INVESTIGATION OF IMPROVED MASKING NOISE FOR THE SPEECH PRIVACY

by

Allen Cho

APPROVED BY SUPERVISORY COMMITTEE:

____
Dr. Issa M. S. Panahi, Chair

____
Dr. John H. L. Hansen

____
Dr. P. K. Rajasekaran

For the glory of God

by the grace of God

INVESTIGATION OF IMPROVED MASKING NOISE FOR THE SPEECH PRIVACY

by

ALLEN CHO, BS

THESIS

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

December 2016

ACKNOWLEDGMENTS

First and above all, I praise God almighty for providing me this opportunity and granting me the capability to proceed successfully. This thesis has come to its present form thanks to the assistance and guidance of several people. I would, therefore, like to offer my sincere thanks to all of them.

Foremost, I would like to express my sincere gratitude to my advisor Dr. Issa Panahi for his continuous support for my Master's degree study and research as well as for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me throughout my research and writing of this thesis. I could not have imagined having a better advisor and mentor for my graduate study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. P. K. Rajasekaran, and Prof. John Hansen for their encouragement, insightful comments, and challenging questions.

My sincere thanks also goes to Speech Privacy Systems for offering me the valuable opportunity to study and research with their groups and leading me to work on diverse exciting projects.

I thank my fellow lab members in SSPRL Group: Jinkyu Park, Chetan Vahanesa, Yu Rao, Anshuman Ganguly, Chandan Reddy and Yiya Hao, for Their stimulating discussions, ideas and suggestions that motivated and developed the future paths of my project and for all the fun we have had in the last two years.

Last but not the least, I would like to thank my family: my parents and Bryce for giving me life in the first place and supporting me spiritually throughout my life, without which my accomplishment would have been impossible.

October 2016

INVESTIGATION OF IMRPOVED MASKING NOISE FOR THE SPEECH PRIVACY

Publication No. ___________________

Allen Cho, MS
The University of Texas at Dallas, 2016

Supervising Professor: Dr. Issa M. S. Panahi, PhD

Many sound masking products for the purpose of achieving speech privacy have emerged on the market in recent years. Over the time, quality of the masking noise has been improved a lot and still is studied for a better method. This thesis introduces a new way of generating masking noise which outperforms the commonly used methods. Using Adaptive Linear Predictive Coding (LPC) we can derive the model for various target sound and create a new masking signal which has exactly the same spectral envelope of the target sound and different phase. Two important aspects in evaluating the performance quality of the masking noise are masking capability and pleasantness. By testing these criteria in objective and subjective ways, we show superior performance of the proposed adaptive method in comparison with several existing techniques.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

As the importance of information has grown considerably high due to the progress of technology, it has also become important to keep the knowledge confidential. Therefore, many ways to achieve privacy has been investigated with different approaches. The privacy does not only include the meaning of preventing people's information and speech from leaking out, but it also means isolating those who do not want to be disturbed by undesired and interfering speech sounds in a working environments. But many of those who are in need of the privacy do not appreciate the existing methods due to some system requirements which would need more development time and higher cost. At the same time, a place in immediate need of privacy requires an affordable and effective system. Speech privacy is also important to those who work in highly sensitive environments so that they can stay focused on their tasks. Therefore, the way we should design a system to provide and maintain good speech privacy is very important [1].

To provide speech privacy, sound masking signal is produced and propagated through the target area so that the private or confidential speech becomes incomprehensible to other people. By adding unstructured sound to where the privacy is needed, we can cover the structured human speech which we want to mask and make it unintelligible [2-6]. The masking sound we generate, which is to be called masking noise, should also result in a pleasant sound to the people in the targeted area. Hence the masking noise itself should not be uncomfortable to listen but should be strong enough to cover the unwanted speech or the irritating sound. Apparently, there exists a trade-off between these two requirements. It can be simply from the loudness of the signal, but it can also be from other components composing the sound signal such as timbre, roughness, sharpness or tonality. Our proposed method to make the best from this trade-off will be discussed later in this thesis. With all these principles and studies in sight, it is hard to have a consistent result because the decision must ultimately be made based on human perception. Although the exact same signal can be played to listeners, opinions and reactions from the listeners can be different. So any speech privacy study and system development should lead to meeting the satisfaction of the target listeners as much as possible. For that, knowledge of the human psychoacoustics has been used by many researchers.

There has been numerous studies done for designing sound masking systems, often aimed at providing speech privacy. Many of such systems are already being used in practical situations in several places such as hospitals, schools and companies. Utilizing the knowledge of human psychoacoustic models and auditory perception and learning from the results of experiments conducted with normal hearing people can be very beneficial to developing a viable sound masking and speech privacy systems.

In this work, we present an improved method of producing an effective masking signal for speech privacy. The experimental result in both subjective and objective test show reliability and strength of the new method. By creating this masking signal adaptively, we can choose the masker to provide speech privacy in various conditions. Adaptive method means that it can also counteract to changes in the noise environment. Many existing sound masking systems have often used simple white noise or pink noise. They are not quite able to change their masking noise in response to varying surrounding noise. In contrast, the proposed method stays alert to the changes around the system and adopt itself efficiently while keeping the comfort of the listeners.

CHAPTER 2

SOUND MASKING

2.1 History of Sound Masking

By considering the early history, we can infer that primitive people were aware of the principle of masking though they did not want privacy by that time. Considering they had not lived near the river or the place with loud noises, we can imagine that the people in that era avoided living near the noisy place because that sound would mask the approach of their enemies or predators. Although they did not know how masking works with the human auditory system, it seems that they at least understood the basic principle of masking! Also, some dentist takes benefit from the effect of masking from the early 1940s. They played random sound, or music, to their patients through earphones to mask the noise of slow speed drills which is annoying to and often terrifies the patients. The very first model of self-contained sound masking system has emerged in the 1960s. They deployed sound producing electronic system in the open office spaces, which seems to have triggered the rapid evolution of sound masking [7-8]. There were only 4 sound sources we can select from the first sound masking system. One was white noise and others were sounds from the nature. The source for the masking noise has been modified many times since the sound masking system has been first introduced [9-10]. Also the target environment and target sound are numerous and different due to expanded usage of sound masking. Environments can be either of open ceiling or closed ceiling, the direction of the sound masking speaker can face upward or downward, size and composition and even material of the wall should also be taken into consideration for a moderate and effective masking functionality [11].

2.2 Overview of recent Sound Masking

The first attempt at sound masking used white noise. White noise was chosen because of its high effectiveness in masking other signals. However, through repeated use it soon became clear that white noise was not an effective way to mask noise. Because people hear the masking signal spread in space, the qualities of pleasantness and masking capability are equally important considerations. However, there are pros and cons to each consideration. In Figure 2.1 we can easily compare the two distinct qualities.

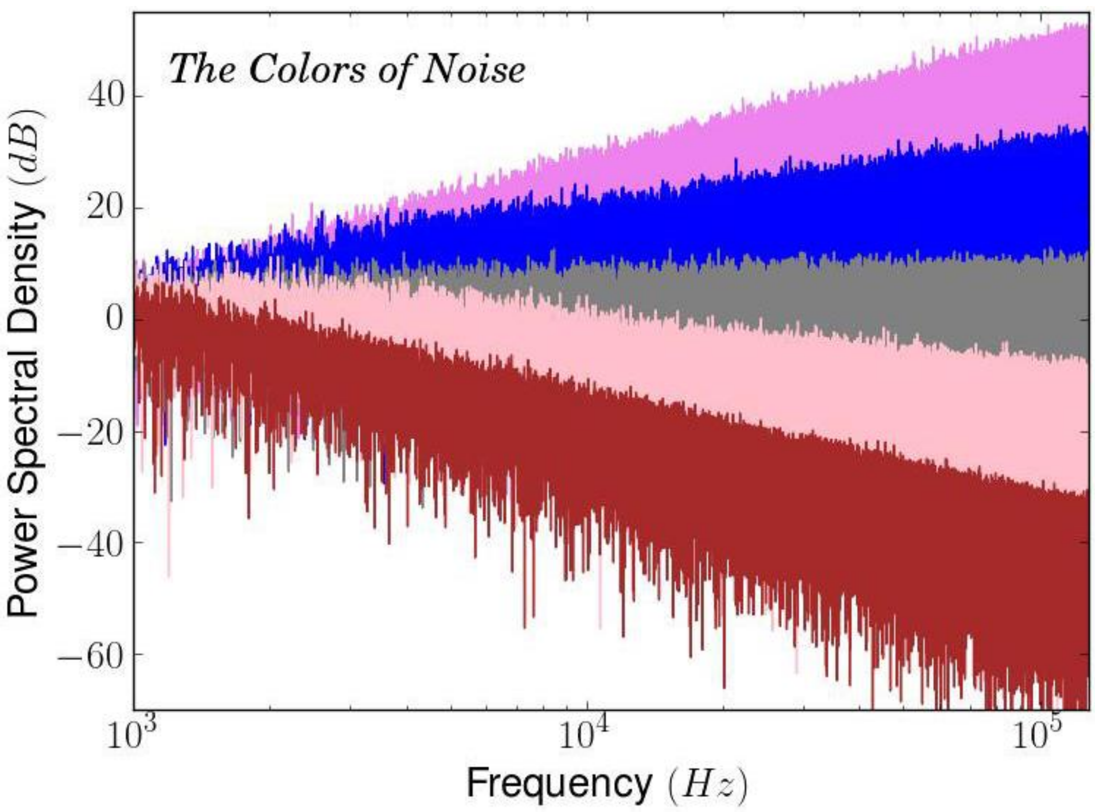


Figure 2.1 Various kind of colored noises [12]

A disadvantage to white noise is its lack of ability to provide aural comfort. Compared with red noise, which has lower energy in the higher frequency range, white noise is better for masking noise but much less pleasant to the ears. Many sound masking companies today attempt to strike a balance between white and red noise. Some companies modify their noise source into a "pink noise". Other companies attempt to filter the output so the user can change the signal for their taste, similar to the way an equalizer function works on a sound system.

The way the emitter and loudspeaker are positioned to generate the masking noise is diverse. System settings should be adjusted according to the purpose of the building. The needs of an educational institution—primarily, the facilitation of concentration—is starkly different from the needs of a rehabilitation hospital, which is primarily rest and calmness. The direction the emitter faces and propagates sound will be different in each location. Other factors to consider in constructing a sound masking system is whether a building has an open or closed ceiling, and the utility of desks and office space.

2.3 Expected progress of future masking technology

A masking system that only covers sound has met its final stage. In addition to recent radical progress on noise handling technologies such as active noise control or sound absorbing panels, we can prospect that another advance could be made. Recently many products have entered the market that combines the functions of these technologies. One product cancels snoring sound around the nose. Another product—bricks absorb HVAC or other environmental noise. Today, noise masking is no longer restricted to an enclosed area.

Considering that sound masking system has been developed in many ways to deal with open or closed spaces, further steps could still be made. For instance, sound masking could be applied in situations where unwanted noise needs to be covered and active noise control can erase extra displeasing sounds.

CHAPTER 3

UNDERSTANDINGS OF PSYCHOACOUSTICS

3.1 Needs for the study of psychoacoustics

Though signal is an absolute value which does not change, hearing is not a purely mechanical phenomenon of wave or vibration delivered over media. It is rather a sensory and perceptual event. As a simple example to explain this, when a person listens to something, the signal sensation they perceive, regardless of how the signal is maintained, will be as different as the number of people who hear. Within the ear, sound is transformed into neural action potentials that then travel to the brain where they are perceived. Developing a fundamental understanding of this psychoacoustic process could help produce the best sound masking outcome.

3.2 Auditory Masking

Auditory masking [13-16] is a phenomenon that occurs when a human perceives two simultaneous sounds and one sound is affected by the other. Auditory masking can be pigeonholed into two categories. In the frequency domain, auditory masking is known as frequency masking [17] or simultaneous masking. In the time domain, auditory masking is called temporal masking. For the purposes of this thesis, frequency masking will be used since the sound masking system will be operated more than temporarily. Acknowledging that human speech contains intelligible content concentrated in the frequency range between 1,000Hz and

3,000Hz, revising the same frequency component of the masking signal would help conclude a better result. As shown in Figure 2.2, a lower frequency signal tends to show a wider extent in masking.

