NANYANG TECHNOLOGICAL UNIVERSITY

Noise Suppression Using Active Noise Control and Masking

Amit Rajora

School of Electrical and Electronic Engineering
Nanyang Technological University

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Master of Engineering

2008

Summary

An innovative solution to make the public/workplace environment more pleasant by combining techniques of active noise control (ANC) and psychoacoustics is discussed in this thesis. A version of the psychoacoustic model-I, suitably modified for operational simplicity, has been deployed by the Integrated ANC-Masking (IANCM) system. The limitations of existing ANC systems, which result in an annoying residual noise, are overcome by applying a soothing and pleasant masker at the exact levels of loudness. A technique for selection of the right masker from a pool of desirable maskers has also been suggested, in addition to giving the end users a choice of selecting their own favorite maskers. Various adaptive algorithms have been compared under several critical factors prior to deciding their suitability for the current task. The system employs a pleasant sound for ascertaining the offline estimate of the secondary path, thereby providing the end-user a completely pleasant experience. The effects of varying the masker levels on the quality of output have been analyzed and compared with a conventional ANC system. Continual tracking of the residual noise enables the system to find the right levels of masker amplitudes. This enables the system to cater for a wide range of sources of unpleasant noise. To further improve the performance of the system, preprocessing the masker for dynamic range reduction has also been employed. In addition to this, statistical methods for clustering have been employed to provide a smooth transition between maskers, thereby avoiding a monotonous experience.

The proposed system is also able to compensate for any changes taking place in the secondary path by incorporating online modeling of secondary path. The system has

been validated for stability and performance for numerous combinations from a large pool of maskers and noise sources. Several practical situations can benefit from the proposed system and a few of them have been suggested in this thesis. Lastly, the thesis proposes several areas of research for further improvement of the IANCM system.

Acknowledgements

I would like to express my gratitude to several people who have contributed in many ways to support and encourage me throughout the period of my candidature. I wish to extend my gratitude to the following people:

(a) My sincere thanks to my project supervisor, Associate Professor Dr. Gan Woon-Seng, for his continual support, advice, guidance, and encouragement. He has sown the seeds of the many innovative ideas behind the development of Integrated Active Noise Control and Masking System. He was instrumental in providing a very friendly and stimulating working environment, which permitted me to work on this thesis.

(b) I was fortunate to find the company of sincere friends in the DSP Lab like Mr Joseph Tan Ee Leng, Mr Dahyanto Harliono, Mr Vincent Wang Liang, Mr Nguyen Dinh Quy, Mr Reuben Johannes and Mr Nay Oo. Their presence made the place interesting for doing research and studying. Special thanks also go out to Iti Chaturvedi, Piyush Mundra, Akshay Mutha and Rajkumar for their immeasurable help.

(c) I am grateful to both technicians in DSP Lab, Mr Yeo Sung Kheng and Mr Ong Say-Cheng for their support and help in logistic and administrative matters.

(d) I am greatly indebted to my wife, Sneh for her belief and never ending support for my efforts. She has shown a very deep understanding and unflinching support for my study in Singapore. I would like to dedicate this thesis to my son Arjun, whose smile always made my determination stronger.

(e) Greatest of all, I would like to thank GOD for giving me the strength to embark on this endeavor.

Table of Contents

List of Figures

Chapter 1

Introduction

Psychoacoustics, the science of studying the interaction of sound and human auditory system has been around for more than five decades and has been used in applications such as the vastly popular MP3 digital audio coding, digital audio watermarking, hearing aids, suppression of audio artifacts and loudspeaker design among other applications. However, psychoacoustics has primarily made its mark in digital audio compression. This thesis aims to break new grounds in noise suppression using psychoacoustic parameters in conjunction with active noise control (ANC) systems. ANC systems [1] have been refined for more than seventy years, after the basic idea was invented by Paul Leug. There are two ways to achieve noise reduction. The first approach is using passive systems, which are based on the absorption and/or reflection properties of materials, and perform well for frequencies higher than 1 kHz. These techniques are somewhat expensive, obstructive and not effective at low frequencies. The second method is ANC which shows good performance for lower frequencies (less than 500 Hz) and can be achieved at a low cost. Novel techniques utilizing the combined benefits of ANC and psychoacoustics have been introduced in this thesis.

1.1 Motivation

A common problem encountered in public or even private spaces is the incessant noise in

one's surroundings. It is common to encounter construction equipment or engine noise nowadays in urban or even rural settings. The detrimental effects of persistent noise on human health [2] are well known and give us an impetus to find better ways for suppressing noise and making our surroundings pleasant. There is abundant literature available on the topic of active noise control (ANC) since the idea is not new and was patented by Leug in 1936[3]. A basic block diagram of the system [4] using an adaptive filter to control the secondary source output is illustrated in Figure 1.1. A basic narrowband feedforward ANC makes use of the fact that many noises, such as those generated by engines, compressors, motors, fans and propellers are periodic in nature and can use a sensor other than a microphone for providing the reference input.

In practical applications, periodic noise contains harmonics in addition to its fundamental. Further the effectiveness of ANC systems above 500 Hz falls off significantly [5]. Figure 1.2 illustrates the effect of an ANC simulation on a noise source.

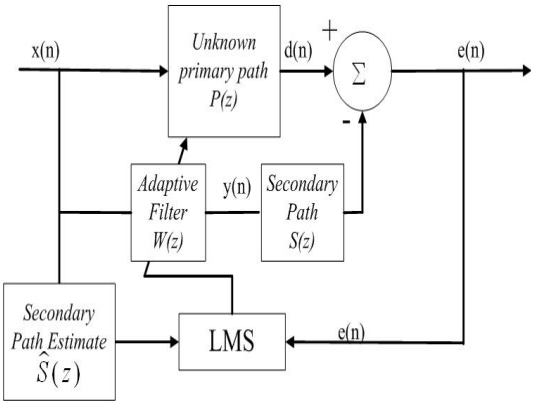

Figure 1.1: ANC using an adaptive filter

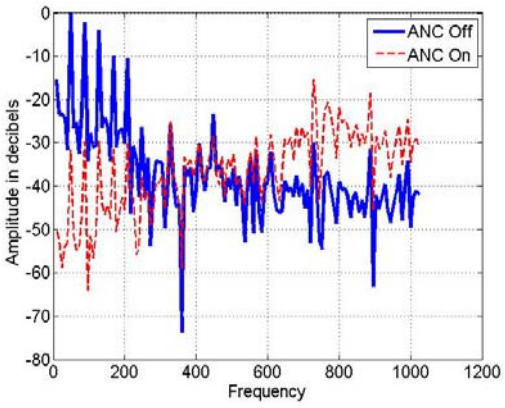

Figure 1.2: Engine noise ANC simulation

It is clearly evident that even in a controlled simulation, ANC is not fully effective.

There is a marked deterioration in the performance beyond 500 Hz. This is because of the accumulation in phase shift with increase in frequency. The errors at higher frequencies cause constructive instead of destructive interference. In addition, the quantization and round off errors due to the limitations of hardware also contribute towards the errors. This limitation provides the motivation for finding an alternative solution. As the current ANC systems are not able to completely decimate the unwanted noise, the residual noise is sought to be masked by a pleasant sound in the least intrusive manner possible.

1.2 Objective

The objective of the research is to integrate the best practical ANC techniques with psychoacoustics and process the residual noise so as to make it pleasant to the human ear. The psycho-acoustical models acceptable under ISO/IEC 11172-3 are taken as guidelines for the sensitivity and discriminatory characteristics of the human ear.

1.3 Major contribution of the thesis

The thesis advances the relatively unexplored field of combining psychoacoustic techniques with ANC and utilizing them for providing a pleasant auditory experience to the user. Innovative techniques for selecting the most suitable masker from an available pool of masker's database and a smooth transition from one masker to another are proposed in this thesis. A novel technique of using pleasant sounds for training the system has also been implemented in the proposed ANC-masking system. The proposed

system demonstrates its robustness by its ability to handle abrupt changes of the unwanted noise by employing well known automatic gain control (AGC) techniques.

1.4 Literature Study and thesis outline

To understand the behavior of sound waves, some acoustics theory is needed. Sound is a pressure wave traveling through a medium, for example air. Sound is usually generated by the vibration of an object, like speakers or human vocal cords. The vibrating object radiates a pressure wave into the adjoining medium. Sound is defined by many factors and the most important factors are outlined below.

Frequency is the number of cycles that the periodic signal completes in one second. The unit of the frequency is Hz (Hertz). Frequency characterizes the tone or pitch of a sound, for example a bass cord sounds at a lower frequency than a violin cord. The wavelength is inversely related to frequency. Thus, higher frequencies have lower wavelengths. The amplitude is the maximum amount of pressure at any point in the sonic wave. Finally, the sound-propagation velocity depends on the type, temperature and pressure of the medium through which it propagates. The propagation speed u can be calculated by using equation (1.1), where γ the adiabatic exponent (Approximate value 1.4), R the gas constant (8.3144 Joules per Kelvin per Mole), T the absolute temperature and M the molar mass:

$$u = \gamma RTM \quad . \tag{1.1}$$

The propagation velocity of sound waves in air is about $344ms^{-1}$ at 20 degree centigrade. The absolute value of the sound intensity (I) can be described in Watts per square meter or W/m^2. Bel is the logarithm of the ratio of two values which can be used to compare the power of two noise sources. The Bel unit is considerably large for practical usage; hence decibel is used more frequently. Decibel provides a relative measure of sound intensity. The unit is based on powers of 10 to give a manageable range of numbers to encompass the wide range of the human hearing response, from the minimum threshold of hearing (20 micro Pascal) to the threshold of pain at some ten million times that intensity. In view of this wide range, logarithmic ratios are preferred over linear relationships.

The sound pressure level (SPL) describes the logarithmic ratio of actual sound pressure to the reference pressure of 20 micro Pascal. The threshold of hearing lies at 10^{-12} $W/m^{2.}$ Just noticeable difference (JND) is the minimum difference in the sound pressure intensity to be noticed by an average human ear. Care should be exercised in that this term is not exactly quantifiable and differs across frequencies and people. Sound absorption coefficient describes the efficiency of the material or the surface to absorb the sound. The ratio of the absorbed sound energy to the incident energy is the sound absorption coefficient. Noise reduction coefficient (NRC) is the arithmetic average of the sound absorption coefficients at 250, 500, 1000, and 2000 Hz. This average is rounded to the multiples of 0.05.

Psychoacoustics is the science that seeks to explain the acoustic signal processing by the human ear and brain. The highly nonlinear and complex nature of the human auditory system has been studied extensively over the last 50 years. Its use in data

compression algorithms and reduction of audio artifacts in consumer devices are the most popular applications. It has also been used for many other signal processing applications such as audio watermarking [6], transmission of wireless signals/communications [7], and speech enhancement [8]. This thesis proposes another novel application of psychoacoustics; residual noise masking for ANC systems.

This thesis aims to combine the advantages of psychoacoustics with practical ANC systems. Psychoacoustic masking of the residual noise, offline and online modeling of secondary path for Integrated ANC-Masking (IANCM) systems and pre-processing of masker by dynamic range reduction have been developed. The outline of the thesis is as follows.

Conventional ANC systems and the most popular algorithms used for their implementation are explained and analyzed in Chapter 2. The need to reduce the residual noise is recognized. Chapter 3 introduces the concepts of masking and psychoacoustics. Industry prevalent noise measurement and weighting standards are analyzed and judged for their suitability in the proposed IANCM system. Important factors for gaining a pleasant auditory experience are studied and the psychoacoustical model-I [9] which has been deployed for the system has been discussed in some detail. Chapter 4 presents the proposed IANCM system. Methods for selection of the most suitable masker from the available corpus of pleasant sounds have been proposed. The effects of inclusion of masker audio into ANC systems are discussed. Successful implementation of using pleasant signals for training the system has been highlighted. The robustness of the system has been demonstrated with the widely used least mean square (LMS) and normalized least mean square (NLMS) algorithms. The system has been enhanced with

automatic gain control (AGC) to cater for changes in the maskee. This additional feature has increased the system capabilities to handle time varying maskees too. Practical utility of the system to handle everyday situations, such as snore noise, household appliances and air conditioner noise have been demonstrated via computer simulations. Chapter 5 incorporates some improvements in the design of the IANCM system. These improvements include the reduction in dynamic range to minimize masker volume, online modeling to cater for changes in secondary path and transitioning of maskers to reduce listener fatigue. The possibility of using reverberation for minimizing masker loudness has also been explored. Chapter 6 lists the conclusions drawn from the research and outlines the scope for further research.

Chapter 2

Active Noise Control

Active noise control (ANC) systems work on the principle of superposition. ANC is achieved by introducing an appropriate anti-noise through secondary sources using signal processing algorithms. Feedback or feedforward techniques are used to get an estimate of the primary noise from its point of source and an anti-noise signal is generated before it reaches the destination, where a quiet zone is desired. Signal estimation techniques can be used to produce an anti-noise signal, if the primary noise were stationary or a periodic signal. However, most acoustical noises like automobile noise are non-stationary, thereby requiring the employment of an adaptive filter. ANC systems are currently being employed for automotive, mass transportation, military, industrial and domestic appliances [4]. An ANC system consists of the physical system to be controlled, the sensors that sense the primary disturbance and monitor the residual error, actuators that physically alter the plant response, and a controller to process the sensor signals and alter the input to actuators. In this chapter, we will introduce different types of controllers used for these systems and their commonly used algorithms.

2.1 Types of ANC

ANC uses an electrically generated acoustic sound field to reduce the unwanted noise. If the control is dynamic and adapts to the changes in the unwanted sound field, then ANC

must be an adaptive system. Such a system is ideally suited to cancel noise in the range of 0-500 Hz [4] – [5] as it faces several technical difficulties at frequencies higher than 500 Hz. First of all, for higher frequencies, the spacing requirements for sensors and speakers become prohibitive due to the smaller wavelength. Secondly, the real time processing requirements make ANC techniques unmanageable. Thirdly, passive treatments become more effective at frequencies higher than 500 Hz [5] and often provide an adequate solution without the need for active control. ANC can broadly be classified into:

(i) Broadband Feedforward ANC.

(ii) Narrowband Feedforward ANC.

(iii) Broadband Feedback ANC.

(iv) Narrowband Feedback ANC.

2.1.1 Broadband feedforward ANC:

These are systems that have a single secondary source, a single reference microphone and a single error microphone. In this system (shown in Figure 2.1), the reference microphone samples the incoming signal for processing by the electronic controller, which also drives the control source i.e. the canceling loudspeaker.

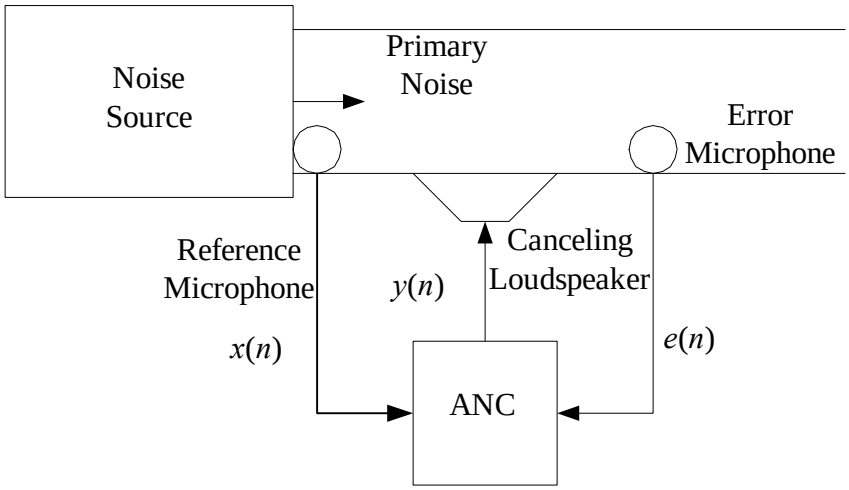

Figure 2.1: Broadband feedforward ANC

The effectiveness of the cancellation is measured by an error sensor located at a specific position. The system exhibits responses to a measured disturbance in a pre-defined way. The objective of the ANC controller is to minimize the residual error signal $e(n)$. For effective implementation of a feedforward controller, the disturbance must be measurable and the effect of the disturbance to the output of the system must be known. In addition, the time it takes for the disturbance to affect the output must be longer than the time it takes the feedforward controller to affect the output. If these conditions are met, feedforward ANC can turn out to be extremely effective. This basic broadband ANC system can be described as an adaptive system identification framework as shown in Figure 2.2.

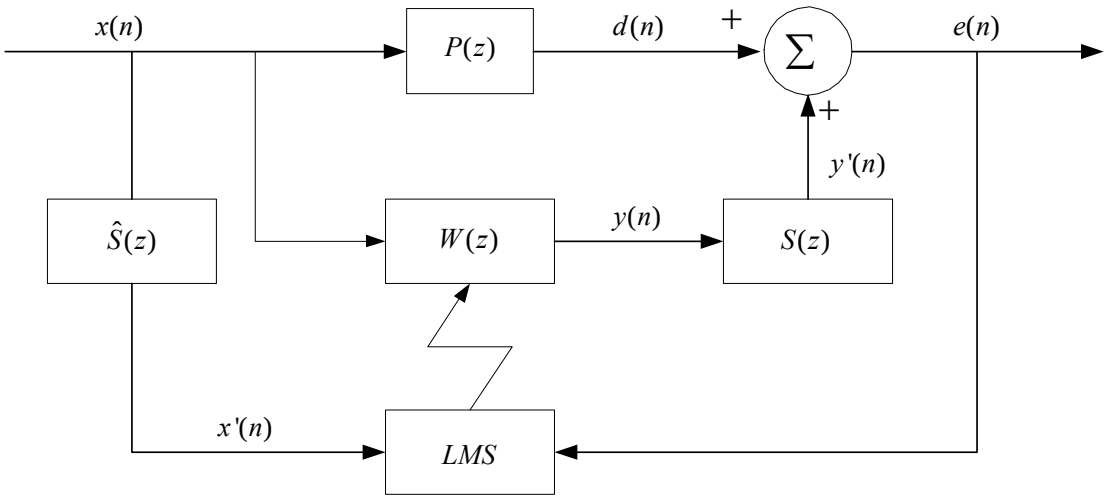

Figure 2.2: Adaptive feedforward ANC

Essentially, an adaptive filter $W(z)$ is used to estimate an unknown plant $P(z)$ which consists of the acoustic response from the reference sensor to the error sensor. The objective of the adaptive filter $W(z)$ is to minimize the residual error signal $e(n)$. It is necessary to compensate for the secondary path transfer function $S(z)$ from the output of the adaptive filter till the point where the error signal is measured. The introduction of the secondary path transfer function in a system using the conventional LMS algorithm prevents the convergence of the adaptive filter [5]. This can be solved by placing an estimated secondary path filter in the reference signal path prior to the weight update of the LMS equation, as shown in Figure 2.2. This is known as the filtered-X LMS algorithm [4]. In practical applications, the secondary path transfer function $S(z)$ is unknown and must be estimated by an additional filter $\hat{S}(z)$. Therefore,

$$x'(n) = \hat{s}(n) * x(n), \tag{2.1}$$

where $\hat{s}(n)$ is the impulse response of $\hat{S}(z)$.

2.1.2 Narrowband feedforward ANC:

Many noise sources such as engines, compressors and motors are periodic in nature. For such cases, direct observation of the mechanical motion using an appropriate sensor can be used to provide an electrical reference signal which consists of the primary frequency and all the harmonics of the generated noise. Thus, the problem of acoustic feedback from the antinoise output to the reference microphone is removed. The basic block diagram of the narrowband feedforward ANC system [10] is shown in Figure 2.3. The periodicity of the noise means that the causality constraint is not very stringent [5].

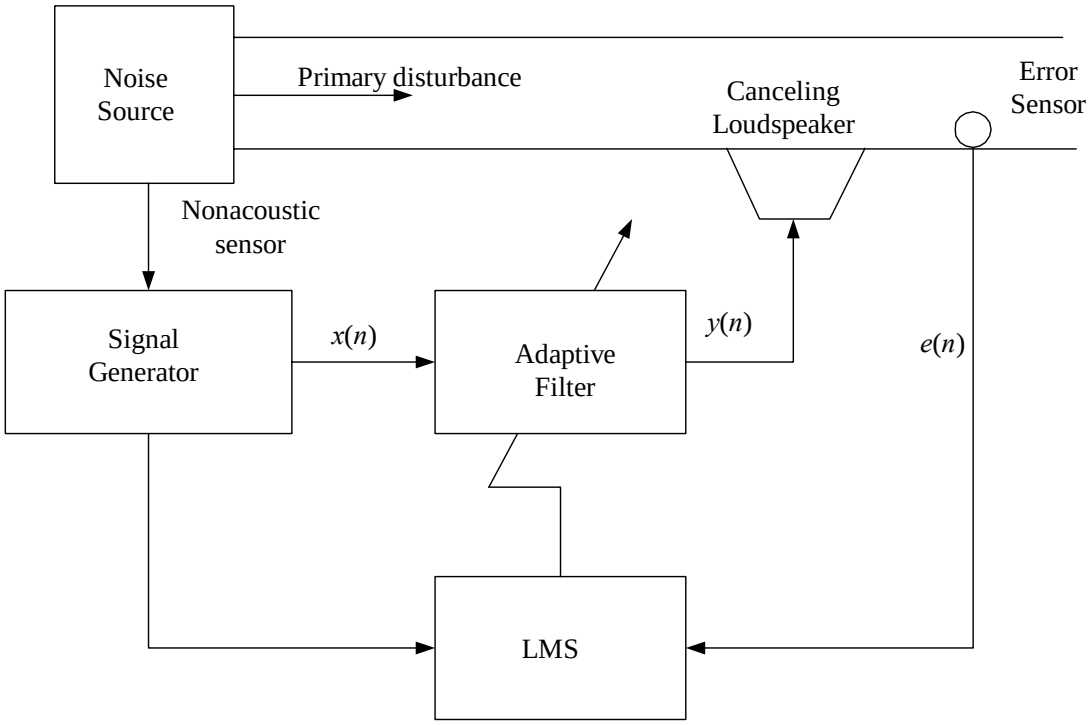

Figure 2.3: Narrowband feedforward ANC

2.1.3 Broadband Feedback ANC:

In practical applications of ANC, it is not possible to accurately predict all the factors and their effect on the output of a system. While feedforward systems require knowledge of the incoming disturbance to generate an anti-noise signal, feedback systems attenuate the disturbance as it occurs. The basic block diagram of a classical feedback ANC system [4] is shown in Figure 2.4. In feedback control, the variable required to be controlled is measured. This measurement is compared with a given datum. The controller takes this error and manipulates the necessary variables to compensate for the error. The advantage of this type of control is that it is simple to implement.

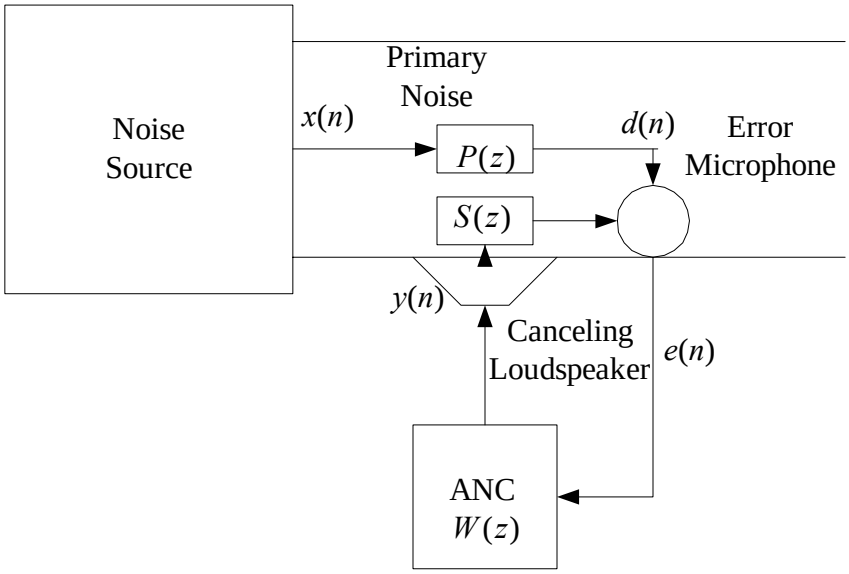

Figure 2.4: Broadband Feedback ANC

The performance of ANC can be determined by the frequency-domain analysis of the residual signal $e(n)$.The autopower spectrum is given by [4]

$$S_{ee}(\omega) = [1 - C_{dx}(\omega)]S_{dd}(\omega), \tag{2.2}$$

where $C_{dx}(\omega)$ is the magnitude squared coherence function between two wide sense stationary random processes $d(n)$ and $x(n)$, and $S_{dd}(\omega)$ is the autopower spectrum of $d(n)$. In order to have a low residual error, we need high coherence at frequencies in the primary disturbance. To circumvent this problem, feedback systems work on the residual effects of a disturbance after its occurrence. They process the error signal picked up by the error sensor. To minimize acoustic delays and increase feedback control performance, control source and error sensor should be as close together as possible. In Figure 2.4, if $W(z)$ is the transfer function of the controller, $D(z)$ is the transfer function of the primary disturbance and $S(z)$ is the transfer function of the secondary path, then the z-transform of the error signal can be expressed [11] as

$$E(z) = D(z) - S(z)W(z)E(z). \tag{2.3}$$

Thus, we get

$$E(z) = \frac{D(z)}{1 + S(z)W(z)}. \tag{2.4}$$

Therefore, the closed loop transfer function $H(z)$ from the primary noise to the error signal is given as [11]

$$H(z) = \frac{E(z)}{D(z)} = \frac{1}{1 + S(z)W(z)}. \tag{2.5}$$

A feedback control system requires no knowledge of the source or nature of the disturbances. As long as an adjustment is being made in the correct direction, the control system removes the effect of an external disturbance. The disadvantage is that the disturbance has to pass through the system before it is eliminated. Hence, the stability of this system is lower than the feedforward system. An adaptive feedback ANC system basically estimates the primary noise and uses it as a reference signal. In figure 2.4, the reference signals $x(n)$ is given as [5]

$$x(n) \simeq e(n) + \sum_{m=0}^{M-1} \hat{s}_m y(n-m), \tag{2.6}$$

where $e(n)$ is the error signal, $y(n)$ is the control signal, M is the order of the finite impulse response(FIR) filter used to estimate the secondary path, and $\hat{s}_m$ (for m = 0,1,….M-1) represents the coefficients of the FIR filter.

2.1.4 Narrowband Feedback ANC:

In this type of control system (shown in Figure 2.5), a feedback sensor generates a signal in response to the error signal which is fed to a narrow-band feedback controller. The controller has a frequency detector, a bandpass filter controlled by the frequency detector and a feedback control generator [11]. The frequency detector determines the peak frequency of the excitations. A bandpass filter passes only the peak frequency signals in a narrow frequency band from the feedback sensor. The feedback control generator receives the filtered signal from the bandpass filter and produces a control output used by

an actuator to attenuate the excitations. Several bandpass filters can be used in parallel to attenuate several frequency bands if the excitation source has multiple frequency peaks. This configuration allows the canceling signals to move across the frequency spectrum in response to the excitations.

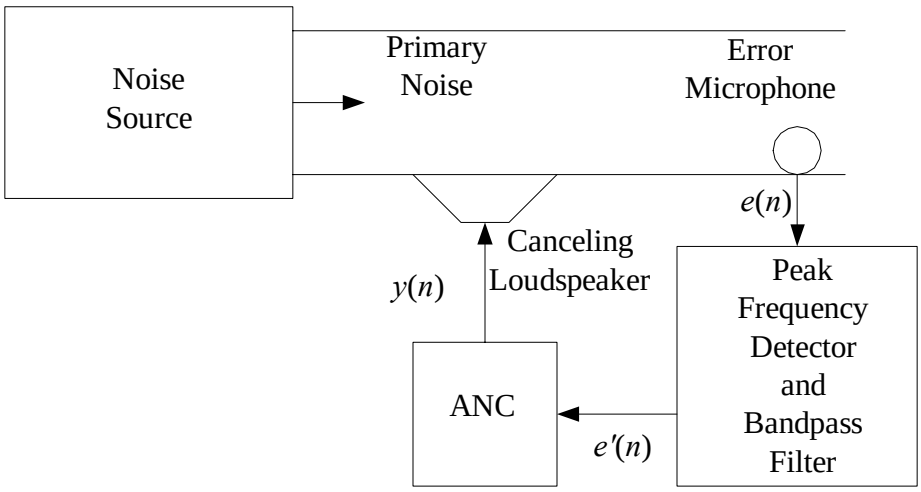

Figure 2.5: Narrowband Feedback ANC

2.2 ANC Algorithms

Many different adaptive algorithms have been introduced over the years for the implementation of ANC systems. Due to restrictions in the scope of this work, we will only introduce two most commonly used adaptive algorithms; the least mean square and the normalized least mean square algorithms.

2.2.1 The Least Mean Square (LMS) algorithm:

The LMS algorithm was introduced by Widrow and Hoff in 1959 [12]. This algorithm uses the estimates of the required corrections from the reference signal and the error signal. The LMS algorithm incorporates an iterative procedure that makes successive corrections to the weight vector in the direction of the negative of the gradient vector which eventually leads to the minimum mean square error.

