

# Blind Dereverberation using Maximum Kurtosis of the Speech Residual

*Vikrant Tomar*

*McGill ID: 260394445*



telecommunications &  
signal processing  
laboratory

Department of Electrical & Computer Engineering  
McGill University  
Montreal, Canada

February 20, 2012

---

© 2012 Vikrant Tomar

## Abstract

Reverberation in speech, caused by room reflections, is problematic specially for hands-free telephonic applications in a confined space. The problem is even severe for hearing impaired people. Therefore blind speech dereverberation is an important research area. The task is to remove reverberation from the output of a room, where the room impulse response, as well as the clean speech signal is unknown. The method discussed herein maximizes the fourth order cumulant, referred to as Kurtosis, of the Linear Prediction (LP) residual of the speech to remove reverberation from the degraded speech.

---

## Acknowledgements

I wish to thank Prof. Peter Kabal for his guidance, support, and unconventional brilliance. In addition, many discussions with people in the TSP lab, especially Hoda Daou, and Marc-Antoine Rondeau have helped in the development of this work.

## Preface

The report will cover analysis, implementation, and test of the maximum kurtosis based algorithm used to deconvolve room impulse response. Primary metrics of the algorithm analysis, discussed herein, include the ability to deconvolve, convergence speed, and the number of required multiplications. The implementation of the algorithm is done on MATLAB R2009a in a 64-bit environment.

In the report, the natural algorithm is denoted with **log**; otherwise the base is stated. Throughout the report, the whole sequence of a signal will be denoted by a vector written with bold letters, *i.e.*  $\mathbf{x}$ , whereas  $x(n)$  will correspond to an element in that sequence. If both time and vector indices are used the variable  $\mathbf{h}_j(k)$  is the  $j$ 'th coefficient at time index  $k$ , where the bold face represents a vector. Literature references are represented with numbers in IEEE format, e.g. [number], and a the full list of references is found in the References section.

The report also includes the MATLAB implementation of the algorithm.<sup>1</sup>

---

<sup>1</sup>The L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> source of this report, and related MATLAB codes can be downloaded from <http://www.ece.mcgill.ca/~vtomar/reports/dsp2>.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation and Problem Statement . . . . .	1
1.2	Approach . . . . .	2
1.3	Related Literature . . . . .	3
<b>2</b>	<b>Project Definitions and Background</b>	<b>4</b>
2.1	Room . . . . .	4
2.2	Reverberation . . . . .	5
2.3	Speech Production Model . . . . .	5
2.4	Performance Metrics . . . . .	6
2.5	Higher Order Statistics . . . . .	7
<b>3</b>	<b>Dereverberation</b>	<b>9</b>
3.1	Linear Prediction of Speech . . . . .	9
3.2	Maximum Kurtosis based Dereverberation . . . . .	10
3.3	Complexity of the Algorithm . . . . .	12
<b>4</b>	<b>Experiments and Results</b>	<b>14</b>
4.1	Kurtosis and LP residual of reverberant speech . . . . .	14
4.2	Dereverberation Experiments . . . . .	15
4.2.1	Experiment 1 . . . . .	16
4.2.2	Experiment 2 . . . . .	18
4.2.3	Experiment 3 . . . . .	18
4.3	Summary . . . . .	19

<b>Contents</b>	<b>iv</b>
<b>5 Discussion and Conclusion</b>	<b>23</b>
<b>References</b>	<b>25</b>
<b>A Matlab Codes</b>	<b>27</b>

---

---

## List of Figures

2.1	Block diagram of speech convolution with room impulse response. . . . .	4
2.2	The simplest case of one reflective surface. . . . .	5
2.3	The direct and reflected signals in time domain. . . . .	6
2.4	Block diagram of source-filter model of speech production. . . . .	6
3.1	Block diagram of steepest ascent algorithm used for dereverberation. . . .	11
3.2	Modified block diagram to avoid LP artifacts in signal reconstruction from residual. . . . .	12
4.1	Kurtosis vs distance between the source and microphone. . . . .	15
4.2	Waveforms of LP residuals of the clean, reverberant and dereverberated speech - Exp 1. . . . .	16
4.3	Waveforms of clean, reverberant and processed speech - Exp 1. . . . .	17
4.4	Spectrogram of clean, reverberant and processed speech - Exp 1. . . . .	20
4.5	Waveforms of LP residuals of the reverberant and dereverberated speech - Exp 2. . . . .	21
4.6	Waveforms of clean, reverberant and processed speech - Exp 2. . . . .	21
4.7	Waveforms of LP residuals of the reverberant and dereverberated speech - Exp 3. . . . .	22
4.8	Waveforms of clean, reverberant and processed speech - Exp 3. . . . .	22

# Listings

A.1	Matlab code for comparing kurtosis of varying reverberation . . . . .	27
A.2	Matlab code for the kurtosis maximizing algorithm . . . . .	30

# Chapter 1

## Introduction

Sound waves travel as waveforms, which are then reflected by various surfaces, and objects in the room; such reflected multipath signals form a delayed and attenuated feedback of the source signal to the sink/Mic. Such a feedback is referred to as reverberation, and causes degradation of the speech intelligibility.

### 1.1 Motivation and Problem Statement

Speech reverberation, mixed with environmental noise, is regarded as one of the primary issue in the speech capturing in a confined space, for example, business offices. Normally, degraded (additive or reverberant) speech is processed assuming that the degradation has long term stationary characteristics relative to speech; many methods of speech enhancement have been proposed based on this concept. Unfortunately, due to sharp changes in spectrum within a speech frame and across the frames, the resulting processed speech produces significant audible distortions. Such noise reduction is accomplished at the cost of quality. Thus it is required to look at the methods focusing on characteristics of speech for enhancement of degraded speech rather than the degradation itself.

Over the years, several methods of speech dereverberation based on a simplified discrete model of speech production have been proposed. The basic model consists of an excitation source, and a time-varying vocal tract filter. Such as model can easily be modeled by an Auto-Regressive (AR) linear prediction (LP) technique. The inverse LP filter gives the LP-residual, which is a close approximation of the excitation signal. The motivation for this project is the observation that in reverberant environments, the LP residual contains

the original impulses followed by several other peaks due to multi-path reflections. Thus, dereverberation can be achieved by modifying the spectral envelope and/or the excitation signal.

## 1.2 Approach

It has been empirically established that, for clean voiced speech, LP residuals have strong peaks corresponding to glottal pulses, whereas for reverberated speech such peaks are spreaded in time [1], in other words, LP residual of reverberated speech is a time-spread version of the relatively more *peaky* LP residual of clean speech. Thus, amplitude spread, in a degraded signal, can be seen as a reverberation metric. Recent researches have suggested to look at kurtosis, which is a degree of peakedness of a distribution, as a reasonable measure of reverberation [2, 3]. The goal of this project is to study an on-line gradient-ascent algorithm to maximize LP residual Kurtosis, as proposed in [4]. In such an attempt, there will be more emphasis on the speech – and enhancement for human intelligibility – than on the degradation during the enhancement. The project will focus on single microphone setup for it is more practical as well as challenging. Furthermore, the scenario considered here is that of Blind Dereverberation, where neither the clean signal, nor the room impulse response (RIR) is known. This is the case in most practical situations.

Based on the targeted application, the problem of dereverberation can further be divided into two major classes, namely dereverberation targeted Automatic Speech Recognition (ASR) systems, or dereverberation targeted at making the signal more intelligible to humans. This work is mainly targeted at the latter class. When we want to remove the dereverberation for better human intelligibility we want to make the signal sound better while protecting the spectral distribution of formants; it is better to leave some reverberation than causing spectral distortion. However, when the target is ASR, the goal is to maximize the Signal-to-Noise ratio by removing as much reverberation as possible; a nice speech – in terms of human perception – is not aspired.

There have been a lot of work done in this area. Some of the interesting methods are presented in the next section.

### 1.3 Related Literature

The RIR, in general, is non-minimum phase. A non-minimum phase filter is a mixture of a minimum phase filter, where all singularities lie inside the unit circle, and a maximum phase filter, where all the singularities lie outside the unit circle. Therefore, the inversion of a non-minimum phase filter will result in a filter which has poles outside the unit circle, such a filter cannot be causal and stable at the same time. Thus the approach available is to estimate a near-inverse RIR filter to cancel the effect of room in the degraded speech. For the same reason, second order statistics cannot be used to reconstruct a direct signal, and therefore, higher order statistics, such as kurtosis, are required.

