MEE 07:33

Speaker Separation Investigation[1]

Fazal-e-Abbas Chaudhry

This thesis is presented as part of Degree of
Master of Science in Electrical Engineering

Blekinge Institute of Technology
August 2007

[1] This research work has been conducted under an occupational training scholarship at the Institute for Telecommunications Research (ITR), University of South Australia, Adelaide, Australia, and was sponsored by the Defence Science and Technology Organization (DSTO), Government of Australia.

External Advisor

Prof. W G (Bill) Cowley:
Institute for Telecommunications Research (ITR)
University of South Australia
SPRI Building,Mawson Lakes Campus
South Australia
Adelaide

University Advisor

Prof. Hans-Jürgen Zepernick:
Blekinge Institute of Technology
SE-371 25 Ronneby,Sweden

Compiled & Organized at

Institute for Telecommunications Research (ITR)
University of South Australia
South Australia 5095
SPRI Building,Mawson Lakes Campus
Adelaide.

And

Blekinge Institute of Technology
SE-371 25 Ronneby
Sweden

Table of Contents

Abstract

This report describes two important investigations which formed part of an overall project aimed at separating overlapping speech signals. The first investigation uses chirp signals to measure the acoustic transfer functions which would typically be found in the speaker separation project. It explains the behaviour of chirps in acoustic environments that can be further used to find the room reverberations as well, besides their relevance to measuring the transfer functions in conjunction with speaker separation. Chirps that have been used in this part are logarithmic and linear chirps. They have different lengths and are analysed in two different acoustic environments. Major findings are obtained in comparative analysis of different chirps in terms of their cross-correlations, specgrams and power spectrum magnitude.

The second investigation deals with using automatic speech recognition (ASR) system to test the performance of the speaker separation algorithm with respect to word accuracy of different speakers. Speakers were speaking in two different scenarios and these were non-overlapping and overlapping scenarios. In non-overlapping scenario speakers were speaking alone and in overlapping scenario two speakers were speaking simultaneously.

To improve the performance of speaker separation in the overlapping scenario, I was working very close with my fellow colleague Mr. Holfeld who was improving the existing speech separation algorithm. After cross-examining our findings, we improved the existing speech separation algorithm. This further led to improvement in word accuracy of the speech recognition software in overlapping scenario.

Acknowledgements

While completing this final report of my research work, I cannot ignore the fact that this is made possible due to keen interest shown by my supervisor Professor W.G (Bill) Cowley in bringing me to this level that looked complex but he gave me a new spirit and I started believing myself. So I really appreciate his attitude towards his profession and his subordinates. Especially I also give credit to Professor Hans-Jürgen Zepernick who made it possible for me to do my research work at Institute for Telecommunications Research (ITR). Professor Zepernick provided me all administrative support besides this opportunity and he kept on checking my visa progress once I applied for visa to Australian Embassy in Berlin. Also Professor Zepernick's book on pseudorandom signal processing helped me a lot in investigating maximal length sequences in context to speaker separation investigation. I would also like to appreciate Mr. Joerg Holfeld who has been my fellow colleague and in this research work and I would like to appreciate his cooperation with me in various aspects of this research work.

I also would like to thank all the staff from ITR, who really cooperated with me in case I asked for it, and they even voluntarily participated in one of the automatic speech recognition system that I tested along with Mr. Holfeld. I really enjoyed my time in ITR environment, especially the barbeques and monthly tea breaks other then my research work. I would also like to appreciate ITR as a whole for giving me an opportunity to work with them. I would like to thank Dr. Ahmad Hashemi who provided me hardware on behalf of DSTO to do ASR testing. And last but not the least I would like to thank my family back at home in motivating me throughout this research work.

Table of Figures

List of Tables

List of Abbreviations

ASR	Automatic Speech Recognition
MLS	Maximal Length Sequences
XOR	Exclusive-OR
ITR	Institute for Telecommunications Research
DSTO	Defence, Science, Technology Organization
NIST	National Institute of Standards and Technology
HMM	Hidden Markov Model
OL	Overlapping
NOL	Non-overlapping
ME	Mean Error
MA	Mean Accuracy
ASIO	Audio Stream Input Output
USB	Universal Serial Bus
WDM	Windows Driver Model
LED	Light Emitting Diode
IR	Impulse Response

Chapter 1

1 Introduction

1.1 Scope of the Thesis

This research report generally describes two important areas of signal processing which is behaviour of different chirps as excitation signals in room impulse response scenarios. This part of my research work will constitute second and third chapters. Second chapter talks about theoretical background relevant to important room impulse parameters that are useful in relation to knowing the theory of impulse response of a room. Matlab commands that are useful in generating room impulse response can be found separately in Appendix B for the interest of readers. I have described the important parameters like maximal length sequences and chirp along with auto-correlation and cross-correlation in a profound way that includes explanations, diagrams and mathematical derivations that gives a classic view to the reader about them. The third chapter describes real time measurements of different chirps using some audio editing tools and hardware used for playing and recording these chirps, followed by their analysis using Matlab as the tool behind this analysis. It also describes various acoustic parameters that I have utilized in analyzing the behaviour of different chirps in Matlab.

The second part of my research work is about automatic speech recognition system, which is nowadays being used for various speech applications. I have limited my work to speaker separation using automatic speech recognition system. I have divided this part into two chapters. The fourth chapter elaborates some background pertaining to automatic speech recognition and the automatic speech separation software that have been used for this purpose, followed by testing of this software to see the word accuracy of different speakers in two different scenarios. These scenarios are mentioned in detail in Appendix C and are very interested to read for the readers who are curious about these scenarios The fifth chapter describes the improvement in word accuracy of the speech recognition software using an adaptive speech separation algorithm and its brief description. In this chapter, I have compared word accuracy of one of the existing adaptive algorithms with an enhanced version of the same algorithm that has been developed by my colleague Mr. Holfeld. This second part of my thesis work is a joint effort made by both of us. We kept on cross-examining each other before came up with improvements in word accuracy of the existing algorithm, by an enhanced version of this algorithm.

1.2 Background

Nowadays, due to advancement in the speech processing and identifying technology, so many organizations are using automatic speech recognition systems (which are basically application software), in order to keep records of their meetings so that they can easily recover the data whenever they want to review their previous meetings. An automatic speech recognizer system can automatically transcribe the data for each participant in the conference, as a result this software is becoming popular but still there are some flaws and bugs that require some research work. One of them is the scenario when a speaker who is speaking and suddenly he/she gets interference from another speaker sitting next to him/her. As a result the automatic speech recognition system will also transcribe the data of the primary speaker. But due to interference of the other speaker, it will not be able to accurately transcribe the speech. This is due to abnormalities in the mixed speech identified by the software. The bottom line can be implementing some kind of speech adaptive algorithm that basically improves the transcription of this software. One of the most important speech separation techniques are least mean square algorithm, recursive least mean algorithm, blind source separation.

Room impulse response is another area of interest in the field of signal processing and sound and vibration analysis. There are many ways to determine room impulse response. One of them is to analyze meeting room impulse response via multi channel and mono channel measurements. Especially for room acoustic environment, it is important for specialists to come up with some innovative ideas like room impulse response of different rooms to see how important it is to build rooms that are free from background noise. This noise may be added along with other interferences and will cause the signal to become an acoustic mixture of a true signal with interference. And as a result, if automatic speech recognition software is running there, it will not be able to identify the speaker's voice accurately in terms of word accuracy so it is very important to measure and analyze room impulse response using some excitation signals. This method of room impulse response can also be effectively used in other applications like improving acoustic environment within a submarine for sonar communication.

Chapter 2

2 Acoustic Channel Measurement

2.1 Sound and its characteristics

According to [8] and my own opinion:

Sound is a type of longitudinal wave, in which the particles oscillate in the same direction of wave propagation. Transmission of sound waves through vacuum is not possible. This transmission requires medium in the form of solid, gas or liquid. At frequency range between 20 and 200000 Hz sound is audible to human ears, but the human perception of sound varies from person to person. Sound waves can be subsonic or ultrasonic depending upon their frequency range in comparison with audible waves. If frequency of sound waves is below the range of audible waves, then it falls in category of subsonic wave and if greater then audible then it is ultrasonic.

2.2 Reflection of Sound

These concepts are based on [3] and my own visualization:

Room acoustics is important once it comes to measure room impulse response. A sound wave is a spherical wave with aperture that gradually vanishes after being originated from a sound producing source like loud speakers, musical instruments. A sound wave is similar to light wave in terms of propagation, except having a different propagation velocity.

Suppose we follow a sound ray originating from a sound source on its way through a closed room. Then it is being observed that it is reflected not once, but many times, from the walls, the ceilings, floor and other shining surfaces. This succession of reflection continues until the ray arrives at a perfectly absorbent surface. But even if there is no perfectly absorbent area in our enclosure, the energy carried by the ray will become vanishingly small after some time. This is because during this free propagation in air as well as with each reflection, a certain part of it is lost by absorption. Sound wave propagates through air as a longitudinal wave. The speed of sound is determined by the properties of the air, and not by the frequency or amplitude of the sound. Sound waves, as well as most other types of waves, can be described in terms of the following basic wave phenomena. If a sound wave strikes a plane surface, it is usually reflected from it. This process takes place according to the principle of reflection. It says that sound wave remains in the plane including the incident ray and the normal to the surface, and that the angle between the incident rate and reflected ray is halved by the normal to the wall. In a room if a sound is being generated it may give rise to non-uniform distributions of sound energy as a result of its reflection from a surface especially a concave surface. So we can detect echoes and other abnormalities as result of reflection. Normally energy of the sound wave is not completely reflected but in different proportions depending upon the kind of surface it strikes since different surfaces have different reflection properties. It may happen that some part of the energy is absorbed in the wall and some being reflected back. Sound reflections can be observed in a room. For example, let us consider a sound source generating a sound. It can be a single high power speaker and the sound is being fed to a microphone so there can be many acoustic paths from which sound can reach microphone besides a direct part which is the path sound will take to reach from loud speaker to microphone via air at some specific pressure and temperature. The other paths are known as in direct acoustic paths which also fed sound into microphone but with delayed and changed amplitude of the original sound; these are known as sound reflections. The number of reflections produced in the room depends upon reflection surfaces present in the room for example smooth doors, ceilings and other reflecting surfaces (see Figure 1).

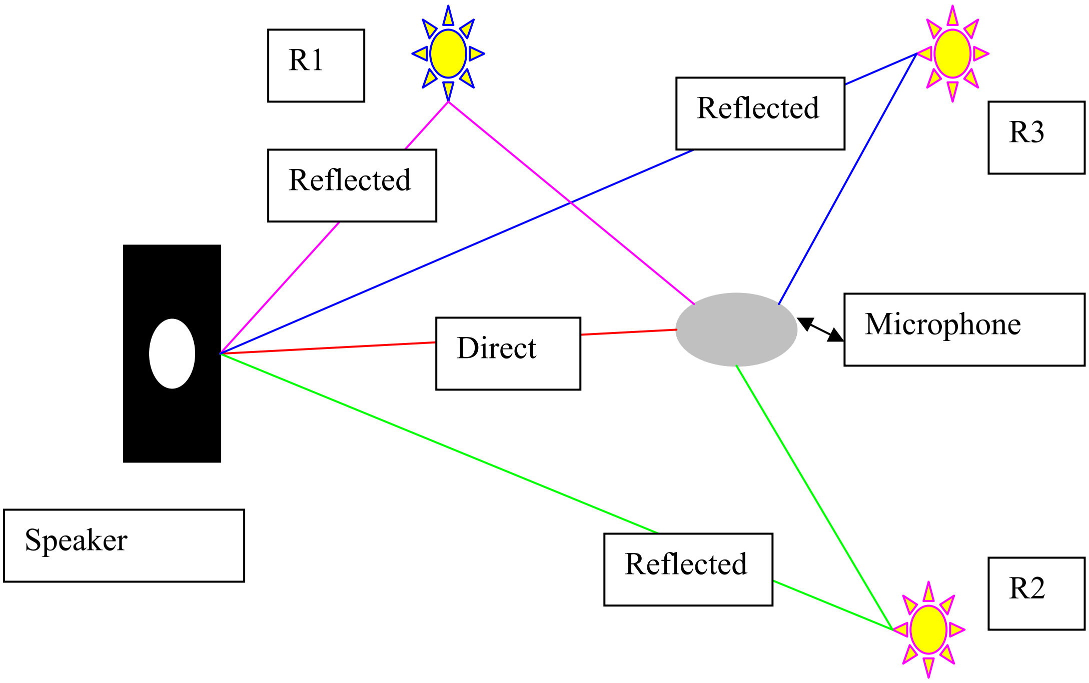


Figure 1   Sound Acoustics

In Figure 1, R1, R2 and R3 are basically reflections from three reflection surfaces present in a rectangular room. This figure indicates the direct path between the transmitter (Speaker) and the receiver (Microphone). Similarly other three reflections are depicted above. It illustrates how the reflections are being received by microphone after some different delays. All these reflections cannot be the exact replica of the original signal rather modified version of the original signal.

2.3 Deterministic Signal

These are the type of signals that have no uncertainty with respect to their amplitude values at any time. Deterministic signals are also referred to as waveforms. With a deterministic signal each value of the signal is fixed and can be determined by a mathematical expression, rule, or table. Because of this, the future values of the signal can be calculated from past values. Some of the important deterministic signals for impulse response generation are maximal length sequences and chirps which are described in the following sequence.

2.3.1 Maximal Length Sequences

Maximum length sequences are a kind of sequences that are pseudorandom in nature and are used for various applications in signal processing. One of the important sequences are binary maximum length sequences and their length is equal to (2^n-1), where n is basically the length index in this case. The number of ones in a maximal length sequence is one greater then number of zeros that is why the sum of maximal length sequence or its length is equal to (2^n-1).

The main component of a maximal length sequence is an exclusive-or gate, which is a digital logic gate that behaves as shown in Table 2.1.

Table 2.1: Exclusive-OR Truth Table

Input		Output
A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

Where XOR in the above table is symbol that is being used for exclusive-or operation between states A and B.

With correct positioning of the exclusive-or, it is possible to obtain different kind of maximal length sequence signals. A really important property of this signal is that by auto-correlation itself, it is possible to obtain impulse response and this will be demonstrated in this same chapter.

One of the important logical gate exclusive-or can be useful in generating maximal length sequences of different lengths. It all depends upon its positioning. Another important component of maximal length sequence is a linear feedback shift register. A linear feedback shift register is one whose input bit is a linear function of its previous state. Since the operation of this register is deterministic, the sequence of values produced by the register is completely determined by its current or previous state. Likewise, because the register has a finite number of possible states, it therefore produces a cycle that repeats itself. A diagrammatical visualization of linear feedback shift register is shown in Figure 2.

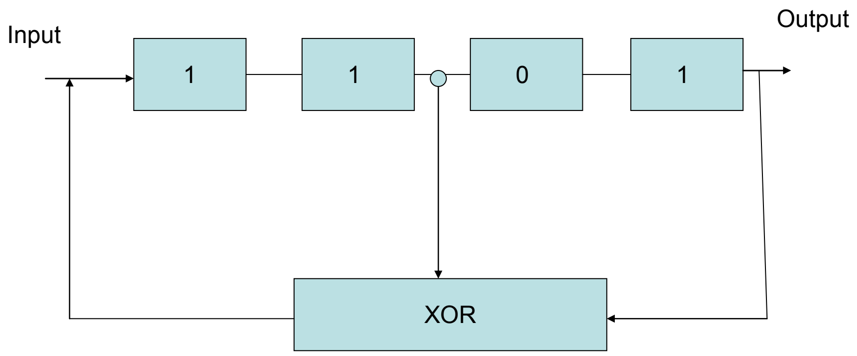


Figure 2 Linear Feedback Mechanism

2.3.2 Chirps

Chirps are basically defined as a specific range of frequencies sweeping in a finite time interval with a defined sampling rate. They are very easy to generate and perform better in acoustic environments where there are non-linearities that are being added to an excitation signal compared to other excitation signals like maximal length sequences. They are also periodic and deterministic like maximal length sequences. They have better signal to noise ratio in case of low frequencies and can be used to exploit the behaviour of loud speaker response in conjunction with speaker separation. I have decided to use chirps in order to investigate their behaviour in different room acoustic environments.

2.4 Auto-correlation

From [2] the auto-correlation can be understood from the explanation below:

Let x (t) be an energy signal which is assumed to be complex valued. Auto-correlation of this complex valued energy signal is being defined as the function of the energy signal x (t) for a lag τ

$$Rxx(\tau)= \int x(t)x^*(t- \tau)d\ \tau \qquad (2.1)$$

According to (2.1), the auto-correlation function $Rxx(\tau)$ provides a measure of similarity between the signal x(t) and its delayed version $x(t- \tau)$. Here the delayed version of the original signal is delayed by an amount τ, which basically shows similarity between the transmitted and delayed version of this transmitted signal, the less value of τ, shows a good auto-correlation so its one of the important property of the excitation signals and due to this they are used for room excitation.

2.5 Cross-correlation

As mentioned earlier, auto-correlation function provides measure of similarity between a signal and its own time delayed version. In a similar way cross-correlation is a function which depicts measure of the similarity between one signal and time delayed version of a second signal. Let x (t) is a transmitted signal and y (t- τ) the delayed version of the received signal so the cross-correlation between these two signals is given as

$$Rxy(\tau)= \int x(t)y(t- \tau)dt \qquad (2.2)$$

Another way of describing cross-correlation between the two energy signals can be described by

$$Ryx(\tau)= \int y(t)x(t- \tau)dt \qquad (2.3)$$

2.6 Real Time Auto and Cross-correlation Scenario

Figure 3 illustrate the system implementation for two different excitation signals generating auto-correlation and cross-correlation in conjunction with room impulse response.

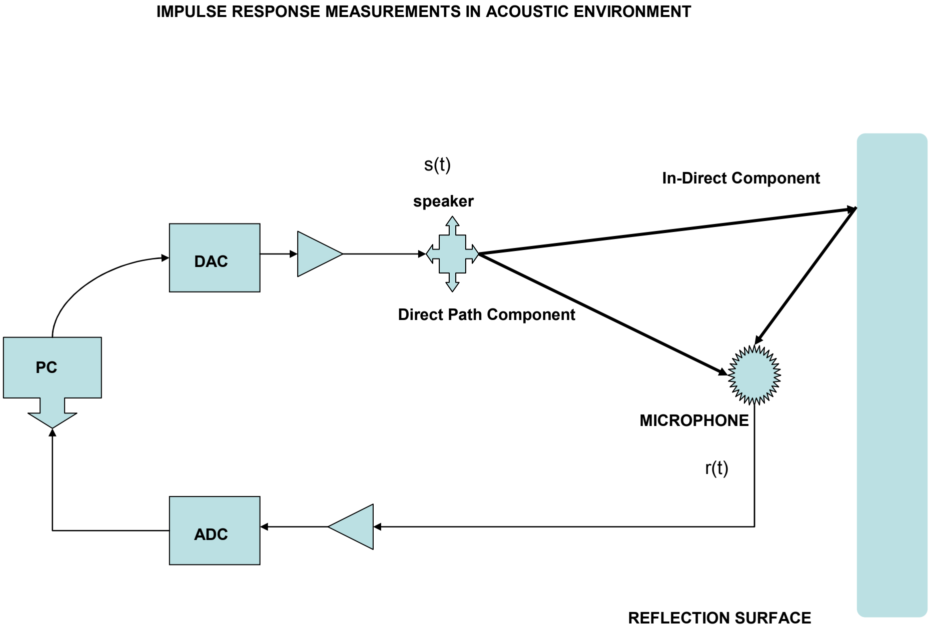


Figure 3 Impulse Response in Acoustic Environment

Here I have used a maximal length sequence and a chirp for carrying out analysis of auto-correlation and cross-correlation respectively. Both these deterministic signals are being represented as s(t) in time domain and r(t) as the received sequence in time domain so its basically implementation of LTI (Linear Time Invariant) system using deterministic signals. For maximal length sequence, the length of the binary sequence in this case is 262143 and its duration is 5.84 seconds and the sampling frequency is 44.1 kHz , along with 16 bit mono for bit resolution that are suitable for such kind of measurements. I have used distance of 1 meter between the microphone and speaker. So in above figure in case of maximal length sequence, the sequence is transmitted as s(t) from the speaker but before this it is being passed through a digital to analog converter (DAC), since it has to be analog to pass through the speaker in continuous time duration until it ends. From here we can assume that its auto-correlation is given as

$$E[s\ (t-\tau).s\ (t)] = \delta\ (t) \tag{2.4}$$

From the above equation we can say that auto-correlation is the multiplication of the original maximal length sequence with its own delayed version and this can be visualized in Matlab by a command wavread (Appendix B). This transmitted sequence which is being saved as an audio extension by another command wavwrite (Appendix B) is now being simulated by a Matlab source code named room_dem3 (Appendix F). We can visualize the central portion of the auto-correlation as shown in Figure 4.