The conventional LMS algorithm is a stochastic implementation of the steepest descent algorithm. It simply replaces the cost function by its instantaneous coarse estimate [13]. In the method of steepest descent, which is a deterministic method to find the minimum of the error performance surface using an adaptive feedback approach [13], the biggest problem is the computation involved in computing the covariance matrices in real time. The LMS algorithm, on the other hand simplifies computation by using the instantaneous values of covariance matrices instead of their actual values, i.e. computation of correlation function and matrix inversion are not required. However, the algorithm requires *apriori* information for the reference signal [13]. We can see from Figure 2.6 that the error signal at the n^{th} time instant is given by a difference of the desired signal $d(n)$ and the estimated signal at n^{th} time instant. The estimated signal is the data signal $x(n)$**,** passed through an arbitrary set of coefficients represented by $w(n)$. In order to obtain the mean square error (MSE) of the LMS algorithm, we compute the expectation of the squares of the instantaneous values given by the expression [13],

$$E\{(d(n)-w^T(n)x(n))^2\}.$$

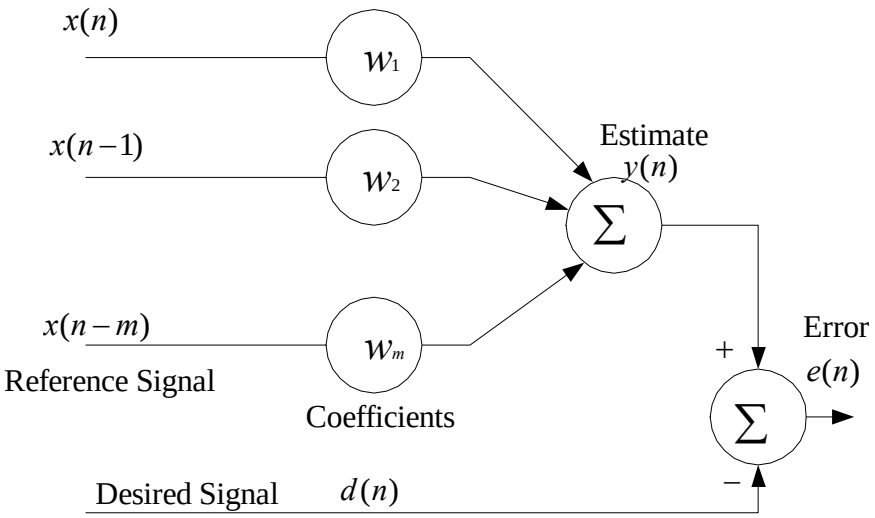

Figure 2.6: Comparison of estimate with desired signal

In order to minimize the MSE, we need to minimize the expression obtained above. Solving for its magnitude as zero, we get the required coefficient vector as [13]

$$w^* = \frac{E\{x(n)d(n)\}}{E\{x(n)x(n)^T\}} \quad , \tag{2.7}$$

This is the well known Wiener solution. This results in the steepest descent algorithm, with coefficient updates given as [13]

$$\begin{aligned} w(n+1) &= w(n) + 2\mu[E\{x(n)d(n)\} - w(n)E\{x(n)x(n)^T\}] \\ &= (I - 2\mu E\{x(n)x(n)^T\})w(n) + 2\mu E\{x(n)d(n)\} \end{aligned} \quad , \tag{2.8}$$

where μ is the step size. However, to avoid the computation of the correlation function for matrix inversions, we use the instantaneous values of covariance matrices instead and thus get the coefficient update using the following equation [13]

$$w(n+1) = w(n) + 2\mu x(n)e(n) \ , \tag{2.9}$$

This method finds a minimum, if it exists, by adjusting the filter coefficients to minimize the error. The iterative correction of the coefficient vector eventually leads to the solution of the minimum value of the mean squared error. The LMS algorithm is first initiated with some arbitrary value for the weight vector and converges to a minimum value when the below condition is satisfied [13]

$$0 < \mu < \frac{1}{\lambda_{\max}} \ , \tag{2.10}$$

where $\lambda_{\max}$ is the largest eigenvalue of the autocorrelation matrix. When the eigenvalues of the correlation matrix are widespread, convergence is slow. The eigenvalue spread of the correlation matrix is estimated by computing the ratio of the largest eigenvalue ($\lambda_{\max}$) to the smallest eigenvalue ($\lambda_{\min}$) of the matrix. If μ is chosen to be very small, then the algorithm converges very slowly. A large value of μ may lead to a faster convergence, but may be less stable around the minimum value. The fastest convergence is achieved with the value of [13]

$$\mu = \frac{1}{\lambda_{\min} + \lambda_{\max}} \ . \tag{2.11}$$

2.2.2 Normalized LMS (NLMS).

The NLMS algorithm may be viewed as a LMS algorithm with data-dependant adaptation step size. It improves the convergence speed in a non-static environment. Hence, we use optimal normalized step size at each step instead of a constant μ given in (2.9). The normalized step size of the NLMS algorithms is given as

$$\mu(n) = \frac{\alpha}{Lp(n)}, 0 < \alpha < 2\,, \tag{2.12}$$

where L is the order of the filter, α is the normalized step size and $p(n)$ is the estimated power of the signal at the n^{th} time instant and the weight update is given as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)\mathbf{x}(n)e(n)\,. \tag{2.13}$$

A comparison of the two algorithms in MATLAB [14] simulation is carried out to verify the above mentioned theoretical derivations. A total of 10,000 independent runs are carried out for each of the two algorithms under identical conditions when implementing ANC and the output is displayed in Figure 2.7.

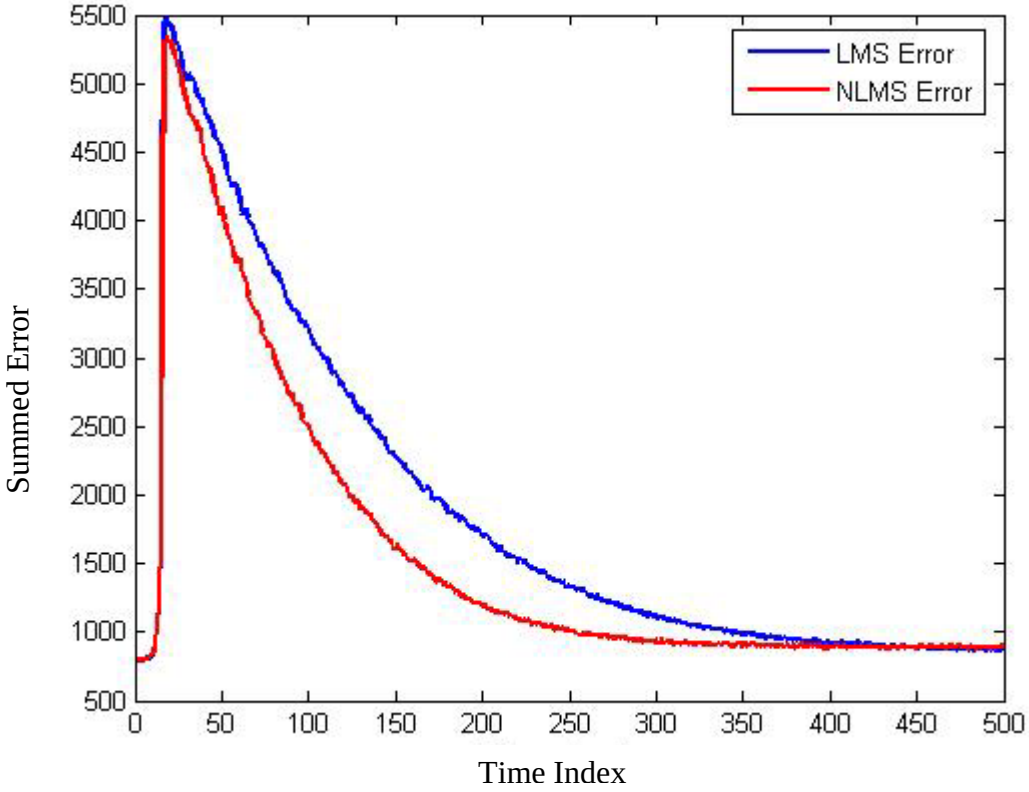

Figure 2.7: LMS-NLMS error comparison over 10000 runs

From Figure 2.7, we notice that the NLMS filter converges faster towards the steady state error as compared to LMS filter. However, in this case, these results are only valid for random Gaussian noise. We have to verify whether this observation also holds true for the type of data signal we are likely to encounter in the ANC system. So a narrowband engine noise is used and the algorithms are compared for a better performance in system identification. As can be observed from Figure 2.8, the NLMS algorithm provides a faster convergence as well as lower steady state error in system identification with a narrowband engine noise used as the input signal. This provides encouragement to carry out further simulations using the NLMS algorithm in addition to the vastly popular LMS and incorporating it in any future system.

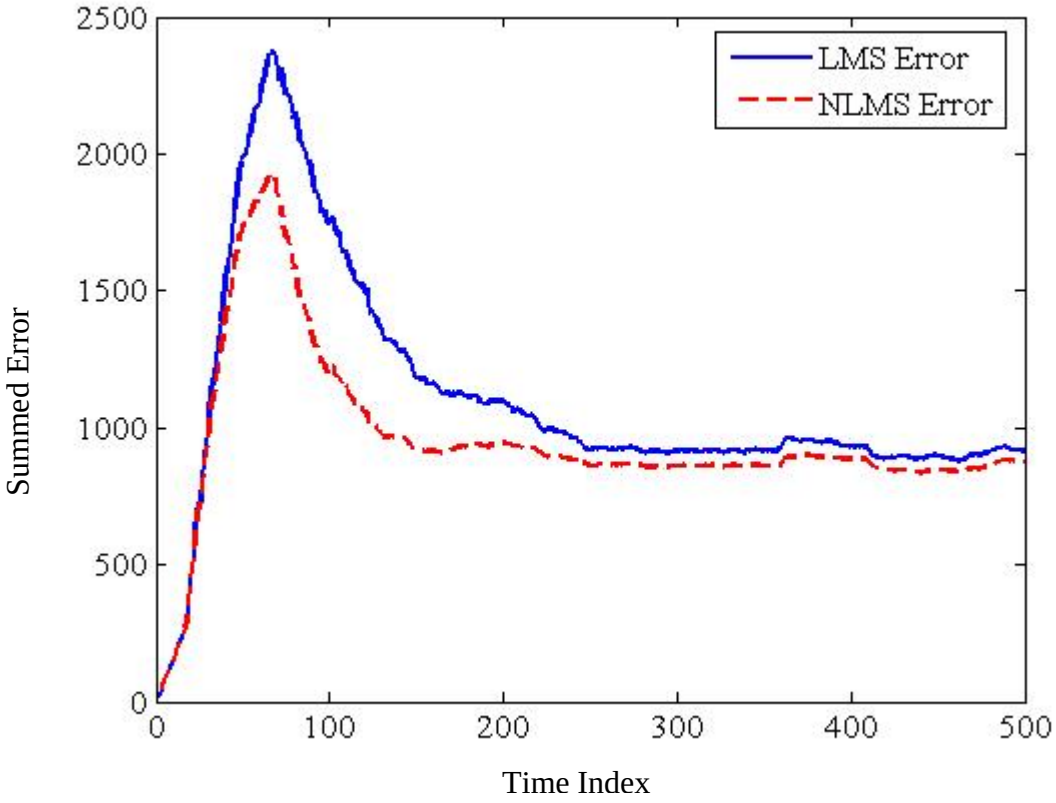

Figure 2.8: Comparison of LMS-NLMS algorithms with narrow band engine noise

2.3 ANC Applications

As mentioned earlier, ANC has widespread potential applications but so far it has been commercially popular mainly in the following fields:-

2.3.1 Propeller aircraft

Noise generated from light aircraft is primarily periodic, consisting of a fundamental and multiple harmonics of the rotor frequency. The constant-speed (and variable pitch) blades

of the engine produce relatively constant frequencies. Active noise control in aircraft cabins is well documented and its application in larger commercial twin engine aircraft is becoming increasingly popular, with approximately 250 ANC aircraft in service in early 1999[15].

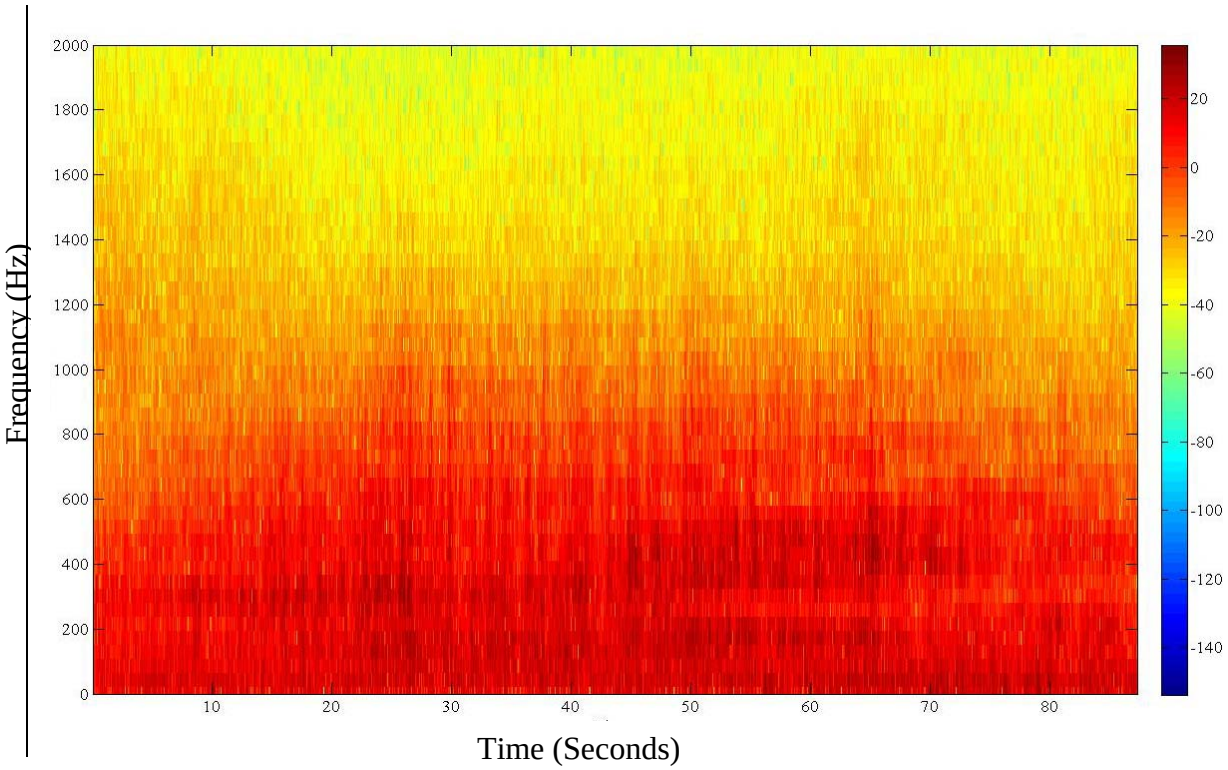

Figure 2.9: Airplane noise spectrum

Figure 2.9 displays the spectrum of the recorded noise of an airplane passing overhead at an altitude of about 8 miles. The adjacent color bar reflects the intensity levels. As can be seen, the maximum energy is concentrated in the band centered around 300 Hz. Further, more than ninety percent of the energy is available below 500 Hz. The provision of a reference source such as a tachometer signal makes feedforward ANC feasible in a small single engine aircraft. The frequency is also low enough (below 500

Hz) to avoid the complexity of the sound field encountered at higher frequencies. Further, the limitations of the active noise control system at higher frequencies are also avoided.

2.3.2 Automotive cabins

Automobile active noise control is quite popular in technical literature. Road noise inside automobiles has been described [16] as a "challenging low frequency problem". The major source of noise is tyre interaction with the road, transmitted to the enclosed cabin via the suspension. Even though the revolving tyres and the engine revolutions produce some periodic frequency components, the irregular road surface generates an enveloping broadband signal. As a reference taken from the wheel axle vibration is sufficiently far enough from the cabin for the feedforward controller to predict the sound field, road noise is a good candidate for adaptive feed forward control. In contrast with the airborne noise of a light aircraft which is dominated by strong periodic tones, the structure borne road noise of an automotive vehicle is nearly impossible to predict above 200 Hz.

Figure 2.10 gives us a very good idea of the nature of the noise inside an automobile cabin. Even though a large portion of the noise energy is in the low frequency (0-200 Hz) region, high intensity peaks can be observed even at frequencies as high as 700 Hz and 1200 Hz. Designing a system which removes the low frequency noise as well as unpredictable noise at higher frequencies is a very difficult task. Thus, controlling automobile road noise is a very challenging problem.

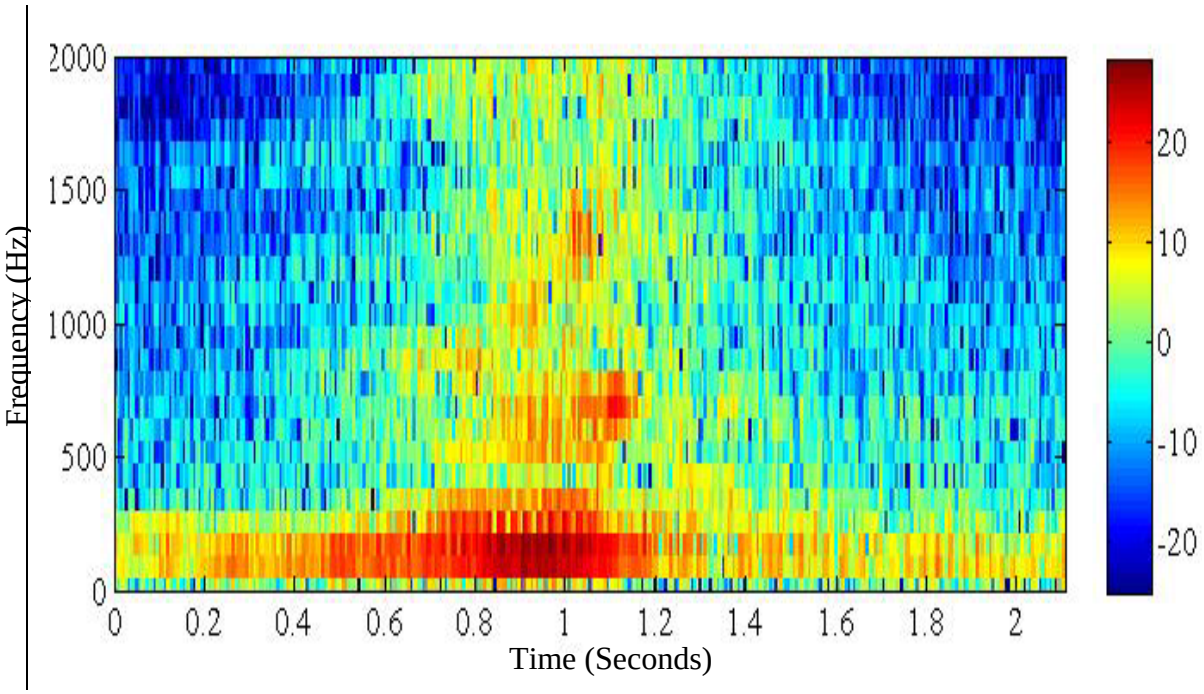

Figure 2.10: Automobile noise spectrum

2.3.3 Active headsets

Within the noisy environment of an aircraft or armored tank cabin, headsets are the most obvious solution to the noise problem as they are worn by pilots for communication and already fitted with speakers, thereby obviating the need to place canceling speakers. Active noise control headsets typically provide an additional 10 dBA attenuation over the passive reduction provided by the regular headsets [17] for use in armored vehicles.

2.4 Conclusions

This chapter has introduced conventional ANC systems as a background to the integrated ANC-masking (IANCM) system which will be highlighted in chapter 4 of this thesis. There are several other variations of these conventional ANC systems based upon

the bandwidth coverage, number of sensors, methods of combating secondary path and feedback effects, waveform synthesis methods and number of channels among several others. Two popularly implemented algorithms; LMS and NLMS have also been explained and compared. Software simulations implementing the two algorithms have demonstrated that NLMS algorithm leads to lower MSE along with faster convergence with a minimal complexity increase. Thus NLMS algorithm will find a place in the implementation of IANCM system in chapters 4 and 5.

A few practical implementations of ANC systems were also discussed and the difficulties while dealing with broadband, non-stationary or unpredictable noises for cancellation were also highlighted.

Chapter 3

Masking and Psychoacoustic Model

It is essential to have a clear understanding of the human auditory process while designing any audio system. This understanding aids us in designing practical and effective audio devices which take into consideration the capabilities and limitations of the human auditory system. In this chapter we gain an insight into the masking phenomenon and how it has been exploited by psychoacoustic models over the years to bring about revolutionary changes in audio devices. We will also carry out a study and classifications of the different masking sounds and look into the factors which make these masking sound tracks perceptually pleasant or unpleasant to the human brain.

3.1 The human auditory system

The ear is the point of interface between the external sound stimuli and its processing by the brain. The human ear, as represented in Figure 3.1, may be divided into three parts functionally [18]-[19] as:

(i) The outer ear collects the sound and sends it to the ear drum. The dimensions of the auditory canal are tuned to resonate at a frequency around 4 kHz.

(ii) Middle ear converts the air movement in ear canal to fluid movement in cochlea through the movement of three miniature bones.

(iii) Inner ear separates the sounds by frequency. Low frequency sounds travel the farthest inwards in the cochlea and high frequency travels the least distance. The hair cells at those particular points on the cochlea are excited and electrical pulses are sent to the auditory nerves for final processing by the brain. The time period of these pulses is the inverse of the frequency; hence the brain collects the information and perceives a tone.

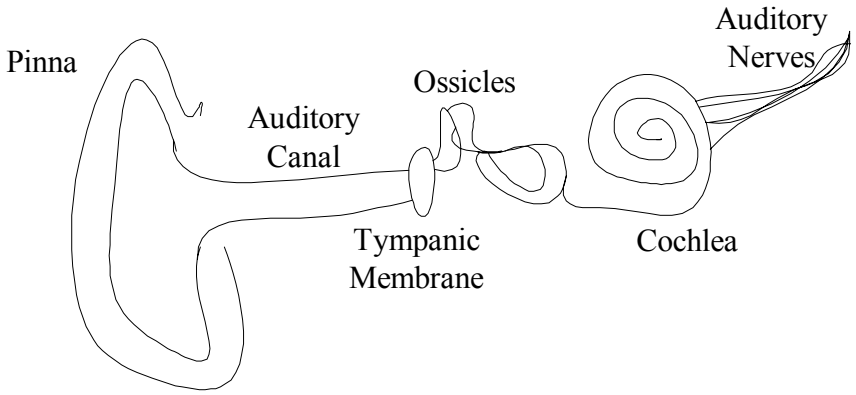

Figure 3.1: Parts of the human auditory system

3.2 The masking effect

A soft sound becoming inaudible due to the presence of a loud sound or characteristics of the central processing is termed as masking [20]. It may be classified into three types:-

(i) Frequency masking

(ii) Temporal masking

(iii) Informational masking

3.2.1 Frequency masking

Figure 3.2 shows a 'masking threshold' curve that represents the audibility threshold for signals in the presence of masking signal. Here, the signals (represented by black bars), which were otherwise well above the hearing threshold, have been masked due to the presence of a loud masker (represented by a white bar). Thus, these signals are effectively useless and can be discarded during coding.

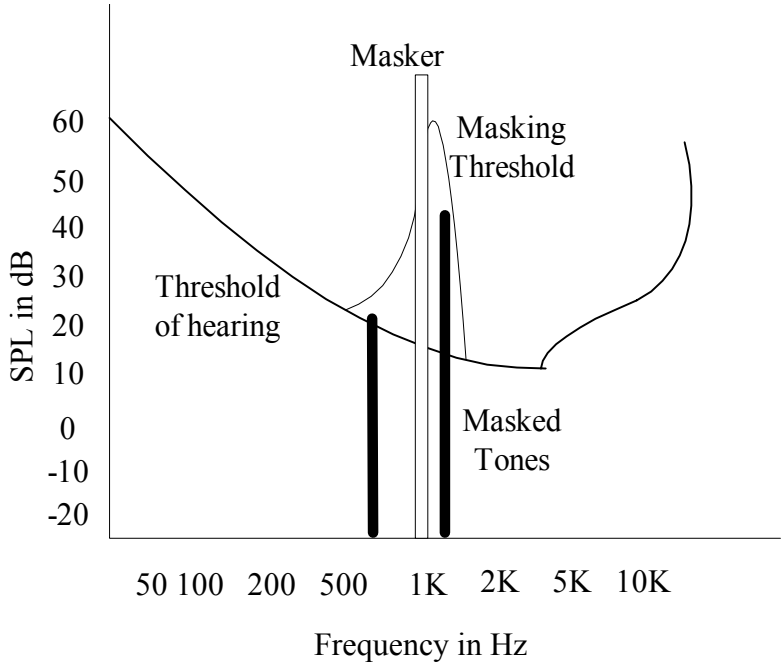

Figure 3.2: Simultaneous Masking

3.2.2 Temporal masking

The masking phenomena can even extend beyond the period of presence of the masker, as illustrated in Figure 3.3. The temporal masking phenomena can be further divided into

two types, namely, pre- and post-masking. Pre-masking is a non-causal phenomenon, such as the effect of an event is felt even before it has occurred. Pre-masking is most effective only for a few milliseconds (2-5 msec) before the occurrence of the strong masking signal. Pre-masking requires a great difference in sound level (around 30-40 dB) and is possible because of slight processing delays in the ear/brain. Post-masking occurs because a loud sound basically drowns out sounds occurring shortly afterwards. In contrast to the short time period of pre-masking, post-masking can occur up to 100 ms after the initial sound. The post-masking duration depends upon the masker level, duration and relative frequency of the masker and probe.

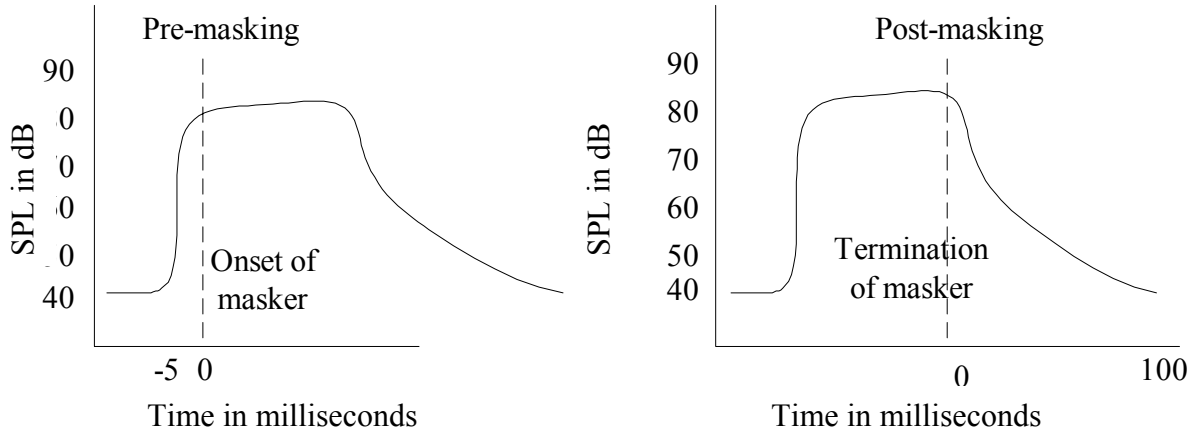

Figure 3.3: Temporal Masking

3.2.3 Informational Masking

This type of masking is lesser understood and it is the subject of considerable debate. Informational masking occurs due to the manner in which the human brain processes the

physical signals received from the auditory system. Two prominent effects caused by the central processing mechanism are the precedence effect and the Haas effect. Wallach et al. [21] presented a very brief sound over headphones with a single reflection that was made to arrive two milliseconds later. The reflection was not heard as a separate sound under these conditions. This fusion of the original sound and its reflection is known as the Haas effect. Further, it was found that by varying the interaural time difference of the reflection, there was no effect on the perceived direction of the whole sound. On the other hand, large changes in the direction were perceived when the interaural time difference of the original sound was altered. This dominance of directional information from the sound received earlier is called the precedence effect. There are many more such examples in literature where our senses lead us into believing something other than the reality. All these effects collectively give rise to what has been termed as 'Informational Masking'.

As an explanation of the masking phenomenon, Fletcher modeled the auditory system as an array of bandpass filters with continuously overlapping passbands of certain widths [22]. This model came about from the fact that in a narrow frequency range around the masker frequency, the masking threshold is flat. The critical bandwidth depends upon the frequency of the masker. Critical bandwidth according to Zwicker and Fastl is stated as

$$\frac{\Delta f}{Hz} = 25 + 75[1 + 1.4(\frac{f_c}{khz})^2]^{0.69} \quad , \tag{3.1}$$

where Δf = critical bandwidth and f_c= centre frequency. Moore and Glasberg [23] proposed a different function to calculate their version of the critical band, termed as the Equivalent Rectangular Bandwidth (ERB):

$$\frac{ERB}{Hz} = 24.7\left(\frac{4.37 f_c}{khz} + 1\right). \tag{3.2}$$

The critical bandwidths predicted by the ERB formula are much narrower at frequencies below 500 Hz compared to that in (3.1), as proposed by Zwicker and Fastl [24]. Frequency can also be mapped to a linear distance measure along the basilar membrane. The unit of length on basilar distance is equivalent to one critical bandwidth. This unit is also known as Bark. Integrating the critical bandwidth formula, we get a critical band rate expression derived by Zwicker and Fastl

$$\frac{Z}{Bark} = 13\arctan\left(\frac{0.76 f}{khz}\right) + \arctan\left[\left(\frac{f}{7.5khz}\right)^2\right]. \tag{3.3}$$

In 1988, Johnston [25] introduced the concept of perceptual entropy to define the average minimum number of bits per frequency sample needed to code a signal without introducing any perceptual difference with respect to the original signal. This was an important step towards the formation of the psychoacoustic model.

