# THE IN-SERVICE DETERMINATION OF THE PRESENCE OF DISTORTION IN A HIGH QUALITY ANALOGUE SOUND SIGNAL

**Stefanus Maré Pr Eng**

**Thesis submitted in compliance with the requirements for the Doctor's Degree in Technology: Engineering: Electrical in the Department of Electronic Engineering at the Durban University of Technology, Mansfield Road, Durban**

**Approved for final submission**

**Date:**

**Supervisor: Professor Vladimir Bajić Pr. Eng., D Eng Sc (EE)**

**Date:**

**Co-supervisor: Professor Kevin Duffy, PhD**

**Declaration:**

**I declare that this thesis is my own work both in conception and execution, except where due acknowledgement is made to others, and has not been submitted previously for any other degree or examination.**

**Stefanus Maré Pr Eng**

## ABSTRACT

Detecting and minimising distortion in audio signals is an important aspect of sound engineering. Distortion of a signal passing through an audio system may be caused by a number of factors and it is necessary to detect these effects for optimal sound. The problem is of interest to users and operators of high quality audio equipment and transmission facilities.

The objective of this thesis was the development of techniques for the blind identification of distortion in a high quality audio signal using digital signal processing techniques. The techniques developed are based on digital signal processing techniques and statistical analysis of a recorded audio signal, which is treated as a random, non-stationary signal.

It has been demonstrated that the following distortions and properties of high quality audio signals can be identified using blind techniques:

a) mechanical defects in recording and playback apparatus leading to wow and flutter in the audio signal ,
b) frequency content and dynamic range attributes of the audio signal, and
c) non-linear distortion of the audio signal.

## Contents

# GLOSSARY OF TERMS, ACRONYMS AND NOTATION

## TERMINOLOGY

| | |
|---|---|
| **Audio signal** | Electrical signal representing an audible sound |
| **Impairment** | An audible distortion |
| **Distortion** | A change to the audio signal envelope |
| **Codec** | Encoder-decoder: Equipment used to convert an analogue signal to a digital signal (and vice versa), while removing redundancy to limit the required transmission bandwidth. |
| **Wow** | Distortion caused by low frequency (<10 Hz) speed variations in the playback or recording apparutus. |
| **Flutter** | Distortion caused by high frequency (10 Hz to 100 Hz) speed variations in playback or recording apparatus |

## ACRONYMS

| | |
|---|---|
| **CSAD** | Cumulative spectral amplitude distribution |
| **DFT** | Discrete Fourier transform |
| **FFT** | Fast Fourier transform |
| **pdf** | Probability density function |

## NOTATION

| | |
|---|---|
| $x(t)$ | Time domain input signal to a system |
| $X(f)$ | Frequency domain input signal to a system |
| $y(t)$ | Time domain output signal from a system |

| | |
|---|---|
| ***Y(f)*** | Frequency domain output signal from a system |
| ***$R_{xx}(\tau)$*** | Autocorrelation of *x(t)* |
| | $= lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)x(t+\tau)dt$ |
| ***$S_{xx}(f)$*** | Power spectrum, the Fourier transform of $R_{xx}(\tau)$ |
| ***$R_{xy}(\tau)$*** | Cross correlation between *x(t)* and *y(t)* |
| | $= lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)y(t+\tau)d\tau$ |
| ***$S_{xy}(f)$*** | Cross spectrum, the Fourier transform of $R_{xy}(\tau)$ |
| ***h(t)*** | The system impulse response function |
| ***H(f)*** | The system frequency response function |
| ***E[ ]*** | The expectation operator |
| | $= \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)dt = \int_{-\infty}^{\infty} xp(x)dx$ |

# CHAPTER 1

## INTRODUCTION

### 1.0 The aim of the research

Within a broadcast studio facility, an audio transmission network or any similar system for the recording, transmission or production of audio signals, control of audio signal quality, i.e. maintaining the listener's perception that the signal is not impaired in any way, is usually done by ensuring that the sound production and transmission equipment operates within acceptable limits. This is achieved through a program of preventative maintenance. In a situation where the equipment is in use 24 hours a day preventative maintenance means having duplicate equipment available, adding to the cost of running a facility.

This research aims to develop techniques whereby the presence of distortion in a high quality audio signal can be detected through the use of blind (or equivalently non-intrusive) digital signal processing techniques. Such techniques will enable the operator of a broadcast facility to continuously, electronically monitor the quality of the audio signal produced in the facility and to take remedial steps should the quality be compromised. In other words the purpose of this research is to investigate the possibility of creating techniques, which could mimic the performance of a human listener in the evaluation of the quality of an audio signal. In a different context Scheirer [Scheirer, 1998] in his PhD proposal summarized the problem of machine perception as follows:' Consider a musically unskilled listener turning on a radio and hearing five seconds of sound from the middle of a previously unheard piece of music. I want to build a computer system that can make the same judgements about this piece of music as a human listener can.' Scheirer's emphasis was on the 'discovery of musical-object-formation heuristics, the analysis of musical-object similarity' and on 'musical pattern recognition systems'. The objective of this research is to identify distortion in an audio signal.

Impairments are those changes (distortions) to an audio signal, which create a subjectively observable change. In other words, an impairment is a signal distortion that can be perceived by a human listener. An impairment may affect the amplitude, phase or spectral content of the signal. Distortion, while it may be quantifiable through measurement, is not of concern to a human listener unless the listener can perceive the distortion as an impairment to the sound. In this text a distinction will be made between an *impairment,* a distortion which can be perceived by a human listener and a *distortion* which may change a signal in some way but which is not necessarily recognized as such by a human listener.

A distortion can be measured using test and measurement techniques, an impairment cannot be measured, it is a subjective assessment made by a human listener. The objective distortion measurement and the subjective impairment are related through extensive listening tests to determine what level of distortion is acceptable in terms of the observed impairment. A lower limit of distortion is then established which will result in acceptable performance.

A distortion, resulting from equipment malfunction, and the resulting impairment in an audio signal will be detected by a listener who will set in motion the procedures for correcting the malfunction. Preventative maintenance in a sound recording studio or broadcast production studio is aimed at preventing equipment malfunctioning and maintaining an impairment free signal.

Blind signal processing, as used in this text, also referred to as non-intrusive assessment [Picovici and Mahdi, 2003], refers to situations where information about a system is derived from the signal at the output of the system without access to a reference signal, for example the reference may be the signal at the input to the system. In that sense this work reports on blind identification of distortion in analogue audio signals.

The problem of blind assessment of signal quality has been the subject of research in the speech communications community [Picovici and Mahdi, 2003].

## 1.1 Delimitations

The emphasis in this research is not on determining equipment malfunction, but on determining whether the signal has been distorted. Clearly equipment malfunction may lead to distortion, but inappropriate use of perfectly sound equipment may also lead to distortion, e.g. overdriving an amplifier. If the distortion can be detected then steps can be taken to trace the source of the distortion.

Where possible the magnitude of the distortion that has been detected will be related to that which would be obtained using standard measurement procedures. This is not always possible. In some cases standard measurement procedures do not exist, e.g. for the dynamic range of a signal, or the standard procedures are limited in scope, e.g. the amount of non-linear distortion depends on the input signal level. Standard measurements of non-linear distortion are done under set conditions of input signal level.

Relating the distortions that are detected to impairments, which can be heard by a human listener, requires that listening tests be carried out. This is outside the scope of this research. This research aims to develop techniques for the detection of distortion. Since the techniques are applicable to actual audio signals and not to deterministic signals it is not always possible to relate the detected distortion quantitatively to a measurement. The only measurement that can be carried out on an actual audio signal is a listening test where the listener will venture an opinion as to whether the distortion is an audible impairment to the signal.

## 1.2 Application of the research

*"In a competitive telecommunications market where price differentials have been minimized, quality of service has become very important".* [Picovici and Mahdi, undated].

This comment is also applicable to radio broadcasting, audio recording and audio transmission systems. The quality of the audio signal has to be determined in a way that does not involve taking equipment out of service unnecessarily. This can be achieved through non-intrusive quality assessment, i.e. assessment of the quality of the signal directly from the signal itself.

### 1.3 The background to the research project

Some distortion mechanisms give rise to impairments which are readily observable while others are less easily detected by a human observer. Conventional measurement techniques have been designed to control the amount of those distortions that are most troublesome and the distortions that do not give rise to readily observable impairments are not part of routine measurements.

In Chapter 2 signal distortion is discussed in detail. The concept of linear and non-linear distortion is explained, as is the difference between these two sources of distortion. Both types of distortion give rise to changes in the shape of the signal. Linear distortion leads to differential time delay and differential amplification of the spectral components that make up the signal. Non-linear distortion on the other hand results from differential changes in the amplitude of the signal dependent on the signal amplitude.

Linear distortion leading to differential time delay, i.e. non-linear phase vs frequency response, does not necessarily give rise to an audible impairment in a real audio signal [Hartmann, 1998; Koya, undated.; Preiss, 1982].

Linear distortion arising from differential amplification of frequency components, i.e. the amplitude vs frequency response of a system is not constant, results in changes to a signal that are readily audible to a listener. Such changes may be introduced to create a particular ‘sound’, for example audio signals with different high frequency content are described in subjective terms such as: “bright”, “flat”, “dull”.

Distortion arising from mechanical imperfections in recording and playback apparatus, known as wow and flutter, may not be very relevant in a post analogue age. However many recordings made on analogue equipment exist and these are still in demand by the public. Detection of the presence of wow and flutter distortion is therefore still relevant.

Dynamic range is one of the properties of an audio signal that is much used or abused but which is not measured. The relevance of dynamic range to sound quality is discussed.

Chapter 3 addresses the topic of psychoacoustics and how it relates to the perception of audio signals. Loudness perception, pitch perception and spectro-temporal masking effects are described.

Consonance and dissonance are important aspects of the human auditory system in terms of the perception of distortion as being something that sounds ‘unpleasant’. An overview of these effects forms part of Chapter 3.

