The same superscripts indicate that there is no significant mean threshold difference between delays. On the other hand, the different superscripts indicate that mean thresholds between delays are different. The results can also be observed in figure 4.7 and figure 4.8.

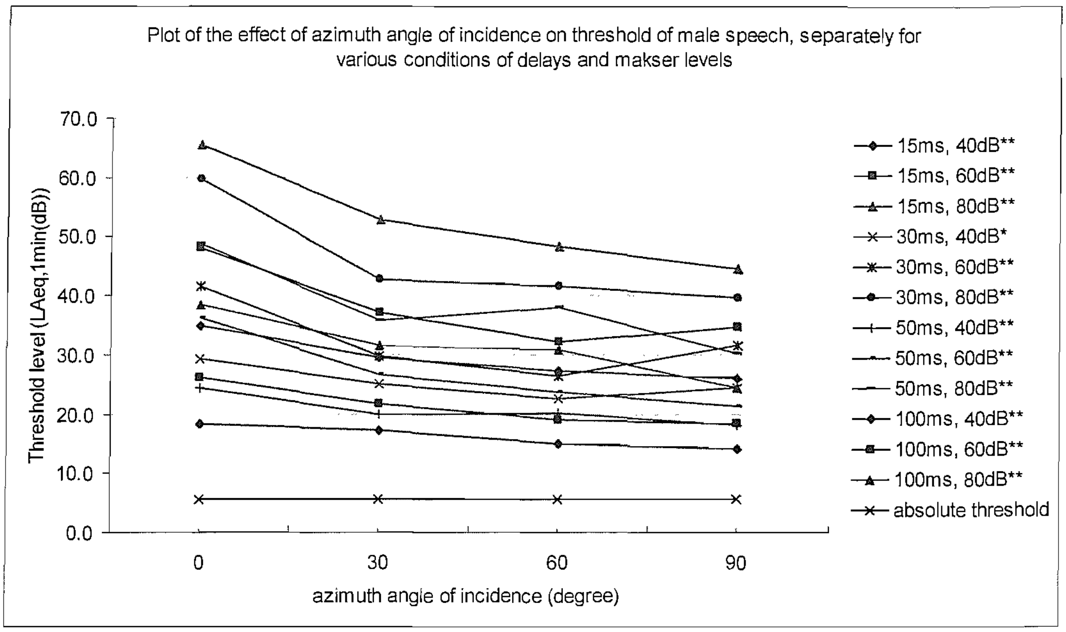


Figure 4.7: The effect of azimuth angle of incidence on thresholds for male speech, separated for various conditions of delay and masker levels. (* $p < 0.05$, ** $p < 0.01$)

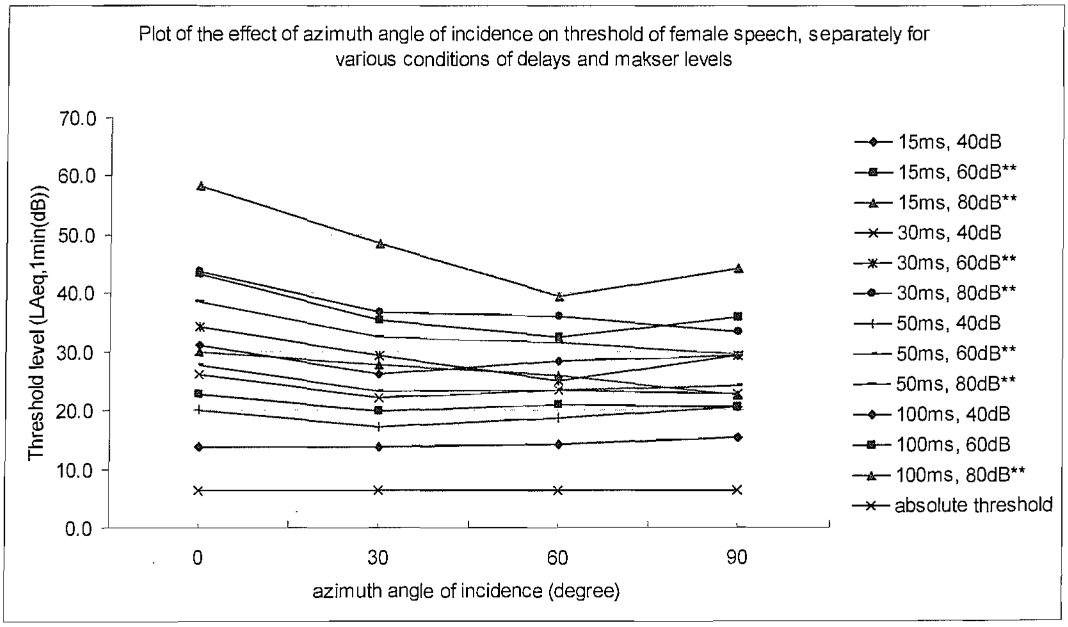


Figure 4.8: The effect of azimuth angle of incidence on thresholds for female speech, separated for various conditions of delay and masker levels. (* $p < 0.05$, ** $p < 0.01$)

From figure 4.7 and figure 4.8, the findings show that the threshold for male speech and female speech is found to be poorest when the masker arrived at 0 degrees or when masker arrived at the same azimuth angle as the signal. The mean thresholds decline when the signal and masker were separated by increasing the incidence angle of the masker to 30 degrees, 60 degrees, and 90 degrees. When we look closely to the threshold of female speech when masker arrived from 30 degrees, 60 degrees, and 90 degrees, the lowest threshold level might not always be measured when the masker arrived from 90 degrees. This might have been due to the characteristics of speech, especially fundamental frequencies of speech and the silent interval. These characteristics of speech will provide different conditions of binaural masking levels for different conditions when the masker arrived from 30 degrees, 60 degrees, and 90 degrees (See binaural masking level difference, BMLD in chapter 2).

4.3.5 Masker level

To investigate the effect of level, the hypothesis is that masking thresholds obtained for masker levels of 40 dB, 60 dB, and 80 dB are different. The repeated measures factored in this case there are three – 40 dB, 60 dB, and 80 dB. From repeat measurement, the average epsilon of <0.75 indicates that sphericity is violated (see details in chapter 3). Therefore the results of the Greenhouse-Geisser corrections (Univariate) test are used. The univariate results and specific dependent variables (various conditions of speech type, azimuth angle of incidence, and delay) were measured. The tabled results are provided below.

Table 4.8: Threshold level comparing masker levels of 40 dB, 60 dB, and 80 dB under various conditions of speech type, azimuth angle of incidence, and delay by using univariate ANOVA.

	40 dB (mean±SD)	60 dB (mean±SD)	80 dB (mean±SD)	F-Value	Significance
male, 0 degrees,15 ms	35.0 ± 1.9^a	48.0 ± 2.3^b	65.5 ± 1.1^c	395.108	.000**
male, 0 degrees,30 ms	29.2 ± 2.8^a	41.5 ± 2.8^b	59.7 ± 0.9^c	191.329	.000**
male, 0 degrees,50 ms	24.3 ± 1.5^a	36.2 ± 2.5^b	48.7 ± 3.2^c	99.351	.000**

male, 0 degrees,100 ms	18.3±2.0[a]	26.2±1.7[b]	38.5±2.2[c]	164.344	.000**
male, 30 degrees,15 ms	29.5±0.7[a]	37.2±2.2[b]	52.8±2.7[c]	215.384	.000**
male, 30 degrees,30 ms	25.0±3.5[a]	29.8±3.2[b]	42.7±3.2[c]	230.049	.000**
male, 30 degrees,50 ms	19.8±3.1[a]	26.7±2.4[b]	35.7±3.1[c]	262.146	.000**
male, 30 degrees,100 ms	17.2±1.7[a]	21.8±2.2[b]	31.5±3.4[c]	122.794	.000**
male, 60 degrees,15 ms	27.3±1.6[a]	32.3±1.6[b]	48.3±2.4[c]	217.530	.000**
male, 60 degrees,30 ms	22.6±3.1[a]	26.5±3.3[b]	41.5±2.8[c]	160.738	.000**
male, 60 degrees,50 ms	20.1±3.0[a]	23.6±2.7[b]	38.0±2.0[c]	343.955	.000**
male, 60 degrees,100 ms	15.0±3.0[a]	19.1±3.5[b]	30.8±3.6[c]	161.707	.000**
male, 90 degrees,15 ms	26.0±1.0[a]	34.6±3.2[b]	44.5±3.7[c]	126.705	.000**
male, 90 degrees,30 ms	24.3±1.8[a]	31.5±3.5[b]	39.6±1.8[c]	45.539	.000**
male, 90 degrees,50 ms	18.1±2.9[a]	21.3±3.4[b]	30.3±1.5[c]	67.005	.000**
male, 90 degrees,100 ms	14.1±2.4[a]	18.4±2.8[b]	24.3±2.1[c]	93.270	.000**
female, 0 degrees,15 ms	31.0±3.1[a]	43.3±3.2[b]	58.2±2.8[c]	100.149	.000**
female, 0 degrees,30 ms	26.0±3.6[a]	34.2±3.1[b]	43.7±2.1[c]	136.530	.000**
female, 0 degrees,50 ms	20.0±2.6[a]	27.7±2.6[b]	38.5±2.5[c]	124.987	.000**
female, 0 degrees,100 ms	13.8±3.3[a]	22.7±2.3[b]	30.0±3.7[c]	178.937	.000**
female, 30 degrees,15 ms	26.2±3.6[a]	35.3±3.0[b]	48.5±1.8[c]	151.733	.000**
female, 30 degrees,30 ms	22.0±1.3[a]	29.2±3.6[b]	36.8±2.9[c]	48.292	.000**
female, 30 degrees,50 ms	17.2±2.5[a]	23.2±3.6[b]	32.3±2.3[c]	169.005	.000**
female, 30 degrees,100 ms	13.8±1.3[a]	19.7±3.4[b]	27.7±3.5[c]	70.510	.000**
female, 60 degrees,15 ms	28.3±1.0[a]	32.3±2.5[b]	39.5±2.7[c]	53.618	.000**
female, 60 degrees,30 ms	23.3±2.1[a]	25.0±2.6[b]	36.0±3.2[c]	108.260	.000**
female, 60 degrees,50 ms	18.7±2.3[a]	23.5±2.7[b]	31.5±2.9[c]	184.807	.000**
female, 60 degrees,100 ms	14.2±2.8[a]	21.0±3.0[b]	25.8±4.4[c]	78.645	.000**
female, 90 degrees,15 ms	29.2±2.6[a]	35.7±3.0[b]	44.2±3.3[c]	86.028	.000**
female, 90 degrees,30 ms	22.7±3.2[a]	29.3±2.5[b]	33.3±2.8[c]	133.455	.000**
female, 90 degrees,50 ms	20.5±2.8[a]	24.0±2.8[b]	29.5±3.2[c]	50.275	.000**
female, 90 degrees,100 ms	15.2±2.8[a]	20.5±2.8[b]	22.5±2.8[c]	38.047	.003**

* $p<0.05$ ** $p<0.01$

Following the univariate ANOVAs indicated that a statistically significant difference is observed among masker levels of 40 dB, 60 dB, and 80 dB at a significant level ($p<0.01$). Since the threshold is significantly affected by the masker level, what has to be determined is which masker levels differ significantly for thresholds under each condition. Therefore, each pair of levels was tested using paired a samples t-Test. The results of the pair t-Test have been summarized in table 4.8 in term of superscripts a, b, c, and d. The same superscripts indicate that there is no significant mean threshold difference between levels. On the other hand, the different superscripts indicate that mean thresholds between levels are different. The results can also be observed in figures 4.9 and 4.10.

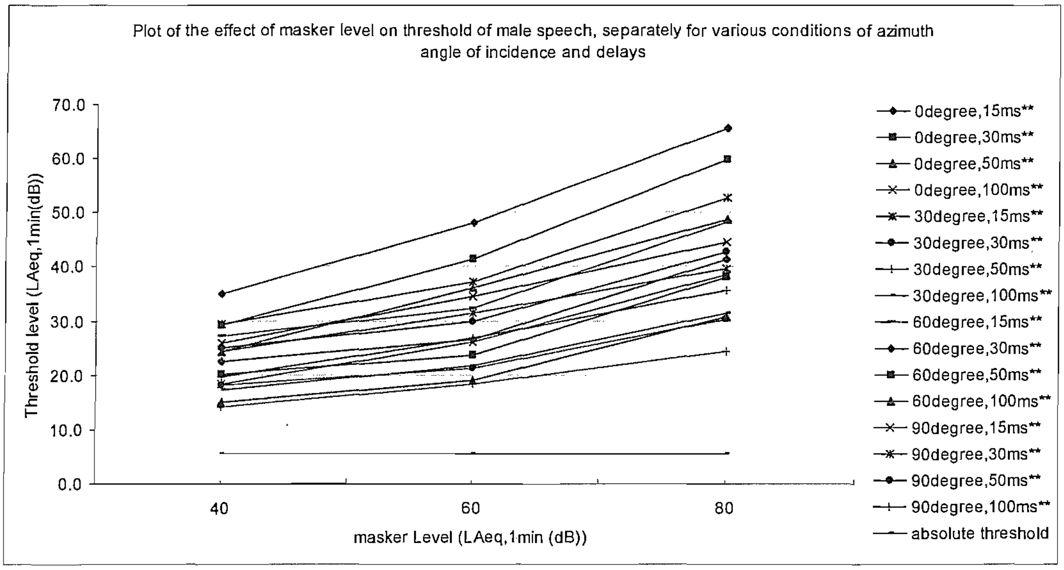


Figure 4.9: The effect of masker level on thresholds for male speech, separated for various conditions of azimuth angle of incidence and delay. (* p <0.05, ** p<0.01)

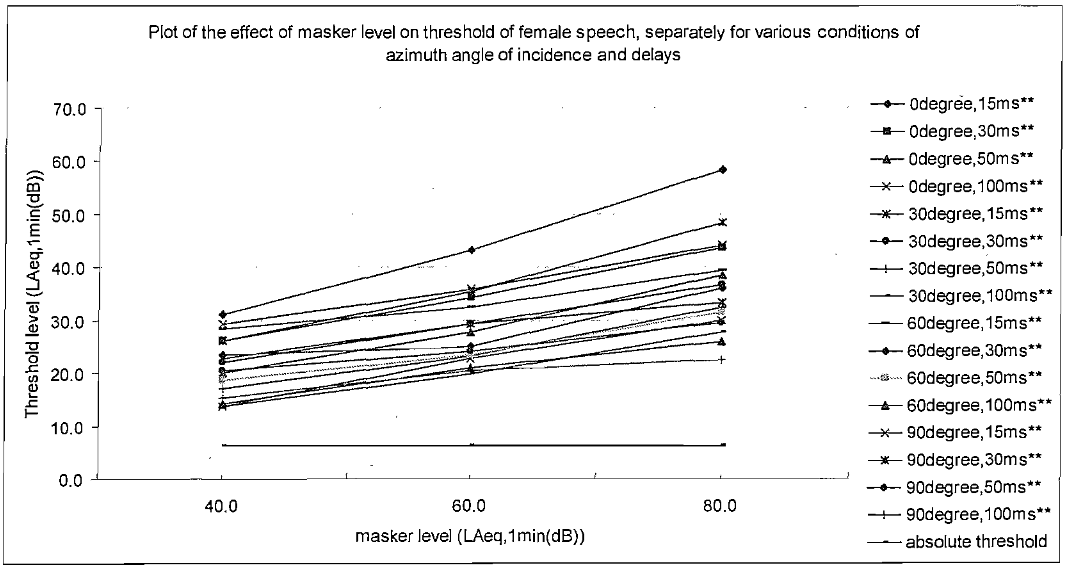


Figure 4.10: The effect of masker level on thresholds for female speech, separated for various conditions of azimuth angle of incidence and delay. (* p <0.05, ** p<0.01)

From figure 4.9 and figure 4.10, the findings show that the mean threshold rises as the masker level increases toward 80 dB at a significant level (p <0.01) under all conditions investigated. However the increasing masking threshold investigated in this study is found to be non-linear. For example, in cases of male speech and a masker arriving at 0 degrees azimuth angle at a time delay of 15 ms, the signal to masker ratio was -5 dB when the masker level was 40 dB. The signal to masker ratio was reduced to -12 dB and -14.5 dB when the masker levels were 60 dB and 80 dB respectively. According to these calculations, this would suggest that, although the threshold was increased according to the increase in the masker level, that the signal detection became better at a higher masker level. If it is assumed that the relationship between masking threshold and masker level is liner, an assumption line of 1 dB/ dB will be plotted in the results of male speech (figure 4.9) as shown in figure 4.11.

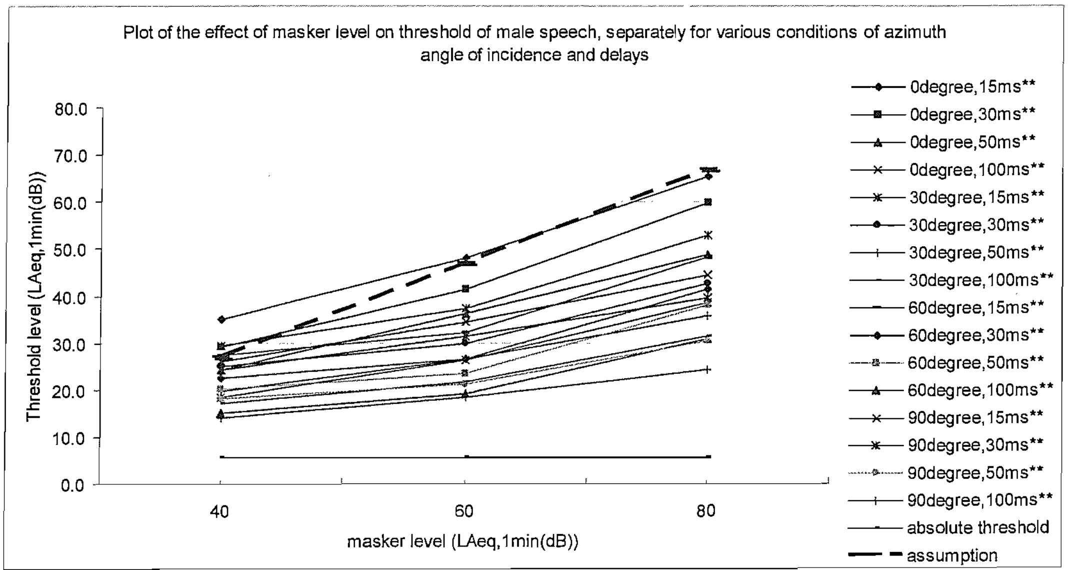


Figure 4.11: The effect of masker level on the threshold of male speech, separated for various conditions of azimuth angle of incidence and delay with the prediction line sloped 1 dB/ dB. (* p <0.05, ** p<0.01)

From the above figure, it shows that the masking threshold is raised at 40 dB, giving a non-linearity pattern as the result. This non-linearity curve may be caused by the ambient noise inside the room which raised the threshold of speech when the masker level is 40 dB. As well, the internal noise generated inside the auditory system may have affected the threshold when the masker level was 40 dB because internal noise generated inside the auditory system was only found when the masker level was low (Moore, 1997).

CHAPTER 5

Experiment 2: *Type of programme: The detectability of speech signal when speech is masked by delayed same-speech under headphone listening condition.*

5.1 Introduction

The main objective of this experiment is to contribute to understanding how speech signals are detected under the influence of simultaneous masking effect and non-simultaneous masking effect on male and female speech when speech is masked by delayed same-speech. Following the first experiment, one variable interested was the type of programme as speech. The characteristics of the speech masker may affect the detectability of the signal when the speech signal was presented together with the delayed same-speech masker. A number of studies report on speech masked by speech that the masked threshold levels are always affected by the combination of simultaneous and non-simultaneous masking effects (Steven et al, 1946; Miller, 1947;

Hawkins and Steven, 1950; Spiegel, 1987). Unfortunately, the detectability of a speech masker signal under the influence of a simultaneous masking effect and non-simultaneous masking effect has not yet been understood.

The simultaneous masking effect when speech is masked by broad band noise has been detailed by many investigators such as Steven et al. (1946), Miller (1947), Hawkins and Steven (1950). Their results show that the masking effect of noise on speech is similar to the masking effect of pure tone on pure tone. That is, the masking threshold increases when the level of the masker is increased, with an upward spread masking in frequency. However this pattern might not be the same as that of speech on speech because a speech masker is not represented in a fixed acoustics pattern as noise (Moore, 1982). It is composed of variable waveform patterns (see figure 5.1).

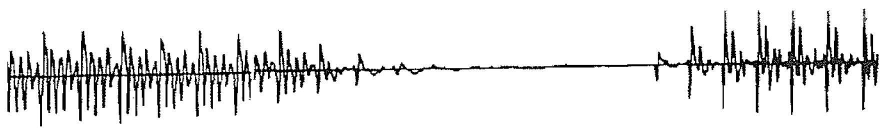


Figure 5.1: Speech waveform.

From the figure above, it can be seen that the characteristics of the speech masker are difference from board band noise. Speech is composed of vowels, consonants, and intervals of silence. The silent intervals in speech may provide an important cue in speech signal detection. The effect of a silent gap between two maskers was investigated by Wilson and Carhart in 1971. They observed that the level of the masking threshold depends mainly on the silent gap's duration. A smaller release from masking has been found to correlate with a smaller silent gap. On the other hand, a greater release from masking has been found to correlate in a bigger silent gap. The masking threshold may be as low as the absolute threshold when the gap is bigger than 400ms. Following the previous research, it is clear that the silent interval or silent gap between utterances in a speech masker play an important role in this signal detection and provide some release from masking.

Nevertheless, Spiegel (1987) argued that the low amplitude consonants in speech could also provide a cue in signal detection because they act similarly to a silent gap. It is known that the low frequency masker is more effective in masking signals than the higher frequency masker (Steven et al., 1946; Miller, 1947). That means that low amplitude consonants can hardly mask the lower frequency sound when they overlap in time. In other words, it is possible that the lower frequency elements in a signal may be detected in the low amplitude consonants of the masker. Unfortunately, this assumption has not yet been investigated.

Consequently, in order to understand how speech signals are detected in speech maskers, this experiment was carried out into two sessions.

The first session was designed to test the ability of the subject to detect speech signals in the 'silent gap' between utterances in a running speech masker in comparison with those in the low amplitude consonants of the running speech masker.

In the second session, the experiment was carried out to test the effect of upward spread masking, or the effect of simultaneous masking when the speech signal was detected in the low amplitude consonant parts of speech. The experiment in this session was designed to test this effect by filtering the running speech into a range of frequency bandwidth conditions for both male and female speech. Headphone listening was used in the experiment because it is more convenient than loudspeaker listening when the relative angle between the masker and the direct sound is not an issue. In addition, headphone listening automatically negates any variations of head movement while listening.

5.2 Experimental procedure

5.2.1 Experimental design

As described previously, this experiment was designed to examine the relative role of simultaneous and non-simultaneous masking on speech detection; therefore the experiment was divided into two parts. The first part was the investigation of the effect of a silent gap speech masker, whereas the second part was the investigation on an upward spread masking effect when speech was presented together with a delayed same-speech signal. Both parts of the experiment were conducted under headphone listening conditions in an acoustics booth.

The masking threshold level of speech was measured while the amplitude of delayed sound was set at a comfortable listening level 60dB $L_{Aeq,1min}$. This level was calibrated as described in section 3.2.4.3. The delay time was chosen to be 15, 30, and 50ms. These delay times were long enough to assure incoherence of the two sounds. No coloration effects occurred (see section 3.2.3.3).

In the first session of this experiment, the ability to detect speech in a silent gap and in low amplitude consonants was investigated. In order to carry out this experiment, the speech signal and speech masker in this session were edited digitally by eliminating the silent gap using "Gold wave" software. Silence intervals wider than 15ms were removed because it was considered that only the silent intervals wider than 15ms contributed to masking release when forward and backward masking effects were combined (Wilson and Carhart, 1971). Then, the masking thresholds of normal speech and of speech without a silent gap were measured and analyzed.

In the second session, the detection of speech in the low amplitude region of consonant speech was investigated. Upward spread masking in frequencies was examined. Speech without silent gaps was chosen in this session instead of normal speech in order to avoid the possibility of detecting the speech signal in the silent interval. In the other words, without a silent interval, speech can be only detected in syllabic parts of a speech signal, with components of speech maskers easily observed. Therefore, speech without a silent interval was filtered into frequency bands.

To filter speech into frequency bands, the concept of critical bands was first considered. Critical band concept is always referred to in simultaneous and non-simultaneous effect masking effects, as these effects occur within critical bands as described in section 2.2.1 and 2.2.2. Unfortunately, speech can not be filtered into the critical bands because the speech energy within each critical band is too low. Furthermore, under critical band frequencies, speech would be filtered into too many frequency bands without providing any sensitive results, as a speech waveform in any critical band might be the same as that in the adjacent critical bands.

Without using critical band frequency, speech should be filtered into frequency bands that are wider than the critical bands. One option is to investigate the characteristics of speech, including vowels and consonants. For vowels, the frequencies are classified into formants; whereas, for consonants, the frequencies cannot be classified into a range of frequencies because it varies from low frequencies to high frequencies depending on the type of the consonant used (see section 2.10.2). Consequently, in order to filter speech into frequency bands, the frequencies of vowels were taken into consideration. The first and second formant frequencies of vowels were used as references for frequency bands because it is the most important formants used in identify the characteristic of vowels (Kent & Read, 1992). According to table 2.3 (in section 2.10.1), it shows that the average first formants frequencies of vowels are between 280Hz and 860Hz, the average second formants frequencies of vowels are between 820Hz and 2230Hz. Based on these frequency ranges, the speech in the second session of this experiment was filtered into three frequency bandwidths below 800Hz as the low frequency, between 800Hz and 2,000Hz as the mid frequency, and above 2,000Hz as the high frequency.

In this session of the experiment, the filtered frequency speech was used for both signal and masker. The normal speech was not used as a masker with the filtered signal because the masked threshold presented with a normal speech masker may become lower than that of the filtered signal presented with a filtered speech masker, due to the "comodulation masking release (CMR)" effect as described in section 2.4.3.

5.2.2 Stimuli and Calibration

Stimuli were running speech of male and female speech as described in section 3.2.4.2. These stimuli were edited digitally (see section 3.2.4.2.1). In the first session, the silent gaps in the speech were removed. In the second session, speech without silence intervals was filtered into three frequency bands as described in the previous section. After all the stimuli were filtered digitally, the frequency responses were measured using an HP frequency analyzer in order to ensure that the stimuli were filtered correctly (see section 3.2.4.2.1).