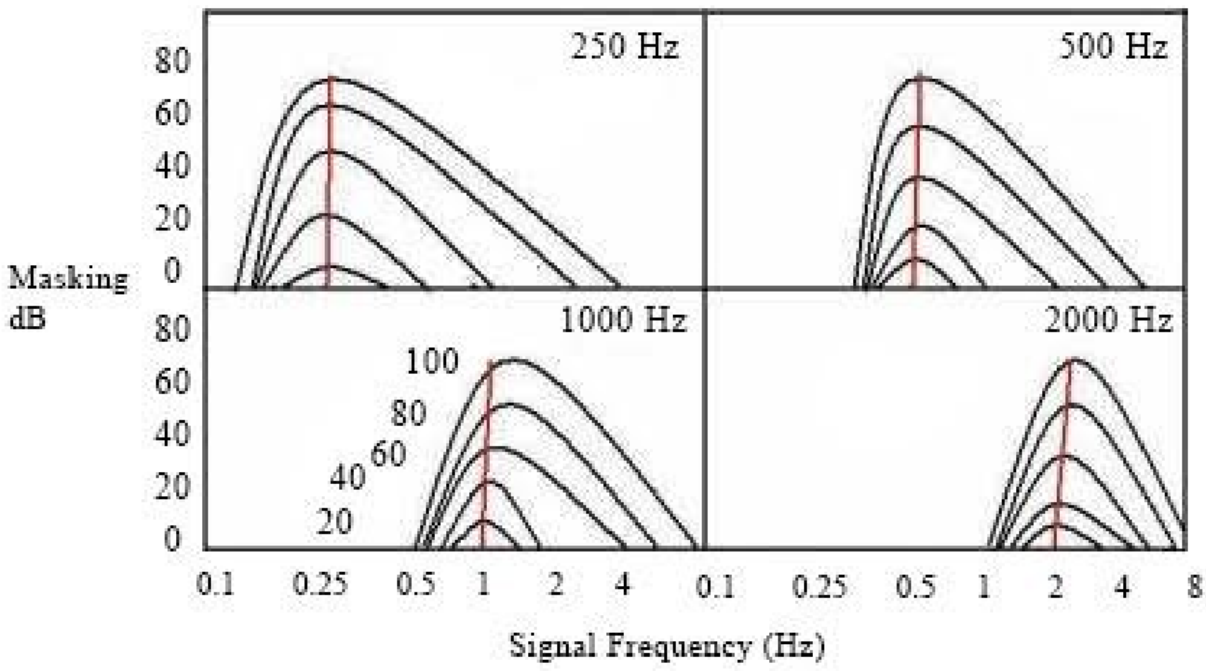


Figure 3.1 Series of masking patterns which is also known as audiograms [18]

Four attributes impact psychoacoustic effect in simultaneous masking. These are described in the audiograms. First, the signal must have its own critical bandwidth over the general frequency range. While the same intensity of a single tone is played, the range of masking differs with different frequencies. Second, the signal must consist of a lower frequency. As we stated above, human speech has much power concentrated in a lower frequency, somewhere between 1,000Hz and 3,000Hz. Therefore a lower frequency signal is more likely to mask compared to a higher one. Third, varying intensity levels can also have an effect on masking. The lower slope of a masking filter in which single tones form becomes more flat as decibel level increases, whereas the higher slope becomes steeper. Both higher and lower slopes broaden as the intensity of the input level increases. These observations are illustrated in Figure 3.2.

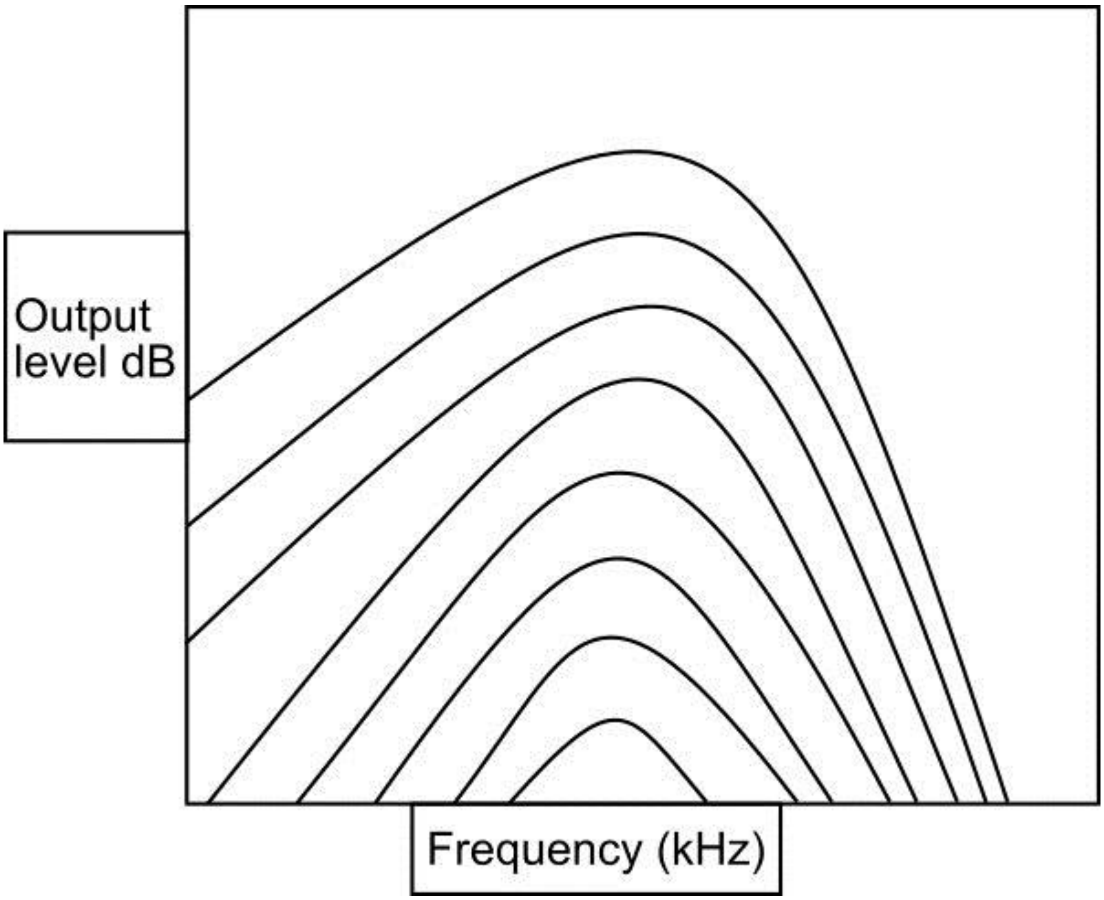


Figure 3.2 The effect of intensity on masking [8]

Auditory masking requires all four attributes. Engaging in auditory masking without taking into account any of these factors could ruin one's ability to find proper ways of masking, especially in the case of a single sinusoidal tone. An actual situation will be more complex. Because the signal we are dealing with is a sound, it contains multiple kinds of intensity over various frequencies.

3.3 Weber-Fechner Law

Certain sound can be more easily perceived in a silent environment than in a place with noise. In other words, the difference between masking threshold and desired masking level will change according to time and environment. With the changing frequency spectrum and the variable intensity of the target sound, auditory masking that does not cause discomfort can be accomplished with an optimum signal [19].

3.4 Sensory Adaptation

Sensory adaptation explains why and how sound masking can be used in the actual field. It does not matter how comfortable the signal is for someone to hear; eventually a continuously exposed stimulus will become inconvenient. At this point, the phenomenon called sensory adaptation goes into play. To avoid the problems stated above, we need to refine the masking signal to make it sound pleasant. Exposure to a masking signal for a long time is not going to introduce another inconvenience that requires more sensory adaptation. Neural adaptation or sensory adaptation [20] is a change over time in the responsiveness of the sensory system to a constant stimulus.

The sensitivity of human sensory receptors changes relative to a change in external stimuli. When the stimulus, the masking signal, is applied continuously, people can adapt to the sound. At that point they will neither detect the masking signal nor take offense to it. However, since a significant change of stimulus could abruptly interrupt sensory adaptation, sound must be changed gradually. A sound masking system is equipped with an initial ramp-up function. This function prevents the system from starting in loud volume. A ramp-up feature will help the listener not recognize the sound masking system is on, thus making it possible to apply discontinuous stationary masking noise without arousing discomfort.

3.5 Zwicker’s Model of Sensory Pleasantness

It seems impossible to quantify numerically the pleasantness of what humans hear since it is already stated that every individual has a different structure of hearing sensation and distinctive way of perceiving. However, there is a way to measure the comfort level humans accept quantitatively. The model of Sensory Pleasantness was first presented by Dr. Zwicker in 1990 in his book, *Psychoacoustics* [21]. This model allows us to numerically evaluate, compare, and measure the abstract concept of how people feel when they hear sound [22]. Zwicker’s equation consists of four different parameters, each of which examines the property of sound. The four elements of the Sensory Pleasantness model are as follows.

3.5.1 Loudness

$$N = \int_0^{24\,Bark} N' dz \tag{1}$$

We call the value we need to calculate the total loudness “specific loudness” with the symbol N’. Loudness N will be the integral of specific loudness over critical band rate from 0 Bark to 24 Bark. Figure 3.3 explains the procedure used to compute total loudness [23].

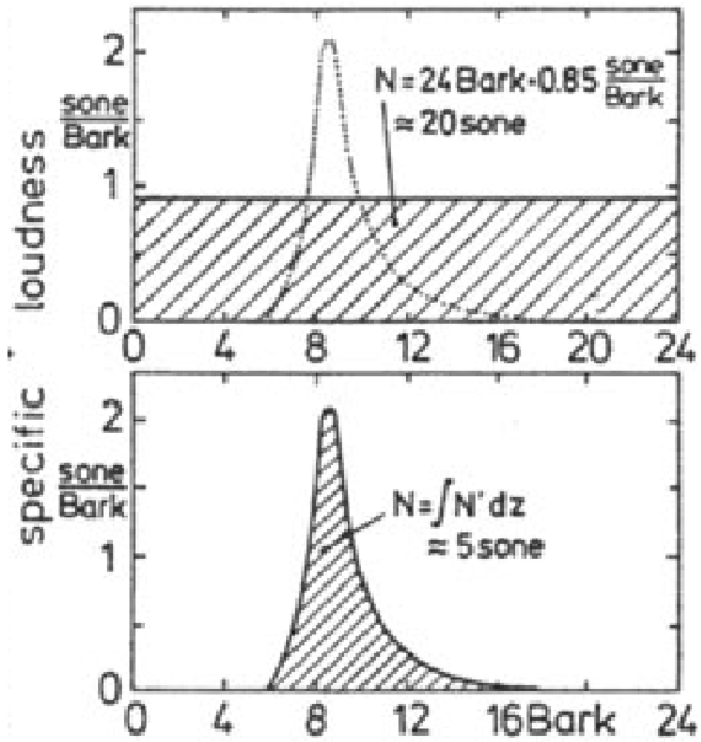


Figure 3.3 Developing total loudness out of specific loudness. The area of hatched region corresponds to the total loudness. [21]

3.5.2 Sharpness

$$S = \frac{0.11 \int_0^{24\,Bark} N' g(z) z dz}{\int_0^{24\,Bark} N' dz} acum \quad (2)$$

In developing a model for sharpness, it is helpful to treat it independent of spectral structure. The overall spectral envelope is the factor that decides sharpness. In Figure 3.4, it is shown that sharpness increases as the frequency and critical band-rate goes up. But it is not follow linearity. Also in Figure 3.5 is the weighting factor $\mathrm{g(z)}$, which varies based on conditions.

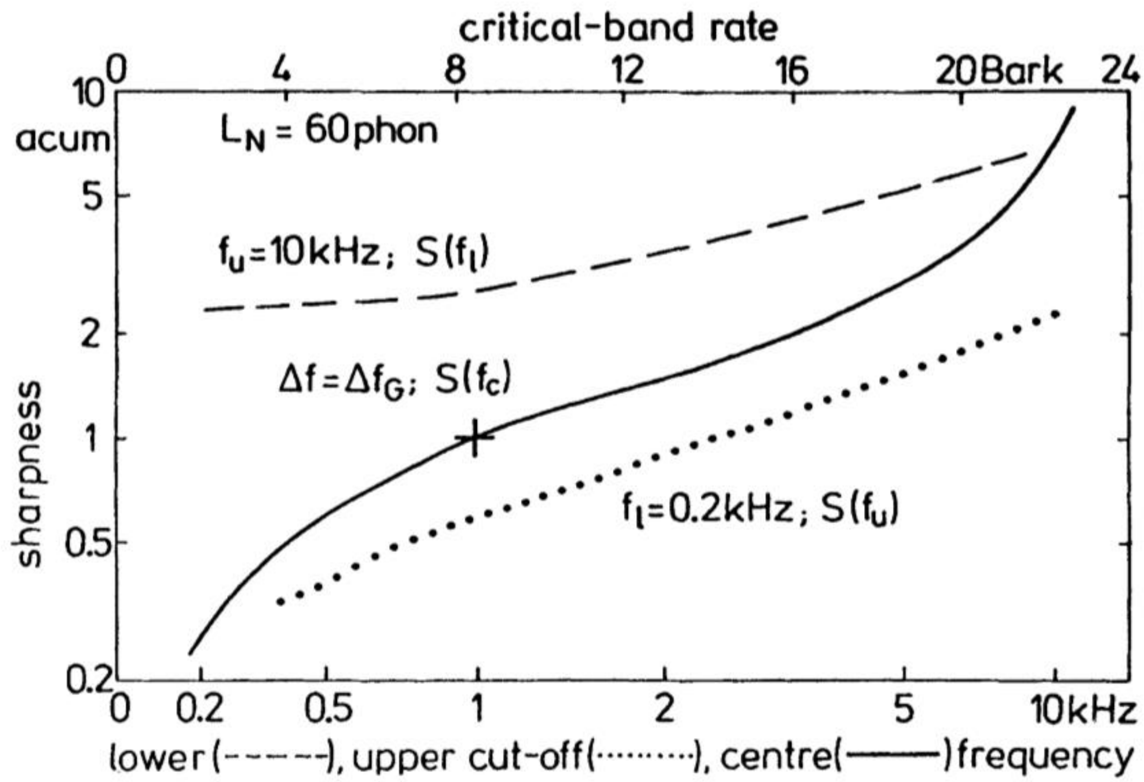


Figure 3.4 Sharpness of narrow-band noise as a function of center frequency of band pass noise with an upper cut-off frequency of and a lower cut-off frequency of 0.2 kHz [21]