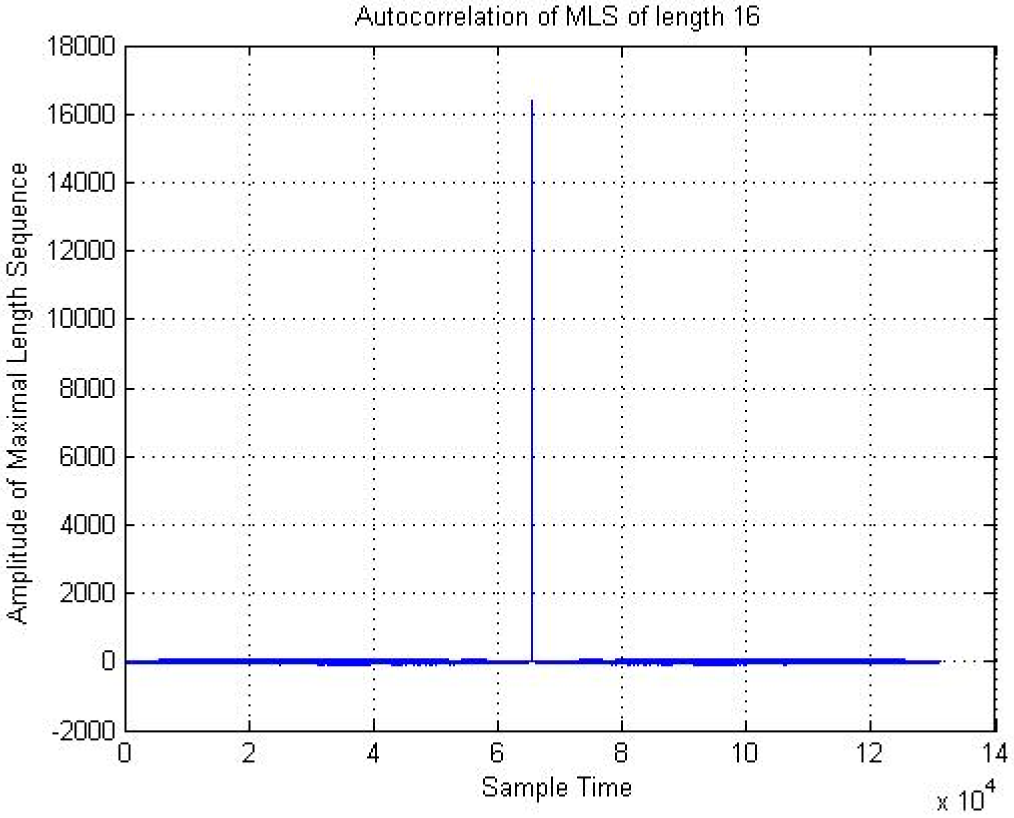


Figure 4 Auto-correlation of MLS

Similarly in case of cross-correlation, I have used the same setup, but in this case I am using a chirp for the measurement. Also, I am analyzing its cross-correlation with the received signal. The chirp is sweeping between 200 Hz and 14000 Hz, at a sampling rate of 44.1 kHz and time duration of 10 seconds. Mathematical notation for the cross-correlation is elaborated below:

$$r\ (t) = s(t)*h(t)$$

So r(t) is basically the convolution of the transmitted sequence s(t) with its impulse response h(t), that follows

$$\int E\{s\ (t-\tau)\}.r\ (t)\ d\ \tau = E\{s\ (t-\ \tau) \int s\ (t-\lambda)\ h\ (\lambda)\ d\lambda\} \qquad (2.5)$$

$$= \int E\ \{s\ (t-\ \tau)\ s\ (t-\lambda)\}\ h\ (\lambda)\ d\lambda$$

$$= \int \delta\ (\tau\ -\lambda)\ h\ (\lambda)\ d\lambda$$

$$= h\ (\tau)$$

The limit of integral here is - ∞ to + ∞.

The above derivation gives an impulse response derived from auto-correlation of s (t) which in this case is a chirp.

$$Rsr = E \int s\ (t)\ r\ (t-\ \tau)\ d\ \tau \qquad (2.4)$$

Actually, the cross-correlation above tells us the estimation of impulse response in terms of the original impulse followed by the delayed and less amplitude of the received sequence that combined together produces cross-correlation.

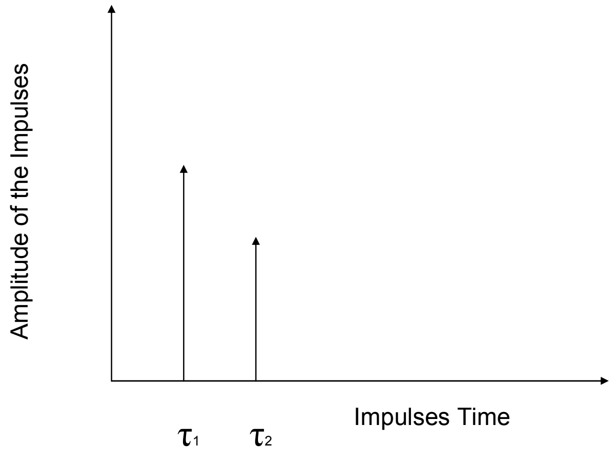


Figure 5 Impulse Response Estimation

Figure 5 illustrates that the original impulse occurs at τ_1 and then another version of the received sequence as τ_2 which is delayed with less amplitude of the transmitted sequence.

Cross-correlation of a logarithmic chirp is illustrated in Figure 6 which basically is a simulation in Matlab .

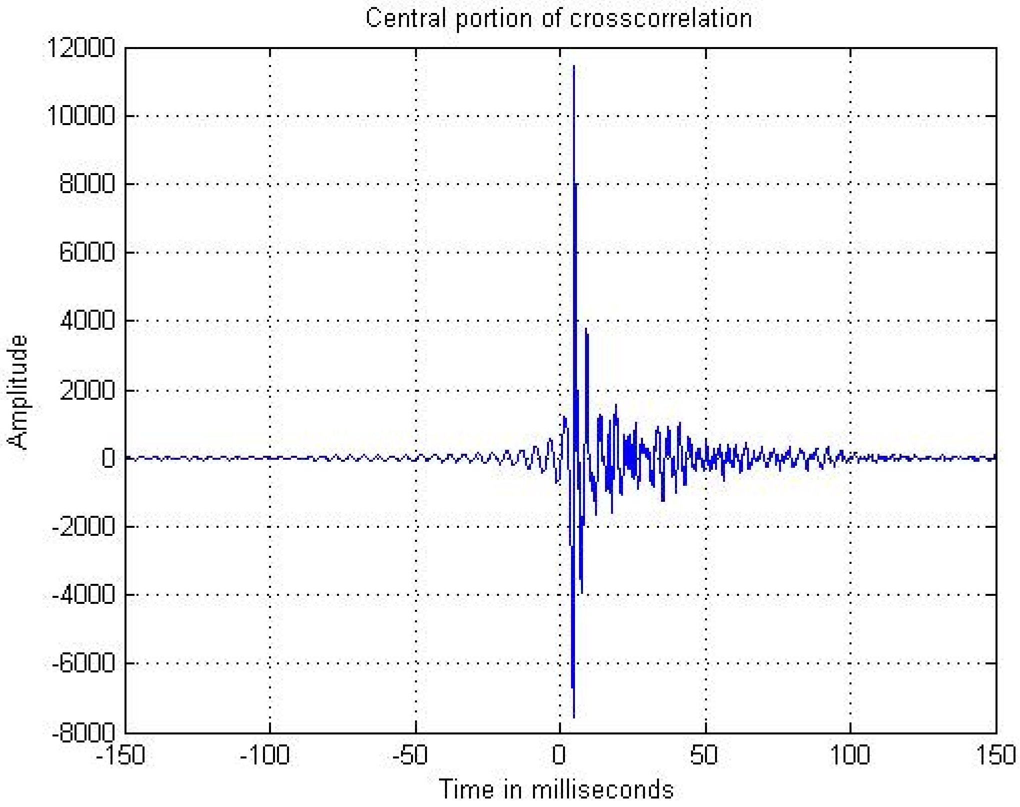


Figure 6 Cross-correlation of a Logarithmic Chirp

2.7 Specgrams

Another important parameter for analyzing the room impulse response is visualizing specgram of the excitation signal in Matlab. This provision is applicable to chirps only, so with the help of chirps we can investigate the received sequence in terms of the harmonic distortion and ambient noise obtained from an acoustic environment which has a kind of setup that makes the visibility of these chirps along with their harmonics and other noise. Chirps can be linear or logarithmic in this case. Below are two specgrams that I have obtained from different scenarios named Gallery and Room-7. Further details of specgram can be found in Appendix B.

Gallery Specgram of a linear chirp as shown in Figure 7 which is linear through out its occurrence, with duration of 10 seconds and start and stop frequencies of (200-14000 Hz). Also in this specgram of the received sequence we can see harmonics along with interference in the form of ambient noise. Also the presence of a direct component is prominent at 0.6 Hz that looks like the normalized version of the sampling frequency. Another important observation is that of presence of another direct part component along with noise component stretching along x-axis. This can be another area of investigation in conjunction with future work.

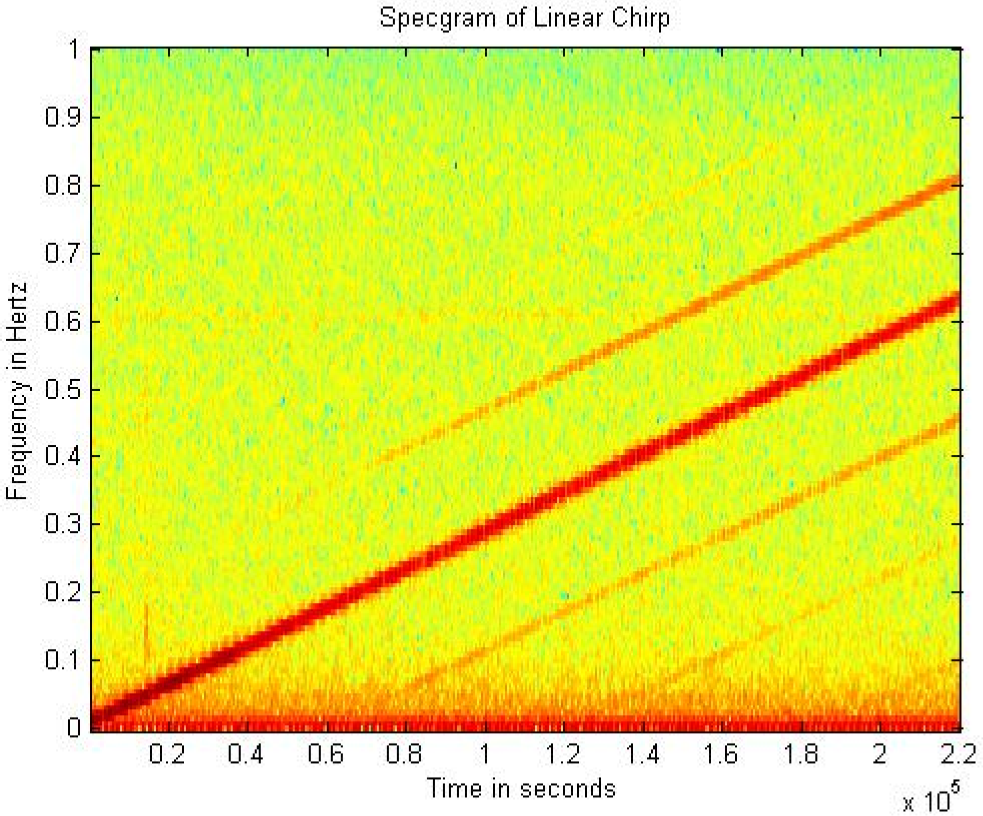


Figure 7 Specgram of a Linear Chirp

Figure 8 shows the Room-7 Specgram of a logarithmic chirp that is going exponentially in amplitude from lower to high frequencies, with duration of 10 seconds and frequency spectrum (200-14000 Hz). In this specgram of the received sequence, we can see three harmonics along with the direct path received signal. Sampling frequency is the same (44.1 kHz) that is being used for all the measurements. Another important observation is that one of the harmonic is occurring ahead then original signal. This can be another area of investigation in conjunction with future work.

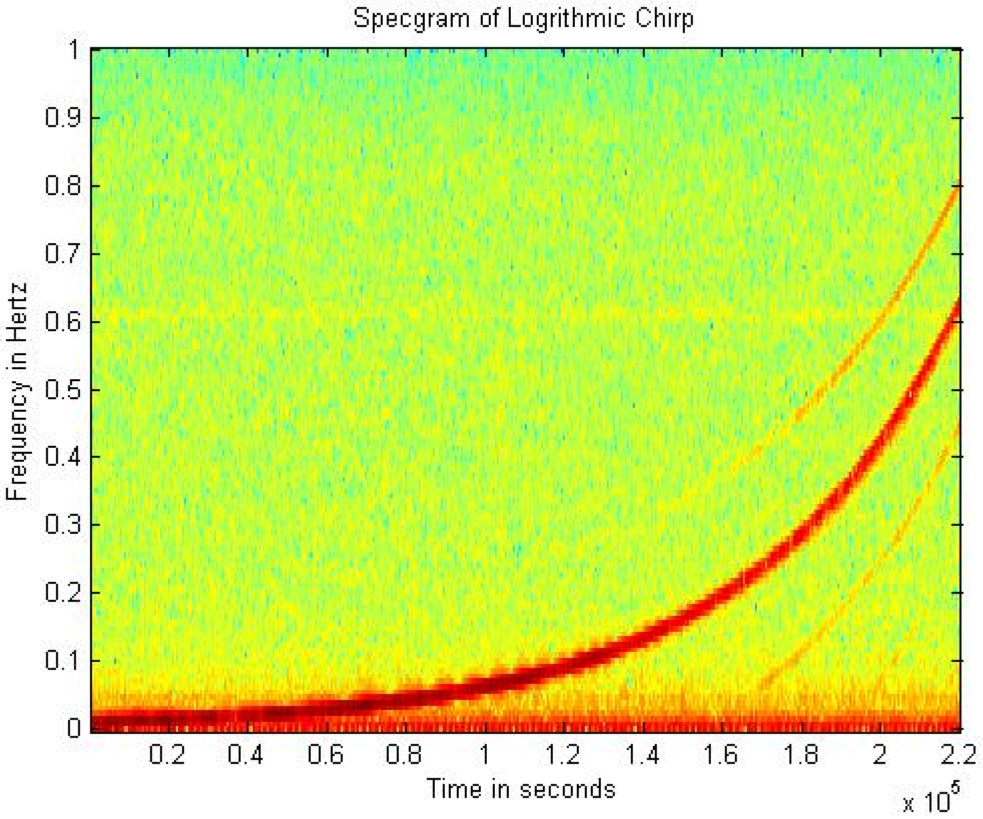


Figure 8   Specgram of a Logarithmic Chirp

CHAPTER 3

3 Real Time Impulse Response Measurements

3.1 Excitation Signals for Impulse Response Measurement

In acoustic path measurements, one of the important parameters is measuring the impulse response (IR) of a system in terms of its linearity. In case of an ideal system, we can get impulse response by a direct measurement, simply by applying a chirp (linear or logarithmic) and recording the resulting impulse response. In actual scenario, however, such measurements will usually suffer from either a poor signal to noise ratio due to the low energy contents in the short pulse or suffer from overloading in case it is increased to improve signal to noise ratio.

3.2 Chirps as Excitation Signals

Chirps are often used in area of room acoustics and audio engineers are carrying out extensive testing of these chirps in order to utilize them for impulse response of various systems for different applications. Chirps whose frequency increases at a linear rate in time are known as linear chirps and chirps whose frequency increases exponentially are known as logarithmic chirps. Once a linear chirp is being generated, harmonic distortions present in its impulse response tends to spread over the time axis but in case of logarithmic chirps, these distortions spread at very precise time. Also in a logarithmic chirp of long duration each distortion in the form of harmonics can have an independent impulse response without overlapping to other harmonic. For example, a second order harmonic or reflection of a longer duration logarithmic chirp can have its own impulse response and same goes with third order harmonic of a log chirp. In terms of signal to noise ratio, logarithmic chirps have good signal to noise ratio especially in its lower frequency components. So I have used two scenarios for carrying out measurements of chirps of different durations and types and these are mentioned is the following section.

3.3 Gallery Scenario

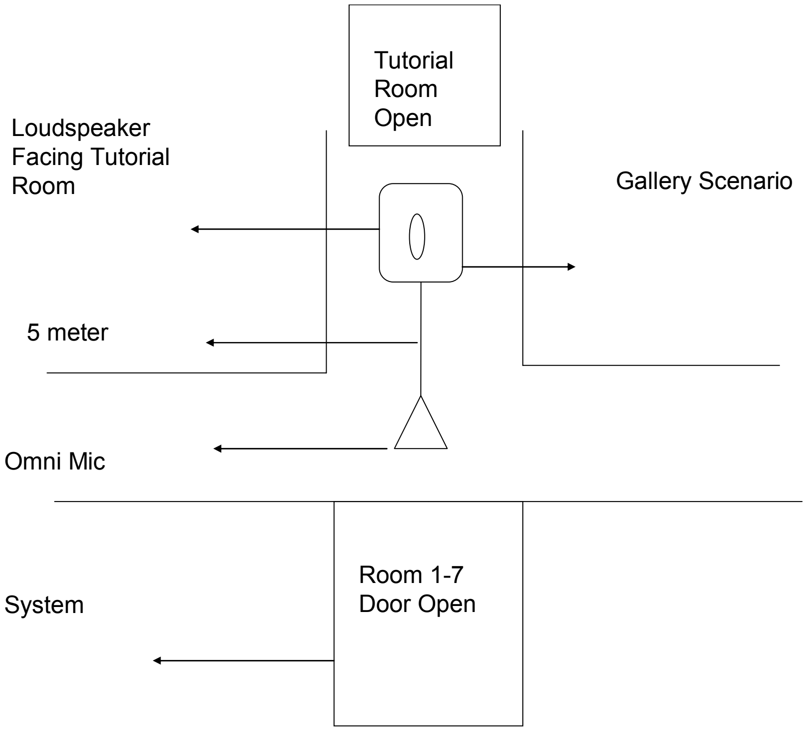


Figure 9 Chirps IR Generated in ITR Gallery

Description of transmitted linear chirp = tx_linear_swp.wav

The transmitted sweep spans between 200 to 14000 Hz, with a sampling frequency of 44100 Hz. This is a high quality sampling frequency for acoustic measurements, and is consider as standard or default sampling frequency for most of the sound cards that are used for acoustic measurements. Both the softwares “Cool Edit” and “Cakewalk” are using the same sampling rates. Their resolution is 16 bits mono, which is again a standard that actually describes greater bit depths in acoustic measurements and results into good signal to noise ratios and improved flexible range. The duration of this sweep is 10 seconds. This sweep is being created by using some Matlab commands, like ‘chirp’ by defining some parameters like time span and frequency span and then reading them in Matlab by using ‘wavread’ command to save them as audio data. All these relevant commands can be found in Appendix B. Also this transmitted audio file is being normalized to make it free from audio clipping that causes distortion in the transmitted file.

Description of received linear chirp = Gallery _linear_swp.wav

The name Gallery is being suggested because this measurement is being done in a gallery outside the rooms to see reverberant components and is being depicted in Figure 9.

The number of samples used was 441001 for cross-correlation between tx(transmitted) and rx (received) files and sampling rate was 44100 Hz. Volume of the amplifier was 4 and volume of sound card also 3 (SiS7018). Sensitivity of microphone was 70 percent and distance between microphone and omni-directional microphone was 5.06 meter but speaker was not facing the microphone directly but in opposite direction. Duration of the chirp was 10 seconds. Direct monitor volume output was 50 percent.

3.3.1 Cross-correlation

The central part of this cross-correlated linear sweep as shown in Figure 10 has some important findings. We can see a weak impulse response of the actual transmitted signal occurring just after 0 milliseconds, followed by impulse response of the received sequence around 17 milliseconds. A multi-path component can be seen around 60 millisecond which gives an estimation of the original impulse that I have described earlier in Chapter 2. Ambient noise is stretched along with the cross-correlated sequence because of the acoustic environment. This is why we have to zoom the cross-correlated sequence and its different regions along the time axis scale, so future work can be using reverberant chambers or acoustic labs to see the impulse response and its estimation with very small acoustic noise. Intermodulation products along with reverberant components can be seen in this case. The time scale width has been adjusted to 300 milliseconds.