Table 5.1: Calibrated level of signal and masker in various conditions

Stimuli	signal calibrated level ($L_{Aeq,1min}$)	masker calibrated level ($L_{Aeq,1min}$)
male speech	60.3	59.8
male speech without gap	59.6	59.3
male speech below 800Hz (L)	59.4	60.8
male speech between 800-2000Hz (M)	60.6	60.6
male speech above 2000Hz (H)	60.5	60.3
female speech	59.9	60.0
female speech without gap	59.9	60.3
female speech below 800Hz (L)	59.3	60.0
female speech between 800-2000Hz (M)	60.5	60.5
female speech above 2000Hz (H)	60.3	59.8

After all stimuli were filtered into different frequency bands, it was observed that the level of each stimulus was different from the other because of the dynamic range of speech. To avoid bias from changing the level of the signal and masker in various conditions during the experiment, the rms level of these filtered stimuli were equalized digitally using software before recording on a DAT (Sony DTC-1000ES).

Finally the signal and masker were calibrated at 60dB $L_{Aeq,1min}$ using the apparatus described in section 3.2.4.2.2, before being presented to the subject through the headphones. However, the level of the stimuli could not be calibrated at exactly 60dB $L_{Aeq,1min}$ because the level of speech over time is varied. Consequently, all stimuli were calibrated at the level of 60±1 dB $L_{Aeq,1min}$. the signal- and masker-calibrated levels of all stimuli are shown in the table above.

5.2.3 Experimental session

Table 5.2: Experimental session part 1

male speech

no.	signal (frequency range)	masker (frequency range)	Delay (ms)
1	Normal	Normal	15
2	Normal	Normal	30
3	Normal	Normal	50
4	Without gap	without gap	15
5	Without gap	without gap	30
6	Without gap	without gap	50

female speech

no.	signal (frequency range)	Masker (frequency range)	Delay (ms)
1	Normal	Normal	15
2	Normal	Normal	30
3	Normal	Normal	50
4	without gap	without gap	15
5	without gap	without gap	30
6	without gap	without gap	50

Table 5.3 Experimental session part 2

				Signal-Masker condition				
1	Lo	Lo	15	$(SL\text{-}ML)_{15}$	1	Lo	Lo	15
2	Lo	Mid	15	$(SL\text{-}MM)_{15}$	2	Lo	Mid	15
3	Lo	Hi	15	$(SL\text{-}MH)_{15}$	3	Lo	Hi	15
4	Mid	Lo	15	$(SM\text{-}ML)_{15}$	4	Mid	Lo	15
5	Mid	Mid	15	$(SM\text{-}MM)_{15}$	5	Mid	Mid	15
6	Mid	Hi	15	$(SM\text{-}MH)_{15}$	6	Mid	Hi	15
7	Hi	Lo	15	$(SH\text{-}ML)_{15}$	7	Hi	Lo	15
8	Hi	Mid	15	$(SH\text{-}MM)_{15}$	8	Hi	Mid	15
9	Hi	Hi	15	$(SH\text{-}MH)_{15}$	9	Hi	Hi	15
10	Lo	Lo	30	$(SL\text{-}ML)_{30}$	10	Lo	Lo	30
11	Lo	Mid	30	$(SL\text{-}MM)_{30}$	11	Lo	Mid	30
12	Lo	Hi	30	$(SL\text{-}MH)_{30}$	12	Lo	Hi	30
13	Mid	Lo	30	$(SM\text{-}ML)_{30}$	13	Mid	Lo	30
14	Mid	Mid	30	$(SM\text{-}MM)_{30}$	14	Mid	Mid	30
15	Mid	Hi	30	$(SM\text{-}MH)_{30}$	15	Mid	Hi	30
16	Hi	Lo	30	$(SH\text{-}ML)_{30}$	16	Hi	Lo	30
17	Hi	Mid	30	$(SH\text{-}MM)_{30}$	17	Hi	Mid	30

18	Hi	Hi	30	(SH-MH) 30	18	Hi	Hi	30
19	Lo	Lo	50	(SL-ML) 50	19	Lo	Lo	50
20	Lo	Mid	50	(SL-MM) 50	20	Lo	Mid	50
21	Lo	Hi	50	(SL-MH) 50	21	Lo	Hi	50
22	Mid	Lo	50	(SM-ML) 50	22	Mid	Lo	50
23	Mid	Mid	50	(SM-MM) 50	23	Mid	Mid	50
24	Mid	Hi	50	(SM-MH) 50	24	Mid	Hi	50
25	Hi	Lo	50	(SH-ML) 50	25	Hi	Lo	50
26	Hi	Mid	50	(SH-MM) 50	26	Hi	Mid	50
27	Hi	Hi	50	(SH-MH) 50	27	Hi	Hi	50

Each subject attended a total of two sessions. In each session, the experiment lasted for about 45 minutes, including a short break in order to avoid fatigue. In each session there were 34 conditions, including both parts of the experiment and the absolute threshold. The first session of the experiment was the male speech whereas the second session of the experiment was the repeat of the first session with female speech stimuli. This has been done in order to make the subject familiar with the stimuli in each session, thus the subjects were be able to detect the stimuli in different frequency bands. Results of part 1 and part 2 are shown in the table 5.2 and table 5.3.

5.2.4 Apparatus

The masking threshold was measured using the apparatus already described in section 3.2.4.1. This apparatus has been described in detail in section 3.2.4.2 together with its calibration in section 3.2.4.3. The stimuli used in this experiment were male and female running speech that was defined in section 5.2.2.

5.2.5 Subject

According to power analysis calculation (see section 3.3.3.2), the number of subjects required in the first part of the experiment was six, while the second part required seven. In order to make it more convenient, the number of subjects required for both part should be equal. However, for this experiment, there were seven subjects available; therefore, seven subjects attended the experiment in both parts. They consisted of five male and two female university students, 21-33 years of age. Two of them had had some experience in psychoacoustic experiments and took part in the pilot test.

All subjects were tested as having normal hearing before taking part in the experiment. The hearing test procedure was described in section 3.3.4.1.2. All subjects have a hearing threshold of less than 20dB in both ears at all frequencies tested.

5.2.6 Procedure

In the experiment, subjects were instructed to press the response button when they heard or thought that they heard the first indicator of speech. Prior to the actual tests, the subjects trained to make sure that they were familiar with the procedure.

Details of the operational procedure were described in section 3.2.4.4.

The run of measured thresholds consisted of 10 reversals. The level of the first 4 reversals had not been recorded. Only the levels for the last 6 reversals were read and recorded. The mean of the last 6 levels was taken as the threshold for that orientation as explained in section 3.4.

5.3 Results

In this experiment, the aim is to understand how the speech signal was detected in the speech masker during simultaneous and non-simultaneous masking effects. The subsidiary aim is to identify the dominant type of masking and to examine the effect of time intervals between the signal and the masker, or the delay in speech detection under headphone listening conditions. In this experiment, the study was conducted into two parts. The first part was the investigation of the effect of a silent gap in the speech masker, whereas the second part was the investigation on an upward spread masking effect when speech was presented together with delayed same-speech. The results for both parts of this experiment were taken by measuring the threshold using the method described in section 3.4. The results and analysis of data collected for both tests will be described below.

5.3.1 Part 1: Investigation on effect of silent gap

5.3.1.1 Absolute threshold

At the beginning of the first section of this experiment, the absolute thresholds of both male and female speech with a silent gap and male and female speech without a silent gap were measured. The mean of the absolute threshold for male speech with a silent gap was 13.0dBA; 13.3dBA for male speech without silent gap; 12.4 dBA for female speech with silent gap and 12.7dBA for female speech without a silent gap. From the results, it can be seen that the absolute threshold under headphones condition are higher than that under free field condition by approximate 6dB (see experiment 1). A portion of this difference is due to head (and body) diffraction effects. The absolute threshold measured in the free field reflects the minimum audible sound as presented in a sound field via a loudspeakers or "Minimum audible field (MAF)" response. These results reflect the binaural sensitivity of hearing, where the sound source was located directly in front of the listeners (0 degree azimuth). However in this experiment, the absolute threshold was measured when the stimuli were presented via headphones rather than by loudspeaker in a sound field. This is called the "Minimum audible pressure (MAP)". The results from MAP will be higher than MAF. That indicates less sensitive hearing. This is due to head and body diffraction effect. It is sometimes called the "missing 6dB" due to the result of using headphone (Durrant and Lovrinic, 1995). Therefore the results between two experiments cannot be compared. From the results in this experiment, the distribution of data was tested for normality using the Shapiro-Wilk test (which is more reliable when subjects are less than 50) as shown in table 5.4.

Table 5.4: The test of normality using Shapiro-Wilk test

	(mean±SD)	statistic	Sig.
Male with silent gap	13.0±0.6	.914	.426
Male without silent gap	13.2±0.6	.869	.183
Female with silent gap	12.4±0.7	.919	.464
Female without silent gap	12.7±0.8	.867	.176

* p <0.05, ** p<0.01

From table 5.4, it may be assumed that the data is adequately normal (p>0.05) to perform a parametric test. At first glance, it appears that the mean absolute threshold of male speech without a silent gap was a little bit higher than that of female speech with a silent gap. To investigate the difference among these mean absolute thresholds, repeated measures were used. The factors in this case are male speech with a silent gap, male speech without a silent gap, female speech with a silent gap, and female speech without a silent gap. The epsilon of less than 0.75 indicated that the sphericity was violated (see details in chapter 3). Therefore, the results of the Greenhouse-Geisser corrections (Univariate test) are used. The tabled results provided below.

Table 5.5: Absolute threshold level comparing male speech with a silent gap, male speech without a silent gap, female speech with a silent gap, and female speech without a silent gap.

	Male with silent gap (mean±SD)	Male without silent gap (mean±SD)	Female with silent gap (mean±SD)	Female without silent gap (mean±SD)	F	Sig.
Absolute threshold	13.0±0.6	13.2±0.6	12.4±0.7	12.7±0.8	.343	.705

From table 5.5, the univariate ANOVAs indicated that no difference is observed among the absolute threshold. This important statistic means the absolute threshold among these speeches was the same level.

5.3.1.2 *Signal type*

To investigate the effect of signal type, the hypothesis is that masking thresholds obtained for male speech and female speech are different. The repeated measures that factor in this case are male speech and female speech. The epsilon of 1.000 indicates perfect sphericity. The sphericity assumption is met (see details in chapter 3). Of the multivariate test results given, Wilks' Lambda is used as shown in Table 5.6.

Table 5.6: Multivariate analysis of male and female data

Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.
SPEECH	Pillai's Trace	.933	2.310(a)	6.000	1.000	.465
	Wilks' Lambda	**.067**	**2.310(a)**	**6.000**	**1.000**	**.465**
	Hotelling's Trace	13.860	2.310(a)	6.000	1.000	.465
	Roy's Largest Root	13.860	2.310(a)	6.000	1.000	.465

a Exact statistic
b Design: Intercept Within Subjects Design: SPEECH
c Tests are based on averaged variables.

From the table 5.6, the multivariate analysis yields significance at the 0.465 level for speech types. That means no significant effect was found for speech type (Lambda = 2.310, p = .465). Therefore it would be considered that the threshold of male and female speech were at the same level. The results can also be seen in figure 5.2 and figure 5.3

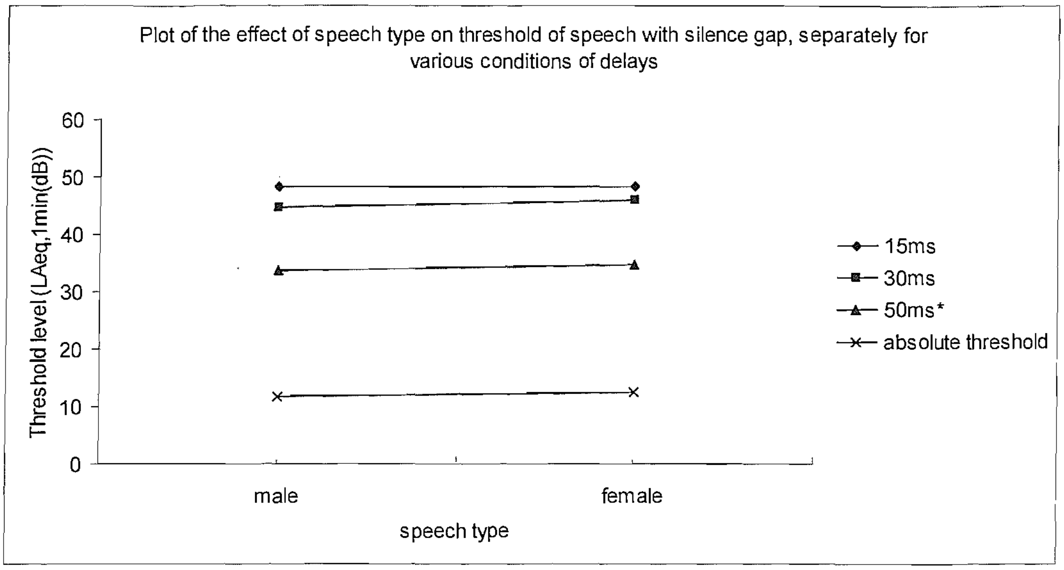


Figure 5.2: Effect of speech type on threshold when signal and masker are the speech with a silent gap, separated for various conditions of delay. (* p <0.05, ** p<0.01)

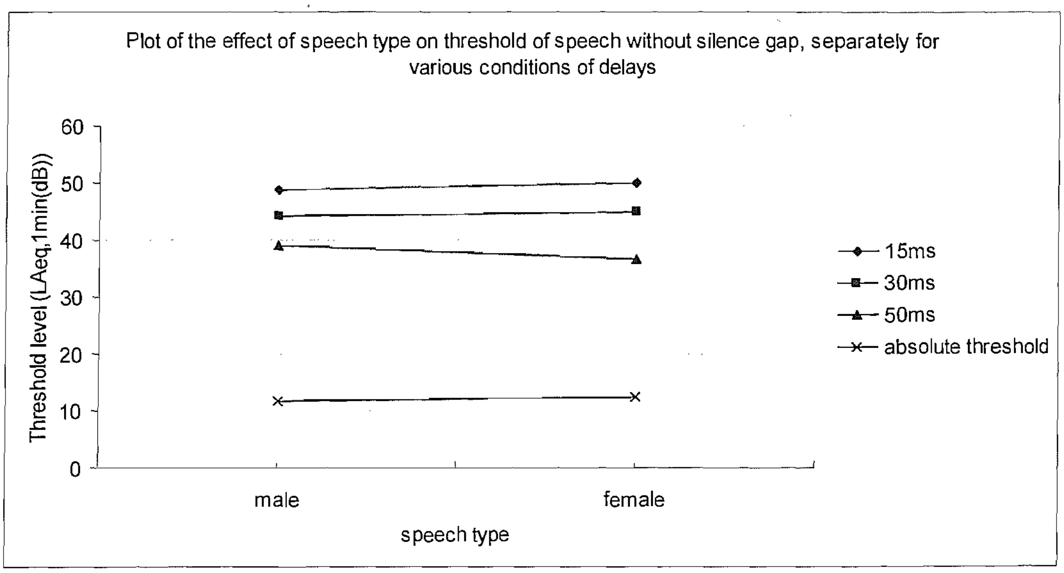


Figure 5.3: Effect of speech type on threshold when signal and masker are the speech without a silent gap, separated for various conditions of delay. (* p <0.05, ** p<0.01)

From the figure 5.2 and figure 5.3, it can be observed that the threshold of male and female speech is slightly different when the delay was 50ms. In order to have a closer look at the results, the univariate results and specific dependent variables (various conditions of delays, and silent gap conditions) were measured. The tabled results are provided below.

Table 5.7: Threshold level comparing male speech to female speech under various conditions of azimuth angle of incidence, delays, and masker's levels by using univariate ANOVA.

	male (mean±SD)	female (mean±SD)	F-Value	Significance
with gap, 15ms	48.3±2.6	48.4±2.3	0.020	.893
with gap, 30ms	44.6±0.5	45.8±1.2	5.535	.057
with gap, 50ms	33.7±1.8	34.6±0.9	6.357	.045*
without gap, 15ms	48.7±2.1	50.0±0.0	2.802	.145
without gap, 30ms	44.1±1.3	45.0±0.0	3.050	.131
without gap, 50ms	39.0±1.0	36.6±3.0	3.047	.131

* $p<0.05$
** $p<0.01$

The univariate ANOVAs indicated that a statistically significant difference is observed between male and female speech at a significant level ($p = 0.05$) only when both male and female speech with a silent gap (normal speech) was delayed for 50ms ($F=6.357$, $p < 0.05$). That means the difference in the threshold between male and female might be influenced by delays and silent gaps. For further clarification, the effect of delay and silent gap will be investigated.

5.3.1.3 *Time interval between signal and masker or delay*

To investigate the effect of delay, the hypothesis is that masking thresholds obtained for time intervals between a signal and masker of 15ms, 30ms, and 50ms are different. The repeated measures factored in this case were at four levels – 15ms, 30ms, and 50ms. The average epsilon is <0.75, indicatings that sphericity is violated (see details in chapter 3). Therefore the results of the Greenhouse-Geisser corrections (Univariate test) are used. The univariate results and specific dependent variables (various conditions of speech type, azimuth angle of incidence, and masker's levels) were measured. Tabled results are provided below.

Table 5.8: Threshold level comparing delays of 15ms, 30ms, and 50ms under various conditions of speech by using univariate ANOVA.

	15ms (mean±SD)	30ms (mean±SD)	50ms (mean±SD)	F-Value	Sig.
Male with silent gap	48.3±2.6[a]	44.6±0.5[b]	33.7±1.8[c]	102.184	.000**
Male without silent gap	48.7±2.1[a]	44.1±1.3[b]	39.0±1.0[c]	113.237	.000**
Female with silent gap	48.4±2.3[a]	45.8±1.2[a]	34.6±0.9[b]	214.687	.000**
Female without silent gap	50.0±0.0[a]	45.0±0.0[b]	36.6±3.0[c]	103.176	.000**

* $p<0.05$
** $p<0.01$

The univariate ANOVAs indicated that a statistically significant difference is observed among delays of 15ms, 30ms, and 50ms at a significant level ($p<0.01$). Since the threshold is significantly affected by delays, what has to be determined is which delays differ significantly for the threshold under each condition. Therefore each pair of delays was tested using a paired sample T-Test. The results of paired T-Test have been summarized in table 5.8 in term of superscripts a, b, c, d. The same superscripts indicate that there is no significant mean threshold difference between delays. On the other hand, the different superscripts indicate that mean thresholds between delays are different. The results can also be observed in figure 5.4.

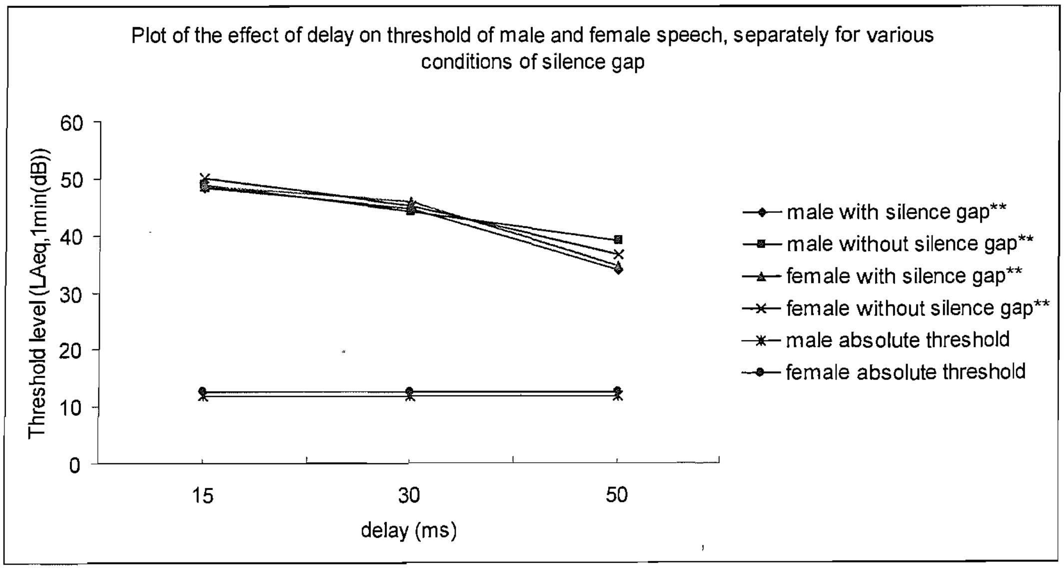


Figure 5.4: The effect of delay on thresholds for male speech and female speech, separately for various conditions of silent gap. (* $p<0.05$, ** $p<0.01$)

From the figure 5.4, it can be seen clearly that the mean thresholds decline as the time interval between signal and masker increased toward 50ms at a significant level ($p < 0.01$) under all conditions investigated.

5.3.1.4 *Silent interval or silent gap*

To investigate the effect of a silent gap, the hypothesis is that masking thresholds obtained for speech with a silent gap and speech without a silent gap are different. The repeated measures in this case are two levels: speech with a silent gap and speech without a silent gap. The epsilon of 1.000 indicates perfect sphericity. That is, the sphericity assumption is met (see details in chapter 3). Of the multivariate test results given, Wilks' Lambda is used as shown in Table 5.9.

Table 5.9: Multivariate analysis of speech with silent gap and speech without silent gap data

Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.
GAP	Pillai's Trace	.942	2.715(a)	6.000	1.000	.434
	Wilks' Lambda	**.058**	**2.715(a)**	**6.000**	**1.000**	**.434**
	Hotelling's Trace	16.292	2.715(a)	6.000	1.000	.434
	Roy's Largest Root	16.292	2.715(a)	6.000	1.000	.434

a Exact statistic
b Design: Intercept Within Subjects Design: GAP
c Tests are based on averaged variables.

From the table 5.9, the multivariate analysis yields significance at the 0.434 level for both speech with a silent gap and speech without a silent gap. That means no significant effect was found for the silent gap (Lambda = 2.715, p = .434). Therefore it would be considered that the threshold of speech with a silent gap and of speech without a silent gap were at the same level. The results can also be seen in figure 5.5 and figure 5.6

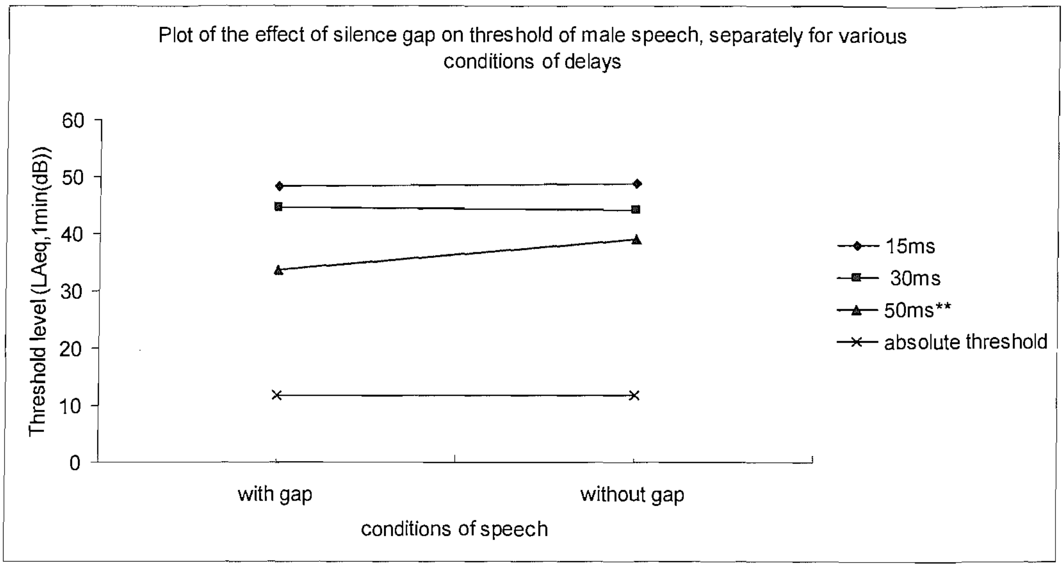


Figure 5.5: Effect of a silent gap on thresholds for male speech, separated for various conditions of delay. (* p <0.05, ** p<0.01)

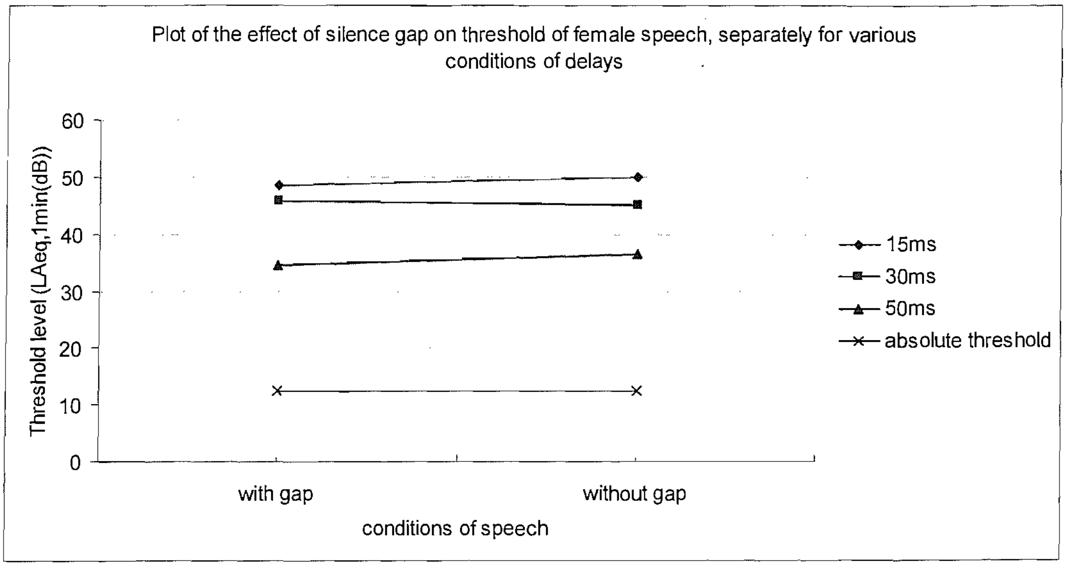


Figure 5.6: Effect of a silent gap on threshold for female speech, separated for various conditions of delay. (* p <0.05, ** p<0.01)

From the figure 5.5 and figure 5.6, it can be clearly seen that the threshold of male speech without a silent gap was different from that of male speech with a silent gap when the delay was 50ms. Therefore, in order to have a closer look at the results, the univariate results and specified dependent variables (various conditions of speech types and delays) were measured. The tabled results are provided below.