The idea of using kurtosis to remove reverberation from speech was at first proposed by Tanrikulu *et. al.* in [3], as Least-mean kurtosis. However, the authors did not exploit any speech specific properties of the input signal. Later, Gillespie *et. al.* in [4] proposed an LMS-like gradient maximizing algorithm that maximizes the kurtosis of the LP residuals of the speech signal to the clean speech. LP residual has been used as an efficient metric for the reverberation in speech, and many different algorithms has been proposed utilising LP residual for reverberant speech enhancement. Authors in [1], for example, have utilised an LP residual weighting scheme which enhances the regions with high signal-to-reverberation ratio in a speech signal. Other methods of dereverberation include spectral subtraction, as used in [5]. In [6], authors have proposed to use the CELP postfilterfor dereverberation.

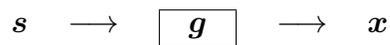
## Chapter 2

# Project Definitions and Background

This chapter present a brief introduction to widely used concepts in this report.

### 2.1 Room

In this section the room and its properties are described. The room is described a the filter,  $\mathbf{g}$ , between the source and sink *i.e.* speaker and microphone. The clean output of the speaker is the speech,  $\mathbf{s}$ , and the signal at the sink is the reverberated speech  $\mathbf{x}$ . The effect of ambient environment on speech can be modeled as a convolution in time of the speech signal and the RIR [7]. Furthermore, the system can also be modeled as containing additive noise [1]. This is illustrated in Fig. 2.1 and Eq. 2.1.



**Fig. 2.1** Block diagram of speech convolution with room impulse response.

$$x(n) = g(n) * s(n) + w(n) \quad (2.1)$$

Where,  $n$  is the time index, and  $w(n)$  is the additive noise.

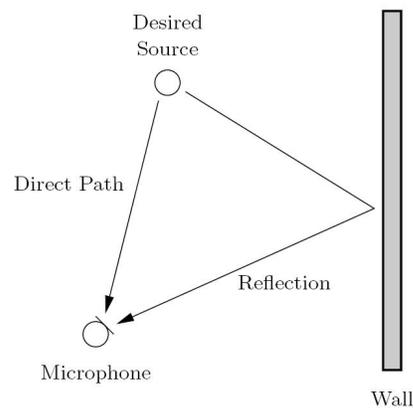
In this project, different room models were used. For many experiments, simplified room filters were defined using the famous image-source method. However, the implementation proposed by Lehmann and Johansson [8], which promises to address the problem of anomalous tail decay in the original image-source method proposed by Allen and Berkley [9], was used. Furthermore, real life RIRs were downloaded from Aachen university's Aachen

Impulse Response database [10], and convolved with clean speech signal to produce reverberated speech.

## 2.2 Reverberation

When a person is speaking in a regular room, the listeners will not only perceive the direct speech signal, but also various multipath copies of it created by reflections on the room walls and other objects. The multipath signals are delayed and possibly attenuated as compared to the direct signal. This phenomenon is known as reverberation. A very simple scenario with only one reflective surface is illustrated in Fig. 2.2. To tackle this degradation, it is of interest to minimize the effect of the room.

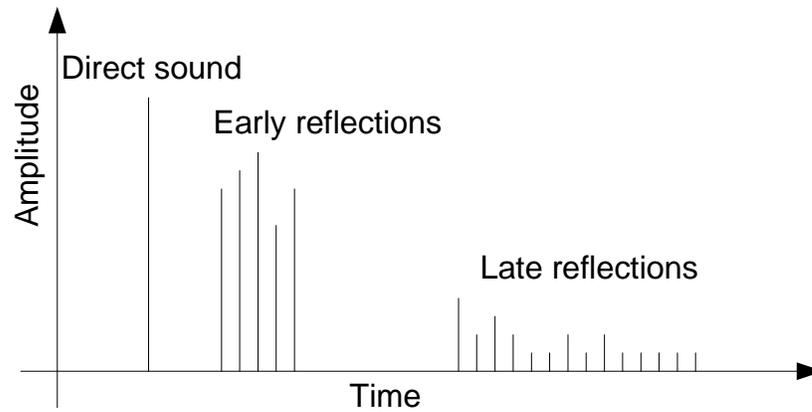
In time domain, such effects can be divided into early and late reflections. Early reflections can be defined as first 50-100 ms of the reflection. Early reflections are good for the speech intelligibility for humans, as they provide information related to the acoustic environment, such as, size of the room, and the position of the speaker in the room. Fig. 2.3 illustrates the direct, early and late signals.



**Fig. 2.2** The simplest case of one reflective surface.

## 2.3 Speech Production Model

The vocal tract can be modeled as an autoregressive (AR) process. The voiced sounds are generated when the input is given as quasi-periodic pulses, also referred to as the glottal

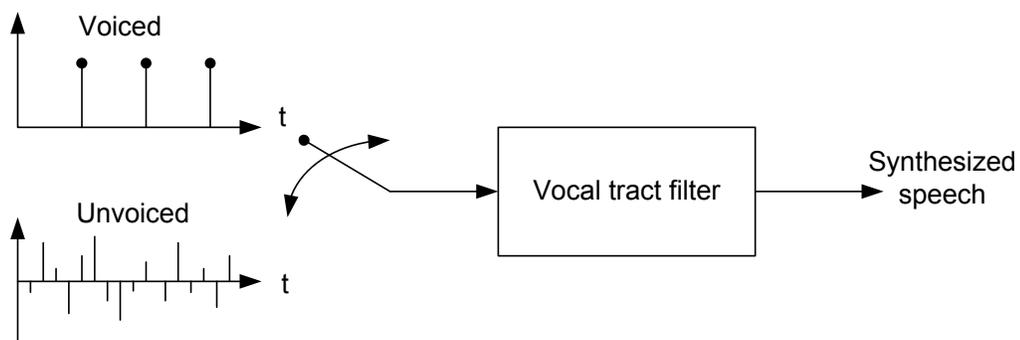


**Fig. 2.3** The direct and reflected signals in time domain.

pulses. Unvoiced sounds are made when the input is white noise. This simple source-filter model is illustrated in Fig. 2.4.

## 2.4 Performance Metrics

The selected metrics for algorithm evaluation are computation complexity, measured in the number of used multiplications, spectrogram, and subjective quality measures such as, feedback from many listeners. Spectrogram gives visual cues of formant spreads in time because of reverberation. Both computational complexity and subjective quality are important metrics for an hearing aid application, which has limited computational powers at its disposal. Therefore, it is important that the algorithms are as effective as possible, and giving as clear output as possible. Other important metrics are execution time (measured



**Fig. 2.4** Block diagram of source-filter model of speech production.

in seconds or clock cycles) and the amount of required memory for program and data storage. The execution time and memory requirements will however not be analysed in this project, but an estimate of how far the algorithms are from a real time application will be considered. Also, mean-square error between the original and reconstructed signal is not used here, because the system is *not* driven to minimize the mean-square error, and minimum mean-square error does not necessarily correspond to better sounding speech.

## 2.5 Higher Order Statistics

The main advantage of using higher order statistics follows the hypothesis that the received signal at the Mic,  $\mathbf{x}(n)$ , can be considered as composed of a Gaussian distributed, and a non-Gaussian distributed component. One important assumption here is that the input clean speech is non-Gaussian. A reverberated speech signal is a multipath signal, and it is represented as weighted, delayed, and summed copies of the same signal. Hence from central limit theorem (CLT) – which states that the distribution of the sum of independent and identical distributed (iid) signals is approximately Gaussian – the ambient noise and reverberation added to the signal because of convolution with RIR, can be considered as being Gaussian distributed.

All Gaussian distributed signals have higher order statistics equal to zero. The idea is therefore to establish a cost function that maximizes the higher order statistics of the reverberated signal, which entails that the processed signal should obtain a pdf that is non-Gaussian, that is, the room impulse response has been removed.