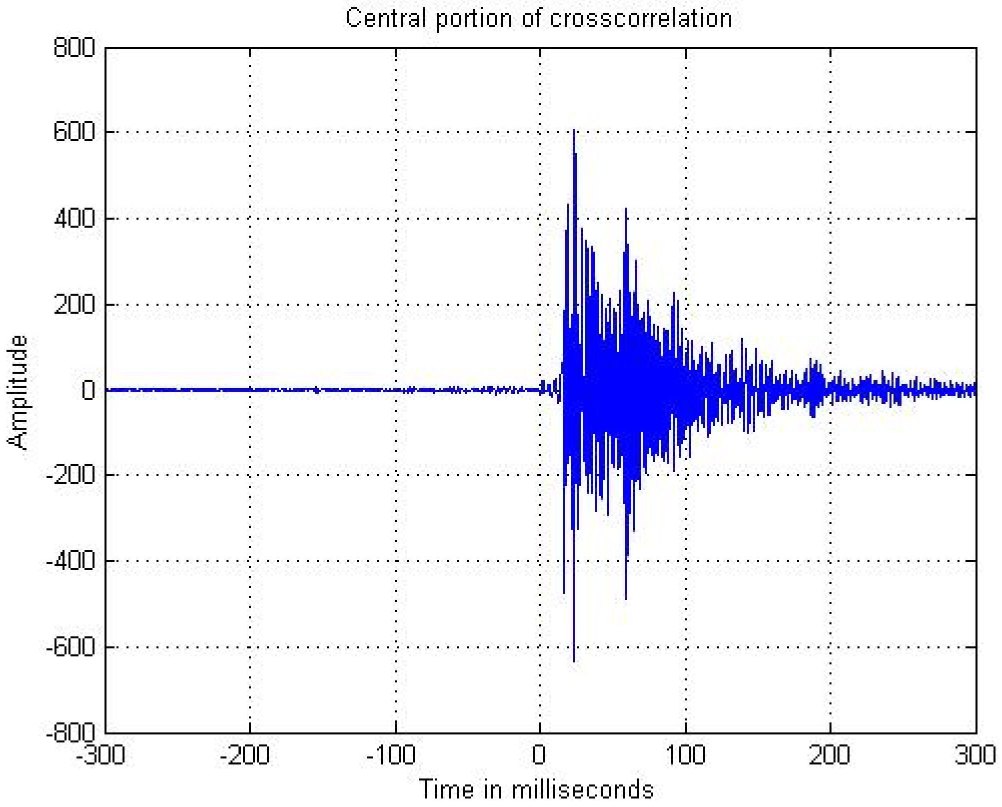


Figure 10 Cross-correlation of a Linear Chirp in Gallery Scenario

3.4 Room -7 Scenario

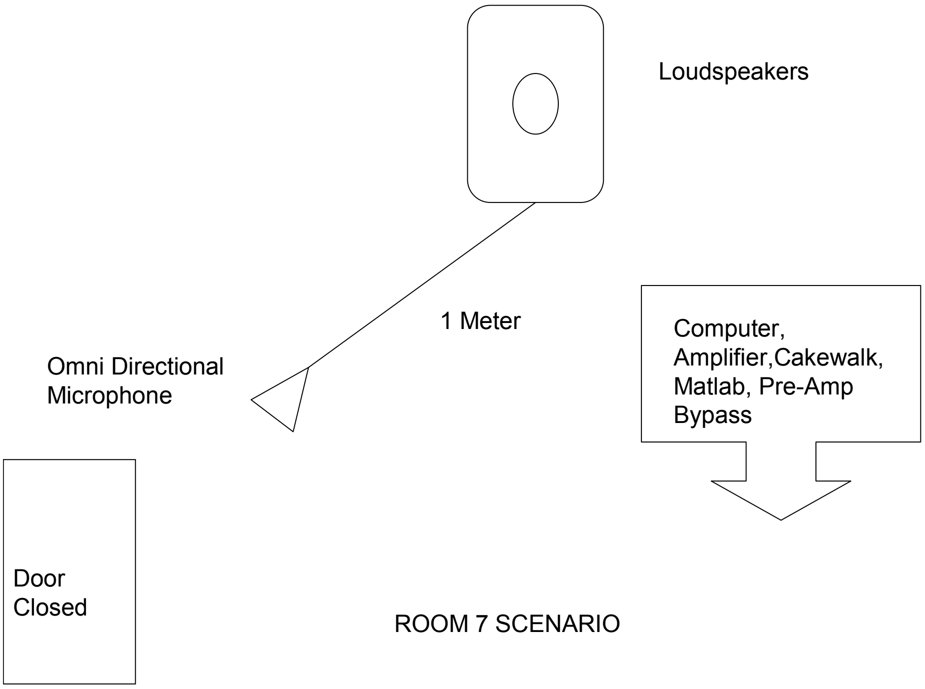

Loudspeakers
1 Meter
Omni Directional Microphone
Computer, Amplifier,Cakewalk, Matlab, Pre-Amp Bypass
Door Closed
ROOM 7 SCENARIO

Figure 11 Impulse Response of the Chirps Generated in the ITR Room-7

Description of transmitted linear chirp= tx_linear_swp.wav

Description of this transmitted chirp is same as that of the one in Gallery Scenario mentioned earlier but here acoustic environment was different as shown in Figure 11.

Description of received linear chirp = 007_linear_swp.wav

Description of this received chirp is same as that of the one in Gallery Scenario mentioned earlier but here acoustic environment was different as shown in Figure 11.

3.4.1 Cross-correlation

Middle part of the cross-correlation as depicted in Figure 12 is giving explanation about the received impulse response of the transmitted linear chirp, which is appearing around 5 milliseconds on time scale. This is followed by appearance of its multi-path component around 5.8 millisecond. Then, we can see intermodulation products appearing along with some more delayed components of the received impulse response. Between 40 to 50 milliseconds time scale, strange acoustic peaks can be seen between some of the multi-path components which can be reverberant components.

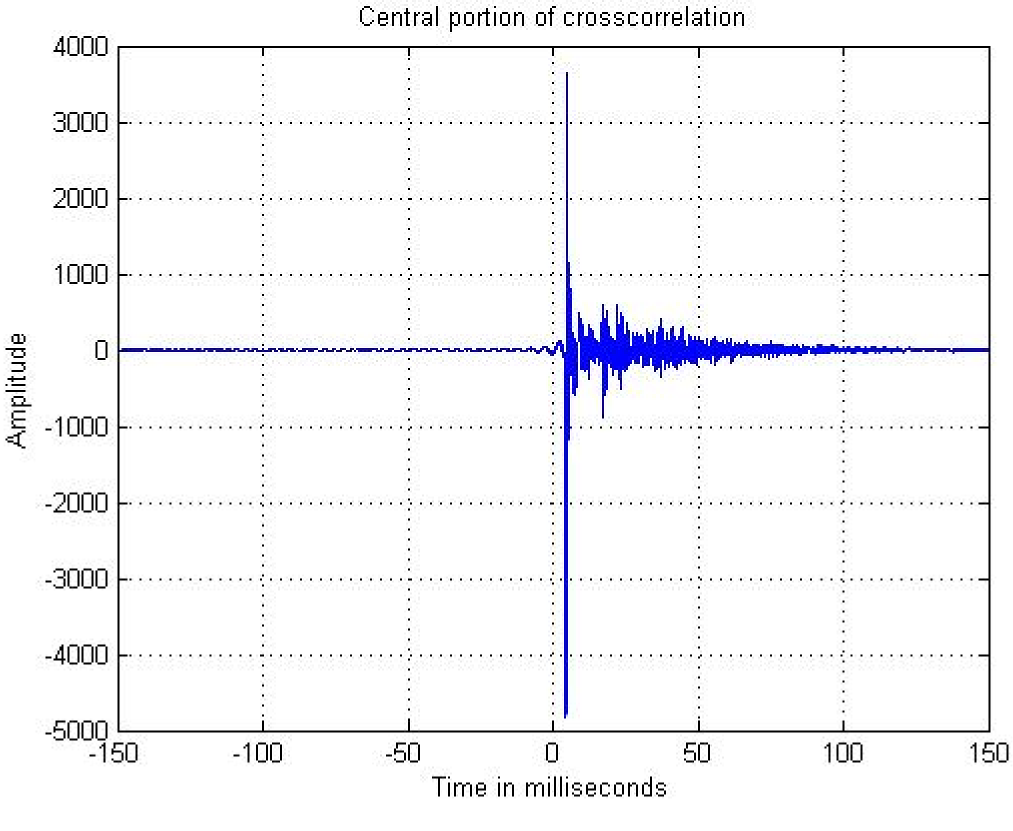


Figure 12 Cross-correlation of a Linear Chirp Generated in Room-7

3.5 Comparison of two chirps

Below is the description of two audio files saved in wav format, these are basically logarithmic in nature. I have made a comparison of these two chirps in terms of their cross-correlation, specgram and power spectrum density. Diagram for the cross-correlation is being shown and diagrams of specgrams and power spectrum magnitude can be accessed via Appendix B.

Description of transmitted logarithmic chirp = tx_log_chirp1.wav

Description of this transmitted chirp is same as that of the one mentioned earlier in two scenarios but the major difference is that it is logarithmic in nature.

Description of received logarithmic chirp = 007_log_chirp1.wav

Description of this received chirp is same as that of the one mentioned earlier in two scenarios but the major difference is that it is logarithmic in nature.

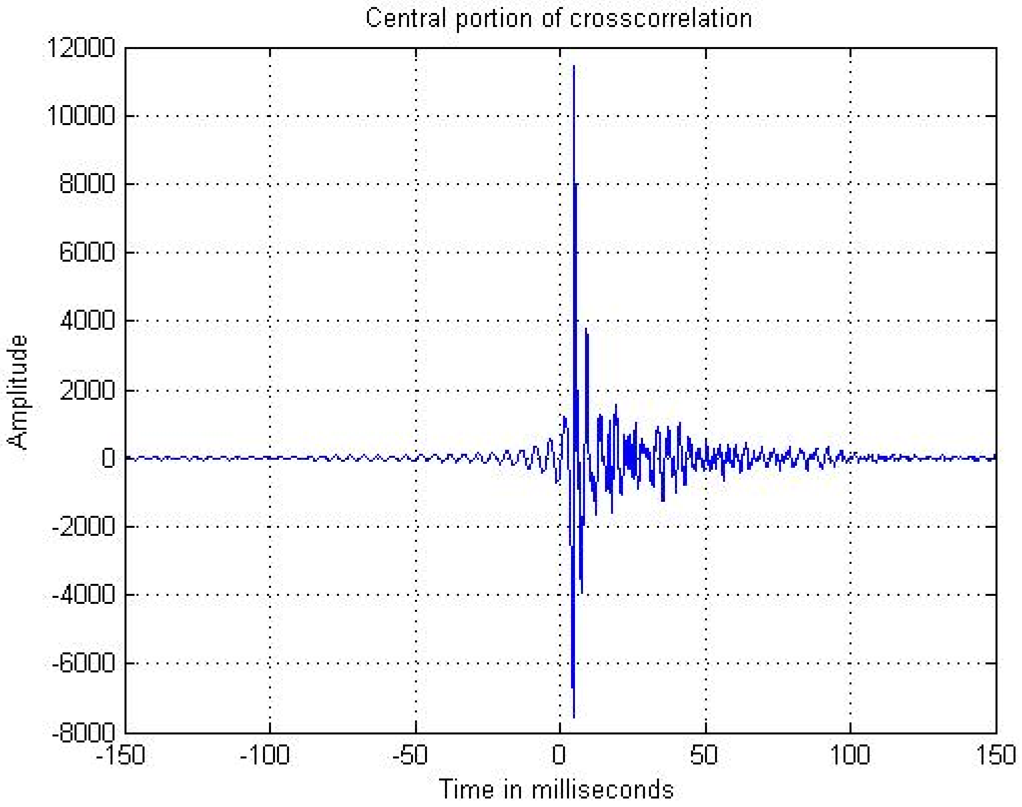


Figure 13 Cross-correlation of a Logarithmic Chirp in Room-7

Description of short transmitted logarithmic chirp = tx_log_chirp1_short.wav

The transmitted sweep spans between 200 to 4000 Hz, with a sampling frequency of 44100 Hz. The duration of this sweep is 2 seconds. Rest of its specifications are same as those of the earlier mentioned chirps

Description of short received logarithmic chirp = 007_log_chirp1_short.wav

The received sweep spans between 200 to 4000 Hz, with a sampling frequency of 44100 Hz. The duration of this sweep is 2 seconds.

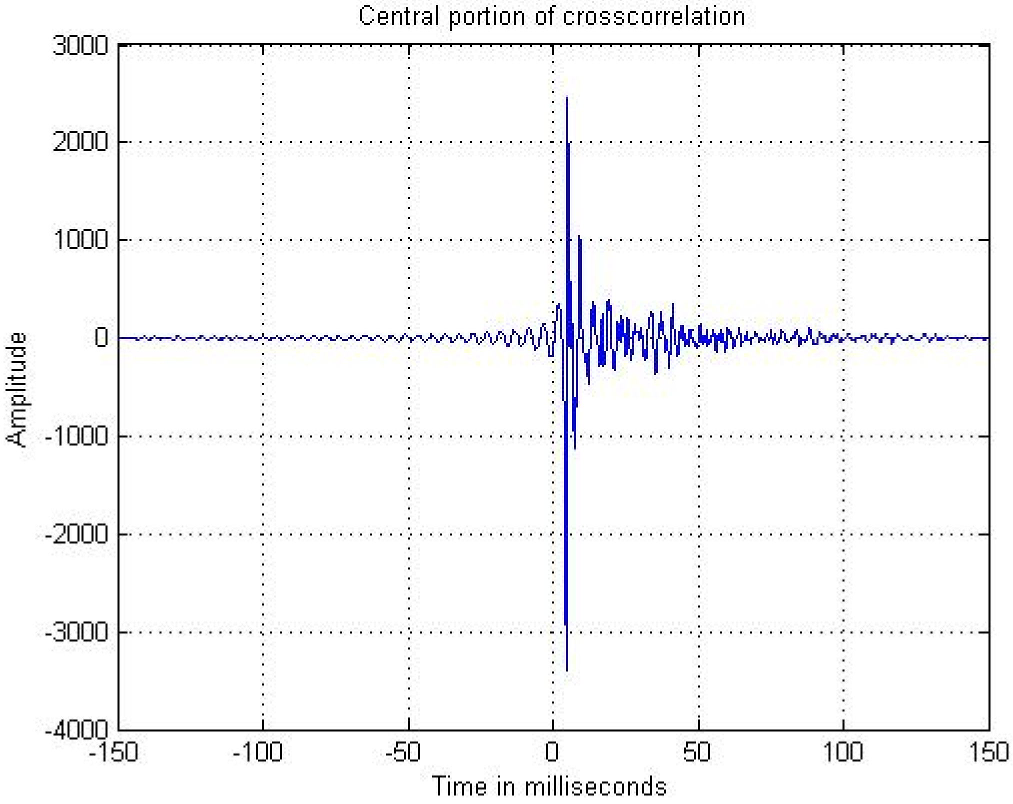


Figure 14 Cross-correlation of a Short Logarithmic Chirp in Room-7

3.5.1 Comparative Analysis of two different logarithmic chirps

Table 3.1: Comparison between two different logarithmic chirps in Room-7

tx_log_chirp1.wav	tx_log_chirp1_short.wav
007_log_chirp1.wav	007_log_chirp1_short.wav
Its duration is 10 seconds	Its duration is of 2 seconds
Its frequency span is from 200-14000 Hz	Its frequency span is from 200-4000 Hz
Logarithmic Nature	Logarithmic Nature
Number of samples is 441001	Number of samples is 88201
Specgrams (Figure 8)shows that the	The specgram (A-3: Appendix A) in this

received signal consists of received signal which is rising exponentially and magnitude of frequency overlaps is decreasing as it spans over its frequency range 200-14000 Hz. Harmonics of the received sequence are also appearing and are four in total. A direct strange component is also appearing at 0.6 Hz that looks a normalized version of the sampling frequency. Acoustic noise is there and more in magnitude near low frequency components, this ambient noise can be due to the acoustic complexity of the room. Another strange component appearing along x-axis, which also has acoustic noise in it. This chirp is of 10 seconds duration	case suggests ambient noise more, and is continuously mixing with received actual signal. Harmonic distortions are not visible here, but here again the strange direct component is visible at 0.6 Hz along y-axis. The duration of this sweep is 2 seconds and the frequency range is 200 to 4000 Hz. Also in this case the strange direct component is stretched along x-axis with ambient noise in it. Acoustic noise is dominating the low frequency components of this chirp
The impulse response Figure 13 obtained from cross-correlation of two signals, defined above is around 5 milliseconds. Estimation of this impulse response is occurring at 8 milliseconds. Strange acoustic peaks are present and prominent after second order reflection till 100 milliseconds. These strange peaks should be reverberant components.	The impulse response Figure 14 obtained from cross-correlation of two signals, defined above is around 5 milliseconds. Estimation of this impulse response is occurring between 8 and 9 milliseconds. In this case the acoustic peaks are there but not too much, one of the multi-path components is occurring between 40 and 42 milliseconds. Intermodulation products are also there.
Power spectrum magnitude(A-4: Appendix A) is distributed, it's more towards low frequency components decaying from 0 till -28 decibel between 0 till 5000 Hz and the decay is rapid at 14000 Hz as this is the end point of the sweep so there is a sudden decay without any acoustic peaks between them. Power spectrum magnitude is flat between 5000 Hz and 10000 Hz	In Power spectrum magnitude (A-5: Appendix A), acoustic peaks are there at the beginning of the decay and then it goes down sharply without any acoustic peak from -22 decibel till -50 decibel. Since this is the end point for this short logarithmic sweep and is 4000 Hz. So in this case, the power spectrum magnitude relevant to this frequency range is decayed within the first frequency block that is from 0 till 5000 Hz.

3.5.2 Comparative Analysis of two different linear chirps

Table 3.2: Comparison between two different linear chirps in Gallery

tx_linear_swp.wav	tx_linear_swp_short.wav
Gallery_linear_swp.wav	Gallery_linear_swp_short.wav
Its duration is 10 seconds	Its duration is 2 seconds
Its frequency span is from 200-14000 Hz	Its frequency span is from 200-4000 Hz

Number of samples is 441001	Number of samples is 88201
Specgram (Figure 7) shows that the received sequence (yellow) consists of a linear impulse response along with its harmonics (four in totals) and acoustic noise (red). One of the harmonic is occurring before the direct part component. There is another direct part component intercepting the actual received impulse at 0.6 Hz that looks normalized version of the sampling frequency. Presence of another strange component stretched all along the x-axis is another important thing to be investigated for future purpose, this component also has acoustic noise in it.	The received signal (A-2: Appendix A) is linear at 0.2 Hz with its harmonic close to 0.4 Hz occurring before the actual received chirp. Here again a direct part component of the received at 0.6 Hz and a strange direct component, mixed with ambient noise is stretched along x-axis
The impulse response (Figure 10) is around 16 milliseconds, and its delayed counter parts are appearing between 58 and 66 milliseconds and they are three in total. Strange acoustic peaks occurring between these multi-path components.	The received impulse response (A-6: Appendix A) around 16 milliseconds is followed by a delayed version of this impulse but equal in amplitude to the impulse response can be seen around 23.8 milliseconds. In between them are some inter-modulation products. Two multi-path component can be seen between 58 and 59.5 milliseconds, so with acoustic environments like this we can see lots of multi-path components of the direct part received sequence
Power specrtrum magnitude (A-7: Appendix A) is flat between 200 and 14000 Hz but also the presence of acoustic peaks is there. Most of the power response is between -10 decibel and -40 decibel	Power spectrum magnitude (A-8: Appendix A) is decaying with some acoustic peaks in between and then, there is a sharp decay which is very steep and it starts at -30 decibel and reaches till -60 decibel. There after it becomes flat except for a strange impulse occurring between 10 kHz and 15 kHz

CHAPTER 4

4 ASR Theory and Its Practical Analysis

4.1 Idea Behind Speech Recognition System

The following text is taken from [5].

“Speech is used to communicate information from a speaker to a listener. Speech, hearing is an essential part of the speech chain. The idea behind speech production is that the speaker wants to communicate to a listener. The speaker does this through a series of neurological processes and muscular movements to produce an acoustic sound pressure wave that is received by a listener’s auditory system, processed and converted back to the neurological signals. To achieve this, a speaker forms an idea to communicate, converts that idea into a linguistic structure by choosing appropriate words or phrases based on learned grammatical rules in conjunction with the particular language, and finally adds any additional local or global characteristics such as pitch accuracy or stress to emphasize aspects important for overall meaning. Once, this is accomplished the human brain produces a sequence of motor commands that move the various muscles of the vocal system to produce the desired sound pressure wave. This acoustic wave is received by the talker’s auditory system and converted back to a sequence of neurological pulses that provide necessary feedback for proper speech production. This allows the talker to continuously monitor and control the vocal organs by receiving his or her own speech as feedback. Any delay in this feedback to the ears can also cause difficulty in proper speech production. The acoustic wave is also transmitted through a medium normally air, to a listener's auditory system. The speech perception process begins when the listener collects the sound pressure wave at the outer ear, converts this into neurological pulses at the middle and inner ear, and interprets these pulses in the auditory cortex of the brain to determine what idea was received. We can see that in both production and perception, the human auditory system plays an important role in the ability to communicate efficiently. The auditory system has both strength and weaknesses that become more apparent by analyzing human speech production. For example, one silent feature of auditory system is selectivity in what a person wishes to listen to. This permits the listener to hear one individual voice in the presence of several persons talking at the same time for example in a party, and this phenomenon is known as cocktail party effect. A disadvantage of the auditory system is its inability to distinguish signals that are closely spaced in time or frequency. This occurs when two tones that are spaced close together in frequency, one masks the other, resulting in the perception of a single tone. There are many interrelationships between production and perception that allows individuals to communicate among one another".