Table 5.10: Threshold level comparing speech with a silent gap to speech without a silent gap under various conditions of speech type and delays by using univariate ANOVA.

	with gap (mean±SD)	without gap (mean±SD)	F-Value	Significance
male, 15ms	48.3±2.6	48.7±2.1	0.636	.455
male, 30ms	44.6±0.5	44.1±1.3	0.815	.401
male, 50ms	33.7±1.8	39.0±1.0	33.503	.001**
female, 15ms	48.4±2.3	50.0±0.0	3.128	.127
female, 30ms	45.8±1.2	45.0±0.0	3.460	.112
female, 50ms	34.6±0.9	36.6±3.0	4.173	.087

* p<0.05
** p<0.01

Following table 5.10, the univariate ANOVAs indicated that a statistically significant difference is observed between speech with a silent gap and speech without a silent gap at a significant level ($p < 0.01$) only when they are male speech with a delay of 50ms ($F=33.503$, $p < 0.01$).

From multivariate, the finding indicates that vowel sounds in speech could have been detected during the masker period represented by silent gap or lower amplitude consonant sounds. But together with the univariate ANOVAs, it appears that the masking threshold levels of speech with and without a silent gap for a delay of 50ms are different. This finding could be related to the typical duration of actual speech component sounds in relation to simultaneous masking effects and critical band concepts. Also, when editing the male speech stimuli to remove the silent gaps, it was noted that some silent gaps were longer in duration than some of the low amplitude consonants present. That means the masking threshold of speech that contains long silent gaps was lower than that of speech that contains a short duration of low amplitude consonants.

5.3.2 Part 2: Investigation on upward spread masking effect

5.3.2.1 Absolute threshold

At the beginning of the experiment, the absolute thresholds of low-, medium- and high frequency filtered male speech and low- medium-, and high-frequency filtered female speech were measured. The distribution of data was tested for normality using the Shapiro-Wilk test (which is more reliable when the subjects are less than 50) as shown in table 5.11.

Table 5.11: The test of normality using Shapiro-Wilk test

	(mean±SD)	statistic	Sig.
Male low frequency	13.0±0.6	.914	.426
Male mid frequency	13.2±0.6	.869	.183
Male high frequency	13.0±0.6	.914	.426
Female low frequency	13.2±0.6	.869	.183
Female mid frequency	12.4±0.7	.919	.464
Female high frequency	12.7±0.8	.867	.176

From table 5.11, it may be assumed that the data is normal enough ($p>0.05$) to perform a parametric test. On first glance, it appears that absolute thresholds of low-, medium- and high frequency filtered male speech and low- frequency filtered female speech were a little bit higher than that of medium- and high frequency filtered female speech. To investigate the differences among these mean absolute thresholds, the repeated measures were used. An epsilon of less than 0.75 indicated that the sphericity is violated (see details in chapter 3). Therefore the results of the Greenhouse-Geisser corrections (Univariate test) are used. The tabled results provided below.

Table 5.12: Absolute thresholds of low-, medium- and high frequency filtered male speech and low- medium-, and high-frequency filtered female speech by using univariate ANOVA.

	Male			Female				
	low freq. (mean±SD)	mid freq. (mean±SD)	hi freq. (mean±SD)	low freq. (mean±SD)	mid freq. (mean±SD)	hi freq. (mean±SD)	F	Sig.
absolute threshold	13.0±0.6	13.2±0.6	13.0±0.6	13.2±0.6	12.4±0.7	12.7±0.8	.594	.620

From the table 5.12, the univariate ANOVAs indicated that no statistically significant difference is observed among the absolute thresholds.

5.3.2.2 ***Signal type***

To investigate the effect of signal type, the hypothesis is that masking thresholds obtained for male speech and female speech are different. The two repeated measures used in this case are male and female speech. The epsilon of 1.000 indicates perfect sphericity. That is the sphericity assumption is met (see details in chapter 3). Of the multivariate test results given, Wilks' Lambda is used as shown in Table 5.13.

Table 5.13: Multivariate analysis of male and female data

Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.
SPEECH	Pillai's Trace	.971	5.678(a)	6.000	1.000	.311
	Wilks' Lambda	**.029**	**5.678(a)**	**6.000**	**1.000**	**.311**
	Hotelling's Trace	34.069	5.678(a)	6.000	1.000	.311
	Roy's Largest Root	34.069	5.678(a)	6.000	1.000	.311

a Exact statistic
b Design: Intercept Within Subjects Design: SPEECH
c Tests are based on averaged variables.

From the table above, the multivariate analysis yields significance at a level of .311 for speech types. That means no significant effect was found for speech types (Lambda = 5.678, p = .311). Therefore, it would be considered that the threshold of male and of female speech were at the same level. The results can also be seen in figure 5.7, figure 5.8 and figure 5.9.

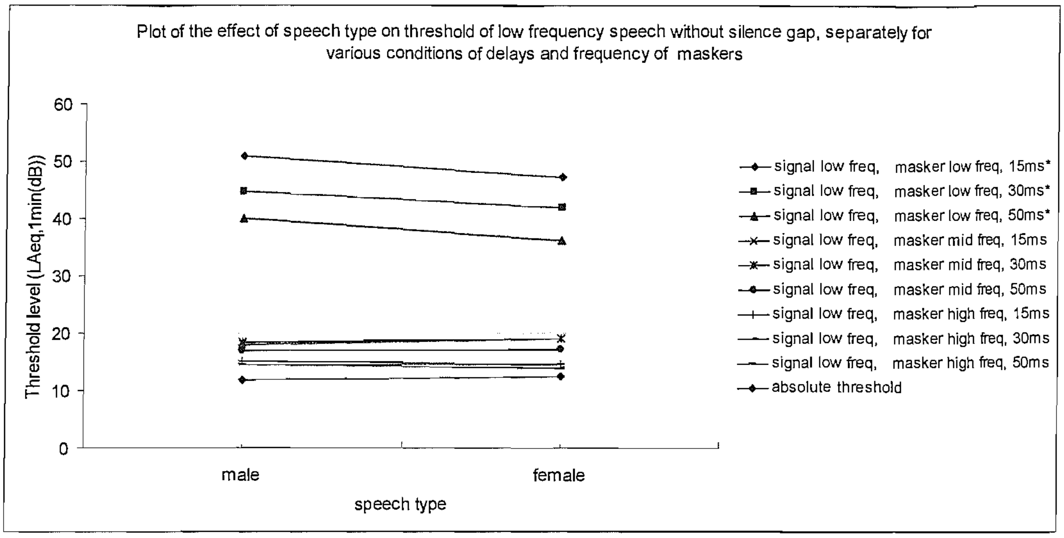


Figure 5.7: Effect of speech type on threshold of low frequency speech without silence gap, separately for various conditions of delays and frequency of maskers. (* p <0.05, ** p<0.01)

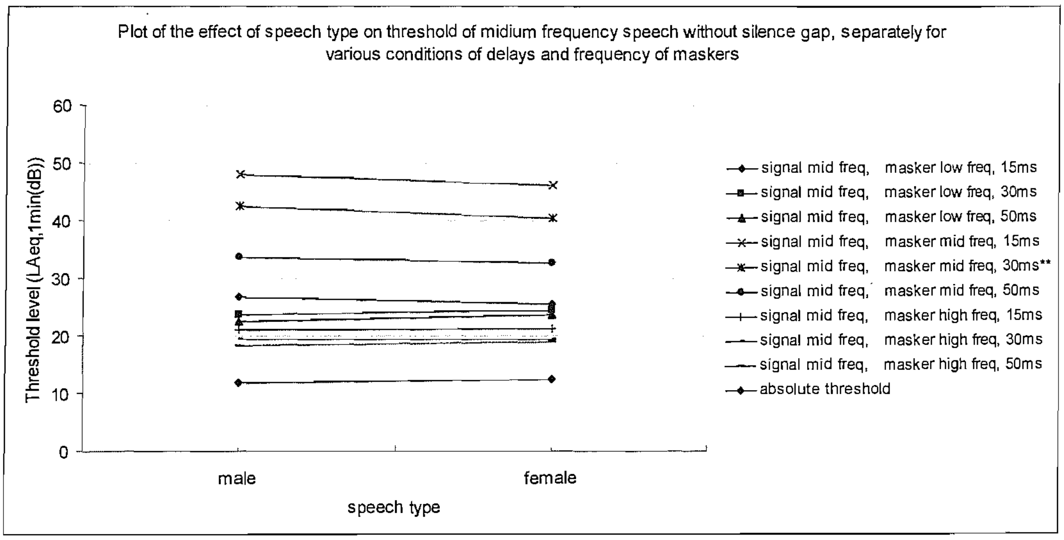


Figure 5.8: Effect of speech type on threshold of medium frequency speech without silence gap, separately for various conditions of delays and frequency of maskers. (* p <0.05, ** p<0.01)

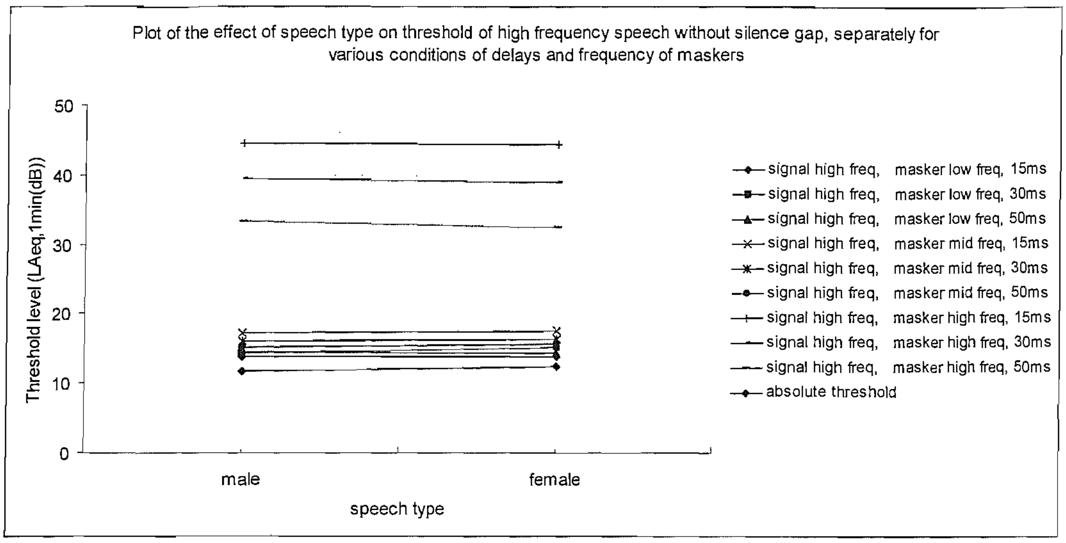


Figure 5.9: Effect of speech type on the threshold of high frequency speech without silence gap, separately for various conditions of delays and frequency of maskers. (* p <0.05, ** p<0.01)

From figure 5.7, figure 5.8, and figure 5.9, it can be observed that the threshold of male and females is slightly different when the signal and masker were in the same frequency range. In order to have a closer look at the results, the univariate results and specify the dependent variables (various conditions of a signal's frequency, masker's frequency, and delays conditions) were measured. The tabled results are provided below.

Table 5.14: Threshold level comparing male speech to female speech under various conditions of signal's frequency, masker's frequency, and delay by using univariate ANOVA.

		male (mean±SD)	female (mean±SD)	F-Value	Significance
signal low freq,	masker low freq, 15ms	50.8±2.2	47.1±2.0	10.937	.016*
signal low freq,	masker low freq, 30ms	44.6±0.5	41.9±2.3	9.323	.022*
signal low freq,	masker low freq, 50ms	40.0±0.7	36.2±3.6	6.682	.041*
signal low freq,	masker mid freq, 15ms	17.9±1.9	19.0±1.9	1.303	.297
signal low freq,	masker mid freq, 30ms	18.4±0.9	19.0±1.1	.797	.406
signal low freq,	masker mid freq, 50ms	17.0±1.4	17.1±1.6	.031	.867
signal low freq,	masker high freq, 15ms	15.1±2.4	14.5±2.1	1.647	.247
signal low freq,	masker high freq, 30ms	14.4±1.2	13.9±1.9	.398	.551

signal low freq,	masker high freq, 50ms	14.4±1.4	14.5±1.3	.037	.853
signal mid freq,	masker low freq, 15ms	26.8±2.1	25.6±2.5	.837	.395
signal mid freq,	masker low freq, 30ms	23.7±3.1	24.4±3.1	.433	.535
signal mid freq,	masker low freq, 50ms	22.5±3.5	23.7±4.3	.582	.474
signal mid freq,	masker mid freq, 15ms	48.0±0.9	46.2±1.2	22.605	.003**
signal mid freq,	masker mid freq, 30ms	42.5±1.7	40.5±4.6	2.131	.195
signal mid freq,	masker mid freq, 50ms	33.6±1.3	32.6±5.4	.215	.660
signal mid freq,	masker high freq, 15ms	21.1±1.8	21.3±1.6	.046	.837
signal mid freq,	masker high freq, 30ms	19.3±1.6	19.3±2.5	.000	1.000
signal mid freq,	masker high freq, 50ms	18.4±2.0	19.0±3.1	.219	.656
signal high freq,	masker low freq, 15ms	13.8±2.0	13.8±2.0	.002	.967
signal high freq,	masker low freq, 30ms	14.3±1.4	15.2±2.3	.999	.356
signal high freq,	masker low freq, 50ms	14.5±2.2	14.3±2.9	.031	.867
signal high freq,	masker mid freq, 15ms	17.2±2.2	17.6±2.2	.096	.768
signal high freq,	masker mid freq, 30ms	16.0±2.1	16.3±2.7	.065	.808
signal high freq,	masker mid freq, 50ms	15.2±2.5	15.6±3.0	.081	.786
signal high freq,	masker high freq, 15ms	44.5±0.8	44.4±0.9	.066	.805
signal high freq,	masker high freq, 30ms	39.4±2.0	39.0±4.1	.062	.811
signal high freq,	masker high freq, 50ms	33.3±1.5	32.5±0.5	2.066	.201

* p<0.05

** p<0.01

From multivariate, the finding indicates that the threshold of male speech is the same as the threshold for female speech. But the univariate ANOVAs showed that a statistically significant difference is observed between male and female speech at a significant level ($p = 0.05$) when both male and female speech are low frequency with a delay of 15ms ($F=10.931$, $p < 0.05$), 30ms ($F=9.323$, $p < 0.05$), and 50ms ($F=6.682$, $p < 0.05$). The univariate ANOVAs also indicated that a statistically significant difference is observed between male and female speech at a significant level ($p = 0.01$) when both male and female speech are mid frequency with a delay of 15ms ($F=22.625$, $p < 0.01$). This finding could be related to the typical duration of actual speech component sounds in relation to the duration of each phoneme. For tested material, male speech tends to have longer phonemes than that of female speech. Therefore, when speech stimuli were filtered into different frequency bands, it was noted that the stimuli may have different in relative phoneme duration and gap duration. That means the masking threshold of male speech that contains a long duration of phonemes (as a masker) is higher than that of female speech that contains a short duration of phonemes (as masker), because the masking threshold is affected by the duration of the masker, especially when signal frequency and masker frequency are the same.

5.3.2.3 *Time interval between signal and masker or delay*

To investigate the effect of delay, the hypothesis states that masking thresholds obtained for time intervals between signal and masker of 15ms, 30ms, and 50ms are different. The factors in this case are four levels: 15ms, 30ms, and 50ms. The average epsilon of <0.75 indicates the sphericity is violated (see details in chapter 3). Therefore the results of the Greenhouse-Geisser corrections (Univariate test) are used. The univariate results and specific dependent variables (various conditions of speech type, azimuth angle of incidence, and masker's levels) were measured. The tabled results are provided below.

Table 5.15: Threshold level comparing among delays of 15ms, 30ms, and 50ms under various conditions of speech by using univariate ANOVA

	delay 15ms (mean±SD)	Delay 30ms (mean±SD)	delay 50ms (mean±SD)	F	Sig.
male, signal low freq, masker low freq	50.8±2.2[a]	44.6±0.5[b]	40.0±0.7[c]	141.099	.000**
male, signal low freq, masker mid freq	17.9±1.9[a]	18.4±0.9[a]	17.0±1.4[a]	2.278	.160
male, signal low freq, masker high freq	15.1±2.4[a]	14.4±1.2[a]	14.4±1.4[a]	.811	.450
male, signal mid freq, masker low freq	26.8±2.1[a]	23.7±3.1[a]	22.5±3.5[a]	6.325	.034*
male, signal mid freq, masker mid freq	48.0±0.9[a]	42.5±1.7[b]	33.6±1.3[c]	318.761	.000**
male, signal mid freq, masker high freq	21.1±1.8[a]	19.3±1.6[a]	18.4±2.0[a]	4.363	.051
male, signal high freq, masker low freq	13.8±2.0[a]	14.3±1.4[a]	14.5±2.2[a]	.633	.483
male, signal high freq, masker mid freq	17.2±2.2[a]	16.0±2.1[a,b]	15.2±2.5[b]	9.781	.015*
male, signal high freq, masker high freq	44.5±0.8[a]	39.4±2.0[b]	33.3±1.5[c]	73.223	.000**
female, signal low freq, masker low freq	47.1±2.0[a]	41.9±2.3	36.2±3.6	113.878	.000**
female, signal low freq, masker mid freq	19.0±1.9[a]	19.0±1.1[a]	17.1±1.6[a]	3.592	.096
female, signal low freq, masker high freq	14.5±2.1[a]	13.9±1.9[a]	14.5±1.3[a]	1.312	.302
female, signal mid freq, masker low freq	25.6±2.5[a]	24.4±3.1[a]	23.7±4.3[a]	2.733	.134
female, signal mid freq, masker mid freq	46.2±1.2[a]	40.5±4.6[a]	32.6±5.4[b]	37.292	.000**
female, signal mid freq, masker high freq	21.3±1.6[a]	19.3±2.5[a]	19.0±3.1[a]	4.736	.032*
female, signal high freq, masker low freq	13.8±2.0[a]	15.2±2.3[a]	14.3±2.9[a]	.987	.370
female, signal high freq, masker mid freq	17.6±2.2[a]	16.3±2.7[a,b]	15.6±3.0[b]	8.345	.009**
female, signal high freq, masker high freq	44.4±0.9[a]	39.0±4.1[b]	32.5±0.5[c]	44.710	.000**

* p<0.05 ** p<0.01

The univariate ANOVAs indicated that a statistically significant difference is observed among delays of 15ms, 30ms, and 50ms at a significant level (p<0.01). Since the threshold is significantly affected by delays, what has to be determined is which delays differ significantly for the threshold under each condition. Therefore, each pair of delays was tested using a paired sample T-Test. The results of the pair T-Test have been summarized in table 5.15 in terms of superscripts a, b, c, d. The same superscripts indicate that there is no significant mean threshold difference between delays. On the other hand, the different superscripts indicate that mean thresholds between delays are different. The results can also be observed in figure 5.10 and figure 5.11.

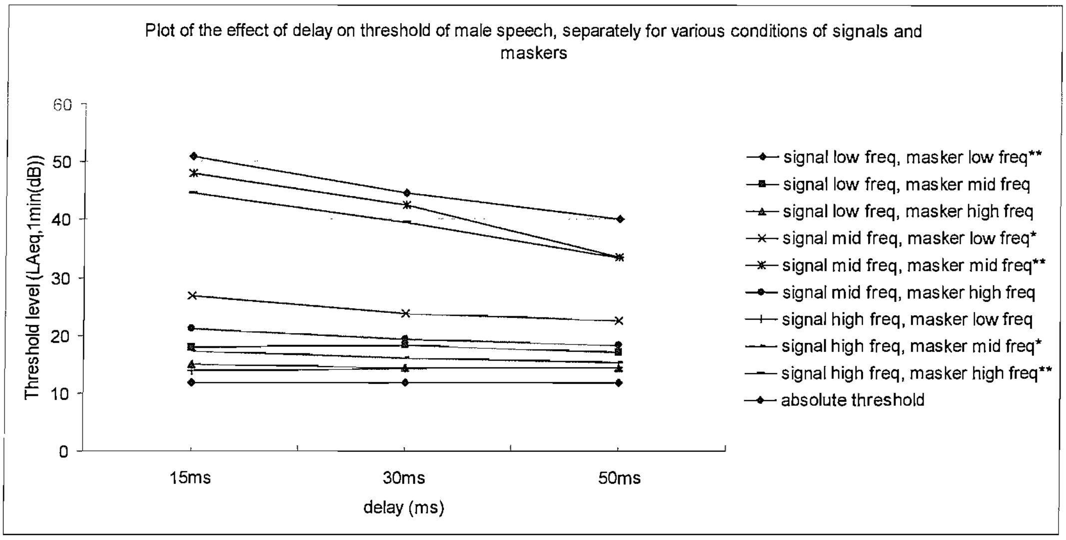


Figure 5.10: The effects of delay on threshold for male speech, separated for various conditions of a silent gap. (* p <0.05, ** p<0.01)

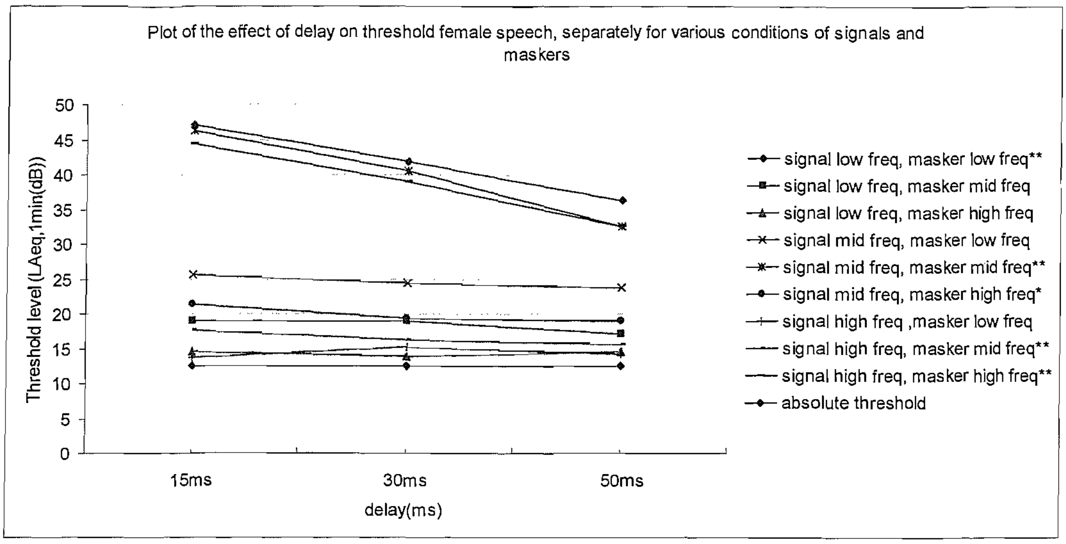


Figure 5.11: The effect of delay on threshold for female speech, separated for various conditions of a silent gap. (* $p < 0.05$, ** $p<0.01$)

From figure 5.10 and figure 5.11, it can be seen clearly that the results can be separated into two groups. One is comprised of the masking thresholds when a signal's frequency and masker's frequency are the same, and the other is that the masking thresholds when signal's frequency and masker's frequency are difference.
When a signal's frequency and masker's frequency are the same, the univariate ANOVAs indicated that a statistically significant difference is observed among delays of 15ms, 30ms, and 50ms at a significant level ($p<0.01$). These masking thresholds were similar to those measured for normal speech (see section 5.3.1.3) that is consistent with known non-simultaneous masking effects (see chapter 2). When the threshold patterns for each frequency were observed, it can be seen that the thresholds of the high frequency are found to be lower than that of low and mid frequency. These differences may be a result of differences in relative gap duration when the signals have been filtered into different frequency bands.

When the signal's frequency and masker's frequency are different, the univariate ANOVAs indicated that no statistically significant difference is observed among delays of 15ms, 30ms, and 50ms under almost all conditions tested. This suggests that only simultaneous masking effects could have been occurring under these circumstances. There is some indication of a small upwards spread of the masking effect when the frequency of the masker was adjacent to the frequency of the signal; for example, the masking threshold when the signal is mid frequency and the masker is low frequency was higher than that of the masking threshold of the remaining differentiated masker and signal frequencies.

CHAPTER 6

EXPERIMENT 3: Azimuth angle of incidence: *The role of head rotation on masking effects under free - field listening conditions*

6.1 Introduction

The main objective of this experiment is to enhance our understanding of the effect of masker azimuth angle of incidence on speech signals and the role that head rotation plays when speech is presented with a delayed same speech in free-field. Following the first experiment, the results showed that the masked threshold of normal speech when both the speech signal and speech masker were presented from the same azimuth angle was higher than that of normal speech when the speech signal was present from the front and speech masker was presented from a 30 degree, 60 degree, and 90 degree azimuth angle. The findings suggest that the results could be separated into two groups. One group was the worst when signal and masker are presented from the same direction at 0 degrees, while the other group was the threshold level that was separated from masking when the signal and masker are presented separately. Unfortunately, in the first experiment, the azimuth angles were chosen to be 0degrees, 30 degree, 60 degree, and 90 degree; therefore the transformation between the two groups of the threshold levels could not be investigated. In addition, the subjects were allowed to move their head naturally in the first experiment. This might have affected the threshold level due to possible changing binaural cues owing to the rotation of the head. Nevertheless, the effects of head rotation on speech masking have not yet been investigated.

Head rotation occurs naturally and is difficult to avoid. It plays an important role in spatial hearing, especially in sound localization. It is reported by Blauert (1993) that when a listener moves his head, the acoustic interference of the sound wave around the head changes, with both the cue Inter-aural intensity difference and Inter-aural time differences change accordingly. The changes of both inter-aural intensity and time differences due to head rotation improve the ability to determine the direction of sound incident. This makes it possible to determine of the direction of the sound source with great accuracy. Furthermore, the head rotation also leads to a reduction of confusion in front-back reversals (Kendall, 1995).

Consequently, many studies have investigated the effect of head rotation on localization only; none of these studies mentions the effect of head movement on un-masked binaural incidents.