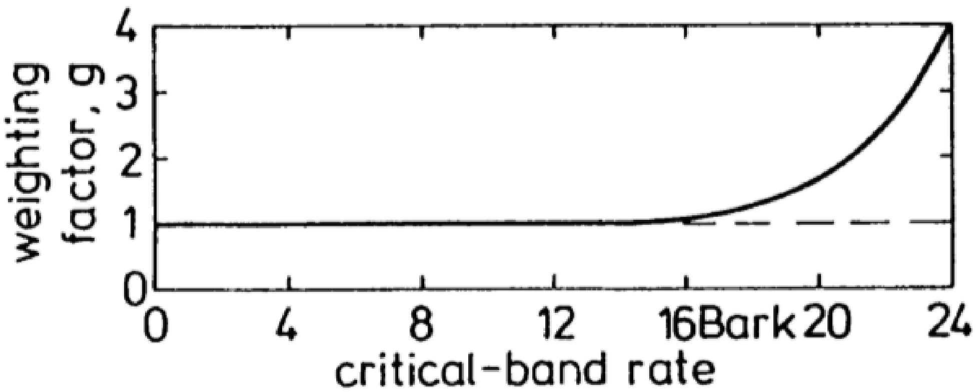


Figure 3.5 Weighting factor for sharpness as a function of critical-band rate [21]

3.5.3 Roughness

$$R = \frac{0.3 f_{mod}}{kHz} \int_0^{24\ Bark} \frac{\Delta L_E(z) dz}{\frac{dB}{Bark}} asper \tag{3}$$

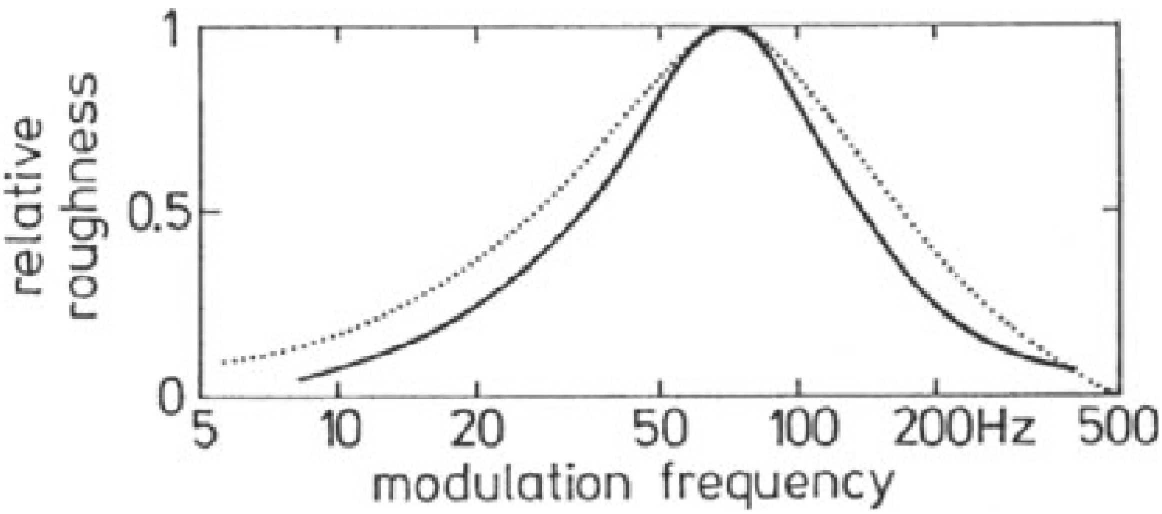


Figure 3.6 Relative roughness with different modulation frequency Solid line stands for the values measured subjectively and dotted line is the result out of calculation. [21]

In Figure 3.6 roughness relies on modulation frequency.

$$\mathrm{R} \sim \mathrm{f}_{\mathrm{mod}} \Delta L \tag{4}$$

ΔL stands for the masking depth. Roughness should be computed over the critical bands.

$$\mathrm{R} \sim \mathrm{f}_{\mathrm{mod}} \int_0^{24Bark} \Delta L_E(z) dz \tag{5}$$

Using the boundary condition that a 1 kHz tone at 60 dB and 100%, 70 Hz amplitude-modulated produces the roughness of 1 asper, we get the model for roughness R which can be used to calculate the roughness for any sound.

3.5.4 Pleasantness

$$P = e^{-0.7R} e^{-1.08S} (1.24 - e^{-2.43T}) e^{-(0.023N)^2} \quad (6)$$

Whether a sound is perceived as pleasant or unpleasant depends on the relationship of the listener to the sound. It can be measured subjectively as well as with the physical parameters of the specific sound [24]. Ignoring these non-acoustic influences, we develop a model of sensory pleasantness based on the relation between the characteristics of the human hearing system on the one hand and the physical parameters of sound on the other [25]. Sensation sharpness, roughness, tonality and loudness of human beings can be approximated based on the equation developed above. Starting with the sensation of loudness, it is possible to get a close value for actual loudness. Other sharpness and roughness can also be computed based on that. The sensation of tonality is not able to be estimated with an equation in a numerical manner. Therefore it should be approximated through subjective measurement. The model of Sensory Pleasantness enables direct comparison among various masking noise candidates to determine which sounds are more comfortable [26]. Unfortunately, for now the model can be applied only to stationary signals. Masking signals combined with background music or other nonstationary masking signals are not able to be measured with the equation.

CHAPTER 4

PROPOSED METHOD TO GENERATE IMPROVED MASKING NOISE

4.1 Appropriate SPL for better sound masking

Sound pressure level (SPL) is a function of sound, distance, direction, and source signal characteristics. The SPLs of different sounds can be observed from Table 4.1. Each SPL has a different duration. It is known that the average SPL of conversation among humans is in a range between 40 dB to 60 dB. The sound masking product already on the market registers an SPL that does not exceed 50 dB.

Some may be concerned that the way multiple sound mixes with ambient noise, environmental noise and masking noise, If we assume the background noise is 60 dB at most and even if the masking noise is over 50 dB, the summation of multiple noises does not yield 110 dB but 80. So it is not unprecedented to consider this problem in building a sound masking system. Though SPL satisfies the required level of appropriate masking noise, there are still a few more things that need to be considered. For example, the spectral envelope has a direct relation with the actual feeling of how the sound will be felt to people. This relationship needs to be developed according to external standards of efficiency and effectiveness.

Table 4.1 Typical noise level chart

SPL (dBA)	Type of source sound
120	Whistle at 1 m distance, test run of a jet at 15 m distance
	Threshold of pain, above this fast-acting hearing damage in short action is possible
110	Siren at 10 m distance, frequent sound level in discotheques and close to loudspeakers at rock concerts, violin close to the ear of an orchestra musicians (greatest level)
100	Frequent level with music via head phones, jack hammer at 10 m distance
90	Angle grinder outside at 1 m distance
	Over a duration of 40 hours a week hearing damage is possible
80	Very loud traffic noise of passing lorries at 7.5 m distance, high traffic on an expressway at 25 m distance
70	Level close to a main road by day, quiet hair dryer at 1 m distance to ear
65	Bad risk of heart circulation disease at constant impact is possible
60	Noisy lawn mower at 10 m distance
55	Low volume of radio or TV at 1 m distance, noisy vacuum cleaner at 10 m distance
50	Refrigerator at 1 m distance, bird twitter outside at 15 m distance
45	Noise of normal living; talking, or radio in the background
40	Distraction when learning or concentration is possible
35	Very quiet room fan at low speed at 1 m distance
25	Sound of breathing at 1 m distance
0	Auditory threshold

Table 4.2 Permissible exposure time to the sound due to Sound Pressure Level

Sound Pressure Level	**Sound pressure**	**Permissible Exposure Time**
115 dB	11.2 Pa	0.46875 minutes (~30 sec)
112 dB	7.96 Pa	0.9375 minutes (~1 min)
109 dB	5.64 Pa	1.875 minutes (< 2 min)
106 dB	3.99 Pa	3.75 minutes (< 4 min)
103 dB	2.83 Pa	7.5 minutes
100 dB	2.00 Pa	15 minutes
97 dB	1.42 Pa	30 minutes
94 dB	1.00 Pa	1 hour
91 dB	0.71 Pa	2 hours
88 dB	0.50 Pa	4 hours
85 dB	0.36 Pa	8 hours
82 dB	0.25 Pa	16 hours

4.2 Linear Predictive Coding

Using the knowledge of psychoacoustics and psychophysics, it is possible to create a new masking signal. Research has demonstrated up to this point that early attempts at sound masking used white noise or pink noise for their masking noise. However, those signals are inadequate. When it comes to reality, the feelings of the listeners and their privacy should also be taken into consideration.

Our proposed method can cover all the problems that need to be addressed. Since its spectrum is almost the same as the original target spectrum, masking efficiency will outdo the utility of the previous colored noise (white noise or pink noise). If the pink or white noise is used, there shall be some specific frequency zone that the masking noise covers in relationship to the distance between the frequency and signal. As long as the frequency spectrum duplicates the target spectrum, masking leaves no room for a leaking point. Discomfort in hearing can be also reduced since this method doesn't yield useless frequency which could cause disturbances. Linear Predictive Coding (LPC) uses the autocorrelation method of autoregressive modeling to find coefficients of the filter [27]. The autocorrelation method data assumes that signal samples beyond the size of x are 0. LPC calculates the least square solution to

$$\mathrm{Xa} = \mathrm{b} \tag{7}$$

Where

$$\mathrm{X} = \begin{bmatrix} x(1) & 0 & \cdots & 0 \\ x(2) & x(1) & \ddots & \vdots \\ \vdots & x(2) & \ddots & 0 \\ x(m) & \vdots & \ddots & x(1) \\ 0 & x(m) & \ddots & x(2) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & x(m) \end{bmatrix}, a = \begin{bmatrix} 1 \\ a(2) \\ \vdots \\ a(p+1) \end{bmatrix}, b = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{8}$$

When m is the size of the x, we solve the least square via normal equation method.

$$\mathrm{X}^{\mathrm{H}} X a = X^H b \tag{9}$$

Then we get to the Yule-Walker equation.

$$\begin{bmatrix} r(1) & r(2)^* & \cdots & r(p)^* \\ r(2) & r(1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r(2)^* \\ r(p) & \cdots & r(2) & r(1) \end{bmatrix} \begin{bmatrix} a(2) \\ a(3) \\ \vdots \\ a(p+1) \end{bmatrix} = \begin{bmatrix} -r(2) \\ -r(3) \\ \vdots \\ -r(p+1) \end{bmatrix} \tag{10}$$

Where

$$\mathrm{r} = [r(1)\, r(2) \ldots r(p+1)] \tag{11}$$

is an estimate of the autocorrelation needed to compute cross-correlation. The Yule-Walker equation can be solved with the Levinson-Durbin algorithm. Once the coefficients successfully from the signal, we now have coefficients to make a filter which refines original source signal into the white signal as described in Figure 4.1.

Figure 4.1 Finite Impulse Response whitening filter that pre-whitenes given input signal into whitened output

Our proposed method was to come up with a new process that inverts this method. Unlike converting source signals into white noise, white noise is our desired signal. The whitening filter that we have from LPC is

$$\mathrm{A(z)} = \sum_{\mathrm{k}=0}^{\mathrm{p}} \left(\mathrm{a_k z^{-k}}\right) \tag{12}$$

Given that this filter is finite impulse response (FIR), this equation can be reverted to get a reciprocal of the filter.

$$\mathrm{H(z)} = \frac{1}{\mathrm{A(z)}} \tag{13}$$

The reciprocal filter of the original will do the exact opposite of the original.

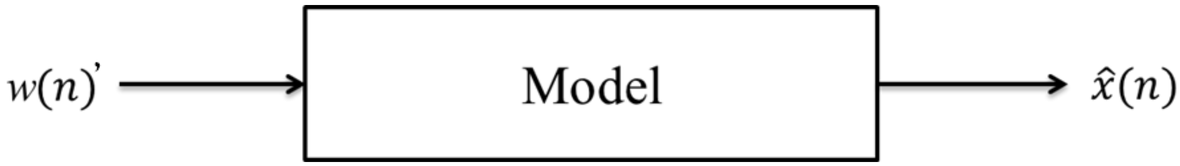


Figure 4.2 Inverse of the whitening filter which generates new masking signal out of any given white noise to be used for reproducing our proposed method.