3.3 Study of masking sound and classifications

Prior to undertaking any study which involves psychoacoustic considerations, there is a need to understand the ambient noise prevalent in any particular setting. The results of various studies are available in literature. For example, ambient noise levels in school classrooms are usually around 60 dB [26]; while Hodgson et al [27] found that university classrooms have average noise levels of 44 dBA. One universally acceptable noise is the Hoth noise which can be described as acoustic random noise that has the power density spectrum incorporated in IEEE Standard 269-1992[28]. The spectrum of Hoth noise is designed to simulate typical ambient room noise over time and can be seen in Figure 3.4.

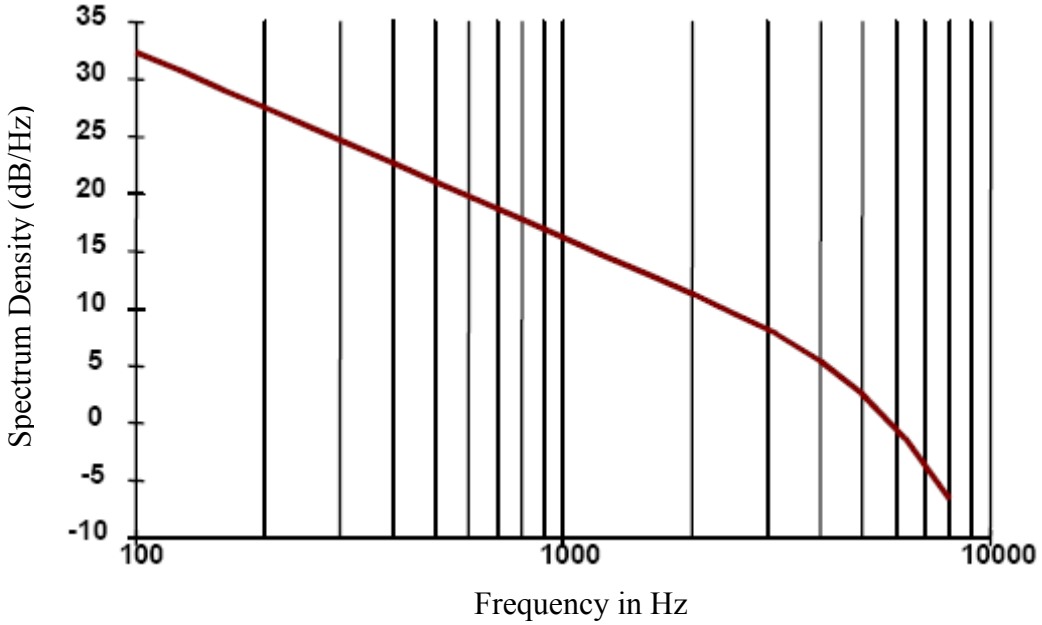

Figure 3.4 Hoth noise spectrum distribution (Extracted from [29])

The typical noise spectrum of various locations as published by Hoth [29] shows that the spectrum level decreases about 5 dB per octave. The measurements were taken in several business/residential/educational locations. It is surprising that the spectrum energy distribution does not vary much across different locations. Except for running water which has a spectrum similar to white noise, the residential noise sources also have an average spectrum similar to Hoth noise.

For the purpose of classification, different masking sounds may be segregated into various categories such as random signals, speech-like signals, white noise, pink noise, tonal signals and compound signals among others.

Truly random signals may be obtained in nature; however they may be difficult to reproduce for the purpose of experimentation. In practice, many practical noise generators produce pseudo-random signals. If the period of such signals is very long compared to the analysis period, and if the analysis period does not bear any correlation to the generator period, then for all practical considerations, these signals may be classified as random. These signals may be described by their statistical characteristics, such as the long-term power spectral density and probability density functions.

White noise has a constant spectral density per Hertz. Some noises in nature (such as running water) exhibit white noise characteristics. Pink noise has a power spectral density that decreases 3 dB per octave. This noise is also found in nature (for example, in the sounds emitted by some insects.) and is perceived to be softer than white noise. Speech-like signals include artificial conversational speech, as well as synthesized and real speech signals. The artificial voice is a continuous speech signal with a frequency range of 90 Hz to 8.9 kHz. Pauses may need to be inserted to emulate the on-off

characteristics of conversational speech. Tonal noise comprises noise which has its energy concentrated predominantly near one or a few frequencies. Also, we have the compound signals which contain various combinations of the signals above and constitute the bulk of the auditory output**.**

Qualitatively, we may also classify sounds as being continuous or intermittent, steady or fluctuating, bearing some information or devoid of any information, with pure sounds or having a complex composition.

3.4 Important factors for a pleasant experience

One important criterion for selecting any masking signal should be its pleasantness. Though quantifying the pleasantness of any signal is a very difficult task, studies have been attempted to compare the subjective preference of various signals. Bricker et al. [30] subjectively evaluated 166 different signals. Results showed that signals with the highest ratings had prominent components in the 500 Hz to 200 Hz region and had a modulation frequency of 20 Hz. According to Hunt [31], the 500-4,500 Hz region was found suitable for meeting pleasantness requirements in noisy environments. Another study concluded that the ideal frequency spectrum is within the range of 400 Hz to 4,000 Hz, with at least one component below 1,000 Hz [32]. Other studies claimed that anxiousness is induced by sounds having an energy peak around 2,500 Hz [33], and low frequencies are associated with sadness [34].

From these studies it is difficult to conclude that any particular combination is likely to be pleasant to most of the listeners. In addition to the studies on pleasantness,

studies on the unpleasantness of sounds have also been conducted. Halpern et al. carried out some experiments [35] and found that the scraping of a garden tool across a slate was adjudged to be the worst sound. Further separation of the low frequency and high frequency components revealed that the mid frequency components (3,000 to 6,000 Hz) were responsible for the unfavorable response.

Subsequent tests on monkeys (cotton-top tamarins) [36] revealed, that in contrast to humans, they exhibited no preference for any particular sound. When given a choice between the scraping sound and equal amplitude white noise, the monkeys showed equal revulsion to both sounds. The monkeys also preferred silence over soothing music. These results led to the conclusion that the unpleasantness/pleasantness was the result of an acquired behavior and not inherent. This hypothesis is in agreement with some unrelated studies [37]-[38] which bring out that the pleasantness/annoyance is dependant upon the observer's experience of the source of the sound.

3.5 The Psychoacoustic Model

The psychoacoustic models were developed as a result of the efforts to understand the behavior of the human ear in presence of complex pieces of sound. However, it is essential to have an understanding of the basic weighting curves which were in existence prior to the discussion of the psychoacoustic models.

3.5.1 A-Weighting

The A-weighting curve is an approximate inversion of the Fletcher-Munson's equal loudness curve of 40 phons [39]. The characteristic of this curve can be observed in Figure 3.5. This curve is generally used to filter the input when making noise measurements to get a better idea of how the noise would sound to a human being. The attenuation of the sound signal with an A-weighted filter seeks to replicate the human ear response. It is part of a family of curves (A, B, C, D) defined in the standards for adjusting the output of sound level meters for specific applications. This curve is commonly used for the measurement of environmental noise and industrial noise [40].

A-weighting is only valid for relatively quiet sounds and for pure tones. The B and C curves are used for louder sounds while the D curve is used in assessing loud aircraft noise (IEC 537) [41]. Figure 3.5 displays the A- weighting curve.

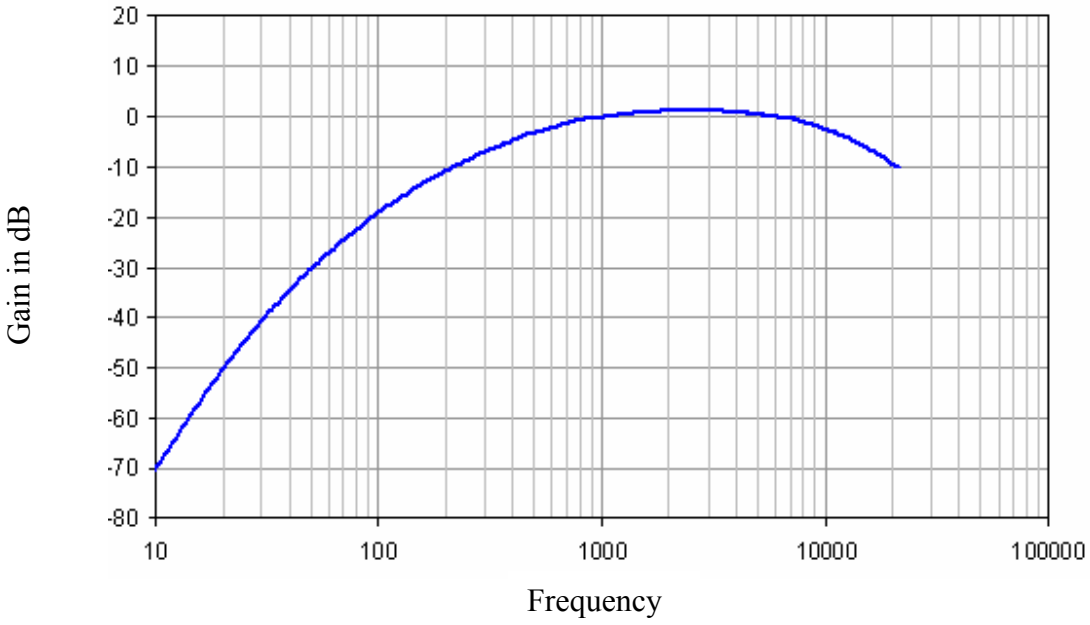

Figure 3.5: The A-weighting curve (Extracted from [40])

While the A-weighting curve has been widely adopted for environmental noise measurement, it does not really give valid results for noise because of the way in which the human ear analyzes sound. For example, humans are considerably more sensitive to noise in the region of 6 kHz than to other tones of equivalent level [18]. The transfer function of A-weighting gain curve is given as

$$G_A(s) = \frac{7.4\times10^9.s^4}{(s+129.4)^2(s+676.7)(s+4636)(s+76655)^2}. \quad (3.4)$$

The A-weighting value in decibels is given by

$$W_A = 10Log[\frac{1.56\,f^4}{(f^2+107.65^2)(f^2+737.86^2)}]+10Log[\frac{2.24\times10^{16}\,f^4}{(f^2+20.6^2)^2(f^2+12194.22^2)}], \quad (3.5)$$

where f is the frequency in Hz and W_A is the weighting to be applied in decibels.

3.5.2. ITU-R 468 Weighting

With advances in psychoacoustic measurements, the A-weighting was found to be inaccurate, and it gave way for the ITU-R 468 weighting [42]. These curves were developed from the CCIR-468-1 recommendations [43]. As can be seen from Figure 3.6, these curves differ from the A-weighting curves mainly in the 5-8 kHz region and give figures that are usually about 11dB worse than A-weighted. Thus, if two different systems use different weightings and get the same figure for noise reading, then the

system using ITU-R 468 weighting has an 11 dB lesser noise than the system using A-weighting.

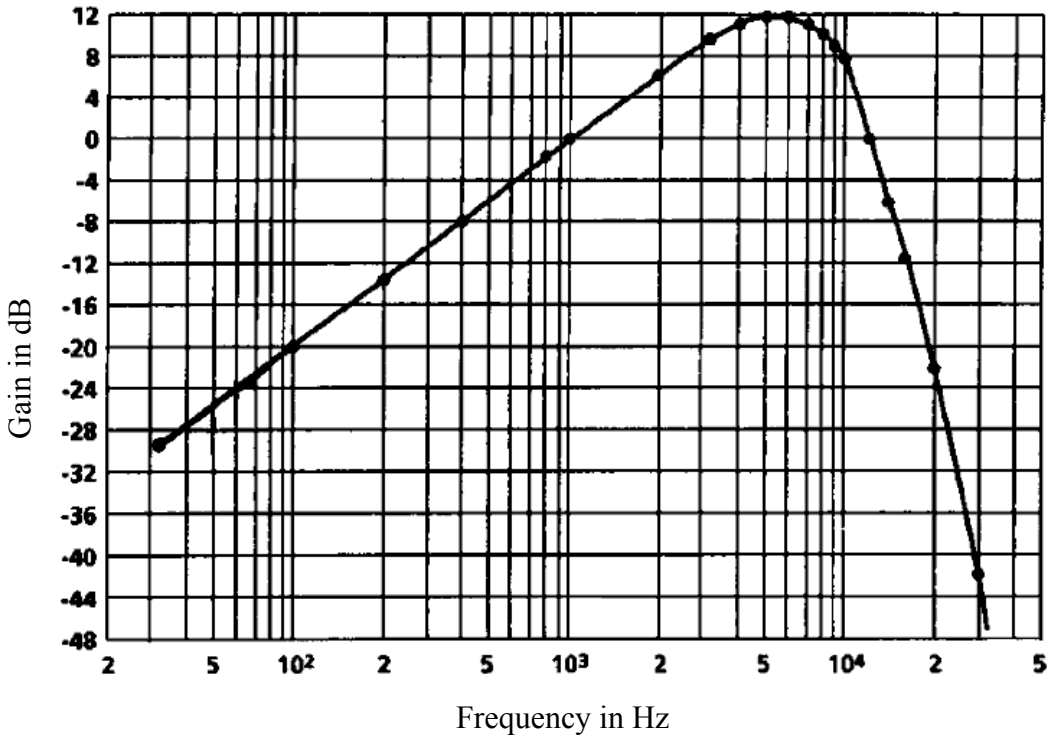

Figure 3.6 ITU-R 468 weighting curve (Extract from [42])

3.5.3. Psychoacoustic model I

The psychoacoustic model-I was developed in 1993[9] and has since been adapted successfully for lossy compression of music, digital watermarking and designing of hearing aids among other applications. Figure 3.7 illustrates the steps involved in the entire process for calculating the Global Masking Threshold (GMT) for any given signal in time domain.

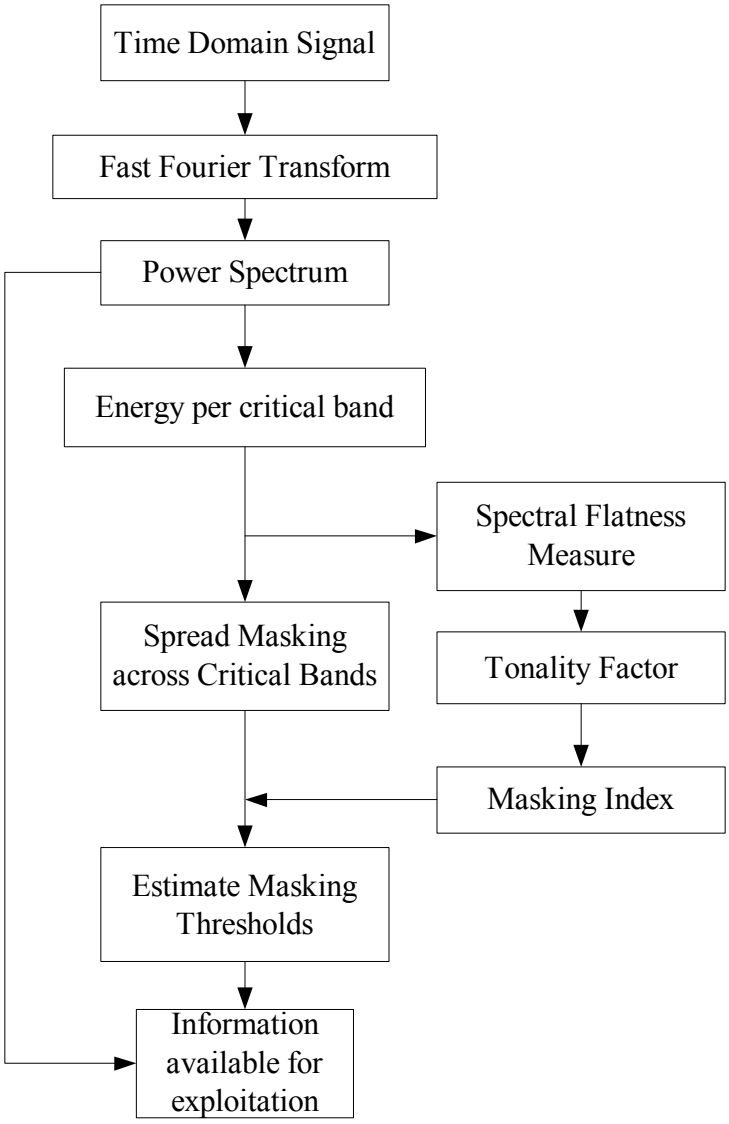

Figure 3.7: Signal processing flowchart to estimate the Masking Thresholds

Depending on the application, different layers of the coding system with increasing encoder complexity and performance can be used. An ISO MPEG Audio [44]-

[46] layer-N decoder is able to decode bit-stream data which has been encoded in Layer N and all layers below N. Layer-I contains the basic mapping of the digital audio input into 32 sub-bands, fixed segmentation to format the data into blocks, a psychoacoustic model to determine the adaptive bit allocation, and quantization. Layer II provides additional coding of bit allocation, scale factors and samples. Different framing is used in this layer. Layer III introduces increased frequency resolution based on a hybrid filter bank. It adds a different (non-uniform) quantizer, adaptive segmentation and entropy coding of the quantized values. For the sake of computational simplicity, we will consider layer I [46] for our current application. This model has been formulated by calculations based upon the data gathered from psychoacoustical experiments carried out during the basic research in this field and incorporated in ISO standards. The individual steps are explained subsequently.

Step1: Calculation of spectrum

According to the ISO/IEC 11172-3 for layer I, the input audio signal is transformed to frequency domain using a 512-point fast Fourier transform (FFT) and supports sampling frequencies of 32 kHz and 44.1 kHz for high fidelity audio signals. The signal is passed through a Hanning window prior to the transformation to avoid the edge effects. The Hanning weighting is given by

$$h(i) = 0.5 * \{1 - \cos[2\pi * i/(N-1)]\} \qquad 0 \le i \le N-1, \qquad (3.6)$$

with $X(k)$ being the sound pressure level of the spectral line in decibels with index k of the FFT for $k = 0...N/2$. Normalization to the reference level of 96 dB Sound Pressure Level (SPL) is carried out so that the maximum value corresponds to 96 dB [46].

Step2: Determination of the sound pressure level

In each subband m, the sound pressure level $L_{sb}[m]$ is calculated in the following manner. SPL of the maximum amplitude FFT spectral line which corresponds to that particular subband and the maximum scale factor for that subband are compared and the greater value is considered. The scale factor is a multiplier that sizes the samples to make full use of the range of the quantizer. Each scale factor has a 6 bit representation. The calculation of scalefactor for each subband is done every 12 subband samples. The maximum of the absolute value of these 12 samples is determined. The next largest value from a table in ISO 11172-3 is used as the scalefactor. Hence,

$$L_{sb}[m] = \max\{X(k), 20\log_{10}(scf_{\max}[m] * 32768) - 10\}. \quad (3.7)$$

The expression $scf_{\max}[m]$ represents the scale factor of subband m within the frame.. The "−10 dB" term corrects for the difference between peak and RMS level. The sound pressure level $L_{sb}[m]$ is computed for every subband.

Step3: Considering the threshold in quiet

The threshold in quiet $L_{Tq}[m]$, also called absolute threshold, is available from the ISO tables. These tables have been formulated using the results of extensive subjective tests performed under certain conditions and are part of the ISO standards [9].

Step 4: Finding tonal and non-tonal components

The tonality of a masking component has an influence on the masking threshold. For this reason, it is necessary to discriminate between tonal and non-tonal components. This step starts with the determination of local maxima, then extracts tonal components (sinusoidal components) and calculates the intensity of the non-tonal components within the bandwidth of a critical band.

The bandwidth of the critical bands varies with the center frequency with a bandwidth of about only 0.1 kHz at low frequencies and with a bandwidth of about 4 kHz at high frequencies. It is known from psychoacoustic experiments that the ear has a better frequency resolution in the lower frequency region compared to the higher frequency region. To determine if a local maximum may be a tonal component, a frequency range around the local maximum is examined. To classify spectral lines $X(k)$ as tonal or non-tonal, the following three operations are performed:

(i) A spectral line $X(k)$ is labeled as a local maximum if

$$X(k) > X(k-1) \text{ and } X(k) \geq X(k+1) \tag{3.8}$$

(ii) Listing of tonal components and calculation of the sound pressure level. A local maximum is put in the list of tonal components if

$$X(k) - X(k+j) \geq 7\,dB, \tag{3.9}$$

where the absolute value of index j may vary from 2 to 6 dependent on the location of the masker in the critical band.

(iii) Listing of non-tonal components and calculation of the power. The non-tonal (noise) components are calculated from the remaining spectral lines. To calculate the non-tonal components from these spectral lines $X(k)$, the critical bands are determined using the ISO tables. Within each critical band, the powers of the spectral lines (remaining after the elimination of tonal components) are summed to form the sound pressure level of the non-tonal component corresponding to that critical band. The spectral line with index k, which lies nearest to the geometrical mean of the critical band frequencies, is selected to be the frequency index for the new non-tonal component

Step 5: Decimation of tonal and non-tonal masking components

Decimation is a procedure that is used to reduce the number of maskers which are considered for the calculation of the global masking threshold. The decimation is performed using the following two steps.

(i) Tonal or non-tonal components are considered for the calculation of the masking threshold only if they exceed the threshold of quiet. The values for threshold of quiet are obtained from the ISO tables [9].

(ii) The components with the highest power are retained, and the smaller components are removed from the list of tonal components. For this operation, a sliding window in the critical band domain is used with a bandwidth of 0.5 Bark.

Step 6: Calculation of individual masking thresholds

Of the original N/2 frequency domain samples, only a subset of the samples is considered for the global masking threshold calculation. The samples to be used are demarcated in ISO standards [9]. Every tonal and non-tonal component is assigned the value of the index that most closely corresponds to the frequency of the original spectral line $X(k)$. The psychoacoustic model-I uses a different linear spreading function for the higher and the lower frequencies [9]. The spreading function is defined as

$$B(dz, L) = -17dz + 0.15L(dz - 1)\theta(dz - 1) \qquad \text{for } dz \geq 0 \qquad (3.10)$$

and

$$B(dz, L) = -(6 + 0.4L)|dz| - (11 - 0.4L)(|dz| - 1)\theta(|dz| - 1) \qquad \text{for } dz < 0. \qquad (3.11)$$

Here $B(dz, L)$ is the spreading function given by L, which is the sound pressure level of the masker.

In addition, the level of the masker is also considered in ascertaining the masking thresholds. The masking patterns are shifted depending upon the tonality factor, which is determined by the spectral flatness measure. This leads to a different masking index shift for tonal maskers compared to noise-like maskers. This shift is expressed as [46]:

$$\Delta_T(z) = -6.025 - 0.275z \text{ dB} \tag{3.12}$$

and

$$\Delta_N(z) = -2.025 - 0.175z \text{ dB}. \tag{3.13}$$

Here Δ_T and Δ_N denote the shifts for tonal maskers and non-tonal maskers, respectively. The frequency of the masker is represented on the Bark scale, and the masking threshold can be given as:

$$M_{T,N}(L, z_i, z_j) = B(z_i - z_j, L) + \Delta_{T,N}(z_j). \tag{3.14}$$

Here, M is the masking threshold of a masker with level L and a frequency index z_j on the Bark scale. z_i denotes the maskee frequency index on the Bark scale . The value of $\Delta_{T,N}$ depends upon the masker tonality.

Step7: Calculation of the global masking threshold

The global masking threshold at every frequency sample is derived from the upper and lower slopes of the individual masking thresholds of each of the tonal and non-tonal maskers and from the threshold in quiet. The global masking threshold is computed by

summing the powers corresponding to the individual masking thresholds and the threshold in quiet.

$$M_G(z_i) = 10\log_{10}\left[10^{\frac{M_q(z_i)}{10}} + \sum_{j=1}^{m} 10^{\frac{M_{T_j}(L_j, z_i, z_j)}{10}} + \sum_{k=1}^{n} 10^{\frac{M_{N_j}(L_k, z_i, z_k)}{10}}\right], \qquad (3.15)$$

where M_q denotes the absolute threshold in a Bark scale index z_i, M_{T_j} denotes the masking threshold of the tonal masker at the Bark scale index z_j, and M_{N_j} represents the masking threshold generated due to the presence of a non-tonal masker at the Bark scale index z_k. m is the number of tonal and n is the number of noise-like maskers. This final step allows us to compute the masking threshold for every subband on a frame-by-frame basis depending upon our requirement. Once we have this information, then we know the level below which another sound will not be perceived by humans. This information can be exploited to hide unwanted noise below these levels.

3.6 Conclusions

In this chapter, a functional description of the human auditory process is provided. Various studies exploring human preference for certain frequencies over others have thrown up some inconclusive results. Based upon this and some other studies which link the likeability of a sound with its source lead to the tentative conclusion that pleasantness of a sound is dependant upon the listener's past association with a similar sound. This will be taken as a starting point in this thesis whenever evaluating factors for the pleasantness of a masker.

The psychoacoustic models have been framed specifically for digital audio systems and are easier to use in DSPs. The psychoacoustic model takes into account the effects of sampling, quantization, frequency transform, windowing, critical bands, power spectra, assessing of the tonal/non-tonal components, masking and responses of the human ear at various frequencies. Modifying the A-weighting and ITU-R468 curves for all these will again require a rigorous subjective testing process, whereas the psycho-acoustical models have already been validated after the required testing procedures via subjective measurements and successfully used in MP3 and other audio coding and high quality lossy compression. In view of the above, it would only be suitable that the psychoacoustic models are used for further experiments as the A-weighting and ITU-R468 weighting have been found to be inaccurate over certain frequency bands and require extensive reworking for digital audio processing.

Lastly, while carrying out any design it is also necessary to characterize the type of background in which the device will be operating. Various studies have carried out a spectral analysis of the background noise. Based on these studies, Hoth noise may be considered as the prevalent background noise.

Chapter 4

Integrated ANC and Masking

In the earlier chapters, the phenomena of masking and technique of active noise control (ANC) are described. While ANC and masking both have proven their effectiveness in reducing/suppressing unwanted sounds, there are some inherent limitations with each method. In this chapter a combination of the two processes is introduced. The proposed system integrates the two processes so that individual limitations of either method are overcome in the combined system. The sequence of operation of this integrated ANC and masking (IANCM) system will be described as follows:

(I) Estimation of residual noise obtained with a conventional ANC system.

(II) Selection of the most suitable masker

(III) Estimation of secondary path

(IV) Integrated ANC-Masking operation

(V) Automatic control of gain as residual noise changes

In this operation, the first three steps are carried out offline (pre-processing) and the last two steps are executed online. In section 4.1, the process of masker selection and residual noise estimation is explained. Section 4.2 deals with secondary path effects and its estimate for the system. Integrated ANC-Masking is described in Section 4.3 and finally, Section 4.4 introduces an innovative method to deal with the changes in maskee.

4.1 Masker selection process

We have seen earlier in chapters 1 and 2, that ANC or masking alone is insufficient to mitigate the problems associated with noise. Hence, an IANCM system is being proposed. For this combination, we need to select a suitable masker. The three basic conditions for an effective masker are:

(I) The underlying noise should be transparent to an average listener.

(II) The masker should be played at the lowest volume possible.

(III) The masker should sound pleasant to the listener.

As a pre-requisite to the selection process, we need a reliable estimate of the residual noise obtained with a conventional ANC system. This is required because we would like our selected masker to effectively mask the residual noise without being too loud. This estimate can be obtained in two ways; either a separate ANC system just for estimating the reduced noise or an inaccurate residual noise estimate by an attenuated version of the original noise can be used. The first approach unnecessarily makes the system more complex and expensive and the second gives us inaccurate results. To avoid complexity, the second approach is used initially by the system and refined subsequently. The proposed system has the advantage that it can be turned into a conventional ANC system just by tweaking one of the parameters. This property will be demonstrated in Section 4.4. This characteristic of the system makes it simple, accurate and efficient.

A psychoacoustic processor [46]-[47] is used to objectively rank a collection of potential maskers. The process of ranking the maskers is explained in the next paragraph.

Although this objective measurement may fulfill the first two conditions, the third condition is based upon subjective assessment and may differ from person to person.

The psychoacoustic processor divides the incoming sound samples into frames of 512 samples each and processes the signal frame wise according to the procedure [46] in Chapter 3 and according to steps outlined in Section 3.5.3. Thus, for every frame, we get a set of 32 values denoting the SPL of each of the 32 sub bands. Similarly, 32 values of the GMT, one for each sub band, are obtained. The psychoacoustic processor first takes in the estimated noise to be masked, calculates its sound pressure level (SPL) and stores its values. Next, each potential masker is compared with the estimated noise signal. The processor calculates the GMT for all the frequency bands. For example, the SPL and GMT for a frame of recorded noise and the corresponding frame of recorded music are illustrated in Figures 4.1(a) and 4.1(b), respectively. The solid line illustrates the SPL and the dashed line represents the GMT. In order to ensure that the noise is not heard, the SPL of the noise has to be below the GMT of the music. Thus, the solid line of Figure 4.1(a) should fall below the dashed line of Figure 4.1(b). This ensures that the noise is fully masked by the music in this particular frame.