In the past, the masking effect has always been studied under conditions in which the subject's head is held still, mechanically immobilized or with signals presented over a headphone. These studies have been carried out in order to understand how the binaural cues affect the masking phenomenon. In general, their results indicate that the binaural masking released is mainly provided by inter-aural time and intensity differences due to the head shadow effect (Brokhorst and Plomp, 1988) as described in section 2.5.4.

With head movement, the pattern of binaural masking released might be different from the pattern of binaural masking released under a fixed head condition because rotating the head changes the binaural cues accordingly. Many authors, for example Koenig (1950), have reported that there are changes in Inter-aural time and intensity difference due to head movement. If this is the case, it is possible that the head movement might affect the masking threshold of speech in the present with the same-speech delayed by time.

This experiment was carried out to test the effects of head rotation at various fine incident angles of azimuth. The masked threshold levels of the direct sound were measured in the presence of direct sound and the simulated delayed same-speech through the loudspeakers in an anechoic room. The subjects were instructed to either keep their head still against the head rest or to move their heads from one side to the other.

6.2 Experimental procedure

6.2.1 Experimental design

The experiment was designed to examine the effects of head rotation on the masking effect. The experiment was divided into two sessions: head-fixed and head-moving. Both sessions were conducted under the constant condition of direct sound, while the reflection was changed in incident angles of 0, 2, 4, 6, 8, 10, 15, 30, and 90 degrees (left side only). The direct frontal sound and the reflection delayed with respect to the direct sound were produced in an anechoic room using two loudspeakers, each at a distance of 2 meters from the center of the listener's head.

The masking threshold level of direct sound was measured while the amplitude of reflected sound was set at a comfortable listening level 60dB $L_{Aeq,1min}$ at the subject's head. The delay time was set to be as short as possible in order to avoid providing extra cues in signal detection. However it has to be long enough to assure incoherence of the two sounds. Therefore the delay of reflected sound was chosen to be 15ms with respect to the direct sound.

6.2.2 Stimuli and Calibration

Stimulus was male speech that had been used in the previous two experiments. The stimulus was edited digitally by using the apparatus described in section 3.2.4.2., in order to create a 15ms delay in the right channel as a masker, while the left signal remained non-delayed. After the stimulus was edited, they were recorded into DAT (Sony STS-1000ES) via the sound card of a personal computer. The signal was recorded for about 2 minutes per session. Both signal and masker were played back to the subjects via loudspeakers (KEF C35) in an anechoic room. The level of masker held constant at 60dB $L_{Aeq,1min}$ at the subject's head position, using the calibration method as described in section 3.2.4.3. Then, the masking threshold level of the speech signal was controlled and measured by the experimenter using an audiometer (Kamplex AD-27).

6.2.3 Experimental session

Each subject attended two sessions. One was head-fixed, the other was head-moving. In each session, the experiment lasted for about 20 minutes. There were a total of 10 conditions in each session, including absolute threshold of speech.

To avoid the possibility of the learning effect, the subjects who attended this experiment were separated into two groups. The first five subjects of group one attended the head-fixed session first, followed by the head-moving session. The second five subjects of group two attended the head-moving session before the head-fixed session. The experimental session is shown in the table 6.1.

Table 6.1: The experimental session

Head-fixed									
azimuth (degree)	0	2	4	6	8	10	15	30	90
Head-moving									
azimuth (degree)	0	2	4	6	8	10	15	30	90

6.2.4 Apparatus

The masking threshold was measured using the apparatus already described in section 3.2.4.1. This apparatus used to produce the stimuli have been described in detail in section 3.2.4.2.1 together with its calibration in section 3.2.4.3. The stimuli used in this experiment were running male and female speech that was defined in section 3.2.4.

6.2.5 Subject

According to power analysis calculation (see section 3.3.3), the number of subjects required in this experiment was seven. However, there were ten subjects available for the experiment. All subjects were university students and staff members between the ages of 21-33. The ten subjects consisted of eight males and two females. Two males had some experience in psychoacoustic experiments, as they took part in the pilot test.

All subjects were tested as having a normal hearing before taking part in the experiment. The hearing test procedure is described in section 3.3.4.1. All subjects have a hearing threshold of less than 20dB in both ears in all frequency tests.

6.2.6 Procedure

Measurements were carried out with the subject seated in an anechoic room. The absolute thresholds of speech were measured at the beginning of the first session.

In the head-fixed session, subjects were instructed to keep the head still by using the head rest. Subjects were also instructed to report to the experimenter if they noticed they might have moved their head even slightly, allowing for a re-measurement of the threshold.

In the session when head-moving was permitted, some subjects expressed concern because they did not know how much to move their head, or in which direction he has to turn. This issue was addressed by instructing the subject to move his head around the area between the sound source and the masker positions.

To measure the threshold, subjects were instructed verbally to press the response button when they heard, or thought they heard, speech from the direct sound at the front. Prior to the actual tests, the subjects trained to make sure that they were familiar with the procedure.

The details of the procedure are described in section 3.2.4.4.

The run of threshold measuring consisted of 10 reversals. The level of the first 4 reversals was not recorded. Only the levels of the last 6 reversals were read and recorded. The mean levels of the last 6 reversals were taken as the threshold for that orientation as explained in section 3.4.

6.3 Results

Because the experiment was designed to investigate the effects of head rotation on threshold levels, together with the effects of azimuth angles of incidence, the results will be measured using three factors: absolute threshold, head rotation, and azimuth angle of incidence.

6.3.1 *Absolute threshold*

At the beginning of the experiment, the absolute thresholds of the speech were measured. The means of the absolute threshold for tested speech was 5.7dBA. The lowest absolute threshold measured was 3.3dBA, whereas the highest absolute threshold measured was 9.2dB. The distribution of data was tested for normality using the Shapiro-Wilk test (which is more reliable when subjects are less than 50) as shown in table 6.2.

Table 6.2: The test of normality using Shapiro-Wilk test

	(mean±SD)	statistic	Sig.
Absolute threshold	5.7±1.7	.870	.100

From table 6.2, it may be assumed that the data is adequately normal, as the significant level is more than 0.05.

6.3.2 *Head rotation*

To investigate the effects of head rotation, the hypothesis is that masking thresholds obtained for head-fixed test and head–rotated tests are different. The repeated measures factor into this case as there are only two levels; head-fixed and head-rotated. With only two factors, the epsilon is always 1.000, which indicates perfect sphericity. That means the sphericity assumption is met. Therefore the multivariate test results gives, Wilks' Lambda is used as shown in Table 6.3.

Table 6.3: Multivariate analysis of head rotation data

Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.
HEAD	Pillai's Trace	.998	53.822(a)	9.000	1.000	.105
	Wilks' Lambda	**.002**	**53.822(a)**	**9.000**	**1.000**	**.105**
	Hotelling's Trace	484.396	53.822(a)	9.000	1.000	.105
	Roy's Largest Root	484.396	53.822(a)	9.000	1.000	.105

a Exact statistic
b Design: Intercept Within Subjects Design: HEAD
c Tests are based on averaged variables.

From table 6.3, the multivariate analysis yields significance at the 0.105 level for head rotation. That means no significant effect was found for head rotation (Lambda = 53.822, p= .105). Therefore, it may be assumed that the threshold of speech when the head is fixed and the threshold of speech when the head is moving were the same. The effect of head rotation on masking threshold under various conditions of azimuth angle of incidence and delay was plotted in the figure 6.1.

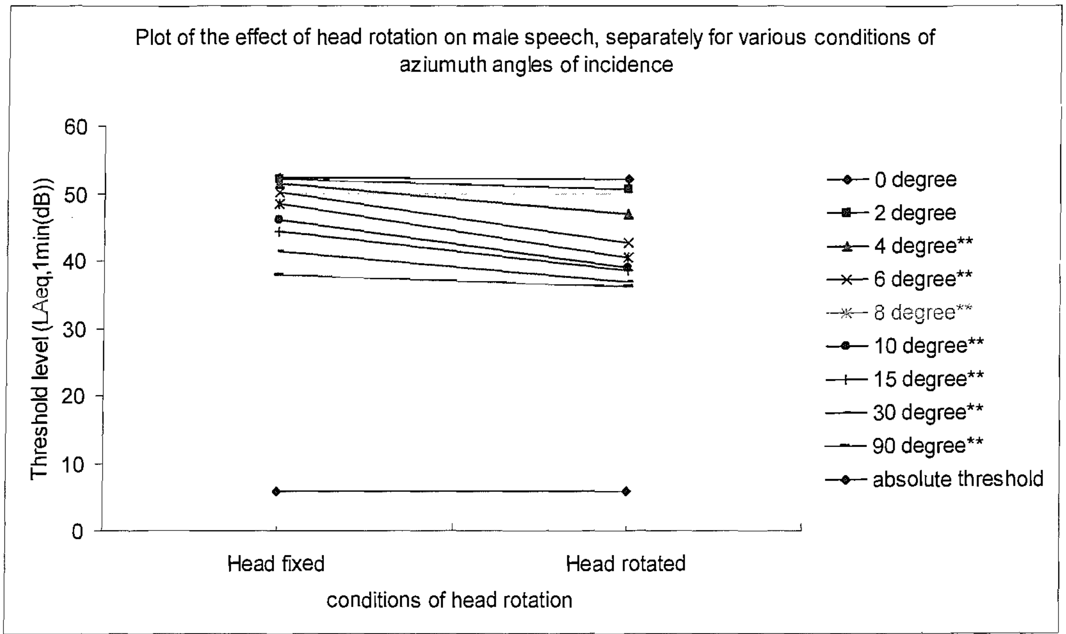


Figure 6.1: Effect of head rotation on threshold, separately for various conditions of azimuth angle of incidence (* $p < 0.05$, ** $p < 0.01$)

Although the multivariate indicated that the thresholds of speech with the head-fixed and thresholds of speech when the head is moving were the same, from figure 6.1 it can be seen that the difference between threshold of speech when the head is fixed and the threshold of speech when the head is moving is small when the masker arrived from a 0 degree azimuth angle of incidence. However, the difference was increased when the azimuth angle of incidence was increased. Until the masker arrived from a 90 degree azimuth angle of incidence, the difference between the threshold of head-fixed speech and the threshold of head-moving speech was small. In order to have a closer look at the results, the univariate results, specificlly the dependent variables (various conditions of azimuth angle of incidence and delays), were measured. The table 6.4 results provided below.

Table 6.4: Threshold of speech comparing head-fixed and head-moving tests under various conditions of azimuth angle of incidence by using univariate ANOVA

	Head-fixed (mean±SD)	Head-moving (mean±SD)	F-Value	Significance
0 degree	52.3±0.8	52.0±1.0	0.310	.591
2 degree	52.0±2.0	50.5±2.3	3.857	.081
4 degree	51.3±1.8	46.8±1.2	38.368	.000**
6 degree	50.0±2.4	42.6±2.8	120.049	.000**
8 degree	48.3±2.0	40.5±3.0	67.047	.000**
10 degree	46.0±1.7	39.0±2.1	87.425	.000**
15 degree	44.3±1.7	38.5±1.7	78.049	.000**
30 degree	41.3±1.8	36.8±1.1	59.754	.000**
90 degree	37.8±1.4	36.1±1.2	11.125	.009**

* p<0.05
** p<0.01

Following the univariate ANOVAs showed that a statistically significant difference is observed between the threshold of speech when the head is fixed and the threshold of speech when the head is moving at a significant level of $p < 0.01$ when the azimuth angle of incidence is 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees. That means the threshold of speech during head-fixed tests and the threshold of speech during head-moving tests were the same when the masker arrived from a 0 degree or 2 degree azimuth angle of incidence. When the azimuth angle of incidence was increased to 4 degrees, the threshold of speech during the head-fixed test and the threshold of speech during the head-moving test were observed to be statistically different. These differences increased when the azimuth angle of incidence increased toward 90 degrees.

These findings suggested that head rotation causes some changes in binaural cues and leads to a lower threshold level when azimuth angles of incidence are more than 4 degrees. On the other hand, the threshold when the azimuth angle is 2 degrees for both head-fixed and head-moving tests was observed to be as high as the poorest masking threshold when the azimuth angel was 0 degrees. One possible explanation might be that when the azimuth angle was 2 degrees or less, the two auditory events could not be separated because of the sound localization blur (Blauert, 1997).

6.3.3 *Azimuth angle of incidence*

To investigate the effect of azimuth angle of incidence, the hypothesis states that masking thresholds obtained for azimuth angles of incidence of 0 degrees, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90degrees are different. The repeated measures in this case are nine levels: 0 degrees, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees. The average epsilon is <0.75, indicating that the sphericity is violated (see details in chapter 3). Therefore the results of the Greenhouse-Geisser corrections (Univariate test) are used. The univariate results and specifics of the dependent variables (various conditions of speech type, delays, and masker's levels) were measured. The tabled results are provided below.

Table 6.5: Threshold level comparing azimuth angles of incidence of 0 degrees, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees under conditions of head-fixed and head rotation testing by using univariate ANOVA

Degree	0 Mean ±SD	2 Mean ±SD	4 Mean ±SD	6 Mean ±SD	8 Mean ±SD	10 Mean ±SD	15 Mean ±SD	30 Mean ±SD	90 Mean ±SD	F	Sig.
Head-fixed	52.3 ±0.8	52.0 ±2.0	51.3 ±1.8	50.0 ±2.4	48.3 ±2.0	46.0 ±1.7	44.3 ±1.7	41.3 ±1.8	37.8 ±1.4	117.72	.000**
Head-moving	52.0 ±1.0	50.5 ±2.3	46.8 ±1.2	42.6 ±2.8	40.5 ±3.0	39.0 ±2.1	38.5 ±1.7	36.8 ±1.1	36.1 ±1.2	134.51	.000**

* p<0.05
** p<0.01

The univariate ANOVAs indicated that the statistically significant difference observed among azimuth angles of incidence of 0 degrees, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees are different at a significant level $p<0.01$ for all conditions. Since the threshold is significantly affected by azimuth angle of incidence, what needs to be determined is which pair of azimuth angles introduce the significantly difference in masking thresholds. Therefore each pair of azimuth angles was tested using paired sample t-Tests. The results of the pair t-Tests have been summarized in the table below.

Table 6.6: Threshold level comparing azimuth angle of incidence of 0 degrees, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees using pair t-Tests during head-fixed testing by using paired sample t-Tests

	0 degree	2 degree	4 degree	6 degree	8 degree	10 degree	15 degree	30 degree	90 degree
0 degree		0.429	2.499*	3.250**	6.000**	11.180**	12.829**	19.900**	29.000**
2 degree			1.964	2.228	6.708**	9.000**	9.858**	12.836**	19.000**
4 degree				1.627	4.811**	7.584**	9.635**	15.492**	20.250**
6 degree					2.090	4.311**	6.273**	11.389**	21.000**
8 degree						3.250**	7.236**	9.635**	16.837**
10 degree							3.280**	6.042**	15.461**
15 degree								6.000**	15.922**
30 degree									8.573**
90 degree									

* p<0.05 ** p<0.01

Table 6.7: Threshold level comparing azimuth angle of incidence of 0 degrees, 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees using pair t-Test during head-moving testing by using paired sample t-Tests

	0 degree	2 degree	4 degree	6 degree	8 degree	10 degree	15 degree	30 degree	90 degree
0 degree		1.616	9.000**	9.322**	10.173**	15.451**	17.676**	27.041**	24.665**
2 degree			6.708**	15.077**	15.492**	13.554**	36.000**	25.521**	20.502**
4 degree				7.032**	8.135**	11.063**	21.604**	20.953**	17.072**
6 degree					4.081**	3.922**	9.154**	6.978**	6.340**
8 degree						1.913	4.000**	3.779**	4.296**
10 degree							0.885	2.725**	4.232**
15 degree								3.416**	3.827**
30 degree									1.742
90 degree									

* p<0.05 ** p<0.01

Following the univariate ANOVAs and pair sample t-Tests, the results can also be observed in figure 6.2.

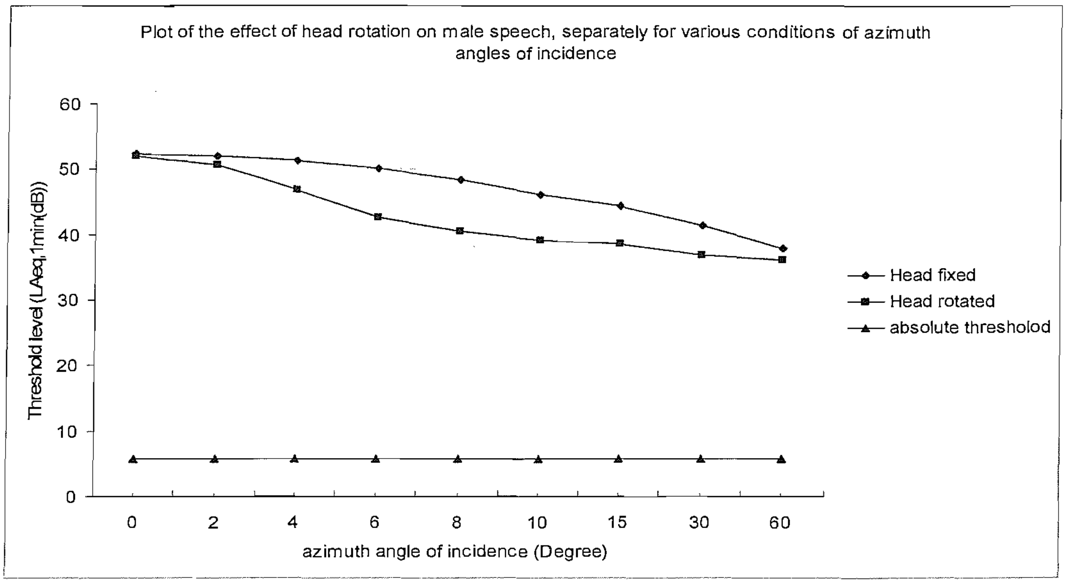


Figure 6.2: Threshold for male speech with various incidence angle of masker.

The finding showed that the azimuth angle of incidence observed affects the masking threshold. The thresholds have been found to be poorest when the masker arrived at 0degrees or at the same angle as the signal. When the signal and masker are separated by increasing the incidence angle to 2 degrees, 4 degrees, 6 degrees, 8 degrees, 10 degrees, 15 degrees, 30 degrees, and 90 degrees, the threshold was reduced. Although these reductions in threshold levels could be observed for both thresholds under head-fixed and head rotation testing, their reduction patterns were slightly different as shown in figure 6.2. For head-fixed testing, the masking threshold is gradually decreasing toward 90 degrees. On the other hand, for head-moving testing, the threshold is rapidly decreasing, especially when the masker arrived from between 4 degrees and 8 degrees. The statistically significant difference is also observed between pairs of azimuth angles of incidence of 2 degrees and 4 degrees, 4 degrees and 6 degrees, and 6 degrees and 8 degrees at a significant level of $p<0.01$. This suggests that head rotation might cause some change in binaural cues and lead to a lower threshold level when the masker arrived from some azimuth angles of incidence.

CHAPTER 7
DISCUSSION

7.1 Experimental discussion

The main objective of this study was to contribute to understanding the phenomenon of speech being masked by delayed reflections of the same-speech signal in a room which is not completely anechoic. Three experiments were conducted to measure the masking threshold of speech signals in the presence of simulated delayed same-speech maskers.

7.1.1 First Experiment

In the first experiment, the following were all measured:

- the effects of the type of programme (male speech vs. female speech)
- the time difference between the signal and the delayed speech masker
- the direction of the masker relative to the signal (the azimuth angle)
- the level difference between the speech signal and the delayed same-speech masker

7.1.1.1 Type of the programme (male speech and female speech)

The results from the first experiment did not show any statistically significant differences between the male and female voices when the data was averaged across all the other test conditions. However, more detailed statistical analysis suggested significant differences between male and female speech under test conditions where the speech and delayed speech masker both arrive from the same azimuth angle. Where the speech and delayed speech masker arrive from different azimuth angles, there were no statistically significant differences (in most cases). This finding does not appear to have been confirmed in any previous research. The most likely explanation is that the differences were due to the small changes in average frequency spectra between typical male and female speech. Zwicker & Henning (1985) describe some of the differences between typical male and female speech. Rabiner and Juang (1993) reported that male speech tends to have more energy at frequencies below 250Hz with fundamental voice pitches from around 80Hz up to 240Hz. Rabiner and Juang reported that female speech tends to have more energy at slightly higher frequencies than male speech, with significant energy up to 5,000Hz and above, and with voice pitches from around 140Hz up to 240Hz. It see ms likely that the results of the first experiment were affected by differences in voice pitch between the overall frequency spectra of male and female speech, as shown below.

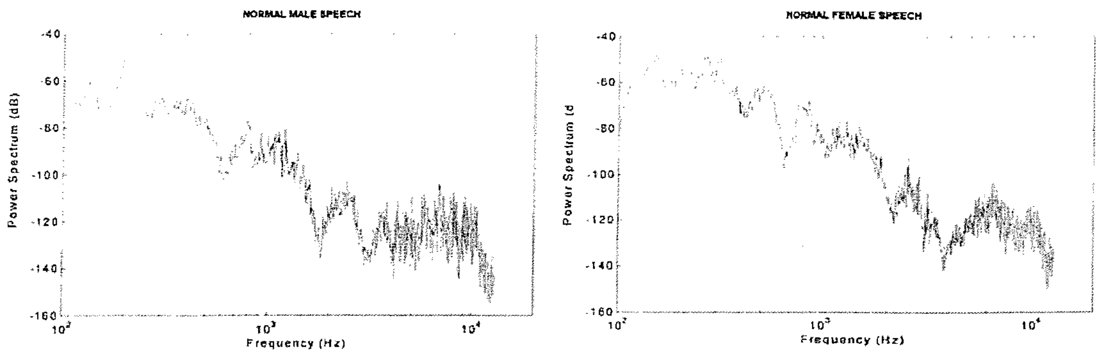


Figure 7.1: Overall frequency spectra of male speech and female speech in this study.

According to the measured overall frequency spectra of male and female speech, the female speech has more energy at higher frequencies than male speech. The voice pitches for female speech is generally between 150Hz to 320Hz, whereas the voice pitches for male speech hovers around 200Hz. That means there are some differences in average frequency spectra between male and female speech that are used in this study. Therefore, it may be argued that the significant difference between male and female masking thresholds is that the first experiment was affected by the frequency spectra of the speech. This argument is found to be consistent with the works of Zwicher and Henning (1984), Gilkey and Good (1995), and Van de par and Kohlrausch (1999), as they stated that the masking threshold depended strongly on the frequency of the signal. However the overall frequency spectrum of speech may not be the only factor that affects the masking threshold. The significant difference between male speech and female speech in this experiment could also have been due to several other factors such as the tendency for the speaker to leave slightly longer silent gaps between utterances, thereby changing the peak to the long term r ms (root mean square) ratio of the speech.

7.1.1.2 Time interval between signal and masker (15 ms, 30 ms, 50 ms, and 100 ms)

The second variable investigated in the first experiment was the time difference between the signal and the masker (15 ms, 30 ms, 50 ms, and 100 ms). This variable had a significant effect on observed masked thresholds. In an experiment on forward and backward masking, Wilson and Carhart (1971) showed that the masked threshold depended on the time gap between the signal and the masker. They reported that the masked threshold dropped when the time interval between the signal and masker was increased. The results of the first experiment were consistent with Wilson and Carhart's findings: i.e. The masked threshold was highest at the shortest delay of 15 ms, reducing with increasing delay up to 100 ms.

This finding is consistence with the well know "precedence effect", as described in chapter 2 (Haas, 1972). According to the precedence effect, the speech signal and the delayed speech masker would be heard as one sound when the delay is short. That means a masking effect caused by short delay provides good sound quality. In the other words, the minimum delay from speech signal to the same-speech masker is always preferred; not in order to avoid masking effects caused by the delayed speech masker, but in order to get the precedence effect or masking effect. Without the masking effect, the delayed speech masker could be heard separately from the speech signal, as an echo. Not only can the unwanted echo be heard, but also the perceptual fusion between the speech sounds might be observed, as both the speech signal and the delayed same-speech masker were audible simultaneously in the auditory stream (Bregman, 1994). This perceptual fusion may degrade the sound quality and the speech intelligibility. The effect perceptual fusion will be further discussed in general discussion.

7.1.1.3 Masker's azimuth angle of incidence (0 degrees, 30 degrees, 60 degrees, and 90 degrees)

The direction of a masker relative to the direct sound was the third variable investigated as the azimuth angle of incidence (0 degrees, 30 degrees, 60 degrees, and 90 degrees). The findings showed that the masking threshold was found to be poorest when the delayed speech masker arrived at 0 degrees, or the same azimuth angle of incidence, as the speech signal. When the speech signal and the delayed speech masker were separated by increasing the incidence azimuth angle of the masker to 30 degrees, 60 degrees, and 90 degrees, the masking thresholds were reduced. These results were consistent with the work reported by Seberi et al. (1991) and Gilkey and Good (1995). Seberi et al. (1991) reported that the threshold was found to be highest when the signal and masker are spatially coincident, and a 9 dB reduction in masking was observed; whereas, Gilkey and Good (1995) reported the same finding, but with a reduction in masking for high frequencies as much as 18 dB. However in this study, the reduction of the mean masking threshold was found to be between 1 dBA (male, 100 ms, 40 dB) and 17 dBA (male, 30 ms, 80 dB). This could have been due to the materials used and the testing conditions in the experiment, because both previous experiments did not use running speech; in these tests, artificial vowels, consonants, and noise were used.

7.1.1.4 Level (40 dBA, 60 dBA, and 80 dBA)

The last variable investigated in the first experiment was the level between the speech signal and the delayed speech masker (40 dBA, 60 dBA, and 80 dBA). The findings indicated that the level between the speech signal and the delayed speech masker affected the masking threshold. This finding was consistent with the previous study (Zwicker and Henning, 1984) which investigated the masking threshold in the presence of a masker. The results in this experiment indicated that the mean masking threshold was increased by approximately 6 dBA as the masker level increased from 40 dBA to 60 dBA, and the mean masking threshold was further increased by approximately 10 dBA when the masker level increased from 60 dBA to 80 dBA. It showed that increases in the masking threshold are non-linear. For example, in the case of male speech with a masker arriving at 0 degrees azimuth angle with a time delay of 15 ms, the signal to noise ratio was -5 dB when the masker level was 40 dB. The signal to noise ratio was reduced to -12 dB and -14.5 dB when the masker levels were 60 dB and 80 dB (see figure 4.11). This would suggest that, although the threshold was increased according to the increase in the masker level, the ability to detect the speech signal became better with a higher masker level. Indeed, the increase in ability to detect the signal was also non-linear. This non-linearity found was consistent with the data obtained by Zwicker and Henning (1984). One possible factor that might lead to this non-linearity was that when masker level was 40 dBA, the masking threshold was high because of the background noise in the room, together with the internal noise generated inside the auditory system. The effects of this internal noise were only found when the masker level was low (Moore, 1997).