4.2 Hidden Markov Model

This text is also from [5] and some of my comments.

In the last few decades lot of research work has been done on hidden Markov model. The hidden Markov model provides foundation for many successful laboratory and commercial speech recognition systems. "Hidden Markov model is in fact, a stochastic finite state automation, used to model a speech utterance. The utterance may be a word, a sub-word unit, a complete sentence or paragraph. In small vocabulary systems, the hidden Markov model tends to be used to model words, whereas in large vocabulary systems, the hidden Markov model is used for sub-words units like phones. In order to introduce the operation of the hidden Markov model, however, it is sufficient to assume that the unit of interest is a word. In case of hidden Markov model, we can analyze word in term of a set of test features, like t(1), t(2), t(3),…,t(i),…,t(I).

So we can refer to the string of test features as the observations or observables since these features represent the information that is observed from the incoming speech utterance. A hidden Markov model is normally associated with a particular word or other utterance so it is a kind of finite set machine capable of generating observation strings. A given hidden Markov model is more likely to produce observation strings that would be observed from real utterances of its associated word. During the training sequence, hidden Markov model is being taught the statistical make up of the observation strings for its dedicated word. During the recognition phase, given an incoming observation string, it is imagined that one of the existing hidden Markov model produced the observation string. The word associated with the hidden Markov model of highly likelihood is declared to be the recognized word.

A diagrammatic elaboration of hidden Markov model is shown in Figure 15. It shows about a typical hidden Markov model with four states, and each of these states is labelled by an integer. The structure or topology, of the hidden Markov model is determined by its allowable state transitions. The hidden Markov model is imagined to generate observation sequences by jumping from state to state”.

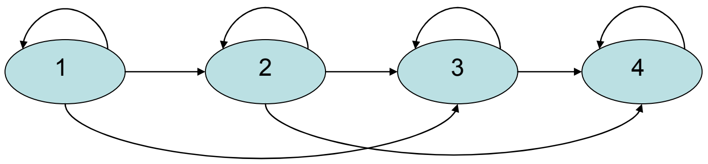


Figure 15 Four State Hidden Markov Model

4.3 Speaker Separation Testing via ASR

The testing for speaker separation using dragon software was carried out in tutorial room of ITR and numbers of speakers were ten. Eight were male speakers and 2 were female speakers. Each of the speakers was supposed to undergo a training sequence by the Dragon Speaking Software. They were being told about this whole testing scenario along with precautions relevant to speaking in front of the microphone. It took almost 25 minutes for each speaker to under go all the speaking scenarios. The distance between reference speaker and a uni-directional microphone one was 30 centimetre (cm). The distance between the interfering speaker and the reference speaker was 50 cm. The interference speaker was using omni-directional microphone. This testing was basically conducted in context to the testing done by

DSTO (the sponsors for this research work) in which they used the same software (Dragon Naturally Speaking) to check the word accuracy of this software. They created user profiles for each speaker using sampling frequency of 11025Hz and 16 bit resolution. They used two separate references for the speakers. Five speakers read one reference text while the other five read different reference text. The text of all these speakers was then compared and this comparison was based on a scoring program called Sclite from the US National Institute of Standards and Technology (NIST) to give the percentage word accuracy results. They were using two scenarios during their trials on ASR. The first was non-overlapping scenario, in which the speakers were reading text independently irrespective of interference and second scenario was overlapping scenario where interference is being introduced to the speakers. The average word accuracy for ten speakers was 70.6 percent that reduced to 29.9 once interference was being introduced. In order to improve the word accuracy of the overlapping scenarios they used a speech separation algorithm. Once they implemented this algorithm in their recorded data files relevant to overlapping scenarios the word accuracy improved from 29.9 to 62.9 percent. As a whole their speech separation experiment is based on three scenarios. In our test of the ASR we had four scenarios which were non-overlapping, overlapping, using DSTO algorithm to improve word accuracy and using our algorithm (ITR) the enhanced version of the existing algorithm. In our test, 7 out of ten speakers were having Australian accent and three were having Indian accent. The reference test was in English, and it consisted of 277 words and for interference we were using a different text.

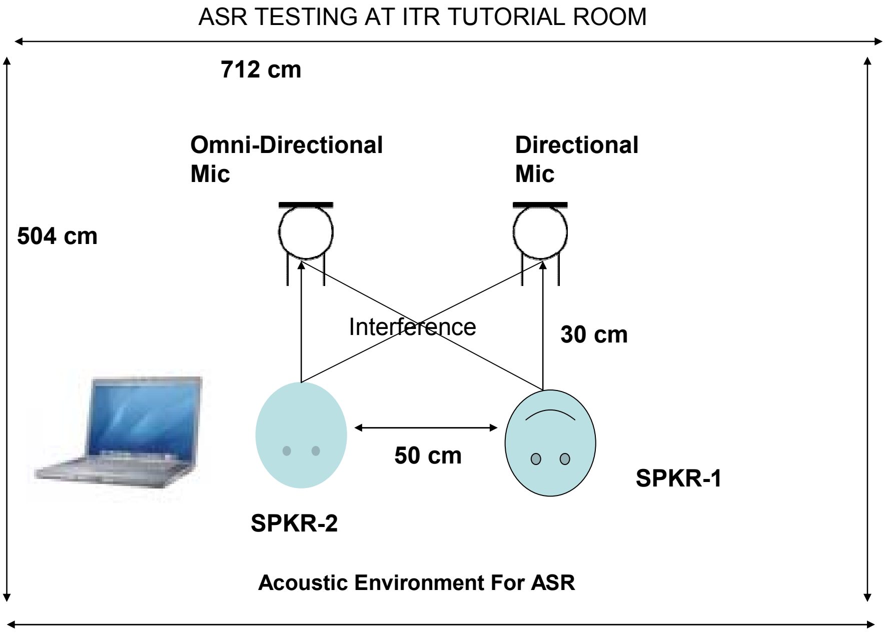


Figure 16 ASR Acoustic Environment

Figure 16 is depicting automatic speech recognition using Dragon Professional Naturally Speaking Software installed in the computer which is also running audio simulations using audio editing software Cakewalk, for speech recording scenarios for the participating speakers.

4.4 Transcription of Audio Data

Once the recording of all four scenarios for each speaker was completed. We started transcription of this audio data which was saved as wav files after being recorded in Cake-walk that we used for recording. The sampling rate and resolution bits of all these wav files were adjusted (using gold wave which is another audio editing tool) according to what was being used by the dragon software for transcription. It took short time in transcribing non-overlapping scenario, but longer time in overlapping scenario which was obvious due to interference from the other speaker. The reference text and the interference text that we used and transcriptions of all the speakers are described below for the two scenarios which are Non-overlapping and Overlapping. The details for all the speakers are attached in Appendix-B.

4.5 Reference Text

[10] Reference text can be found in Appendix- B.

4.5.1 Description

The reference text was of 277 words and three commands were also used by the speakers which the dragon software transcribe as commands while transcribing the audio data.

Three commands were full stop, new line and comma. All these were highlighted for speaker's convenience. Total number of lines was 17. Reference text was quite easy to read but speakers were told about these commands and also to go through them before going for all the scenarios.

4.6 Interference Text

[11] Reference text can be found in Appendix- B.

4.6.1 Description

The interference text was to be read by the interfering speaker in case of overlapping scenarios. Here again the user was spontaneously speaking but also using the three commands that were used in reference text as well and are being highlighted.

Then these all recorded audio files were being transcribed one by one and then two scenarios (Non-overlapping and Overlapping) were compared with the reference text to find word accuracy for each speaker along with the errors being made by both speakers and the software. Then next step would be to increase the word accuracy of non-overlapping scenarios using algorithms used by DSTO and ITR and then comparing word accuracy from algorithms used by both the organizations. Following are the texts resulted from transcription of both non-overlapping and overlapping scenarios for each speaker.

4.7 Observations

After carefully going through the transcriptions for both the scenarios in case of the participating speakers, I have some observations that can be useful in improving the word accuracy of the dragon naturally speaking software and they are following. (Note that they have been taken from all the speakers).

4.7.1 Non-Overlapping Scenario

It happens that mistake made by the speaker is being indicated by the software while transcribing, for example in one case the speaker missed out a word from the reference text while reading it, as a result this was indicated by software. In this case, the actual line was "Some times the author will put key ideas in the margin" so one of the speaker missed the word key. As a result the software transcribed this sentence as "put clear ideas". So the remedy can be while speaking during testing scenarios, the speaker makes it sure that he will not miss any of the word; else there will be an error.

In case of using commands like "full stop new line" it happens some time that the software consider it part of the sentence and transcribed the reason can be that the speaker casually spoke the commands as they looked part of the sentence. In one of the sentence which was "In most cases you know what you are looking for, so you are concentrating on finding a particular answer" so once the speaker started reading sentence and once he used the command words at the end of this sentence which were "full stop new line", these two commands were not correctly understood by the software and the resultant sentence was transcribed as "In most cases you know what you are looking for, so you are concentrating on finding a particular answer will stop new line" so it is obvious that these commands should be spoken in such a manner that ASR system can understand them and will not write them as part of the spoken sentence.

One common mistake that I believe is difficult to be corrected, was transcribing a word "Reading off" to "Reading of" which is a bug in this software, since it sounds the same for both these cases.

Grammatical mistakes are also the ones that can be corrected by training the software about their use. The observation that I found about this is taken from the original sentence "Reading off a computer screen has become a growing concern". Once this sentence was read, its transcribed version came up with this, "Reading off the computer". So it has a flaw of mixing small words like 'a' or 'the' with each other or in some other situation.

The ASR also does some insertions while transcribing the audio data, like in one of the sentence which was originally spoken as "In most cases, you know what you are looking for, so you are concentrating on finding a particular answer". Once this sentence was read by the speaker, the ASR inserted a new word from its vocabulary database and the sentence appeared as "In most cases, and you know". So this can be one of the bugs in the software.

Another important parameter of errors made by dragon software was deletion of words. It can be found in one of the original sentences here, which was spoken as "Once you have scanned the document you might go back and skim it". So dragon transcribed it as "Once you have scanned document you might go back and skim it". So here the word 'the' is being deleted by the software. This was observed in number of scenarios.

Substitutions of actual words with own hypothesized word was another finding. There were lots of sentences where actual word was being substituted by a new word. In one of the sentence that was spoken originally as "Scanning involves moving your eyes quickly down the page seeking specific words and phrases". Once transcribed it appeared as "Scanning involves moving arise quickly". Here the actual words "your eyes" were being substituted as 'arise". This was another bug, I discovered. In order to get rid of these bugs like substitutions, the best way is to train this software with new words so that probability of substituting a word that was previously strange to this software can be minimized.

4.7.2 Overlapping Scenario

In this scenario, the probability of word accuracy was reduced and probability of errors was increased due to interference from other speaker started speaking with the reference speaker at the same time. As a result the software found it hard to transcribe the audio data that was being corrupted by interference. It took longer time to transcribe the audio data but also, when it got stuck between two different human voices, it started making errors like deletion, insertion, substitution and generation of its own words that were inserted in place of the actual words. Some of the new observations in overlapping scenario were noticed. Most of them were observed at the start of the sentence.

In one of the overlapping scenario, the original sentence spoken was "First, they are an aid in locating new terms which are introduced in the chapter". And the transcribed version of this transmitted sentence appeared as "First, they and 18 locating new terms which are introducing the chat of". My observation was that "are an aid in" were being replaced by software with its own version of "and 18" that sounds similar to them. Since there was interference, it was interpreted like that.

Once the ASR software is unable to find the exact word from its analysis, it starts selecting from its own database on and come up with a new sentence instead of the actual sentence. In one of the overlapping scenario, the speaker read the following sentence "Look for words that are bold faced, italics or in a different font size, style or colour". Since there was interference from other speaker who was speaking using an omni-directional microphone at 360 degree angle then the reference speaker who was using a microphone at 115 degree and that too a uni-directional microphone so the uni-directional microphone was getting interference of the speaker (interfering) at an angular direction between 0 and 115 degree. The ASR software replaced this sentence by its own generated sentence which has some words different then the originally transmitted reference sentence and it appeared in the transcribed file as "The looks of words that are bold faced, by telex or rooms are different on size, style or co lour". These kinds of errors have more probability of more occurrences in overlapping scenario.

The above scenarios are mentioned in the form of a summary in Table G1 at the end of the appendices. It describes both the overlapping and non-overlapping scenarios in terms of errors and word accuracy for each speaker concluded by mean of the errors and word accuracy to before using speech separation algorithm to extract information in overlapping scenarios.

CHAPTER 5

5 Implementation of Speech Separation Algorithm

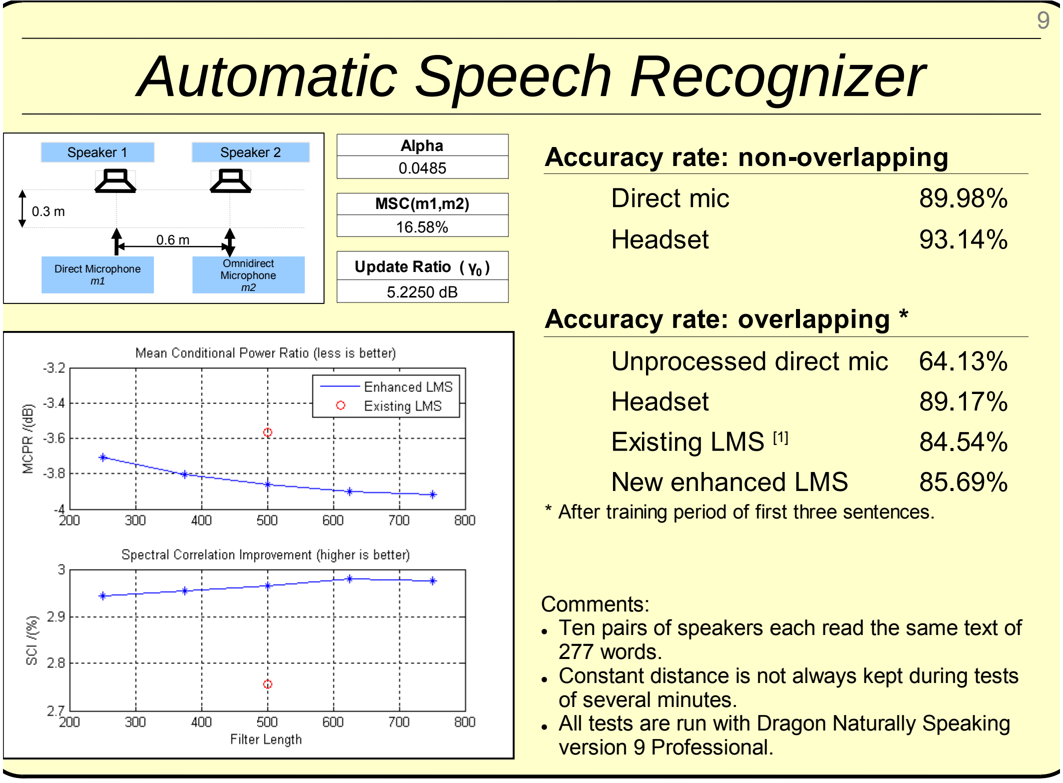
5.1 After Processing Scenario (ASR)

After careful analysis of these two scenarios, they were being passed through speech separation algorithms used by DSTO and ITR. Then, we observed improvement in our algorithm which was also the main aim of speaker separation. We were using a speech adaptive algorithm known as LMS (Least Mean Square) algorithm which was also being used by DSTO. Our algorithm showed an improvement in word accuracy after speech separation from interference speech signal. Since the same overlapping scenarios were being used that are mentioned before processing scenario. I have considered one of the non-overlapping scenario for analysis in terms of after processing using our, which is ITR algorithm and DSTO algorithm followed by the improvement table made by my fellow colleague Mr. Holfeld to show where our (ITR) algorithm has improved the existing (DSTO) algorithm. This example is of speaker 7 whose details are mentioned above.

The summary in Table G2 in this scenario after processing can be accessed at the end of the appendices. This table shows how our algorithm improves in terms of sentence. Especially, after three sentences it dominates the existing algorithm (DSTO) which dominates in case of first three sentences.

5.2 Summary Visualization

The visualization of automatic speech recognizer is shown below with description:



9

Automatic Speech Recognizer

Alpha
0.0485

MSC(m1,m2)
16.58%

Update Ratio (γ_0)
5.2250 dB

Accuracy rate: non-overlapping

Direct mic	89.98%
Headset	93.14%

Accuracy rate: overlapping *

Unprocessed direct mic	64.13%
Headset	89.17%
Existing LMS [1]	84.54%
New enhanced LMS	85.69%

* After training period of first three sentences.

Comments:

- Ten pairs of speakers each read the same text of 277 words.
- Constant distance is not always kept during tests of several minutes.
- All tests are run with Dragon Naturally Speaking version 9 Professional.

Figure 17 ASR Word Accuracy

[1]Description

Figure 17 gives an overall view of the automatic speech recognizer results that are being obtained in non-overlapping scenarios and overlapping scenarios. One of the observations is that of the speakers that were using head-set while performing ASR testing. In this case the accuracy rate shows almost same result for both non-overlapping and overlapping scenarios. Since our task was to enhance the existing algorithm and that speaker does not need to put on head-set while speaking, we have improved the accuracy rate from existing algorithm by an improvement factor of 1.15 percent. Although this is not too much but of course its what our

[1] This description is basically a common work that I performed with my fellow colleague Mr. Holfeld and we presented this research work in a seminar in Australia. This was basically a poster presentation, which was sponsored by some of the leading research organizations in South Australia.

objective was, and further improvement can be done using beam-forming technique or blind source separation.

Another finding as for as above diagram is concerned, is of representation of both the algorithms (the existing and enhanced) in terms of time domain and frequency domain. In frequency domain we have defined the relationship between frequency of speaker 2 before and after the implementation of speech separation algorithm. It suggests that a large filter length speech separation algorithm performs better than a small filter length algorithm. However, its convergence is slower than the small filter length algorithm. In time domain analysis of this algorithm its about measuring the mean conditional power of speaker 2.We have defined five different length filters for our speech separation algorithm that makes it more flexible than existing algorithm which is using a single filter length of 500 in its implementation. In the figure some other important parameters are ‘Alpha’, which is an update counter that does counting for number of updates or samples in terms of mean conditional power of the interfering speaker once its voice activity has been detected.

6 Conclusion and Future Work

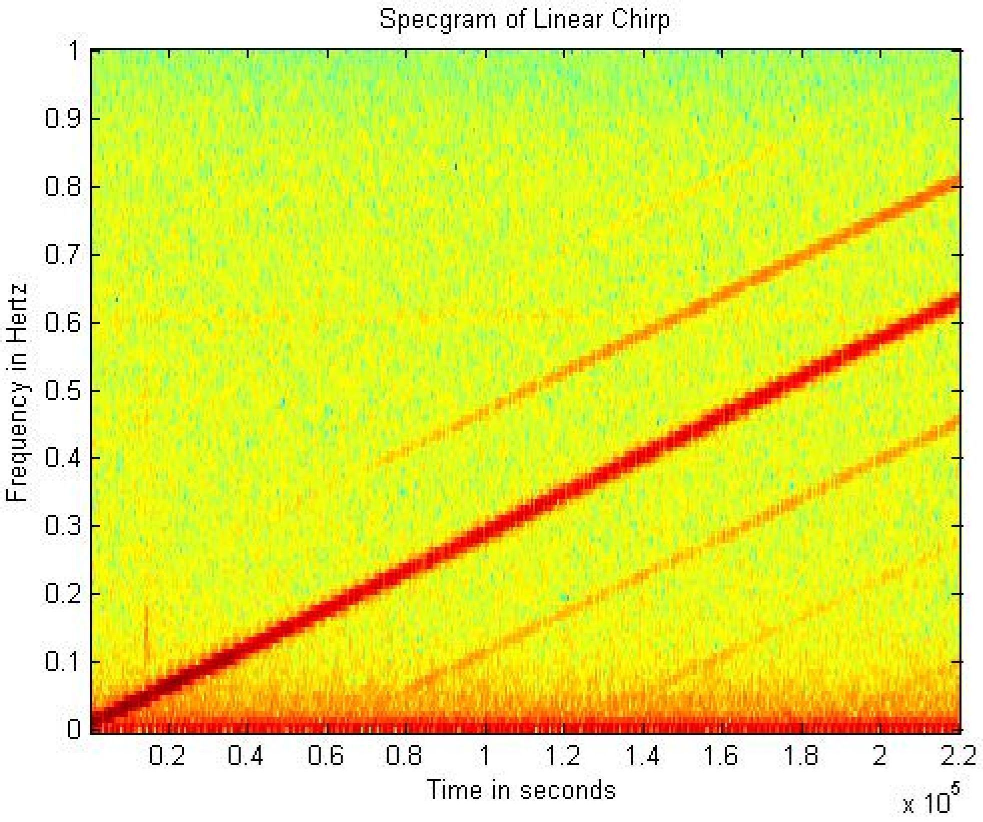
The reverberations are important in acoustic environments where they can be a source of interference for scenarios like speaker separation that is basically used to separate original speech from other speech. Future work in this regard can be analysis of reverberations being introduced as interference to a speaker and then we can estimate the impact of impulse response on speech in terms of various parameters like power estimate between the speech and reverberations. They can play significant role in this kind of scenario where environment is suitable for generating lot of acoustic multiple reflections of the speech that can lead to further investigate this area of signal processing. The automatic speech recognition system plays important role in speaker separation recognition, but still there are areas of work that can effectively enhance its efficiency in conjunction with word estimation and transcription. In fact it cannot transcribe a perfect replica of the speaker’s voice since voice varies from person to person. But still investigation in this area of research work is possible, by testing it in different acoustic environments and with different linguistic features that are relevant to speech processing.