In higher order statistics, we would want to focus on the 3<sup>rd</sup> and 4<sup>th</sup> order cumulants. The  $n^{\text{th}}$  order cumulants  $c_n$  of a random variable  $X$  are defined by the cumulant generating function, which is logarithm of the moment-generating function.

$$g(t) = \log(E[e^{tX}]) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!} = \mu t + \sigma^2 \frac{t^2}{2} + \dots \quad (2.2)$$

Cumulants are then given by derivatives at zero of  $g(t)$ . A more general definition of the  $n^{\text{th}}$  order cumulant for a non-Gaussian stationary random process  $x(k)$  is given in [11, Eq.19]:

$$c_n^x(\tau_1, \tau_2, \dots, \tau_{n-1}) \triangleq m_n^x(\tau_1, \tau_2, \dots, \tau_{n-1}) - m_n^G(\tau_1, \tau_2, \dots, \tau_{n-1}) \quad (2.3)$$

Where, the parameter  $m_n^x$  is the  $n^{\text{th}}$  order central mean of  $x(k)$ , and  $m_n^G$  the  $n^{\text{th}}$  order moment of a Gaussian signal with mean and autocorrelation equal to those of  $x(k)$ . The  $\tau$ 's are different delays, and  $c_n^x$  is the  $n^{\text{th}}$  order cumulant of  $x(k)$ . It is important to note here that the above equation is not valid for  $n = 2$ . The 2<sup>nd</sup> order cumulant is auto-covariance of  $x(k)$ . It is clear from Eq. 2.3 that if  $x(k)$  is Gaussian distributed, the cumulants will be zero. This prove that the higher order cumulants are not affected by the Gaussian noise. Finally, it is worth mentioning three commonly used parameters, which are defined for zero delay, and zero mean (or central mean), *i.e.*, have moments equal to cumulants.

$$\begin{aligned} c_2^x(0) &= E[x^2(n)] = \sigma_x^2 \\ c_2^x(0, 0) &= E[x^3(n)] = \gamma_x^3 \\ c_4^x(0, 0, 0) &= E[x^4(n)] - 3(\sigma_x^2)^2 = \gamma_x^4 \end{aligned} \quad (2.4)$$

Where  $\sigma_x^2$  is the variance,  $\gamma_x^3$  is the skewness, and  $\gamma_x^4$  is the **kurtosis**. The  $3(\sigma_x^2)^2$  factor is important to note as the kurtosis of a normally distributed signal equals to  $3(\sigma_x^2)^2$ . A more widely used flavour of kurtosis is the **normalized kurtosis**, defined as:

$$\kappa_x = \frac{\gamma_x^4}{(\sigma_x^2)^2} = \frac{E[x^4(n)]}{(\sigma_x^2)^2} - 3 \quad (2.5)$$

Thus the normalized kurtosis of a normally distributed signal equals to zero. Kurtosis is a measure of “peakedness”, *i.e.*, a signal with many large value in the middle and small values at the tails has a positive kurtosis.

This concludes the brief presentation of higher order statistics and this chapter. The developed concepts and definitions will be utilized in the forthcoming chapters.

# Chapter 3

## Dereverberation

Having acquired sufficient background in previous chapters, we would look at the kurtosis based method of dereverberation in this chapter.

### 3.1 Linear Prediction of Speech

A Speech signal can be expressed as a linear combination of its past samples. Based on the source-filter model discussed in Section 2.3, the clean speech can be modeled as an output of an all-pole process.

$$s(n) = - \sum_{k=1}^p a_k s(n-k) + u(n) \quad (3.1)$$

where  $a_k$ 's are the corresponding filter coefficients, and  $u(n)$  is the glottal pulse excitation signal. Let's say the predicted signal for the above speech be  $\tilde{s}(n)$ , which can also be modeled as an output of an all-pole process.

$$\tilde{s}(n) = - \sum_{k=1}^p b_k s(n-k) \quad (3.2)$$

where  $b_k$ 's are the linear prediction (LP) coefficients. Now, if the speech signal were truly generated by an all-pole filter, Eq. 3.2 would be an exact prediction of the speech signal at all times, except the glottal excitation instants, *i.e.*,

$$\text{For } a_k = b_k; \text{ error in prediction, } e(n) = s(n) - \tilde{s}(n) = u(n) \quad (3.3)$$

This error in prediction is referred to as LP residual. It is evident from Eq. 3.3 that the LP residual whitens the speech signal, and – in ideal conditions – represents the excitation signal.

In a similar fashion, LP of the reverberant speech can be written as,

$$x(n) = - \sum_{k=1}^p h_k x(n-k) + e_x(n) \quad (3.4)$$

where  $e_x(n)$  is the LP residual for the reverberant speech. As reverberation mainly affects the excitation signal, it can be removed by modifying the LP residual in a manner to achieve  $e_x(n) = u(n)$ , and then the clean speech signal can be synthesized from the *cleansed* residual.

### 3.2 Maximum Kurtosis based Dereverberation

In this section the maximum kurtosis based blind dereverberation is discussed. The basic idea is to maximize the kurtosis of LP residual of received reverberant signal to achieve dereverberation. The concept stems from the fact that the LP residual of a speech signal closely approximate the glottal excitation signal, and hence, it has quasi-periodic peaks. These peaks spread in time if reverberation is present/increased, and hence reverberation causes the LP residual of speech to become *less peaky*. Recall from Section 2.5 that kurtosis is a measure of the *peakedness* of a signal. Hence, the kurtosis of the LP residual of a speech signal increases as the reverberation in speech increases. An experimental proof of the same is presented later in Section 4.1.

In [4] Gillispie *et al.* present an adaptive algorithm to maximize the kurtosis of LP residuals. In the steepest-ascent algorithm, the cost function is given as the normalized kurtosis, as in Eq. 2.5. The block diagram for the algorithm is given in Fig. 3.1.

The adaptive filter  $\mathbf{h}(n)$  is controlled by the feedback function  $f(n)$  given by the chosen cost function (described later). And the filtered LP residual  $\tilde{y}(n)$  so achieved is used to synthesize the dereverberated signal  $y(n)$ . An important assumption is made here, that the predictor coefficients obtained from the LP analysis are unaffected by the reverberation, and can be used to synthesize the clean speech from the filtered residual. This may not be true always. Hence, a secondary approach would be to duplicate the adaptive filter coefficients to directly filter the reverberant signal to get the dereverberated speech, as

illustrated in Fig. 3.2.

To derive the adaptation equations, we want to maximize the kurtosis of  $\tilde{y}(n)$ , given by

$$J(n) = \frac{E[\tilde{y}^4(n)]}{E^2[\tilde{y}^2(n)]} - 3 \quad (3.5)$$

which constitutes our cost function. The gradient of  $J(n)$  with respect to current filter is

$$\frac{\delta J}{\delta \mathbf{h}} = \left( 4\tilde{y} \frac{E[\tilde{y}^2]\tilde{y}^2 - E[\tilde{y}^4]}{E^3[\tilde{y}^2]} \right) \tilde{\mathbf{x}} = f(n)\tilde{\mathbf{x}}(n) \quad (3.6)$$

$$\text{and hence, } f(n) = 4\tilde{y} \frac{E[\tilde{y}^2]\tilde{y}^2 - E[\tilde{y}^4]}{E^3[\tilde{y}^2]} \quad (3.7)$$

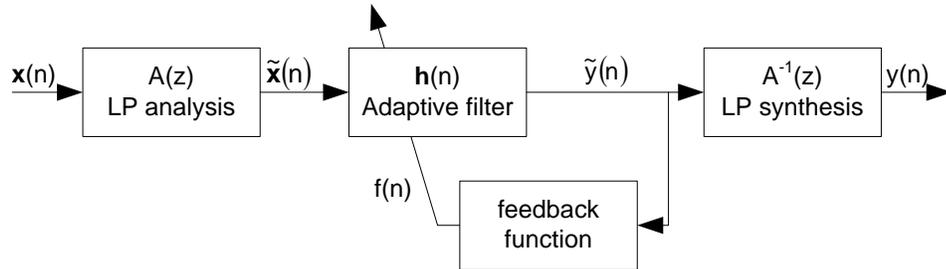
where  $f(n)$  is the desired feedback function used to control the filter update. The update equation can be written as:

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu f(n)\tilde{\mathbf{x}}(n) \quad (3.8)$$