7.1.2 Second Experiment

The second experiment was conducted in order to closer investigate how a speech signal is detected with a delayed speech masker. The experiment was designed in two parts. The first part was to investigate the effect of a silent interval of a speech masker, while the second part was to investigate the upward spread masking effect (simultaneous masking).

7.1.2.1 Part one

The results from the first part of this second experiment indicated that the silent gap in speech does not affect the masking threshold. In other words, the masking threshold of speech with and without a gap was found to be the same. One possible explanation would be that the speech signal could not only be detected in the silent gap of a delayed speech masker but also in the low amplitude consonants in the delayed speech masker. This was consistent with the work of Spiegel (1987).
The type of programme material used – male and female speech – was shown to have no affect on masking threshold in this experiment under headphone listening conditions. This result was also consistent with the previous experiment in that there was no significance difference found between the threshold of male speech and the threshold of female speech.

The time difference between the signal and the masker shown affected the masking threshold in this experiment, which was consistent with the first experiment. These findings suggested that the speech signal was detected at the beginning of each syllable of the delayed speech masker. The speech signal could have been detected in the silent gap of the delayed speech masker (Wilson and Carhart, 1971; Akeroyd, M.A. and Summerfield, 1999) or detected in the lower amplitude consonants of the delayed speech masker (Spiegel, 1987).

However the findings in the first part of this experiment do not preclude the possibility that vowel signals could have been detected during masker periods represented only by lower amplitude consonants sounds. When the results in this part of the experiment was observed closely, it is apparent that the masked threshold level of speech with and without silent gaps with a masker time delay of 50 ms were slightly different. This finding could be related to the typical duration of actual speech component sounds in relation to simultaneous masking and the critical band concept. Also, when editing the speech stimuli to remove the silent gaps, it was noted that some silent gaps were longer in duration than some of the low amplitude consonants present. That means the masking threshold of speech that contained long silent gaps was lower than that of speech that contained short-duration low amplitude consonants. To the extent that certain low amplitude consonants might be considered as equivalent silent gaps for masking purposes, the results could be explained by simultaneous masking phenomena associated with the shorter duration low amplitude consonant gaps, which were still present even when the silent gap had been removed.

7.1.2.2 Part two

In the second part of experiment two, the results clearly showed that there was some difference in the masking thresholds when:

- the speech signal and the delayed speech masker were at the same frequency range
- the speech signal and the delayed speech masker were at a different frequency range

When the speech signal and the delayed speech masker were in the same frequency range, the masking thresholds in all cases were higher than those when the speech signal and the delayed speech masker were in different frequency ranges. This finding could have been correspondent to the phenomenon of simultaneous masking effect (critical band concept) (Moore, 1982).

When the speech signal and the delayed speech masker were in the same frequency range, the masking thresholds were similar to those measured for normal speech. The actual threshold levels appeared to depend on the masker delay in a way that was consistent with known non-simultaneous masking effects (Moore, 1982), with the thresholds decreasing as the masker delay was increased toward 50 ms. Even though the thresholds obtained in this group have shown the same masking released pattern, their threshold levels were slightly different from one another. The thresholds of the high frequency range were found to be lower than that of the low frequency range and the mid frequencies range. These differences may be a result of differences in relative gap duration and masker duration when the speech was filtered into different frequency bands. It appeared that the masker duration affected the threshold level mainly because of the non-simultaneous masking effect (see Zwicker & Fastl, 1990).

When the speech signal and the delayed speech masker were in a different frequency range, the differences between the groups of masking thresholds were observed. It can be seen that, when the delayed speech maskers were in a different frequency range from the speech signal, the masking thresholds were not affected by delay. This suggested that only the simultaneous masking effect could have been occurring under these circumstances. There were some indications of a small amount of upwards spread masking effect when the frequency range of delayed speech maskers was adjacent to the frequency range of the speech signal. For example, when the speech signal was in the mid frequency range and the delayed speech masker was in the low frequency range (SM-ML), the masking threshold was higher than the masking threshold of any other conditions when the speech signal and the delayed speech masker were in a different frequency range (see Figure 5.10 and Figure 5.11).

When the speech signal was in the high frequency range and the delayed speech masker was in the low frequency range (SH-ML), and when the speech signal was in the low frequency range and the delayed speech masker was in the high frequency range (SL-MH), the masking thresholds were found to be only slightly higher than that of the absolute threshold. These finding were counter to established wisdom that low frequencies, such as vowels, may not always mask high frequencies, like consonants, in the context of overlap masking within the range of parameters tested.

7.1.3 Third experiment

Following the first experiment, the results showed that the masked threshold of speech when the speech signal and the delayed speech masker were presented from the same azimuth angle was higher than that of speech when the speech signal was presented from the front and the delayed speech masker from a 30, 60, and 90 degree azimuth angle. The findings suggested that the results of the threshold level could be separated into two groups. One was worst when both the signal and masker were presented from 0 degrees, and the other was the threshold level that was released from masking when the signal and masker are presented separately. Therefore, the third experiment was designed to investigate the transformation between two groups of the masking threshold levels affected by azimuth angle of incidence. Furthermore, in the first experiment, the subjects were allowed to move their head in a natural manner, which might have affected the threshold level as well. Consequently the third experiment studied the effect of head rotation on the masking effect with respect to azimuth angle of incidence.

The findings showed that the head rotation affected the masking threshold. When the head rotated, the masking thresholds were lower than they would have been without a head rotation. This indicated that the head rotation provided an extra cue in the release from masking. One possible explanation would be that, during the head rotation, the binaural cues, including inter-aural time difference and inter-aural intensity difference, were changed. This changing of binaural cues during the head rotation has also been reported by Thurlow, Mangle, and Runge (1967), Brokhorst and Plomp (1988), and Wightman et al (1994). This change could have an affect on the masking threshold due to the condition of "Binaural Masking Level Difference (BMLD)" (see chapter 2).

For the azimuth angle of incidence, the masking threshold was found to be poorest when the masker arrived at 0 degrees, or the same angle as the signal, which was consistet with the first experiment. When the signal and masker are separated by increasing the incidence angle of the masker to 2, 4, 6, 8, 10, 15, 30 and 90 degrees, the masking thresholds were reduced. These results are also found to be consistent with the first experiment and the work conducted by Seberi et al. (1991), and Gilkey and Good (1995). In this study, the reduction in mean masking thresholds was found to be about 14 dBA (head fixed) and 16 dBA (head rotated).

When the azimuth angle of incidence was investigated further, the findings showed that the masking threshold was found to be poorest when the delayed speech masker arrived from 0 degrees, for both head fixed and head rotating tests. In other words, there was no binaural masking released when the speech signal and the delayed speech masker were present from the same azimuth angle. This had also been observed by Saberi et al (1991). However when the azimuth angle of incidence was increased toward 90 degrees, the masking threshold decreased. For the head fixed test, the masking threshold gradually decreased. For the head rotating test, the masking threshold was rapidly decreased, especially when masker arrived from 4 to 10 degrees. This suggested that head rotation caused some change in binaural cues and led to a lower threshold level at some masker azimuth angles of incidence. The masking thresholds when the azimuth angle was 2 degrees for both head fixed and head rotated tests were observed to be as high as the poorest masking threshold when the azimuth angle was 0 degrees. Tests with head rotation did not provide any more masking release than did head fixed tests, nor was there any statistically significant difference found for either condition. One possible explanation might be that, with an azimuth angle of 2 degrees or less, the two auditory events could not be separated, which was considered as a sound localization blur angle (Blauert, 1997). The findings also showed that when the azimuth angle was increased toward 90 degrees, head movement provided a significant reduction in the masking threshold. This suggested that head rotation caused the masking threshold to decrease when the incidence angle of the speech signal and the delayed speech masker were presented spatially apart by between 4 and 90 degrees.

In summary, the findings from all three experiments indicate that the speech signal masked by a delayed same-speech masker was more complicated than the speech signal masked by a broadband noise masker. With a delayed same-speech masker, the speech signal could be detected in both the silent gap of the speech masker and the low amplitude consonants of the speech masker (see experiment 2). To detect the speech signal in the low amplitude consonants part of the delayed speech masker, the results from this study contradicted the established wisdom that low frequencies as vowels may not always mask high frequencies as consonants, in the context of overlap masking within these experiments (see part 2 of the experiment 2). The findings also suggested that the masking threshold of a speech signal was affected by non-simultaneous masking when the speech signal was presented with a delayed same-speech masker. Based on the non-simultaneous masking effect, that means that the time delay between the speech signal and the delayed speech masker was one of the most important variables. The findings indicated that the masking threshold was reduced when the delay time between the speech signal and the delayed speech masker was increased. These results were also found to be true for both male speech and female speech, as there was no significant difference between the masking thresholds of male speech and the masking thresholds of female speech, which were observed in both experiment 1 and 2. Regarding the level of the delayed speech masker, the masking thresholds were found to increase as the level of delayed speech masker increased (see experiment 1). When the spatial condition was taken into consideration, the findings showed that the azimuth angle of incidence affected the masking thresholds (see experiment 1 and experiment 3). The masking threshold has been found to be poorest when the delayed speech masker arrived at 0 and 2 degrees. When the speech signal and the delayed speech masker were separated by increasing the incidence angle of the delayed speech masker toward 90 degrees, the masking thresholds were reduced. This effect can be most clearly observed when subjects rotated their head.

7.2 General Discussion

7.2.1 Summary of the experiments

This study started with the broad problem that reverberation time provides a relatively crude indication of the amount of delayed reflections, but it does not always discriminate very well between rooms of very different acoustical quality. Many researchers have attempted to devise superior indicators of acoustical quality, but none of these has yet been fully successful. This study addressed the problem by studying the effect of delayed same-speech masking under controlled laboratory test conditions. The findings suggest that the initial time difference between the speech signal and the reflection, level of reflections with respect to the speech signal, and the direction of reflections as compared to the direct sound are significant. The roles of their relationships are shown as a diagram in figure 7.2.

In figure 7.2, it shows the diagram of the masking effect when a speech signals and a delayed same-speech masker are presented simultaneously. The main variables that strongly affect the masking thresholds are the initial time difference between the speech signal and the delayed same-speech masker, and the direction of the delayed speech masker as compared to the direct speech sound. Two conditions of the initial time difference between the speech signal and the delayed same-speech masker were selected according to the precedence effect (see chapter 2). They are:

- shorter than 20 ms
- longer than 20 ms

The two conditions of azimuth angle of the delayed speech masker were selected according to the masking release angle when the listener moves their head naturally (see experiment 3). They are:

- 2 degrees or less
- more than 2 degrees

Consequently, there are total of 4 conditions under the interactions between the initial time difference between the speech signal and the delayed same-speech masker and the direction of the delayed speech masker as compared to the direct speech sound.

The diagram also shows the possible effects occurred when speech signal is masked and when it is unmasked. When the speech signal is unmasked, the auditory scene analysis is introduced in this diagram. It can be seen clearly from the experiment 2 that vowels do not always mask the consonants. That means it is possible that two auditory events may always occur simultaneously. In order to perceive one or the other, the auditory stream segregation may play the important role. The stream segregation is the method by which our brains group together or "fuse" multiple sensations of acoustic stimuli because they are understood to have come from a common source. If any wanted stimulus is segregated, that stimulus will be perceived. However, if any two stimuli cannot be segregated, it may cause perceptual fusion. This perceptual fusion caused by two stimuli occurred simultaneously may reduce the speech intelligibility. Assmenn (1999) shows that the percentage of the words correct can be as low as 50 when two speeches were presented simultaneously. Consequently, the perceptual fusion may lead to the bad sound quality. Please note that the sound quality based on masking effect in this study refers to the speech clarity and the speech intelligibility in the present of the reflection.

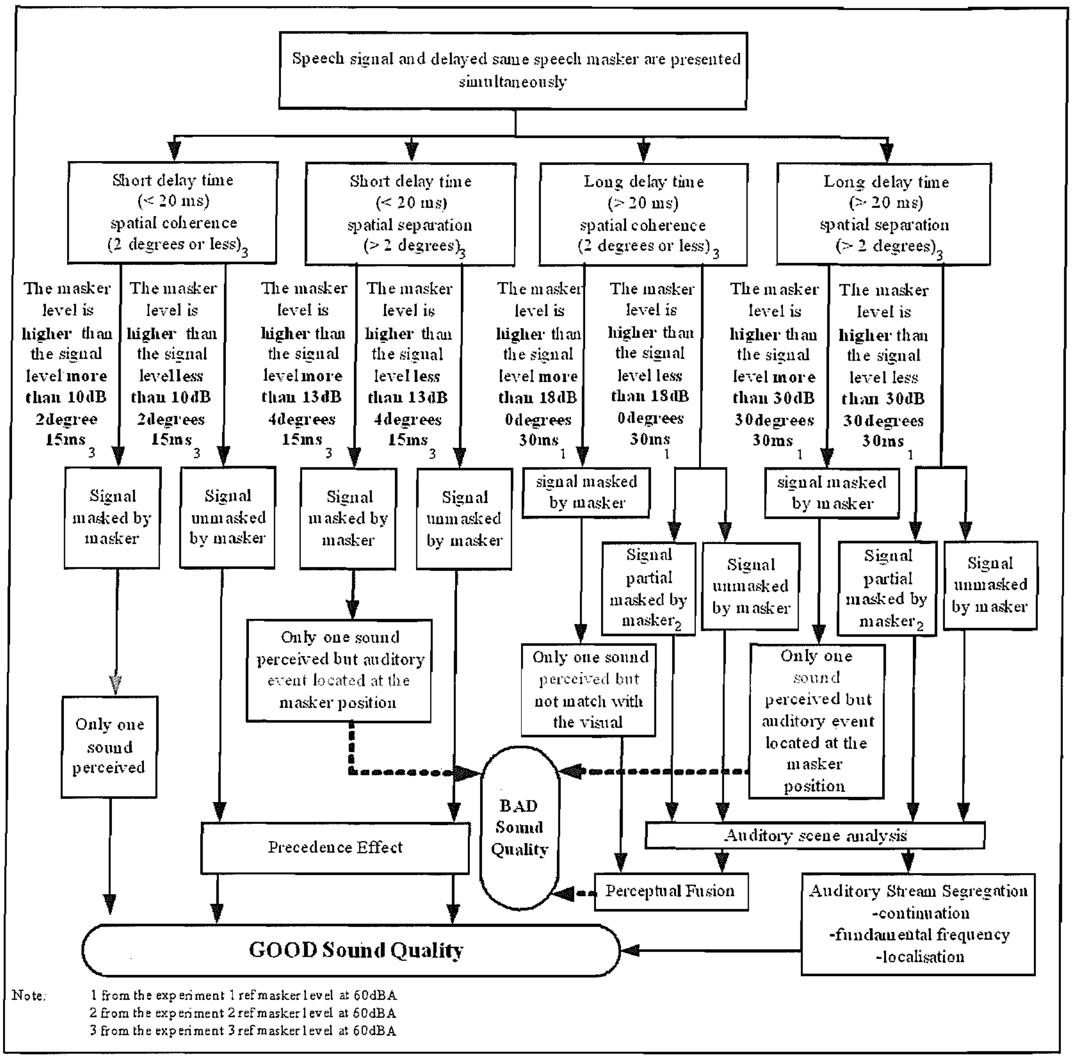


Figure 7.2: Diagram of the masking effect when a speech signal and a delayed same-speech masker are presented simultaneously.

From the above diagram, the first condition had a short delay time (less than 20 ms) and a spatial coherence (2 degrees or less). The speech signal is masked by the delayed speech masker when the level of the delayed speech masker is higher than the speech signal by more than 10 dBA (with a 2 degree azimuth angle of incidence and a 5 ms delay), according to the results of experiment 3. Under these conditions, this creates the impression that there is only one speech source. A clear speech signal is perceived, which can be identified as good sound quality. On the other hand, if the perceived, which can be identified as good sound quality. On the other hand, if the level of the delayed speech masker is higher than the speech signal by less than 10 dBA (with a 2 degree azimuth angle of incidence and a 15 ms delay), according to the results of the experiment 3, the delayed speech masker does not mask the speech signal. This creates the impression that the speech signal that arrived first is the source of the sound. Its sound pressure level will be enhanced by the delayed speech masker because of the precedence effect. Clear, loud speech is perceived, which is also identified as good sound quality.

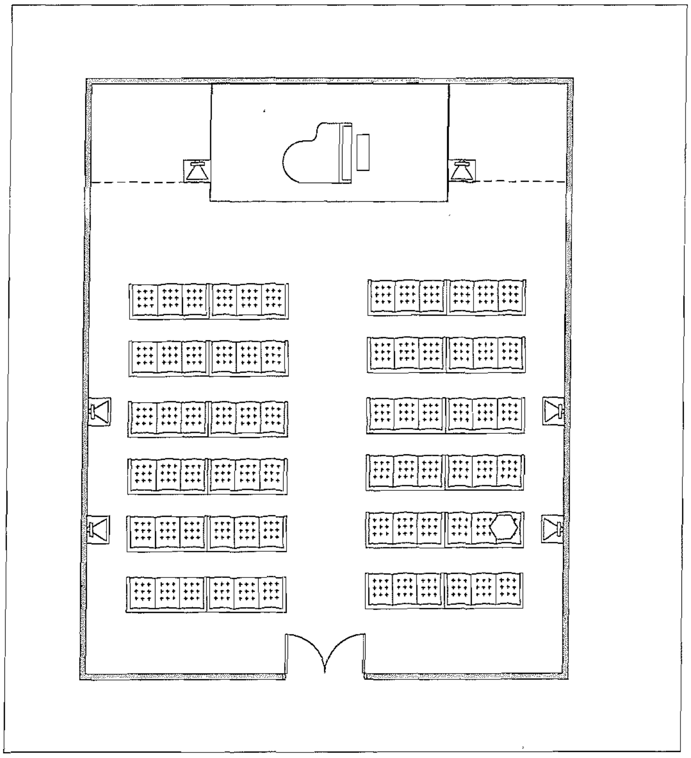


Figure 7.3: Sound reinforcement system with frontal loudspeakers and column loudspeakers.

The second condition was under a short delay time (less than 20 ms) and a spatial separation (more than 2 degrees). The speech signal is masked by the delayed speech masker when the level of the delayed speech masker is higher than the speech signal by more than 13 dBA (with a 4 degree azimuth angle of incidence, and a 15 ms delay), according to the results of the third experiment. Under these conditions, this creates the impression that there is only one speech signal which arrived from the delayed speech masker's azimuth angle of incidence. Clear speech is perceived, but the image of the speech source will be shifted to the incidence azimuth angle of the delayed speech masker. This can be seen as bad sound quality. This effect may not be found in a real room, but it can be found in the reinforcement system. In figure 7.2, it shows a reinforcement system that is designed by using a pair of frontal loudspeakers together with two pairs of column loudspeakers on the side wall. The purpose of this design is to provide equal volume coverage around the room. In general, a short delay time will be applied to the column loudspeaker in order to avoid echo. Although a clear sound is perceived from the loudspeakers in this design, unfortunately the perceived sound source will be shifted to the position of the column loudspeakers when the audience is sitting close to the column loudspeaker, for example at position number 1 in figure 7.2.

Once again, when the second condition is under a short delay time (less than 20 ms) and is spatially separated (more than 2 degrees), and the level of the delayed speech masker is higher than the speech signal by less than 13 dBA (with a 4 degree azimuth angle of incidence and a 15 ms delay), according to the results of the third experiment, the delayed speech masker does not mask the speech signal. Under these conditions, a listener may hear two identical speech sounds from two directions. The speech signal created at the front arrives first, followed by the delayed speech masker. Therefore, to the listener, this creates the impression that the speech signal arriving first is the source of the sound because of the law of first wave front. This means that the correct perception of the sound source is provided, with good sound quality being assumed under these conditions, even though partial masking or unmasking is present.

The third condition had a long delay time (more than 20 ms) and spatial coherence (2 degrees or less); the speech signal was masked by the delayed speech masker when the level of the delayed speech masker was higher than the speech signal by more than 18 dBA (with a 0 degree azimuth angle of incidence and a 30 ms delay) according to the results of experiment 1. These conditions create the impression that there is only one speech signal. Under these conditions, clear speech is perceived. Unfortunately, if the delay time is too long, the listener might hear sound that is not matched with the lips of the speaker. This may cause perceptual fusion caused by vision (McGurk and McDonald, 1976), leading to bad sound quality. Some sounds may be confused throughout the sentence; for example a visual /ga/ combined with an aural /ba/ is often heard as /da/ (Bergman, 1990).

Under these conditions, the speech signal might be partially masked or unmasked by the delayed speech masker when the level of the delayed speech masker is higher than the speech signal by less than 18 dBA (with a 0 degree azimuth angle of incidence and a 30 ms delay), according to the results of experiment 1. The partial masking means that only some phonemes are masked, while some remain unmasked. This may occur due to the upward spread masking effect (see experiment 2 part 2), especially when the frequency of the signal and the frequency of the masker are adjacent. Under these conditions, the speech signal and the delayed speech masker will be audible simultaneously. This is also true when the speech signal is unmasked by the delayed speech masker. Under these circumstances, the sound quality may depend on auditory scene analysis. The sound quality might be bad if there is some perceptual fusion, meaning that the speech may be degraded. On the other hand, the sound quality might be good if the speech signal is segregated from the delayed speech masker and the speech signal is perceived to be clear.

The last condition was under a long delay time (more than 20 ms) and spatial separation (more than 2 degrees). The speech signal is masked by the delayed speech masker when the level of the delayed speech masker is higher than the speech signal by more than 30 dBA (with a 15 degree azimuth angle of incidence with a 30 ms delay), according to the results of experiment 1. These conditions create the impression that there is only one speech signal arriving from the delayed speech masker's azimuth angle of incidence. That means that clear speech is perceived, but the image of the sound source will be shifted to the incidence azimuth angle of the delayed speech masker, similar to condition two with a short delay and a spatial separation. This can be assumed as bad sound quality. This effect is unlike when the delay is short. With a long delay, this effect may by found in some large auditoriums. Generally, in a large space, there will be a position called "critical distance". The critical distance is the position at which the sound pressure level of the signal and the sound pressure level of the reflections are equal. In other words, it is a point in space where the echo from a signal is at the same volume as the source. If a listener is further away from the sound source than the critical distance and is close to the reflection surface, he may perceive only one sound, as the signal is masked, but an image of the sound source might be shifted to the direction of the reflection. It can be assumed that this is bad sound quality.

Finally, in the last condition, the speech signal might be partially masked or unmasked by the delayed speech masker when the level of the delayed speech masker is higher than the speech signal by less than 30 dBA (with a 30 degree azimuth angle of incidence and a 30 ms delay). Under these conditions, the speech signal and the delayed speech masker will be audible simultaneously. Under these circumstances, the sound quality may depend on auditory scene analysis. The sound quality might be bad if there are some perceptual fusions, which means that the speech might be degraded. On the other hand, the sound quality might be good if the speech signal is segregated from the delayed speech masker and the speech signal is perceived as clear.

It can be seen that under conditions where time delay is longer than 20 ms, whether azimuth angles are 2 degrees or less, or more than 2 degrees, the perceptual fusion might be found when the signal is not masked by the delayed masker. So far, it may be agreed that a long delay will provide bad sound quality; nevertheless, it is too soon to jump to conclusions. The perceptual fusion might be not be found under some conditions. This is supported by Bregman (1990), who reported on the process which allows man to perceive the speech signal when two or more sounds are audible simultaneously. It is known as "auditory stream segregation".

Auditory stream segregation is a perceptual process in which relationships are formed between different sensations, and the effect of relationships on what is included and excluded from our perceptual descriptions of distinct auditory events. It is the method by which our brains group together or separate multiple sensations of acoustic stimuli. According to Bregman (1990), the auditory system performs sound separation by employing various cues for grouping and separating, using continuation, fundamental frequency, and localisation methods.

Following the concept of auditory stream segregation, if two or more sounds are presented simultaneously and cannot be segregated, we can assume that the sound quality will be bad because this perceptual fusion may reduce the speech intelligibility. Assmenn (1999) shows that the percentage of the words correct can be as low as 50 when two speeches were presented simultaneously. And the percentage of the words correct was increased when a clue (fundamental frequency) was added in the stimulus. Consequently, the perceptual fusion may lead to the bad sound quality. On the other hand, if speech signals are presented together with a delayed speech masker, but the speech signal can be segregated, we can assume that the sound quality will be good because the speech signal is perceived clearly. Please note that the sound quality based on masking effect in this study refers to the speech clarity and the speech intelligibility in the present of the reflection.

From the previous discussion, it can be seen that, with a short delay time, the masking effect is not an issue. Even under conditions wherein the speech signal is masked by a delayed speech masker arriving from an azimuth angle of more than 2 degrees with respect to the signal, it creates the impression of a direct sound. Luckily, this effect may not be found in a real room. Not only is the masking effect a matter of sound quality, but the azimuth angle of incidence (whether 2 degrees or less, or more than 2 degrees) does not significantly affect the sound quality. Under these short delay conditions, the sound quality is always assumed to be good. On the other hand, with a long delay time, the masking effect is always unwanted. If the speech signal is fully masked when the speech signal and the delayed speech masker arrive from the same direction, a listener might perceive a sound that does not match with the lips of the speaker. This might cause perceptual fusion because of the influence of vision on sound, which is known as the "McGurk effect" (McGurk and Mcdonald, 1976). If the speech signal is fully masked when the speech signal and the delayed speech masker arrive from differene directions, a perceived sound image might be shifted to the direction of the masker. Under both conditions mentioned, we can assume the sound quality to be bad. Nevertheless, when the speech signal is not fully masked, the sound quality may not always be good as, without the masking effect, the speech signal and the delayed speech masker might be audible simultaneously and might cause some perceptual fusion. Therefore, under conditions where the time delay is longer than 20 ms, whether the azimuth angle is 2 degrees or less, or more than 2 degrees, the sound quality can always be assumed to be bad when the speech signal is fully masked by the delayed speech masker; and when the speech signal is not masked by the delayed speech masker, the sound quality may be good or bad. This depends on the ability to segregate the signal from the masker. If the speech signal can be segregated from the delayed speech masker, the sound quality can be assumed to be good; but if the speech signal cannot be segregated from the delayed speech masker, the sound quality can be assumed to be bad because of some perceptual fusion. Consequently, the masking effect may be significantly affected by delayed time, but may not be significantly affected by the azimuth angle of incidence.