If the signal from the same white noise is recovered, there could be a chance to get the original signal, which contains many speech components and factors that may disrupt listeners. In the sound masking system, white noise is made from the processing unit by generating random numbers. It is expressed as $\mathrm{w(n)}'$ in Figure 4.2 and is different from the white noise regarding $\mathrm{w(n)}$ the original filter. Though the spectral envelope is almost the same, because white noise includes the same magnitude over all frequencies, the phase of the white noise is totally different between $\mathrm{w(n)}$ and $\mathrm{w(n)}'$. As we looking into the power spectrum of the two signal $\mathrm{x(n)}$ in Figure 4.3 and $\hat{x}(n)$ in Figure 4.4, it is almost hard to discern the difference between the signals given that the order of the filter is over 50.

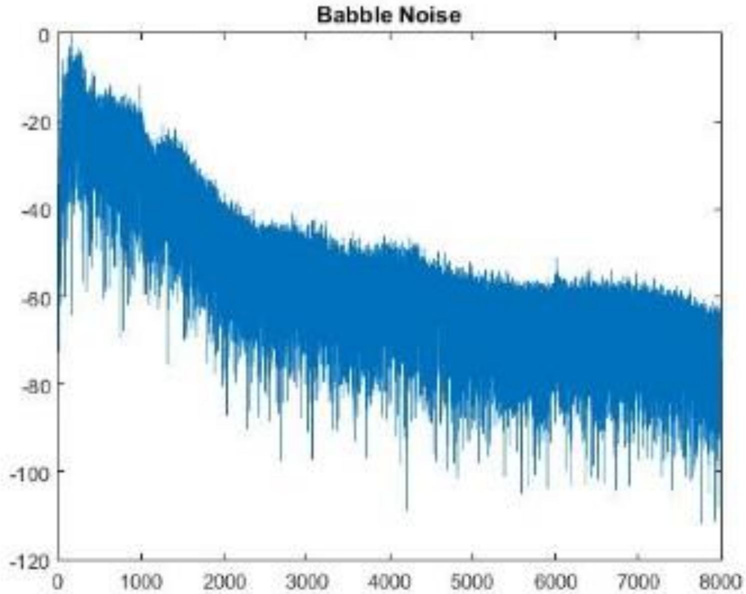


Figure 4.3 Frequency spectrum of the babble noise

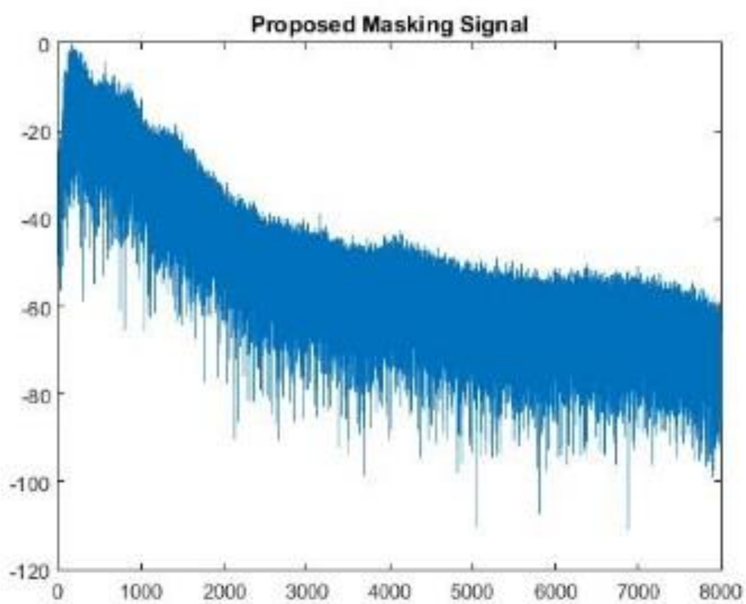


Figure 4.4 Frequency spectrum of the masking noise generated from proposed method

Seeing that the result shows no considerable contrast between lower order and higher order, the order of the filter does not need to go beyond 100. Though higher order could yield a more refined structure of the signal, the reason to avoid more order is because the system is designed to run adaptively in real time. If the masking noise is heard, there still remains a point to be fixed or improved. In Chapter 5 introduces a couple of processes to improve the performance of the system.

CHAPTER 5

PROCESS TO IMPROVE PERFORMANCE

5.1 A-Weighting Filter

The critical frequency range differs from the property of speech. The vowel sound is delivered predominantly in the frequency of 250Hz to 500Hz while a consonant is delivered in the frequency of 2,000Hz to 4,000Hz. Sensitivity with regard to frequency is represented on the equal-loudness contour and speech intelligibility index as shown in the Figures below. The A-weighting filter enhances the specific frequencies that have speech components. Its original purpose was to shape the perception of the way humans feel noise. By adapting this filter to the proposed masking signal, the speech-containing frequency range can be emphasized.

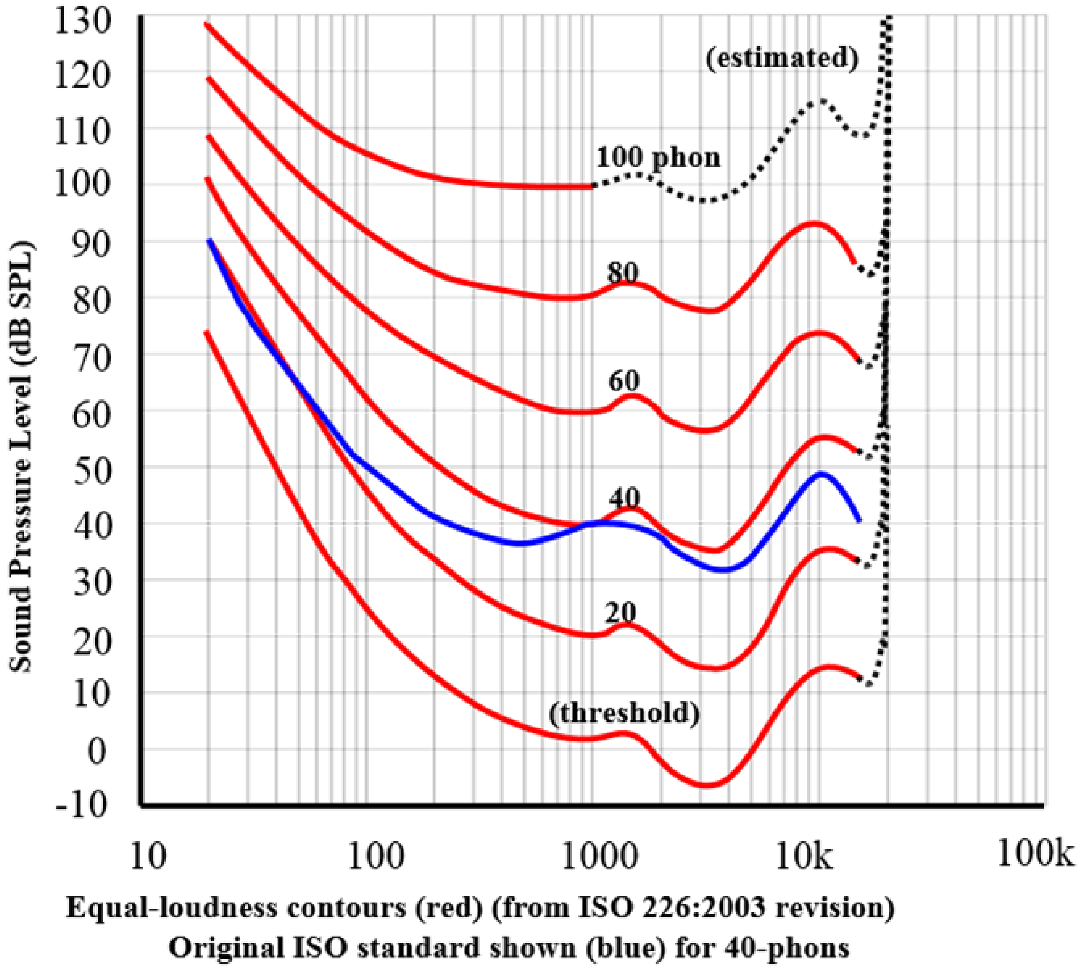


Figure 5.1 Equal loudness contour.

The equal-loudness contour [29] is a measure of sound pressure in dB SPL over frequencies. The loudness level is measured with the unit of phon. Each contour has a different form or shape due to its various phons. It has recently been revised with the progress of telecommunication and is shown in the Figure 5.1. From analyzing the graph, it is evident that the human auditory system is most sensitive to sound occurring within the 2 kHz to 5 kHz frequencies. In addition, subjects perceived the same degree of loudness at the 3 kHz tone with minimum sound pressure.

Table 5.1 Weighting dB values with 1/3 octave band center frequency

1/3 Octave Band Center Frequency (Hz)	**A-Weighting dB (dB)**
160	-13.4
200	-10.9
250	-8.6
315	-6.6
400	-4.8
500	-3.2
630	-1.9
800	-0.8
1000	0
1250	0.6
1600	1
2000	1.2
2500	1.3
3150	1.2
4000	1
5000	0.5
6300	-0.1
8000	-1.1
10000	-2.5

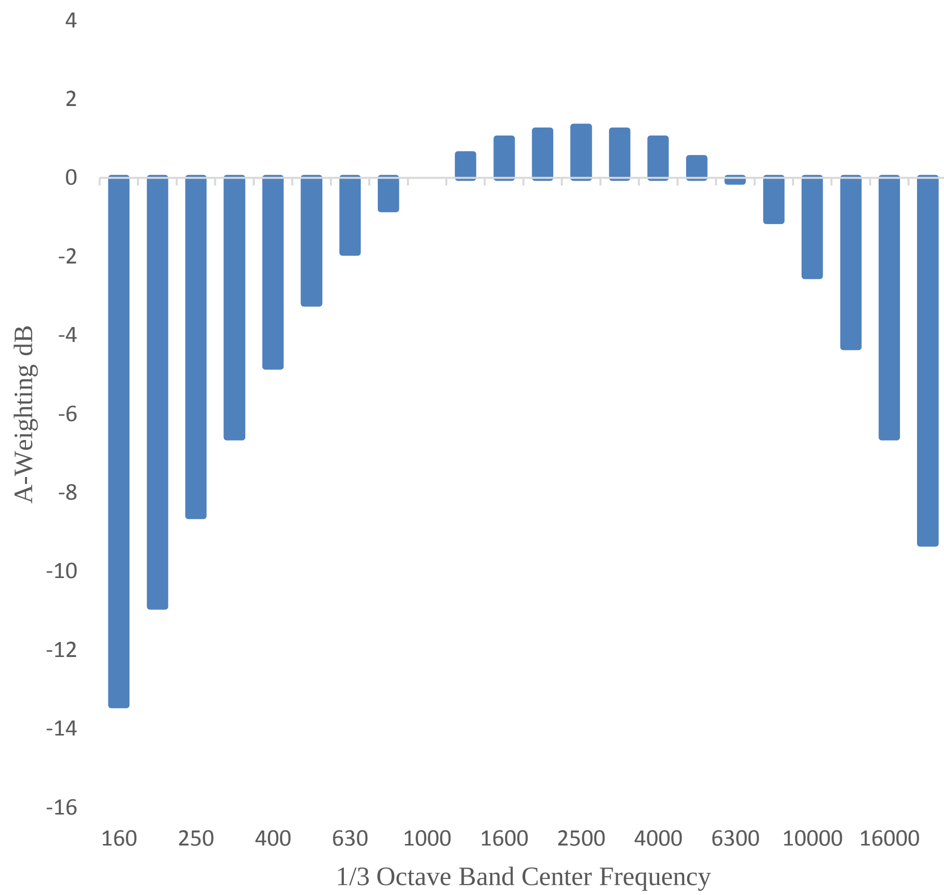


Figure 5.2 A-Weighting values varies with frequency.

$$W_A = 10Log\left[\frac{1.562339f^4}{(f^2+107.65265^2)(f^2+737.86223^2)}\right] + 10Log\left[\frac{2.242881\times10^{16}f^4}{(f^2+20.598997^2)^2(f^2+12194.22^2)^2}\right] \tag{14}$$

Where W_A is weighting value to be applied in dB and f is frequency in Hz.