Chapter 4 deals with the way audio signal distortion is conventionally measured, i.e. through stimulus response testing. The background to conventional audio quality measurements and their relationship to the subjective perception of audio quality is discussed. Conventional quality measurements are structured according to a preventative maintenance paradigm. Equipment is regularly taken out of service and after performance measurements on the equipment, using standardized test signals, the necessary adjustments are made and the equipment is put back in service. Due to the competitive nature of telecommunications services this approach may not be the most appropriate and hence blind assessment of quality has become the focus of research [Picovici and Mahdi: 2003; Jin and Kubichek: 1996; Gray et al: 2000]

Four approaches to quality determination of audio signals are mentioned and three are discussed in more detail:

a) Conventional stimulus-response approaches to controlling the quality of audio signals [Hueber et al, 1976; King, 1979; Kuni, 1997; Leinonen et al, 1977 ]

Conventional testing makes use of well defined test signals which are applied to the system under test. The output of the system is then analysed to quantify the change that has occurred to the test signal. This change is related to a subjective impression of impairment through limits that have been set for the particular test.

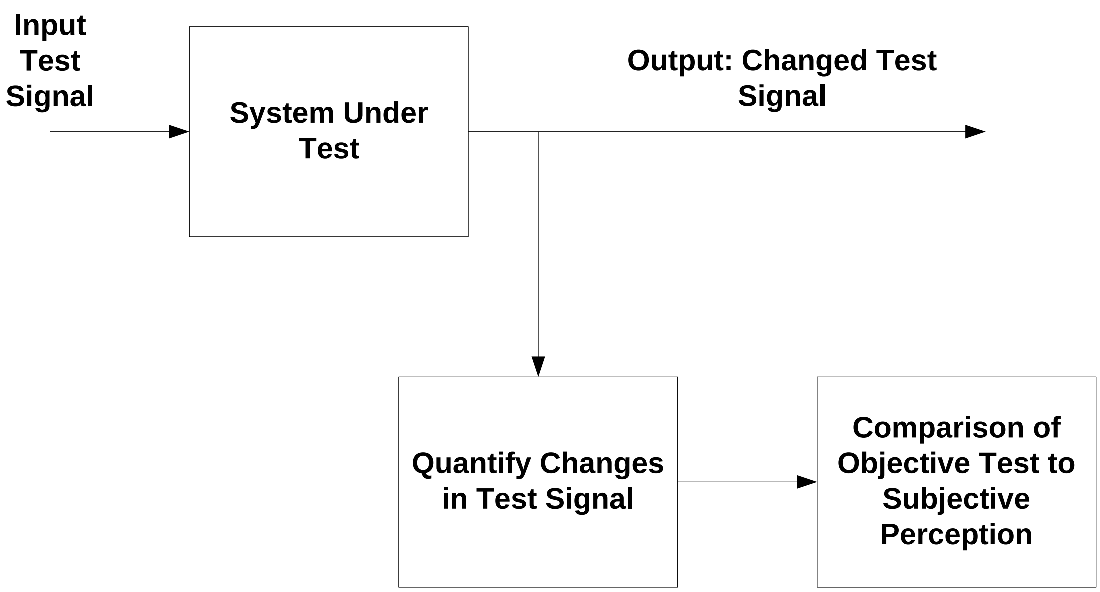


**Figure 1.1:** Conventional stimulus-response testing

b) Modeling of the human auditory system [Beerends, 1998; Beerends & Stemerdink, 1992; Hollier et al, 1995; Robinson and Hawksford, 1999; Yang et al: 1997; Yang et al: 1998 ].

In testing based on a model of the human auditory system the input test signal to the system is an audio signal that has been mapped to a perceptual dimension. The output signal is mapped to the same perceptual dimension and the differences are evaluated in terms of their audibility.

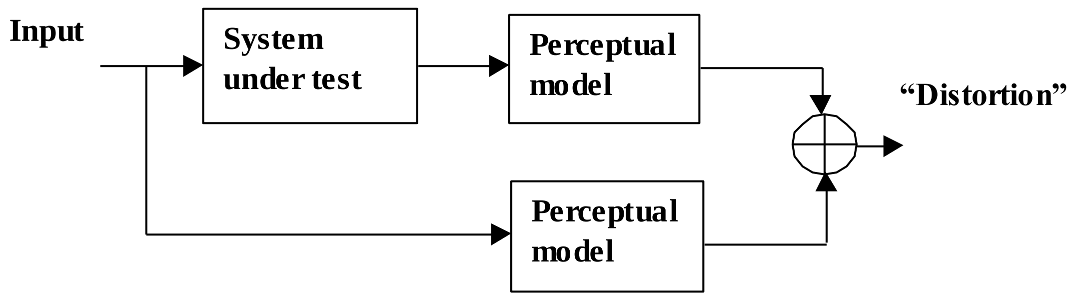


**Figure 1.2:** Perceptual model based testing

c) Coherence testing [Kates, 1992; Totzek & Preiss, 1987].

In coherence testing the input signal can be the actual audio signal. This means that the stimulus is authentic and the results relate to real service conditions.

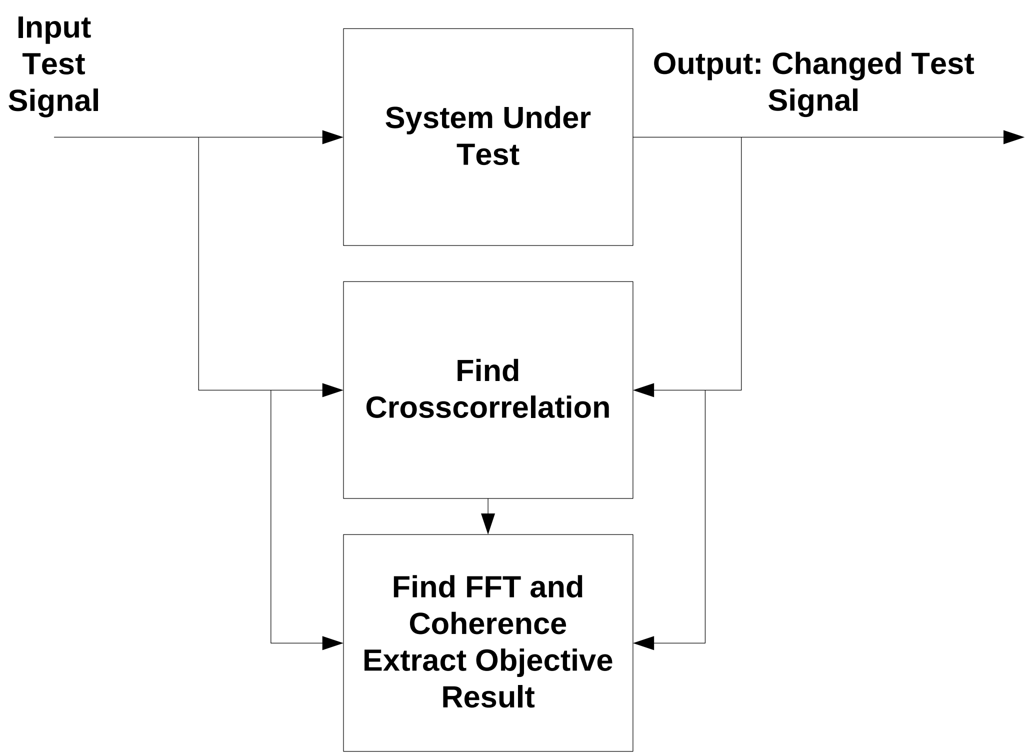


**Figure 1.3:** Coherence testing

d) Automated testing

This is not considered as a topic on its own since automated testing is based on one of the methodologies discussed above in (a), (b) and (c). Automated testing is done largely without the intervention of a human operator.

### 1.4 Overview of the research

The first of the techniques developed through this research relates to the detection of distortion resulting from imperfections in recording and playback apparatus, colloquially known as ‘wow and flutter’. Chapter 5 is devoted to theoretical analysis of the distortion caused by mechanical imperfections in apparatus and results obtained in blind detection of this distortion. A technique for the detection of ‘wow and flutter’ distortion, based on the autocorrelation of the distorted audio signal was developed and publicised.

Frequency content and dynamic range and their influence in determining the perceived quality of a signal are discussed in Chapter 6. A method for determining the frequency content, and the dynamic range as a function of frequency, is described. This method is given the name CSAD (Cumulative Spectral Amplitude Distribution).

In Chapter 7 the problem of the identification of non-linear distortion in an audio signal is addressed. Since the audio signal is frequently generated by non-linear systems, e.g. musical instruments, conventional means, like higher order spectral analysis [Hinich & Patterson, 1995; Brillinger & Irizarry, 1998], as applied to the identification of non-linearity in time series could not be used. A method based on pattern recognition using neural networks was developed and publicised.

The techniques developed are in the engineering domain rather than in the psychoacoustic domain. As such they address things like ‘wow and flutter’, dynamic range, frequency content and non-linear distortion rather than their perceptual equivalents.

# CHAPTER 2

## AUDIO SIGNAL DISTORTION

### 2.0 Background

The term ‘distortion’ as used here refers to a change in the shape (waveform envelope) of an audio signal as it passes through an audio processing system, i.e. a recording or sound production facility or a transmission system. (A transmission system can be anything from a single amplifier to a 1000 km fibre optic link.)

The signal of concern here is an audio signal, i.e. sound that has been transformed into an electrical signal. The audio signal is intended for use by human listeners and they will perceive a distortion in the audio signal as an impairment to the sound that they are hearing. Not all distortions are the causes of audible impairments.

Distortion may result from a system that is linear or one that is non-linear.

Distortion measurements relate a subjective effect to an objective measurement. Most distortion measurements are done using deterministic test signals, this is problematic since an audio signal is not deterministic, it is complex, occupies a wide frequency range and a wide dynamic range and it is not a stationary signal. Ideally distortion measurements should be done using an audio signal, or a signal that is comparable to an audio signal, and impairments should then be identified through listening tests on the audio signal.