In conclusion, the findings in this study suggest that the initial time between direct sound and reflection, and the direction of reflection as compared to the direct sound can be significant. The speech quality can be assumed to be good with short delay whether the direct speech signal is masked or not. Unfortunately, under the condition studied the delayed same speech masker from different direction was always louder than the direct speech signal. It can be seen that the condition in which the reflection is louder than the direct sound has almost never been the case in real auditoria; but it can be found when sound reinforcement system is used. Interestingly, when the sound reinforcement system with multi-loudspeakers are used, the short delay may not always provide the good sound quality. The location of each loudspeaker installed can affect the sound quality. Using column loudspeakers, the audience who sits next to it can have the auditory image shifted to the location of the loudspeaker instead of the location of the talker in the front. This can be assumed that the sound quality is bad. For further development of objective parameter determining speech quality based on masking effect in a room, the multi-reflections is required for further study. The sound quality may be affected by a pattern of reflections (the delay) and the directions of the reflections. With determination of direction, some existing objective parameters such as Lateral Fraction (LF) and Late Lateral Level (GLL) might be importance. However, from this study, the quality of speech may depend on auditory scene analysis when the initial time between direct sound and reflection is long. That means there are many factors, such as fundamental frequency, continuation, and localisation, involved. Consequently, to develop a precise indicator of speech quality, a method of measuring speech quality might require a computer-based system, which can be complicated. It may not depend on the geometry of a room and its finishes alone. In addition, it also may not be able to be used for a room design.

CHAPTER 8
CONCLUSION AND SUGGESTIONS

8.1 General conclusion

This study has looked at the problem of reverberation time providing a relatively crude indication of the amount of delayed reflections. It does not always discriminate very well between rooms of quite different acoustical quality. Even though many researchers have attempted to devise superior indicator of acoustical quality, none of them has yet been fully successful. Therefore, this study has addressed the problem by studying the effects of delayed same-signal masking under controlled laboratory test conditions. The main objective of this study was to contribute to understanding the phenomenon of speech being masked by delayed reflections of the same speech sounds in any room which is not completely anechoic. Three experiments were conducted to measure the masking threshold of speech signals in the presence of simulated delayed same-speech maskers.

After three experiments, the findings suggest that there is no statistical significant difference observed between the thresholds of male speech and the thresholds of female speech. On the other hand, three factors – the initial time difference between the speech signal and the reflection, the level of reflections with respect to the speech signal, and the direction of reflections as compared to the direct sound – all have a significant influence on the threshold levels of speech.

In the first experiment, the role of each variable on the masking threshold can be explained individually. When *the level of the masker* is increased, the masking threshold increases. When *the time interval between the signal and the masker* is increased, the masking threshold decreases. When the signal and masker are spatially coincident, masking thresholds have been found to be the highest. When *the azimuth angle of incidence between the signal and the masker* is increased, the masking threshold decreases.

In experiment 2, the findings suggest that the speech signal could be detected in the silent gap and in the low amplitude consonant parts of speech because there is no significant difference found between the masking threshold of speech with or without a silent gap. In this experiment (the second part) we saw that, when the signal and the masker are in the same frequency range, the masking thresholds are higher than when the signal and the masker are in different frequency ranges. Their masking thresholds were similar to those measured for normal speech and both appear to depend on the masker delay. On the other hand, when the signal and the masker are at different frequency ranges, the masking thresholds measured are not affected by delay. A small upwards spread masking effect has also been observed when the frequency range of the signal and of the masker are adjacent. Interestingly, when the signal is at a high frequency and masker is at a low frequency (SH-ML), the masking thresholds are found to be only slightly higher than that of the absolute threshold. It suggests that the low frequencies as vowels may not always mask high frequencies as consonants in the context of overlap masking within the range of the tested parameters.

In the last experiment, the findings indicate that head rotation affects the masking threshold level. This suggests that the head rotation could be a factor that affects the perception of sound in a room when a person listens to the sound naturally in space. The results show that high masking threshold levels were found when the masker arrived from 0 degrees and 2 degrees under both head fixed and head rotating conditions. For head fixed testing, when the azimuth angle is increased toward 90 degrees, the masking threshold gradually decreases. For head rotation, when the azimuth angle is increased toward 90 degrees, the masking threshold rapidly decreases, especially when the masker arrives from between 4 to 10 degrees.

From these three experiments, it can be seen that sound quality based on the masking effect in a room may depend on two main variables- the time interval between signal and reflection and the azimuth angle of incidence of the reflection.

In a small room with a short delay time, the masking effect is not an issue. If the speech signal is partially masked or unmasked by the delayed speech masker (reflection), the delayed speech masker will be audible simultaneously. Under these circumstances, a listener may hear two identical speech sounds from two directions. The speech signal created at the front arrives first, followed by the delayed speech masker. To the listener, this creates the impression that speech signal arriving first is the source of the sound because of the law of first wave front, or the precedence effect. That means the correct perception of the sound source is provided and good sound quality is also assumed. On the other hand, if the speech signal is masked by the delayed speech masker, it will create the impression of a direct sound arriving from the azimuth angle of the delayed speech masker. When the azimuth angle of incidence of the delayed speech masker is less than 2 degrees, the correct perception of the sound source is evident, as is good sound quality. When the azimuth angle of incidence of the delayed speech masker is more than 4 degrees, the perceptual location of sound source is shifted to the direction of the masker and the poor sound quality is assumed. Luckily, this effect may not found in a real room. It can be found in a reinforcement system. Thus, based on the masking effect, in a small room with short delay conditions, the quality of sound is always assumed to be good.

In a large room with a long delay time, the masking effect is always unwanted. If the speech signal is fully masked when the speech signal and the delayed speech masker arrive from the same direction, a listener might perceive a sound that does not match with the lips of the speaker. This might cause perceptual fusion because of the influence of vision on sound, which is known as "McGurk effect" (McGurk and Macdonald, 1976). If the speech signal is fully masked when the speech signal and the delayed speech masker arrive from a different direction, a perceived sound image might be shifted to the direction of the masker. Under both conditions mentioned, the sound quality may be assumed to be bad. Nevertheless, when the speech signal is not fully masked, the quality of sound can be not easily determined. It depends on the ability to differentiate the signal from the masker, as the speech signal and the delayed speech masker are simultaneously audible. If the speech signal can be segregated from the delayed speech masker, the sound quality may be assumed to be good; but if the speech signal cannot be segregated from the delayed speech masker, the sound quality may be assumed to be bad because of some perceptual fusion.

Consequently, the initial time difference between the signal and reflection significantly affects sound quality based on the masking effect. This study shows that the sound quality in any room may be assumed to be good when the delay is short, Unfortunately, under the condition studied the delayed same speech masker from different direction was always louder than the direct speech signal. It can be seen that the condition in which the reflection is louder than the direct sound has almost never been the case in real auditoria; but it can be found when sound reinforcement system is used. Interestingly, when the sound reinforcement system with multi-loudspeakers are used, the short delay may not always provide the good sound quality. The location of each loudspeaker installed can affect the sound quality. Using column loudspeakers, the audience who sits next to it can have the auditory image shifted to the location of the loudspeaker instead of the location of the talker in the front. This can be assumed that the sound quality is bad. For further development of objective parameter determining speech quality based on masking effect in a room some existing objective parameters such as Lateral Fraction (LF) and Late Lateral Level (GLL) might be importance because the direction of the reflection is significant. However, from this study, the quality of speech may depend on auditory scene analysis when the initial time between direct sound and reflection is long. That means there are many factors, such as fundamental frequency, continuation, and localisation, involved. Therefore, to develop a precise indicator of speech quality, a method of measuring speech quality might require a computer-based system, which can be complicated. It may not depend on the geometry of a room and its finishes alone.

8.2 Suggestions for further work

In this study, the findings show that the time interval between the speech signal and the delayed speech masker, the level of the delayed speech masker, and the incidence azimuth angle of the delayed speech masker affect the masking threshold of the speech signal. The findings also suggest that the sound quality based on masking effect in a room may assumed to be good when the delay is short, but remain inconclusive when the delay is long. Further work may be needed to look into the effect of multi-reflections in order to relate these results to a more natural, real world setting because this experiment was conducted using simulated methods to represent a delayed same-speech reflection through loudspeakers in an anechoic room or through headphones, which do not exist in a natural, everyday setting.

In order to achieve the goal in developing a method to measure the sound quality in a room based on the masking effect, further study on auditory scene analysis may need to be taken into consideration. In auditory scene analysis, there are many factors involved, such as fundamental frequency, continuation, and localisation. Therefore, to develop a precise indicator of speech quality, a method of measuring speech quality might require a computer-based system, which can be complicated. A computer-based system may reach its conclusion based on many factors, and not simply room geometry and finish alone.

REFERENCES

Ando, Y. 1977. ***Subjective preference in relation to objective parameters of music sound fields with a single echo.*** J. Acoust. Soc. Am. 62, 1436-1441.

Ando, Y. and Gottlob, D. 1979. ***Effects of early multiple reflections on subjective preference judgments of music sound fields.*** J. Acoust. Soc. Am. 65, 524-527.

Aoki, S. and Houtgast, T. 1992. ***A precedence effect in the perception of inter-aural cross correlation.*** Hearing Research. Vol.59, 25-30.

Akeroyd, M.A. and Summerfield, A.Q. 1999. ***A binaural analog of gap detection.*** J. Acoust.Soc.Am. 150(0), 2807-2820.

Assmann, P.F. 1999. ***Fundamental frequency and the intelligibility of competing voices.*** Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, Aug. 1-7, 1999, pp. 179-182.

Barwick, J. 1994. ***Sound Recording practice.*** Oxford University Press: Oxford.

Beranek, L.L. 1986. ***Acoustics***, The American Institute of Physics: USA.

Beranek, L.L. 1996. ***Acoustic and musical qualities***. J. Acoust.Soc.Am. 99(5), 2648.

Blauert, J. 1997. ***Spatial hearing***. The MIT press, London: England.

Boer, E. de 1966 ***Encoding of frequency information in the discharge pattern of auditory nerve fibres.*** Inter.Audiol.8, 547.

Bolt, R.H. and Macdonld, A.D. 1949. ***Theory of speech masking by reverberation.*** J.Acoust.Soc.Am. 21(6), 577-580.

Bradley, J. S. and Soulodre, G.A. 1995. ***The influence of late arriving energy on spatial impression.*** J.Acoust.Soc.Am. 97(4): 2263-2271.

Breebaart, J., van de Par, S., and Kohlrausch, A. 2001a. ***Binaural processing model based on contralateral inhibition. I. Model structure.*** J.Acoust. Soc. Am. 110, 1074–1089.

Breebaart, J., van de Par, S., and Kohlrausch, A. 2001b. ***Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters***. J. Acoust. Soc. Am. 110, 1089–1105.

Breebaart, J., van de Par, S., and Kohlrausch, A. 2001c. ***Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters.*** J. Acoust. Soc. Am. 110, 1105–1118.

Bregman, A.S. 1990. ***Auditory scene analysis: The perceptual organization of sound.*** Cambridge, MA: MIT Press.

Bronkhorst, A.W. and Plomp, R. 1988. ***The effect of head-induced interaural time and level differences on speech intelligibility in noise.*** J.Acoust.Soc.Am. 83, 1508-1516.

Burger, J.F. 1958. ***Front-back discrimination of the hearing system.*** Acousitca 8, 301-302

Burgtorf, W. 1963. ***On the subjective effect of sound fields in rooms (backwards masking, phantom sound sources).*** Acoustica 13, 86-91

Burgtorf, W., Wagener, B. 1968. ***Masking by means of subjectively diffuse sound fields.*** Acoustica 14, 254-265.

Carhart, R., Tillman, T.W. and Greetis E.S. 1969a. ***Release from multiple maskers: effect of interaural time disparities.*** J.Acoust.Soc.Am. 45, 411-418.

Carhart, R., Tillman, T.W. and Greetis E.S. 1969b. ***Perceptual masking in multiple sound background.*** J.Acoust.Soc.Am. 45, 694-703.

Cherry, E.C. 1953. ***Some experiments on the recognition of speech, with one and with two ears.*** J.Acoust.Soc.Am. 25, 975-979.

Cohen, J. 1988. **Statistical power analysis for the behavioral sciences (2nd edition).** Hillsdale, NJ: Erlbaum.

Colburn, H.S. and Durlach, N.I. 1978. ***Handbook of perception (Carterette, E.C. and Friedman, M.P. editor).*** Academic Press: New York.

Culling, J.F. and Summerfield, Q. 1995. ***The role of frequency modulation in the perceptual segregation of concurrent vowels.*** J.Acoust.Soc.Am 98(2), 837-846.

Damaske, P. 1970. ***Direction dependence of the spectrum and the correlation function of the signals received at the ears.*** Acoustica 22, 191-204.

Davis, D. and Davis, C. 1987. ***Sound system engineering.*** Haward W. Sams&co: USA.

Deatherage, B.H. and Evan, T.R. 1969. ***Binarual masking: backward, forward, and simultaneous masking effect.*** J.Acoust.Soc.Am. 46, 362-371.

Dolan, R. and Trahiotis, C. 1970 ***Binaural interaction in backward masking.*** J.Acoust.Soc.Am. 47, 131.

Dolan, T.R. 1968. ***Effects of masker spectrum level on MLD at low frequencies.*** J.Acoust.Soc.Am. 44, 1507-1512.

Duifhuis, H. 1973. ***Consequences of peripheral frequency selectivity for non simultaneous masking.*** J.Acoust.Soc.Am. 54, 1471-1488.

Duifhuis, H. 1971. ***Audibility of high harmonics in a periodic pulse. II. Time effect.*** J. Acoust. Soc. Am. 49, 1155–1162.

Durlach, N.I. 1972. ***Foundations of modern auditory theory***, vol. 2. Academic Press: New York.

Durlach, N.I. 1963. **Equalization and cancellation theory of binaural masking level differences**. J.Acoust.Soc.Am. 35, 1206-1218.

Durrant, J.D. and Lovrinic, J.H. 1995. **Bases** of **Hearing Science**, 3rd Ed., Williams and Wilkins, Baltimore.

Ebata, M., Sone, T., and Nimura, T. 1968. ***Improvement of hearing ability by directional information*** . J. Acoust.Soc.Am. 43, 289-297.

Elliot, L.L. 1971. ***Backward and forward masking.*** Audiology. 10, 65-76.

Elliot, L.L. 1962. ***Backward and forward masking of probe tones of different frequencies.*** J.Acoust.Soc.Am. 34(8), 1116-1117.

Egan, J.P. and Hake, H.W. 1950. ***On the masking pattern of a simple auditory stimulus.*** J. Acoust.Soc.Am. 22, 622-630.

Egan, M.D. 1972. ***Concepts in Architectural Acoustics***. McGraw-Hill: USA.

Flanagan, J.L. and Watson, B.J. 1966. ***Binaural unmasking of complex signals***. J. Acoust.Soc.Am. 40, 456-458.

Fletcher, H. 1940. ***Auditory Patterns.*** Rev. Mod. Phys. 12, 47-65.

Gelfand, S.A. 1998. ***Hearing***. Marcel Dekker,inc.: New York.

Gilkey, R.H. and Good, M.D. 1995. ***Effect of frequency on free-field masking.*** Human factors. Vol.37(4), 835-843

Good, M.D., Gilkey, R.H. and Ball, J.M. 1997. ***The relation between detection in noise and localization in noise in the free field.*** Binaural and spatial hearing in real and virtual environments. Erlbaum: New York

Gray, P. 1991. ***Psychology***. New York, NY: Worth Publishers, Inc.

Green, D.M. and Henning G.B. 1969. ***Audition.*** Ann.Rev.Psychol.20, 105-128.

Haas, H. 1972. ***The influence of single echo on the audibility of speech***. J.Audio Eng.Soc.20, 145-149.

Hall, J.W., Haggard, M.P., and Fernandes, M.A. 1984. ***Detection in noise by spectro-temporal pattern analysis.*** J. Acoust. Soc. Am. 76, 50-56.

Hamilton, P.M.1957. ***Noise masked threshold as a function of tonal duration and masking noise band width***. J. Acoust.Soc.Am. 29, 506-511.

Handel, S. 1989. ***Listening.*** The MIT press: London.

Hawkins, J.E. and Steven, S.S. 1950. ***The masking of pure tones and of speech by white noise.*** J.Acoust.Soc.Am. 22, 6-13.

Heringa, Ir.P.H., Melkemijer,Ir.R.A. and Pentz,Ir.V.M.A. 1988. ***Concert hall acoustics.*** Proceeding of the Institute of Acoustic, Paris, 331.

Hidaka, T. and Beranek, L. and Okano, T. 1995. ***Interaural cross correlation (IACC), lateral fraction (LF), and sound energy level (G) as partial measures of acoustical quality in concert halls.*** J. Acoust Soc. Am. 98, 988-1007.

Hirsh, I. J. 1948. ***The influence of interaural phase on interaural summation and inhibition.*** J. Acoust. Soc. Am. 20, 536-544.

Hirsh, I. J. and Burgeat, M. 1958. ***Binaural effect in remoter masking.*** J. Acoust. Soc. Am. 30, 827-832.

Jeffe,C., Scarbough, P., Copper, R and Clement, P. 1999. ***Excellent acoustics and a low RT? A case study of the recent renovation of the Kennady center concert hall, Washington DC.*** Acoustica. vol.85, S61.

Jeffress, L.A., Blodgett, H.C., and Deatherage, B.H. 1952. ***The masking of tone by white noise as a function of the interaural phases of both components.*** J. Acoust.Soc.Am. 24, 523-527.

Jordan, V. L., 1969. ***Room Acoustics and Architectural Acoustics—Development in Recent Years***. Applied Acoustics, 2, 77-99.

Katz, J. 1985. ***Handbook of clinical audiology***. Williams & Wilkins

Keet, W. de V. 1968. ***The influence of early lateral reflections on the spatial impression.*** Proceedings of the 6th International Congress on Acoustics, Tokyo. E53-E56.

Kendall, G.S. 1995. ***A 3-D sound primer: Directional hearing and stereo reproduction***. Computer Music Journal 19 (4, Winter): 23-46

Kent, R.D. and Read, C. 1992. ***The acoustics analysis of speech.*** California: Singular Pub. Group: USA

Knudsen, V. O. 1929. ***The hearing of speech in auditoriums.*** J.Acoust.Soc.Am. 1, 56-82.

Koenig, W. 1950. ***Subjective effects in binaural hearing.*** J. Acoust.Soc.Am. 22, 61-61.

Koenig, A. H., Allen, J. B. and Berkley, D. A. 1977. ***Determination of masking-level difference in a reverberant environment.*** J.Acoust.Soc.Am. 61(1), 1374-1376.

Kopčo, N. 2003. ***Spatial hearing, Auditory sensitivity, and pattern recognition in noisy environment.*** Dissertation, Boston University.

Kurtović, H. 1975. ***The influence of reflected sound upon speech intelligibility.*** Acoustica 33, 32-39

Kuttruff, H. 1979. ***Room acoustics.*** Applied science publishers Ltd.: London.

Lambert, R.M. 1974. ***Dynamic theory of sound source localization.*** J. Acoust. Soc. Am. 56 (1), 165-171.

Langford, T.L. and Jeffress, L.A. 1964. ***Effect of noise cross correlation on binaural signal detection.*** J.Acoust.Soc.Am. 36, 1455-1458.

Lawrence, A. 1989. ***Acoustics and the built environment.*** Elsevier applied science: London.

Lewis, R. J. 2000. ***Power analysis and sample size determination: concepts and software tool.*** The 2000 annual meeting of the society for academic emergency medicine (SAEM) in San Francisco, California.

Licklider, J. C. R. 1948. ***The influence of interaural phase relation upon the masking of speech by white noise***. J.Acoust.Soc.Am. 20, 150-159.

Lipsey, M.W. 1990. ***Design Sensitivity: Statistical Power for Experimental Research.*** Sage Publishing: USA.

Machin D, Campbell MJ, Fayers P, Pinol, A.1998. ***Statistical Tables for the Design of Clinical Studies, Second Edition.*** Blackwell, Oxford.

Maekawa, Z. and Lord, P.1994. ***Environmental and architectural acoustics.*** E&FN Spon: UK.

Marshall, A.H. 1968a. ***Levels of reflection masking in concert halls***. J. Sound Vib. 7(1), 116-118.

Marshall A.H. 1968b. ***Concert hall shapes for minimum masking of lateral reflections***. The 6th International congress on acoustics, Tokyo.

Marshall, A.H. 1967. ***A note on the importance of room cross section in concert halls***. J. Sound Vib. 5(1), 100-112.

McEachern, R. 1992. ***How the ear really works***. 1992 IEEE Int. Symposium on Time-Frequency and Time-Scale Analysis, Victoria, BC. Proceedings, 437-440.

McFadden, D. 1968. ***Masking level differences determined with and without interaural disparities***. J.Acoust.Soc.Am. 44, 212-213.

McGurk, H. and MacDonald, J. 1976. ***Hearing lips and seeing voices***. Nature, Vol 264(5588), 746–748

Miller, G.A. 1974. ***The masking of speech***. Psychological Bulletin 44, 105-129.

Moore, B.C.J. 1997. ***An introduction to the psychology of hearing (4^{th} edition).*** Academic press, inc.:London.

Moore, B.C.J. 1982. ***An introduction to the psychology of hearing.*** Academic press, inc.:London.

Nábělek, A.K.1988. ***Identification of vowels in quiet, noise and in reverberation by hearing impaired listeners.*** J.Acoust.Soc.Am. 84, 476-484.

Nábělek, A.K. and Dagenais, P.A. 1986. ***Vowel errors in noise and in reverberation by hearing impaired listeners.*** J.Acoust.Soc.Am. 80, 741-748.

Nábělek, A.K., Letowski, T.R., and Tucker, F.M. 1989. ***Reverberant overlap-and self-masking in consonant identification.*** J.Acoust.Soc.Am. 86(4),1259-1265.

Peterson, G. E., and Barney, H. L. 1952. ***Control Methods Used in a Study of the Vowels.*** J Acoust Soc Am. 24, 175-185.

Pickett, J.M. 1959. ***Backward masking.*** J. Acoust.Soc.Am. 31, 1613-1615.

Rabiner, L. and Juang, B. 1993. ***Fundamentals of speech recognition.*** Prentice Hall.

Rabiner, L.R., Laurence, C.L., and Durlach, N.I. 1966. **Further results on binaural masking and the EC-model.** J.Acoust.Soc.Am. 40, 62-70.

Robinson, D.E. and Jeffress, L.A.1963. ***Effect of varying the interaural noise correlation on the detectability of tone signals.*** J.Acoust.Soc.Am. 35, 147-152.

Rodgers, C.A.P. 1981. ***Pinna transformation and sound reproduction.*** J.Audio Eng. Soc. 29, 226-234.

Roederer, J.G. 1995. ***The physics and psychophysics of music: An introduction.*** Third Edition. New York, NY: Springer-Verlag.

Saberi, K., Dostal, L., Sadralodabai, T. Bull, V. And Perrott, D.R. 1991. ***Free field release from masking.*** J.Acoust.Soc.Am. 90(3), 1355-1370.

Santon, F. 1987. ***Detection of a pure sound in the presence of masking noise, and its dependence on the angle of incidence of noise. Relation with the speech reception threshold.*** Acoustica 63, 222-228

Schenkel, K.D. 1964. ***On the dependence of absolute threshold of perception on the interaural phase angle of the test signal.*** Acoustica 14, 337-346.

Schroeder, M. R., 1965. ***New Method of Measuring Reverberation Time.*** J.Acoust.Soc.Am. 37, 409

Shavelson, R.J. 1988. ***Statistical reasoning for the behavioral sciences, 2nd edition.*** Allyn & Bacon: Boston.

Spiegel, M.E. 1987. ***Speech masking I: simultaneous and non simultaneous masking with stop/d/ and flap /f/ closures.*** J.Acoust.Soc.Am. 82(5), 1492-1502.

Stevens, S.S., Miller, J., and Truscott, I. 1946. ***The masking of speech by sine waves, square waves, and regular and modulated pulses.*** J.Acoust.Soc.Am. 18, 418-424

Thomas, L. and Kerbs, C.J. 1997. ***A review of statistical power analysis software.*** Bulletin of the Ecological Society of America 78(2), 126-139.

Thurlow, W.R., Mangel J.W., and Runge P.S.1967. ***Effects of induced head movements on localization of direct sound.*** J.Acoust.Soc.Am. 42, 489-493.

Thurlow, W.R. and Mergener, J.R. 1971. ***Effect of stimulus duration on localization of direction of noise stimuli.*** J. Speech and Hearing Res.13, 826-838.

Van de par, S. and Kohlrausch, A. 1999. ***Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters.*** J.Acoust.Soc.Am. 106(4), 1940-1946.

Webster, F.A.1951. ***The influence of interaural phase on masked thresholds.*** J. Acoust.Soc.Am. 23, 452-462.

Wegel, R.L. and Lane, C.E. 1924. ***The auditory masking of one sound by another and its probable relation to the dynamics of inner ear.*** Phys.Rev. 23, 266-285.

Wightman, F. L. and Kistler, D. J. and Andersen, K. 1994. ***Reassessment of the role of head movements in human sound localization.*** J.Acoust.Soc.Am 95(5), 3003-3004.

Wilson, R.H. and Carhart, R. 1971. ***Forward and backward masking: interactions and additivity***. J.Acoust.Soc.Am. 49(4), 1254-1263.

Yoshida, T. 1994. ***The Rewritable MiniDisc System***. Processdings of the IEEE vol. 82, No.10. 1492-1500

Yost, W. A. and Mapes-Riordan, D. 1997. ***Head movements in localization.*** J.Acoust.Soc.Am. 101(5), 3105.

Yost, W. A. and Walton, J. 1977. ***Hierarchy of masking-level difference obtained for temporal masking.*** J.Acoust.Soc.Am. 61(1), 1376-1379.