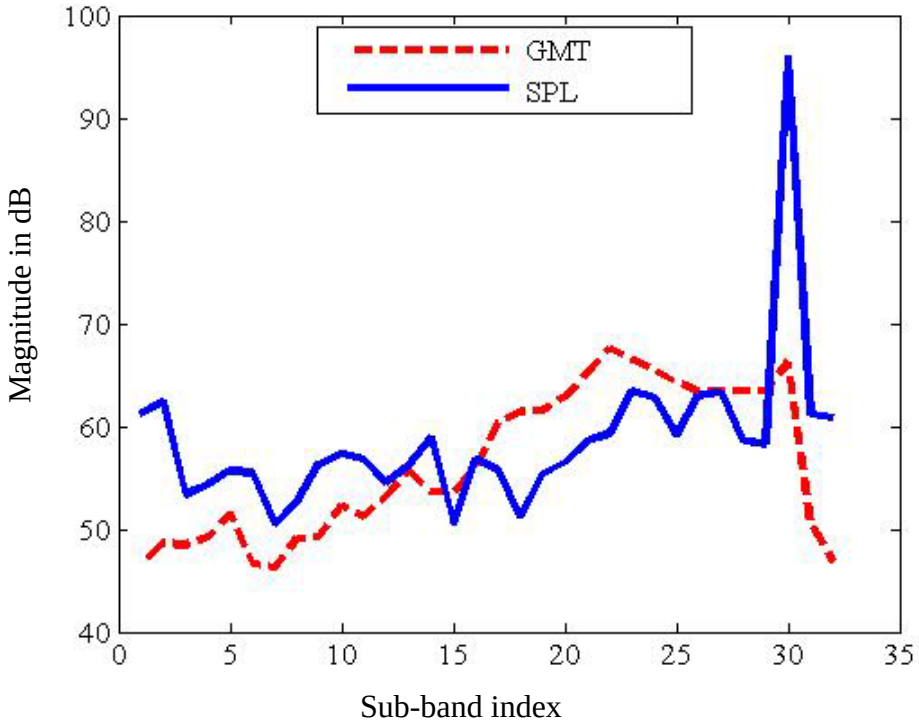

Figure 4.1(a): Noise SPL and GMT

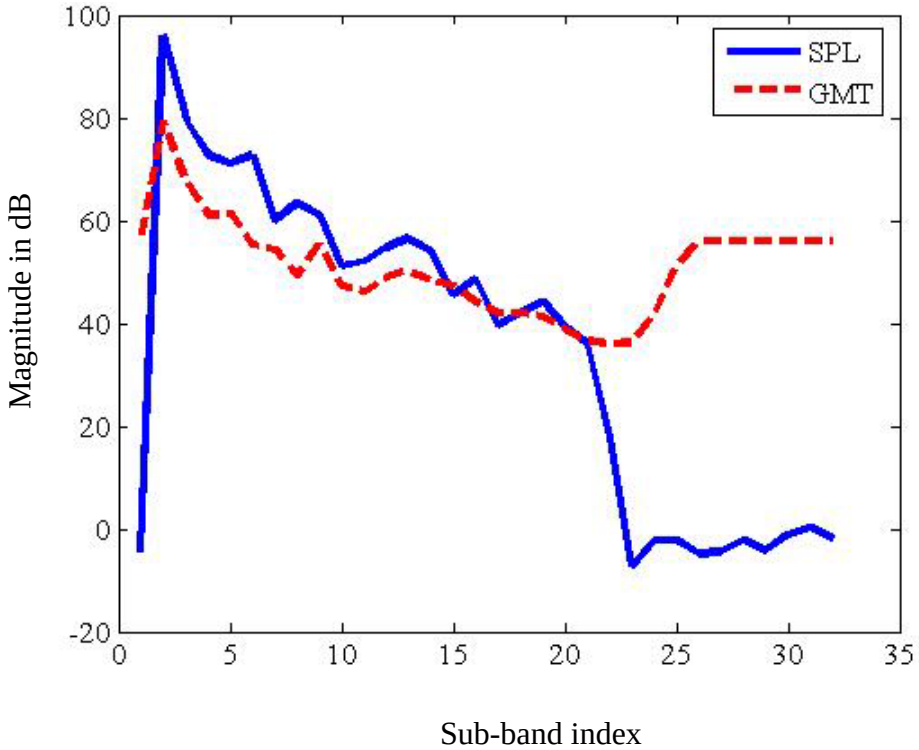

Figure 4.1(b): Music SPL and GMT

Next, the process of ranking the various maskers is explained. In order to satisfy the loudness constraint, one would like to have as low volume of the masker as possible, for a fixed value of noise. This is likely to be achieved if the waveforms of the maskee SPL and the masker GMT are similar in the corresponding frames.

For example, let us consider *n* frequency points of interest and their SPL and GMT values for the masker and the maskee. For these *n* points acquired for masker and maskee, the differences in decibels are calculated, and the masker is amplified by the maximum of the difference values. This ensures that the masker GMT and maskee SPL are coinciding at one point and the masker GMT is above the maskee SPL at other remaining points. Thus, the first condition for a suitable masker is fulfilled. Now the individual differences at all points between the masker GMT and maskee SPL are summed together to give a masking score for that particular masker, as given below:

Let $\mathbf{x} = (x_1, x_2, ...x_n)$ be the maskee SPLs.

Let $\mathbf{y} = (y_1, y_2, ...y_n)$ be the GMTs of a potential masker.

The matching process for $\mathbf{x}$ and $\mathbf{y}$ is as follows:

Step 1: d=x-y

Step 2: Increase gain of $\mathbf{y}$ by Max ($\mathbf{d}$)

Step 3: Again calculate **d=y-x**

Masking Score $S = \sum_{i=1}^{n} d_i$, where d_i is the *i-th* component of $\mathbf{d}$.

This process is repeated for all frames of all potential maskers. The masker with the least score is ranked first, the second lowest score ranked next, and so on for all the

maskers. Once a ranking is available, a pleasant sounding masker can be chosen from the best suited (low score) maskers. Only the selected masker signal and the amplification required for the masker to effectively overcome the maskee is passed on to the system for real-time processing. A small amplification of 1 or 2 dB can be optionally added to the selected masker to produce a better result.

Although systems integrating ANC and psychoacoustic processing have been implemented before, they have been mostly used in systems where the noise and audio are processed electrically [48], [49]. Once the audio (desirable sound, such as soothing music or sounds from nature) and the noise (undesirable, such as engine noise) are out in the acoustic domain, the complexity of combining ANC and masking increases a notch of difficulty. An application of this method has been used in [50] to cancel out snore noise and mask the residual with a pleasant noise to minimize disturbance to the snorer's partner. This thesis extends the work in [50] using ANC and masking techniques for handling a wider range of noises, with as low volume as possible. To keep the implementation simple, a narrowband engine noise is considered first. Figure 4.2 (a) illustrates the spectrum of the recorded engine noise on a Bark scale. The narrowband nature of the signal is further highlighted in Figure 4.2(b) which covers a limited frequency range. It can be observed from these figures that considerable attenuation can be achieved in the sound pressure levels just by the reduction of the fundamental noise and its harmonics. Subsequently, more complex waveforms are considered for suppression using the proposed system.

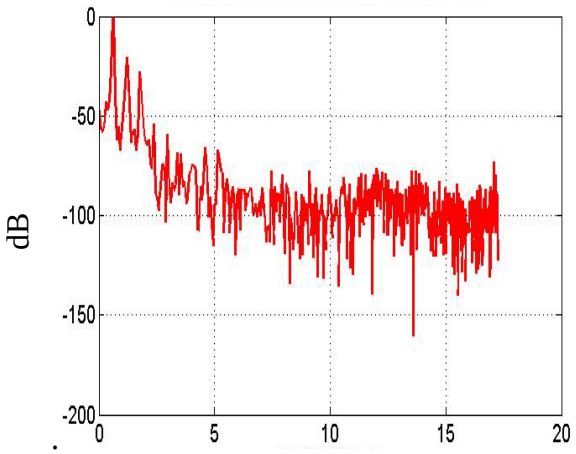

Figure 4.2(a): Engine Noise in Bark scale

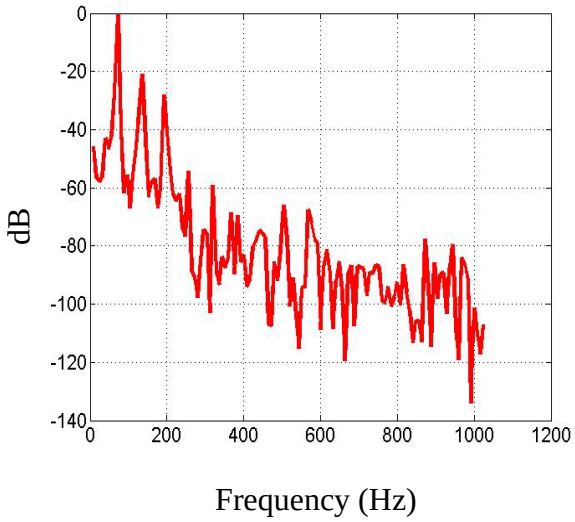

Figure 4.2(b): Engine Noise in Hz

4.2 Secondary path modeling using masking signal:

Most available control algorithms for ANC require identification of the secondary path. The error sensor picks up the residue of the acoustic superposition between the canceling loudspeaker output and the primary noise. The secondary path includes the digital-to-analog (DAC), reconstruction filter, power amplifier, speaker, acoustic path from loudspeaker to error microphone, error microphone, preamplifier, anti-aliasing filter, and analog-to-digital converter (ADC). Therefore, to get a stable adaptive filter, it is necessary to compensate for the secondary-path transfer function, as can be seen in Figure 2.2(Chapter 2). Hence, this is a vital component of any ANC system. Figure 4.3 illustrates the technique for modeling the secondary path.

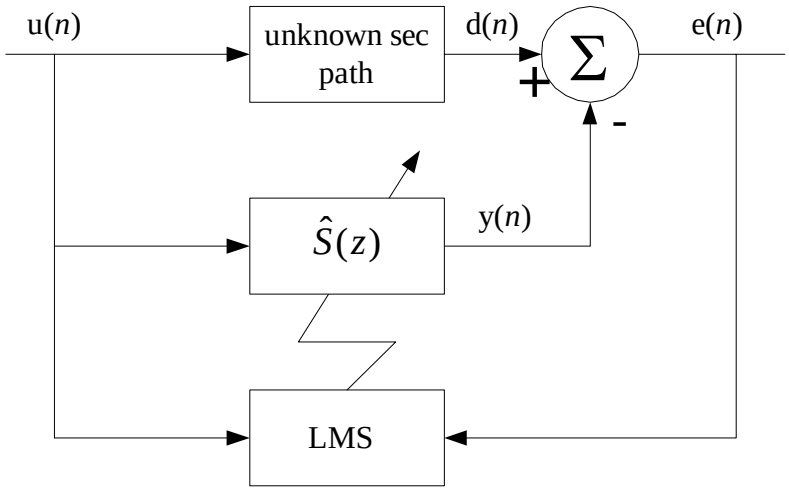

Figure 4.3: Secondary path modeling

A known signal u(*n*) is passed through the unknown secondary path and the adaptive filter is updated iteratively so that it resembles the actual secondary path. The magnitude of the actual and estimated secondary path coefficients shown in Figure 4.4(a) and the rapidly converging error in Figure 4.4(b) illustrate that a reasonably good estimate can be obtained using the LMS algorithm after a certain period of time (n > 400).

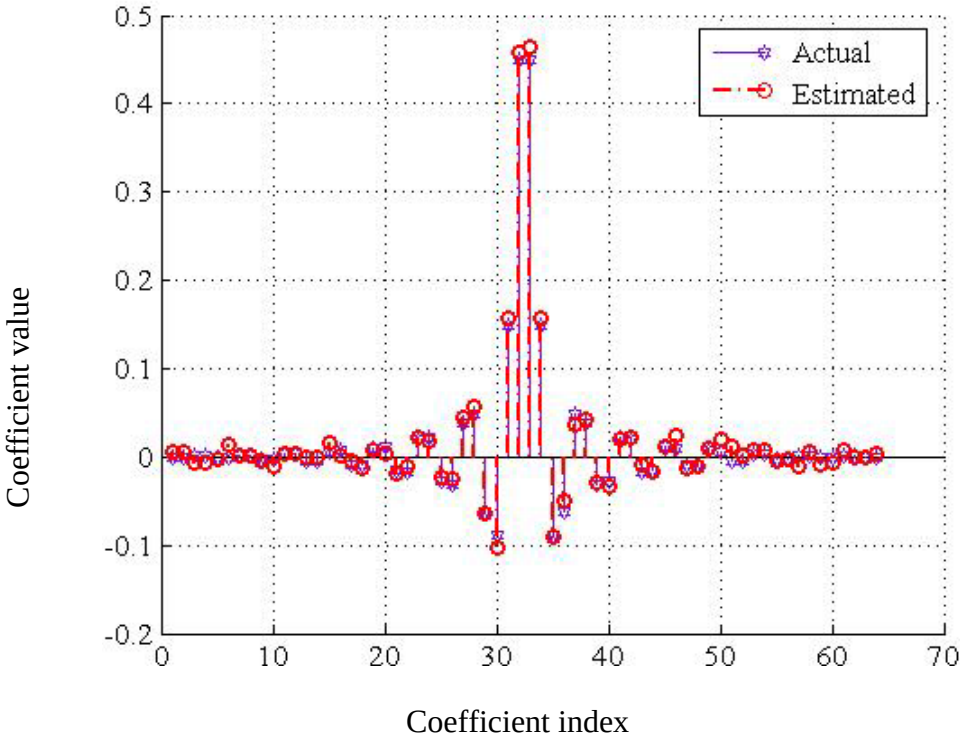

Figure 4.4(a): Actual and estimated secondary path coefficients

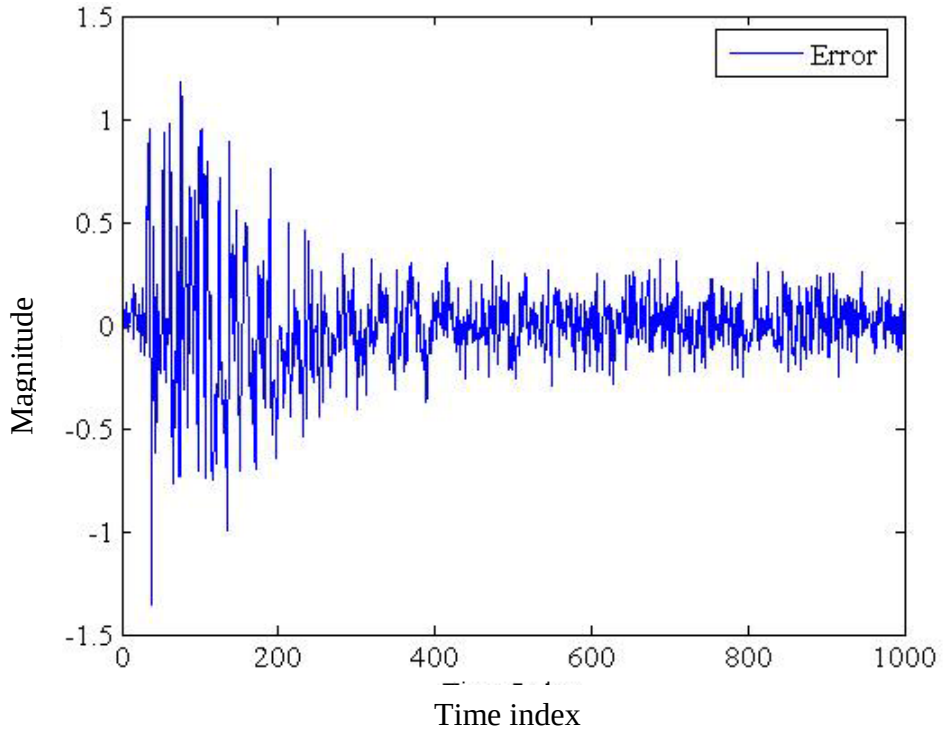

Figure 4.4(b): Error due to inaccurate estimation of secondary path

While any adaptive FIR filtering algorithm could be used for designing the secondary propagation path estimate, the LMS or the NLMS algorithms are often used due to their simplicity and robustness [4]. A comparison of the error signal obtained using the secondary path estimates of the LMS and NLMS algorithms can be seen in Figure 4.5. The parameters used here for the LMS algorithm are:(i) length of secondary path modeling filter is 64, (ii) step size is 0.008 and (iii) input signal is a pseudo-random noise chosen from a normal distribution with mean zero, variance one and standard deviation of one. The NLMS algorithm uses the same parameters except for the step size value of 1.

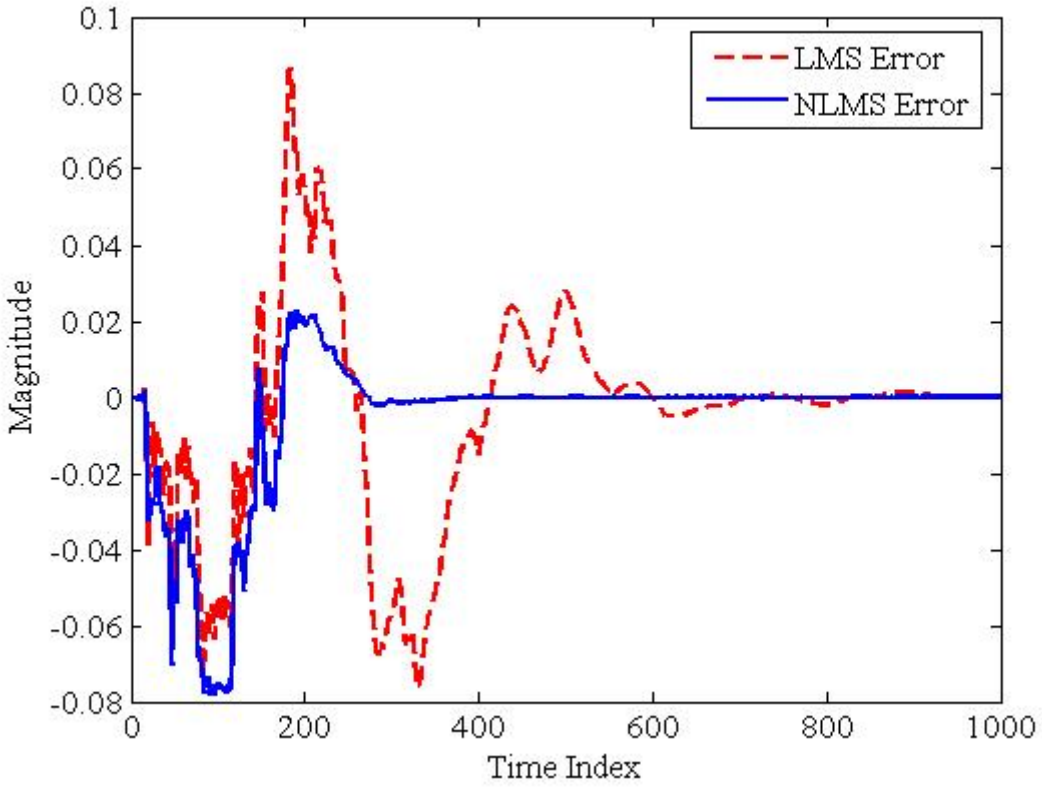

Figure 4.5: Error signal comparison of LMS and NLMS algorithms

Usually, white noise is used as the training signal, as it covers a wide spectrum and the actual noise sought to be cancelled may not be available. As white noise sounds unpleasant (like a hiss) to the ear, the selected masker (classical music), with wide frequency spectrum is used for the training. We can observe from Figures 4.5(a) and 4.5(b) that even though the error is initially larger for the system using pleasant sound instead of white noise, the error quickly converges to a small and comparable value after 0.08 second (Corresponding to 800 units on the x-axis). Thus, a pleasant masker can be used without any significant reduction in the system performance.

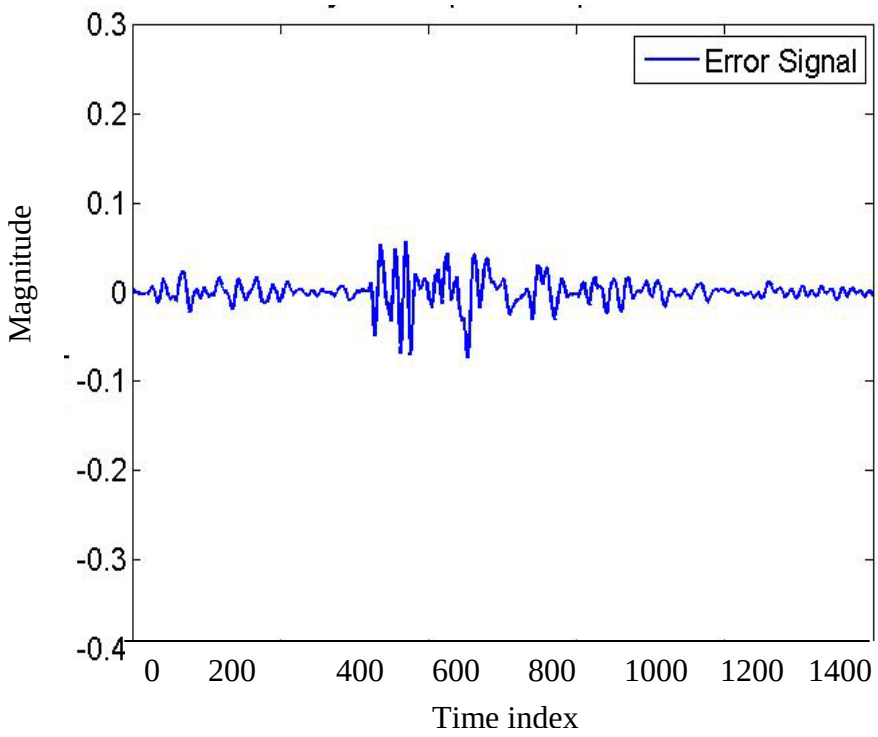

Figure 4.5(a): Modeling error using white noise

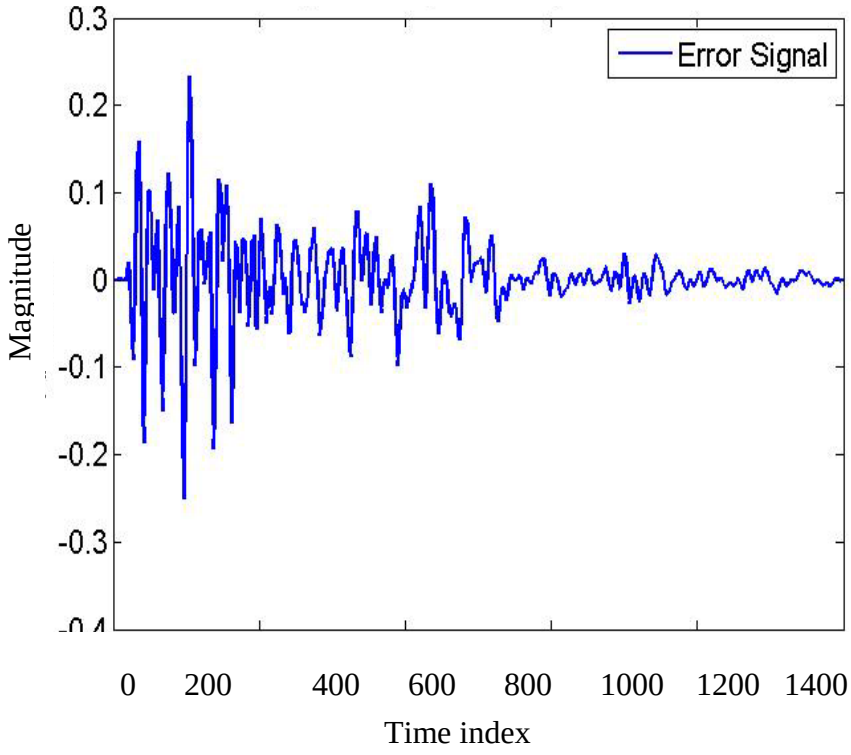

Figure 4.5(b): Modeling error using pleasant sound

4.3 Integrated ANC and Masking:

Once a suitable masker has been selected and a reasonably good estimate of the secondary path has been obtained, the building blocks for the IANCM system is proposed. Figure 4.6 illustrates the IANCM system.

The IANCM system has a structure similar to the conventional ANC system. In fact, if the gain value of g is set to zero, the system is reduced to a conventional Filtered-x Least Mean Square (FXLMS) ANC system. A reference input $x(n)$ is passed through an adaptive filter $W(z)$ and the output is fed to the canceling speaker. The output passes through the secondary path $S(z)$ and interferes with the primary disturbance output

through the primary path represented by $P(z)$. The dashed box represents the flow of the signals in the acoustic domain (open environment).

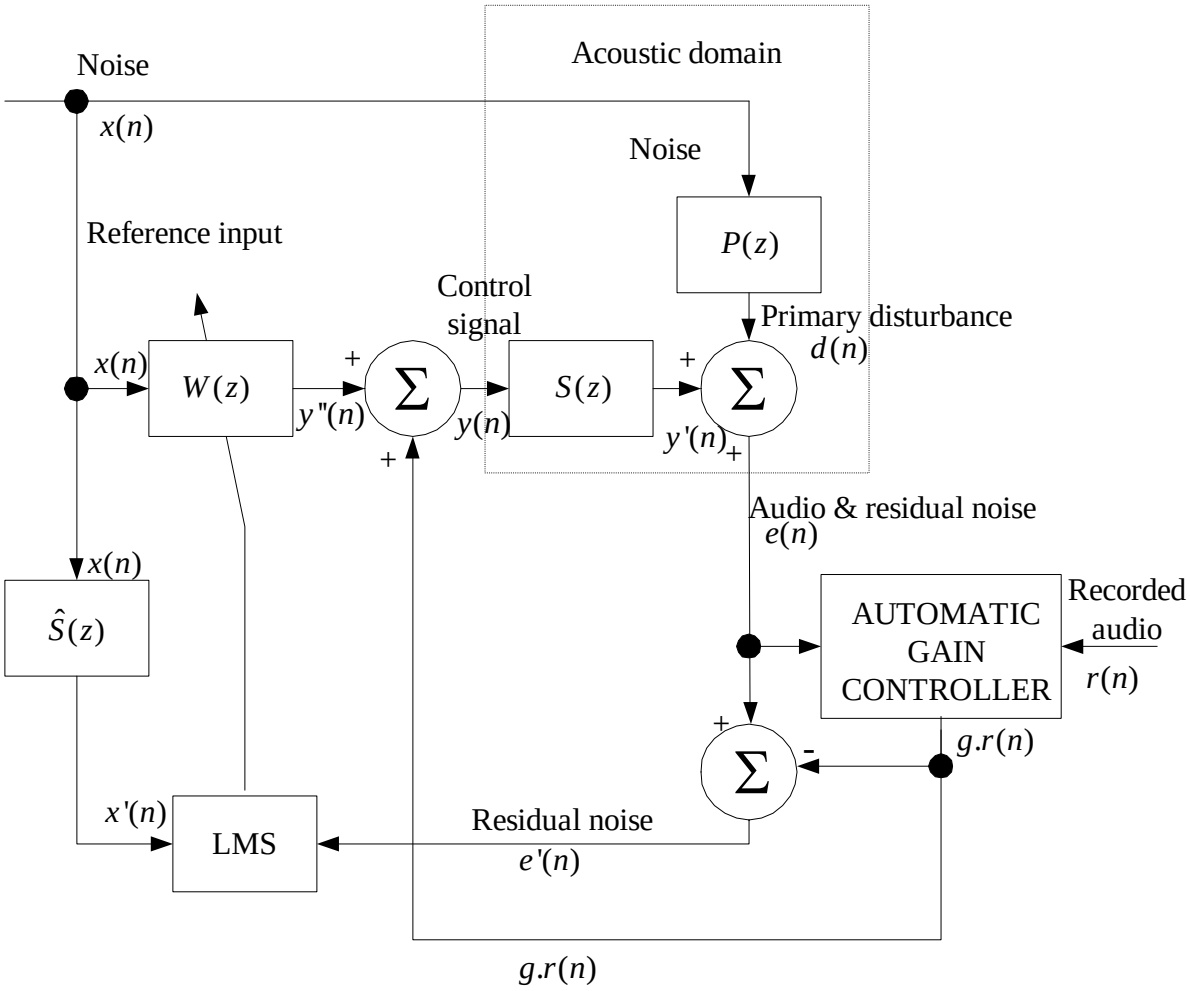

Figure 4.6: Integrated ANC-Masking System

Equation (4.1) illustrates the superposition effect at the error microphone

$$e(n) = y'(n) + d(n). \tag{4.1}$$

The result of the superposition of the two waves is further processed by the system. To compensate for the effects of the secondary path the reference input is also passed through an estimate of the secondary path. The LMS controller receives the filtered reference signal and the error signal as inputs, and based on these signals, alters the values of the coefficients of the adaptive filter. The adaptive filter processes the input noise reference signal and its output is combined with the scaled audio signal to give the control signal. This control signal $y(n)$ is given as input to the canceling speaker. The secondary path $S(z)$ represents the DAC, the ADC and the entire set of components/acoustic path in between. The output after the filter representing the secondary path is termed $y'(n)$ and this interferes with the primary disturbance $d(n)$ to produce the error signal $e(n)$.

The objective is to ensure that the sound obtained after the superposition, in addition to being minimized in magnitude, resembles audio more than the noise. Therefore, the outcome of the superposition of the control signal after the secondary path and the primary disturbance should sound like the chosen audio to the human ear. To achieve this objective, we compare it with the desired signal and obtain a pseudo error. The system now works in the conventional manner and tries to minimize this pseudo error, thereby ensuring that the actual error sounds like the chosen masker.

$$e'(n) = e(n) - g.r(n), \tag{4.2}$$

Here $r(n)$ is the recorded audio being used as the masker. The calculation of gain constant g for the masker will be explained in Section 4.4. The symbol '.' in (4.2) denotes element multiplication. The residual noise $e'(n)$ is further processed in a manner identical to a conventional ANC system [4]. Furthermore, the audio input is fed to the canceling speaker along with the output of the adaptive filter. Thus, the adaptive filter coefficient update is given as

$$w(n+1) = w(n) + \mu.e'(n)x'(n). \tag{4.3}$$

The IANCM system performance on a narrowband engine noise using constant spectrum classical noise will be examined next. The simulations of the system are carried out in Matlab [14]. The length of the adaptive filter is chosen as 256, which is the same as the length of secondary path modeling filter. A suitable step size of 0.003 is used in the simulation. A constant envelope classical music (cm31.wav in appendix B) has been used as masker and also for training the system. The recorded noises of an engine have been used as noise sources.