### 2.1 Linear Distortion

*Linear distortion* changes the amplitude and/or phase relationships between the frequency components of the signal, so that the shape (waveform envelope) of the signal is changed [Preiss, 1976, Preiss, 1982]. Linear distortion is not signal amplitude dependent. It is a function of the amplitude and phase relationship of the spectral components of the signal. If the phase and amplitude of the output signal components have changed relative to the relationships that existed at the input to the system, then the signal has been distorted. The same components will be present at both the input and the output. Referring to Figure 2.1, for distortionless transmission it is required that:

$$y(t) = kx(t-T) \tag{2.1}$$

Where:

y(t) is the output signal from the system

x(t) is the input signal to the system

k is a constant amplification factor

T is a constant time delay.

This will ensure that the output waveform is a copy of the input waveform with a possible time delay and a possible change in amplitude.

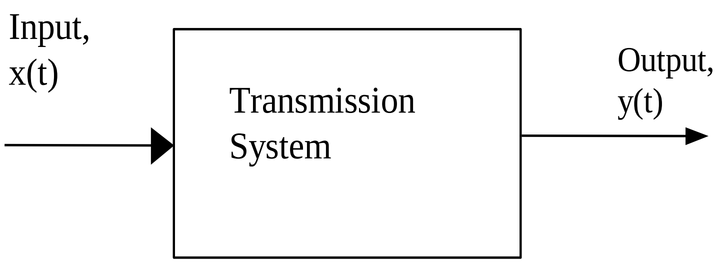


**Figure 2.1**: Signal transmission

In the frequency domain the relationship for distortionless transmission will be the Fourier transform of equation 2.1.

$$\begin{aligned} Y(f) &= \int_{-\infty}^{\infty} y(t)\, e^{-j2\pi ft} dt \\ &= kX(f)\, e^{-j2\pi fT} \end{aligned} \tag{2.2}$$

Where:

X(f) is the Fourier transform of x(t), the input signal.

Since $X(f) = |X(f)| \angle \Phi(f)$, the output spectrum will be:

$$Y(f) = k|X(f)| \angle (\Phi(f) + 2\pi f T) \qquad (2.3)$$

The signal will therefore be distorted if either or both of the following happen in passing through the system:

a) The constant, *k*, in equation 2.3 is not a constant but is a function of frequency.

b) The constant delay, *T*, in equation 2.3 is not a constant but is a function of frequency.

Both of these conditions commonly occur in transmission equipment. To cite one example, all signals are band limited by the frequency response of the equipment used in a transmission system. This implies that the frequency transfer function does not have an amplification that is constant across all frequencies, but that it is constant up to a certain limit and for frequencies higher than that limit the amplification will decrease.

The program managers of broadcast programs may also use specialised equipment to modify the audio signal to achieve a desired ‘sound’ for their programme. The modification could involve compression [Katz, undated] to increase the apparent loudness, or the boosting of high frequencies to achieve a ‘bright’ sound and so on. These modifications are distortions since the original waveform has been modified, however the listener may not necessarily classify the distortion as an impairment.

### 2.1.1 Measuring linear distortion

Linear distortion is measured through swept frequency or spot frequency measurements of the system amplitude versus frequency and phase versus frequency transfer characteristics. An input signal with amplitude within the known linear operating range of the system is applied. The input signal will be a single frequency sinusoid. The input and output signals' amplitude and phase are measured and displayed as a graph of amplitude or phase versus frequency, over the range of frequencies of interest.

The amplitude vs. frequency and phase vs. frequency transfer characteristics are then compared to the desired constant amplitude vs. frequency and linear phase vs, frequency characteristics of a distortionless system. The system is then rated as 'acceptable' or 'unacceptable' based on comparison with a norm established through subjective evaluation of the perception of the signal as being impaired or not.

### 2.2 Non-Linear Distortion

*Non-linear distortion* is signal amplitude dependent. This means that the amplification factor in equation 2.1 is not constant for all signal amplitudes. The amplification is often lower for high level input signals than for low level input signals due to amplifier saturation. Non-linear distortion results in the generation of new signal frequency components that add to the existing components resulting in a change in the signal envelope. "Non-linear distortion changes the frequency content of the input signal such that energy is transferred from one frequency at the input to more than one frequency at the output." [Temme, undated]

Harmonic distortion may or may not be experienced as an impairment to the audio signal due to the psychoacoustics of human hearing. Short duration distortion may not be audible, low frequency distortion is less objectionable than distortion at high frequencies, even harmonic distortion may under certain circumstances even be pleasing to a human listener [Temme, undated].

#### 2.2.1 Measuring non-linear distortion

Distortion is a relative measurement. It is expressed as a percentage of the power in the harmonic components compared to the power in the fundamental plus the harmonics. This is known a Total Harmonic Distortion (THD) [Temme, undated].

$$\%THD = \frac{\sqrt{H_2^2 + H_3^2 + H_4^2 + ..H_N^2}}{\sqrt{H_1^2 + H_2^2 + H_3^2 + H_4^2 + ...H_N^2}} \times 100 \qquad (2.4)$$

Where:

$H_1$ is the amplitude of the fundamental.

$H_N$ is the amplitude of the Nth harmonic.

This is the method used in commercial distortion analysers. This method will give an accurate enough reading where the distortion is small, say, less than 10%, which would be the case in audio equipment. This method is a measurement of the ratio of distortion to signal plus distortion.

The correct method of measuring harmonic distortion compares the distortion products to the input signal, it is thus described as follows [ITT, 1973]:

$$\%THD = \frac{\sqrt{H_2^2 + H_3^2 + \ldots H_N^2}}{\sqrt{H_1^2}} X100 \qquad (2.5)$$

In this method the input signal amplitude has to be accurately set otherwise the measurements will give erroneous results.

## 2.3 Wow and Flutter

### 2.3.1 Description

Wow and flutter is the term used for frequency variations in recorded audio signals, caused by speed variations in the transport mechanism of the recording or playback medium. The frequency variations give rise to pitch (subjective perception of frequency) variations, which are readily observed by a human listener. These frequency variations can be compared to Doppler shift in a signal transmitted from a moving object.

In an audio tape recorder the magnetic tape is pulled past the recording or playback head through the combined action of the capstan and the pinch roller. The capstan is a small diameter steel shaft and the pinch roller is a larger diameter rubber wheel. The tape is pinched between these two and the rotation of the capstan moves the tape. Any variations in the speed of rotation of the capstan causes speed variations in the tape.

The capstan speed of rotation can be affected by dirt, worn bearings, misalignment of the axis of the capstan and so on.

Irregularities in the capstan cause high frequency speed variations, known as flutter, and irregularities in the pinch roller cause low frequency speed variations, known as wow. The speed variations translate into frequency variations in the audio signal (See Chapter 5).

Similar distortion effects are present in the mechanism which rotates a vinyl record past the playback head.

The audibility of wow and flutter distortion will depend on the peak frequency deviation and the frequency of the deviation (Zwicker and Fastl, 1990). The greater the frequency deviation caused by the irregularity the more noticeable the distortion. Irregularities causing a deviation at a frequency around 4 Hz are most disturbing.

Digital audio codecs based on perceptual coding techniques may introduce distortion similar to wow and flutter (Shlien and Soulodre, 1996).

### 2.3.2 Measuring wow and flutter

Wow and flutter in analogue recordings is measured using standard pre-recorded tapes or vinyl records. A sine wave of standard frequency is recorded on the tape or record and the frequency deviation is measured on playback. The deviation from the standard frequency is expressed as a percentage, and this is then the measure of wow and flutter [AES, 1982]. For example if the recorded signal is a 3 kHz sine wave and it is found to vary between 2994 Hz and 3006 Hz then the wow and flutter distortion would be 0.2 %.

### 2.3.3 Current relevance of wow and flutter distortion

Wow and flutter is a distortion that has largely disappeared with digital recording techniques, however worldwide sales of vinyl long playing records (LPs) still matches sales of compact discs (Mock, 2004), and there are numerous analogue recordings which are being re-mastered into digital format. In the process of re-mastering wow and flutter is one of the distortions that need to be identified and corrected.

Research into this problem is ongoing as is clear from the program of the AES 117$^{th}$ Convention (Thursday, October 28 until Sunday, October 31, 2004. Moscone Convention Centre, San Francisco, CA, USA) where the following two papers will be presented:

**1 Wow and Flutter Compensation Employing Spectral Processing of Audio**—*Andrzej Czyzewski, Przemyslaw Maziewski, Marek Dziubinski, Andrzej Kaczmarek, Bozena Kostek*, Gdansk University of Technology, Gdansk, Poland.

**2 Correction of Wow and Flutter Effects in Analog Tape Transfers**—*Jamie Howarth,* Plangent Processes, Nantucket, MA, USA; *Patrick Wolfe*, University of Cambridge, Cambridge, UK.

Compensation/correction of wow and flutter in recordings is a time intensive process and it would be beneficial to have a simple technique for the identification of those recordings that require correction.

### 2.4 Dynamic Range

#### 2.4.1 Description

Audio signal dynamic range is defined as:

*the difference in loudness between the loudest peak and the "noise floor" of your equipment* [Graham, undated].

We could also define audio signal dynamic range without reference to the inherent capabilities of the equipment involved by referring to dynamic range as:

*The difference between the loudest and quietest sounds in a recording, or in a live performance, or the difference between the highest and lowest amplitude signals in a recording.*

The dynamic range of the human ear is approximately 120 dB if taken as the range between the threshold of hearing and the threshold of pain. Musical instruments have a dynamic range, which is much smaller as shown in the following table [Rossing et al, 2002].

| **Instrument** | **Maximum Dynamic Range (dB)** |
|---|---|
| Violin | 40 |
| String Bass | 30 |
| Recorder | 10 |
| Flute | 30 |
| Clarinet | 45 |
| Bassoon | 40 |
| Trombone | 38 |

**Table 2.1:** Maximum dynamic range of musical instruments

The dynamic range of individual instruments does not give an indication of the dynamic range of an orchestral performance or of a combination of instruments. One instrument may occupy the lower part of the useable dynamic range of the human listener while another may produce a sound almost at the threshold of pain.