Zurek, P.M. 1993. ***Acoustics Factors Affecting Hearing Aid Performance.*** College-Hill Press: Boston.

Zurek, P.M., Freyman, R. L., and Balakrishnan, U. 2004. ***Auditory target detection in reverberation.*** J.Acoust. Soc. Am. 111(4), 1609-1620.

Zwicker, E. and Fastl, H. 1990. ***Psychoacoustics***. Springer Verlag Berlin Heidelberg: Germany.

Zwicker, E. and Henning, G. B. 1984. ***Binaural masking-level differences with tone masked by noise of various bandwidth and levels.*** Hearing Research. Vol.14, 179-183.

APPENDIX 1

APPARATUS USED IN THIS STUDY

Experiment 1

	model	serial number
Mini-disc SHARP	T-20	91216912
Audiometer Kamplex	AD-27	5539
Amplifier H&H	TPA-50	I 5204
Amplifier H&H	TPA-25	B258
Speaker KEF	C-35	023590, 018667
Sound level meter DAWE	D-1421D	04310029
Calibrator DAWE	D-1411E	12412538

Experiment 2 (part 1 and 2)

	model	serial number
DAT Player Sony	DTC-1000ES	50I787
Audiometer Kamplex	AD-27	5539
Attenuator Hatfield	2120	59914
Mixer YAMAHA	DMP-7	I834
Headphone STAX	SRD-X	05968
Sound level meter B&K	2218	257647
Artificial Ear B&K	4152	1025329
Calibrator B&K	4230	1026449

Experiment 3

	model	serial number
DAT Player Sony	DTC-1000ES	50I787
Audiometer Kamplex	AD-27	5539
Amplifier H&H	TPA-50	I 5204
Amplifier H&H	TPA-25	B258
Speaker KEF	C-35	023590, 018667
Sound level meter DAWE	D-1421D	04310029
Calibrator DAWE	D-1411E	12412538

APPENDIX 2

FREQUENCY ANALYSIS FOR TEST STIMULI IN EXPERIMENT 2

Using Analyzer Hewlett-peckard type 3566A serial no. 2911A00263

Part 1

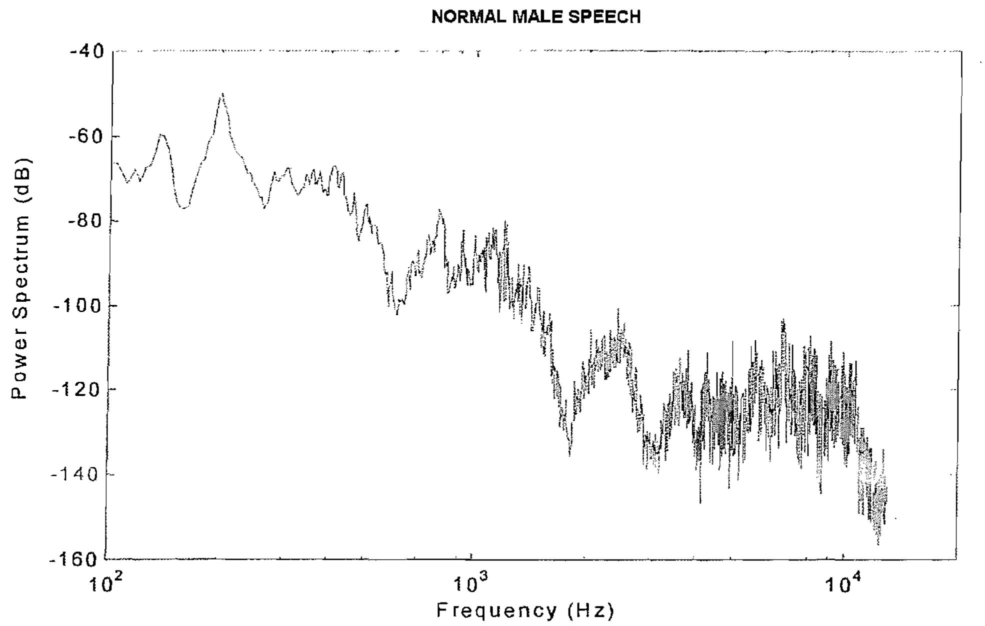


Figure a2.1 Frequency analysis for normal male speech.

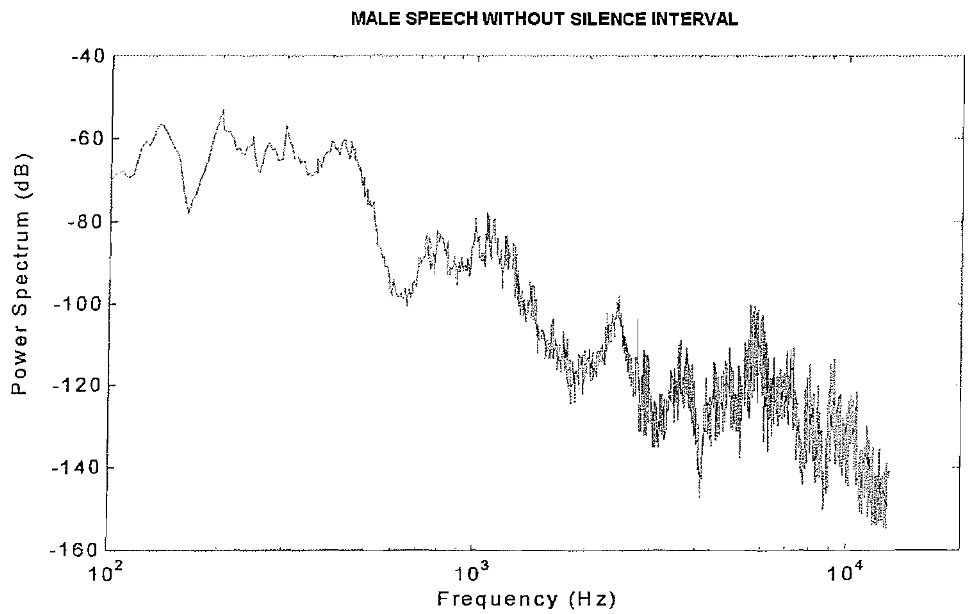


Figure a2.2 Frequency analysis for male speech without silence gap.

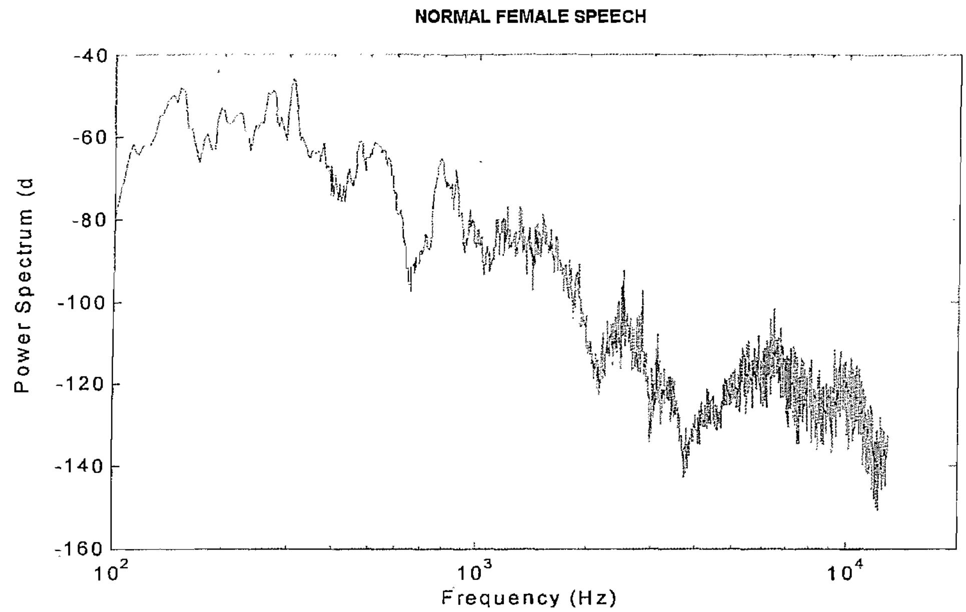


Figure a2.3 Frequency analysis for normal female speech.

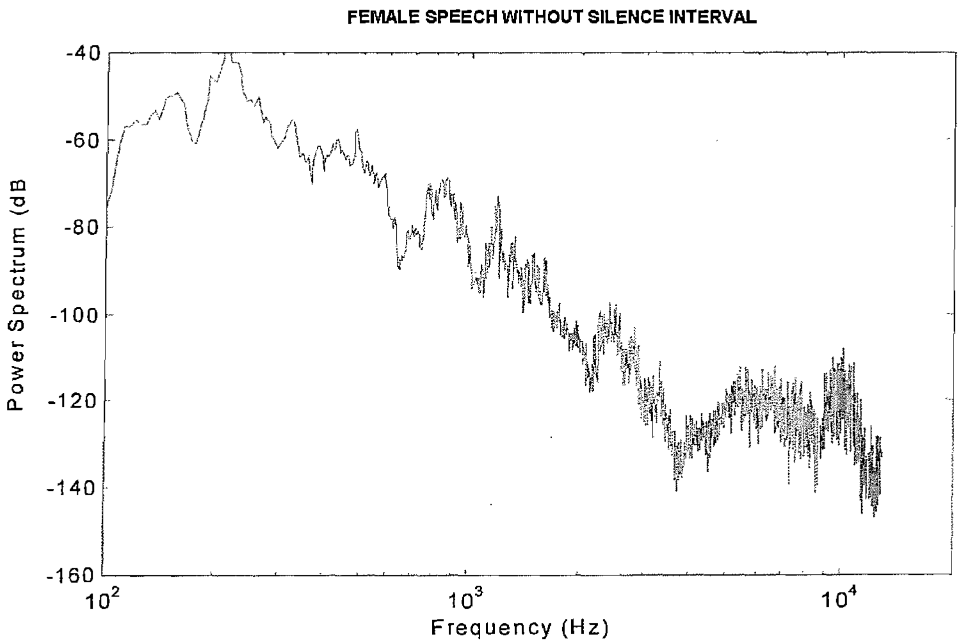


Figure a2.4 Frequency analysis for female speech without silence gap.

Part 2

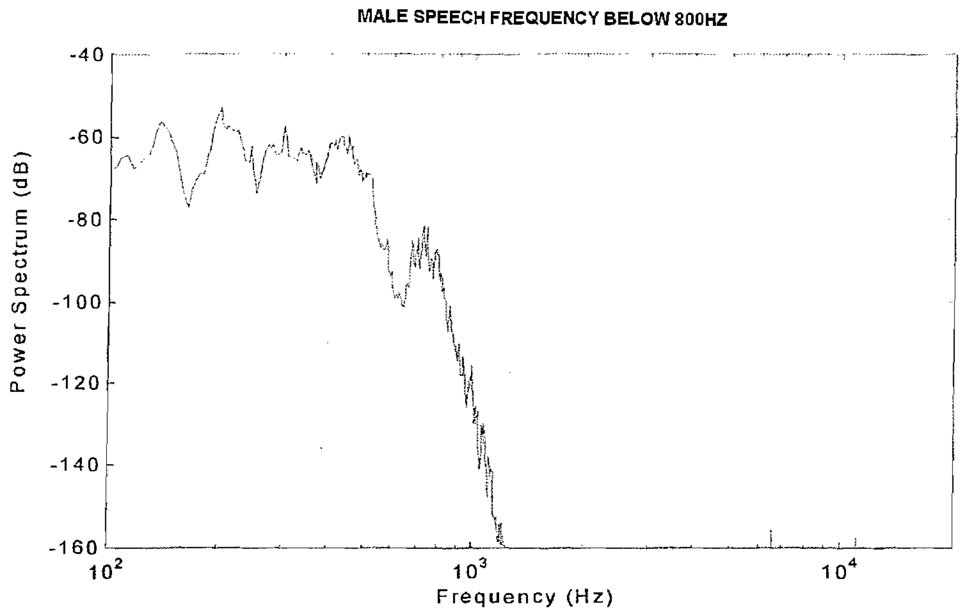


Figure a2.5 Frequency analysis for male speech edited frequency below 800Hz.

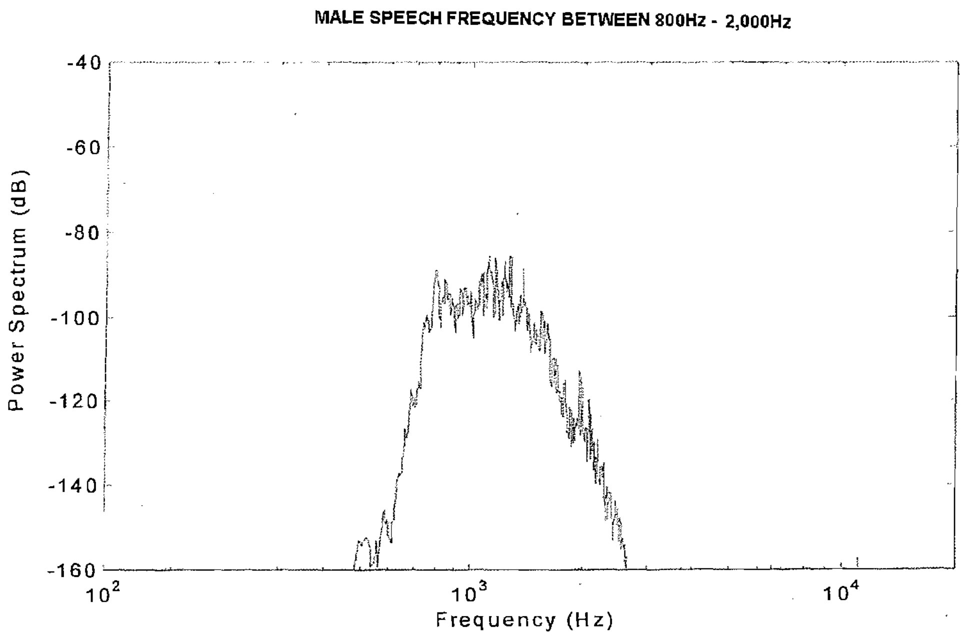


Figure a2.6 Frequency analysis for male speech edited frequency between 800Hz-2,000Hz.

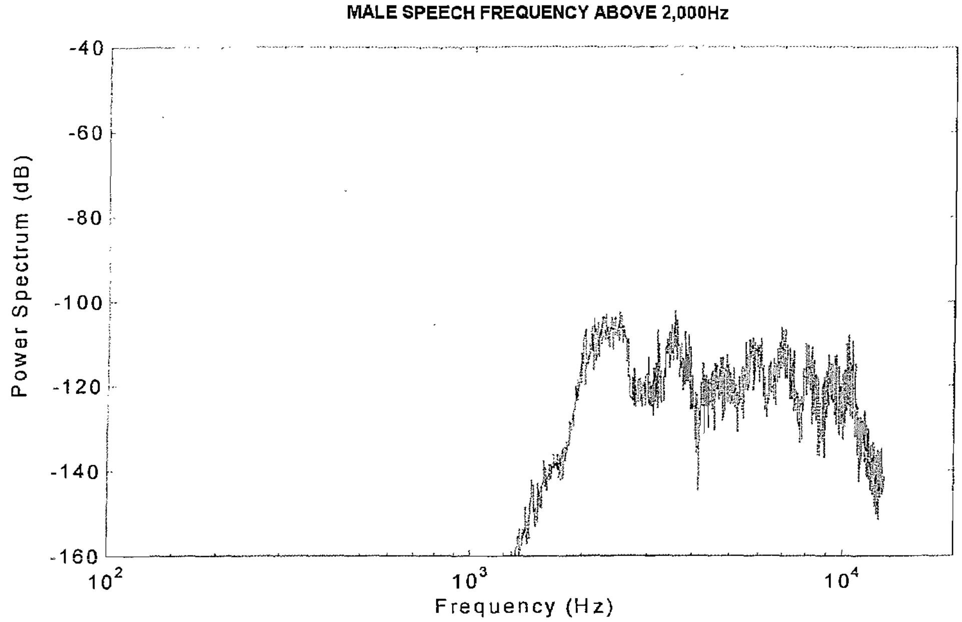


Figure a2.7 Frequency analysis for male speech edited frequency above 2,000Hz.

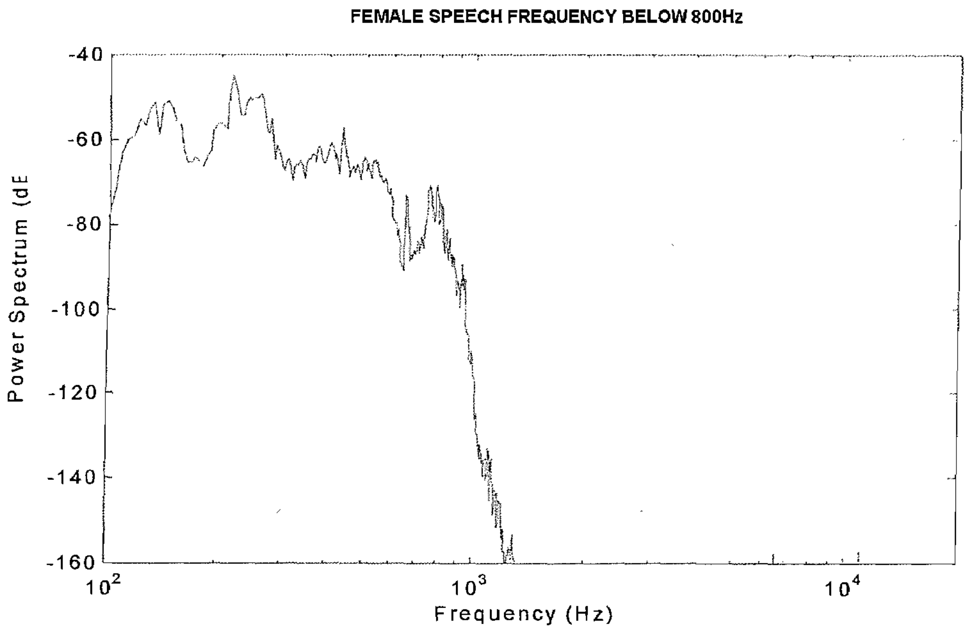


Figure a2.8 Frequency analysis for female speech edited frequency below 800Hz.

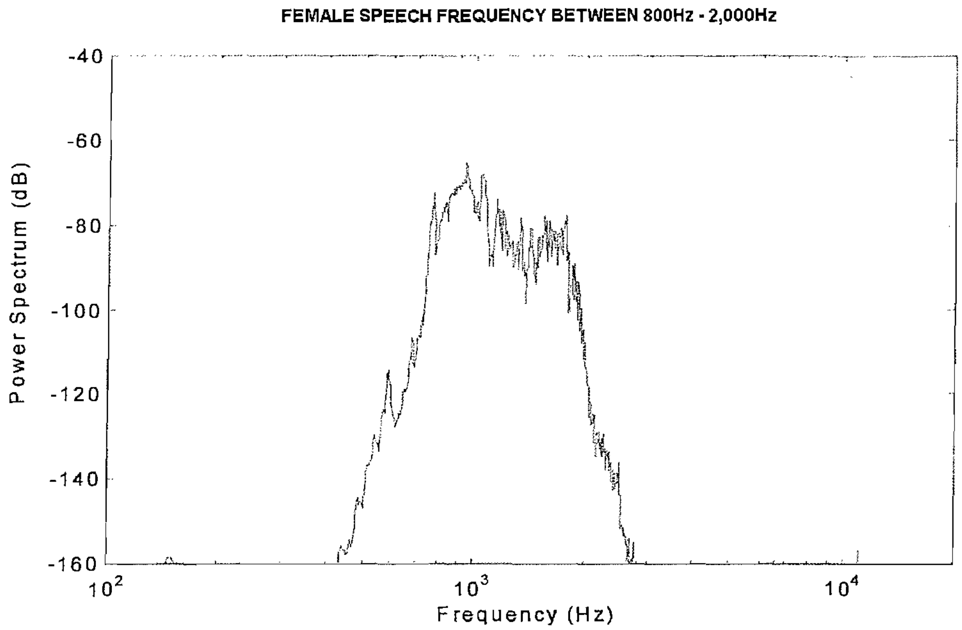


Figure a2.9 Frequency analysis for female speech edited frequency between 800Hz-2,000Hz.

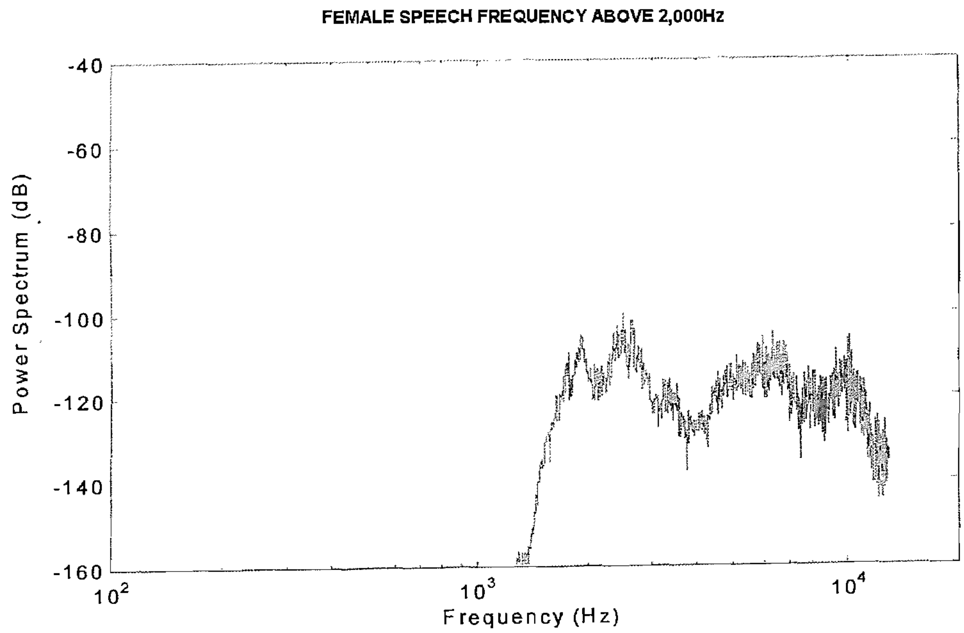


Figure a2.10 Frequency analysis for female speech edited frequency above 2,000Hz.

APPENDIX 3

DATA COLLECTION SHEET

Absolute threshold

SUBJECT NUMBER:________

THRESHOLD OF HEARING

frequency

	250	500	1k	2k	4k	8k	
-10							
-5							
0							
5							
10							
15							
20							
25							
30							
35							
40							
45							

ABSOLUTE (UNMASKED) THRESHOLDS OF SPEECH

SOURCE: *male speech X*
female speech O

INTERVALS

	1	2	3	4	5	6	
-10							
-5							
0							
5							
10							
15							
20							
25							
30							
35							
40							
45							

AVERAGE LEVEL:___dB(A)

Experiment 1

DELAY:________

AZIMUTH ANGLE:________

SOURCE: *male speech ____X*
female speech O

SOUND FIELD AUDIOGRAM

MASKER LEVEL:__40__dB MASKER LEVEL:__60__dB MASKER LEVEL:__80__dB

INTERVALS
1 2 3 4 5 6

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70

INTERVALS
1 2 3 4 5 6

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70

INTERVALS
1 2 3 4 5 6

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70

AVERAGE LEVEL:___dB(A) AVERAGE LEVEL:____dB(A) AVERAGE LEVEL:____dB(A)

Experiment 2 part 1

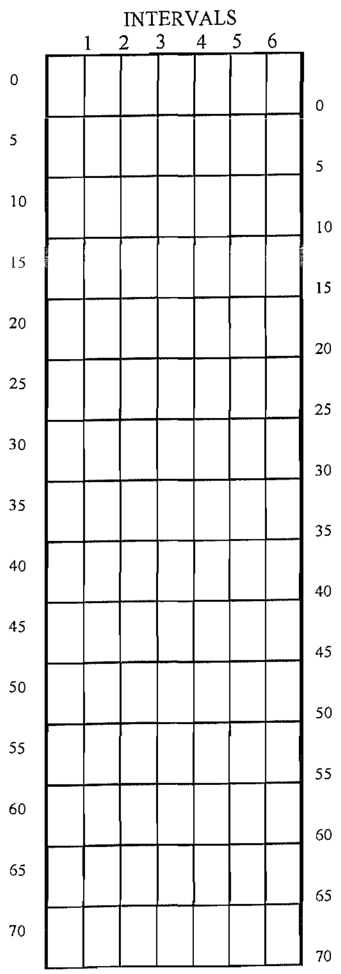
SUBJECT NUMBER:________

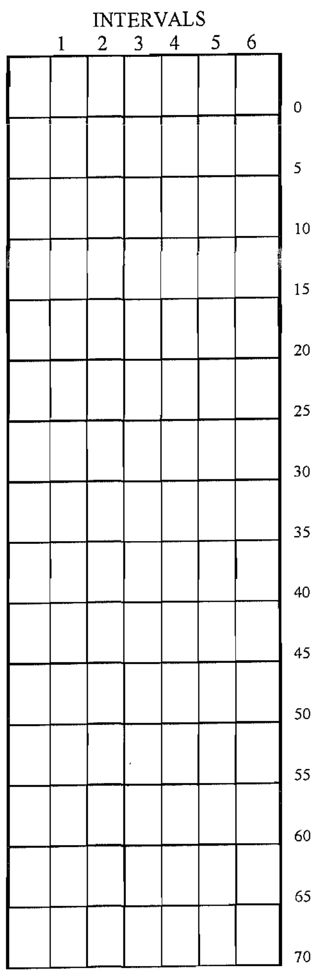
WITH / WITHOUT GAP

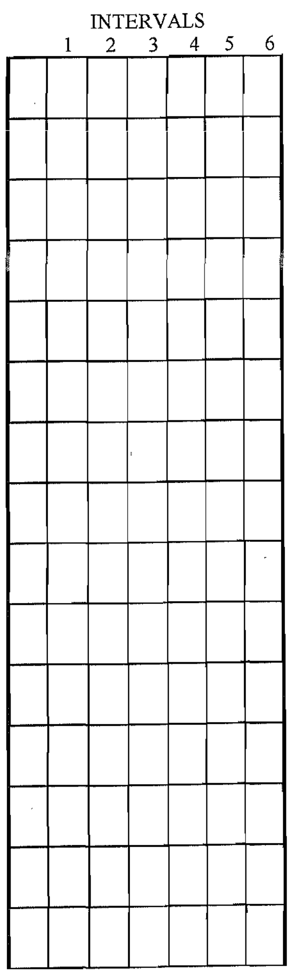
SOURCE: *male speech X*
female speech O

AUDIOGRAM

MASKER DELAY:__15__ms MASKER DELAY:__30__ms MASKER DELAY:__50__ms







AVERAGE LEVEL:___dB(A) AVERAGE LEVEL:____dB(A) AVERAGE LEVEL:____dB(A)