As can be seen from Figure 4.7(a), the overall envelope of the undesirable noise has reduced. It can be observed from the simulations, that on an average, there is about a 10 dB reduction. Figure 4.7(b) illustrates the spectral content of the processed output in relation to the original noise and the output of a conventional ANC system. It can be observed that in the lower frequencies, the IANCM system performance is comparable to an ANC system. In the higher frequencies, the magnitude level increases and masks out the high frequencies environmental noise. Compared with ANC, IANCM results in

relatively pleasant residual noise. Comparison of the IANCM output with the masker reveals that their spectrum content is similar. The similarity extends throughout the entire human auditory range. This is mandatory if the listener is to experience a pleasant auditory experience. Hence the twin objectives of reducing the undesired noise, as well as managing the residuals so that they sound pleasant, have been achieved.

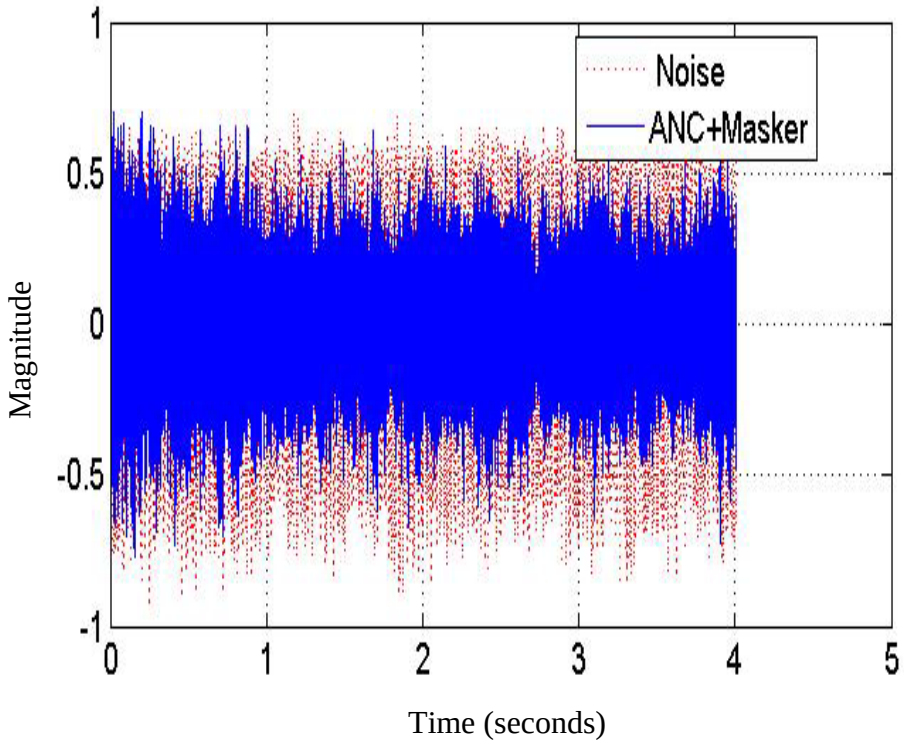

Figure 4.7(a): Amplitude reduction with IANCM

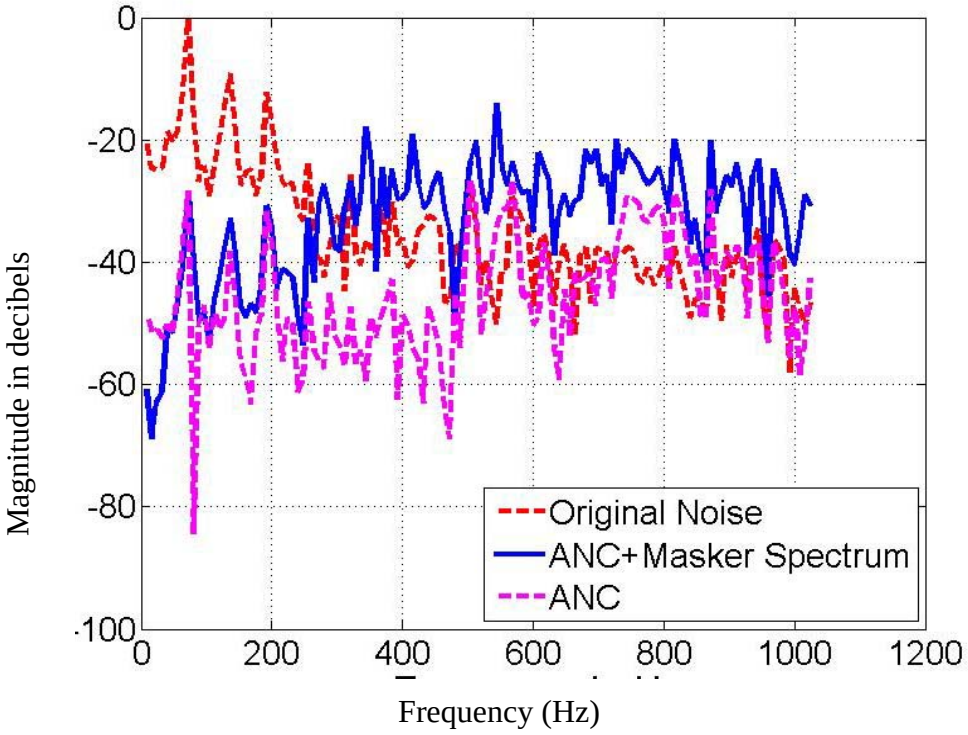

Figure 4.7(b): Spectrum comparison plot

However, one major problem faced in the IANCM system is that the sampling frequency used in a narrowband ANC, is quite low (a few kHz). In contrast, any pleasant noise covers a gamut of frequencies and ceases to be pleasant the moment it is restricted to a very narrow band. Thus, there is a mismatch between the sampling frequency requirements of the noise and the masker. A solution suggested in [48] is to equalize the sampling rate by up sampling ANC part. High quality music is available at a sampling rate of 44.1 kHz or higher. However, implementing ANC at this high rate may not be practically possible. Thus a compromise by down sampling the masking portion to a practical limit may be necessary.

4.4 Time-varying Maskee (Noise Source):

The solution for a narrowband, near stationary engine noise has been discussed so far. However, in a practical scenario, the engine rpm is likely to vary, causing a change in the characteristic of the maskee (or noise). Hence, there is a need for the system to adapt to any such variation. The spectrum plots of the same engine noise at two rpms will be very different due to the change in the fundamental and the harmonics. It can be inferred that the difference in the maskees may result in different maskers being found suitable.

To avoid the complexity of changing masker when the noise source is changing, it is proposed that automatic gain control (AGC) of the selected masker levels be performed, so that condition I stated in Section 4.1 remains satisfied. It can also be assumed that general trends/characteristics of noise source are not changing too drastically. The signal flow for the proposed AGC is illustrated in Figure 4.8. It is pertinent to mention that this is not a continuous process and the change in gain takes place only on receiving an external trigger such as an electronic signal indicating a significant different in rpm of the engine.

After the IANCM system has converged to the new rpm's residual noise, the AGC comes into operation. As can be seen from Figure 4.6 and 4.8, the AGC receives the error signal and the recorded audio as inputs and produces the amplified or attenuated (depending upon the value of gain, g) audio signal as the output. The initial value of g is considered as unity. Here, $e(n)$ consists of both the recorded audio and residual noise components. Thus, the difference between $e(n)$ and $r(n)$ results in a residual noise. The

objective now is to obtain a scaled version of the audio, which is able to mask this residual noise, without being louder than necessary.

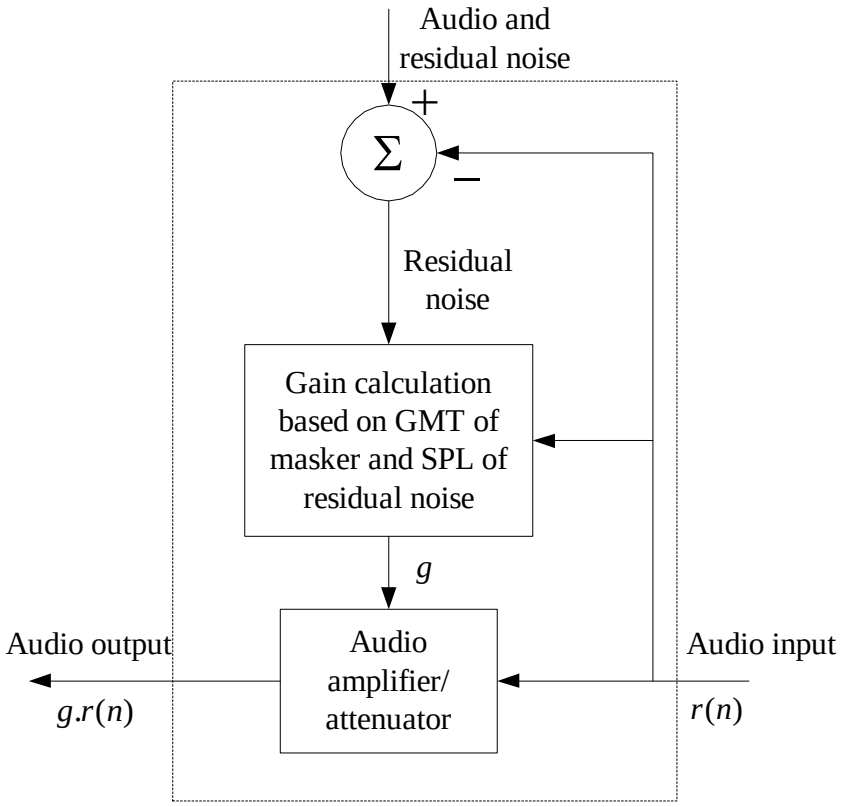

Figure 4.8: Signal flow for Automatic Gain Control of Masker

The value of gain g is calculated using the steps explained in Section 3.5.3. Let $\mathbf{m}$ denote the set of SPLs of the residual noise $e'(n)$ and $\mathbf{n}$ denote the set of GMTs of the current volume of audio, $g.r(n)$, then we obtain a value

$$t = \max(\mathbf{m}\text{-}\mathbf{n}). \tag{4.4}$$

The new value of gain is given by

$$\begin{aligned} g_{new} &= g_{current} + \Delta, \ \ if \ \ \ t \geq threshold. \\ g_{new} &= g_{current} - \Delta, \ \ if \ \ \ t < threshold. \end{aligned} \tag{4.5}$$

The values of Δ and the threshold are selected such that the system remains responsive to any changes, without fluctuating too much. The effect of the gain control on noise cancellation and output spectrum quality remains to be seen. Figures 4.9 (a) and (b) demonstrate the effect of processing on the initial rpm and Figures 4.10 (a) and (b) demonstrate the result after the rpm has been switched over. It can be seen that there is no deterioration of the system performance. In fact, Figure 4.11 clearly demonstrates that the system continues to try and catch up with the reduction in decibel levels achieved with a conventional ANC system. The system is able to improve on the quality of the residual noise without compromising much on the reductions in quantity. Thus, it can safely be stated that a reasonable solution for adapting to changes in the maskee has been obtained.

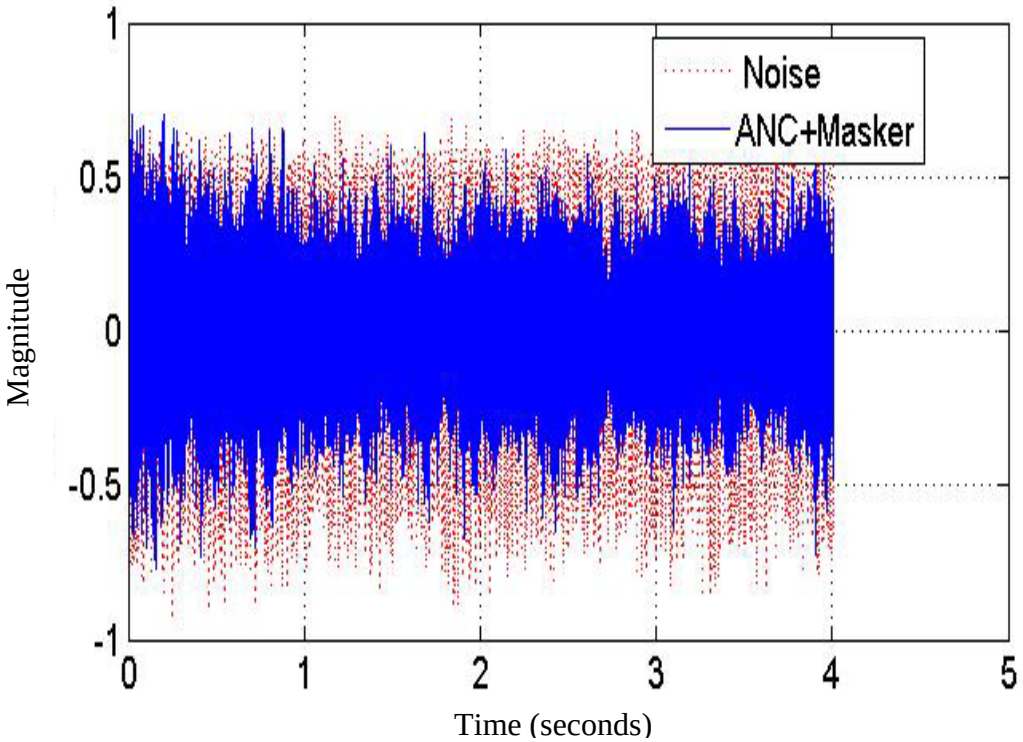

Figure 4.9 (a): Reduction in noise magnitude (Initial rpm)

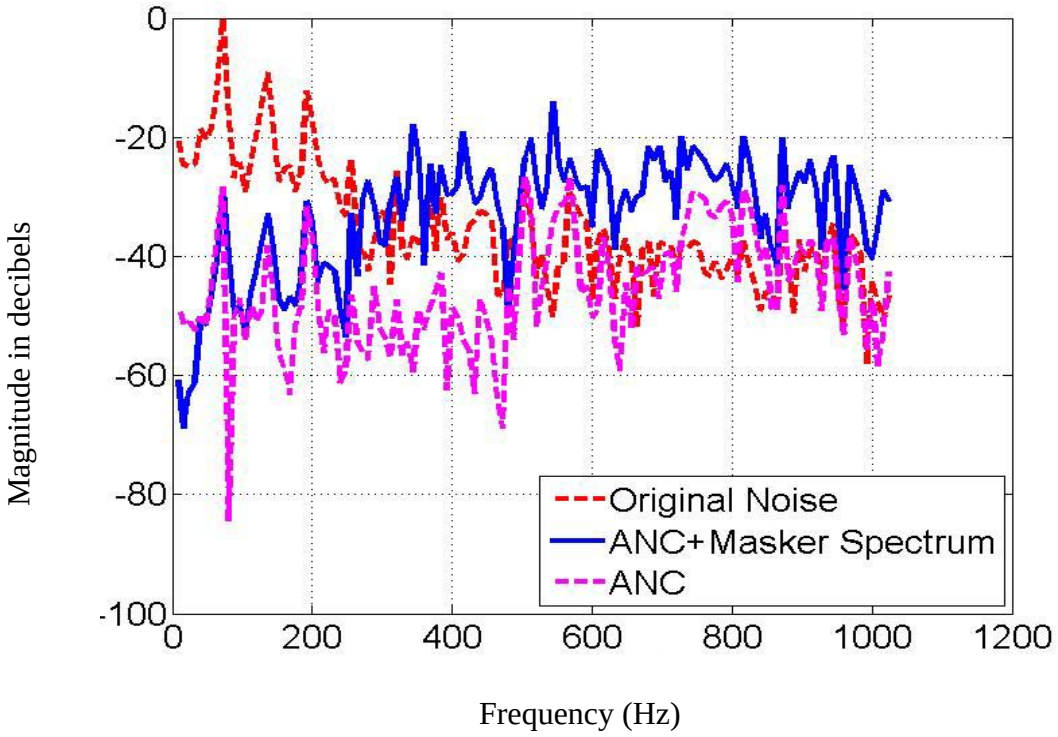

Figure 4.9 (b): Comparison with ANC system in the frequency domain

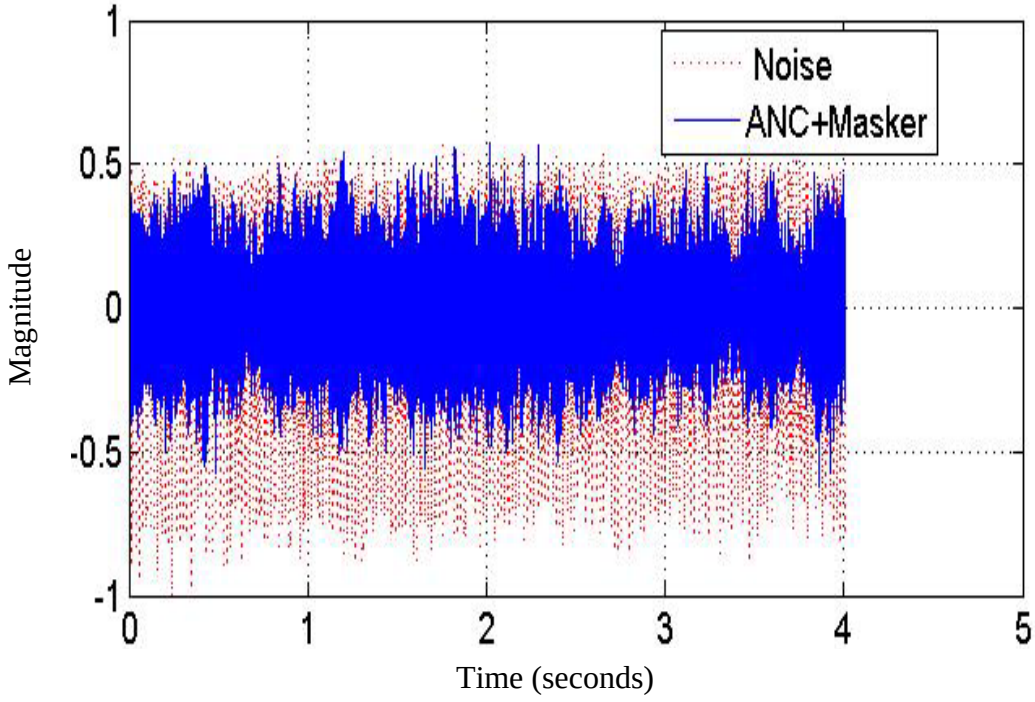

Figure 4.10 (a): Reduction in noise magnitude (Changed rpm)

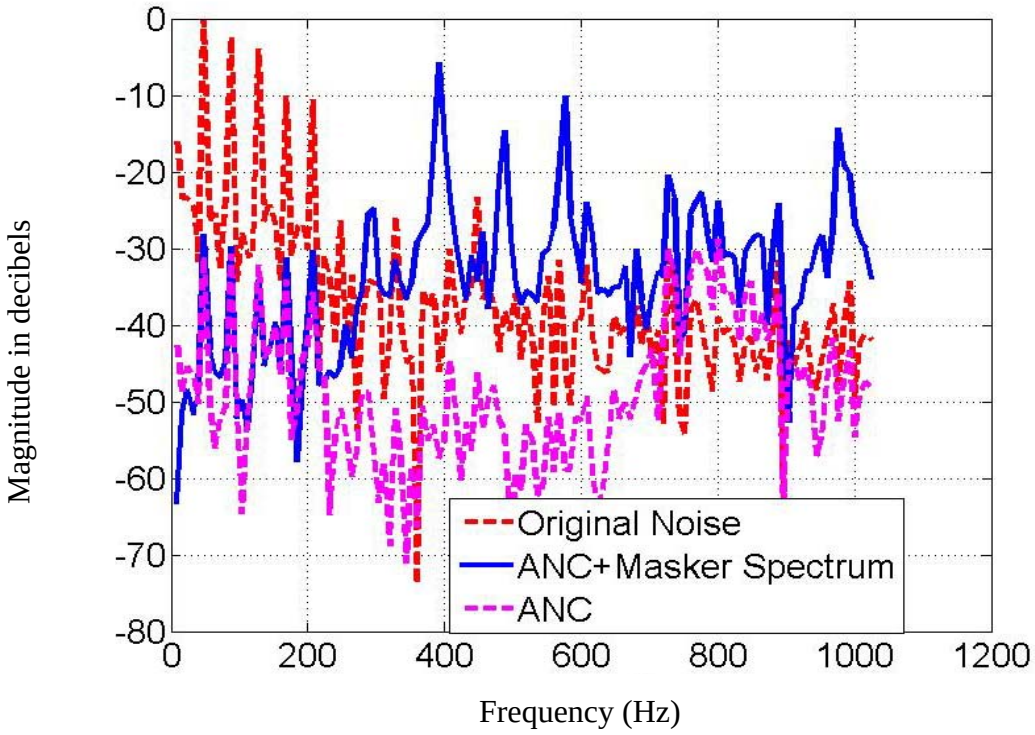

Figure 4.10 (b) Comparison with ANC system at changed rpm

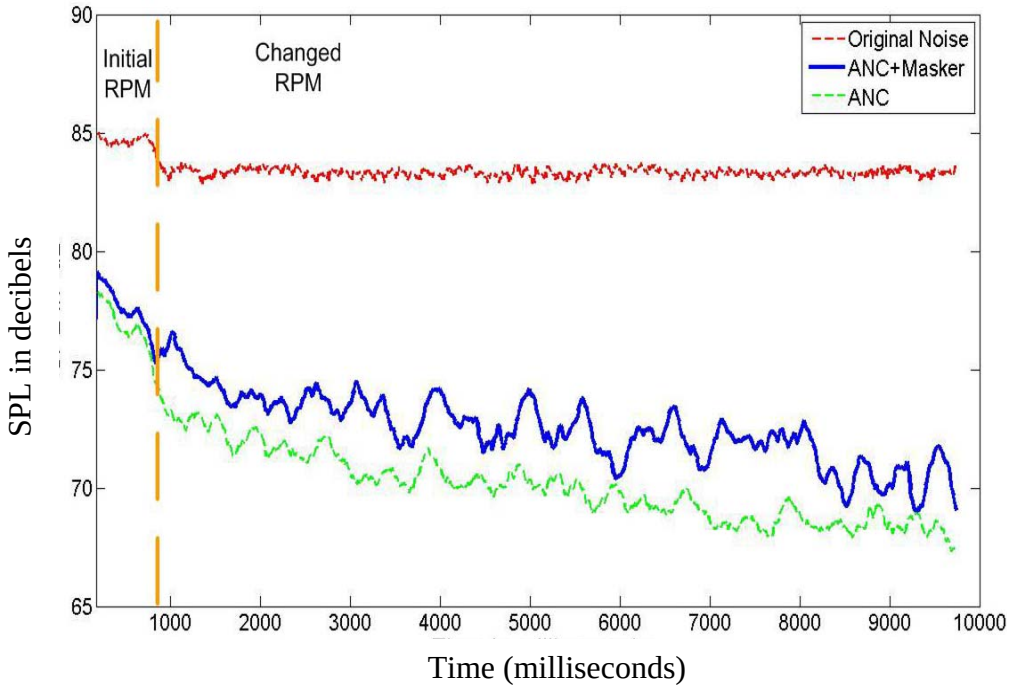

Figure 4.11: Effect of IANCM on the SPL of the signal

4.4.1 Effect of varying gain on the output

Based upon the results of the simulations, it can be deduced that the gain value is a very important parameter of the IANCM system and decides the behavior of the system. If the gain is kept at a very low value, then the maskee is audible and the result is an unpleasant noise. On the other hand, a high gain means that the masker is too loud and the benefits of ANC are lost. Hence, the AGC should be able to find an optimum gain value for a pleasant listening experience.

Figures 4.12 (a) through 4.12 (d) further highlight the role played by the gain value. As seen in Figure 4.12 (a), the maskee (recorded engine noise, engine1.wav in appendix B) has most of its energy concentrated below 200 Hz. The tonal nature of the noise is apparent from the dark straight lines in the lower frequencies. Figure 4.12(b) displays the wavy, melodic form of the masker (classical music, cm30.wav in Appendix B). A conventional ANC system, as is visible from Figure 4.12(c), is successful in reducing the low frequency components, however the energy seems to have shifted to the higher frequencies. The straight streaks, characteristic of an unpleasant noise, are still visible and can be heard (engine1_anc.wav in Appendix B). On the other hand, the output of the IANCM system (engine1_cm30.wav in Appendix B) looks like a low energy version of the masker spectrum. The wavy pattern and discontinuous patches, similar to the masker are clearly discernible.

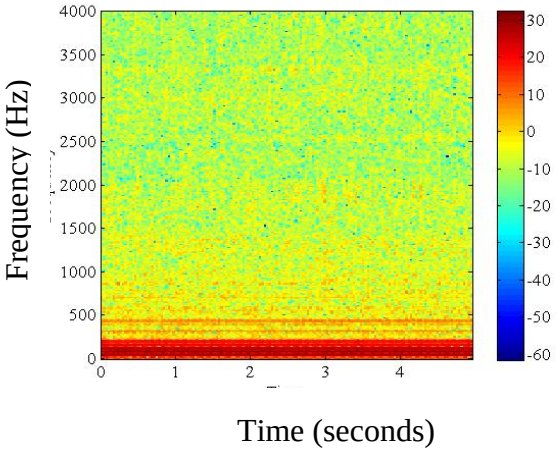

Figure 4.12(a): Maskee spectrum

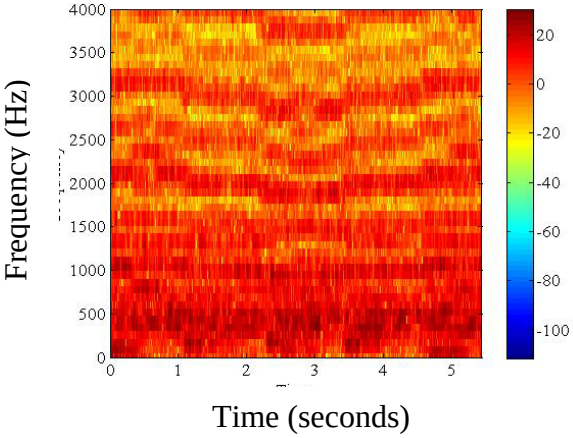

Figure 4.12 (b): Masker spectrum

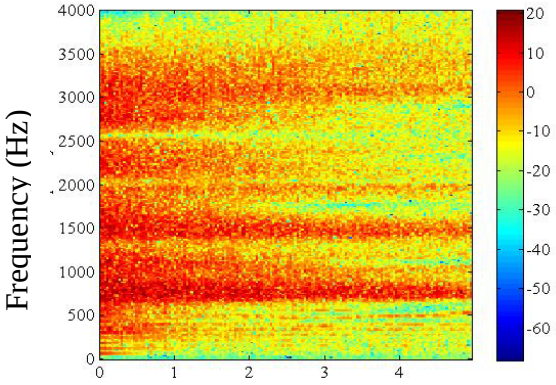

Figure 4.12(c): ANC output

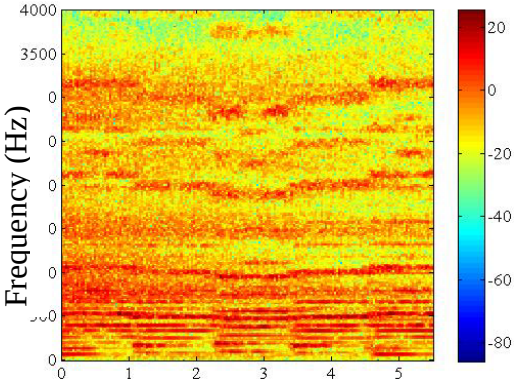

Figure 4.12(d): IANCM output

4.4.2 Time-variant broadband maskee

In the previous Section, the ability of the system to cater for any changes from one narrow band noise to another was demonstrated. It can be observed from Figures 4.13 (a)-(d) that the system fares reasonably well even when faced with a time-variant broadband noise (cabin noise of a car moving on road, car1.wav in Appendix B).