The dynamic range of classical music is generally much wider than that of popular music. The useable dynamic range will depend on the listening situation. Ambient noise in high-density city housing will reduce the useful dynamic range compared to a concert hall or a recording studio. Fletcher [Fletcher, 1942] estimated the required dynamic range to be 100 dB in a residential listening environment. Fletcher's work is unlikely to yield the same results today as the environmental noise levels in residential areas have increased substantially.

Fielder [Fielder, undated] proposes a dynamic range of 109 dB under professional playback circumstances.

The dynamic range of an audio signal cannot be divorced from the practice of compression. Compression is done in two ways:

1) Reduction of audio dynamic range, so that the louder passages are made softer, or the softer passages are made louder, or both. Examples include the limiters used in broadcasting, or the compressor/limiters used in recording studios. (Katz, 2003)

2) Digital Coding systems which employ data rate reduction, so that the bit rate (measured in kilobits per second) is less. Examples include the MPEG (MP3) or Dolby AC-3 (now called *Dolby Digital)* systems. (Katz, 2003)

The reduction of audio dynamic range is of interest but data rate reduction is not of interest here.

Dynamic range reduction is a method employed by broadcast program managers to increase the apparent loudness of their programs and by the music industry to achieve the same ends [Graham, undated; Katz., 2003; Carroll, undated].

Audio equipment displays sound levels through the use of VU (Volume Units) or PPM (Peak Program Meters) meters to allow control of sound levels. These meters do not measure loudness and are more useful in controlling the peak deviation of an FM transmitter. Audio signals with widely differing dynamic ranges may give the same readings on a PPM meter but to a listener there will be an appreciable difference in loudness. To overcome this broadcasters and the recording industry make use of dynamic range processors to reduce the range between the loudest and softest sounds.

Current dynamic range control equipment make use of multi-band processing [Octiv, 1999]. This avoids the problem of ‘breathing’ where, for example, recording a drum and flute together results in changes in loudness for the flute every time the drum is played, or where background noise is increased in level whenever there is no other signal present.

### 2.4.2 Measuring dynamic range

Dynamic range measurements most often refer to the first definition in section 2.4 above, using the noise floor as an indication of the lower amplitude that can be used in a signal and the onset of non-linear distortion as an indication of the higher limit. This definition restricts itself to the dynamic range of the equipment (system) and does not consider the dynamic range of the signal.

Measuring the dynamic range of the equipment is readily accomplished once the limits have been defined for the onset of non-linear distortion as the upper limit, e.g. 1% THD. This form of dynamic range measurement gives an indication of what can be achieved but does not say anything about what has been achieved. The measurement indicates what the equipment is capable of but does not measure how the dynamic range of the equipment has been utilised by the signal passing through. In audio signal processing the dynamic range of the signal is most important from the perspective of the human listener.

When the dynamic range of an audio signal has been changed it could be said that the signal has been distorted. With current dynamic range compressors the distortion is unlikely to be such that it exceeds the allowable limits of, say, non-linear distortion. If a human listener were to listen to the signal before and after compression they may not be able to observe a difference between the two signals, an expert listener however would detect the difference.

### 2.5 Summary

Four types of distortion that affect audio signals have been discussed and the techniques used to measure these distortions have been explained. It is clear that not all measurable distortions give rise to observable impairments to the sound as perceived by a human listener. Some types of distortion may in fact be used by sound engineers to achieve a particular 'sound'. The most common process being the use of compression to increase the loudness of the signal.

The measurement techniques mentioned here are standard techniques. Their use is in ensuring that the equipment used in sound recording, playback and transmission does not cause objectionable impairment to the signal. All of these techniques are based on the 'stimulus-response' paradigm.

# CHAPTER 3

## THE PSYCHOACOUSTICS OF HUMAN HEARING

### 3.0 Introduction

Psychoacoustics is the study of the relationship between the physical properties of a sound and the hearing sensations evoked in a human listener. Of interest are the relationships between what can be physically measured and the human perception of that physical quantity and how the human auditory system reacts to particular combinations of sounds. Aspects that will be discussed are:

a) frequency in Hz and the sensation of frequency denoted by pitch,
b) signal amplitude or sound intensity and the sensation of loudness,
c) the spectro-temporal masking effects of spectrally or temporally proximate signals, and
d) the concept of consonance and dissonance.

The human auditory system can be described as follows [Gold & Morgan, 2000]:

a) The outer ear terminates at the eardrum and affects the acoustics of sounds reaching the ear from the outside.

b) The middle ear performs mechanical impedance transformation, from the malleus (driven by the eardrum) to the stapes (driving the inner ear fluids).

c) The basilar membrane acts like a bank of mechanical tuned circuits, with resonances over the range of auditory signals.

d) Motion of the basilar membrane is transmitted to the stereocilia of the hair cells and this leads to the firing of peripheral auditory neurons.

e) Auditory nerves adapt to a stimulus. At first they will spike vigorously in response to a new stimulus then they will settle to a steady state as the stimulus is maintained.

f) Each auditory nerve has a characteristic frequency, which is a function of its place on the basilar membrane.

g) Various non-linearity’s exist in the auditory system leading to effects such as limited dynamic range of a nerve, masking effects and combination tones.

The response of the human auditory system is not linearly related to the external stimulus. If the intensity of a pure tone is doubled the human listener does not perceive it to be twice as loud, if the frequency of a pure tone is doubled the human listener does not perceive that the pitch has doubled. The same applies to the perception of the duration of a tone.

The subjective perception of a stimulus is affected by other parameters. For example, the perceived frequency (known as the pitch) will be affected by the intensity and the spectral content of the stimulus

The auditory system behaves as though it is made up of a bank of filters. The centre frequencies of the filters being the characteristic frequencies of the auditory neurons. The filter bandwidth changes with frequency, filters with higher centre frequency have larger bandwidth. At frequencies above 1 kHz the filters appear like ¼ octave filters.

The filter model of hearing gives rise to the concept of critical bandwidths. Within the critical bandwidth components interact leading to masking, loudness summation consonance and dissonance.

## 3.1 Perception of sound intensity, loudness

Loudness is defined as the subjective intensity of a sound [Verhey, 1999]. A subjective measure of loudness is the sone [Gold & Morgan, 2000]. The relationship between the sound pressure, p, and the loudness in sones, S, is given by:

$$S \propto p^{0,6}$$

A sone value of one is set to be the loudness of a 1kHz tone at an intensity of 40 dBSPL. Since intensity is proportional to the square of pressure:

$$S \propto I^{0,3}$$

Sounds at the same sound pressure level (dBSPL) but at different frequencies do not sound equally loud. Equal loudness contours (loudness in phons) show that sounds at 4 kHz are loudest, i.e. the curves have their lowest point on an equal loudness contour at a frequency of 4 kHz. Equal loudness contours do not have the same shape, at higher sound pressure levels they tend to be flatter than at low sound pressure levels and the perceived loudness of tones at different frequencies shows less variation at high intensity than at low intensity.

Loudness summation is an effect involving the critical bands. A signal will sound louder if its bandwidth extends over more than one critical band, this indicates that the output from the different critical bands is summed [Verhey, 1999].

## 3.2 Perception of sound frequency, pitch and timbre

Pitch is the frequency perception of a pure tone, "Pitch is the feature of a sound by which listeners can arrange sounds on a scale from 'lowest' to highest'." [Scheirer, 2000]. This is the common definition of pitch, but as Terhardt [Terhardt, 2000] says, " … pitch is an auditory attribute not only of single, isolated tones, but also of tones that are accompanied by additional sound, and also of *multiple* simultaneous tones (sounds)." In our daily lives we do not encounter many situations where we hear a single tone, sounds are more complex, even in music.

The pitch of a single tone can be explained by the 'place' theory of pitch perception where the ear acts as a 'spectrum analyser'. The place theory runs into problems when more complex sounds are involved, the pitch sensation may even be evoked by tones that are not part of the complex signal, e.g. missing fundamental in a repetitive signal.

Timbre is the perception of a complex signal, i.e. one that has more than a single spectral component. Timbre allows a listener to discriminate between sounds having the same pitch, duration and loudness, i.e. a listener can tell whether a sound comes from a violin or a flute, or a listener can identify a person by listening to them speaking.

The exact mechanism of pitch perception is still under debate. The theoretical and experimental developments leading to the present state of knowledge are well summarized by Cartwright et al. [Cartwright el al, 1999] as follows:

- Ohm [1843] proposed that pitch perception was due to the ability of the ear to perform Fourier analysis on acoustical signals. This required that a component at frequency $\omega_0$ must be present in the incoming stimulus in order for a corresponding pitch sensation to be experienced.
- Seebeck [1843] showed that the fundamental could be removed from the spectrum of a periodic sound without affecting the perceived pitch.
- Von Helmholtz [1863] reinforced Ohm's view asserting that the ear acts as a rough Fourier analyzer and hypothesized that this analysis is performed in the basilar membrane. Non-linearity's of the ear could account for the missing fundamental through the generation of difference combination tones. Difference combination tones are equal to the difference in frequency between successive harmonics (or partials). This would equal the fundamental tones.
- Von Békésy [1960] demonstrated experimentally that Helmholtz's hypothesis is essentially correct, i.e. the basilar membrane effects a rough Fourier analysis of the incoming stimulus.
- Schouten et al [1962] demonstrated experimentally that the missing fundamental is not a difference combination tone by shifting the spectrum of a complex signal by an amount $\Delta f$ while keeping the difference between the spectral components equal to $\omega_0$, the supposed fundamental. The perception of a listener was that the pitch had shifted by the amount $\Delta f$.
- Current models are the spectral model of Cohen et al [1995] and the temporal model of Meddis and Hewitt [1991].