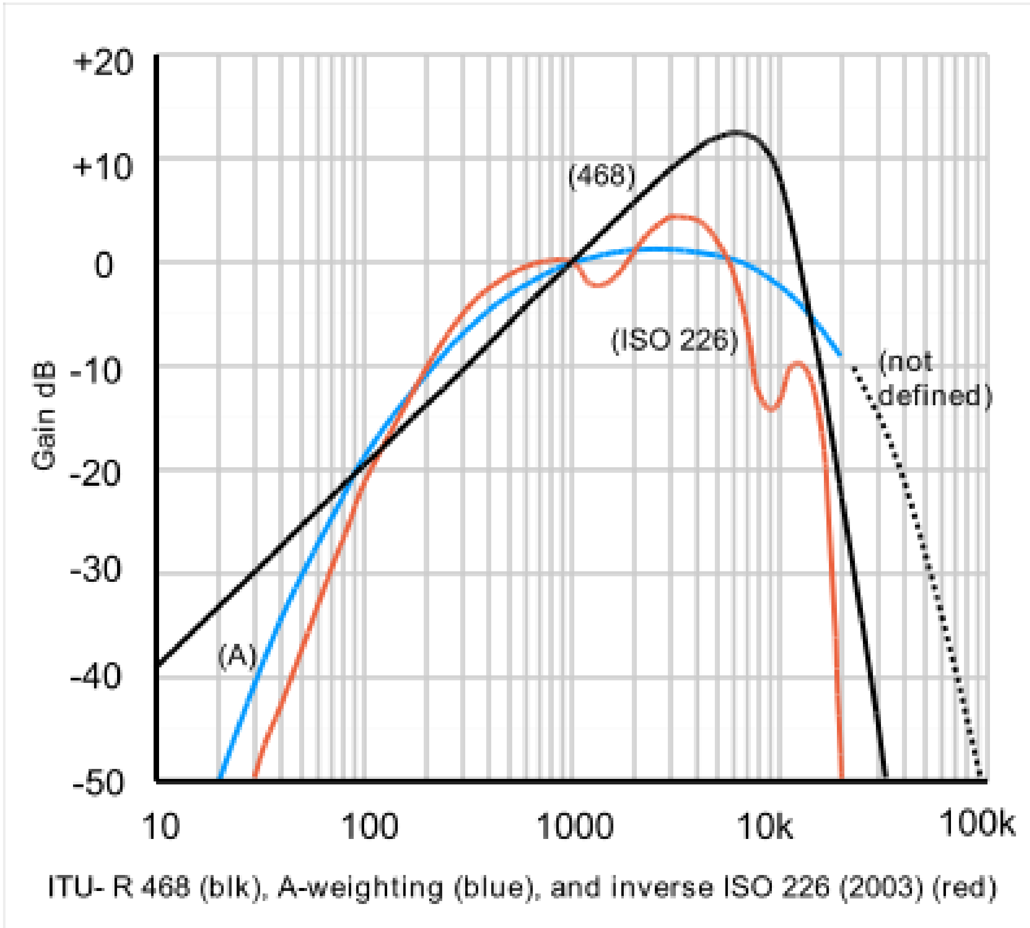


Figure 5.3 Inverse of equal-loudness contour and A-weighting function.

An A-weighting filter [30] is based on the equal-loudness contour of 40 phon. The shape of A-weighting filters is close to that of inverse contour for 40-phon equal-loudness graph. The figures below show how A-weighting filters can modify the signal [31-32]. For original equal-loudness contour and A-weighting filter, the graph is explained in a logarithmic rather than linear scale. Figure 5.4 is the proposed masking signal from Chapter 4. Figure 5.5 is the signal to which has been applied the A-weighting filter of Figure 5.4. This figure shows that the filter enhances frequency around 2 kHz to 5 kHz while suppressing other frequencies.

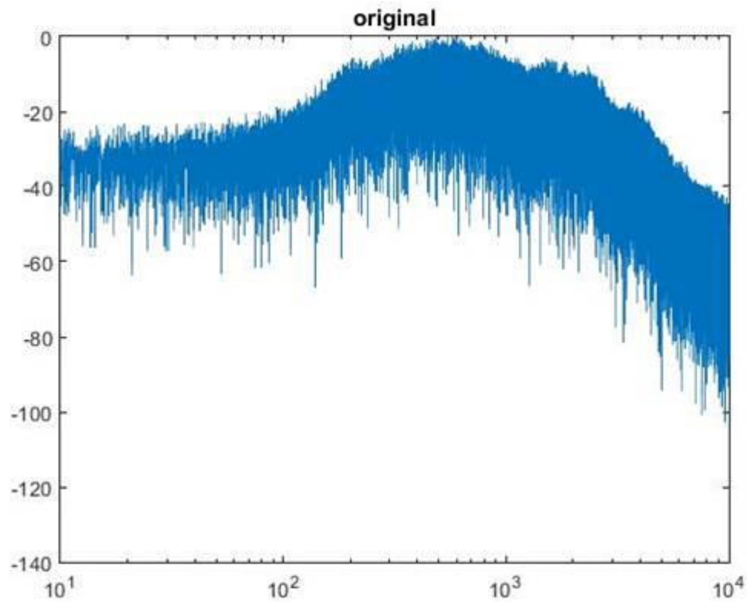


Figure 5.4 Original proposed masking noise

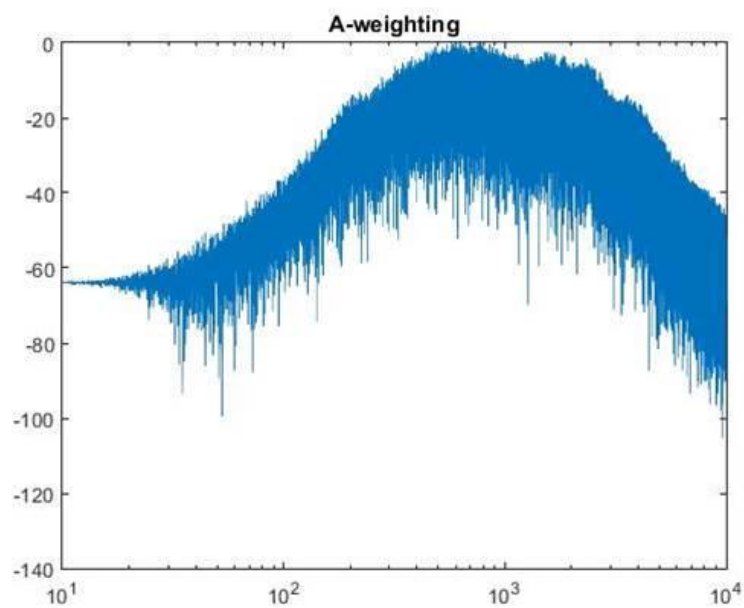


Figure 5.5 A-weighting filter applied masking noise

Masking capability is greatly enhanced when it is characterized by the same sound pressure level. How to estimate and assure the performance of this criterion will be the topic in Chapter 6.

5.2 Moving Average Filter

The outcome of the first proposed masking noise was good masking proficiency. Listening to the very first signal out of the LPC yielded a sound that was comprised of little crackling sounds. This crackling sound causes an unpleasant experience for listeners. The crackling sound was a phenomenon that resulted from a discontinuity in amplitude with neighboring signal samples.

Simply a sudden change of signal can affect a signal’s pleasantness. This discontinuity was eliminated by adding another filter called the Moving Averaging Filter [33].

$$Y(i) = \frac{1}{2N+1}\big(y(i+N) + y(i+N-1) + \cdots + y(i-N)\big) \tag{15}$$

N is the number of neighboring data points and is called span. This filter smooths the signal by replacing each data point with the average of its adjacent ones. According to Figure 5.6, the filter induces an abnormal form of spectrum in the output signal. An analysis of the equation in the frequency domain will easily explain why.

$$\mathrm{H[f]} = \frac{\sin(\pi f(2N+1))}{(2N+1)sin(\pi f)} \tag{16}$$

Figure 5.7 shows the frequency response of the given filter.

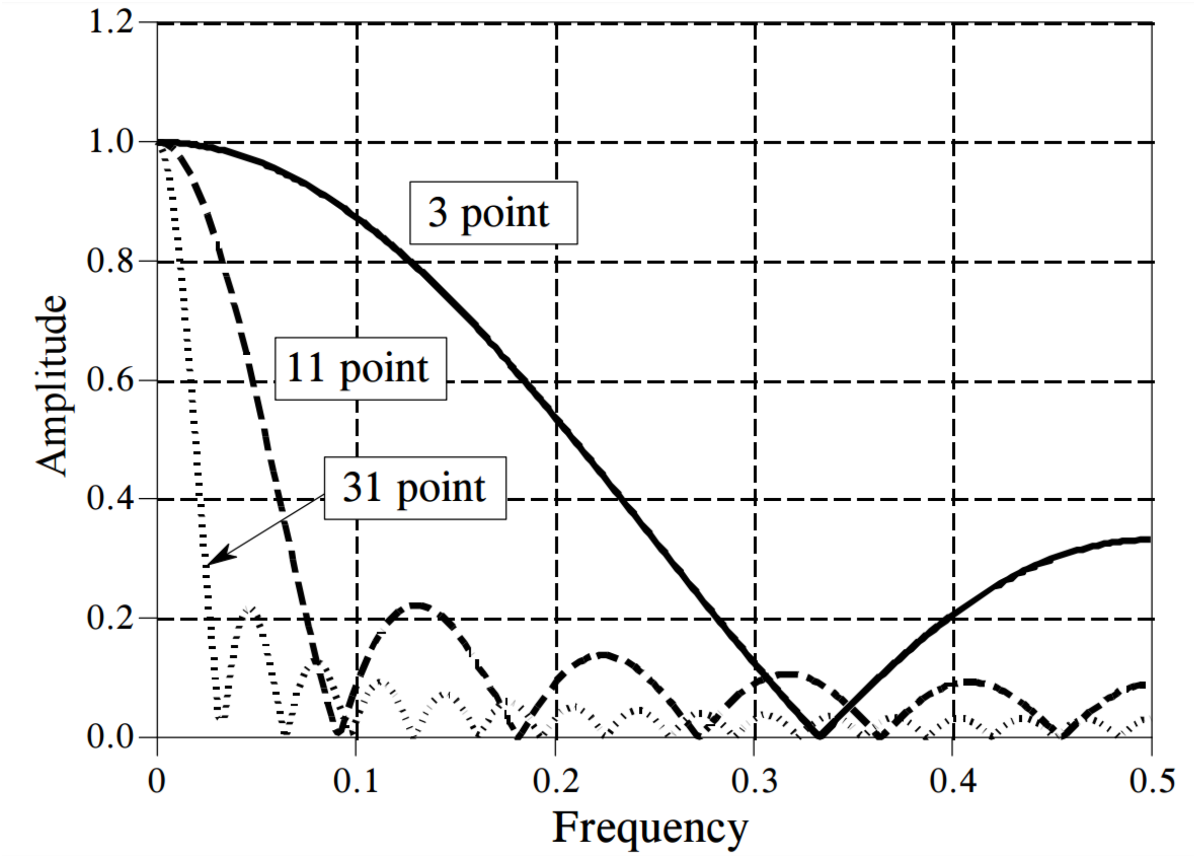


Figure 5.6 Frequency response of the moving average filter with different span length.

The large span may destroy the intended frequency spectrum of masking noise. N=5 is the ideal number to make the best of both pleasantness and masking capability. In an actual hearing, this process could possibly remove all of the crackling sounds. At last, Figure 5.7 is the final output signal through a moving average filter with proposed masking noise.

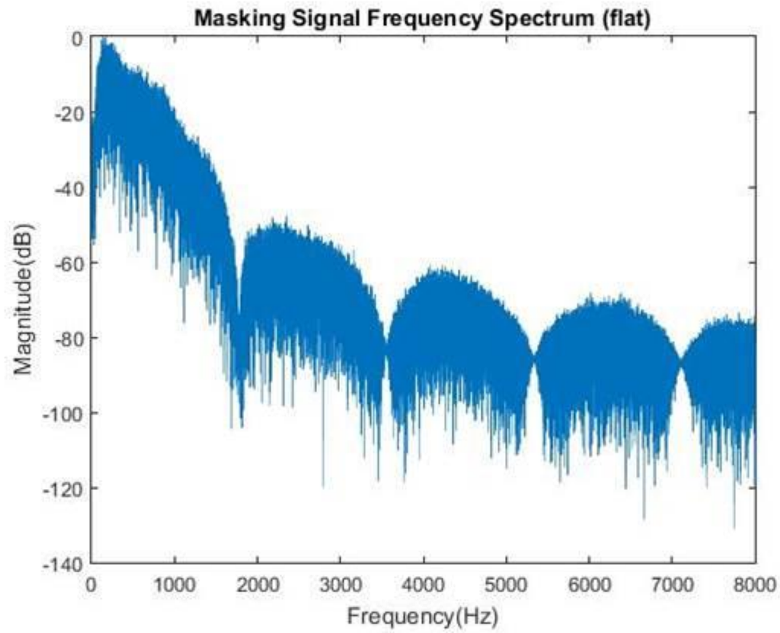


Figure 5.7 Masking noise with moving average filter.