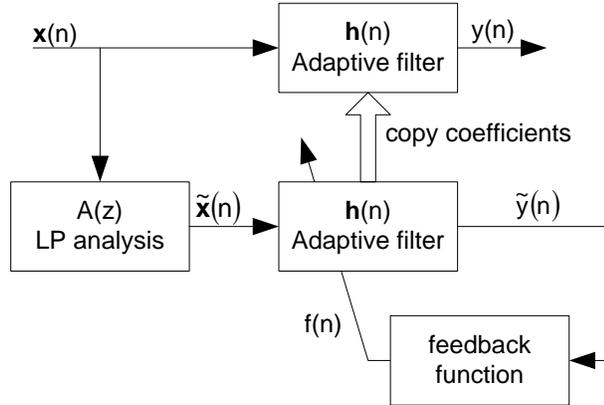
where  $\mu$  is the step-size. The expected values can be calculated recursively, as following:

$$\begin{aligned} E[\tilde{y}^2(n)] &= \beta E[\tilde{y}^2(n-1)] + (1-\beta)\tilde{y}^2(n) \\ E[\tilde{y}^4(n)] &= \beta E[\tilde{y}^4(n-1)] + (1-\beta)\tilde{y}^4(n) \end{aligned} \quad (3.9)$$

The parameter  $\beta$  is the weighing factor in the recursive update, and controls the smoothness of the moment estimates.



**Fig. 3.1** Block diagram of steepest ascent algorithm used for dereverberation.



**Fig. 3.2** Modified block diagram to avoid LP artifacts in signal reconstruction from residual.

### 3.3 Complexity of the Algorithm

In this section, the computation complexity of the algorithm is discussed. Before the analysis of the algorithm is made it is noted that this could be optimized, *e.g.*, by using pre-calculations of often used variables, using look-up tables, and optimizing with regards to parallelism, because the algorithm utilize summations which can be calculated in parallel. Execution speed is critical because the application is a hearing aid, where real time execution is required. Using parallel computation will help increase the execution speed significantly. It will, however, require a processor which is capable of performing the calculations in parallel.

The kurtosis maximization algorithm can be divided into four separate calculations defined in Eq. 3.7, 3.8, and 3.9. Furthermore it is also necessary to calculate the output signal  $y(n)$  of the filter once per algorithm update. In the following the number of multiplications required to compute each of the equations is determined.

The filtering resulting in  $y(n)$  is defined as

$$y(n) = \mathbf{h}^T \mathbf{x} \quad (3.10)$$

The number of taps in the filter  $\mathbf{h}(n)$  is equal to  $L$  and therefore Eq.3.10 requires  $L$  multiplications to be computed. The number of required additions is not used in this simple cost analysis.

The filter update equation, given in Eq.3.8, requires one multiplication in scaling (mul-

tiplication by  $\mu$ ) the feedback  $f(n)$ , and multiplying this result with the input vector  $\mathbf{x}(n)$  requires  $L$  multiplications because length of  $\mathbf{x}(n)$  is  $L$ , thus total  $L + 1$  computations.

The feedback function is given in Eq.3.7. Squaring of  $\tilde{y}(n)$  requires one multiplication, and because the multiplication with the constant 4 can be included into the step size  $\mu$  in Eq.3.8, the nominator can be calculated using three multiplications. The denominator requires two multiplications, and because it is assumed that a division requires the same number of cycles as a multiplication, even though it is a rough approximation, the total number of multiplications is six for this equation.

The first expectation operation,  $E[\tilde{y}^2(n)]$  defined in Eq.3.9 requires two multiplications, because the squaring of  $\tilde{y}$  has already been made in Eq.3.7, and the latter estimate  $E[\tilde{y}^2(n)]$  requires three multiplications, because  $\tilde{y}^4(n) = \tilde{y}^2(n) \cdot \tilde{y}^2(n)$ , where  $\tilde{y}^2(n)$  is known.

Hence, total number of calculation required to once update the filter =  $L + (L + 1) + 6 + 2 + 3 = 2L + 12$ . Using the O-notation the complexity is  $O(L)$ .

This concludes the chapter, and next we will look at some experiments and results.

# Chapter 4

## Experiments and Results

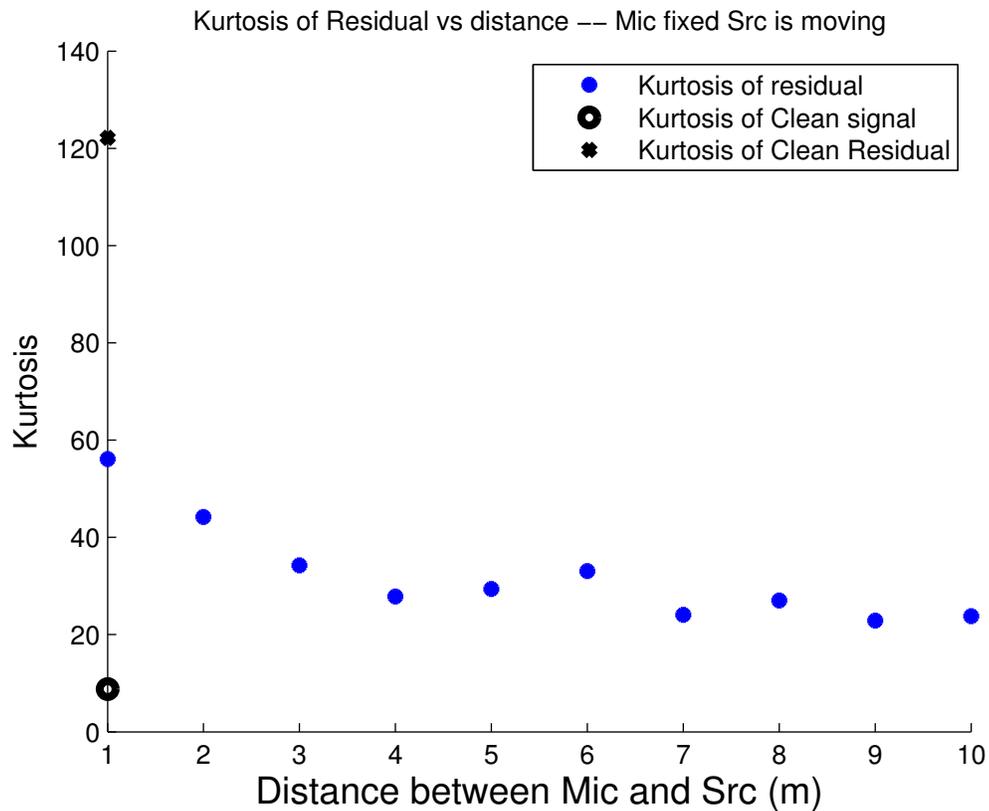
In this chapter, experimental setup and various results are discussed.

### 4.1 Kurtosis and LP residual of reverberant speech

To verify that the kurtosis of the LP residual of a speech signal decreases with reverberation, a room environment was simulated, using the algorithm proposed in [8], with following details.

- Dimensions:  $4 \times 13 \times 4$  in  $m^3$ .
- Mic Position: At  $[2,2,2]$
- Source Position: Moving from  $[2,3,2]$  to  $[2,12,2]$
- Reverberation time,  $T_{60} = 0.4$  seconds.

A clean speech of 8000 Hz sampling frequency was convolved with the RIR. The kurtosis of the LP residual of this output was calculated, and plotted against the distance between source and mic, as depicted in Fig.4.1. As the distance between the source and mic increases, the reverberation in the received speech increases, and the kurtosis of its residual decreases. This is evident from the figure. It should be noted that one should expect similar results if the source was fixed at  $[2,2,2]$ , and the mic was moving. Furthermore, it can be seen from the figure that the kurtosis of LP residual of clean speech is very high as compared to that of the reverberated residual, and that the kurtosis of actual signal is not very high.



**Fig. 4.1** Kurtosis vs distance between the source and microphone.

This small experiment establishes that the kurtosis of LP residual decreases with increase in reverberation.