Perceptual pitch detection in complex sounds makes use of the correlogram [Slaney and Lyon, 1990; Scheirer, 1999; Granqvist and Hammarberg, undated]. The correlogram is a three dimensional representation of the autocorrelation of an audio signal. The three dimensions are: time, amplitude and frequency (calculated from the correlation lag).

The correlogram is constructed from short time autocorrelations calculated for different frequency bands within the bandwidth of the audio signal. The results can be viewed in time sequence to identify components that are correlated over time. This allows the identification of different sound sources that contribute to the audio signal since changes in a one source over time will not be correlated to changes in any other source.

While pitch, as such, does not give rise to distortion, a change in pitch, as occurs in wow and flutter distortion, can be construed to be distortion. Similarly if the playback rate is different from the rate at which the sound was recorded the pitch of the sound will be shifted and the result will be distortion.

Timbre, the combination of individual sounds into a complex whole, is also worth studying in terms of identifying sounds which are distorted. Mostly we can assume that sounds produced for listening are pleasing to the listener, however certain sounds when combined seem to blend while others clash. Scheirer [2000] mentions that the sounds from a clarinet and a horn will blend while those from an oboe and a trumpet will clash. Sound processing, feedback, deliberate distortion of electric guitars, and so on may give rise to sound sensations that are not necessarily pleasant.

### 3.3 Spectro-temporal masking

The ability to hear a sound depends on the presence of other sounds within the same critical bandwidth, and also on the presence of sounds outside that critical bandwidth. When two tones are present simultaneously they may not both be heard. This effect is more pronounced for tones that are close in frequency, i.e. in the same critical band. The louder a tone the larger the frequency range over which it will mask another tone.

Masking may also occur even when the two tones are not present simultaneously. This is known as temporal masking. Such masking decreases with an increase in the time difference between the two tones.

The listener uses cues in complex sounds to improve perception. If a tone and noise are combined in the same critical bandwidth and the noise is modulated in some way, e.g. the level fluctuates periodically, then the masking threshold can be measured. If noise with the same modulation is introduced in adjacent critical bands the masking threshold changes in such a way that a higher level of noise is required to achieve the same masking as before. This phenomenon is known a co-modulation release (CMR) [Scheirer, 2000].

Differential loudness, a just noticeable difference, between two tones is affected by the presence of other tones. If each of two tones are surrounded by identical supplementary tones then the just noticeable difference becomes smaller, it is easier to hear a difference when unchanged tones are present for contrast [Scheirer, 2000].

Some masking effects may be affected by the non-linearity of human hearing. For example while a strong tone may mask a weaker tone the non-linearity could give rise to an inter-modulation product, which may not be masked (Robinson and Hawksford, 2000).

It is clear that perception is affected by the complexity of a sound.

## 3.4 Consonance and dissonance

Central to the question of the measurement of distortion in audio signals is the perception of a signal as being distorted. If a human listener did not perceive a signal to be 'distorted' then there is no need to measure the amount of distortion, what cannot be heard is of no consequence. Dissonant sounds are not pleasing to the human listener, but a distinction should be drawn between musical dissonance and sensory dissonance [Terhardt, 1974]

Musical consonance is a cultural concept, and psychoacoustic consonance or sensory consonance, a sensory concept [Terhardt, 1974]. Sensory consonance or its opposite sensory dissonance is the same for all people while musical consonance depends on the aculturalisation process. In music, what sounds pleasing may be culturally determined. Human speech on the other hand may provide a better basis for sounds that can be classified as distorted or undistorted. Could a human produce a distorted sound? Is the ability to discriminate between sounds, as pleasant or unpleasant, related to survival in primitive societies?

Schwartz et al [2003] state that consonance judgments, "… arise from the statistical structure of naturally occurring periodic sound stimuli. An analysis of speech sounds shows that the probability distribution of amplitude-frequency combinations in human utterances predicts both the structure of the chromatic scale and consonance ordering." This indicates that dissonance may be associated with the unusual, or the unpleasant. 'Roughness' is one way of describing the sensation of dissonance and this is one of the indicators used for describing pathological voice quality [Tsai, 2004].

In music, notes may be grouped together in combinations at different frequencies, known as chords, or grouped together in time, known as melodies, or a combination of both. Groupings are often made on the basis of creating a whole that sounds pleasant as opposed to unpleasant. Combinations that sound pleasant are consonant while unpleasant grouping are dissonant or rough.

The consonance of two pure tones is related to their spectral separation and their absolute frequency. This is shown in the dissonance curve in Figure 3.1.

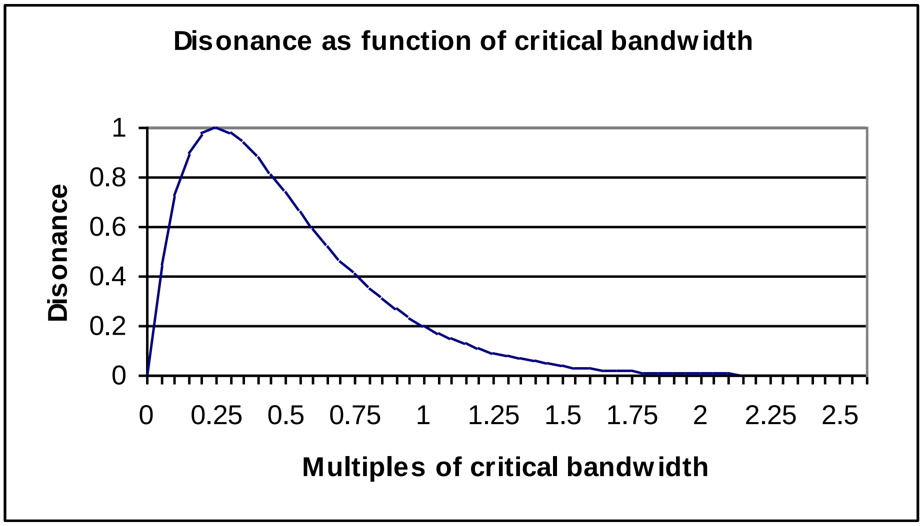


**Figure 3.1:** Dissonance in musical tones as a function of critical bandwidth

The dissonance curve is derived as follows. When two tones with small frequency difference are presented simultaneously, one frequency with a beating (periodic) change in amplitude is heard. When the frequency difference is increased the rapid fluctuations in amplitude cannot be followed and the sound gets an unpleasant or rough character [Leman, 2000] as the dissonance between the two tones becomes apparent.

Plomp [1965] found that the maximum dissonance occurred at about 25% of the critical bandwidth (25 Hz at frequencies below 500Hz and at about 4%-5% of the frequency in the range above 500 Hz).

Computation of sensory dissonance is done according to two models [Leman, 2000]:
* The curve-mapping model: All frequency intervals or frequency component pairs present in the spectrum of the sound are mapped to a psychoacoustical curve (see Figure 3.1). The dissonance of a complex sound is then defined to be equal to the sum of the individual dissonances. This model does not take into account temporal effects, the amplitudes of components and their phases, and noise like sounds.
* Auditory modelling: This approach is intended to overcome the shortcomings in the curve-mapping model. Leman [2000] has extended this model in the so-called

synchronization index model. Tsai [2003, 2004] discusses the shortcomings of the curve fitting approach as applied to sub harmonics and their contribution to roughness.

The auditory modelling approach is better suited to real world signals.

### 3.5 Summary

The psychoacoustics of the perception of sound is important in that the final determinant of the quality of an audio signal must be how it sounds to a human listener. The concepts of consonance and dissonance, first articulated by Helmholtz [Benson, 2004; Leman, 2000], provide a basis on which to approach the analysis of sounds in terms of their pleasing or not effect on a human listener.

Through evolution the human auditory system has evolved to perceive some sounds as pleasing and others as not pleasant, possibly through association with danger or disease, e.g. an upper respiratory tract infection causes the voice to sound rough. Research into consonance starting with Pythagoras through Helmholtz has concentrated more on the consonance of combinations of single frequencies and on what is termed musical consonance. Recent work has focused more on real world signals other than music.

The task of relating consonance in a broad sense, i.e. sensory consonance, involving real world sounds, to measurements of distortion in an engineering sense, is an area of research that has not received much, if any, attention.

# CHAPTER 4

## APPROACHES TO THE MEASUREMENT OF AUDIO SIGNAL QUALITY PARAMETERS

### 4.0 Introduction

Objective measurements of audio system parameters like frequency response, non-linear distortion, time base errors and signal to noise ratio are of great importance to audio system engineers. Objective measurements are used to determine the performance of the system and individual components of the system and as such are important in maintaining the quality of the audio signal. An audio signal will at some time be presented to a human listener as sound and objective measurements of audio signal quality must be related to subjective sound quality, the quality perceived by a human listener.

Conventional measurement procedures require that the system (or individual components of the system) be taken out of service in order for the system parameters to be measured using specially designed test signals. Advances in measurement techniques, and changes in audio signal processing requirements, have necessitated the use of the transmitted audio signal as the test signal. For example the almost universal use of digital transmission systems for the distribution of sound signals has given rise to compression techniques to limit the required bandwidth. The compression only works on an audio signal, tests signals give false results, and so testing a sound transmission system using compression requires an audio signal as input.

The conventional test procedures and the more recent procedures have one factor in common, the use of stimulus-response testing, see Figure 4.1. This means that both the input signal to the system and the output signal from the system must be available for the measurement to be carried out. The measurement then involves some kind of a comparison between the input and output signals, the input signal acting as an undistorted reference signal. Examples of this type of measurement procedure are frequency response measurements, distortion measurements and gain or attenuation measurements.