5.3 Multiband Equalizer

Sensational preference differs from gender, age and other numerous reasons. The proposed method could result in a better response than the previous one. But it is impossible to make the best signal satisfying everyone who hears. By adapting equalizer [34-36] onto the signal that we make, we can allow users to manipulate the sounds so that they can change it for their own favor. In this process, we also consider that human perception in hearing differs from each frequency. Dividing frequency into 20 different frequency segments not in linear scale but in logarithmic scale, non-uniformity among frequency can be adjusted. We use 1/3 octave bands to compose frequency bins because 1/3 octave band resembles most closely to a human auditory system. On each frequency bin, we add the gain function so we can alter the frequency range which we want to make a change.

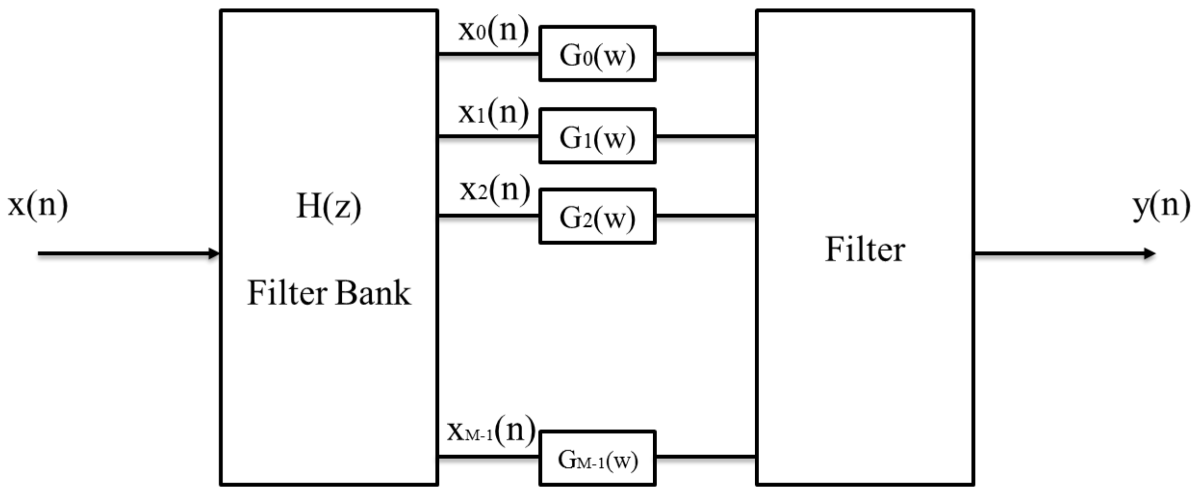


Figure 5.8 Block diagram for Multiband filter equalizer.

CHAPTER 6

EXPERIMENTAL RESULT

6.1 Criteria to Estimate the Performance

There are many criteria for deciding whether the sound masking system performs well or poor with regard to human perception preferences. Two of the most common criteria are masking capability and aural pleasantness. Masking capability measures how much of the unwanted noise has been covered or made unintelligible with the masking signal. Aural pleasantness is applied to human subjects. As discussed earlier, there is a trade-off between capability and pleasantness, because one directly affects the other. This chapter will examine and verify how the proposed signal can be made most effective in light of this trade-off.

6.1.1 Masking Capability

This system's primary goal was to add specific noise to the environment, make other ambient noise, and eliminate unwanted speech. For that reason the value of the masking system depends on its masking level and efficiency. The speech intelligibility index (SII) [37-38] has been widely used as a means for estimating speech intelligibility under additive stationary noise condition. The coherence and speech intelligibility index that is referenced in this thesis—which includes peak and center clipping distortion and coherence between the input and output signals enables us to examine the effect of noise and distortion in intelligibility. This procedure produces the signal's coherence SII in three separate regions: low, mid, and high level. Also, this procedure estimates the total intelligibility from a weighted combination of these three SII values. CSII value will be used as an objective test method to give quantitatively expressed way of measuring intelligibility. Following is the equation from which CSII value can be calculated.

$$CSII = \frac{\sum_{k=1}^{K=23} W(k) \times SNR^{CSII\,final}(k)}{\sum_{k=1}^{K=23} W(k)} \tag{17}$$

To test this equation in realistic cases we will adopt this signal to people in an actual field situation. Subjects will be 20 normal hearing people under the constant power of masking signals. Speech samples are from IEEE corpus and all sentences are clearly understandable. Test rule follows the standard mean opinion score (MOS) in all cases.

Table 6.1 Mean Opinion Score rating scheme for measuring masking capability

MOS	**Quality**	**Intelligibility**
5	Excellent	Fully understand the speech
4	Good	Understand most of the speech
3	Fair	Understand half of the speech
2	Poor	Hard to understand the speech
1	Bad	Barely understood

The goal of the field test is to make the target speech not understandable. So the MOS score for intelligibility should be lower for a good performing signal. The test will ask 20 people to listen to the proposed masking noise and score how much they understand the original speech with and without the additive masking signal. An average from these 20 people will be compared to see which signal works better in an aspect of masking.

6.1.2 Pleasantness

In spite of the effect from sensory adaptation in human hearing sensation, when a sound feels disturbing the usability will drop down drastically. As long as the masking capability is guaranteed, the sound should be as pleasant to the ear as possible. This criterion is adopted as one of the most important standards in determining performance. This test again follows the MOS test scheme for subjective tests to examine pleasantness for those who listen. For this factor, high test score indicates better performance.

Table 6.2 Mean Opinion Score rating scheme for measuring pleasantness

MOS	**Quality**	**Perceptibility**
5	Excellent	Perceptible without annoyance
4	Good	Perceptible but slightly annoying
3	Fair	Little annoying but perceptible
2	Poor	Annoying and barely perceptible
1	Bad	Very annoying

Other than the subjective test, another criterion is needed to yield constant results regardless of experimental conditions. To accomplish this criterion, the objective test from Zwicker's model of sensory pleasantness from Chapter 3 will be used. The model consists of four different characteristics of sound. These characteristics will provide us with an objective standard from which to pick the best outcome among masking noise candidates.

6.2 Test Setup

To verify the performance of the proposed method, we will conduct a subjective test and objective test. We will compare the proposed method with white noise, pink noise and formerly used masking products. Through speakers installed in the ceiling, masking noise will be played and another loudspeaker will play human speech. Equipment position is explained in Figure 6.1. Human speech will be understood perfectly in the silence. Subjects will score whether the masking noise is pleasant and speech under the masking noise is intelligible. Scoring procedure will follow the Mean Opinion Score (MOS) measurement. A subjective test will be done for 20 normal-hearing people. To show the result in a quantitative way we will use an objective measuring process as well as two criteria, intelligibility and pleasantness. Coherence and the Speech Intelligibility Index (CSII) and Zwicker's sensory pleasantness model will be used in the tests.

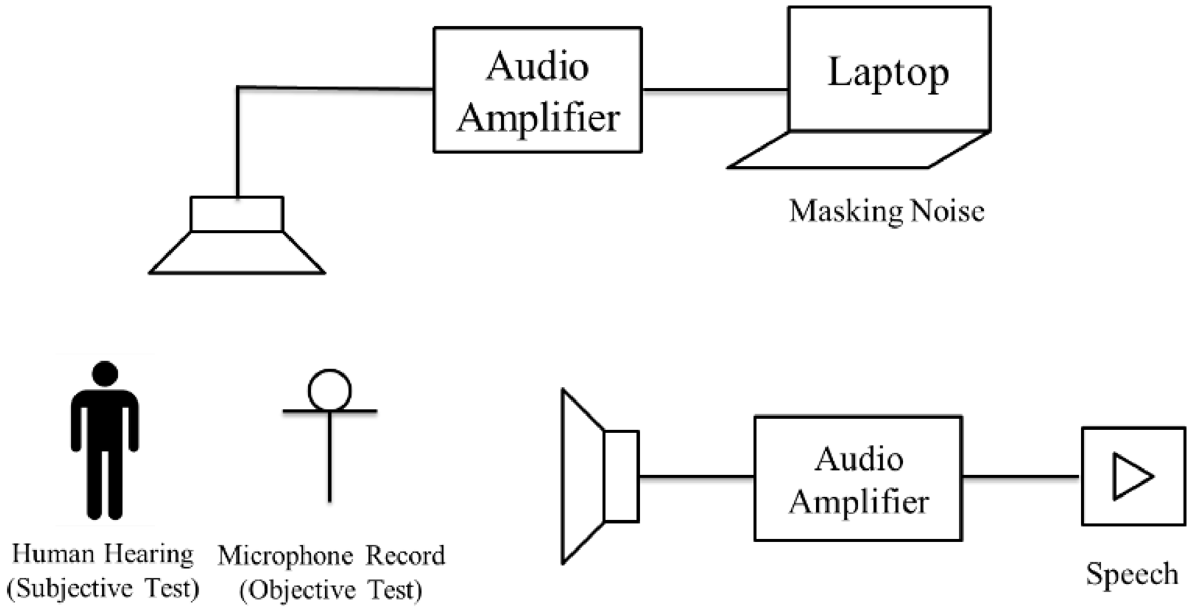


Figure 6.1 Abstract diagram for the test set up.

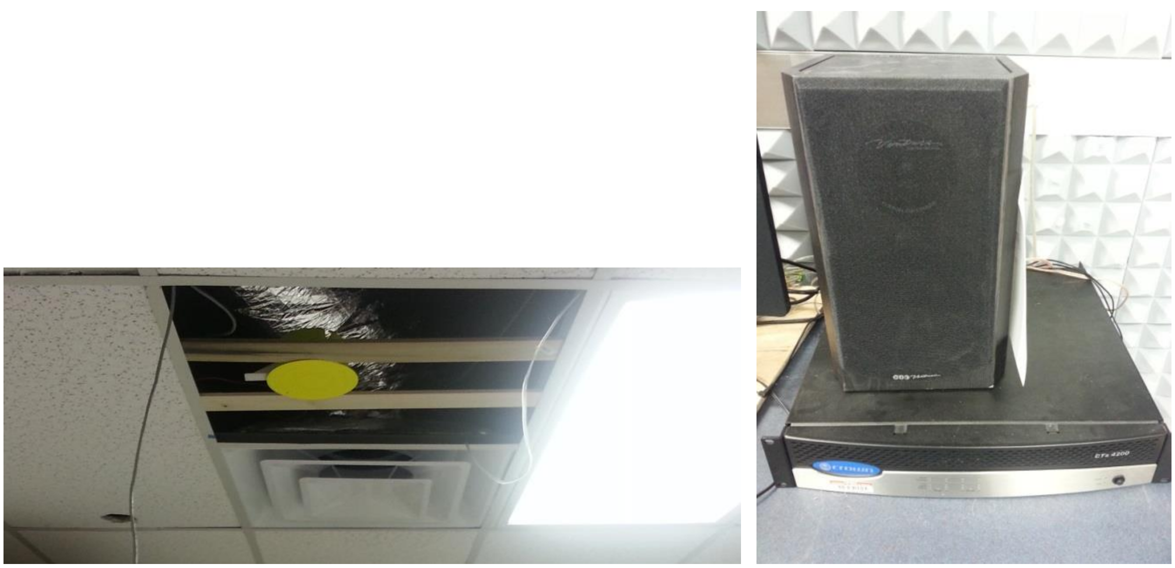


Figure 6.2 Actual test set up photo of loudspeakers for playing masking noise and speech.