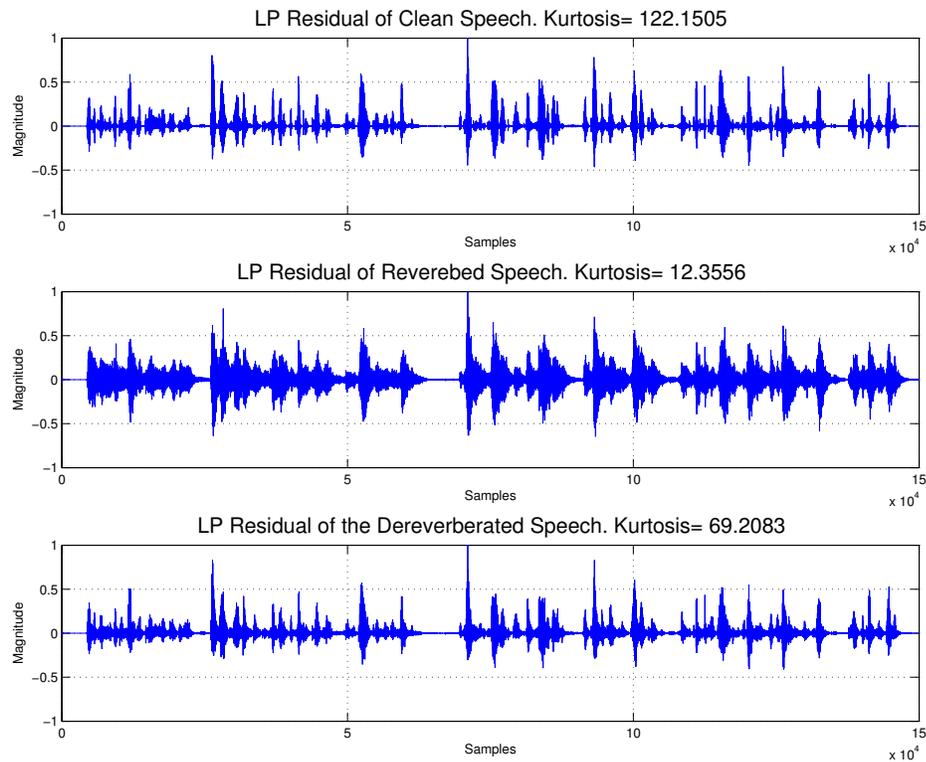
## 4.2 Dereverberation Experiments

In this section, three dereverberation experiments are discussed, where a separate reverberant room impulse response is simulated for each experiment. For all the experiments below  $\beta = 0.9$  was used. For LP residual a hamming window of size 256, and filter-tap length of 20 was used.

### 4.2.1 Experiment 1

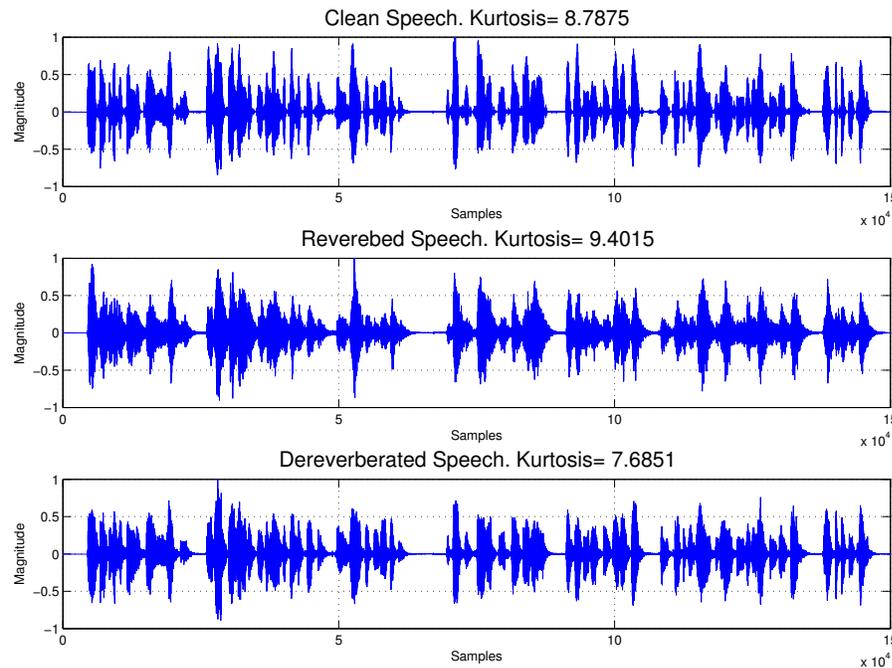
In the first experiment, a room of  $4 \times 4 \times 4$  meter cube dimension was considered, with reverberant time,  $T_{60} = 0.7$  seconds. A clean speech sampled at 8000 Hz was taken to be originating from a source located at  $[2,2,2]$ , and the microphone was situated as  $[2,3,2]$ . To understand the performance of the algorithm, waveforms of the signals, as well as LP residuals were plotted. Spectrogram was analyzed using the “wave-surfer” software. Furthermore, many people were asked to rate the improvement in reverberation for subjective measures. The results are discussed below.

- The LP residuals of the clean as well as reverberant, and then the dereverberated output were calculated. The waveforms are depicted in Fig. 4.2.



**Fig. 4.2** Waveforms of LP residuals of the clean, reverberant and dereverberated speech - Exp 1.

- To visualize the effect of reverberation in the speech signal, the waveforms of the clean, reverberant, and processed speech were plotted, as illustrated in Fig. 4.3.



**Fig. 4.3** Waveforms of clean, reverberant and processed speech - Exp 1.

- In all the figures above, the kurtosis of the respective signal has also been marked above its waveform. It is apparent from the plots that reverberation spreads the signal energy in time making the LP residual less peaky; the kurtosis values of different LP residuals also align with the results. It is also evident that the LP residual of the filtered output is much closer to that of the clean speech, and the kurtosis value has also increased.
- Spectrogram of all the three waveforms were plotted using “wave-surfer” software. The same are shown in Fig. 4.4. It can be seen in the figure how different formants are spread in time in the reverberant speech, and that the filtering process produces a much cleaner picture for the dereverberant speech.
- **Subjective Feedback:** When listened to the reverberant and processed speech, the

output was perceptibly better than the input, in terms of significant mitigation in reverberation.

### 4.2.2 Experiment 2

In this experiment, the case of blind dereverberation is tried. A reverberant speech from ITU-T Wideband database was taken, for which no clean signal or room information was available; the signal was sampled at 16000 Hz. To understand the performance of the algorithm, waveforms of the signals, as well as LP residuals were plotted; and many people were asked to rate the improvement in reverberation for subjective measures. The results are discussed below.

- The LP residuals of the reverberant, and then the dereverberated are depicted in Fig. 4.5.
- On the same lines as of experiment - 1, the waveforms of the reverberant, and processed speech were plotted, as illustrated in Fig. 4.6.
- Again, the kurtosis of the respective signal has also been marked above its waveform. It is apparent from the plots that the LP residual of the filtered output is more peaky; the kurtosis values also increase with decrease in reverberation.
- **Subjective Feedback:** When listened to the reverberant and processed speech, the output was perceptibly better than the input, in terms of significant mitigation in reverberation.