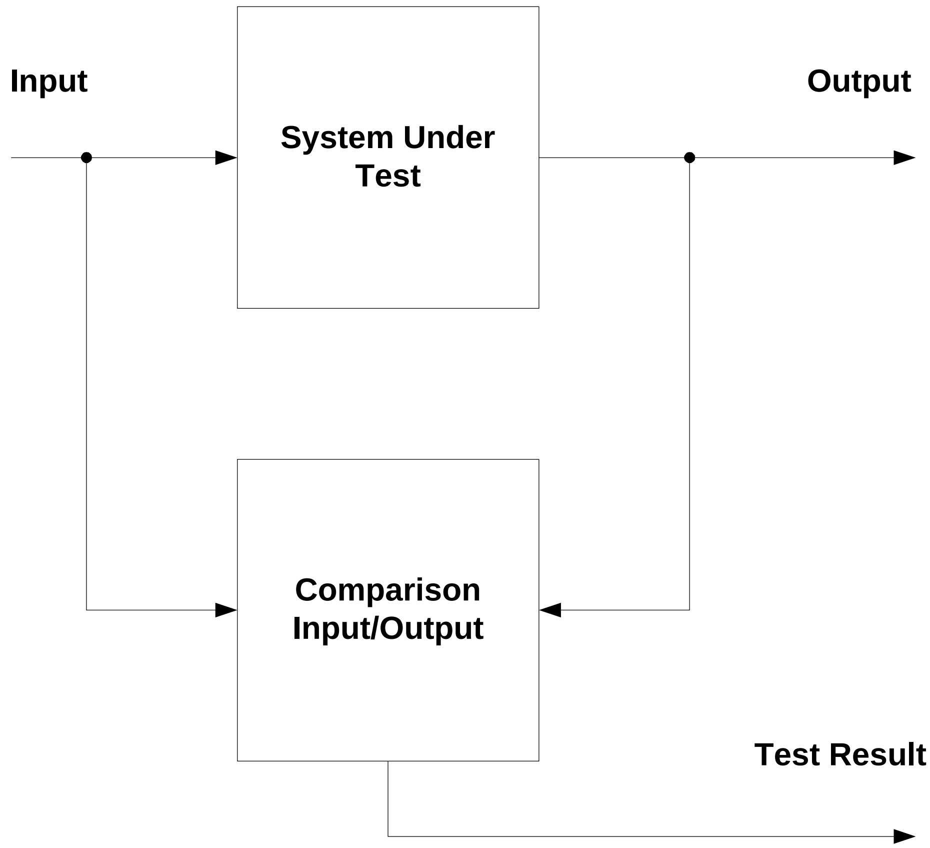


**Figure 4.1:** Stimulus-response testing

In this chapter various approaches to the use of measurements as a tool for maintaining audio signal quality are discussed.

## 4.1 Background

Stimulation of a transmission network with well-defined, deterministic, test signals, suited to the application of the network, and measurement of the change in these test signals at various points in the network, is the method most frequently used to maintain the desired quality of performance of the network [Weaver, 1971]. However, this method restricts the tests that may be carried out on the network to those for which test signals have been defined, and in the case of sound and vision signals intended for perception by humans, to those for which measurements determining the level of subjectively observable impairment have been devised [Allnatt, 1983; Atkin, 1991; Weaver, 1971]. In this type of testing the tacit assumption is that the test signal is completely known, that changes in its parameters can be measured and quantified, and that these changes are related to subjectively observed impairments. "If we wish to 'measure' what we 'hear' then we must deal with subjective perception and the illusion of sound" [Heyser, 1976A]. There are however problems in finding the relationship between an objective measurement and subjective quality as perceived by a human listener [Beerends & Stemerdink, 1992A; Heyser, 1976A; Heyser, 1976B].

The closer the test signal approximates the signals that may be transmitted over the network during normal use, the closer the measurement comes to reproducing the actual operating conditions. The frequency response of an amplifier can be measured using a sinusoidal input signal at different frequencies and comparing the input and output signals at each frequency. This procedure does not tell us much about the response of the amplifier to a complex, dynamically changing signal such as a typical audio signal, nor to the response when the input signal amplitude is different from that at which the test is conducted.

Deterministic test signals are inherently artificial and do not reflect the variety of signals that may be transmitted across a network. Attempts have been made to model the programme signals on a typical audio programme transmission circuit [Ehara, 1977, 1982; CCIR]. Test signals reflecting the complexity of these models are reported in the literature [Hollier et al,1993].

Transmission networks are not time invariant. Operator accessible parameters may be varied, transmission conditions may change, and circuit parameters may deviate with time, temperature and other ambient conditions. As a result of these changes the network operator must determine suitable intervals for conducting tests on the network in order to maintain acceptable performance. When gradual deterioration in circuit performance takes place between scheduled maintenance tests, the network operator may not be aware that the system is no longer performing adequately until users of the network complain about the performance.

In a broadcast network the signal source may also be the source of distortion. A tape machine, record player, amplifier or other piece of equipment may not be aligned or may be malfunctioning. This piece of equipment may be used infrequently and may not be subject to regular maintenance. The cause of signal distortion would therefore not be traced if the transmission network is subjected to testing. However the end user, the listener, would be aware that the signal has been impaired.

Some performance parameters in high quality audio networks are not amenable to determination through the use of deterministic test signals [Lipshitz & Vanderkooy, 1981]. Such parameters are often determined by listening tests where an experienced listener(s) makes a subjective judgement on the quality of the signal under controlled conditions. Listening tests are also employed as a matter of course in the daily operation of a broadcast system since (amongst other factors) the equipment is in daily, continuous use and often is not taken out of service until a failure occurs. Quality determination using test signals is thus done only after repairs have been affected and not on a regular basis.

## 4.2 Distortion: Linear and Non-Linear

In passing through a transmission system a signal will be distorted if the spectral characteristics of the signal are changed. This change may result from linear or non-linear distortion.

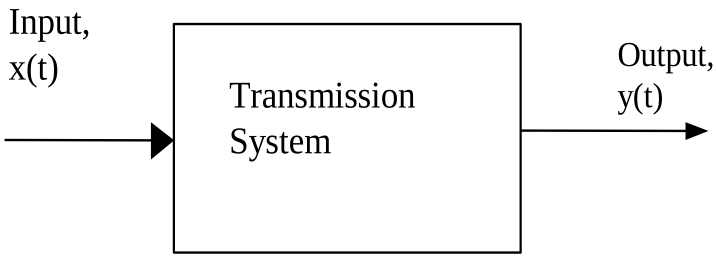


**Figure 4.2**: Signal transmission

*Linear distortion* changes the amplitude and/or phase relationships between the frequency components of the signal, so that the shape (envelope) of the signal is changed [Preiss, 1976]. Linear distortion is not signal level dependent. It is a function of the amplitude and phase relationship of the spectral components of the signal. If the phase and amplitude of the output signal components have changed relative to the relationships that existed at the input to the system, then the signal has been distorted. The same components will be present at both the input and the output. Referring to Figure 4.1, for distortionless transmission it is required that:

$$y(t) = kx(t - T) \tag{4.1}$$

Where:

$y(t)$ is the output signal from the system

$x(t)$ is the input signal to the system

$k$ is a constant amplification factor

$T$ is a constant time delay.

In terms of the frequency and phase transfer functions we have:

$$Y(f) = \int_{-\infty}^{\infty} y(t)\, e^{-j2\pi ft} dt = kX(f)\, e^{-j2\pi fT} \tag{4.2}$$

Where:

$X(f)$ is the Fourier transform of $x(t)$, the input signal.

Since $X(f) = |X(f)| \angle \Phi(f)$, the output spectrum will be:

$$Y(f) = k\,|X(f)|\,\angle(\Phi(f) + 2\pi fT) \qquad (4.3)$$

The signal will therefore be distorted if either or both of the following happen in passing through the system:

a) The constant, *k*, in equation 4.3 is not a constant but is a function of frequency.

b) The constant delay, *T*, in equation 4.3 is not a constant but is a function of frequency.

*Non-linear distortion* may be signal level dependent. This means that the amplification factor in equation 4.1 is not constant for all signal amplitudes. The amplification is often less for high-level input signals than for low-level input signals due to amplifier saturation. Non-linear distortion results in the generation of new signal frequency components that add to the existing components resulting in a change in the signal envelope. "Non-linear distortion changes the frequency content of the input signal such that energy is transferred from one frequency at the input to more than one frequency at the output." [Temme, undated]

### 4.3 Conventional test procedures

*The difference between input and output provides intelligence regarding the system under test.*

The above sentence sums up the conventional approach to performance testing of audio signal transmission and recording equipment. This approach is known as the *stimulus-response* approach to performance testing. In stimulus-response testing the system under test is stimulated with a known input signal. The output from the system is observed and changes affected by the passage of the stimulus signal through the system are noted and quantified. The input signal may be a single sinusoidal waveform, or it may be a complex combination of different signals, or it may even be an example of a typical signal that the system has been designed to transmit. What matters is that the input signal is available for comparison with the output.

When stimulus-response testing is used in performance evaluation of transmission systems used for communication to human beings, an attempt is made to relate the tests to aspects of human perception, so that the tests will identify changes in the signal that may be perceptible to humans. In other words, an objective measurement of a subjective experience is sought.

Measurement results have to be repeatable and comparable; therefore the input (test) signal is standardised so that measurements on different equipment can be compared. Test signals have evolved over time. Early testing used static signals i.e. signals whose characteristics do not change over time. Single frequency sine waves are typical of these test signals [Heuber, et al.1976; King, 1979; Kuni, 1997]. Such signals are adequate for identifying changes in linear time invariant systems, e.g. amplifiers, assumed to be linear over their operating range; filters, and so on. However when the system's characteristics are intentionally designed to be non-linear and time varying (audio conditioning equipment such as automatic level controllers are examples), different test signals have to be devised [Leinonen et al. 1977]. These signals more closely resemble the actual audio signals, intended for transmission over the transmission system during its normal operation. The test signal must contain the salient features of the information signal to be transmitted and it must be analysed in terms of the perceptual significance of errors that may be introduced by the system under test [Hollier et al. 1993].

Conventional test procedures characterise the device under test using deterministic test signals. Conventional test signals rely on a correlation between the distortion of the test signal by the device under test, and the human perception of distortion. Conventional test procedures are not good at characterising time varying or non-linear systems [see also Hollier et al. 1993] .