From [14] “Chirps are better to use then other excitation signals like maximal length sequences in acoustic noise”, although there is still a debate going on in this field of signal processing. People have come up with different ideas based on their analysis of these excitation signals. After going through my investigation in comparing these excitation signals as for as room impulse response is concerned, I am convinced that chirps perform better than maximal length sequences, especially in acoustic environment where existence of non-linearities and time variances is common. Here maximal length sequences will not be able to give a perfect impulse response that can be used for the scenarios like speaker separation or acoustic environment inside a submarine to improve sonar communication. This can also lead to another investigation as for as the suitability of excitation signals is concerned and it is possible by going through a comparative analysis of these two methods in reverberation chambers that are specifically build for this area of acoustics.

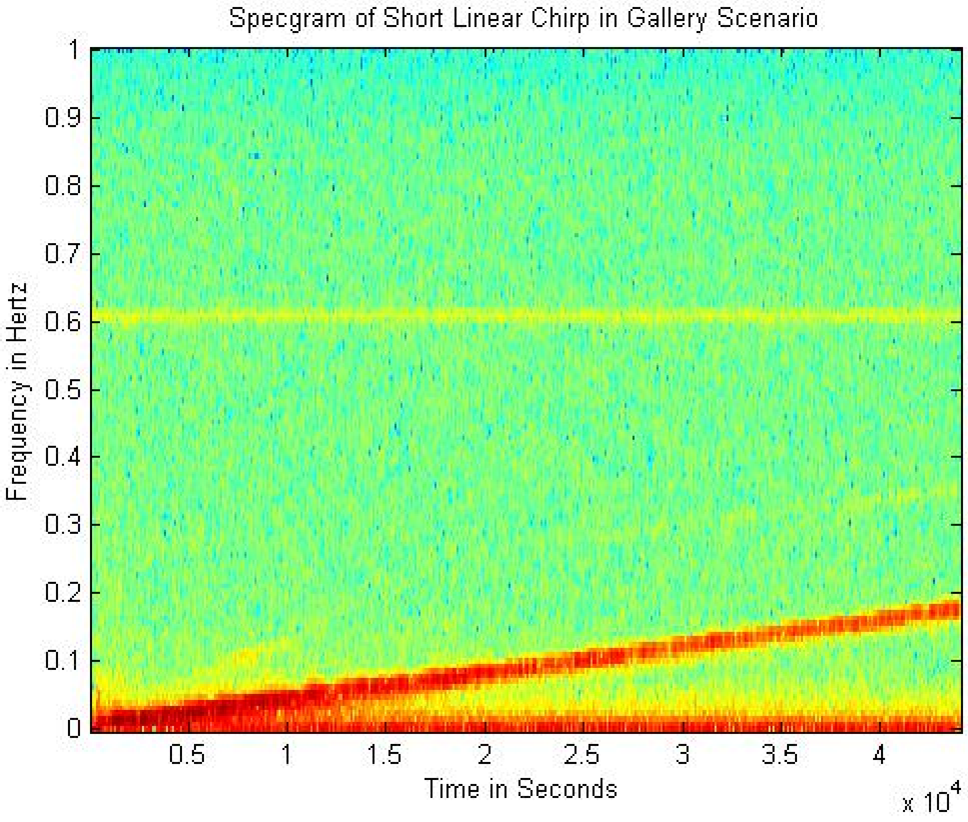
7 Appendices

Appendix A: Figures

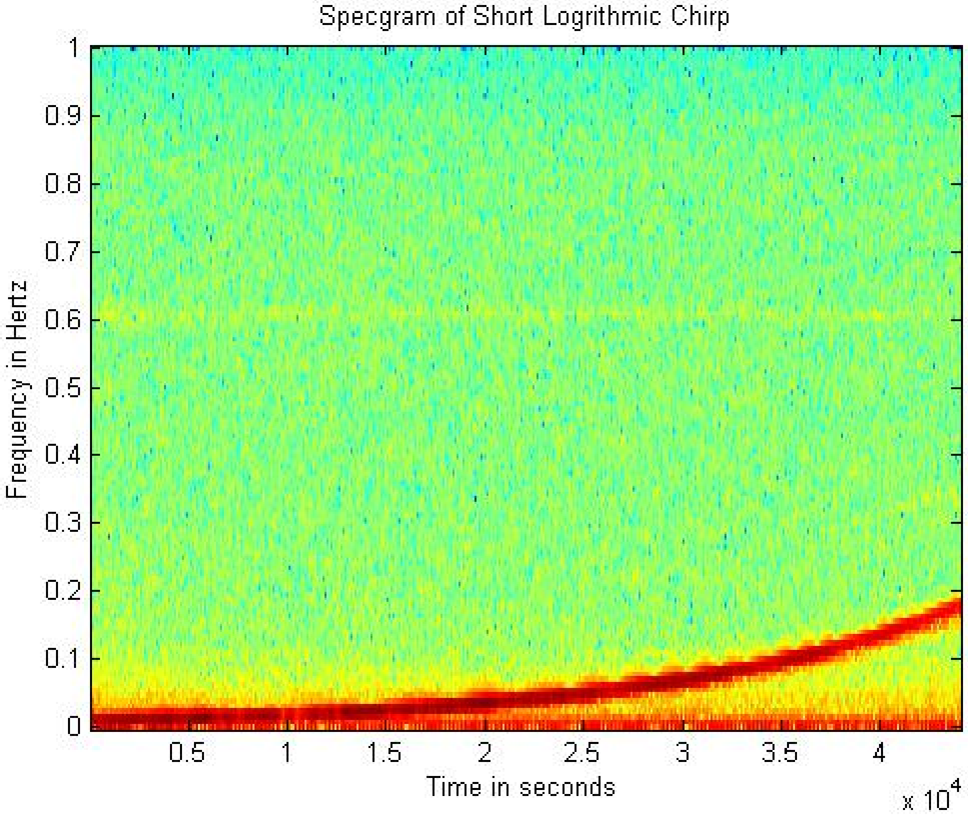
A-1: Received Linear Sweep Spegram obtained in Room-7



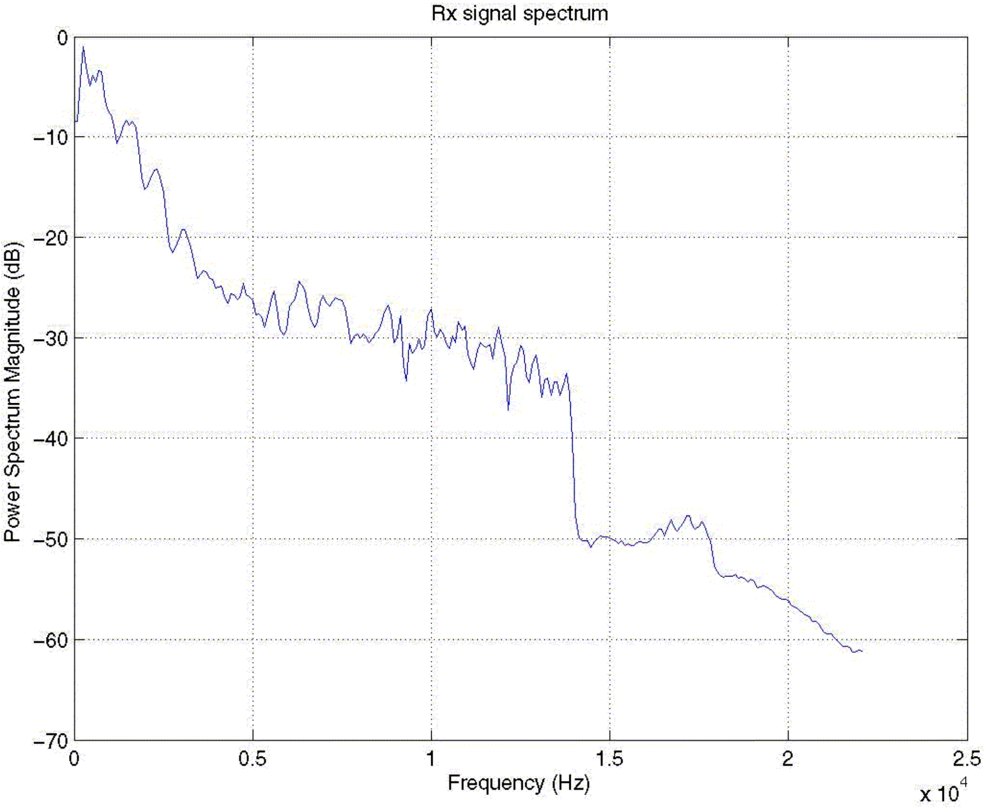
A-2: Received Short Linear Sweep Specgram obtained in Gallery



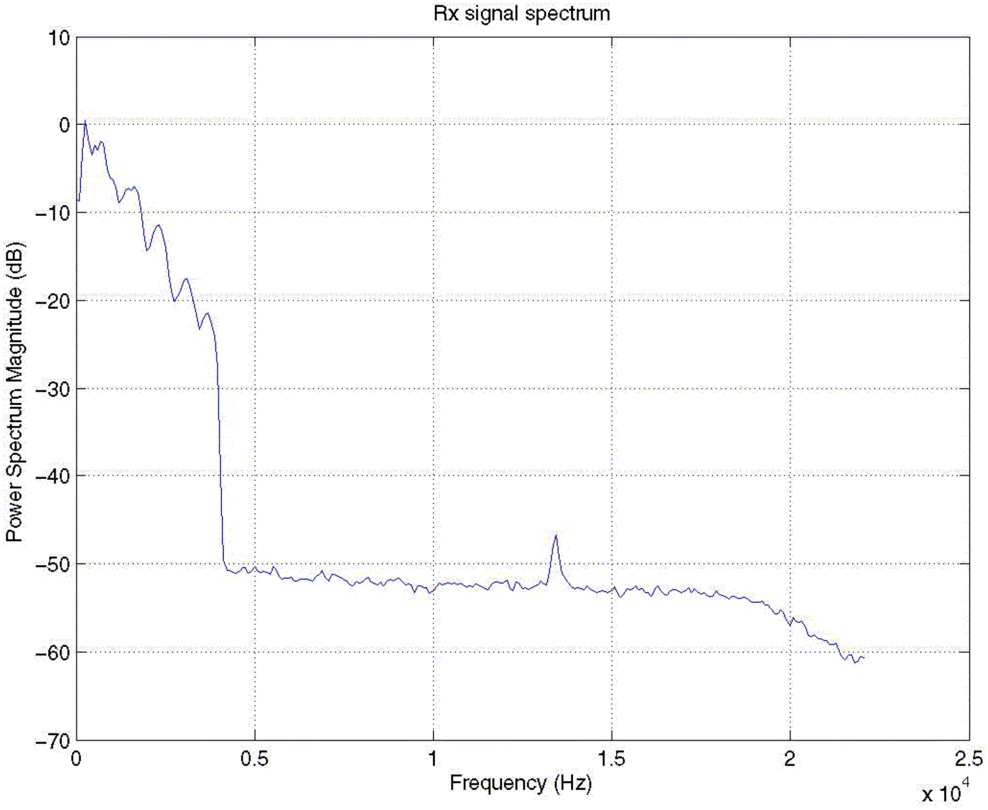
A-3: Received Specgram of Short Logarithmic Chirp (Room-7)



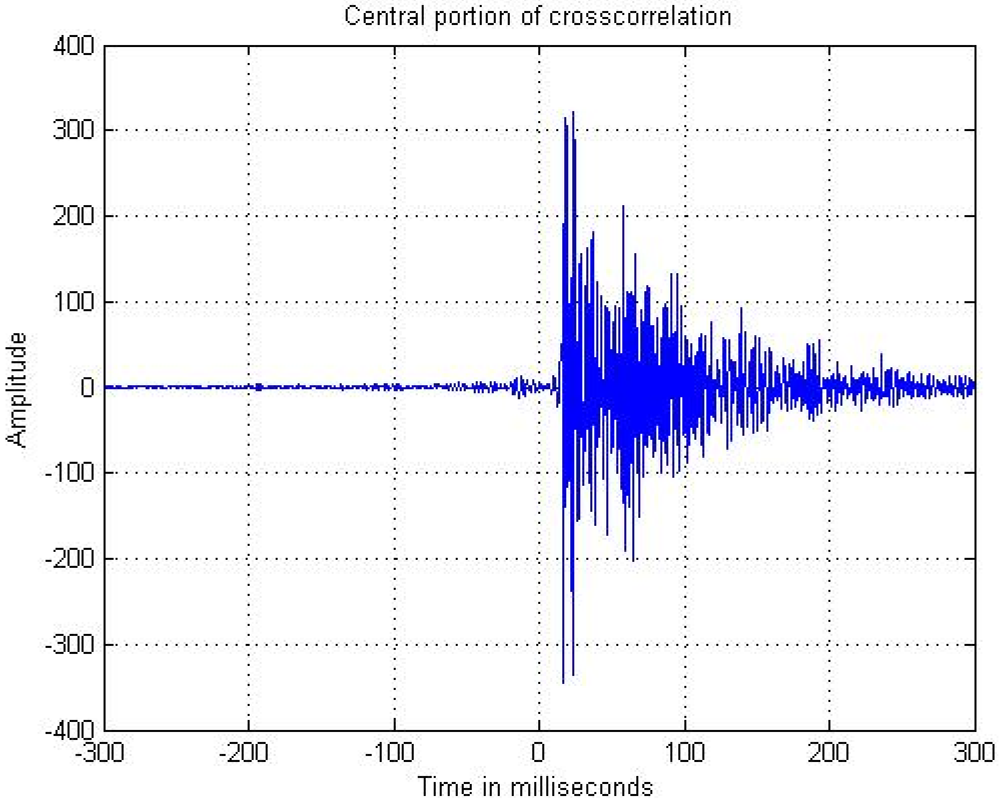
A-4: Power Spectrum Magnitude of Log Chirp (Room-7)



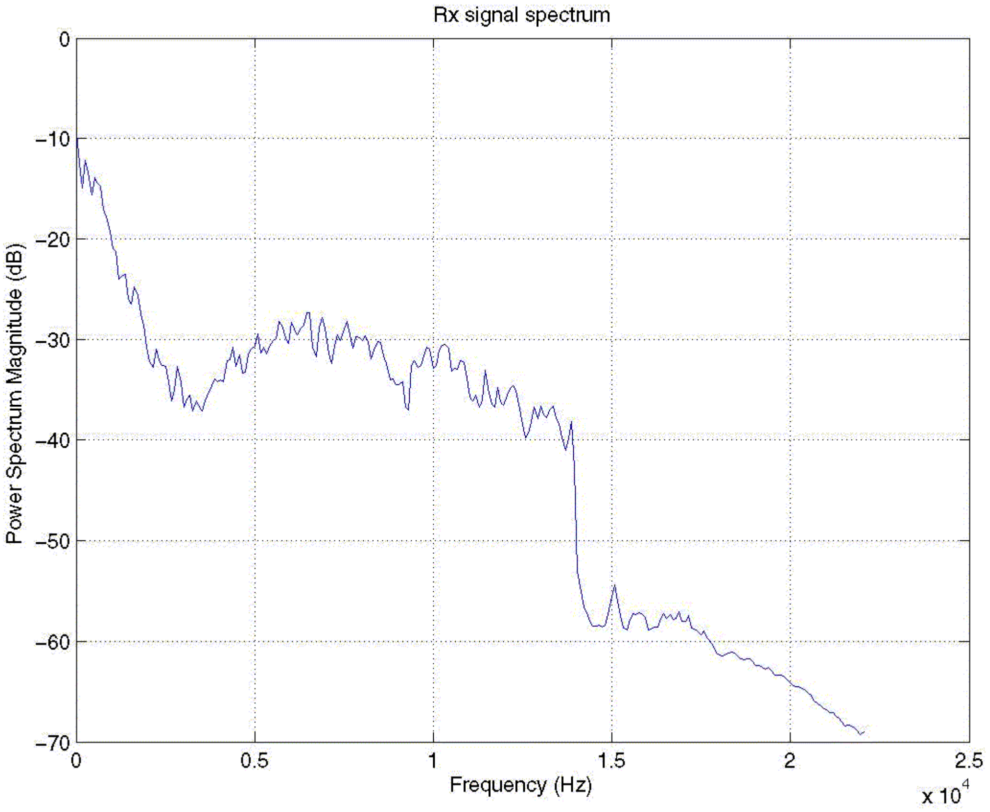
A-5: Power Spectrum Magnitude of Received Log Short (Room-7)



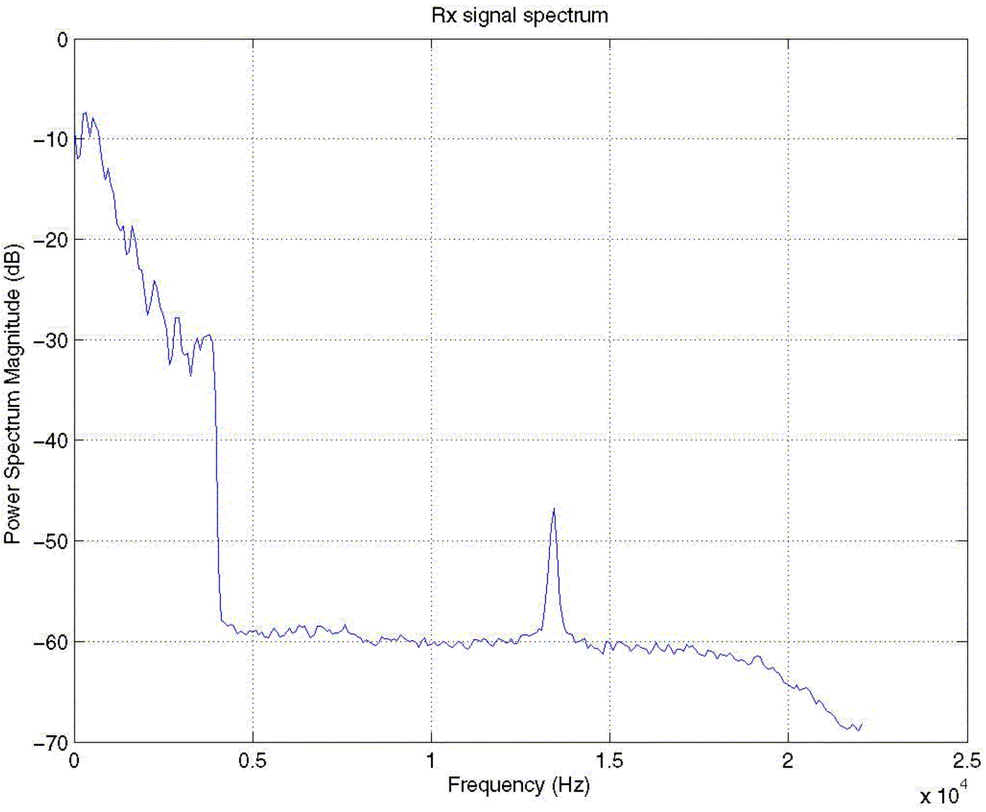
A-6: Cross-correlation of Short Linear Sweep (Gallery)



A-7 Power Spectrum Magnitude of Linear Sweep (Gallery)



A-8: Power Spectrum Magnitude of a Short Linear Sweep (Gallery Scenario)



Appendix B: Matlab Commands Relevant to IR

The simulating tool for generating impulse response of a room is Matlab. Some of the commands relevant to impulse response are given in the sequence.