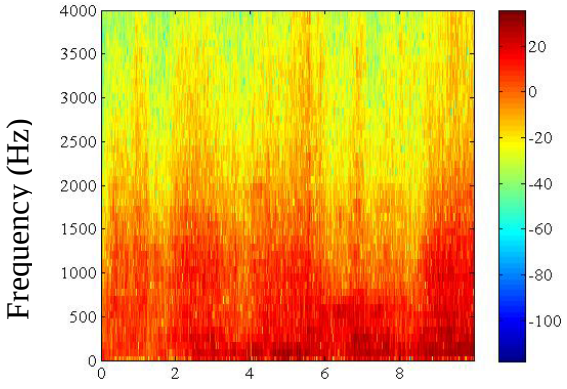

Figure 4.13(a): Maskee spectrum

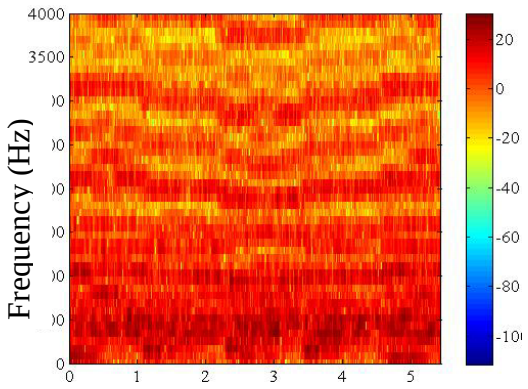

Figure 4.13 (b): Masker spectrum

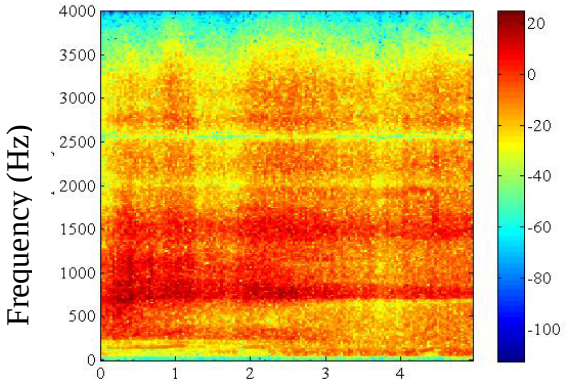

Figure 4.13(c): ANC output

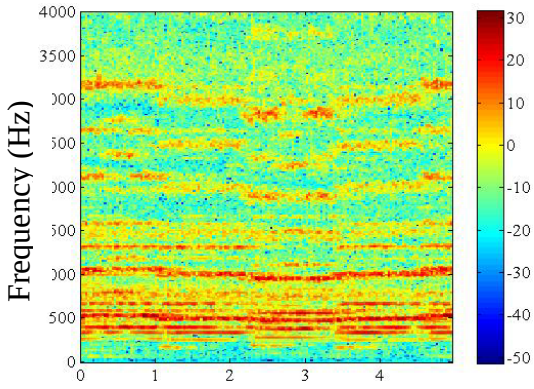

Figure 4.13(d): IANCM output

4.5 Practical utility of the system

The efficacy of the IANCM system has already been highlighted for narrowband engine noise, abruptly changing engine noise and a slightly wider band car noise earlier. In this section, the utility of the system is explored for noises occurring in day-to-day activities of an average household and analyze the nature of the noise, as well as the characteristics of a suitable masker. In addition, another psychoacoustic effect is introduced for exploitation by the IANCM system.

While performing daily household chores, many sounds such as the vacuum cleaner, blender, air conditioner or even a partner's snore noise are encountered. Some of these have been accepted as a way of modern life, while others may be quite annoying. Alternatively, some noises may be unacceptable at certain times of the day and quite alright at other time slots. For example, normally the noise of a hand blender is not annoying when no one is disturbed, but when a baby is sleeping it might not be acceptable. Snore noise has been found to have an adverse effect on the relationship between partners and there have been some attempts [50], [51] to alleviate these symptoms. However, the field still has considerable scope for improvement.

The waveform and spectrum of a typical snore noise (snore.wav in Appendix B) are shown in Figure 4.14. The two halves of the image show the typical inhale-exhale pattern. The spectrum shows the peak of the decibel levels (darkest lines) to be in the vicinity of 80 Hz, corresponding to the resonant frequency of the human tissues vibrating to produce the snore noise. The continuous (in time) streaks of high energy are typical of the sounds that are usually annoying to the human brain.

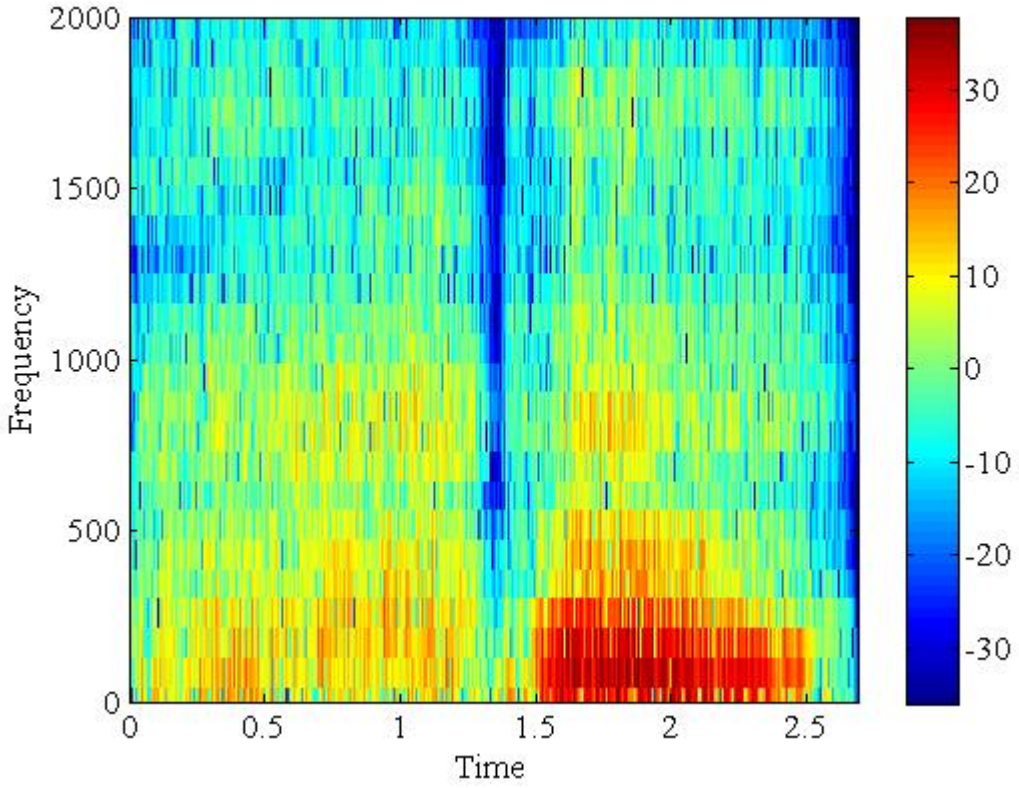

Figure 4.14: Snore waveform and spectrum

Going further, we analyze a few other sounds such as vacuum-cleaner (Figure 4.15), hand-blender (Figure 4.16) and air-conditioner (Figure 4.17). The spectrum of a typical vacuum-cleaner (Royal Dirt devil Model 085360 acquired from http://freesound.iua.upf.edu, vacuum.wav in Appendix B) shows its dominant frequency at around 300 Hz with a prominent harmonic near 1400 Hz. Figures 4.16 and 4.17 show that the hand-blender (downloaded from http://freesound.iua.upf.edu, handblender2.wav in Appendix B) emits noise mainly in the frequency region of 800 to 900Hz, and the air-conditioner (downloaded from http://freesound.iua.upf.edu, aircon.wav in Appendix B) emits sound energy almost uniformly in the region below 2000 Hz.

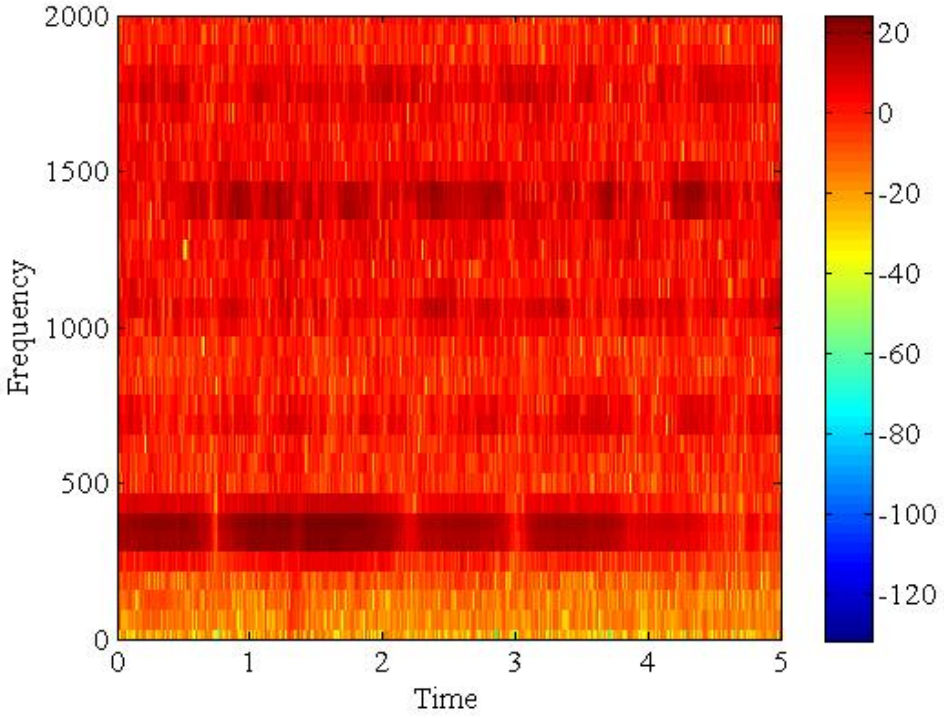

Figure 4.15: Vacuum cleaner spectrum

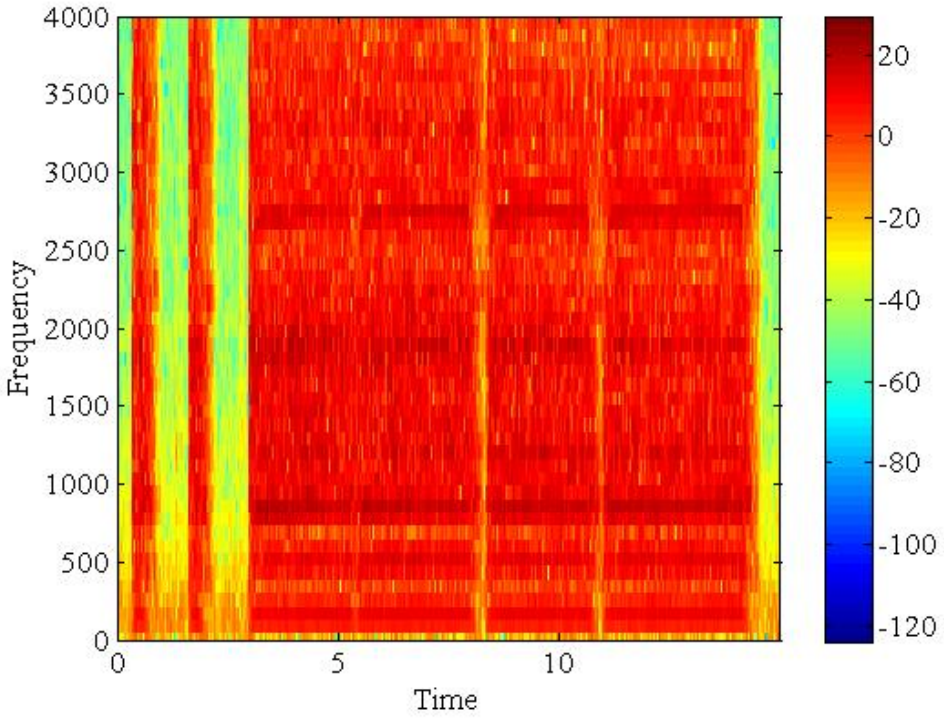

Figure 4.16: Hand-blender spectrum

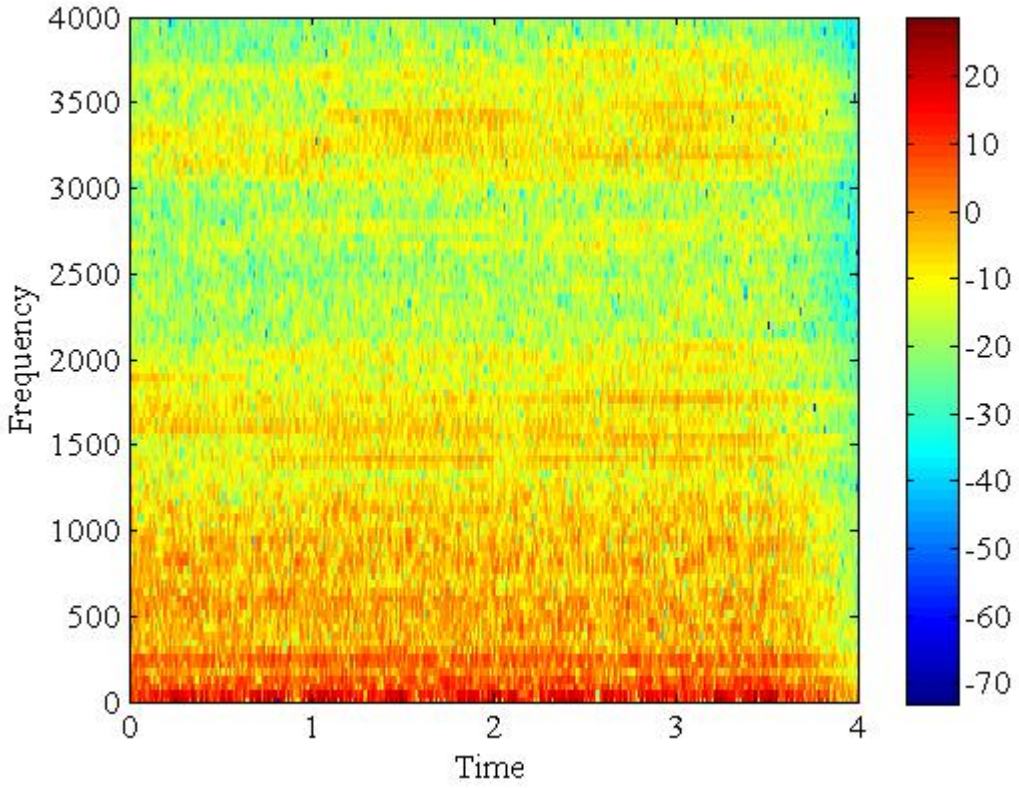

Figure 4.17: Air-conditioner spectrum

From observing all above noise spectrum, a common pattern emerges between all such unpleasant sounds. One or more high intensity, unbroken, dominant streaks of noise energy against an evenly spread out noise energy can be discerned. In contrast, the pleasant sounds, as can be seen in Figure 4.12 (b) and Figure 4.18 display a different type of pattern.

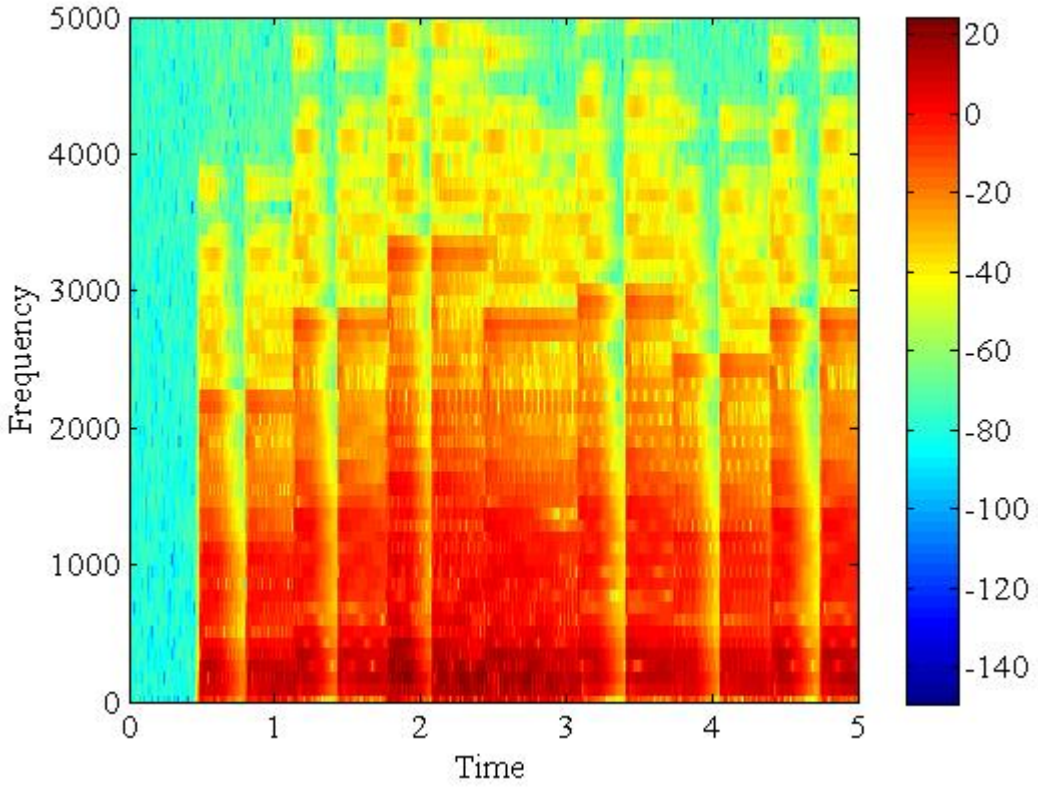

Figure 4.18: Piano notes spectrum

Observing the spectrum of pleasant sounding piano notes (“London bridge is falling down” downloaded from http://freesound.iua.upf.edu, piano.wav in Appendix B) we can see the contrast with unpleasant noises. As opposed to the high decibel streaks of noises, we observe rung like patterns embedded in nearly empty spaces. It appears as if the sound energy is climbing or descending in steps placed in silence. These high decibel ‘islands’ make their presence felt and then disappear from the scene giving way to a different frequency and so on. Thus, the human brain keeps on receiving different stimuli at certain intervals and the sound is perceived as pleasant [52]-[53]. Of course, this is a highly simplified explanation and the actual auditory process and human perception is much more complex and needs much more work, before it is fully revealed.

After observing the characteristics of the unpleasant noises and the pleasant maskers, another psychoacoustic characteristic that can be exploited, is the effect due to reverberation. Some piano notes get louder after they are listened to for a while [54]. On critical observation using any software such as Matlab or Sonic Visualiser [55], one can see that sound is not 'physically' getting louder but is only being perceived as being louder. A moderately strong fundamental, after being reinforced by the harmonics, appears to get louder even though the decibel levels do not change. Hence, the perceived loudness is due to reverberation and not an inherent increase in intensity.

Exploiting this property, maskers having a high portion of their energy in the regions of the harmonics generated by the maskee can be used for IANCM systems. For example, the snore sound (downloaded from http://freesound.iua.upf.edu, snore.wav in Appendix B) of Figure 4.14 has its dominant frequency at 80 Hz.

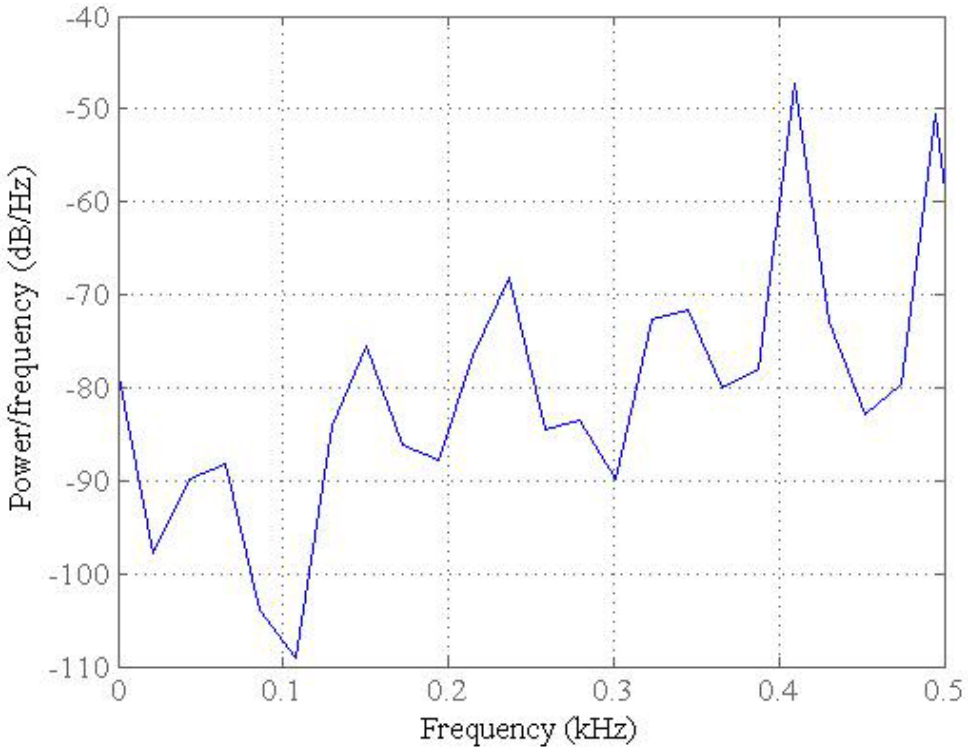

Figure 4.19: Tibetan chant power spectral density estimate

Now, looking at Figure 4.19 which displays the spectrum of a chant by Tibetan monks (downloaded from http://freesound.iua.upf.edu, tibetanchant.wav in Appendix B), it can be observed that it has high concentrations of energy in the 400 Hz region and other integral multiples of 80 Hz (such as 160 Hz, 240 Hz, 320 Hz and 480 Hz). It is widely believed that such chants have a soothing effect on the human brain [56]. Now as the energy is concentrated in integral multiples of 80 Hz, one can infer that a low decibel level of Tibetan chants is likely to sound louder in the presence of snore noise. This is because of the reverberations, due to matching harmonics. Thus, the objective of keeping the overall decibel levels low as well as masking the unpleasant sound can be achieved if the chant is used as masker to mask snore noise.

Running simulations in Matlab, with the maskee as snore noise (80 Hz dominant frequency) and Tibetan chants (400 Hz dominant frequency) as masker, satisfactory results are obtained (tibetanchant_snore.wav in Appendix B). Comparatively low levels of masker are able to successfully mask the snore sound. On comparing Figure 4.20 with Figure 4.14, we can see that that a very high degree of reduction in the decibel levels has been achieved and the character of the spectrogram has also changed. Most of the unpleasant streaks have been replaced with isolated energy spots. Many more such combinations of unpleasant sounds can be made with maskers which have frequencies that are integral multiples of the maskee frequencies. Thus, IANCM can be performed at lower decibel levels without any decrease in performance.

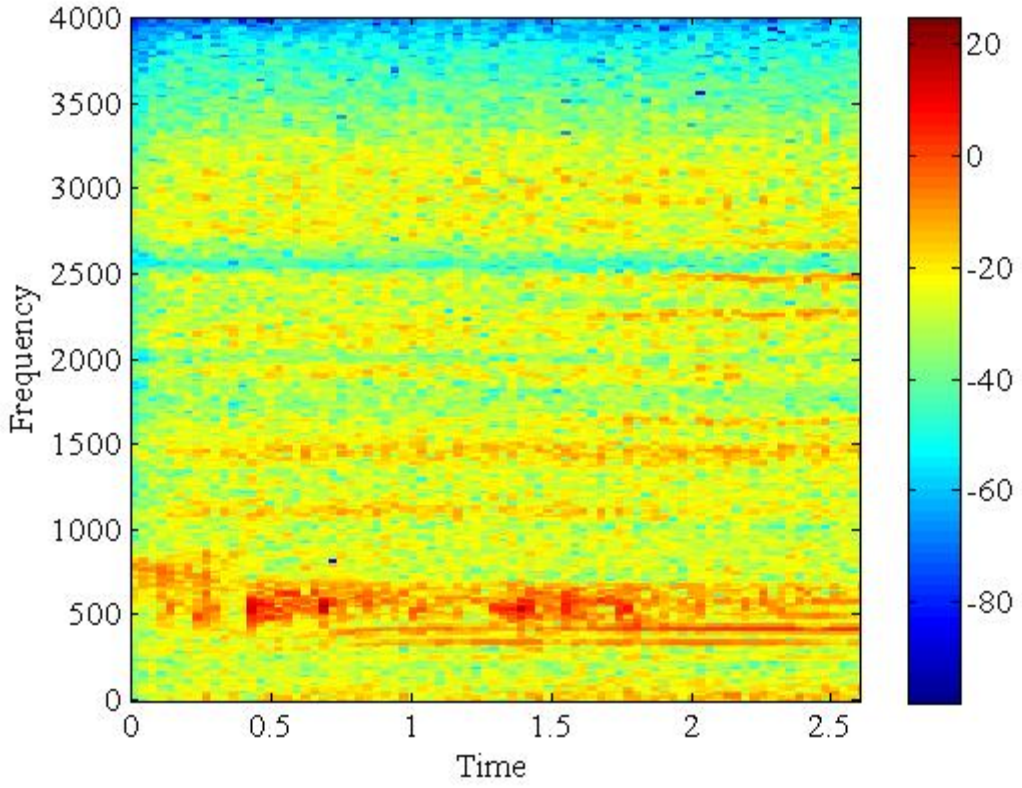

Figure 4.20: IANCM with snore as maskee and Tibetan chants as masker

It is pertinent to mention here that the discussions so far pertain to suppression of unwanted signals at one point in space. Unfortunately, it is also possible that at other points in the sound field, the pressure of the noise and canceling signal is in phase and constructive interference occurs, thereby increasing the sound level at such points. Thus the problem of global control of noise emerges. This characteristic of ANC systems has been well discussed in the literature [5], [15] and many of the solutions involve the use of multiple sensors to pick up the sound levels at various important locations and multiple loudspeakers located at strategic points to reduce the disturbance. The multiple error sensors necessitate the development of the "Multiple Error LMS Algorithm" put forth by Elliott and Nelson in 1985[57]. A general solution for an active control system with *K*

reference signals, *M* secondary sources and *L* monitor microphones has been discussed by Elliott and Nelson [58]. However, the problem of global cancellation, especially at higher frequencies and indeterminate type of noise is a highly challenging problem and needs a substantial technological breakthrough to achieve widespread commercial application.

4.6 Conclusions

In this chapter, it was demonstrated that in addition to attenuating an unpleasant noise using ANC, the residual can be managed in such a manner, as to render it pleasant to the human ear. This was achieved using the masking effect and other properties of human hearing. The ability to use a pleasant sound such as classical music, for secondary path estimation was also demonstrated. This rich classical music helps to avoid the hiss like white noise usually employed by ANC systems for training the system. AGC was also exploited for taking care of any changes in the engine noise (unpleasant noise) by raising or attenuating the recorded audio (pleasant). The process can be used for myriad applications in the noisy environments encountered commonly these days. The system can be used for domestic as well as public environments. Furthermore, a new method of exploiting psychoacoustics was also introduced to improve the system and expand its utility for daily use.

Chapter 5

Improvements in Integrated ANC-Masking

The IANCM system employs psychoacoustic comparison of the masker and the maskee for selection of the initial masker and subsequently, to adjust the gain of the masker to cater for the changes in maskee. This process has two shortcomings

(I) The consumer is not allowed the freedom of masker selection. The consumers of today are not satisfied by a limited choice [59] and may not like to listen to a repeated masker clip. Thus, an alternate system, where the user can select the masker from a pool, or even upload his/her own favorite masker into the system may be employed.

(II) A second drawback is with the process itself. The masker is chosen on basis of the comparison of the individual frames of the maskee and the masker. The maskee is usually a noise with steady amplitude; however the masker, like any pleasant piece of music is most likely to have periods of silence interspersed with louder segments. While comparing a masker frame containing periods of silence with the steady drone of the maskee, very high and impractical amounts of amplification are required for the maskee SPL to be below the masker's masking threshold. This imposes a serious constraint on the system that needs to be resolved. Hence a different strategy needs to be employed to overcome this shortcoming of the current system.

Simulations on different types of maskees using a broad range of maskers have demonstrated that the system is quite robust and any masker as per individual preference can be used. Thus, we can leave the task of masker selection to the user and the system can just concentrate on the task of varying the gain of the masker to a suitable levels.

In the next section, we introduce a method to keep the masker loudness at a suitable level, ensuring at the same time that the underlying noise is not perceivable.

5.1 Reduction of dynamic range for better performance

The results of the simulation have shown that the integrated ANC and masking is effective for those segments of the masker when the masker is loud. However, when the masker is soft, the residual noise may still be heard. Increasing the overall loudness of the masker would be counterproductive as the processed signal will be too loud and negate the reduction in loudness levels achieved by ANC. Thus, the two requirements are contradictory. However, if we can succeed in reducing the dynamic range of the masker without changing its perceptibility, then we may be able to mask out the residual noise, as well as, be successful in keeping the loudness of the peaks under control. To achieve our objective, audio signal (masker) can be dynamically compressed and its effect is observed for active noise masking and noise control.

5.1.1 Compression of masker

Compressors are classified as dynamic range control devices (DRC) [60]. The aim of a DRC is to increase or decrease the dynamic range of an audio signal without introducing

perceptible distortion. The DRC consists of signal detection by using either peak/root-mean-square or any other criterion; a delay to eliminate overshooting; and an adaptive filtering of the control signal to ensure smooth variations in gain [58]. Compressors can be used to minimize the loudness differences between audio programs. Compression is similar to AGC but operates at a faster time scale. This process increases the loudness of the softer parts of an audio so that it can be heard above the noise floor and decreases the loudness of the louder portions so that it may not disturb the unintended listeners.

As can be seen in Figure 5.1, a compressor reduces the gain of an audio signal if its amplitude exceeds a threshold. The amount of gain reduction is determined by a ratio control. In addition to several other applications, compression is utilized by broadcasting stations, as well as for commercials to achieve maximum possible perceivable loudness while staying within the acceptable norms. Compression is extensively used in the

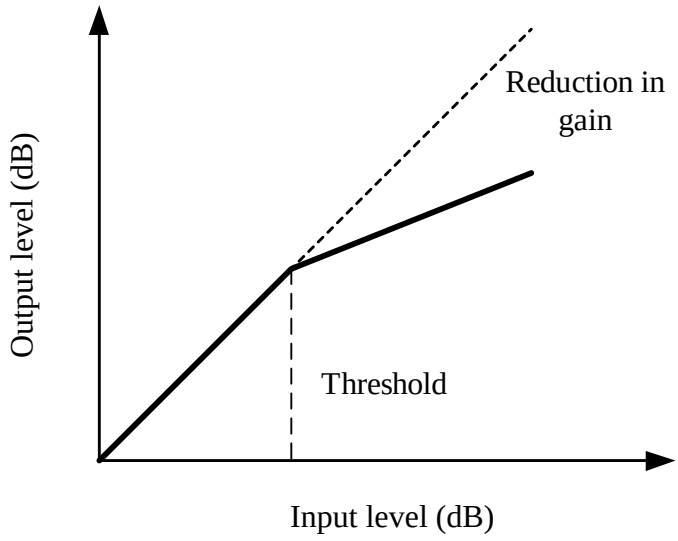

Figure 5.1: Output of a signal after processing by a compressor [58]

automotive audio systems as the noise floor levels are quite high. Once we have done the required compression, we can decide the gain level of the masker based upon its perceived loudness. Automatic selection of the compression ratio is a difficult task; however incorporating DRC into the IANCM system is simpler if the compression ratio is selected manually. Alternatively, a fixed compression ratio can be decided by subjective experiments on different types of maskers. Figure 5.2 shows how the masker can be compressed prior to being introduced into the IANCM system.