### 4.4 Modelling of the human auditory system

In 1992 the following statement was made “Perhaps the most surprising issue in audio coding is the lack of a reliable objective quality criterion” [Paillard et al. 1992]. It would appear that rapid progress was made since then, as Beerends [Beerends 1998] describes a perceptual approach to the determination of audio quality. This approach relies on a model of the human auditory system, which indicates the difference in perceived quality between a reference and the signal being evaluated, see Figure 4.3.

Note that this approach is similar to that which has been referred to as conventional testing, in that the circuit or system is stimulated, and the output is compared to a reference. In this case the reference is the original undistorted audio signal (telephone speech or CD quality audio). Conclusions about the quality of the output signal are drawn from a comparison with the reference via a model of the human auditory system ( Figure 4.3).

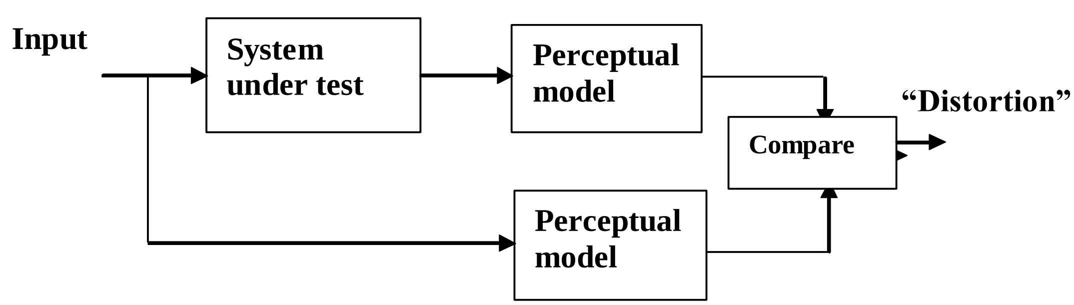


**Figure 4.3:** Perceptual model quality assessment

This method of determining the quality of an audio signal became necessary as digital audio codecs (encoder/decoder) based on knowledge of the human auditory perception system were developed for use in, for example, cellular telephone systems. These codecs made testing using conventional test signals problematic since the encoder would be designed to remove redundant information from the input signal, so as to decrease the required transmission bandwidth. A conventional test signal would be distorted by the processing introduced by the encoder.

The operation of the codec is signal dependent and testing needs to be performed using the actual signal that would be transmitted through the codec, e.g. audio signals [Hollier & Hawksford, 1995]. An audio encoder could, for example, respond differently to audio signals having different frequency content and different transient characteristics e.g. speech or music.

The perceptual approach to audio quality determination does not attempt to characterise the device, or system, under test. It uses an 'ideal' signal as reference (the input signal in Figure 4.3) and, using perceptual models, determines the *audible* differences between the reference and the output signal from the system. To characterise the performance of a system in terms of audio quality a large set of test signals covering the range of possible input signals is required. The subjective quality of the set of test signals must be known.

An alternative approach uses a test signal simulating the typical information signal likely to be transmitted over the system. Again the test signal and the output signal are processed to provide perceptual surfaces, which are then compared. Audible difference surfaces are produced and these give an indication of the distortion introduced by the system [Hollier et al. 1993].

In determining the perceptual response of a human to an audio signal, use is made of knowledge of the human auditory system and the way in which sound is perceived. A summary of pertinent aspects is given in Zwicker & Zwicker (1991). Firstly it has to be recognised that perception is not a linear process. The human auditory response to a change in signal frequency or signal amplitude is non-linear. A doubling in signal amplitude does not result in a doubling in perceived *loudness* to a listener; changes in signal frequency are perceived similarly. To relate signal amplitude to perceived amplitude, i.e. loudness, a logarithmic scale is used and perceived amplitude is expressed in units of *sone*. Perceived frequency, i.e. pitch, is similarly expressed in units of *bark* [Zwicker & Zwicker, 1991].

The second important factor in human perception relates to the masking effect. Signal components of higher level will mask lower level signal components in their proximity both in time and frequency [Pierce, 1983; Howard & Angus, 2001; Hartmann, 1998].

Perceptual model based testing cannot directly predict the perceived quality of a signal, but it can predict aspects of the auditory sensation such as pitch, loudness and masked threshold. The process of mapping a signal into a perceptual surface is described in [Beerends & Stemerdink, 1992B]. In summary the process is as follows:

The signal is windowed in time and FFT transformed to find the power spectrum, $P(t,f)$, as a function of time and frequency. The power spectrum is then transformed to a power spectrum in terms of time and pitch to change from measured frequency to perceived frequency. This function is convolved with a spreading function to find the excitation. The excitation is compressed to give a compressed loudness-time-pitch representation of the original signal. Two signals can then be compared on the basis of their differences in the loudness-time-pitch domains. This gives a better approximation to how a human auditory system will respond to a signal and quantifies differences between signals in terms of parameters relating to the actual functioning of the auditory system. Comparisons can only be made between a time segment of an audio signal and *the same time segment* of the audio signal that has been passed through some audio process, e.g. a coding-decoding process like coding from analogue to MP3 and back to analogue. This is therefore a stimulus-response test, the input is known and the output is compared to the input.

## 4.5 Coherence testing

*'Coherence is a frequency domain measure of common (that is statistically linearly dependent) spectral components between the input and output of a system'* [Totzek & Preiss, 1987].

Coherence testing is a form of stimulus-response testing which uses the audio signal as the stimulus. Coherence testing of audio signals was proposed by Totzek and Preiss [Totzek & Price,1987] as an alternative to the conventional, objective, steady state measurement techniques, based on using sine waves as test signals. In their paper they show that many of the parameters measured using sine wave testing could also be measured using coherence testing. Kates applies coherence testing to the measurement of distortion in hearing aids [Kates, 1992].

Coherence testing is based on the following [Bendat & Piersol, 1971]:
The cross-spectral density function of two sets of random data is the Fourier transform of the cross-correlation of the two data sets. The cross-spectral density is a complex valued function since the cross-correlation is not an even function of time. The cross-correlation is given by:

$$R_{xy}(\tau) = lim_{T\rightarrow\infty} \frac{1}{T}\int_0^T x(t)y(t+\tau)dt \qquad (4.4)$$

where *x(t)* and *y(t)* are the input and output signals respectively.
Taking the Fourier transform of the cross-correlation we get the cross spectral density function:

$$G_{xy}(f) = \int_{-\infty}^{\infty} R_{xy}(\tau)e^{-j2\pi f\tau}d\tau \qquad (4.5)$$

The coherence function is given by [Bendat & Piersol, 1971]:

$$\gamma_{xy}^2(f) = \frac{|G_{xy}(f)|^2}{G_x(f)G_y(f)} \leq 1 \qquad (4.6)$$

where $G_x(f)$ *and* $G_y(f)$ are the power spectral density functions of the input and output signals respectively. The coherence function, $\gamma_{xy}^2(f)$ equals zero for any frequency where the two signals are uncorrelated and equals one where they are correlated.

If the system frequency response function $H(f)$ is known then the cross spectral density is given by:

$$G_{xy}(f) = H(f)G_x(f) \quad (4.7)$$

From this expression the frequency response function can be found if the cross spectral density function is known.

The cross spectrum function is a complex function. The frequency dependent angle of the cross spectrum function represents the phase shift through the system between input and output. This can be used to find the frequency dependent delay, i.e. the group delay of the system.

### 4.6 Automated testing

Automated testing of audio equipment and transmission systems has been made possible by the availability of inexpensive processing power. In the process of automated testing a test signal is inserted into the audio signal during transmission and this signal is detected at various points and analysed.

The test signal consists of a combination of signals which may be used to measure frequency response, signal to noise ratio, intermodulation distortion and so on. The signal may be preceded by a trigger to alert the measuring apparatus to the arrival of a test sequence.

The test sequence may be audible to listeners and as such not acceptable. To avoid listener complaints the test sequence could be hidden in a ‘signature tune’, a time signal or something similar. Audio signals do not have the luxury of the vertical interval found in the television video signal where test signals can be inserted without affecting the viewer in any way.

Automated testing is another form of stimulus-response testing.

### 4.7 Summary

The background to determination of audio signal quality is discussed. Three existing methods of audio system testing are described. The three methods are: Testing using deterministic test signals; Modelling of the human auditory system; Coherence testing.

Conventional testing using deterministic test signals does not directly determine audio signal quality but is used to maintain the performance of the transmission equipment so that the audio signal passing through will not be subjected to more distortion than is acceptable to a human listener.

In audio signal quality determination through modelling of the human auditory system, an attempt is made to relate the measurement directly to the audio quality as perceived by a human listener. The measurement involves the audio signal and a model of the human auditory system. This is a change from the conventional test procedure where the test is validated through listening tests, which relate the objective test to the subjective experience of the distortion that has been measured.

Coherence testing is an objective test procedure. The audio signal is used as test signal. No attempt is made to model the human auditory system and the results of the objective test has to be validated by listening tests to relate the outcome to a subjective listening experience.

# CHAPTER 5

## DETECTION OF DISTORTION DUE TO MECHANICAL IMPERFECTIONS IN SOUND RECORDING AND PLAYBACK APPARATUS (Wow and flutter)

### 5.0 Introduction

This research is concerned with the detection of distortion in audio signals as opposed to measurement of distortion. Detection is carried out by analysing the audio signal, measurement is usually done to quantify the distortion introduced by audio equipment. Measurement techniques will be discussed to illustrate that most conventional techniques cannot be used to detect distortion in an audio signal.

The conventional approach to measuring the magnitude of the distortion introduced by mechanical imperfections in sound recording and playback apparatus, is as follows:

a) Make a distortion free recording of a single sine wave signal,

b) play the recorded signal back on the equipment to be tested, and

c) detect the frequency variations introduced by the mechanical imperfections.