### 4.2.3 Experiment 3

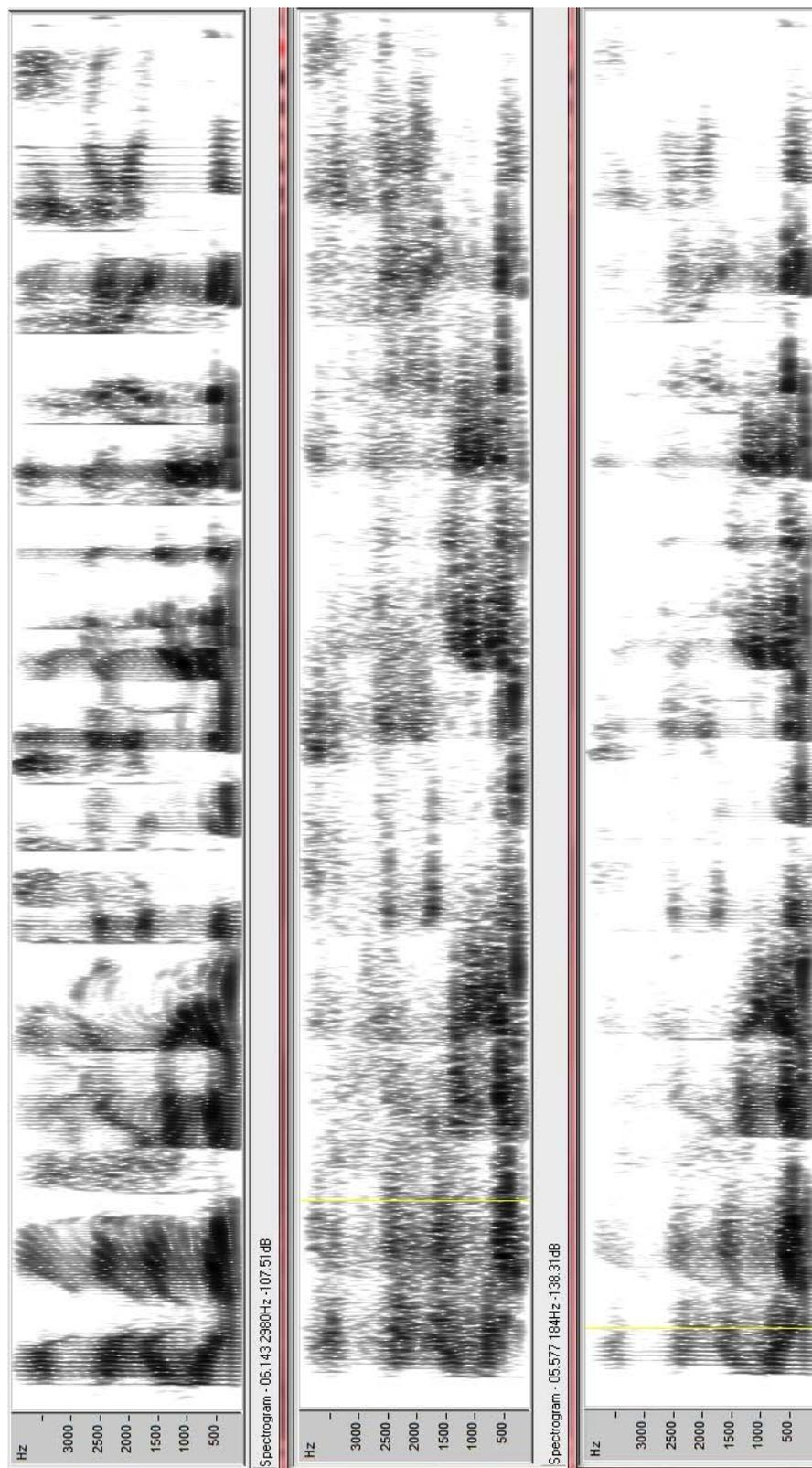
In the aforementioned two experiments the algorithm was used as an offline algorithm, that is, first the final set of filter coefficients was calculated by letting the simulation run for whole length of the residual signal, and then the reverberant speech was filtered through the adaptive filter, using the derived coefficients. However, the primary use of dereverberation algorithms are in real-time scenarios, and hence, for this experiment the filter was applied on-the-fly to the reverberant speech, and a gradual improvement in the quality of speech was noticed. For the purpose of the experiment, same clean speech signal as in Experiment-1 was used here, but for reverberation a real life RIR downloaded from

Aachen Impulse Response database [10] was used. To understand the performance of the algorithm, waveforms of the signals, as well as LP residuals were plotted. The results are discussed below.

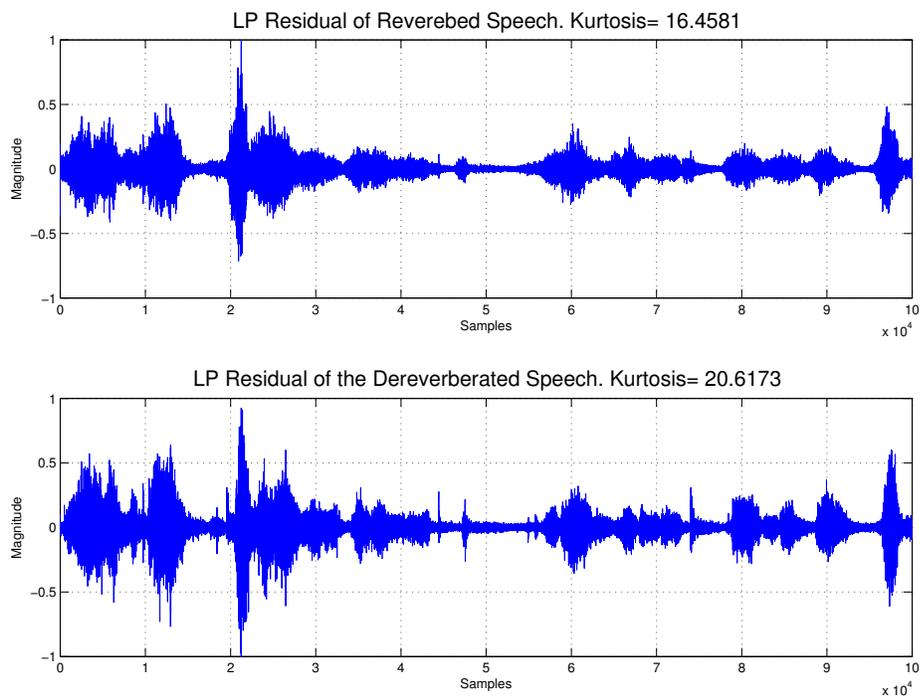
- The LP residuals of the reverberant, and then the dereverberated output are depicted in Fig. 4.7.
- On the same lines as of experiment - 1, the waveforms of the reverberant, and processed speech were plotted, as illustrated in Fig. 4.8.
- Again, the kurtosis of the respective signal has also been marked above its waveform. It is apparent from the plots that the LP residual of the filtered output is more peaky; the kurtosis values also increase with decrease in reverberation.
- **Subjective Feedback:** When listened to the reverberant and processed speech, the output was perceptibly better than the input, in terms of significant mitigation in reverberation. One could hear gradual improvement in the signal quality, as the algorithm learns and adapts the filter coefficients with time; here a gradual improvement in the quality of speech was noticed.

### 4.3 Summary

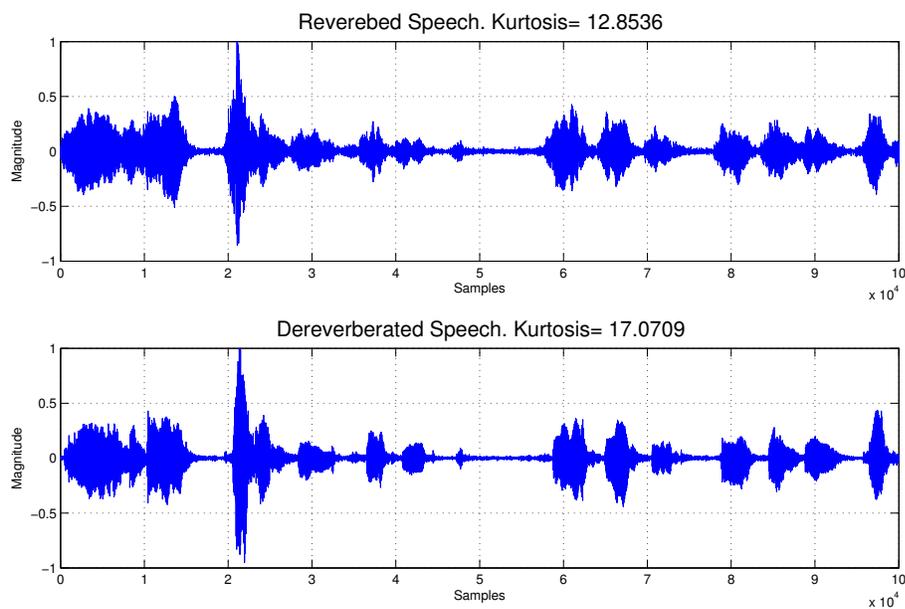
It was proved that the kurtosis of LP residual is a good measure of reverberation in speech. Various experiments were used to validate that the algorithm works well for synthetic as well as real life RIR. Finally experiment 3 proved the algorithm to be suitable for real time scenarios.