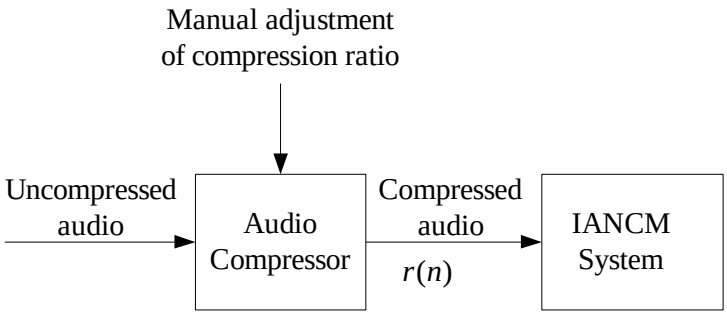

Figure 5.2: Offline pre-processing of masker

Utilizing this technique, the result of compression performed using Matlab [14] can be observed in the time domain. Figure 5.3(a) shows the original waveform of an excerpt of African drum music and Figure 5.3(b) shows the compressed version.

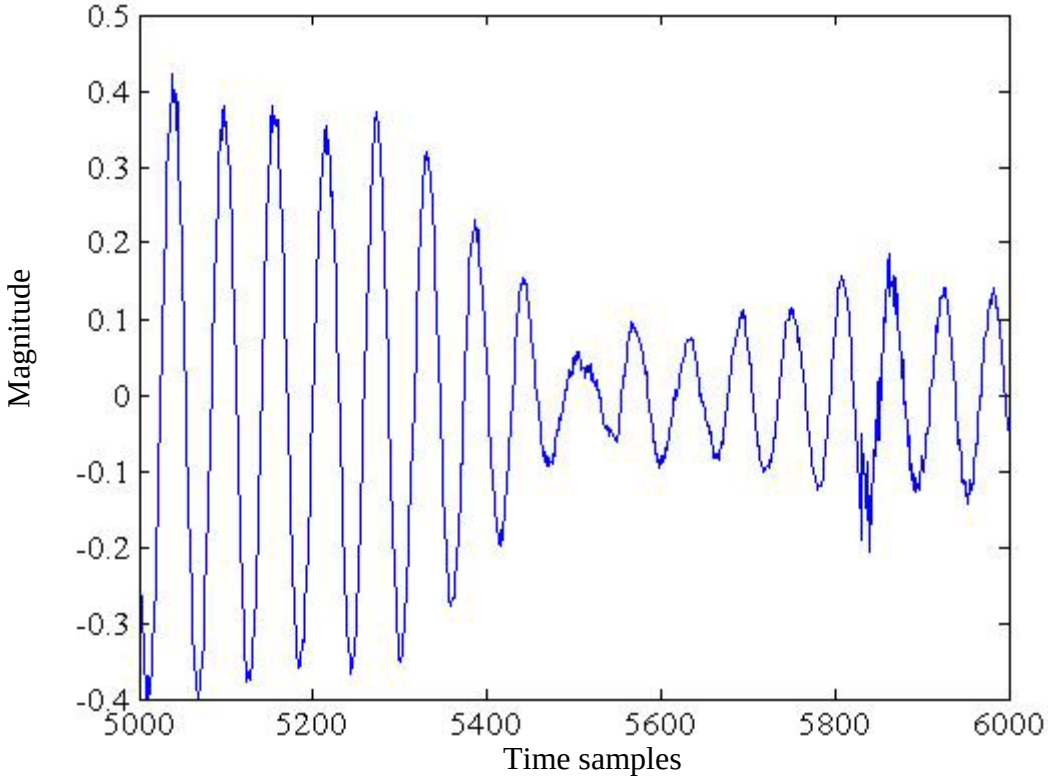

Figure 5.3(a): Uncompressed masker

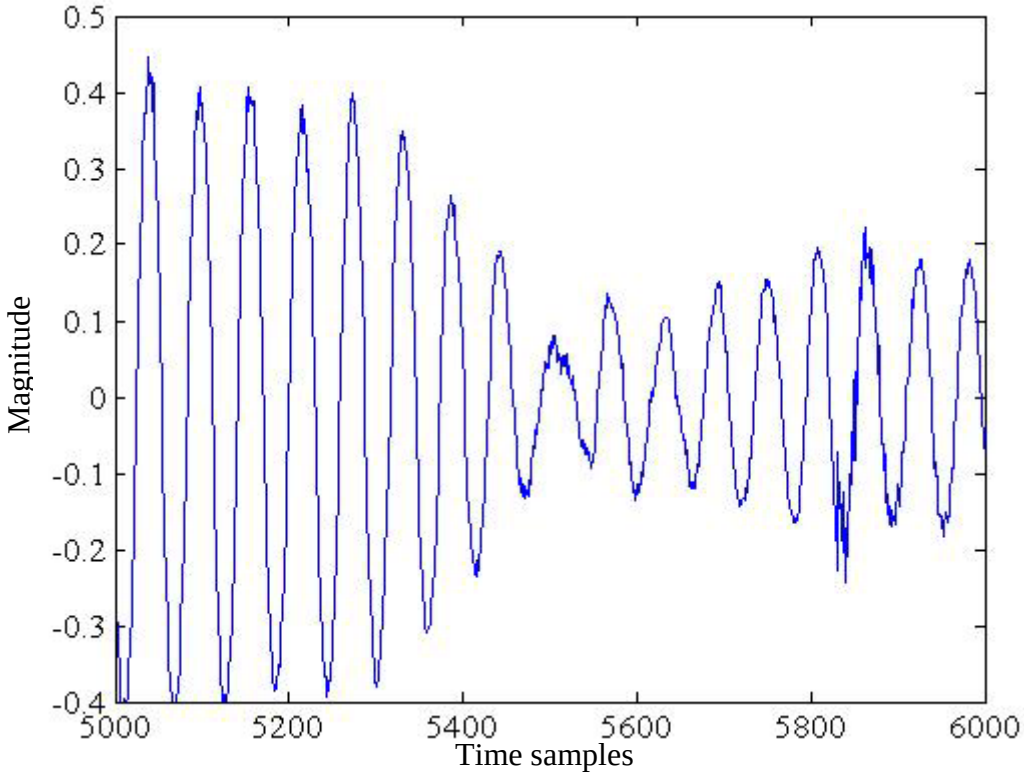

Figure 5.3(b): Compressed masker

These waveforms are almost similar in shape. However, it can be seen that the difference in the amplitudes of the peaks is lesser in case of the compressed waveform. Once it is ensured that the maxima of the compressed signal are the same as that in the original, the effect is that the lower magnitude signals are amplified while the higher magnitude signals remain unaffected. Thus, the number of low-amplitude segments has reduced, and the problem of low-amplitude periods of the masker has been solved. Invariably, due to compression, there might be a slight drop in the quality of the masker. However, flexibility in deciding the threshold level and the compression ratio is available to the user to achieve the desired quality of the masker.

5.2 Time-varying secondary path and online modeling

So far, it has been assumed that the secondary path characteristics are time-invariant, and therefore an offline modeling method can be used for identification purposes and the resulting filter can be used in the online ANC operation [4]. However, it is difficult to assume that in a real life scenario the secondary path is not going to vary with time. Thus the offline estimate no longer accurately represents the secondary path. This is likely to lead to degraded performance or instability. To illustrate this characteristic, secondary path variation is simulated and its result on the system performance is observed. A random variation of not more than 0.1 percent per sample is carried out in the secondary path coefficients. Figure 5.4(a) and 5.4(b) illustrate the difference in the magnitude

response of the original secondary path and the change in the secondary path after 4000 samples.

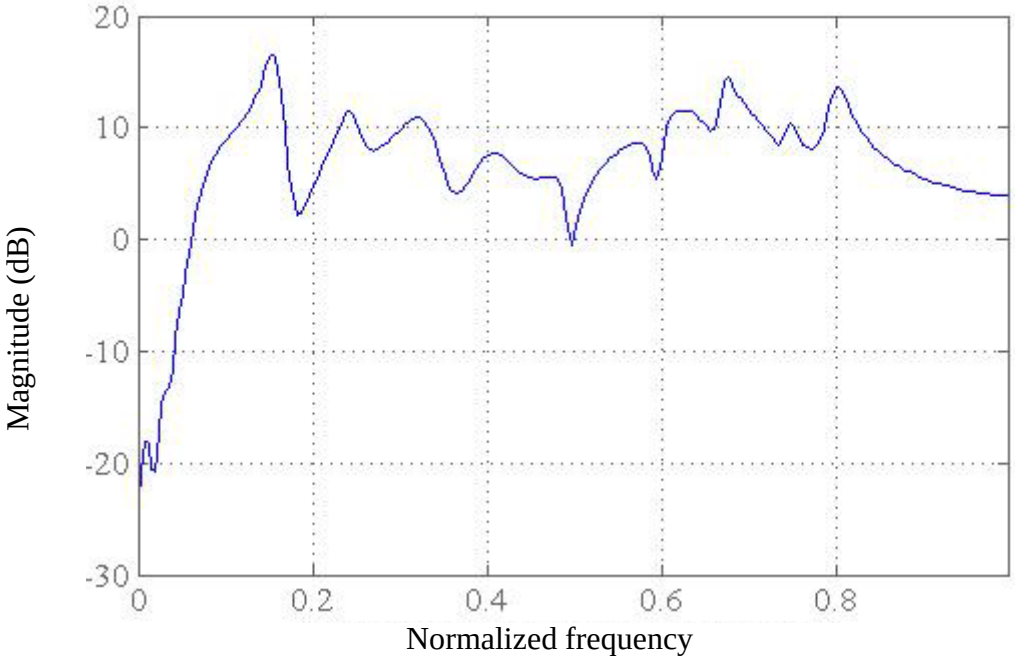

Figure 5.4(a): Magnitude response of the original secondary path

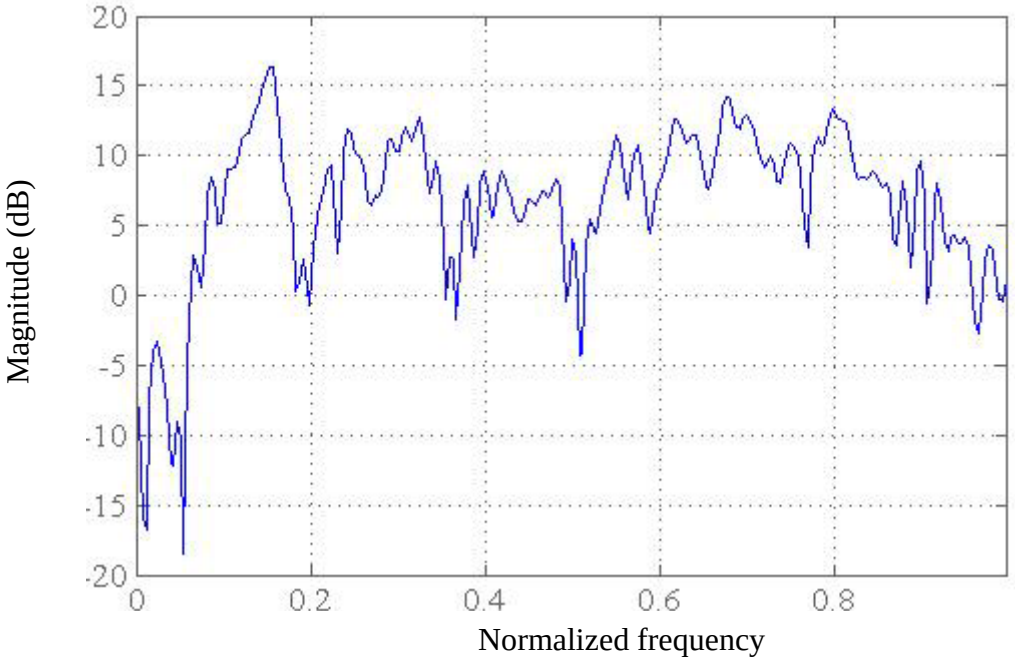

Figure 5.4(b): Magnitude response of the secondary path after 4000 samples

The time-varying secondary path shown in Figure 5.4(b) is used for simulation in the system, whereas the estimate of the secondary path remains the same as used earlier. Figure 5.5 shows the system output when this situation occurs. The first sample on x-axis represents the start of the time when the secondary path starts to vary. It can readily be observed how the change in secondary path is affecting the system performance. The system output starts to diverge after about 2200th sample. This amply illustrates how the ANC system fails to converge when the secondary path varies with time. Continual estimation of the secondary path in parallel to the operation of the ANC system is therefore required for a stable system [5].

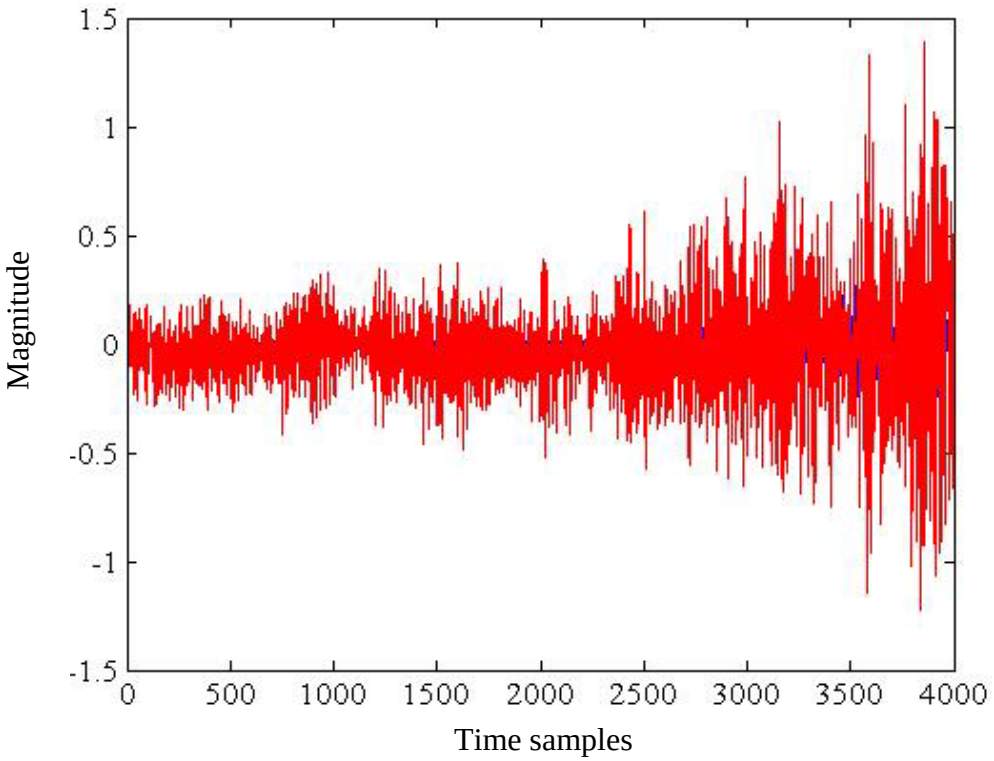

Figure 5.5: Response of the system with a time-varying secondary path.

A solution to the above problem was first proposed by Eriksson [61], using the additive random noise technique. However, adding random noise to the system increases

uncertainty and therefore leads to degradation of the adaptive filter. Many studies [62]-[64] are available on this subject and online modeling of secondary path may be considered a field by itself and a topic for further research. Another approach formulated by Widrow and Stearns [65] can also be used to solve this problem. In this scheme, the control signal also serves as an excitation signal for secondary-path modeling. The coefficients of the adaptive filter are adjusted online to model continuously the secondary path during the operation of the ANC filter. Figure 5.6 illustrates the use of the second technique for online modeling of the changes in secondary path. The performance of the ANC system depends on the efficiency of the online modeling of physical system changes.

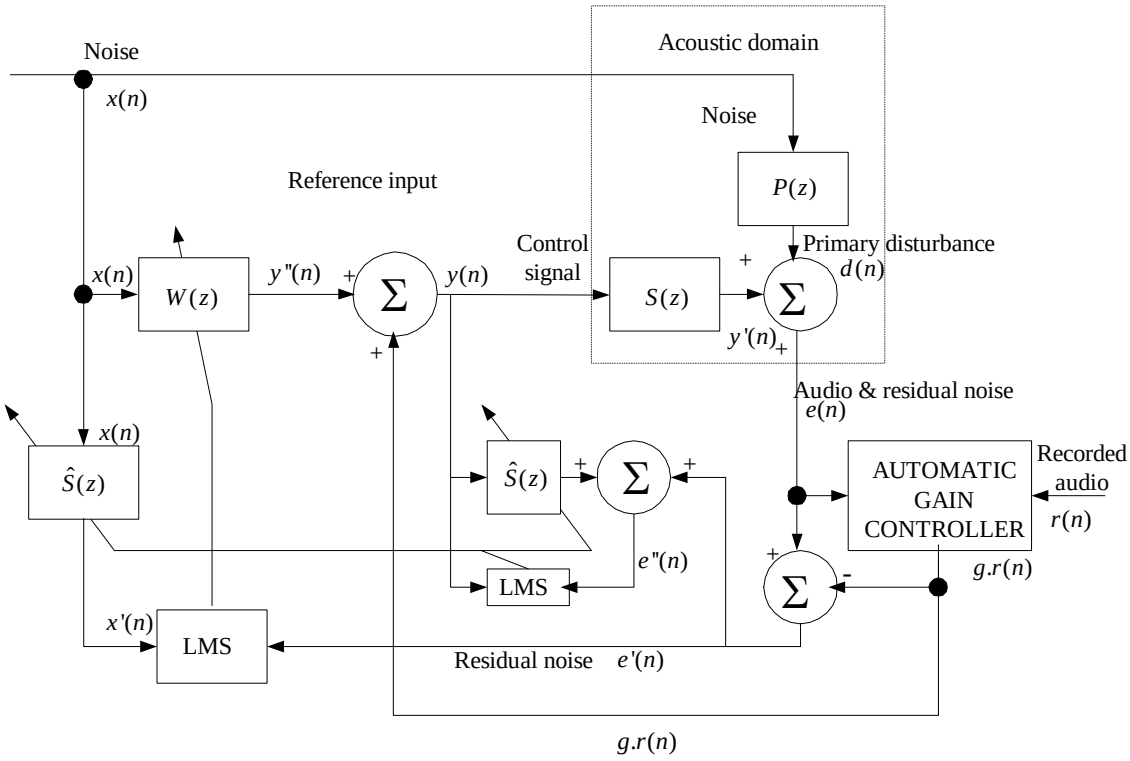

Figure 5.6: Online modeling of secondary path

The developed system is tested with a time-varying secondary path. The online adaptive filter for the secondary path is designed with a filter length of 256 and a step size of 0.04. The maskee is a recording of engine noise and the masker is a classical music. As the system is only a slight modification over the IANCM system of Figure 4.6, there are only two additional equations. The online modeling error $e''(n)$ is the result of the superposition of the control signal passing through the filter, and representing the estimate of the secondary path and the residual noise $e'(n)$. Hence we have,

$$e''(n) = e'(n) + y(n) * \hat{s}(n). \tag{5.1}$$

Thus, the modeling filter is updated by the equation

$$\hat{s}(n+1) = \hat{s}(n) - \mu_m e''(n) y(n). \tag{5.2}$$

Here, μ_m is the step size used for the modeling filter update and the other notations represent their usual meanings as mentioned earlier in Chapter 4.

Figure 5.7 shows the magnitude response of the secondary path and its estimate. It can be seen that the secondary path estimate is close to the actual secondary path. However, one needs to be careful that, it may not be possible for the system to always keep track of the secondary path changes as the online modeling part is not isolated from the ANC controller part of the system and the two mutually affect each other. This is an inherent limitation of the system and an area for further research.

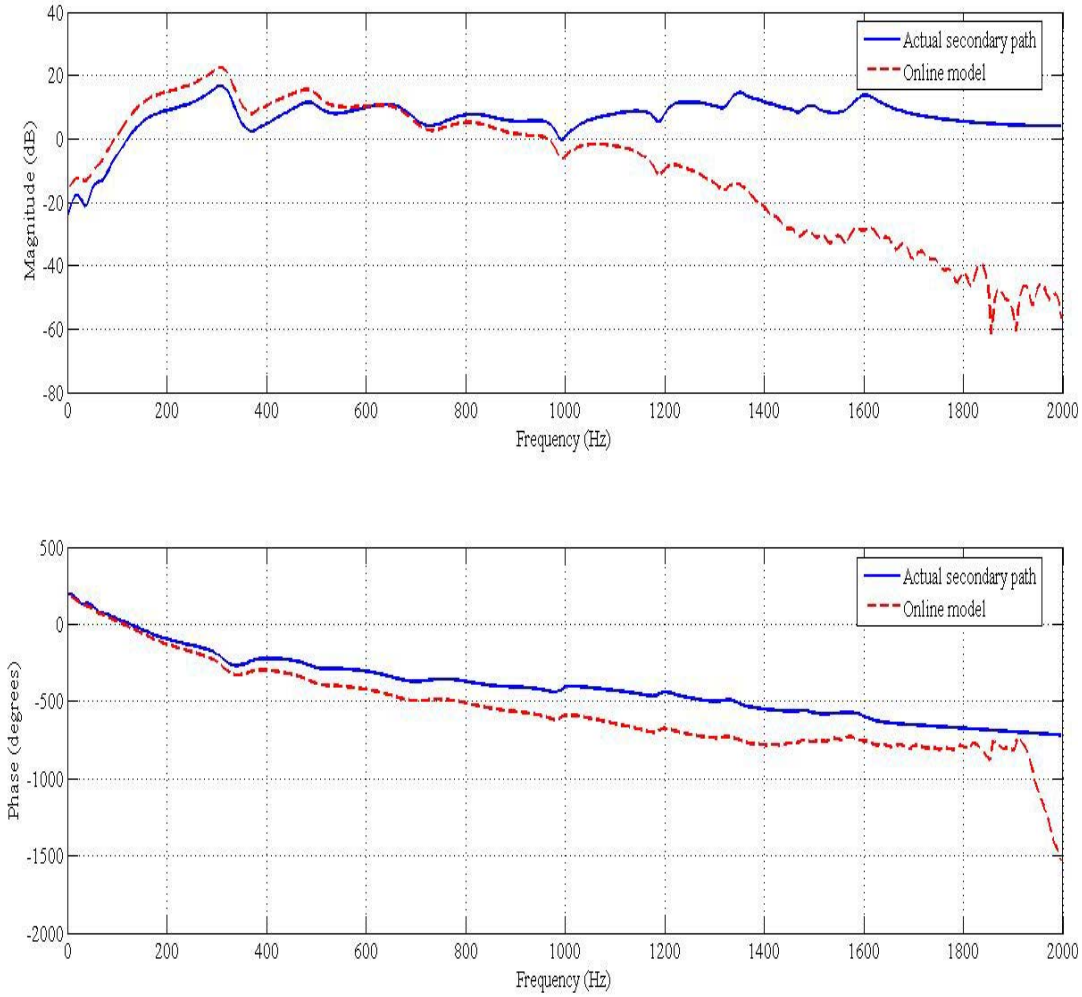

Figure 5.7: Magnitude and phase response of the secondary path and online model

The system is only able to cater for gradual and incremental changes in the secondary path and sudden fluctuations are likely to cause system instability. It can be seen from Figure 5.7 that there is considerable difference between the actual and modeled secondary path for frequencies above 1 KHz. Higher sampling rates of the system may be helpful in alleviating this problem to some extent. Thus, there is a requirement of further improvements in this field if the system has to perform effectively in situations likely to

encounter sudden changes in the secondary path. Figure 5.8 demonstrates the effect of secondary path change on system performance. The sound pressure levels of the system output show a marked increase when the secondary path change is introduced and the system takes quite some time to stabilize. However, it is worth noting that even with the performance degradation, the system output is still below the original noise levels.

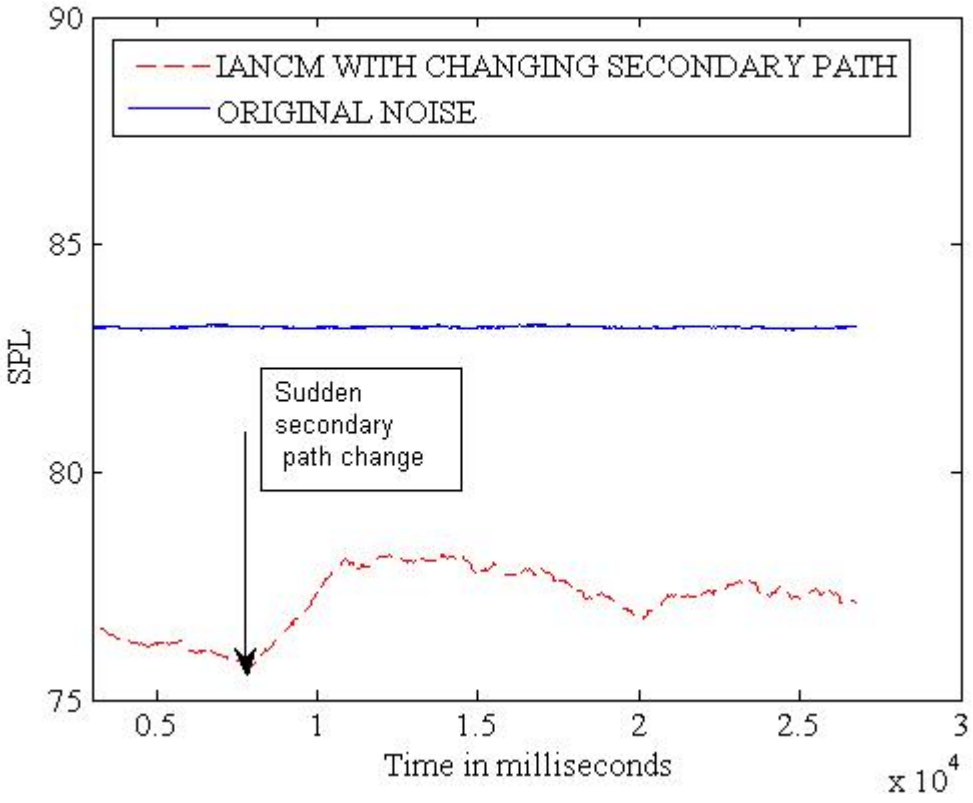

Figure 5.8: Effect of secondary path change on IANCM system performance

5.3 Time-varying maskers to improve the listening experience

One more problem which needs to be tackled in the system is that the unpleasant noise which is sought to be masked is continuously present in the environment. However, the

maskers used are of limited duration. Thus, we have the option of either repeating the masker or changing over to another masker on completion. Studies on human psychology [53], [66] have revealed that continuous and repeated exposure to even a preferred sensation may lead to dislike. This can be easily understood as repeated listening to the same song may put one off over a period of time. Therefore, the second option of changing the maskers one after another needs to be explored. Changing the masker drastically defeats the very purpose of selecting the suitable masker initially. Hence, there is a need to find a way of removing this repetitiveness without changing the nature of the masker too much. We need to gradually switchover to a different masker which is 'close' to the initial masker. However, defining two maskers as 'close' is a difficult task and brings us into the realm of computational intelligence [67]. Appreciation of music is a difficult task even for human beings who are not musically trained or do not have a natural flair for music. This makes us realize the complexity of the task of measuring the 'closeness' of two maskers by a machine. There is a need to make use of the statistical tools already available and define certain features of the maskers which can be extracted automatically and can be used for classification of the maskers. Recently many automatic audio classification schemes have been proposed and audio classification is an active research area [66], [67]. For machine based classification of the maskers, first the requisite features need to be selected. The features commonly used in the acoustic field [68] include; spectral centroid, which is a measure of the centre of the power spectrum of the audio signal. It is computed by finding the mean power spectrum frequency bin and then expressing the index of the bin over the total number of bins; spectral variability which is the standard deviation of the signal's power spectrum; root mean squared power

of the signal, and zero crossings, which counts the number of times within the sample window in which there is a change in the sample sign. Millions of such features can be defined and used for the classification process, however many of these features are irrelevant, that is, they contain no information about the masker and many other features are dependent upon other key features. Thus, we need to carry out the feature extraction and retain only the most relevant features, as processing information for a large number of features becomes practically impossible. Some important features for classification, identified by current studies [69]-[71] to be most effective are psychoacoustic features such as roughness, loudness and sharpness. Roughness is the perception of temporal envelope modulations in the range of about 20-150 Hz and is maximal for modulations near 70 Hz. The estimate of roughness, is based on mid-frequency temporal envelope modulations, and an accurate estimate can only be obtained for relatively long audio frames (> 180 msec) [70]. Loudness is the sensation of intensity and sharpness is a perception related to the spectral density and the relative strength of high-frequency energy. Therefore, the system should be able to extract the values of these relevant features from a large pool of maskers available and process them to arrive at the closeness of the maskers with respect to each other.

Once the most important features have been identified, various maskers are classified into clusters or groups [69]. It can now be reasonably assumed that maskers falling within one cluster will be reasonably close to each other and transition from one masker to another within the same cluster will be smooth and pleasant for the listener.