6.3 Subjective Test Result

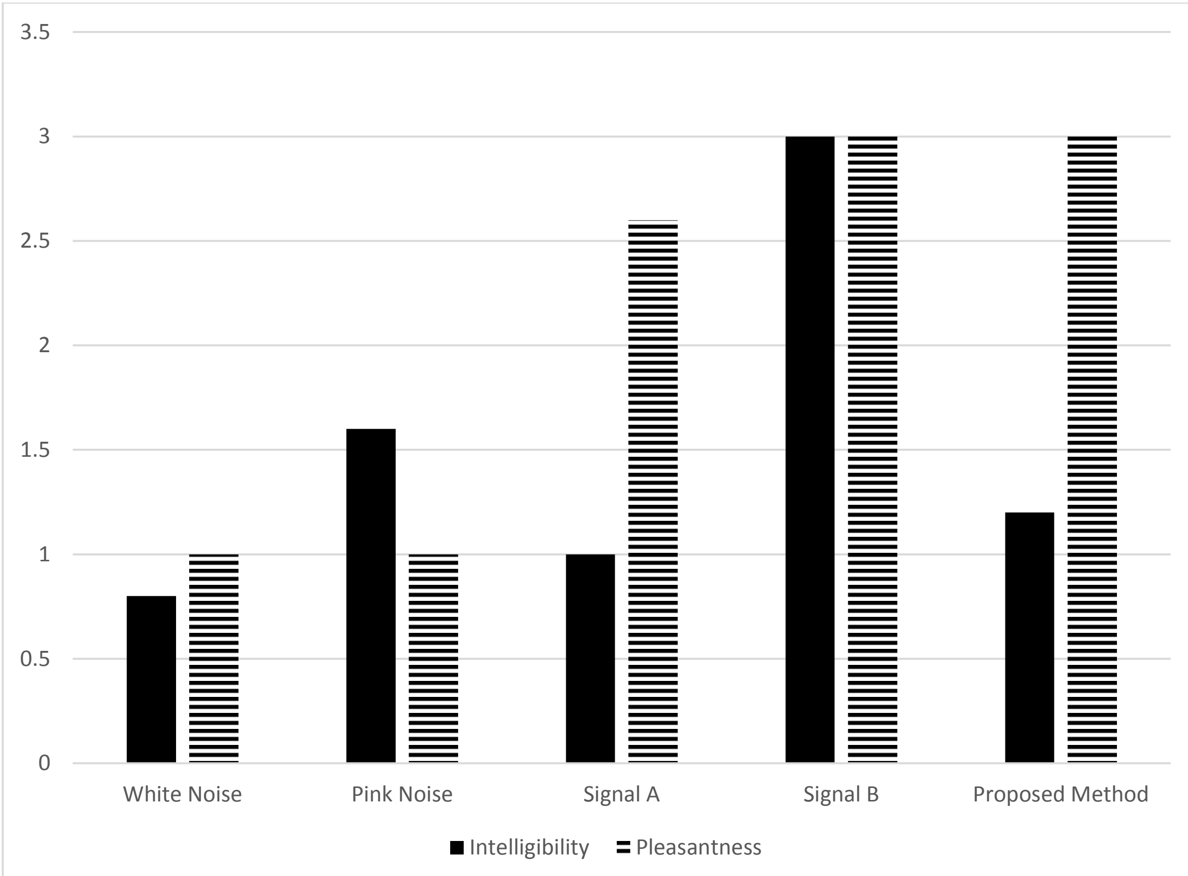


Figure 6.3 Subjective test result.

A subjective test will be done with 20 normal-hearing people. Random sentences of 3 to 5 seconds from long speeches will be played through one loudspeaker and another loudspeaker located in the ceiling will emit different kinds of masking noises. The masking signals that will be compared with the proposed noise are white noise and pink noise. These two signals have been used from the early stage of the masking system. The other two signals will be the noise source from widely used masking system products. In this test, the direction of the speaker will be facing downward directly to the listeners.

6.4 Objective Test Result

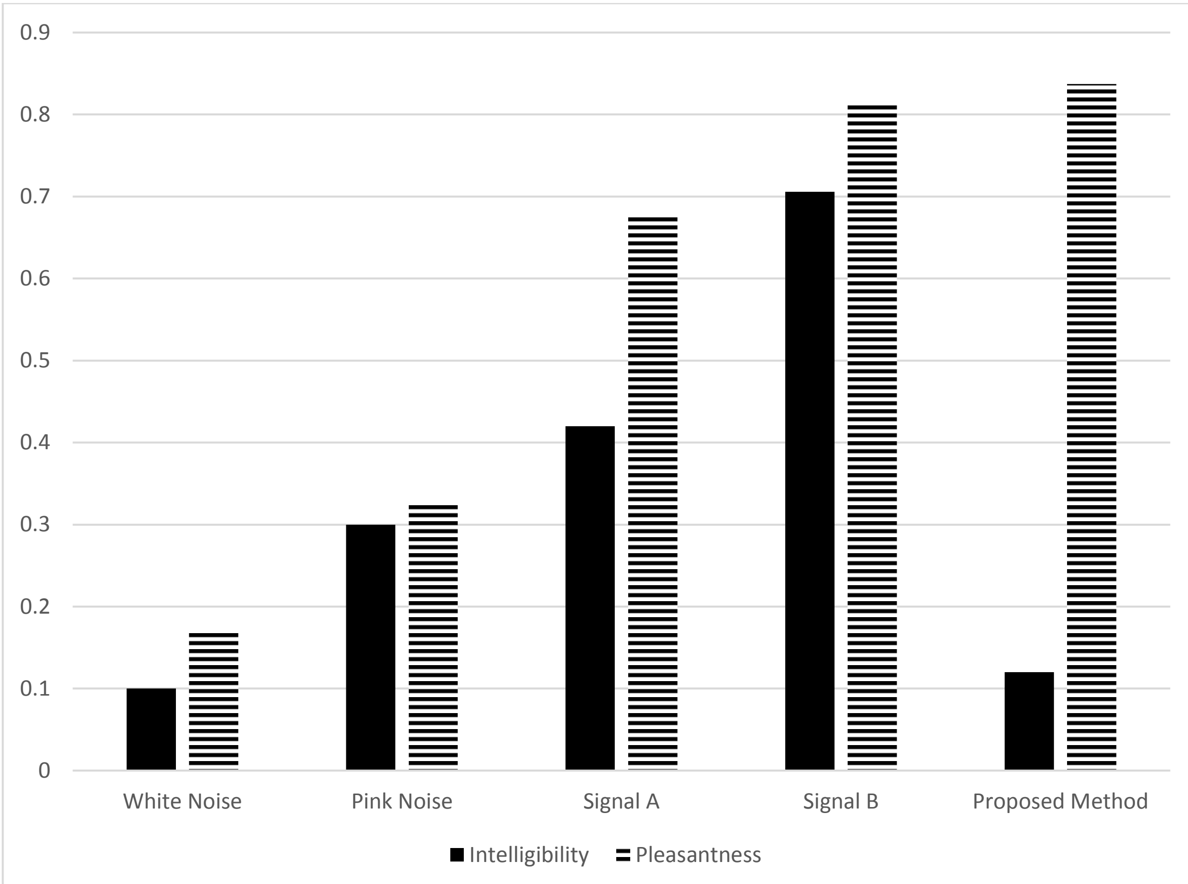


Figure 6.4 Objective test result.

To verify the masking capability and pleasantness of different masking signal candidates, we will also conduct an objective test. For this test, CSII and Zwicker's pleasantness model will be used. The test code is developed in MATLAB code. The accordance between objective test and subjective test method will also be tested. As we can see in Figure 6.4 both test result shows linearity between these two test methods. This result guarantees the credibility of the objective tests.

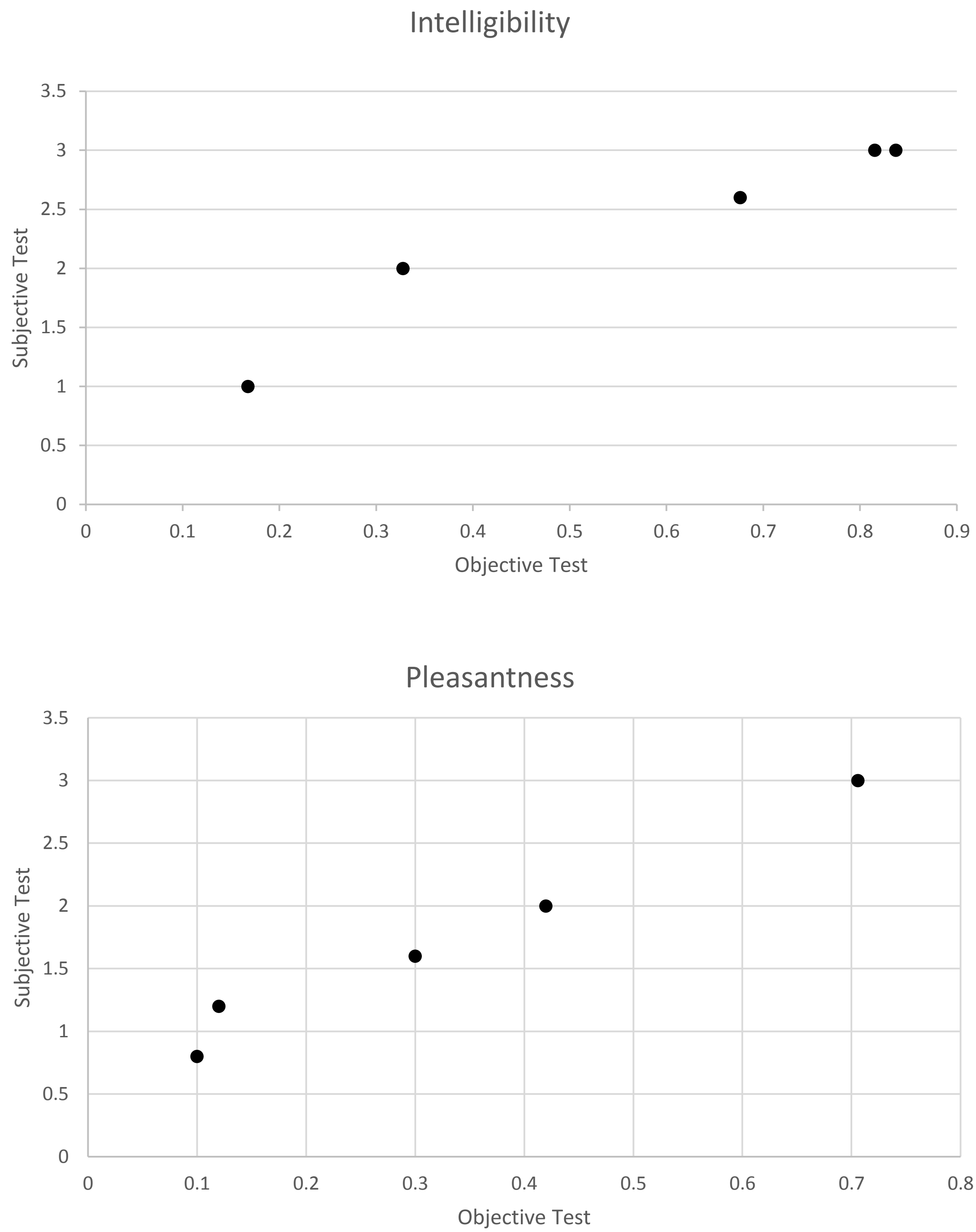


Figure 6.5 The accordance between objective and subjective test results

6.5 Effect of Improving Processes

With these test methods, we will be able to verify the two filter processes used with the proposed masking noise $\hat{x}(n)$. CSII and Zwicker's pleasantness model will be used respectively for A-weighting filter and Moving average filter.

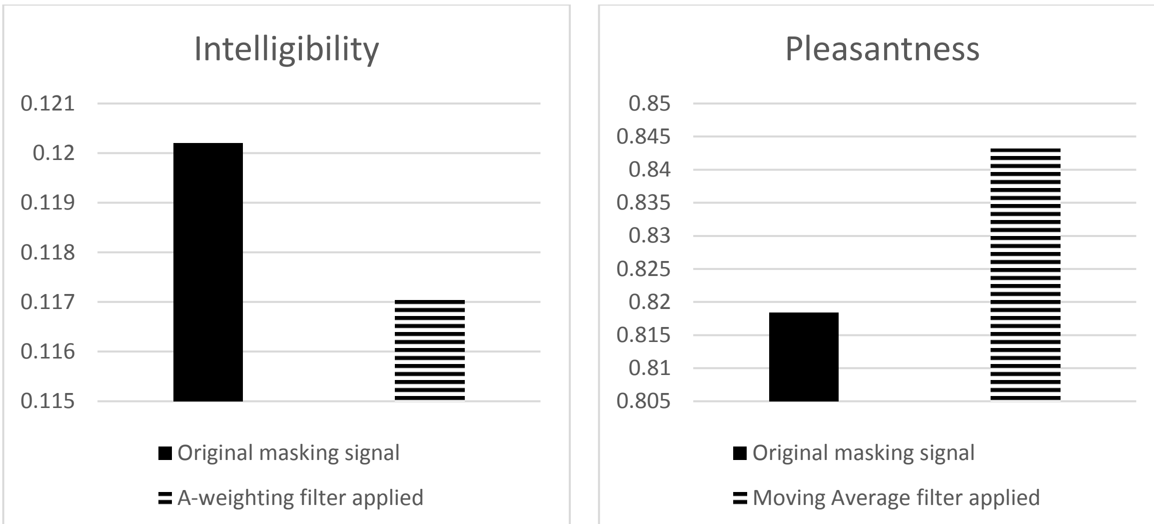
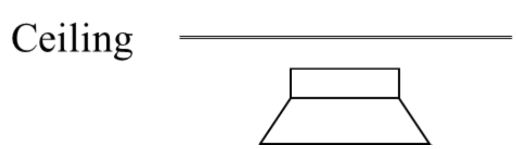
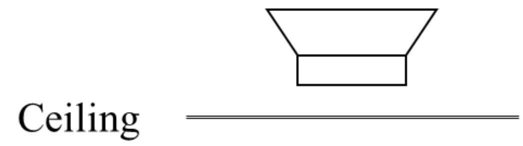


Figure 6.6 Objective test result to verify the improvements.

6.6 Effects of loudspeaker's directivity

For the masking system used in the field, the directivity and composition of the speaker is one of the main factors deciding the sound's characteristic. We will carry on an experiment to validate the consequence of directivity or direction of the loudspeaker. And we will also test the effect of the ceiling. Direct field method in which loudspeakers are facing downward directly to the listeners will have no effect from the ceiling. In an indirect field method, masking noise emitters are located up on the ceiling and facing upward. The sound will be propagated from the speaker to the space above the ceiling and then go back to the room so characteristic of the sound is modified.