AVERAGE LEVEL:___dB(A) AVERAGE LEVEL:____dB(A) AVERAGE LEVEL:____dB(A)

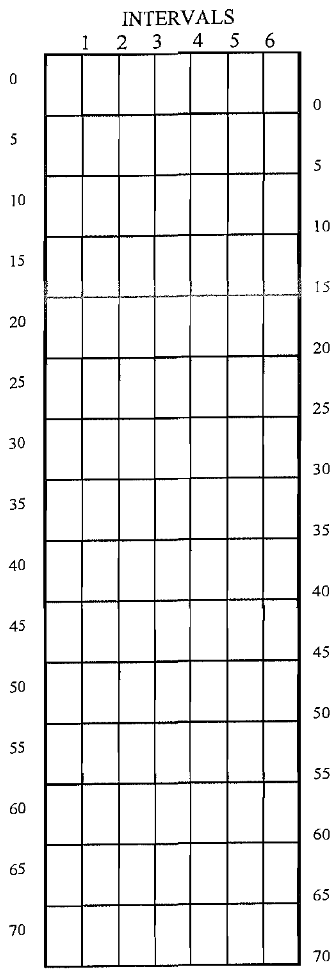
Experiment 2 part 2

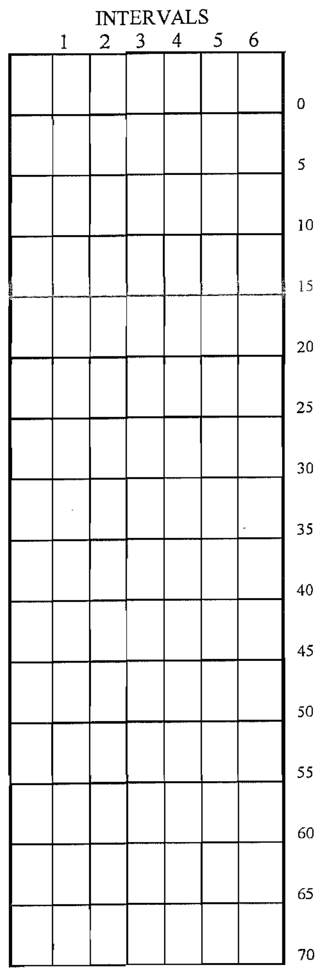
SUBJECT NUMBER:________

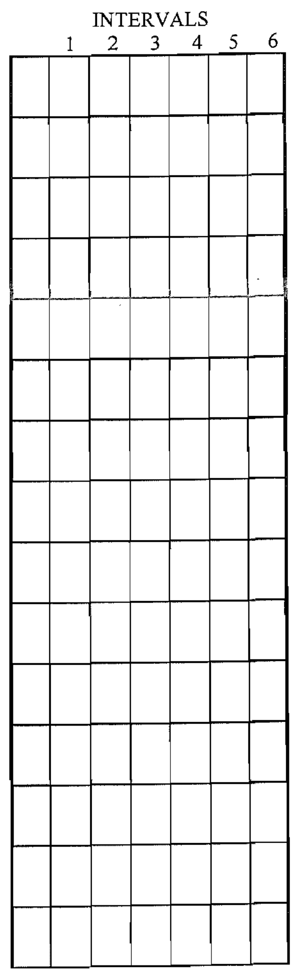
SIGNAL:_______________
MASKER:_______________
SOURCE: *male speech X*
female speech O

AUDIOGRAM

MASKER DELAY:__15__ms MASKER DELAY:__30__ms MASKER DELAY:__50__ms







AVERAGE LEVEL:___dB(A) AVERAGE LEVEL:____dB(A) AVERAGE LEVEL:____dB(A)

AVERAGE LEVEL:___dB(A) AVERAGE LEVEL:____dB(A) AVERAGE LEVEL:____dB(A)

Experiment 3

SUBJECT NUMBER:________

HEAD FIXED / MOVED

SOUND FIELD AUDIOGRAM

AZIMUTH ___________

INTERVALS

	1	2	3	4	5	6	
0							0
5							5
10							10
15							15
20							20
25							25
30							30
35							35
40							40
45							45
50							50
55							55
60							60
65							65
70							70

AVERAGE LEVEL:___dB(A)

AZIMUTH ___________

INTERVALS

1	2	3	4	5	6	
						0
						5
						10
						15
						20
						25
						30
						35
						40
						45
						50
						55
						60
						65
						70

AVERAGE LEVEL:____dB(A)

AZIMUTH ___________

INTERVALS

1	2	3	4	5	6

AVERAGE LEVEL:____dB(A)

APPENDIX 4

PILOT TEST RESULTS

Experiment 1

delay 0	angle 0	SPL 40	angle 0	SPL 60	Angle 0	SPL 80
subject no.	Male	female	Male	female	Male	female
1	35.0	32.5	50.0	37.5	66.6	55.0
2	37.5	34.0	50.0	45.0	65.0	55.0

delay 30	angle 0	SPL 40	angle 0	SPL 60	angle 0	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	27.5	22.5	42.5	30.0	60.0	40.0
2	30.0	30.0	45.0	37.5	60.0	45.5

delay 50	angle 0	SPL 40	angle 0	SPL 60	angle 0	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	22.5	17.5	35.0	25.0	47.5	35.0
2	25.0	22.5	37.5	30.0	45.0	40.0

delay 100	angle 0	SPL 40	angle 0	SPL 60	angle 0	SPL 80
subject no.	male	female	Male	Female	Male	Female
1	17.5	9.0	25.0	20.0	35.0	25.0
2	20.0	17.5	27.5	25.0	40.0	35.0

delay 15	angle 30	SPL 40	angle 30	SPL 60	angle 30	SPL 80
subject no.	Male	female	Male	female	Male	Female
1	30.0	25.0	35.0	30.0	52.5	47.5
2	30.0	32.5	40.0	35.0	55.0	50.0

delay 30	angle 30	SPL 40	angle 30	SPL 60	angle 30	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	25.0	20.0	27.5	25.0	40.0	32.5
2	30.0	22.5	35.0	32.5	45.0	40.0

delay 50	angle 30	SPL 40	angle 30	SPL 60	angle 30	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	17.5	15.0	25.0	17.5	30.0	30.0
2	25.0	20.8	30.0	27.5	40.0	35.0

delay 100	angle 30	SPL 40	angle 30	SPL 60	angle 30	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	15.0	12.5	20.0	20.0	27.5	27.5
2	20.0	15.0	25.0	22.5	30.0	30.0

delay 15	angle 60	SPL 40	angle 60	SPL 60	angle 60	SPL 80
subject no.	Male	female	Male	female	male	Female
1	25.0	27.5	30.0	30.0	45.0	37.5
2	27.5	27.5	32.5	32.5	50.0	40.0

delay 30	angle 60	SPL 40	angle 60	SPL 60	angle 60	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	20.0	25.0	22.5	22.5	40.0	32.5
2	27.5	25.0	27.5	27.5	45.0	40.0

delay 50	angle 60	SPL 40	angle 60	SPL 60	angle 60	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	20.0	17.5	22.5	22.5	35.0	27.5
2	25.0	20.0	27.5	27.5	40.0	30.0

delay 100	angle 60	SPL 40	angle 60	SPL 60	angle 60	SPL 80
subject no.	male	Female	Male	Female	Male	Female
1	12.5	12.5	17.5	15.0	25.0	22.5
2	17.5	17.5	25.0	22.5	35.0	32.5

delay 15	angle 90	SPL 40	angle 90	SPL 60	angle 90	SPL 80
subject no.	Male	female	Male	female	male	Female
1	25.0	30.0	30.0	32.5	45.0	40.8
2	27.5	32.5	37.5	40.0	50.0	47.5

delay 30	angle 90	SPL 40	angle 90	SPL 60	angle 90	SPL 80
subject no.	Male	female	Male	female	male	Female
1	22.5	20.0	30.0	25.0	40.0	30.0
2	22.5	27.5	37.5	32.5	42.5	35.0

delay 50	angle 90	SPL 40	angle 90	SPL 60	angle 90	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	15.0	20.0	20.0	20.0	27.5	25.0
2	20.0	25.0	25.0	27.5	30.0	35.0

delay 100	angle 90	SPL 40	angle 90	SPL 60	angle 90	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	12.5	15.0	15.0	17.5	22.5	20.0
2	17.5	17.5	22.5	25.0	27.5	25.0

Experiment 2 Part 1

subject	male speech delay (ms)			female speech delay (ms)		
	15	30	50	15	30	50
1	47.5	45.0	32.5	50.0	45.0	35.0
2	45.0	45.0	35.0	47.5	47.5	35.0

subject	male speech without silence gap delay (ms)			female speech without silence gap delay (ms)		
	15	30	50	15	30	50
1	50.0	45.0	40.0	50.0	45.0	35.0
2	45.0	42.5	37.5	50.0	45.0	40.0

Experiment 2 Part 2

signal below 800Hz
masker below 800Hz

subject	male speech delay (ms)			female speech delay (ms)		
	15	30	50	15	30	50
1	52.5	45.0	40.0	45.0	40.0	40.0
2	50.0	45.0	40.0	50.0	45.0	40.0

signal below 800Hz
masker between 800Hz-2,000Hz

subject	male speech delay (ms)			female speech delay (ms)		
	15	30	50	15	30	50
1	17.5	20.0	17.5	20.0	20.0	17.5
2	17.5	17.5	15.0	17.5	20.0	17.5

signal below 800Hz
masker above 2,000Hz

subject	male speech delay (ms)			female speech delay (ms)		
	15	30	50	15	30	50
1	15.0	12.5	15.0	15.0	15.0	15.0
2	15.0	15.0	15.0	15.0	12.5	12.5

signal between 800Hz-2,000Hz
masker below 800Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	27.5	27.5	30.0	27.5	25.0	25.0
2	25.0	25.0	22.5	25.0	22.5	22.5

signal between 800Hz-2,000Hz
masker between 800Hz-2,000Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	40.0	42.5	32.5	45.0	37.5	30.0
2	47.5	45.0	35.0	45.0	45.0	32.5

signal between 800Hz-2,000Hz
masker above 2,000Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	20.0	20.0	20.0	22.5	17.5	20.0
2	22.5	20.0	20.0	22.5	22.5	22.5

signal above 2,000Hz
masker below 800Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	12.5	15.0	17.5	12.5	17.5	15.0
2	15.0	15.0	15.0	15.0	17.5	15.0

signal above 2,000Hz
masker between 800Hz-2,000Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	15.0	15.0	15.0	15.0	12.5	12.5
2	20.0	17.5	20.0	20.0	20.0	20.0

signal above 2,000Hz
masker above 2,000Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	45.0	40.0	35.0	45.0	37.5	32.5
2	45.0	40.0	32.5	45.0	40.0	32.5

Experiment 3

		subjects	
Azimuth angle	Head	1	2
0	Fixed	52.5	52.5
	rotated	52.5	52.5
2	fixed	52.5	52.5
	rotated	50.0	50.0
4	fixed	50.0	52.5
	rotated	47.5	47.5
6	fixed	47.5	45.0
	rotated	45.0	40.0
8	fixed	40.0	37.5
	rotated	50.0	52.5
10	fixed	52.5	50.0
	rotated	47.5	45.0
15	fixed	45.0	40.0
	rotated	42.5	40.0
30	fixed	40.0	37.5
	rotated	40.0	37.5
90	fixed	37.5	37.0
	rotated	37.5	37.5

APPENDIX 5

RESULTS

Experiment 1

delay 0	angle 0	SPL 40	angle 0	SPL 60	Angle 0	SPL 80
subject no.	Male	female	Male	female	Male	female
1	34.1	32.5	49.1	38.3	66.6	55.8
2	37.5	34.1	50.8	46.6	65.0	55.8
3	36.6	30.8	48.3	45.0	64.1	57.5
4	33.3	31.6	45.0	44.1	66.6	62.5
5	33.3	25.8	46.6	42.5	65.0	59.1

delay 30	angle 0	SPL 40	angle 0	SPL 60	angle 0	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	27.5	22.5	41.6	31.6	60.0	40.8
2	31.6	30.8	45.0	38.3	60.0	46.6
3	25.0	27.5	43.3	35.8	60.8	43.3
4	30.8	26.6	38.3	30.8	58.3	43.3
5	30.8	22.5	39.1	34.1	59.1	44.1

delay 50	angle 0	SPL 40	angle 0	SPL 60	angle 0	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	22.5	17.5	35.0	25.0	47.5	35.0
2	25.8	23.3	38.3	30.0	44.1	40.0
3	23.3	21.6	39.1	26.6	50.8	40.0
4	24.1	20.0	35.0	25.8	48.3	36.6
5	25.8	17.5	33.3	30.8	52.5	40.8

delay 100	angle 0	SPL 40	angle 0	SPL 60	angle 0	SPL 80
subject no.	male	Female	Male	Female	Male	Female
1	16.6	9.1	25.8	20.0	36.6	25.0
2	20.0	18.3	27.5	25.0	40.8	35.0
3	15.8	14.1	28.3	22.5	37.5	29.1
4	19.1	14.1	24.1	20.8	36.6	29.1
5	20.0	13.3	25.0	25.0	40.8	31.6

delay 15	angle 30	SPL 40	angle 30	SPL 60	angle 30	SPL 80
subject no.	male	Female	Male	female	Male	Female
1	29.1	25.8	36.6	30.8	52.5	47.5
2	30.0	31.6	40.0	39.1	56.6	50.0
3	30.0	25.0	36.6	35.0	49.1	47.5
4	30.0	21.6	34.1	35.8	53.3	46.6
5	28.3	26.6	38.3	35.8	52.5	50.8

delay 30	angle 30	SPL 40	angle 30	SPL 60	angle 30	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	24.1	21.6	28.3	25.8	39.1	32.5
2	30.8	22.5	35.0	33.3	46.6	40.0
3	23.3	23.3	26.6	25.0	40.0	37.5
4	21.6	20.0	28.3	31.6	42.5	35.8
5	25.0	22.5	30.8	30.0	45.0	38.3

delay 50	angle 30	SPL 40	angle 30	SPL 60	angle 30	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	16.6	14.1	24.1	18.3	31.6	30.0
2	25.0	20.8	30.0	28.3	40.0	35.0
3	19.1	17.5	25.8	23.3	34.1	34.1
4	19.1	17.5	25.0	23.3	36.6	30.0
5	19.1	15.8	28.3	22.5	35.8	32.5

delay 100	angle 30	SPL 40	angle 30	SPL 60	angle 30	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	15.8	12.5	20.0	20.8	27.5	24.1
2	20.0	15.8	25.0	24.1	36.6	33.3
3	17.5	14.1	20.8	20.0	30.0	28.3
4	16.6	13.3	20.0	15.0	32.5	25.8
5	15.8	13.3	23.3	18.3	30.8	26.6

delay 15	angle 60	SPL 40	angle 60	SPL 60	angle 60	SPL 80
subject no.	Male	Female	Male	female	male	Female
1	26.6	27.5	30.0	28.3	45.0	36.6
2	29.1	28.3	33.3	33.3	50.8	42.5
3	25.0	30.0	31.6	32.5	50.8	41.6
4	28.3	28.3	32.5	32.5	47.5	36.6
5	27.5	27.5	34.1	35.0	47.5	40.0

delay 30	angle 60	SPL 40	angle 60	SPL 60	angle 60	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	20.8	24.1	22.5	21.6	39.1	33.3
2	27.5	25.0	28.3	28.3	45.0	40.0
3	20.8	25.0	23.3	25.8	39.1	38.3
4	24.1	20.0	29.1	23.3	44.1	32.5
5	20.0	22.5	29.1	25.8	40.0	35.8

delay 50	angle 60	SPL 40	angle 60	SPL 60	angle 60	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	16.6	17.5	22.5	22.5	35.8	28.3
2	25.0	21.6	27.5	27.5	40.8	35.8
3	19.1	20.0	20.0	24.1	36.6	30.8
4	20.8	15.8	24.1	20.0	39.1	30.0
5	19.1	18.3	24.1	23.3	37.5	32.5

delay 100	angle 60	SPL 40	angle 60	SPL 60	angle 60	SPL 80
subject no.	male	Female	Male	Female	Male	Female
1	11.6	12.5	17.5	16.6	26.6	22.5
2	19.1	18.3	25.0	24.1	35.8	32.5
3	12.5	14.1	15.8	22.5	29.1	27.5
4	15.0	10.8	20.0	19.1	33.3	21.6
5	16.6	15.0	17.5	22.5	29.1	25.0

delay 15	angle 90	SPL 40	angle 90	SPL 60	angle 90	SPL 80
subject no.	Male	Female	Male	female	male	Female
1	25.0	30.0	30.8	31.6	44.1	40.8
2	27.5	32.5	38.3	40.0	50.0	48.3
3	25.0	30.0	31.6	35.0	40.0	45.8
4	26.6	25.8	36.6	35.0	45.8	40.8
5	25.8	27.5	35.8	36.6	42.5	45.0

delay 30	angle 90	SPL 40	angle 90	SPL 60	angle 90	SPL 80
subject no.	male	Female	Male	female	male	Female
1	22.5	20.0	30.0	25.8	39.1	30.0
2	22.5	27.5	37.5	32.5	42.5	35.8
3	25.0	24.1	30.0	30.0	39.1	35.8
4	26.6	20.0	28.3	28.3	37.5	30.8
5	25.0	21.6	31.6	30.0	40.0	34.1

delay 50	angle 90	SPL 40	angle 90	SPL 60	angle 90	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	15.8	20.8	19.1	21.6	28.3	25.8
2	20.8	25.0	25.8	28.3	30.8	34.1
3	14.1	19.1	18.3	23.3	30.0	30.8
4	20.0	17.5	24.1	21.6	32.5	28.3
5	20.0	20.0	19.1	25.0	30.0	28.3

delay 100	angle 90	SPL 40	angle 90	SPL 60	angle 90	SPL 80
subject no.	Male	Female	Male	Female	Male	Female
1	12.5	15.0	15.8	17.5	21.6	19.1
2	16.6	18.3	22.5	25.0	27.5	26.6
3	10.8	16.6	15.7	20.0	24.1	22.5
4	15.0	10.8	19.1	19.1	25.0	20.8
5	15.8	15.0	19.1	20.8	23.3	23.3

Experiment 2 Part 1

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	49.1	44.1	33.3	50.0	45.0	35.0
2	45.0	44.1	35.0	47.5	47.5	35.0
3	50.0	45.0	33.3	50.8	45.0	35.0
4	50.0	45.0	35.0	50.0	45.0	35.0
5	50.0	44.1	35.0	45.0	45.8	35.0
6	50.0	45.0	30.0	50.0	45.0	32.5
7	44.1	45.0	34.1	45.8	47.5	35.0

	male speech without silence gap			female speech without silence gap		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	50.8	45.0	40.0	50.0	45.0	34.1
2	45.8	43.3	38.3	50.0	45.0	40.0
3	50.0	45.0	39.1	50.0	45.0	35.8
4	50.0	45.0	38.3	50.0	45.0	35.0
5	48.3	44.1	40.0	50.0	45.0	40.0
6	50.0	45.0	40.0	50.0	45.0	32.5
7	45.8	41.6	37.5	50.0	45.0	39.1

Experiment 2 Part 2

signal below 800Hz
masker below 800Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	52.5	45.0	40.8	45.0	40.0	40.8
2	50.0	45.0	39.1	50.0	45.0	40.0
3	49.1	44.1	40.8	45.8	40.8	34.1
4	49.1	44.1	39.1	45.0	40.0	33.3
5	50.0	45.0	40.0	49.1	42.5	40.0
6	55.0	45.0	40.0	47.5	40.0	32.5
7	50.0	44.1	40.0	47.5	45.0	40.0

signal below 800Hz
masker between 800Hz-2,000Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	17.5	19.1	17.5	20.0	19.1	17.5
2	18.3	18.3	15.8	17.5	20.0	17.5
3	20.0	17.5	16.6	19.1	20.0	17.5
4	19.1	17.5	19.1	19.1	19.1	19.1
5	18.3	18.3	16.6	16.6	17.5	15.0
6	18.3	20.0	18.3	17.5	15.0	15.0
7	14.1	18.3	15.0	18.3	20.0	18.3

signal below 800Hz
masker above 2,000Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	15.0	13.3	15.0	15.8	15.0	15.8
2	15.8	15.8	15.8	14.1	13.3	13.3
3	16.6	14.1	15.0	15.8	15.8	14.1
4	17.5	15.8	15.8	15.0	15.0	15.8
5	15.8	13.3	14.1	15.0	14.1	15.0
6	10.0	13.3	12.5	10.0	15.0	12.5
7	15.0	15.0	12.5	15.8	14.1	15.0

signal between 800Hz-2,000Hz
masker below 800Hz

subject	male speech delay (ms)			female speech delay (ms)		
	15	30	50	15	30	50
1	28.3	28.3	29.1	26.6	25.8	26.1
2	24.1	23.3	21.6	23.3	22.5	23.3
3	27.5	25.8	25.0	25.0	24.1	25.0
4	25.0	22.5	20.8	24.1	23.3	22.5
5	27.5	20.0	19.1	28.3	25.0	25.0
6	30.0	20.0	20.0	22.5	20.0	15.0
7	25.0	25.8	21.6	29.1	30.0	28.3

signal between 800Hz-2,000Hz
masker between 800Hz-2,000Hz

subject	male speech delay (ms)			female speech delay (ms)		
	15	30	50	15	30	50
1	48.3	42.5	32.5	46.6	37.5	30.8
2	47.5	44.1	35.0	45.0	44.1	33.3
3	47.5	40.8	32.5	45.0	37.5	30.0
4	47.5	40.0	32.5	45.0	37.5	30.0
5	47.5	42.5	32.5	47.5	45.8	40.0
6	50.0	42.5	35.0	42.5	35.0	25.0
7	47.5	45.0	35.0	46.6	45.8	39.1

signal between 800Hz-2,000Hz
masker above 2,000Hz

subject	male speech delay (ms)			female speech delay (ms)		
	15	30	50	15	30	50
1	18.3	19.1	18.3	22.5	18.3	20.0
2	22.5	19.1	20.0	23.3	22.5	22.5
3	20.0	18.3	20.8	20.0	20.0	18.3
4	20.8	20.0	17.5	22.5	18.3	20.0
5	20.0	18.3	17.5	21.6	21.6	22.5
6	22.5	22.5	20.0	20.0	15.0	15.0
7	23.3	17.5	15.0	19.1	19.1	15.0

signal above 2,000Hz
masker below 800Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	13.3	15.0	18.3	13.3	16.6	15.0
2	15.0	15.0	15.8	14.1	16.6	15.0
3	13.3	13.3	12.5	15.0	15.8	13.3
4	13.3	13.3	12.5	13.3	16.6	16.6
5	15.0	14.1	14.1	14.1	16.6	18.3
6	10.0	12.5	12.5	10.0	10.8	11.6
7	16.6	16.6	15.8	14.1	13.3	10.0

signal above 2,000Hz
masker between 800Hz-2,000Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	15.0	14.1	13.3	15.0	11.6	12.5
2	20.0	18.3	19.1	20.0	19.1	20.0
3	15.0	13.3	12.5	18.3	16.6	17.5
4	15.0	14.1	12.5	20.0	17.5	17.5
5	15.0	13.3	13.3	17.5	16.6	15.0
6	19.1	17.5	17.5	15.0	15.0	11.6
7	17.5	18.3	14.1	19.1	19.1	17.5

signal above 2,000Hz
masker above 2,000Hz

	male speech			female speech		
subject	delay (ms)			delay (ms)		
	15	30	50	15	30	50
1	45.0	40.0	33.3	44.1	37.5	33.3
2	43.3	40.0	32.5	45.0	40.0	32.5
3	45.0	40.0	32.5	45.0	35.8	31.7
4	45.0	40.0	32.5	45.0	37.5	32.5
5	45.0	40.0	35.0	45.0	47.5	32.5
6	45.0	35.0	35.8	42.5	35.0	32.5
7	43.3	40.8	31.7	44.1	40.0	32.5

Experiment 3

		subjects									
Azimuth angle	Head	1	2	3	4	5	6	7	8	9	10
0	Fixed	52.5	52.5	52.5	50.0	52.5	52.5	52.5	52.5	52.5	52.5
	rotated	52.5	50.0	55.0	50.0	52.5	50.0	52.5	52.5	55.0	50.0
2	fixed	52.5	50.0	52.5	47.5	52.5	50.0	52.5	52.5	52.5	50.0
	rotated	52.5	45.0	50.0	47.5	50.0	52.5	50.0	50.0	50.0	52.5
4	fixed	47.5	45.0	47.5	47.5	50.0	47.5	47.5	50.0	52.5	47.5
	rotated	42.5	45.0	47.5	45.0	47.5	45.0	47.5	45.0	47.5	47.5
6	fixed	42.5	42.5	42.5	45.0	47.5	42.5	45.0	45.0	45.0	45.0
	rotated	42.5	40.0	40.0	40.0	45.0	40.0	40.0	42.5	40.0	42.5
8	fixed	37.5	35.0	37.5	37.5	40.0	37.5	37.5	37.5	37.5	40.0
	rotated	52.5	52.5	52.5	52.5	50.0	52.5	52.5	50.0	52.5	52.5
10	fixed	52.5	45.0	50.0	50.0	52.5	50.0	52.5	52.5	50.0	50.0
	rotated	47.5	45.0	45.0	47.5	47.5	47.5	47.5	47.5	45.0	47.5
15	fixed	45.8	37.5	40.0	42.5	45.0	45.0	45.5	42.5	40.0	42.5
	rotated	42.5	35.0	37.5	37.5	42.5	45.0	42.5	42.5	40.0	40.0
30	fixed	37.5	37.5	40.0	35.0	40.0	42.5	40.0	40.5	37.5	40.0
	rotated	40.0	35.0	37.5	37.5	40.0	40.0	40.0	40.0	37.5	37.5
90	fixed	37.5	35.0	37.5	37.5	37.5	35.0	37.5	37.5	37.0	35.5
	rotated	35.0	35.0	37.5	35.0	37.5	35.0	35.5	37.5	37.5	35.5