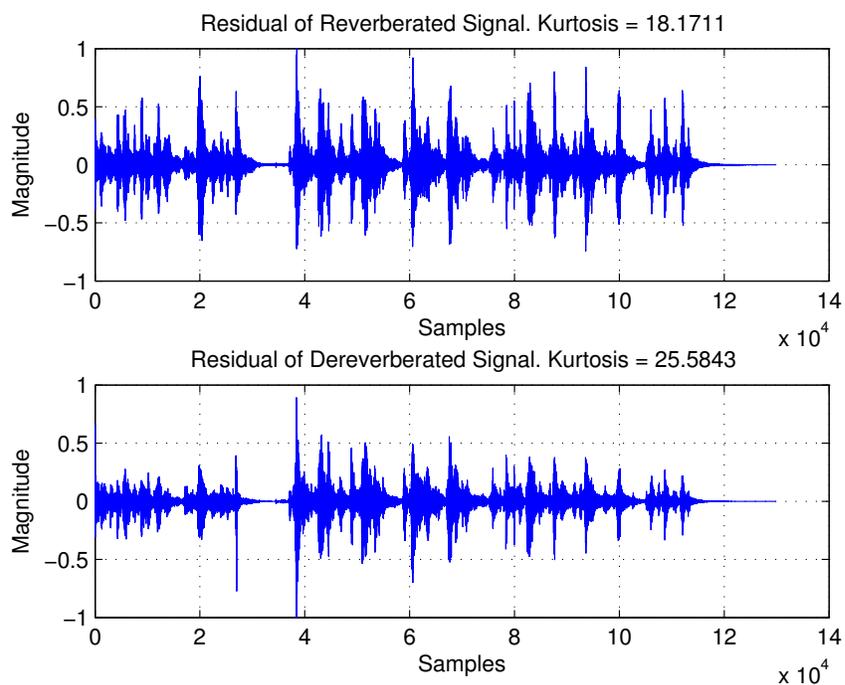
**Fig. 4.4** Spectrogram of clean, reverberant and processed speech - Exp 1.



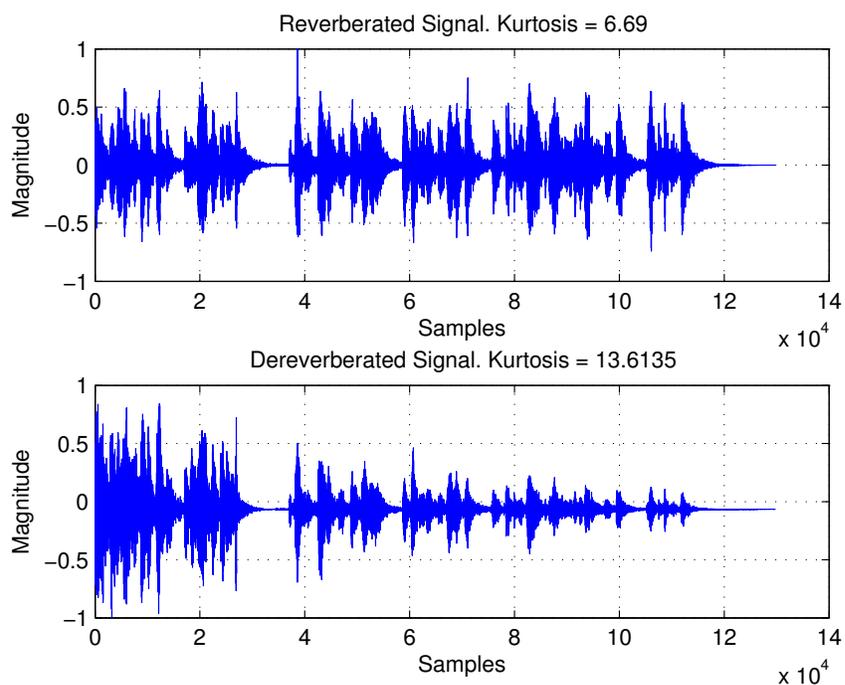
**Fig. 4.5** Waveforms of LP residuals of the reverberant and dereverberated speech - Exp 2.



**Fig. 4.6** Waveforms of clean, reverberant and processed speech - Exp 2.



**Fig. 4.7** Waveforms of LP residuals of the reverberant and dereverberated speech - Exp 3.



**Fig. 4.8** Waveforms of clean, reverberant and processed speech - Exp 3.

## Chapter 5

# Discussion and Conclusion

In this project maximum kurtosis based approaches to the blind speech dereverberation problem was analysed. The applications targeted in this report are hearing aid, and hands-free telephony systems, with a single microphone setup. The method was tested on simulated and real life RIR's with real voiced speech.

The implemented algorithm is based on the work by Gillespie et al., [4], who presented a steepest-ascent solution for the problem. The source signal (the direct speech),  $\mathbf{s}$ , is assumed to be non-Gaussian distributed. The room then causes the observed signal (the reverberated speech),  $\mathbf{x}$ , to be corrupted by delayed and weighted versions of  $\mathbf{s}$ , and introduces Gaussian distributed components. A Gaussian distributed signal has higher order cumulants equal to zero, and the approach is therefore to maximize the kurtosis of the observed signal, such that the room impact on the source signal can be removed.

In conclusion, an interesting method to blind speech dereverberation has been implemented and analysed in detail. Adjustments have been made on carefully performed tests and the final word is that the algorithm is able to significantly reduce the undesired reverberation in speech.

### Future Work

- Authors in [12] claim that an average over several spatially distributed microphones can provide potentially better results, hence, it may be worthwhile to investigate into multi-microphone setups.

- Time domain implementation is prone to slow or no convergence because of all variance in the eigenvectors of autocorrelation matrices of the input signal. Alternate noise robust implementations may be investigated to avoid this issue, such as the subband adaptive method promoted in [4].

## References

- [1] B. Yegnanarayana and P. S. Murthy, “Enhancement of of reverberant speech using LP residual Signal,” *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 267–281, May 2000.
- [2] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of Statistical Learning*. Springer-Verlag, 2001.
- [3] O. Tanrikulu and A. G. Constantinides, “Least-mean Kurtosis: a novel higher-order statistics based adaptive filtering algorithm,” *Electronic Letters*, vol. 30, no. 3, pp. 189–190, Feb 3, 1994.
- [4] B. W. Gillespie, D. A. F. Florencio, and H. S. Malvar, “Speech dereverberation via maximum-kurtosis subband adaptive filtering,” in *ICASSP '01: IEEE International Conference on Acoustics, Speech, and Signal Processing 2001*, vol. 6, pp. 3701–3704, IEEE.
- [5] E. A. P. Habets, N. D. Gaubitch, and P. A. Naylor, “Temporal selective dereverberation of noisy speech using one microphone,” in *ICASSP '08: IEEE International Conference on Acoustics, Speech, and Signal Processing 2008*, p. 45774580, IEEE.
- [6] M. Jeub and P. Vary, “Dereverberation of Speech Signals Based on the Discrete Model of Speech Production,” *20. Konferenz Elektronische Sprachsignalverarbeitung (ESSV) of Studentente zur Sprachkommunikation*, vol. 53, pp. 129–134, Sept 2009.
- [7] M. Wu and D. Wang, “A Two-stage Algorithm for Enhancement of Reverberant Speech,” in *ICASSP '05: IEEE International Conference on Acoustics, Speech, and Signal Processing 2005*, vol. 1, pp. 1085–1088, IEEE.
- [8] E. Lehmann and A. Johansson, “Prediction of energy decay in room impulse responses simulated with an image-source model,” *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, July 2008.
- [9] J. Allen and D. Berkley, “Image method for efficiently simulating small-room acoustics,” *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, April 1979.

- 
- [10] M. Jeub, M. Schaxfer, and P. Vary, “A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms,” in *16th International Conference on Digital Signal Processing (DSP)*, Santorini, Greece, 2009.
  - [11] C. Nikias and J. Mendel, “Signal Processing with higher-order spectra,” *IEEE Signal Processing Magazine*, vol. 10, pp. 10–37, July 1993.
  - [12] N. D. Gaubitch, P. Naylor, and D. Ward, “On the use of Linear Prediction for Dereverberation of Speech,” in *Intl. Workshop on Acoustic Echo and Noise Cancellation*, Kyoto, Japan, 2003.