Table 6.3 Comparison between the directivity of loudspeaker

Direct Field	Indirect Field
	

We will maintain all conditions constant other than the direction of a loudspeaker. White noise will coming from the emitter to clearly distinguish the effect of the ceiling.

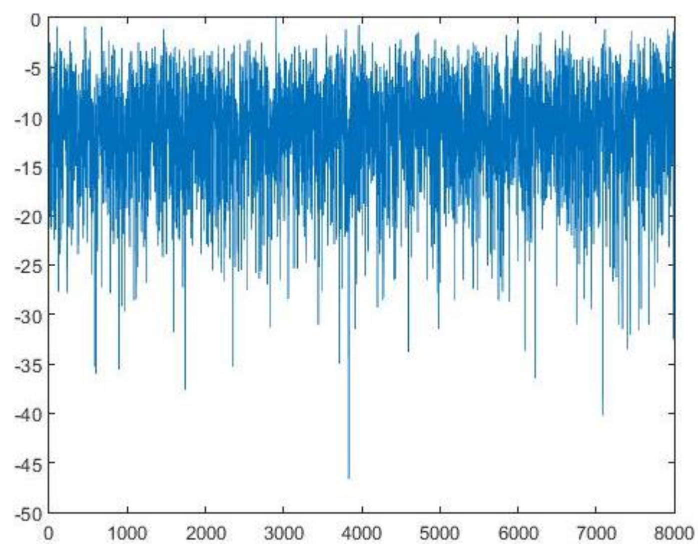


Figure 6.7 Frequency spectrum through direct field method.

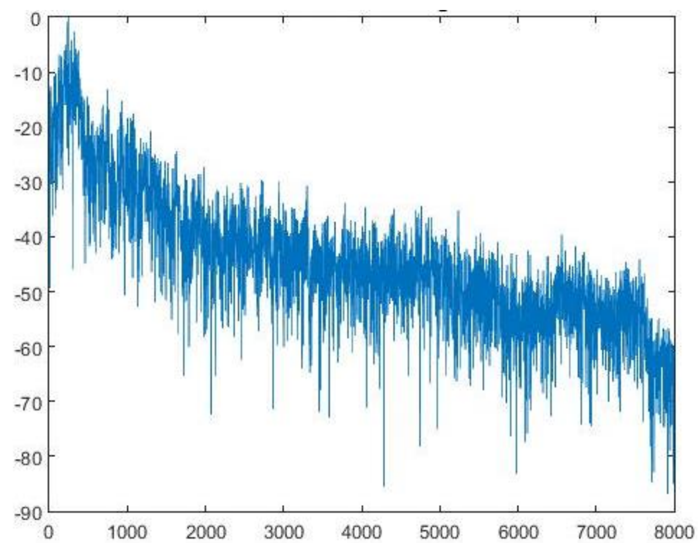


Figure 6.8 Frequency spectrum through indirect field method.

From comparing Figure 6.6 and Figure 6.7, when the loudspeaker emits sound in an indirect field it is affected by the ceiling. Most of the power in higher frequency is attenuated and energy is lost. When the indirect field method is used, we will have to take that knowledge into consideration and do another process in order to guarantee an effective result.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

Coming up with a new way of generating a signal to be used for masking unwanted speech and other noises is essential to helping contribute to an environment marked by better privacy and less distraction. The benefits of the proposed signal are that it can be used as an environmental sound and utilized as input. It is essential to adaptively generate the most appropriate masking signal to cover the ambient noise with better efficiency. Also, it is important to adopt two more filters to enhance the performance of the two important criteria of capacity and pleasantness. To enhance results, other objective deciding factors will be considered in a numerical way. Other than the masking noise itself, how the directivity of the loudspeaker could affect the outcome will also be analyzed.

7.2 Future Work

Sound masking is a comparably old fashioned way of dealing with sound. With active noise control, many other products are emerging and progressing rapidly such as sound absorbing brick and noise canceling machine for small spaces [39-41]. Also, the human ear perceives most sensitively high volume and high pitch sounds. If these sounds even in a small amount could be removed, sound masking won't necessarily have to be that high energy. As this lower energy form of masking occurs, discomfort will be decreased while pleasantness and easiness increases. Along with the progress of these noise related techniques, this experiment could also open other various ways and possible collaborations with other masking systems.

REFERENCES

[1] Allen Cho, Issa M S Panahi, “Investigation of Improved Masking Noise for the Speech Privacy,” Inter-Noise 2016, August 21, 2016.

[2] Sellars P., Behind the Mask, Cambridge: Sound on Sound, retrieved 26 February 2007

[3] K.K. Paliwal and B.T. Lily, “Auditory Masking Based Acoustic FrontEnd for Robust Speech Recognition”, IEEE TENCON-Speech and Image Technologies for Computing and Telecommunication, 1997.

[4] Chanaud, R.C., Concepts of Masking Sound, So. & Video Cont., Feb 1986

[5] Chanaud, R.C., Sound Masking Systems, Sound & Comm. 60-64, Aug 1984

[6] Yamamoto, K., & Nakagawa, S. (2009). Privacy protection of speech information. Proc. in 5th International Conference on Information Assurance and Security, 717-720.

[7] Speech Privacy System Website, http://www.speechprivacysystems.com/what-is-sound-masking/how -to-achieve-speech-privacy/

[8] Wikipedia, “Auditory Masking”, https://en.wikipedia.org/wiki/Auditory_masking

[9] T. Saeki, T. Tamesue, S. Yamaguchi, Study on achieving speech privacy using masking noise, Journal of Sound and Vibration 297 (2006), pp. 1088-1096

[10] O. A. A. Aly and B. O. Kanj, "Investigation of speech masking process using white Gaussian noise," MELECON 2014 - 2014 17th IEEE Mediterranean Electrotechnical Conference, Beirut, 2014, pp. 59-63.

[11] T. Tamesue and T. Saeki, "Sound masking for achieving speech privacy with parametric acoustic array speaker," Soft Computing and Intelligent Systems (SCIS), 2014 Joint 7th International Conference on and Advanced Intelligent Systems (ISIS), 15th International Symposium on, Kitakyushu, 2014, pp. 1134-1137.

[12] Wikipedia, “Colors of noise”, https://en.wikipedia.org/wiki/Colors_of_noise

[13] O. Hazrati, O. Sadjadi, P.C. Loizou, J.H.L. Hansen, "Simultaneous suppression of noise and reverberation in cochlear implants using a ratio masking strategy," Journal of the Acoustical Society of America, Vol. 134, No. 5, pp. 3759-3765, Nov. 2013

[14] J.H.L. Hansen, V. Radhakrishnan, K. Arehart, "Speech Enhancement based on Generalized Minimum Mean Square Error Estimators and Masking Properties of the Auditory System," IEEE Trans. Audio, Speech and Language Processing, vol. 14, no. 6, pp. 2049-2063, Nov. 2006.

[15] A. Natarajan, J.H.L. Hansen, K.H. Arehart, J. Rossi-Katz, "An Auditory Masked Threshold based Noise Suppression Algorithm GMMSE-AMT[ERB] for Listeners with Sensorineural Hearing Loss," EURASIP Journal of Applied DSP: Special Issue on Signal Processing Hearing Aids and Cochlear Implants, vol. 2005, no. 18, pp. 2938-2953, Oct. 2005

[16] R. Sarikaya, J.H.L. Hansen, "Auditory Masking Threshold Estimation for Broadband Noise Sources with Application to Speech Enhancement," EUROSPEECH-99: European Conference on Speech Technology, vol. 6, pp. 2571-2574, Budapest, Hungary, Sept. 1999.

[17] Moore, Brian C. J. An Introduction to the Psychology of Hearing. London: Academic, 1982.

[18] Richard H. Ehmer, "Masking Patterns of Tones," The Journal of the acoustical society of America Vol 31. 1959

[19] Stout, G. F. "The Weber-Fechner Law." A Manual of Psychology.: 199-209.

[20] Webster, Michael. "Evolving Concepts of Sensory Adaptation." F1000 Bio Rep F1000 Biology Reports 4 (2012).

[21] E. Zwicker, H. Fastl, “Psychoacoustics: Facts and Models,” Springer-Verlag Berlin Heidelberg, 3rd ed., 2007, pp. 220–246

[22] Gelfand, S.A., Hearing- An Introduction to Psychological and Physiological Acoustics 4th Ed. New York, Marcel Dekker, 2004.

[23] B. Bauer and E. Torick, "Researches in loudness measurement," in IEEE Transactions on Audio and Electroacoustics, vol. 14, no. 3, pp. 141-151, Sep 1966. doi:10.1109/TAU.1966.1161864

[24] M. Bodden, "Psychoacoustics and Industry: Instrumentation versus Experience?," Acta Acustica united with Acustica, vol. 85, pp. 604-607, 1999

[25] S. Bech and N. Zacharov, Perceptual audio evaluation : theory, method and application. Chichester, England: Wiley, 2006.

[26] Ward Jonckheere, "Masterproef Sound Quality," Elektromechanica, 2007-2008.

[27] J.G. Proakis, D.G. Manolakis, Digital signal processing: principles, algorithms, and applications, Prentice Hall, Upper Saddle River, NJ, 1996.

[28] James J. Hant and Abeer Alwan, "A psychoacoustic-masking model to predict the perception of speech-like stimuli in noise", Speech Communication, 40 (2003) 291–313.

[29] Robinson, D. W., and R. S. Dadson. "A Re-determination of the Equal-loudness Relations for Pure Tones." Br. J. Appl. Phys. British Journal of Applied Physics 7.5 (1956): 166-81.

[30] Richard L., St. Pierre, Jr. "The Impact of A-weighting Sound Pressure Level Measurements during the Evaluation of Noise Exposure." NOISE-CON, 2004

[31] H. Bao, I. Panahi, "Using A-weighting for Psychoacoustic Active Noise Control", 31st Annual International Conference of IEEE Engineering in Medicine and Biology Society, Sep. 2-6, 2009

[32] H. Bao, I. Panahi, "Psychoacoustic Active Noise Control with ITU-R 468 Noise Weighting and its Sound Quality Analysis", 32nd Annual International Conference of the IEEE EMBS, Buenos Aires, Argentina, Aug. 31-Sep.4, 2010.

[33] Ahmad, M., and D. Sundararajan. "A Fast Algorithm for Two Dimensional Median Filtering." IEEE Trans. Circuits Syst. IEEE Transactions on Circuits and Systems 34.11 (1987): 1364-374. Web.

[34] D.R. Morgan, J.C. Thi, "A delayless subband adaptive filter architecture," IEEE Trans. Signal Processing, vol. 43, pp. 18191830, Aug. 1995.

[35] E. R. Ferrara Jr., "Frequency-domain adaptive filtering," in Adaptive Filters, C.F.N. Cowan and P.M. Grant, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1985, ch. 6, pp. 145-179.

[36] H. Bao, I. Panahi, “Psychoacoustic active noise control based on delayless subband adaptive filtering”, ICASSP, 2010.

[37] P. C. Loizou, “Objective Quality and Intelligibility Measure” in Speech Enhancement: Theory and practice, 2nd ed. Boca Raton, CRC press, 2013, ch. 11, sec. 4, pp 561-564.

[38] T.Saeki, S. Yamaguchi, T. Tamesue, Study on evaluation indices of speech privacy, Proceedings of the 2003 Fall Meeting of Acoustical Society of Japan, 2003, pp. 717-718

[39] K. Kondo and H. Sakurai, "Gender-Dependent Babble Maskers Created from Multi-speaker Speech for Speech Privacy Protection," Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on, Kitakyushu, 2014, pp. 251-254.

[40] M. Akagi and Y. Irie, “Privacy protection for speech based on concepts of auditory scene analysis,” in Proc. Inter-noise. New York, NY: I-INCE, Aug. 2012.

[41] T. Modegi, "Auditory masking control system for protecting speech privacy by playing back filtered BGM sounds with flat-panel loudspeakers," SICE Annual Conference (SICE), 2013 Proceedings of, Nagoya, Japan, 2013, pp. 1663-1670.

VITA

Allen Cho, also named as Jungho Cho, was born in Salt Lake City, Utah in December 1989. He studied Electrical engineering at Kyeongpook National University in Korea from 2008 and exceeded his study in University of Texas at Dallas from August 2014. He started working at Statistical Signal Processing Research Laboratory (SSPRL) from February 2015. His project mainly includes sound masking for speech privacy. He is now pursuing his career as an audio engineer to research and develop upcoming smart devices in Samsung Electronics America.

He has submitted the following paper during his research at SSPRL under the advisory of Dr. Issa Panahi:

[1] Allen Cho, Issa M S Panahi, "Investigation of Improved Masking Noise for the Speech Privacy," Inter-Noise 2016, August 21, 2016.