One of the most popular algorithms used for clustering is the K-means algorithm which follows the following simple steps [72]:

(I) Place K points arbitrarily into the space represented by the maskers. These points represent initial group centroids.

(II) Assign each masker to the group that has the closest centroid.

(III) When all maskers have been assigned, recalculate the positions of the K centroids.

(IV)Repeat Steps (II) and (III) until the centroids no longer change.

Mathematically, we minimize the function [72]

$$J = \sum_{j=1}^{k} \sum_{j=1}^{n} || x_i^{(j)} - c_j ||^2, \tag{5.3}$$

such that $|| x_i^{(j)} - c_j ||^2$ is a distance measure between a data point $x_i^{(j)}$ and cluster centroid c_j. In this manner maskers with similar properties are grouped together within a cluster. Clustering is a very resource demanding process and should be done offline prior to the system operation. Figure 5.9 generated using Matlab illustrates the results of clustering by hierarchical clustering with Euclidean distance metric and average linkage.

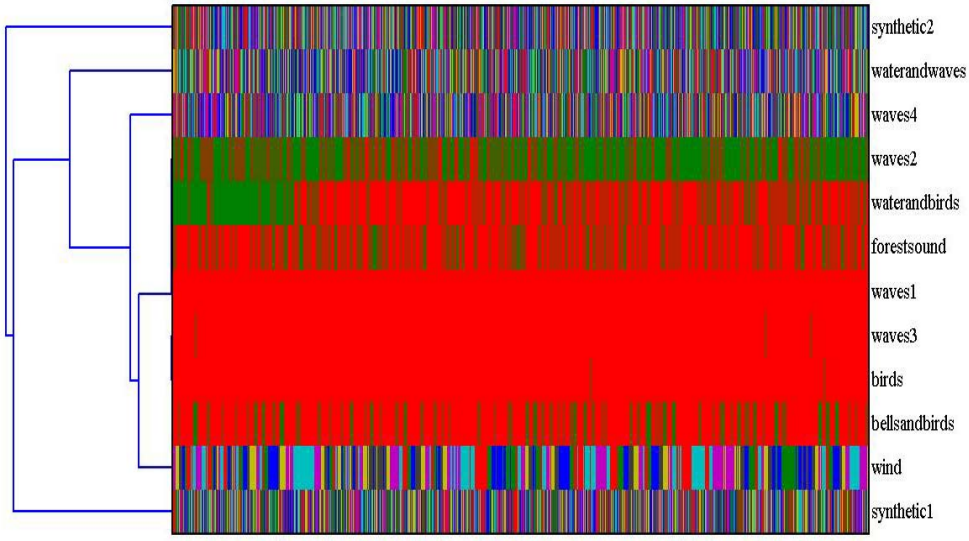

Figure 5.9: Clustergram of some maskers

A limited number of maskers are clustered together as seen in the dendrogram. Of course, there are several other methods of clustering as well, such as self-organizing maps [73], or probabilistic clustering methods [69], however deciding the best method of clustering is outside the scope of the current work.

Alternatively, in case the database of maskers is small enough to be manually manageable then the maskers of similar type can be grouped together and these groups can be segregated from each other. When the system selects any of the maskers from the entire database then the maskers of that particular group are used randomly one by one in the system. This ensures that there are no abrupt changes in the masker type.

In order to ensure that one does not have to start from scratch, there are some readily available tools [74], [75] to classify the maskers into groups. As this grouping is done based on spatial proximity of certain important features of the maskers, it is highly probable that the maskers within a group will be perceived as similar and there will be a

smooth transition from one masker to another. One such widely popular tool is Musicminer [75]. This tool uses signal processing and time series analysis methods to find numerical features that describe audio files with music. This software uses certain selected features and self-organizing maps **(**SOM**)** to visualize a music collection analogous to geographical maps. Similar songs are placed together in valleys; large differences between songs are shown as mountain ranges in between. A screenshot of the audio mapping of 88 different types of songs from its database is shown in Figure 5.10. It can be presumed that songs which are close to each other and not separated by any high ranges are similar and can help achieving a smooth transitioning of the masker. However, this technology is still in its infancy and needs extensive work and subjective confirmation prior to being integrated into the IANCM system. Due to the uncertainties still existing about the human auditory system it is not possible to ascertain the pleasantness of a signal without subjective tests. Different tools available commercially or freely can be compared against each other by large scale subjective evaluation to check their relative performance and the results can be used to decide suitable algorithms for incorporating into the IANCM system.

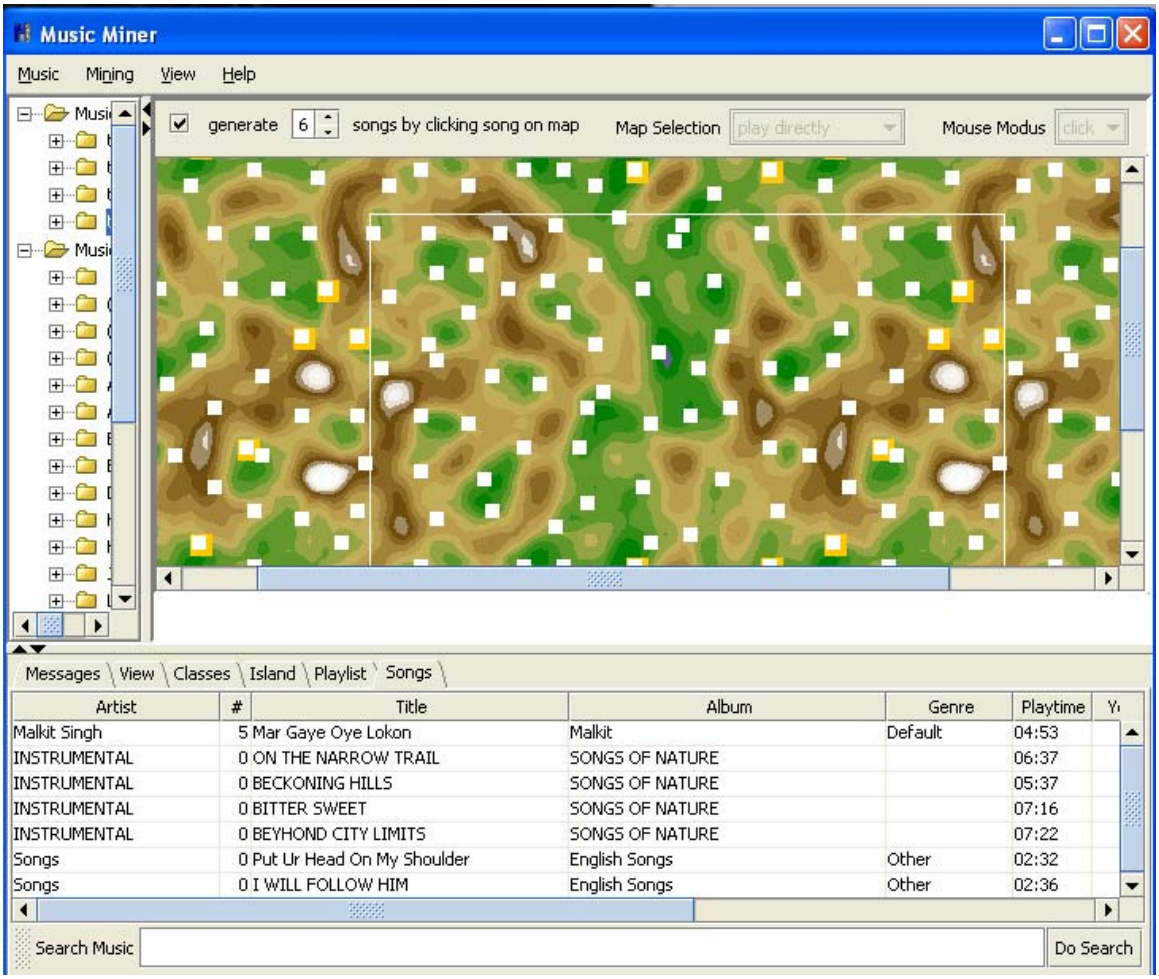

Figure 5.10: Mapping of audio files using Musicminer [75]

5.4 Conclusions

In this chapter, the shortcomings of the IANCM system have been examined and a few solutions have been proposed. In order to cater for the low intensity periods of the masker, the signal was compressed and the results show a decrease in the dynamic range

of the output signal. This technique has been successfully used by radio broadcasting stations [60] and can be easily incorporated into this system.

Further, the problem of instability of the system when faced with a time varying secondary path was highlighted. This area of ANC is still witnessing ongoing research and few available solutions such as additive random noise technique inject new uncertainties in the system. The problem of isolating the secondary path identification filter from the rest of the system needs some innovative solutions and can be considered for future research. One such innovative idea has been proposed in this chapter, however the problem of effectively isolating the secondary path modeling stage from the rest of the system still persists and substantial improvisation on the proposed model needs to be done prior to its effective implementation.

Finally, an innovative solution was proposed for smooth transition from one masker to another. This is necessary as the consumer may not like the same masker to be played repeatedly and an abrupt change from one masker to another may cause the system to perform sub-optimally. Suitable commercially available software has also been identified for segregation of the maskers into groups such that transition from one masker to another within the same group is smooth and aesthetically pleasing to the listener. Alternatively, using any of the available clustering techniques, the available maskers can be grouped such that maskers within a group are similar to each other and different from those in other groups.

Chapter 6

Conclusions and Future work

This chapter presents a summary of the work done so far and suggests some future work. In particular, Section 6.1 highlights the capabilities and shortcomings of the IANCM system, and Section 6.2 suggests some future work and possible improvements to the system.

6.1 Conclusions

This thesis presents innovative methods to expand the domain of ANC systems. Psychoacoustics has been employed in conjunction with ANC techniques to overcome the shortcomings of conventional ANC system. A pleasant sound at appropriate levels has been used to mask the residual noise of an ANC system. This gives us a pleasant residual noise at reasonably low loudness levels. Thus we are able to avoid the unpleasant residual noise of an ANC system and the loud decibel levels of a masker.

A new method for selecting suitable maskers from an available database of maskers has been introduced in this thesis. Thus, it is possible to rank the maskers objectively without human intervention. This property is useful while handling large databases. Furthermore, this method ensures that the selected masker is able to mask the ANC residue at the minimum possible intensity levels. This is done by ensuring that the GMT of the masker is just above the SPL of the maskee. This global threshold helps in

keeping the overall decibel levels low. The practically proven, psychoacoustic model I has been employed for this task, with a few modifications to keep the implementation simple.

The IANCM system has been shown to be effective in suppressing narrow band engine noises by using a combination of ANC and a suitable masker. A wide range of maskers ranging from sounds occurring in nature to musical compositions have been employed in simulations and the system robustness has been ascertained. In addition to the masking function, suitable maskers such as classical music composition extracts can also be used to perform offline secondary path modeling. The ANC system employing a secondary path estimate, obtained by using classical music as a training signal, has performed equally well as the conventional ANC system. This eliminates the need to use white noise as a training signal, thereby avoiding the hiss-like sound prevalent during the training process of a conventional ANC system.

The IANCM system, in addition to dealing with a narrowband noise, has the adaptability to withstand changes in the characteristics of the maskee. The system responds to changes in the maskee by automatically changing the gain of the masker and adjusting the masker levels at a suitable point. Simulation results of the system with even a broadband noise source, such as the cabin noise of a moving car at various speeds have been successfully demonstrated. In addition, the practical utility of the system has been highlighted by using the IANCM system for various household appliance noise scenarios. This includes vacuum cleaner, hand-blender and air-conditioner noise, to name a few.

In addition to masking, another psychoacoustic characteristic of reverberation has also been employed by the system. A fundamental frequency appears to get louder due to its harmonics, even though the decibel levels do not change. For example, a snore sound

which has its dominant frequency at 80 Hz can be effectively masked with relatively lower level of Tibetan chants as the dominant energy of the chants lies in spectral regions which are integral multiples of 80 Hz.

The IANCM system has been made more robust by reducing the dynamic range of the masker. Dynamic range reduction has been achieved by simple compression techniques. This ensures that the levels of the maskers are not too loud and the quieter periods of the maskers do not fall below the maskee levels.

Online modeling of the secondary path has proved to be a difficult task as currently there is no fully effective method [5], [61] for isolating the secondary path variations from the changes in the primary disturbance. This task needs further work and it is an exciting area for research.

An approach to avoid the monotony of a single masker being played repetitively has also been discussed in the thesis. A smooth transition from one masker to another may be achieved by using statistical tools. Maskers may be clustered together based on their physical properties using any of the clustering techniques, such as K-means clustering. Clustering of a few selected maskers has been demonstrated for illustration purpose in Chapter 5.Prevalent feature extraction methods can be used from data mining techniques to decide the relevant features such as loudness, sharpness and roughness amongst various other such features.

6.2 Future work

A few suggestions have been included in this section which could be used for future work on this topic.

6.2.1 Subjective validation of classification and clustering

The techniques discussed so far have been implemented using Matlab simulations and evaluated by a small group of persons. For effective validation, the techniques such as masker classification, feature extraction of maskers and testing of maskers for closeness with respect to each other need to be tested against subjective evaluation by a large number of evaluators. A set of more than 100 persons spread over various age groups and from diverse backgrounds should be used for the subjective evaluations. The results obtained should be compared and correlated with objective methods, thereby ascertaining the effectiveness of the objective methods

6.2.2 Online secondary path modeling

As has been highlighted in earlier chapters; in ANC control systems, the effect of the secondary path needs to be compensated for achieving a stable system. The secondary path can be modeled offline, prior to the operation of an ANC controller. In many practical cases, however, the secondary path can be time-varying, so online modeling is required to ensure the convergence of an ANC system. The most commonly used approach for online secondary path modeling injects an additional noise into the system.

An additive noise is useful to the secondary path modeling process, but it introduces extra noise and increases uncertainty in the operation of the ANC controller. The approach by Widrow and Stearns [65] uses the control signal as an excitation signal for secondary-path modeling. The coefficients of the adaptive filter are adjusted online to model continuously the secondary path during the operation of the ANC filter. However, as seen in Chapter 5, the performance of such systems is poor at frequencies above 1 kHz. Further the system is unable to always keep track of the secondary path changes as the online modeling part is not isolated from the ANC controller part of the system and the two mutually affect each other. Thus, we have the contradictory requirements of increasing the accuracy of the secondary path modeling and the effective adaptation of the ANC controller. Any possible improvements in the IANCM system have to take this factor into account and an innovative solution to solve this problem is required. This field thus has tremendous opportunity for a breakthrough, so that a stable system can be obtained, which can track the changes in secondary path without affecting the ANC controller. However, finding an innovative and effective solution for online modeling of secondary path is beyond the scope of this thesis and further studies and experiments are required to come up with a solution to this problem. This may be taken up as a topic for further research.

6.2.3 Use of ultrasonic components

It is common understanding in psychoacoustics that humans are not able to hear sounds with frequencies above 20 kHz. [76]. The outer and the middle ear act like a filter to

attenuate the ultrasonic frequency components (UFCs) so that the brain does not receive any stimulus. However, studies have proven that if the stimulus is provided to the brain by bypassing the outer and middle ear, then the brain responds to it and is even able to achieve frequency discrimination to the same level as those of audible frequency [77].

Recent research has revealed that though UFCs may not be audible on their own, they have an impact on the quality of the perceived sound in conjunction with the audible frequency [77]. Subjective tests have revealed that people preferred music containing UFCs, over the same music with only audible frequency. Medical imaging has revealed that these UFCs induce a meditative state in the listeners, clinically confirmed by the release of alpha waves in the brain [78]. Thus, in this manner, there is a possibility to provide a better listening experience to the user and overcome the deficiencies of the existing system.

References

[1] P Leug, *Process of silencing sound oscillations*, US. Patent No 2,043,416, 1936.

[2] KD Kryter, *The Handbook of Hearing and the Effects of Noise: Physiology, Psychology and Public Health*, Academic Press, 1994.

[3] D Guicking, Active *Noise Control- A Review Based on Patent Applications*, Proceedings Noise-93, p l53-158, 1993.

[4] SM Kuo and DR Morgan, *Active Noise Control Systems*, Wiley Series, 1996.

[5] SM Kuo and DR Morgan, *Active Noise Control: A Tutorial Review*, Proceedings of the IEEE, Vol. 87, No. 6, June 1999.

[6] L Boney, AH. Tewfik and KM Hamdy, *Digital Watermarks for Audio Signals*, IEEE Int. Conf. on Multimedia Computing and Systems, Hiroshima, Japan, pp. 473-480, 1996.

[7] LW Hinderks, *Method and Apparatus for Adaptive Power Adjustment of Mixed Modulation Radio Transmission*, US Patent: 5,301,363, April, 1994.

[8] D Tsoukalas, M Paraskevas and J. Mourjopoulos, *Speech Enhancement using Psychoacoustic Criteria*, IEEE Trans. Acoust., Speech, Signal Processing, vol. 2, pp. 359-362, April 1993.

[9] ISO/IEC 11172-3:1993, Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio

[10] CH Hansen and SD Snyder, Active *Control of Noise and Vibration*, E&FN 1997.

[11] HF Olson and EG May, *Electronic Sound Absorber*, The Journal of the Acoustic Society of America, Volume 25, pp.1130-1136, 1953.

[12] B Widrow, *Recollections of Norbert Wiener and the first IFAC world congress*, IEEE Control systems magazine, Vol 21(3) pp. 65-70, Jun 2001.

[13] A Feuer and E Weinstein, *Convergence Analysis of LMS filters with uncorrelated Gaussian data*, IEEE trans. on Acoustics, Speech and Signal Proc. Vol 33(1), pp. 222-230, Feb 1985.

[14] www.mathworks.com/products/matlab.

[15] CF Ross and MRJ Purver, *Active cabin noise control*, Active 97, Budapest, Hungary, Aug. 21–23, 1997.

[16] I Kakuhari, K Terai, Y Nakamura, T Inoue and H Sano, *Development of active control system for low frequency road noise*, Proc. Acoust. Soc. Japan., 2000.

[17] DG Zimcik, RM Provencher, R Lecheminant and DAJ McNamara, *Measurement of Noise in Armored Personnel Carriers*, National Research Council of Canada Ottawa (Ontario) Inst for Aerospace Research, 2002.

[18] WA Yost and DW Nielsen, *Fundamentals of Hearing*, Holt, Rinehart and Winston, New York, 1985.

[19] GJ Borden, and KS Harris, *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*, Williams and Wilkins, Baltimore, 1980.

[20] M Bosi, *Introduction to Digital Audio Coding and Standards*, Kluwer Academic Publishers, 2003.

[21] H Wallach, EB Newman and MR Rosenzweig, *The precedence effect in sound localization*, The American Journal of Psychology, pp.315-336, 1949.

[22] J Backus, *The Acoustical Foundations of Music*, W.W. Norton and Company 1977.

[23] BCJ Moore and BR Glasberg, *A revision of Zwicker's loudness model*, Acta Acustica, vol. 82, pp. 335–345, 1996

[24] E Zwicker and H Fastl, *Psychoacoustics: facts and models*, Springer series in informational sciences, 1999.

[25] JD Johnston, *Estimation of Perceptual Entropy Using Noise Masking Criteria*, Proc. ICASSP '88, pp. 2524-2527, 1988

[26] M Ross, and T Giolas, *Effects of three classroom listening conditions on speech intelligibility*, American Annals of the Deaf, *116*, 580–584,1971.

[27] M Hodgson, R Rempel and S Kennedy, *Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures*, The Journal of the Acoustical Society of America 105, pp.226-233, 1999.

[28] IEEE Standard 269, *IEEE Standard Methods for Measuring Transmission Performance of Analog and. Digital Telephone Sets*, 1992.

[29]. D.F. Hoth, *Room noise spectra at subscriber's telephone location*, Journal of Acoustical Society of America, pp.499-504, 1941.

[30] P Bricker, J Flanagan, *Subjective assessment of computer-simulated telephone calling signals*, IEEE Transactions on Audio and Electro acoustics , Volume 18, Issue 1, pp 19-25, Mar 1970.

[31] RM Hunt, *Determination of an Effective Tone Ringer Signal,* 38th Audio Engineering Society Convention, 1970.

[32] RB Archbold, A Ithell and GT Johnson, *The Ideal Characteristics for the Calling Signal of a Subscriber's Telephone Set*, Research Report No. 2l143, Post Office Research Station, Dollis Hill, London, 1967.

[33] JA Ballas, *Common factors in the identification of an assortment of brief everyday sounds*, J. Exp. Psychology: Human Perception and Performance, pp 250-267, 1993.

[34] LN Solomon, *Search for physical correlates to psychological dimensions of sounds*, J. Acoust. Soc. Am 31, pp. 492-497, 1959.

[35] DL Halpern, R Blake and J Hillenbrand, *Psychoacoustics of a chilling sound*, Percept. Psychophysics, 39(2), 77-80, 1986.

[36] J McDermott and M Hauser, *Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate,* Cognition, 94(2), B11-B21, 2004.

[37] G Kidd and CS Watson, *The perceptual dimensionality of environmental sounds*, Noise Control Engineering Journal, 51, 216-231, 2003.

[38] C Lavandier and B Defreville, The contribution of sound source characteristics in the assessment of urban sound-scapes, Acustica 92, 912–921, 2006.

[39] BCJ Moore, *An Introduction to the Psychology of Hearing*, Elsevier Academic Press, 2003.

[40] MT Smith, *Audio Engineer's Reference Book*, Focal Press, 1999.

[41] KL Gee, SH Swift**,** VW Sparrow, KJ Plotkin, JM Downing, *On the potential limitations of conventional sound metrics in quantifying perception of nonlinearly propagated noise***,** Vol 121(1) JASA Express Letters, pp.EL1-EL7, Jan 2007.

[42] International Telecommunication Union- Radio standard 468.

[43] CCIR-468-1 recommendations.

[44] MK Mandal, *Multimedia Signals and Systems,* Springer 2002.

[45] RA Garcia, *Digital Watermarking Of Audio Signals Using A Psychoacoustic Auditory Model And Spread Spectrum Theory*, MS Thesis, University of Miami, 1999.

[46] F Petitcolas, *MPEG Psychoacoustic Model I for Matlab,* article from www.cl.cam.ac.uk

[47] J Boley, *Psychoacoustical Model 2 in Matlab,* article from www.perceptualentropy.com.

[48] WS Gan and SM Kuo, *An Integrated Audio and Active Noise Control Headsets*, IEEE Trans. Consumer Electronics, 48(2), 2002.

[49] S Gustafsson, PJ Martin, and P Vary, *A Psychoacoustic Approach to Combined Acoustic Echo Cancellation and Noise Reduction,* IEEE Trans .Speech and Audio Processing,10(5), 2002.

[50] S Singaraju, *Noise Masking Using Psychoacoustics*, MS Thesis, Northern Illinois University, 2006.

[51] R Gireddy and SM Kuo, *Real-Time Experiment of Snore Active Noise Control*, IEEE Int. Conf. on Control Applications, Oct. 2007.

[52] EM Verveer, H Barry and WA Bousfield, *Change in Affectivity with Repetition,* The *American Journal of Psychology*, Vol. 45, No. 1, pp. 130-134. 1933

[53] GF Stout, *A Manual of Psychology*, University Correspondence College Press, 1899.

[54] D Griesinger, *How loud is my Reverberation*, Audio Engineering Society, 98th Convention 1995.

[55] www.sonicvisualiser.org

[56] B Seaward, *Managing Stress: Principles and Strategies for Health and Well-being*, Jones & Bartlett Pub – 2004.

[57] SJ Elliott and PA Nelson

[58] SJ Elliott and PA Nelson, *Active Noise Control*, IEEE Signal Processing Magazine, pp. 12-35,Oct 1993

[59] H Dezyk and E Slater, *Consumer perceptions,* www.abi.org.uk.

[60] GW McNally, *Dynamic Range control of Digital audio signals,* BBC research department report RD1983/17, 1983.

[61] LJ Eriksson and M Allie, *Use of random noise for on-line transducer modeling in an adaptive active attenuation system*, J. Acoust. Soc. Am., vol. 85, no. 2, pp. 797-802, 1989.

[62] MT Akhtar, M Abe and M. Kawamata, *A New Variable Step Size LMS Algorithm-Based Method for Improved Online Secondary Path Modeling in Active Noise Control Systems*, IEEE Trans. Audio Speech Lang. Proc., vol. 14, no. 2, pp. 720-726, 2006.

[63] M Zhang, H Lan and W Ser, *A Robust Online Secondary Path Modeling Method with Auxiliary Noise Power Scheduling Strategy and Norm Constraint Manipulation*, IEEE Trans. on Speech and Audio Proc. Vol 11-1,pp 45-53, 2003.

[64] M Zhang, H Lan and W Ser, *Cross-updated Active Noise Control System with Online Secondary Path Modeling*, IEEE Trans. Speech Audio Proc., vol. 9, no. 5, pp. 598-602, 2001.

[65] B Widrow and SD Stearns, *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1985.

[66] www.music-ir.org

[67] C McKay, D McEnnis and Ichiro Fujinaga, A *Large Publicly Accessible Prototype Audio Database for Music Research*, Proceedings of the International Computer Music Conference, 2006.

[68] JB MacQueen, *Some Methods for classification and Analysis of Multivariate Observations*, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297, 1967

[69] TM Breuel, *Classification by probabilistic clustering*, IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.2 pp 1333 - 1336, 2001.

[70] Sabin and Schoenike, *HF Radio Systems & Circuits*, Noble, pp. 13-25, 271-290, 1998.

[71] JM Francos, AZ Meiri and B Porat, *A unified texture model based on a 2-D Wold-like decomposition*, IEEE Trans. Signal Processing, Vol. 41, 1993.

[72] JA Hartigan and MA Wong, *A K-Means Clustering Algorithm*, Applied Statistics 28 (1): 100–108. 1979.

[73] T Kohonen, *Self-Organizing Maps*, Springer, 2001

[74] www.soundfisher.com/html/overview.html

[75] www.musicminer.sourceforge.net/research.html

[76] S Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, 1997.

[77] T Oohashi, N Emi, M Honda, Y Yonekura, Y Fuwamoto, N Kawai, T Maekawa, S Nakamura, H Fukuyama and H Shibasaki, *Inaudible High-Frequency Sounds Affect Brain Activity: Hypersonic Effect.* J. Neurophysiol. 83: 3548-3558, 2000.

[78] T Iwaki, M Hayashi, and T Hori, *Changes in alpha band EEG activity in the frontal area after stimulation with music of different affective content*, Percept Mot Skills 84: 515–526, 1997.

Publications

[1] A Rajora, WS Gan, SM Kuo and YK Chong, *Noise Suppression using ANC and Masking*, 14th International Congress on Sound and Vibration, 2007.

Appendix A

Classification of Maskers

NV-Natural sound with variable envelope
NC-Natural sound with constant envelope
SC-Synthetic Noise with constant envelope
SV-Synthetic noise with variable envelope
MV-Music with variable envelope
MC-Music with constant envelope

Serial Number	Maskee (Original noise)	Residual Noise	Masker description	Type of Masker	Masker + Residual Noise (Processed signal) Gain=1	Masker + Residual Noise (Processed signal) Softer due to lower 'g'	Rank using perception	Rank using algorithm
1			Waves lapping shore	NV	Listen	Listen	11	24
2			Waves 2	NV	Listen	Listen	21	16
3			Waves 3	NV	Listen	Listen	20	18
4			Waves 4	NC	Listen	Listen	22	23
5			Bells and birds	MC	Listen	Listen	8	19
6			Birds	NC	Listen	Listen	9	12
7			Forest Sound	NC	Listen	Listen	12	3
8			Water and birds	NV	Listen	Listen	10	2

9	Engine	Residual	Wind and waves	NC	Listen	Listen	23	1
10			Wind	NC	Listen	Listen	30	7
11			Synthetic1	SC	Listen	Listen	29	4
12			Synthetic2	SC	Listen	Listen	28	5
13			Frozen	MC	Listen	Listen	8	6
14			African music	MV	Listen	Listen	7	11
15			Air Conditioner	SC	Listen	Listen	31	22
16			Rain	NC	Listen	Listen	27	15
17			Die another day	MC	Listen	Listen	15	20
18			The river 1	NV	Listen	Listen	14	26
19			The river 2	NV	Listen	Listen	13	9
20			Its my life	MC	Listen	Listen	19	8
21			Always	MV	Listen	Listen	20	21
22			Say	MC	Listen	Listen	18	13
23			Live	MC	Listen	Listen	17	14
24			Crickets	NC	Listen	Listen	25	27
25			White noise	SC	Listen	Listen	26	10
26			Classical 1	MC	Listen	Listen	2	30
27			Classical 2	MC	Listen	Listen	3	31
28			Classical 3	MV	Listen	Listen	4	25
29			Classical 4	MC	Listen	Listen	5	29
30			Classical 5	MC	Listen	Listen	6	17
31			Classical Long	MV	Listen	Listen	1	28

Appendix B

List of audio files

Serial number	Name	Description
1	Engine1.wav	Narrowband noise of an engine
2	cm30.wav	Classical music with low dynamic range
3	engine1_anc.wav	The residual noise obtained after processing engine noise through a conventional ANC system
4	engine1_cm30.wav	The residual noise obtained after processing engine noise through a IANCM system with classical music used as masker
5	car1.wav	Broadband noise inside the cabin of a moving car
6	Snore.wav	Snoring sound
7	vacuum.wav	Vacuum cleaner noise
8	handblender2.wav	Hand blender noise
9	aircon.wav	Air conditioner noise
10	piano.wav	Piano notes
11	tibetanchant.wav	Recording of chants by Tibetan monks
12	tibetanchant_snore.wav	The residual noise obtained after processing snore noise through IANCM system with Tibetan chants used as masker