---

# Appendix A

## Matlab Codes

This section enlists various MATLAB codes used in this project.

### Effect of reverberation on kurtosis

Here the algorithm to compare the effect of increasing reverberation on the kurtosis of LP residual is given, as discussed in Section 4.1. Note that this code requires two external algorithms:

- Algorithm for calculating LP residual of speech. The algorithm is excluded because it is very simple to implement.
- Algorithm to simulate the required room environment. The original codes were downloaded from following link, which were then modified to fit the project's need.

URL = <http://www.eric-lehmann.com/ism.bg.html>

```
1 function kurt_compare
  % Compares the kurtosis of the LP residual for moving Src and Mic
3
  signal='speech_data/Room_clean';
5 xc= wavread(signal);
  rxc = LPres (xc, hamming(256), 128, 20); % Calculates LP residual
7
  dist = zeros(N,1); % Distance between src and mic
9 kurt = zeros(N,2); % Two vectors to store different kurtosis values
11 %% Define Room
  room.Fs = 8000;
```

```
13 room.dim=[4 13 4];
    room.mic_pos= mic;
15 room.src_traj=src;
    room.T60=0.7 ;
17
19 mic = 0;
    src = 0;
21
    %% Moving SRC and MIC
23
    for i = 1: N
25         i
            if i < 11
27                 src = [2 2 2];
                    mic =[2 2+i 2];
29
                    myISM_setup(mic, src); %% This calls the modified ISM)setup function
                        from Lehman-Johansson's image-source method based room simulator
31                    ISM_RIR_bank(myISM_setup(mic, src),strcat('ISM_RIRs_mic',num2str(i),'.
                        .mat'));
                    x = ISM_AudioData(strcat('ISM_RIRs_mic',num2str(i),'.mat'),xc);
33
                    r = LPres (x, hamming(256), 128, 20);
35                    kurt(i,1)= kurtosis(x);
                    kurt(i,2)= kurtosis(r);
37                    dist(i) = norm(mic-src);
            else
39                    mic = [2 2 2];
                    src =[2 2+i-10 2];
41
                    myISM_setup(mic, src)
43                    ISM_RIR_bank(myISM_setup(mic, src),strcat('ISM_RIRs_src',num2str(i),'.
                        mat'))
                    x = ISM_AudioData(strcat('ISM_RIRs_src',num2str(i),'.mat'),xc);
45
                    r = LPres (x, hamming(256), 128, 20);
47                    kurt(i,1)= kurtosis(x);
                    kurt(i,2)= kurtosis(r);
```

```
49         dist(i) = norm(mic-src);
        end
51 end

53 %% Plot Graphs

55 figure
    hold on
57 plot(dist(1:10), kurt(1:10,1), 'x', 'linewidth',2);
    title('Kurtosis of x vs distance — Src fixed Mic is moving')
59 plot(kurtosis(xc), 'ko', 'linewidth',3)
    plot(kurtosis(rxc), 'kx', 'linewidth',3)
61 legend('Kurtosis of x', 'Kurtosis of Clean signal', 'Kurtosis of Clean
        Residual')
    hold off
63

65 figure
    hold on
67 plot(dist(11:20), kurt(11:20,1), 'x', 'linewidth',2);
    title('Kurtosis of x vs distance — Mic fixed Src is moving')
69 plot(kurtosis(xc), 'ko', 'linewidth',3)
    plot(kurtosis(rxc), 'kx', 'linewidth',3)
71 legend('Kurtosis of x', 'Kurtosis of Clean signal', 'Kurtosis of Clean
        Residual')
    hold off
73

75 figure
    hold on
77 plot(dist(1:10), kurt(1:10,2), '*', 'linewidth',2);
    title('Kurtosis of Residual vs distance — Src fixed Mic is moving')
79 plot(kurtosis(xc), 'ko', 'linewidth',3)
    plot(kurtosis(rxc), 'kx', 'linewidth',3)
81 legend('Kurtosis of x', 'Kurtosis of Clean signal', 'Kurtosis of Clean
        Residual')
    hold off
83
    figure ,
```

```

85 hold on
   plot(dist(11:20), kurt(11:20,2), '*','linewidth',2);
87 title('Kurtosis of Residual vs distance — Mic fixed Src is moving')
   plot(kurtosis(xc), 'ko', 'linewidth',3)
89 plot(kurtosis(rxc), 'kx', 'linewidth',3)
   legend('Kurtosis of x', 'Kurtosis of Clean signal', 'Kurtosis of Clean
         Residual')
91 hold off

```

Listing A.1 Matlab code for comparing kurtosis of varying reverberation

## Kurtosis Maximizing algorithm

This section provides the main algorithm utilized for dereverberation in this project, as discussed in Section 3.2. This also requires the LPres.m code to calculate the LP-residual of the speech signal.

```

1 function w = MaxKurto(signal)
   % signal is the "path" to the speech file
3
   [x, FS, NBITS]= wavread(signal); % Reading the speech
5 x_drvb = zeros(length(x),1); %x_drvb would be the filter output

7 sysorder = 20; %Filter-tap length

9 %% Calculate the LP residual
   r = LPres(x, hamming(256), 128, sysorder);
11 % r = LPres(x, ones(sysorder,1), 128, sysorder);

13 %% LMK
   w = ones(sysorder,1)*0.05; % w are the filter-tap coefficients. It is a near
       Identity filter
15 w(1)=1;
   f = zeros(length(r),1); % The feedback function
17
   beta= 0.9; %Smoothness of kurtosis
19
   y = zeros(length(x),1); % output LP residual
21 ey2 = 0; %Expected value of y^2

```

```

ey4 = 0; %Expected value of y^4
23 mu= 0.000388 ; %Step-size

25 for n = 1: length(r)-sysorder+1
    u = r(n:n+sysorder-1);
27 y(n)= w' * u;
    ey2 = beta*ey2 + (1-beta)*y(n)^2; %Expected value of y^2
29 ey4 = beta*ey4 + (1-beta)*y(n)^4; %Expected value of y^4
    f(n)= 4*(ey2*y(n)^2 - ey4)*y(n)/(ey2^3);
31
    w = w + mu*f(n)*u; % updating filter coefficients
33 %% For online algorithm
    v = x(n : n + sysorder - 1);
35 x_drvb(n) = w(:,n+1)' * v;
end
37
%% If not an online algorithm
39 parfor n = 1 : length(x)-sysorder + 1 % Parellelized for loop for better
    performance
        u = x(n : n + sysorder - 1);
41 x_drvb(n) = w' * u;
end
43
%% Writing the .wav file
45 wavwrite(x_drvb,FS,NBITS, strcat(signal(1:end-6), '_ ',room(5:end-4), '_output',
    '.wav'));
sound(x_drvb, FS);
47 wavwrite(x_drvb,FS,NBITS, 'speech_data/xDRoom_meeting-0-0-1.wav');

49 %% Plots

51 hold on
subplot(211)
53 plot(x/max(abs(x))); grid
xlabel('Samples')
55 ylabel('Magnitude')
k=kurtosis(x);
57 title(['Reverberated Signal. Kurtosis = ' num2str(k)])

```

```
59 subplot(212)
   plot(x_drvb/max(abs(x_drvb))); grid
61 xlabel('Samples')
   ylabel('Magnitude')
63 k=kurtosis(x_drvb);
   title(['Dereverberated Signal. Kurtosis = ' num2str(k)])
65
67 rd = LPres(x_drvb, ones(256,1), 128, sysorder); % To calculate the LP
   residual of the filter output
69 figure
   hold on
71 subplot(211)
   plot(r/max(abs(r))); grid
73 xlabel('Samples')
   ylabel('Magnitude')
75 k=kurtosis(r);
   title(['Residual of Reverberated Signal. Kurtosis = ' num2str(k)])
77 subplot(212)
   plot(rd/max(abs(rd))); grid
79 xlabel('Samples')
   ylabel('Magnitude')
81 k=kurtosis(rd);
   title(['Residual of Dereverberated Signal. Kurtosis = ' num2str(k)])
```

Listing A.2 Matlab code for the kurtosis maximizing algorithm