B.1 Wavwrite

This command basically writes data for generating excitation signals like chirps, maximal length sequences. Data can be of different bits according to users own requirement and similarly we can select different sampling frequencies for audio data. Default value for number of bits is 16 and sampling frequency is 8000 Hz. Data can be mono or stereo but in case of stereo it has to be specified as a matrix with two columns. Amplitude value of the audio data should not exceed more then the value [-1, +1], otherwise audio data saved will be clipped off. This condition of audio data applies to the type of data that has number of bits less than 32, standard number of bits represented as N is 8, 16, 24, and 32.The data is saved in a vector that basically defines the type of audio data to be generated. The audio data that uses number of bits that are, 8, 16 and 24 is called type 1 integer pulse code modulation while 32 bit audio data is known as type 3 normalized floating points. In order to normalize the audio data to get rid of clipping, the audio data created can be multiplied by values such as 0.5, 0.75,1.

B.2 Wavread

If we want to import audio data from any location in computer, then the command wavread help in reading audio data. This command basically supports multi-channel data, up to 32 bits per sample. It also supports reading 24 and 32 bit.wav files. Amplitude of the audio data read is the same as that of being supported by wavwrite, i.e. [-1, +1]. It can also be used in a variety of ways to return samples for the audio data. For example it can return only the first N samples from each channel in the file and also it can return the number of samples limited to a particular range from each channel in multi channel environment. For example, a transmitted audio data with .wav extension can be recorded from some audio recording software and then the recorded or received audio data can be imported to Matlab using wavread command and then can be used for various signals processing according to the requirement.

B.3 Auto-correlation

Auto-correlation is a time-domain function that is a measure of how much a transmitted signal resembles a delayed version of itself. The value of auto-correlation can vary between zero and one. A periodic signal has an auto-correlation which is equal to one at zero time delay. Auto-correlation is sometimes used to extract periodic signals from noise. It is also described as a signal with its own time shifted version. Matlab uses a function “xcorr” to estimate the auto-correlation of a periodic signal with itself. For example, a logarithmic chirp, which is an excitation signal, can be created and then its auto-correlation can be computed by using function xcorr to see its auto-correlation.

So auto-correlation a logarithmic chirp can be calculated and plotted in Matlab after creating this excitation signal in a vector and then using the function ‘xcorr’ to compute and then plot it by using command ‘plot’ or ‘stem’ to see the auto-correlation in time domain. The Matlab command for finding auto-correlation of a deterministic signal, say ‘X’ is given as, xcorr(X,X).

B.4 Cross-correlation

Matlab uses a function “xcorr” to estimate cross-correlation sequence between two signals x and y, as already mentioned in theoretical background of this final report.

The function xcorr computes the cross-correlation between the two signals x and y and then returns this cross-correlation sequence. In case of cross-correlation of a chirp following is the simulation representation in terms of plotting of cross-correlation:

The sampling frequency in case of cross-correlation in above case is 44100 Hz and number of bits are 16. I used a special source code to generate this cross-correlation.

First a linear chirp has been created from ‘chirp’ function that generates a longer linear chirp between a certain frequency span. After generating this chirp, I saved it as audio data in wav format by using wavwrite command and I also normalized it. Then I transmitted this chirp and recorded the received version of this transmitted wav file. The transmitted wav file is named as (tx_sweep1.wav) and received wav file is named as (rx_sweep1_mono_left) and I have used audio editing software Cool Edit that plays and records at the same time. I have used a USB Audio interface along with, two different microphones a uni-directional microphone connected as left channel and omni-directional microphone as right channel. Although I saved the received signal in stereo form but then I separated both these channels from Matlab and then cross-correlated the received signal from left channel with original transmitted signal to see the cross-correlation between two signals (tx and ty) which is written as xcorr(tx,ty) where tx is transmitted audio wav file and rx is received audio wav file.

B.5 Specgram

It is analysis of a signal whose frequency is time dependent. It is especially useful in terms of looking at various kinds of chirps that can be used as excitation signals for generation of impulses. Usually the chirps after being transmitted and recorded via some kind of audio editing tool are being analyzed using specgram to see presence of other non-linear components within a chirp, in the form of harmonics, background noise, inter-modulation products and other such kinds of elements in the received signal which can then be extracted in the presence of such elements. It basically calculates the windowed discrete time Fourier transform for a signal in a vector say, ‘y’, using a sliding window. Some important parameters that run this function are ‘nfft’ which specifies the fft length that is being utilized by this function. Nfft get the frequencies at which the discrete-time Fourier transform is computed. Default sampling frequency that it uses is 2 Hz and the window used for this function is Hanning which is equal to the length of nfft. Other important parameter is numoverlap that basically divides the window into half in length. It is also the number of samples by which the section overlaps. The length of the window should be less then or equal to the number of nfft.

B.6 Power spectral density

Power spectral density (PSD) function shows the strength of the variations energy as a function of frequency. In other words, it shows at which frequencies variations are strong and at which frequencies variations are weak. The unit of PSD is energy per frequency and we can obtain energy within a specific frequency range by integrating PSD within that frequency range. Computation of PSD can also be done by computing auto-correlation function and then transforming it. To compute power spectral density of a sequence an object ‘H’ is being initiated by a function “dspdata.psd” in Matlab and data property of this object is set to ‘data’. The frequency spans over the interval [0, π] and sampling frequency is normalized by default. So there is no need of defining another function for normalizing sampling frequency. Data can be a vector or a matrix depending upon ingenuity of the user. The power spectral density is intended for continuous spectra. In this function, the peaks in the spectrum of the received signal does not show reflected power at the frequency that is being used for this spectrum. Rather the integral of the PSD over a given frequency band computes the average power in the signal over such frequency band.

B.7 Convolution

According to convolution, convolving two signals ‘tx’ and ‘ty’ is equal to multiplying their Fourier transform. To get a perfect convolution, the right way is to pad the two signals mentioned above with zeros and ignore round off error. This expression is given below:

```
 T = fft([tx zeros(1,length(ty)-1)])
U= fft([ty zeros(1,length(tx)-1)])
```

Therefore the convolution between these two sequences ‘tx’ and ‘ty’ is given below

```
conv (tx,ty) = ifft (X.*Y)
```

Appendix C: Speaker Separation Testing Scenarios (ASR)

C.1 Reference Text

[10] Following is the reference text used for the recordings.

Please, read the following symbols:

. full stop
. new line
, comma

Reference Text

Scanning is a technique you often use when looking up a word in the telephone book or dictionary..You search for key words or ideas..In most cases, you know what you are looking for, so you are concentrating on finding a particular answer..Scanning involves moving your eyes quickly down the page seeking specific words and phrases..Scanning is also used when you first find a resource to determine whether it will answer your questions..Once you have scanned the document, you might go back and skim it..When scanning, look for the author's use of organizers such as numbers, letters, steps or the words first, second, or next..Look for words that are bold faced, italics or in a different font size, style or color..Sometimes the author will put key ideas in the margin..Reading off a computer screen has become a growing concern..Research shows that people have more difficulty reading off a computer screen than off paper..Although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper..Similarly, scanning skills are valuable for several purposes in studying science..First, they are an aid in locating new terms which are introduced in the chapter..Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary or glossary..Thus a preliminary scanning of the chapters will alert you to the new terms and concepts and their sequence..When you locate a new term, try to find its definition..If you are not able to figure out the meaning, then look it up in the glossary or dictionary..

C.2 Interference Text

[11] Following was the interference text read by the interferer in case of over-lapping scenario.

Please, read the following symbols:

🟩 full stop
🟥 new line
🟦 comma

Interference Text

Once upon a time there was little girl 🟦 pretty and dainty🟩 🟥

But in summer time she was obliged to go barefooted because she was poor 🟦 and in winter she had to wear large wooden shoes 🟦 so that her little instep grew quite red🟩 🟥In the middle of the village lived an old shoemaker’s wife 🟦 she sat down and made 🟦 as well as she could 🟦 a pair of little shoes out of some old pieces of red cloth.🟩 🟥They were clumsy 🟦 but she meant well 🟦 for they were intended for the little girl 🟦 whose name was Karen. 🟩 🟥Karen received the shoes and wore them for the first time on the day of her mother’s funeral. 🟩 🟥They were certainly not suitable for mourning 🟦, but she had no others 🟦 and so she put her bare feet into them and walked behind the humble coffin🟩 🟥. Just then a large old carriage came by 🟦 and in it sat an old lady 🟦 she looked at the little girl 🟦 and taking pity on her 🟦 said to the clergyman: “Look here 🟦 if you will give me the little girl 🟦 I will take care of her.🟩 🟥 Karen believed that this was all on account of the red shoes 🟦 but the old lady thought them hideous 🟦 and so they were burnt.🟩 🟥 Karen herself was dressed very neatly and cleanly 🟩 🟥She was taught to read and to sew 🟦 and people said that she was pretty.🟩 🟥 But the mirror told her 🟦 “You are more than pretty 🟦 you are beautiful.🟩 🟥 One day the Queen was traveling through that part of the country. 🟥 Now the old lady fell ill 🟦 and it was said that she would not rise from her bed again.🟩 🟥🟩 She had to be nursed and waited upon 🟦 and this was no one's duty more than Karen's.🟩 🟥 But there was a grand ball in the town 🟦 and Karen was invited🟩 🟥She looked at the red shoes🟦 saying to her that there was no sin in doing that she put the red shoes on 🟦 thinking there was no harm in that either and then she went to the ballad commenced to dance🟩 🟥 But when she wanted to go to the right 🟦 the shoes danced to the left 🟦 and when she wanted to dance up the room 🟦 the shoes danced down the room 🟦 down the stairs through the street 🟦 and out through the gates of the town.🟩 🟥She danced 🟦 and was obliged to dance 🟦 far out into the dark wood.🟩 🟥Suddenly something shone up among the trees 🟦 and she believed it was the moon 🟦 for it was a face.🟩 🟥 But it was the old soldier with the red beard 🟦 he sat there nodding his head and said: "Dear me 🟦 what pretty dancing shoes.🟩 🟥She was frightened 🟦 and wanted to throw the red shoes away; but they stuck fast.🟩 🟥 She tore off her stockings 🟦 but the shoes had grown fast to her feet.🟩 🟥She danced and was obliged to go on dancing over field and meadow 🟦 in rain and sunshine 🟦 by night and by day.🟩 🟥But by night it was most horrible.🟩 🟥

Note

The overlapping and non-overlapping scenarios for all the speakers are mentioned below in detail, where they read the above texts in these scenarios.

Non-Overlapping Scenario Speaker (2)

Scanning is a technique you offer news when looking at the word in the telephone book or dictionary.

You search for keywords or ideas.

In most cases, you know what you are looking for, so you are concentrating on finding a particular answer will stop new line

scanning involves moving arise quickly down the page seeking specific words and phrases.

Scanning is also used when you first bonded resource to determine whether it will answer your questions.

Once you have scanned the document, you might go back to skim it.

When scanning, you look that the authors use of organizers such as numbers, letters, steps or the words that first, second, or next.

Look the words of the old face, by telex or in a different console is, style or color.

Sometimes the author will put clear ideas in the margin will stop

reading of the computer screen has become a growing concern.

Research shows that people have more difficulty reading of a computer screen and paper.

Although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper.

Similarly, scanning skills are valuable to several purposes in studying science.

First, they are and eight in locating new terms which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary or glossary.

Thus a preliminary scanning of the chapters will alert you to the new terms and concepts and their sequence.

When you locate a new term, try to find its definition.

If you are not able to figure out the meaning, then look it up in the glossary or dictionary.

Overlapping Scenario For Speaker (2)

Scanning is a technical to other moves when looking at a word in a telephone book was cautionary.

Some search keyword or ideas.

While in most cases, you know what you are looking for, so you are concentrating on finding a particular answer.

Scanning involves moving arise @ down the page seeking @ words and phrases.

Standing hills are used when you're under resourced to determine whether it will answer your question.

It will likely have standardised on a, you might go back and skim it will shock

when scanning, who look to the authors who organisers such as numbers, letters, steps or the words first, second and, and ornamental.

While you're or words to old faces, I Or in it and on size, style or colour.

By sometimes you all to put @ ideas into the margins.

What reading of the computer screen has become a growing concern.

Research shows that people have more difficulty reading of the computer screen turnoff paper.

All day they can read and comprehend at the same rate as paper, whose skimming on the computer is much slower than on paper.

Similarly, standing still is a valuable percent returns is studying in clubbing signs.

It first, they are in a in locating new term to introduced in the chapter.

Unless you understand the new term, it is impossible to follow the author's reasoning without dictionary or else to reach.

In just a preliminary scanning of the changes will alert you to the new terms and concepts and message wasn't.

When you locate a new turn, or try to find the definitions of.

If you are not able to figure out the meaning, then look at in the glossary of dictionary.

Non-overlapping Scenario Speaker (3)

Scanning is a technique you often use when looking up the word in the telephone book @ dictionary.

You search for keywords @ radius.

In most cases, you know what you're looking for a, so you're in a concentrating on finding a particular dancer.

Scanning involves moving your eyes quickly down the page seeking specific words and phrases.

Scanning is also used when you first find the resource to determine whether it will answer your questions.

Once you have scanned the document, you might go back and skim it.

When scanning, look for the author’s use of it organises such as numbers, letters, steps of the words first, second, or next.

Look for words that are boldfaced, the tactics or in a different font size, style or other.

Sometimes the author will put key ideas in the margin.

Reading of the computer screen has become a growing concern.

Research shows that people have more difficulty reading of a computer screen than a paper.

Although they can read and complicated comprehend at the same rate as paper skinning skinning on the computer is much slower a than on the paper.

Similarly, canning skills are valuable for several purposes in studying signs.

First, they ardently in locating new terms which are introduced into @ chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning with a dictionary of glossary.

Thus the preliminary scanning of the computer of the chapters will let you to the new terms and concepts and the sequence.

Many locating @ new term, try to find its definition.

If you're not able to figure out the meaning, then look it up in the glossary of dictionary.

Non-overlapping Scenario Speaker (3)

Scanning is a technique you often use when looking up the word in the telephone book @ dictionary.

You search for keywords @ radius.

In most cases, you know what you're looking for a, so you're in a concentrating on finding a particular dancer.

Scanning involves moving your eyes quickly down the page seeking specific words and phrases.

Scanning is also used when you first find the resource to determine whether it will answer your questions.

Once you have scanned the document, you might go back and skim it.

When scanning, look for the author’s use of it organises such as numbers, letters, steps of the words first, second, or next.

Look for words that are boldfaced, the tactics or in a different font size, style or other.

Sometimes the author will put key ideas in the margin.

Reading of the computer screen has become a growing concern.

Research shows that people have more difficulty reading of a computer screen than a paper.

Although they can read and complicated comprehend at the same rate as paper skinning skinning on the computer is much slower a than on the paper.

Similarly, canning skills are valuable for several purposes in studying signs.

First, they ardently in locating new terms which are introduced into @ chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning with a dictionary of glossary.

Thus the preliminary scanning of the computer of the chapters will let you to the new terms and concepts and the sequence.

Many locating @ new term, try to find its definition.

If you're not able to figure out the meaning, then look it up in the glossary of dictionary.

Overlapping Speaker (3)

Scanning is the technique involves a new when looking at murder in the telephone book on the jazzy.

You search for Parivar idea is to go visit of

in most cases him, it will all know what you're looking for, it was a year concentrating on finding a particular and said one.

This is union was moving rise quickly set around a date seeking specific to and rise was a village

is scanning was also used when you first find a resource to determine it is that it will and failure were signs,.

Once your scanned the document was, as he might go back and skin @ .

Your friends and, a glossary authors use of organisers such as numbers, the letters, is on the word was, it is the end, or next.

Look for words that are boldfaced command italics are in a different form life, his trial colour it.

Champagnes in order will put lady is in the margin.

Since the reading of a computer screen you has become a growing concern here.

It is your research shows that people are more difficult reading of a computer screen on and off the grass roots.

24 although they can free and concrete and at the same rate as the, drawing on the computer was much slower than on paper and.

Similarly, it was and still is a valuable for several purposes and standing in.

On the first, who they are and we in locating new terms on which I introduced in the chapter.

Unless you understand the new is ready, it is impossible to follow the orchestration in a dictionary of glossary of.

But as the preliminary scanning of the chapters will let you more than you terms and concerts and the sequence. And

when you locate annual terms, to try to find its resolution should.

If you are not able to figure out the meaning, the end of the Dove in the glossary of the style.

Non-overlapping Speaker (4)

Scanning is a technique you often used when looking up the word in the telephone book or dictionary.

You search for @ words or ideas.

In most cases, you know what you're looking for, so you are concentrating on finding a particular answer.

Scanning involves moving your eyes complete out a page seeking specific words and phrases.

Scanning is also used when you first find a resource to determine whether it will answer your questions.

Once you @ scan a document, you might go back and skim it.

When scanning, look for the author's use of organizers such as numbers, letters, steps all the words first, second, or next stop

look for words that are bold face, by telex or @ a different font size, style or colour.

Sometimes the author will keep with key ideas in the margin.

Reading of the computer screen has become a growing concern.

Research shows that people have more difficulty reading of the computer screen than of paper.

Although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper will stop

similarly, skimming skills of valuables are several purposes in a study in studying science.

first, they are an aid in locating new terms which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary or glossary.

Thus a preliminary scanning of the chapters will alert you to the new terms and concepts and this sequence will stop

when you locate a new term, tried to find its definition.

If you are not able to figure out the meaning, then look it up in the glossary or dictionary.

Overlapping Scenario Speaker (4)

To score technique you often is looking at the word in the telephone book or dictionary.

He said were two ideas.

It most cases, you know what you are looking for, so you're concentrating on finding a particular answer.

It will scanning involves moving the eyes quickly down the page seeking specific words and phrases.

Scheme is also use when you first finer @ resort to determine whether it will answer your questions.

Once you've scanned @ document, you might go back and spin it, which.

When scanning, look from the author's use of organisers such as numbers techno this Scotland steps were worse first, second, or next.

Offer the words of bold face, by telex or in a different @ size from, style or colour.

Sometimes the author will put key ideas in @ margin.

It went up a computer screen has become a growing concern.

Research shows that people have more difficulty reading of the computer screen and off paper. It

although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper.

Similarly, scanning skills available to @ several purposes in studying science.

First, they are an aid in locating new terms which are introduced in the chapters.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary or glossary.

Thus a preliminary scanning of the chapters will alert you to the new terms and concepts and their sequence.

When you locate a new term, try to find this definition.

If you are not able to figure out the meaning, then look it up in the glossary or dictionary.

Non-overlapping Scenario Speaker (5)

Scanning is a technique you often use when looking up a word in the telephone book @ dictionary.

Usage for keywords or ideas.

In most cases, and you know what you are looking for, say you are concentrating on finding a particular answer.

Scanning involves moving your eyes quickly down the page seeking specific words and phrases.

Scanning is also used when you first find a resource to determine whether it will answer your questions.

Once you have scant the document, you might go back and skim it.

When scanning, look for lots as use of organisers such as numbers, letters, steps of the words first, second, @ next.

Look for words that are born faced, italics or in a different font size, style or colour.

Sometimes the otter will put key ideas in the margin.

Reading of a computer screen has become a growing concern.

Research shows that people have more difficulty reading of a computer screen than of paper.

Although they can read and comprehend at the same rate as paper, skimming on the computer is much lower than on paper.

Similarly, Steyning skills are available for several purposes in sterling signs.

First, they are an age in locating new terms which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the authors reasoning without dictionary or a glossary.

Thus a preliminary scanning of the chapters were arrested to the new terms and concepts and the sequence.

When you locate a new term, try refine its definition.

If you're not able to figure out the meaning, then look it up in the glossary of dictionary.

Overlapping Scenario Speaker(5)

Staining his original you often use when looking at a word on the telephone rings in or dictionary.

And are you such work you are our ideas of it. And

his room in most cases, with this you know what you're looking for, say you are concentrating on finding a particular answer it.

It is planning involves in which a rise slightly down the page without seeking specimen words and phrases of soccer

Arabs gaining is also use when you first find a resourceful gentleman where it will answer your question & Sons stopper

no, one solid as he and the government, you might not like asking it to stop

Karen's Jennings who, about four hours as hands of organisers a such as numbers of letters, steps

other words was as jetsetting for more, our necks of online, and served on a battery for words to them in their boldfaced, is the italics or in a different for eyes, a lot of style for colour, story an enquiry and sometimes at the altar will, but he ideas in the noise and.

Eagerly breaking of a computer screen has become a wild constant view.

A particular research shows that you will have no difficulty reading on a computer screen villain of papers was.

Our gratitude all they can wheel can comprehend is the same way is a terminal, skiing on the computer is much slower than if Festival.

So similar level, staining deals lousy variable for several seconds as in stirring son.

But the reverse awoke to they are engaged in locating new jams and which are introduced in the chatter one.

I believe that unless you understand the new terms, it is impossible to honour of his reasoning with originally' Bruce Arnold

Eric and thus of a preliminary scanning of the chatter of the lustre of the new @ and concepts in a sequence you more.

When relocated a new term, tried Finance definition given.

And if you're not able to figure out on the mainland, and again and in the brass ring cottage buried. It will

Non-overlapping Scenario Speaker (6)

Scan @ @ technique you often use when looking up @ words in the telephone book or dictionary.

You search @ keywords or ideas.

In most cases, you know what you are looking for, so you are concentrating on finding a particular answer.

Scanning involves moving arise quickly down the page seeking specific words and phrases.

Scanning is also used when you first find @ resource to determine whether it will answer your questions.

Once you have scanned @ document, you might go back and skim it.

When scanning, look to the author's use of organisers such as numbers, letters, steps or the words first, second, or next.

Look for words that are bold faced, by telex or in @ different font size, style or colour.

Sometimes you offer will put on their key ideas in the margin.

Reading of the computer screen has become a growing concern.

Research shows that people have more difficulty reading of @ computer screen and off paper.

Although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper.

Similarly, scanning skills are valuable to several purposes in studying science.

First, they are an aid in locating new terms which were introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary or glossary.

Thus the preliminary scanning of the chapters will alert you to the new terms and concepts and their sequence.

When you locate a new term, try to find its definition.

If you are not able to figure out the meaning, then look it up in the glossary or dictionary.

Overlapping Scenario Speaker (6)

Scanning is a technique you often use when looking up @ words in the telephone book @ dictionary.

You search @ keywords or ideas.

In most cases, you know what you're looking for, zone you're concentrating on finding a particular answer.

Scanning involves moving arise quickly down the page seeking specific words or phrases.

Sir scanning is also used in the first final resource to determine whether an answer your questions.

Once your scanned @ document, or you might go back and skim it.

A few and scanning, what are the author's views of organisers such as numbers, letters, steps or the words first, second, or next.

The looks of words that are bold faced, by telex or rooms are different on size, style or colour.

She sometimes the author will put key ideas on the margin.

At a ring of a computer screen has become a growing concern.

Research shows that people have more difficulty reading of the computer screen and off paper.

Although they can read and comprehend at the same rate as your paper, skimming on the computer is much slower than on paper.

Is that similarly, it scanning skills are valuable to several purposes in studying science.

First, after they and age in locating new terms which were introduced in the chapter.

While unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary glossary.

Rust a preliminary scanning of the chapters on alert you to the new terms and concepts and their statements.

When you locate a new term, to try to find its definition.

But if you are not able to figure out the meaning, then look it up on the glossary or dictionary.

Non-overlapping Scenario Speaker (7)

@ @ @ Technique you often use when looking at the word in the telephone book @ dictionary.

@ Researcher keywords or ideas will stop

in most cases, you know what you're looking for, so you're concentrating on finding a particular answer

scanning involves moving your eyes quickly down the page seeking specific words and phrases stop

scanning is also used when you @ find a resource to determine whether it will answer your questions.

Much of scan a document, you might go back in skim it

when scanning, look for the authors use of organisers such as numbers, letters, steps of the words first, second, or next.

Look for words of bold faced italics or in a different font @, style or colour.

Sometimes you also will put key ideas in the margin stop reading of a computer screen has become a growing concern

research shows that people have more difficulty reading of a computer screen and of paper

will stop

although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper

similarly, scanning skills of valuable @ several purposes in studying science.

First, they are an age locating new terms which are introduced in the chapter will stop

unless you understand the new terms, it's impossible to follow the authors reasoning without dictionary of glossary stop

thus a preliminary scanning of the chapters will alert you to the new terms and concepts and their sequence.

When you located a new term, tried to find its definition.

If you're not able to figure out the meaning, and look it up in the glossary or dictionary

Overlapping Scenario Speaker (7)

Scanning is a technique @ often used when looking up the word on the telephone @ or dictionary stop

Easter keywords Rienzi's.

In most cases, you find what you're looking for a fun show you're concentrating on finding a particular answer.

You are scanning involves many rise quickly down to seeking specific words and phrases. While

scanning is also yours when you first find a resource to determine whether it were answer your question in our Brompton

once you @ scan a document, you might go back is given.

They scanning, and hopefully also suitable as it is, letters, and Flextech words first, second, or next to them to

look for work in a bold fight on's boring additional sizes, of historical crumble.

Sometimes the author will put clear ideas in the marginal launch in July by a

reading of @ computer scoring has become a growing concern.

Research shows that people who have more difficulty reading of the computer screen than of paper.

Although they can read and comprehend at the same rate of paper, skimming on the computer is much slower: @ @ paper soldiers who are all

similarly hundredth scanning shield your skills or value @ several purposes is that designers.

Personal aeroplane to locate a new terms which are introduced as a chapter. Light version

unless you understand the new tax, it is impossible @ tribunal's reasoning without dictionary of answering.

What as a preliminary scanning of the tractors and will alert you to the new terms and concepts in their statements.

When you locate a new term, tried @ @ its definition.

And if you're not unable to figure out a meaning, then @ @ @ in the glossary or dictionary.

Non-overlapping Scenario Speaker (8)

Scandal @ @ technique you often use when looking at a burden at her from book addiction now and.

He said @ keywords or ideas.

In most cases, the know what you're looking for, so you're concentrating on finding a particular answer.

Scanning involves moving out as quickly done the page seeking specific words and phrases.

Scanning is also be used when you fill had to resort to determine whether it will answer your questions.

Once you've scanned @ document, you might go back and skim it.

When scanning, look for the office use of organisers such as numbers, letters, steps of the words first, second @, next.

Look for words that are boldfaced, italics are in a different font size, style or colour.

Sometimes the offer will prove he had used in the margin.

Reading of a computer screen has become a growing concern.

Research shows that people have no difficulty reading of a computer screen than of the will stop

although they can read in compared with the same rate as paper, skimming on the computer is much lower than on paper.

Similarly, scanning skills are barely above a certain person selling signs.

First, they are in Asia locating new terms which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary of rosemary.

Thus a preliminary scanning of the chapters will allegedly she turns and concerts in the sequence.

When indicated later, try to find its definition.

If we are not able to figure out the meaning, then look it up in a glossary of kitchen.

Overlapping Scenario Speaker (8)

Scandal technique your will spend another word that Sarajevo and it is my will.

Is kilos over eight years ago.

As in most cases, and in all should allocate all, and so you're concentrating on finding a particular answer.

Scandal was rooting out as quickly without a trace of Pacific versus Frazer made.

As well as the tanning is also used to new first annual cost of terminal at largely crossed.

Once you cannot Humana, writer and actor and skidded off.

Warrington and scanning, look for the horses. Organisers such as the numbers, letters, Jackson will all have wards at first, second and third, or next.

There was certainly look from words and at the old faith, it shall extend our own in @ different fathers have, Tiree, in July. By a

sometimes the author with particular ideas in their hearts and the cost of Caroline game by reading of @ computer screen and it has grown concerned in.

The research shows that people have no difficulty reading of a computer screen as men of paper it.

In any of her with a particular region, and @ at the same rate as a, skill and is on the computers was lower than a miraculous.

Though a similarly, asking skills available for severance of the present studying science and arts.

First, the candidate in locating new transmitter and fringes of the Challenger and.

The allies you do understand the new terms, it is impossible that only offers reasonable dictionary was an innovative.

What does and her scanning of the chapters will led to the new Johnson, of the LSE consequences.

Is that when you take a new term, you try to find @ definition as them.

He was if you are not able to figure out the meaning, and delicate in the cousin with its Bramley will stop

Non-overlapping Scenario Speaker (9)

Skating is a technique to offer news when looking up the word in the telephone book or dictionary.

You search for keywords or ideas.

In most cases, you know what you're looking for, so you are concentrating on finding a particular answer stop

scanning involves moving arise quickly down the page seeking specific words and phrases.

Scanning is also used when you first find a resource to determine whether it will answer your questions.

Once you have scanned @ document, you might go back and skim its.

When scanning, look for the author's use of organisers such as numbers letters as numbers, letters, steps all the words first, second, or next.

Look for words that are bold face, italics or in @ different font size, style or colour.

Sometimes you will will put key ideas in the margin stop

reading of a computer screen has become a growing concern.

Research shows that people have more difficulty reading of the computer screen and off paper.

Although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper.

Similarly, scanning scanning skills are valuable for several purposes in starting science.

First, they are an aid in locating new terms which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary or glossary.

Thus a preliminary scanning of the chapters will alert you to the new terms and concepts and their sequence.

When you locate a new term, try to find its definition.

If you are not able to figure out the meaning then, then look it up in the glossary or dictionary.

Overlapping Scenario Speaker (9)

Or scanning the technique to offer new when looking out in a word in the telephone book or dictionary.

Are you searchable teams were nuts or ideas him.

In most cases, and you know what you are looking for, so you you are concentrating on finding a particular answer.

Issues of scanning involves moving your eyes quickly down the page seeking a specific words and phrases.

Scanning is also used when you first find a resort @ determines whether a glancing request tents.

Once you have a scanned @ document, or you might go back and skimming.

What when scanning, look for the author's use of organisers such as numbers, letters @, or to get chortled words first, second, it was second. The new live, it

looked to words that are bold face, her italics or renege from the font styles, and is scrapped statues will look the word is that our gold price, italics or in a different lots to size it, style or colour. As

Singh, sometimes the authors will put key ideas in the margin.

Reading of the computer screen has become a growing concern.

Research shows that people have more difficulty reading of a computer screen than @ paper.

Although they can read and comprehend is the same rate as pavement or, skimming on the computer is much slower than on paper.

There is similarly, scanning films are valuable for several purposes in studying science.

First, they are in aid in locating U-turns of which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary @ glossary.

Last preliminary scanning of the chapters were lurching to the new terms and concepts and their sequence.

When you like add a new turn, try to find its definition. That

if you are not able to figure out the meaning, then lock it up in the glossary or dictionary.

Non-overlapping Scenario Speaker (10)

Scanning is a technique he often use when looking at the word in the telephone book or dictionary.

You search for @ words or ideas.

In most cases, you know what you are looking for, so you are concentrating on finding a particular answer.

Scanning involves moving your eyes quickly down the page seeking specific words and phrases.

Scanning is also used when you first find a resource to determine whether it will answer your questions.

Once you have scanned @ document, you might go back and skim it.

When scanning, looks at the author's use of organisers such as numbers, letters, steps or the first words, second, or next.

Look for the words of the block faced that bald faced, by telex or in a different font size, style or crop colour.

Sometimes the author will put key ideas in the margin will stop

Reading offer computer screen has become a growing concern.

Research shows that people have more difficulty reading offer computer screen that of paper will stop

although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper will stop

similarly, scanning skills are valuable for several purposes in studying science will stop

first, they are in aid they are an aid in locating new terms which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary or glossary.

Thus a preliminary scanning of the chapter will alert you to the new terms and concepts of the sequence.

When you located near term, try to find its definition.

If you are not able to figure out the meaning, then look it up in the glossary or dictionary.

Overlapping Scenario Speaker (10)

Standing is had often use when looking at the word in the telephone book or dictionary.

Some research to achieve worker ideas or stock

in most client cases, all you had to know what you look you're looking for, so you are concentrating on finding a particular answer.

@ Involves moving your eyes canape seek his ticket were and phrases.

Spending is also you use when you first find a resource to determine whether it will to your questions.

Once you have scanned @ document, you might go to installation.

When scanning, looks at the author's use of organisers such as numbers letters, steps of words first, it second, or next.

Hong Kong resident old faithful, I tell it to a different lifestyle route, stylus colour who.

Sometimes the author key ideas into the margin.

Reading offer computer screen has become a growing concern.

The research shows that people have more difficulty reading of a computer screen and off paper.

And though they can read and comprehend at the same rate as painter., to stemming on the computer is much lower than on paper.

Rely on early, scanning skills are valuable was from several purposes and darting signs.

First, they and 18 locating new terms which are introducing the chat of.

Unless you understand the @ terms, it is impossible to follow the author's reasoning without dictionary of glossary will stop

does the preliminary standing of the chapters on the unit of the new terms and concepts and this sequence will stop

when you locate a new term, try to find @ definition.

If you are not able to figure out the meaning, then look it up in the glossary or dictionary will stop

Non-overlapping Scenario Speaker (11)

Scanning is a technique he often use when looking up the word in the telephone book or dictionary.

You search for keywords or ideas.

In most cases, you know what you are looking for, so you are concentrating on finding a particular answer.

Scanning involves moving your eyes quickly down the page seeking specific words and phrases.

Scanning is also use when you first find a resource to determine whether it will answer your questions.

Once you have scanned the document, you might go back and skim it.

When scanning, look for the authors use of organisers such as numbers, letters, steps or the words first, second, or next.

Look for words that are bold face, italics or in the different font size, style or colour.

Sometimes the author will put to the ideas in the margin.

Reading offer computer screen has become a growing concern.

Research shows that people have more difficulty reading off a computer screen and off paper.

Although do they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper.

Similarly, scanning skewers are valuable to several purposes in studying science.

First, they and eight in locating new terms which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the authors reasoning without dictionary or glossary.

Thus a preliminary scanning of the chapters will alert you to @ new terms and concepts and their sequence.

When you locate a new term, try to find its definition.

If you are not able to figure out the meaning, then look it up in the glossary or dictionary.

Overlapping Scenario Speaker (11)

Will need often use when looking out or were @ the telephone book or dictionary.

To use surgical words or ideas were forced

in most cases, it in a what you are looking for, after you are concentrating on finding a particular answer.

Standing involved in moving your eyes quickly down the page seeking specific words and phrases it.

Is gaining is also used when you first defined @ resource to determine whether it will answer your question dots.

Once you have scanned @ document, it might go back and skin.

When scanning, as you look at the author's use of organiser such as numbers, letters, steps all the work's first, and second, or next.

But it is of little words that are bold facelift, italics or in a different font size, or style wall colour.

Sometimes all the key ideas in the margin.

It's a reading of a computer screen and it has become a growing concern that.

Research shows that people have more difficulty reading of a computer screen than of paper.

Although they can read and comprehend at the same rate as placer, skimming a on the computer is much slower than on paper.

Similarly, scanning skewers are valuable for several purposes in space science.

First, they and eight in locating new Jones which a produced in @ chapters.

Unless you understand the new terms, it is impossible to follow the author's reasoning without dictionary glossary.

But at a preliminary scanning @ the chapters will alert you to the new terms and concepts and @ sequence.

When you locate a new term, try to find its definitions.

If you are not able to figure out the meaning, is that they look it up in the dock in the glossary at dictionary.

Notes

The highlighted words are errors made by the ASR software , these mainly describes substitutions, confusion pairs and other phonetic mistakes made by the software and @ is also a kind of error made by software by deleting or inserting word.

Appendix D: After Processing Scenario

D.1 Overlapping Scenario Improvement (DSTO)

6x@Used to 6x@telephone: a dictionary stock

if certain key words are right ears.

In most cases, you find what you're looking for, so you're concentrating on finding a particular answer.

Scanning involves moving your eyes quickly down the page seeking specific words and phrases.

Scanning is also used when you first find a resource to determine whether it will answer your questions

once you @ scan a document, you might go back and skin @.

@ Scanning, look to the author's use of organisers such as numbers, letters, steps of the words first, second, or next

look for words of a boldface, italics or indifferent. It is, star will cover.

Sometimes the author will put the ideas in the margin.

My reading of a computer screen has become a growing concern.

Research shows that people who have more difficulty reading of a computer screen than of paper.

Although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper stopped

similarly, scanning shill skills of valuable @ several purposes in studying science stop

first they relate are locating new terms which are introduced in the chapter.

Unless you understand the new tax, it is impossible to follow the author's reasoning without dictionary @ glossary.

At a preliminary scanning of the chapters will alert you to the new terms and concepts in their sequence stop

when you locate a new term, try to find its definition.

If you're not able to figure out a meaning, then Governor glossary or dictionary.

D.2 Overlapping Scenario Improvement (ITR)

6x@ Use were 6x@ telephone: what dictionary.

You sir keywords or ideas.

In most cases, you find what you're looking for, so you're concentrating on finding a particular answer.

Scanning involves many arise quickly down the page seeking specific words and phrases. My scanning is also used when you first find a resource to determine whether it will answer your questions

once you @ scan a document, you might get back its gimmick.

They scanning, look to the author's use of organisers such as numbers, letters, steps of the words first, second, or next

look for words of a boldface, et al exploring different font sizes, style trouble.

Sometimes the author will put the ideas in the margin.

Reading of a computer screen has become a growing concern.

Research shows that people who have more difficulty reading of a computer screen than of paper.

Although they can read and comprehend at the same rate as paper, skimming on the computer is much slower than on paper stocks

similarly, scanning shield skills of valuable to several purposes in studying science stock first they are later locating new terms which are introduced in the chapter.

Unless you understand the new terms, it is impossible to follow the orders reasoning without dictionary @ glossary.

At a preliminary scanning of the chapters will alert you to the new terms and concepts in their sequence stop

when you locate a new turn, try to find its definition.

If you're not able to figure out a meaning, then a gap in the glossary or dictionary.

Note

The highlighted text includes the mistakes that are present after the implementation of speech separation algorithms and @ is also errors in the form of insertions and deletions.

If the speaker says a wrong word in place of the actual word then it is not consider as a mistake made by the software.

Appendix E: Hardware and Software

This appendix describes the hardware and software used for measurement of impulse response in conjunction with automatic speech recognition. It explains them in detail.

E.1 Computer

The computer used for playing, recording and simulating the acoustic impulse response and also for automatic speech recognition system is equipped with a RAM of capacity 521,784 KB and sound card that I have used both for playing and recording is SIS 7018.

This sound card is compatible with the audio editing software's like Cool Edit, Cakewalk and gives satisfactory performance in terms of playing and recording at the same time.

E.2 USB Audio Interface

It is capable 24-bit, 96 kHz audio input and output. It is enclosed into a compact, solid metal body. Once connected, power is simply taken from the USB bus.

E.2.1 Features of USB Audio Interface

This device hosts audio interfacing technology in a small, durable body for incredibly portable USB audio recording.

It includes a pair of Neutrik XLR/TRS combo inputs with Phantom power, Hi-Z (Guitar) connection, S/P DIF optical I/O, analog outputs, 1/4" headphone output, and MIDI I/O.

With an integrated low-noise, wide-range power supply, this device can be bus-powered with balanced audio inputs & outputs for professional quality.

It is a completely portable recording centre with very clean, high-gain microphone preamps and premium analog components perfect for home studio and field recording.

This device allows more liberty in setting recording levels with the built-in limiter. So the problem of audio clipping is being compensated by this limiter.

This device offers a direct monitor control on the front panel, allowing hearing recording onto the computer without listening through the program.

It comes with WDM, ASIO2.0, & Core Audio drivers for low latency with most audio applications.

It can also use the OS-standard driver, using the ADV switch.

E.2.2 Components

Following are some of the components that can play major role in impulse response scenario.

Combo Input Jacks

These are analog audio input jacks with microphone pre-amplifiers. They accommodate either XLR or phone plugs, in order to connect a variety of equipment. Either balanced or unbalanced signals can be connected. Phantom power of 48 volts is provided for XLR type connections therefore allow condenser microphone to be connected since these kinds of microphones require phantom power.

Input Sensitivity Knobs

They are used to adjust the input level of the signals input to the front panel, i.e. combo input jacks. They adjust sensitivity level of input signal. In case of clipping, they can be used to adjust the amplitude of the audio signal.

Peak/Limiter Indicator

It indicates distortion in the input signal and also whether the limiter is operating. The indicator will function as a limiter indicator. When the input signal exceeds a certain level, the limiter will operate and the indicator will light green. The indicator will function as a peak indicator. In this case using input sensitivity knobs to overcome it as mentioned earlier.

USB Indicator

This will light blue when the USB device is connected to computer via a USB cable and the computer has correctly detected the device.

Stereo/Mono Select Switch

This selects whether the signal input via the combo input jacks is to be monitored in stereo or in mono mode.

Director Monitor Volume

This adjusts the volume at which the signal input via the combo input jacks is output from the headphone jack and the master output jacks. When this is fully turned clock wise, the signal monitored will be at +6 decibel.

Director Monitor Indicator

This indicator will be lit if director monitor is on, and unlit if direct monitor is off. If the digital input switch is on, the director monitor indicator will be extinguished and the direct monitor function will be off.

Limiter Switch

This turns the USB Audio Interface hardware limiter on/off. If a sudden, high volume sound input is fed to the combo input jacks, the limiter applies mild compression to prevent clipping from occurring at the AD converter. Clipping noise will be heard if the input signal exceeds the capacity of the limiter.

Phantom Power Switch

It is an on/off switch for the phantom power supplied to the XLR connectors of the combo input jacks on the front panel. It is important to use this switch in case condensers microphones are being used otherwise it will produce malfunctioning of the device.

USB Connector

It is used to connect USB device with the computer.

Advance switch

This switches the driver operation mode which operates at sampling frequency of 44100 kHz and 16 bit mono only. If this mode is to be used then it is important that the software should be closed and USB cable be disconnected and reconnected after setting this switch. When this switch is on, then we can record, play, and edit audio from sequencer software or audio editing software with high quality and stable. When the switch is off, then it is in standard mode which is mentioned earlier.

96 kHz Rec/Play Select Switch

When using this device at 96 kHz recording or 96 kHz playback. If the setting is to be changed then disconnecting the USB cable and reconnecting it again will bring new setting into activation.

Sample Rate Select Switch

It talks about the sample rate at which audio data is recorded and played back. If its setting is to be changed, then again same procedure of disconnecting and reconnecting USB cable will be implemented here.

E.3 Omni-Directional Microphone

This is a triangular shaped microphone that can pick signal from all directions. In other words its pick up angle is 360 degree. It has an internal pre-amplifier. It is surface-mounted electric condenser microphone and it is being designed for measuring acoustic paths in acoustic environment. It can be mounted on conference tables, stage floors and lecterns. Due to their sensitivity they can pick speech and other sound such as excitation signals like chirps, maximal length sequences and other natural sounds like claps, whistles, bird's chirps, bats chirps and artificial sounds like fire from pistol. They can be used in recording applications.

E.3.1 Placement

To maintain the flattest possible low frequency response and optimum rejection of background noise, the microphone should be placed on the flat surface which should be as large as possible. Surface can be a floor, table or lectern. Microphone should not be placed close to reflecting surface. If it is placed closer to a reflecting surface, then it might result in increased level of reverberant sound.

E.4 Uni-Directional Microphone

These microphones are flexible with their fully adjustable neck. They are designed for easy replacement and are available in cardioid, supercardioid and omni-directional polar patterns. The interchangeable cartridges offer wide frequency response and accurate sound reproduction for a wide variety of applications, such as churches, courtrooms and conference centres. They can be modified for enhanced convenience and consistent sound.

E.4.1 Features

Wide dynamic range and frequency response for accurate sound reproduction across the audio spectrum Interchangeable cartridges provide the right polar pattern for every application.. Shock mount provides over 20 dB isolation from surface vibration noise. Locking flange mount for permanently securing microphone to lecterns, pulpits, or conference tables.

E.5 Audi Editing Tools

I have used two audio editing tools for measuring room impulse response. Their description is given in the sequence.

E.5.1 Cool Edit

I have used Cool Edit Pro (Version 2) for room impulse response measurement. It is user friendly software. It can simultaneously play and record audio data. It is a dynamic digital sound editor for computers that are using window as operating system. It can record and modify a single wav file. It does not require plug-ins from other applications since plug-in support is built-in within this software so it is easy to use DirectX compatible audio plug-ins in this application. It can record and mix up to 128 stereo tracks using any sound card which is compatible with Window operating system. It also supports multiple multi-channel sound cards.

E.5.1.1 How to use

I have used a laptop computer for playing a recording the tracks in Cool Edit, but these were my initial recordings, that basically helped me a lot before I did my final recordings in Cakewalk which is another audio editing tool. So it is basically a kind of testing scenario for me while I was using Cool Edit for simultaneously playing and recording wav files. Laptop that I have used in this case is having a sound card Ali Accelerator and is equipped with a processor with a processing speed of 933 MHz. After installing it, first of all make it sure that the audio data that is going to be imported into this software, is already saved in some folder in computer. After this open the software, from the option box we can adjust wave out and wave in from device properties. In wave out and wave in box, I have selected wave mapper for both these in and out waves. This wave mapper can support different sampling rates and bit resolution. Then after returning to the menu bar, I open the new session from the file menu. It basically asks for the multi track session that is based on sampling frequency that will define the sampling rate of the audio data to be imported. The default sampling rate is 44100 Hz; in case of other sampling rates this software will make a new copy of the sampling rate besides the default sampling rate. I used the default rate through out the impulse response measurements along with 16 bit resolution. Then from main menu bar, go to insert menu and from there use the command 'wave from file' that will import the audio file and it will also briefly describe the size of file, its duration, bit resolution and sampling rate. Then this audio data, a wav file in this case is being adjusted by selecting wave mapper for playing this file, either in stereo or mono channel format. Then it is being highlighted by using the mouse starting from right edge of the wav file to the beginning of the wav file and in this way it will automatically adjust the size for other track that will be used for recording and a perfect synchronization can be obtained between transmitted and received wav files. Each track has three buttons in colour that basically are used for transmission of the file. The button 'S' is used to play a solo wav file, button 'R' is used for recording and 'M' is used to mute the track, in case other track is being used for recording. For playing and recording simultaneously, activate 'S' for the file to be transmitted from the track where it is being imported and for recording select 'R' from other track. This track can be selected in either mono or stereo mode. For playing this wav, the sound card Ali Audio is being used and to record, the USB audio preamplifier is used. Most of the time I was using mono mode for recording but I also used stereo modes as well in my recordings. After recording of the signal, I can also double click the received wav file to analyze it. Similarly I can analyze transmitted wav file. Analysis can be frequency analysis, phase analysis or waveform statistics that include all the necessary details of the wave like minimum sample value, maximum sample value, peak amplitude, clipped samples if there are any, DC Offset, minimum RMS Power, and maximum RMS Power, Average RMS power, actual bit depth and window width.

E.5.2 Cakewalk

It is another audio editing tool. It is a professional tool for playing and recording sound and music on personal computer for various purposes depending upon the ingenuity of the user. It is being designed for audio and production engineers, multimedia and game developers, and recording engineers. It supports most of the popular audio formats that includes wav, midi, mp3. It has built in plug-ins to run and edit the audio data rapidly and efficiently. In order to record audio data the computer monitors the electrical signal generated by a microphone that I am using in this case for recording the generated wave file that I created and then imported into this software. As the signal is caused by a sound, the signal strength varies in direct proportion to the sound's waveform. The computer measures and saves the strength of the electrical signal from the microphone, thus recording the waveform. There are two important parameters of this measuring process. First is the sampling rate, the rate at which the computer saves measurements of the signal strength. According to Nyquist theorem we must measure, or sample, the signal at a rate at least twice that of the highest frequency we want to get. Since humans can hear frequencies well above 10 kHz till 22 kHz, most soundcards and digital recording systems are capable of sampling at much higher rates than that. Typical sampling rates used by audio engineers are 22 kHz, 44.1 kHz, and 48 kHz. The 44.1 kHz rate is called CD-quality**,** since it is the rate used by audio compact discs. The other important aspect of the measuring process is the sampling resolution. The sampling resolution determines how accurately the amplitude of each sample is measured. At present, the music industry has settled on a system that provides 65,536 different values to assign to the amplitude of a waveform at any given instant. Thus, each sample saved by the computer requires 16 bits to store, since it takes 16 bits to store a number from -32,768 to 32,767. The scaling of the electrical input signal level to amplitude value is determined by the audio hardware and by the position of input level control. Before installation to window environment, it should be confirmed that the system used is as per the software's requirement. We can use Window XP or Windows 2000 as operating environment for Cakewalk and a central processing unit with a speed of 800 MHz and 128 Mb RAM, hard disk should be of 100 Mb capacity. Resolution should be 1024x768-16 bit color graphics and audio interfaces. After confirming system specification, it is then installed and then run for audio data editing. Then new project can be opened in which we have variety of options available. From main menu bar we can adjust settings by roaming in option bar and then to audio options, where we can define the playback timing master (soundcard driver normally) and record timing master (an audio pre-amplifier) in case of audio file being measured. Also here we can define sampling rate for the new project along with file bit depth. I have used default sampling rate and file bit depth which are 44100 Hz and 16 bit. Here buffer size can also be adjusted to reduced time delay between the transmitted (played) and received (recorded) wave file. We can also activate wave profiler to profile the sound card installed in the computer. Also we can drivers for both input and output devices. Then audio data file can be imported using import command from file menu of the main menu bar. Once the audio data has been imported in the track, then we can define the input and output devices for this audio data and like Cool Edit, all the tracks have the buttons for playing, recording and mute depending upon how they are being used. To obtain synchronization between played and recorded wave file, we can select punch region and punch points by right clicking the file to be played and then select the punching points. I have done my final recordings for impulse response and also automatic speech recognition testing, using this software.

Appendix F: Source Codes

F.1 Room_Dem3

```
if exist('mode')~=1,  mode =3; end
if mode<3
   Fs = 1e4;    % assume 10 kilohertz  sample rate
   kn = .1;     % assume we add some noise at the microphone
   if mode==2, tx= chirp((1:Fs)/Fs, 100, 1, 1000);   end
   if mode==1, tx= sign(randn(1,Fs));   end
end

if mode<3
   y = conv(tx, [ 1 2 -1  0 .5 .2 ]);     % simulate the low pass response of the loudspeaker

   rx  =conv(y, [zeros(1,20) 1 0 0 0 0 0 0 0 0  0  .5 ]);   % simulate very simple room
   % assuming 2msec delay for line of sight + one echo 1 millisecond later
Nrx = length (rx);
   rx = rx + kn* randn(1,Nrx);                                        % simulate
added noise
else

 [tx, Fs] = wavread(ftx);
```

```
  [rx, Fs2] = wavread(frx);
  if exist('Nlim')==1                    % reduce numb of samples ?
     rx = rx(1:Nlim);
     tx = tx(1:Nlim);
  end

  if Fs==Fs2
     disp([' Sample rate is ', num2str(Fs)]);
  else
     error (' sample rates are not equal! ');
  end
end

Ntx = length (tx);  Nrx = length(rx);
if Ntx == Nrx
  disp([' Number of samples is ', num2str(Ntx)]);
else
  disp(' vector lengths not equal ');
  if Nrx<Ntx,
     disp(' rx is shorter ! ');  tx = tx(1:Nrx);  Ntx = Nrx;



 else
     rx = rx(1:Ntx);    Nrx = Ntx;
  end
```

```
   disp([' New number of tx and tx samples is ', num2str(Ntx)]);
end


figure(1);  plot((1:Nrx)/Fs, rx);        % plot received  signal
title(' Received signal');
figure(4);   psd(rx, 512, Fs);
title (' Rx signal spectrum ');
xlabel(' Frequency (Hz) ');


% do the cross correlation of rx with either x (usual case) or with y for 2 microphone case
xc=xcorr(tx, rx);   %  NB this is the right order for matlab 5 but matlab 7 swaps operands!!
%%   xc=xcorr(y, rx);


N2 = length (xc);
figure (2);  plot((1:N2)/Fs, xc )
title (' Whole crosscorrelation ');


if exist('Twid')~=1,   Twid = 0.01;   end  % default display is +- 10 milliseconds



Nwid = floor (Twid*Fs);


figure(3);   % plot the middle part of the cross correlation
xc_mid = xc(floor(N2/2)-Nwid:floor(N2/2)+Nwid);
plot((-Nwid:Nwid)/Fs*1e3, xc_mid )
```

```
xlabel(' time in millisecs ')
title(' Central portion of crosscorrelation ');
figure(9)
specgram(rx)
```

F.1.1 Description

This source code shows acoustic channel identification using real signals and synthetic signals. The main idea behind this source code is to see whether the cross-correlation of two deterministic looks like that of the synthetic signals which are chirp and a random signal and actually both of them are used as excitation signals for impulse response in room acoustics. They are defined by two different modes. Mode one defines that we are simulating a maximal length sequence synthetically and mode two produces a chirp simulation. Mode three is the one which is actually being used for cross-correlating the two signals to get the room impulse response which should produce a response similar to the synthetic responses obtained from modes one and two. If mode is less then mode three then we can simulate either of the two other modes using sampling frequency of 10kHz and a noise factor that is being added in the microphone synthetically and then we generate these two modes using commands ‘chirp’ and ‘sign(randn()). Then we can convolve either of these two modes (mode1 & mode2) with a vector to see low pass response of a speaker synthetically. Mode three is a flexible mode in a way that we can change the number of samples to see the cross-correlation between the transmitted signal (ftx) and received signal (frx). Another parameter we have defined here is ‘Twid’ that, adjusts the length of cross-correlation in time domain in milliseconds. For true cross-correlation the number of transmitted and received signal samples should be the same, otherwise it will be displayed by this source code that the samples are not equal as an error using a command num2str and the program will not go further until we adjust the length of both these vectors (ftx,frx). This code produces diagrams of cross-correlation in terms of middle of cross-correlation, whole cross-correlation, power spectrum magnitude of the received spectrum and received signal and specgram of the received signals.

F.2 Call_Dem3

This code calls the room_dem3 code inside it to perform cross-correlation between transmitted (ftx) and received (frx) signals. It uses all the sweeps that I have mentioned in my research work.

F.3 Source Code for MLS

```
L=16; S=2^L-1;
sr=zeros (1,L);
sr1(1)=1;sr(3)=1;sr(5)=0;sr(8)=0;sr(11)=1;
sr (16)=1;
register=zeros(1,S);
for Index=1:S;
   operation=xor(xor(xor((sr(16)),sr(15)),sr(13)),sr(4));
   sr(2:L)=sr(1:L-1);
   sr(1)=operation;
   Output(Index)=sr(16);
end
Calculate=~Output;
sequence=0.5*(Output-Calculate);

Fs=44100;
wavwrite(sequence,Fs,16,'mls_16.wav')
```

F.3.2 Description

[1], [12]: It generates an excitation signal of length 16. Where the MLS with length defined as S=2^L-1 runs until it completes the sequence obtained from subtracting the total length of the signal with the defined number of integer, which continuously subtracts the original size until it becomes equal to the actual buffer size and in this case the for loop ends. Inside for loop,

index defines the length of the sequence starting from 1 till S. Then it can be changed into a wav file by a command wavwrite with parameters like sampling frequency and bit rate that would be used for visualising auto-correlation of this maximal length sequence.

Appendix G: ASR Before and After Processing Summary Tables

Table G1: Automatic Speech Recognition testing in terms of word accuracy and errors occurred during non-overlapping and overlapping scenarios

Ref Speakers	**Prof Bill**	**Mr Terry**	**Mr Ian**	**Mr David**	**Mr Nick**	**Ms Christine**	**Mrs Wendy**	**Mr Sridhar**	**Mr Joerg**	**Mr Fazal**		
NOL												
Gender	Male	Male	Male	Male	Male	Female	Female	Male	Male	Male		
Accent	Aus	Aus	Aus	Aus	Aus	Aus	Aus	Ind	Ind	Ind		
Ref Word	277	277	277	277	277	277	277	277	277	277		
Errors	15	28	18	21	12	29	30	40	61	22	27.6	**ME**
Word Accuracy	94.5	89.8	93.5	92.4	95.6	89.5	89.1	85.5	77.9	92	**90**	**MA**
OL												
Ref Word	277	277	277	277	277	277	277	277	277	277		
Errors	86	82	50	41	68	85	79	145	173	183	99.2	**ME**
Word Accuracy	68.9	70.3	81.9	85.1	75.4	69.3	71.4	47.6	37.5	33.9	**64.1**	**MA**
ITR												
Ref Word	277	277	277	277	277	277	277	277	277	277		
Errors	37	42	28	31	18	44	47	55	69	60	43.1	**ME**
Word Accuracy	86.6	84.8	89.8	88.8	93.5	84.1	83	80.1	75	78.3	**84.4**	**MA**
DSTO												
Ref Word	277	277	277	277	277	277	277	277	277	277		
Errors	36	34	26	35	21	38	47	53	85	54	42.9	**ME**
Word Accuracy	87	87.7	90.6	87.3	92.4	86.2	83	80.8	69.3	80.5	**84.5**	**MA**

Note: In the above table following are the abbreviations
ME: Mean Error
MA: Mean Accuracy
OL: Overlapping
NOL: Non-overlapping

Table G2: Improving word accuracy of the overlapping scenarios via speech separation algorithms both ITR & DSTO for final conclusion

Speakers DSTO (OL)	2	3	4	5	6	7	9	10	11	**Mean**
Gender	Male	Male	Male	Male	Male	Female	Male	Female	Male	
Accent	Aus	Ind	Aus	Ind	Aus	Aus	Aus	Aus	Aus	
Ref Words	277	277	277	277	277	277	277	277	277	
Errors	39	53	43	54	26	48	34	43	21	40.1111
Word Accuracy	85.92%	80.87%	84.48%	80.51%	90.61%	82.67%	87.73%	84.48%	92.42%	**85.52%**

Overall	36	62	50	59	X	48	26	42	20	42.875
	35	57	50	64	29	63	27	45	26	44
DSTO (OL)	**87.00%**	**77.62%**	**81.95%**	**78.70%**	**X**	**82.67%**	**90.61%**	**84.84%**	**92.78%**	**84.52%**
ITR	**87.36%**	**79.42%**	**81.95%**	**76.90%**	**X**	**77.26%**	**90.25%**	**83.75%**	**90.61%**	**83.44%**

1 till 3 only										
DSTO (OL)	5	12	7	4	5	8	7	8	4	6.66667
ITR	8	11	12	14	5	24	8	6	8	10.6667
Words	44	44	44	44	44	44	44	44	44	
DSTO (OL)	**88.64%**	**72.73%**	**84.09%**	**90.91%**	**88.64%**	**81.82%**	**84.09%**	**81.82%**	**90.91%**	**84.85%**
ITR	**81.82%**	**75.00%**	**72.73%**	**68.18%**	**88.64%**	**45.45%**	**81.82%**	**86.36%**	**81.82%**	**75.76%**

4 till 16										
DSTO (OL)	29	48	42	50	X	33	16	34	13	
ITR	26	45	37	46	24	32	15	39	15	
Words	184	184	184	184	184	184	184	184	184	
DSTO (OL)	**84.24%**	**73.91%**	**77.17%**	**72.83%**	**X**	**82.07%**	**91.30%**	**81.52%**	**92.93%**	**82.00%**
ITR	**85.87%**	**75.54%**	**79.89%**	**75.00%**	**86.96%**	**82.61%**	**91.85%**	**78.80%**	**91.85%**	**83.15%**

17 till end										
DSTO (OL)	2	2	1	5	X	7	3	0	3	
ITR	1	1	1	4	0	7	4	0	3	
Words	49	49	49	49	49	49	49	49	49	
DSTO (OL)	**95.92%**	**95.92%**	**97.96%**	**89.80%**	**X**	**85.71%**	**93.88%**	**100.00%**	**93.88%**	**94.13%**
ITR	**97.96%**	**97.96%**	**97.96%**	**91.84%**	**100.00%**	**85.71%**	**91.84%**	**100.00%**	**93.88%**	**95.24%**

Note: Following abbreviations are used in the above table
Ref: Reference

OL: Overlapping
NOL: Non-overlapping

Bibliography

[1] H.-J. Zepernick and A. Finger, "Pseudo Random Signal Processing," *Theory and Application*, John Wiley,Chichester, 2005, pp. 173-224.

[2] S. Haykin and M. Moher, "Analog and Digital Communications," John Wiley, Chichester 2006.

[3] H. Kuttruff , "*Room Acoustics,"* Taylor and Francis,2000, pp. 6-88.

[4] L. W. Couch, "*Digital and Analog Communications*," Prentice Hall PTR, July 1992.

[5] J. G. Proakis, J. H. .L. .Hansen, "*Discrete Time Processing of Speech Signals,*" Wiley IEEE Press, 1 Edition, 1999, pp. 99-100 & 677-679.

[6] S. Haykin, "Adaptive *Filter Theory,*" Prentice Hall, September 14, 2001.

[7] Think Quest Team, "Signals Digital Processing",1999, http://library.thinkquest.org/26042/english.html.

[8] The Physics Classroom and Mathsoft Engineering, Inc, "Sound Properties and Their Perception", 2004, http://www.physicsclassroom.com/Class/sound/U11L2a.html.

[9] Wikepedia, "Loudspeaker Acoustics", 2007, http://en.wikipedia.org/wiki/Loudspeaker_acoustics.

[10] Speed Reading Articles, "You Already Use Scanning," 2006, http://www.magicspeedreading.com/words/scanning.html.

[11] A Collection of The World's Fairytales, "The Red Shoes," 2002, http://www.fairytalescollection.com/Hans_Christian_Anderson/The_Red_Shoes.htm.

[12] Maximal Length Sequences, "Example of MLS," 1961, http://www.silcom.com/~aludwig/Signal_processing/Maximum_length_sequences.htm.

[13]J. Holfeld and B. Cowley, "Review of Speaker Separation," Institute for Telecommunications Research, University of South Australia, February 23, 2007.

[14] K. Dogancay, J. Littlefield and A. Hashemi-Sakhtsari, "Performance Evaluation of an Automatic Speech Recognizer Incorporating a Fast Adaptive Speech Algorithm," ICASSP, 2006.

[15] S. Müller & P. Massarani, "Transfer Function Measurements with Sweeps,". *JAES Volume 49 Number 6 pp. 443-471; June 2001*

[16] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-sine technique", *J.AES*, vol.48, p.350, 108th AES Convention, Paris 2000, Preprint 5093.

[17] Twelve Tone System, *Cakewalk* , 2006.

[18] The Sonic Spot, *Cool Edit*, 2006.

[19] Roland USB Audio Interface, *Edirol UA25* , 2005.

[20] The Mathworks, *Matlab Tutorial*, 2006.