The deviation in frequency from the frequency of the test signal is expressed as a percentage of the test frequency. This is then the measured result. A similar process is followed where recording apparatus is to be tested [King, 1979; Kuni, 1997].

In this chapter a mathematical analysis of the distortion introduced into an audio signal by mechanical imperfections in sound recording or playback apparatus is presented. Three alternative methods are presented for detecting and analysing this distortion Two of these methods have been reported in the literature [McKnight & Weiss, 1976; Godsill et al. 1998], and the third method has been developed by the author [Maré, 2004].

## 5.1 The Nature of the Distortion

Mechanical imperfections in analogue audio recording and playback apparatus cause speed variations in the recording or playback medium. For example, the magnetic tape in a tape recorder is pulled past the magnetic head by the capstan. If the capstan does not turn smoothly, perhaps due to worn bearings, the tape will not move past the magnetic head at a constant speed.

If the audio signal has been recorded on a machine exhibiting mechanical imperfections and assuming that a different machine will be used in the playback process, these speed variations will be translated into frequency variations in the resulting audio signal, (possibly) causing an audible impairment to the signal. The mechanical imperfections are not sinusoidal and depend on the characteristics of each individual make and model of machine, e.g. the different drive mechanisms in a turntable, or the different capstan diameters in a tape recorder. Machines of the same make and model would have common characteristics and impart distortions having the same characteristics to the signal. The distortion can be used as a signature imparted to the signal and can be used to identify the individual model of machine on which the recording was made [McKnight & Weiss, 1976].

The effect of speed variations can be analysed as follows. Assume that the recording medium, assumed to be magnetic tape, undergoes speed variations during its passage past the playback head, where the normal (or design) speed, $V_0$, changes periodically, and the peak deviation from normal is $\Delta V$. The resultant speed would be:

$$v(t) = V_0 + d(t) \quad mm/s \tag{5.1}$$

where: $V_0$ is the normal speed in mm/s

$d(t)$ is the function describing the speed variation

$\Delta V$ is the peak deviation in speed = max(d(t))

Assuming further, for ease of analysis, that the speed variation is sinusoidal:

$$v(t) = V_0 + \Delta V \cos(\omega_x t) \quad mm/s \qquad (5.2)$$

where: $\omega_x$ is the deviation frequency
$\Delta V$ and $V_0$ are as above

The amplitude of the recorded signal on the recording medium is a function of (the one dimensional) position on the recording medium. Assuming that a sinusoidal signal has been recorded on the recording medium, the variation of signal amplitude with position on the medium will be given by:

$$m(x) = M \cos(\omega_m x) \qquad (5.3)$$

where: $m(x)$ is the signal amplitude at point x on the recording medium
$M$ is the peak signal amplitude
$x$ is the position on the recording medium in mm from the start
$\omega_m$ is the spatial (signal) frequency in cycles/mm

As the medium moves past the playback transducer, the spatial amplitude dependence is transformed into time dependence, and the signal becomes a function of time. If there are speed variations, either in playback or recording, these will be reflected in the resultant signal frequency. The signal representation on the recording medium, e.g. magnetic flux density, will be transformed into a voltage representation through the transducer reading the recorded signal.

The position on the tape is given by the integral of the tape speed, i.e.

$$x = \int v(t)dt \qquad (5.4)$$

Applying the transformation from position on the tape, to time varying signal, to the expression for *m(x)*, Equation 5.3, transforms *m(x)* into *M(t)* and changes the signal amplitude representation on the tape into a time varying voltage:

$$M(t) = M_t \cos(\omega_m \int v(t)\,dt)$$

$$= M_t \cos(\omega_m V_0 t + \omega_m \Delta V \int \cos(\omega_x t)\,dt) \qquad (5.5)$$

$$= M_t \cos(\omega_s t + \frac{\Delta\omega}{\omega_x} \sin(\omega_x t))$$

where: *M(t)* is the recovered voltage signal as a function of time

$\omega_s = \omega_m V_0$ is the recovered signal frequency in rad/s

$\Delta\omega = \omega_m \Delta V$ is the peak signal frequency deviation as a result of speed variations

$M_t$ is the peak signal amplitude after playback

## 5.2 Spectrum of the Distortion

The signal after distortion has the appearance of a frequency modulated signal (see equation 5.5) [Johns & Rowbotham, 1972], where the mechanical imperfections result in frequency modulation of the audio signal. Analysis of the effect of frequency variations due to tape speed variations is possible if it is assumed that the recorded audio signal is a single frequency sinusoid as has been done above. This analysis also applies to signals that are composed of the sum a number of individual sine waves. When the recorded signal is a normal audio signal e.g. music or speech, then the ‘carrier’ signal, in the analysis above, is a random signal, which sometimes, over short periods of time, can be considered to be composed of a sum of sine waves at different frequencies.

In a frequency modulated signal the peak deviation, of the carrier frequency, is determined by the amplitude of the modulating signal. In the analysis above it is shown that the deviation of the signal frequency from normal, depends on the deviations of the tape speed from the normal. The peak signal frequency deviation, $\Delta\omega$, is determined by the peak deviation, $\Delta V$, in speed from the normal speed, multiplied by the (spatial) signal frequency. The higher the signal frequency, the greater the deviation, and, the larger the imperfection in the playback apparatus, the greater the deviation. Bearing this in mind, the standard analysis for frequency modulation can be applied to derive the spectrum of the reproduced audio signal. The expression in Equation 5.5 is expanded as a sum of Bessel functions resulting in an expression as follows [Haykin, 2001]:

$$M(t) = M_t \sum_{n=-\infty}^{n=\infty} J_n(\beta_f) \cos(\omega_S t + n\omega_x t) \qquad (5.6)$$

where: $M(t)$ is the modulated signal

$\beta_f = \Delta\omega/\omega_x$ is the modulation index

$J_n(\beta_f)$ are the Bessel functions of the first kind of order n

$\omega_S$ is the recorded signal frequency in rad/s

$\omega_x$ is the frequency of the speed variation in the playback medium

$\Delta\omega$ is the peak frequency deviation

## 5.3 Methods of Blind Detection of Flutter Distortion

### 5.3.1 Spectral Analysis [McKnight & Weiss, 1976]

In the analysis of McKnight and Weiss (1976), the peak amplitude of the flutter, i.e. the deviation from normal frequency, is defined as:

$$F = \frac{\Delta f}{f_0} \qquad (5.7)$$

where: $\Delta f$ is the peak signal frequency deviation

$f_0$ is the signal frequency

The flutter is caused by mechanical imperfections causing speed variations in the recording medium, tape or disc. The frequencies at which the speed variations occur are called flutter modulating frequencies, and are denoted by $f_m$ .

If a constant frequency sinewave component is present in the distorted audio signal, then spectral analysis techniques can be employed to find the frequency of the variations in speed of the recording or playback apparatus. Commercial flutter meters analyse the variations in a 3150 Hz test frequency using a frequency demodulator to directly measure the flutter modulating frequencies, $f_{m,}$ and the flutter amplitude $F$ [King, 1979].

In the recordings analysed by McKnight and Weiss (1976) a sine wave frequency component, at the power supply frequency (60 Hz), was present in the recorded signal. The power frequency sine wave provides a single frequency sine wave, similar to the sine wave signal that is used in the normal test procedures. The power frequency signal may be seen as a 'carrier' of the flutter modulation and isolated from the rest of the signal. The parameters of the wow or flutter distortion are found by applying spectral analysis techniques to this isolated signal.

The technique of McKnight and Weiss (1976) does not identify the presence of flutter from the recorded audio signal, it uses the fortuitous presence of a power supply frequency. Normally we would not expect a power frequency component to be present in the recorded audio signal.

### 5.3.2 Estimation of the pitch variation curve [Godsill et al. 1998; Godsill & Rayner, 1998]

Godsill et al. (1998) are concerned with the *restoration* of distorted recordings. Their method is based on determining the 'pitch variation curve' (Note: pitch is the human sensation of frequency). The pitch variation curve describes the way in which the recovered signal has been 'frequency distorted' by the speed variations in the playback apparatus. This curve is estimated using short time discrete Fourier transforms (STDFTs), of the sampled version of the played back audio signal. Expressing this mathematically, the signal with wow or flutter is related to the original, undistorted signal by:

$$M(t) = x(f_m(t)) \tag{5.8}$$

where: $M(t)$ is the distorted signal

$x(t)$ is the undistorted signal

$f_m(t)$ is a time warping function which accounts for the speed variations in the recording or playback mechanism.

If the time warping function, $f_m(t)$ in equation 5.8, is known then its inverse can be found and the distortion corrected.

### 5.3.3 Time domain autocorrelation

This technique for the detection of the presence of frequency variations in a recorded audio signal was developed by the author.

This investigation concerns the *detection* of imperfections in an audio signal and not necessarily the measurement (quantifying) of such a distortion. In detecting the presence of wow and flutter one should note that wow and flutter causes frequency modulation of the recorded signal and, secondly, that the imperfection is of a periodic nature. The period is related to the period of rotation of the capstan in a tape recorder or the period of rotation of a record player drive mechanism. This leads to the hypothesis that wow and flutter can be detected using the autocorrelation function. The autocorrelation of a frequency modulated signal can be found and the result applied to the detection of wow and flutter.

To find the autocorrelation of a frequency modulated signal start with the definition of the autocorrelation of a signal, $m(t)$, [Bendat & Piersol, 1971] and then substitute the

Bessel function representation of an FM signal [Haykin, 2001] for *M(t)*. This results in an expression as follows:

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} M(t)M(t+\tau)\,dt$$

$$where: \; M(t) = M_t \sum_{n=-\infty}^{n=\infty} J_n\,(\beta_f)\cos[(\omega_s + n\omega_x)\,t]$$

(5.9)

$$and$$

$$M(t+\tau) = M_t \sum_{n=-\infty}^{n=\infty} J_n\left(\beta_f\right)\cos[(\omega_s + n\omega_x)(t+\tau)]$$

Since *M(t)* is an infinite sum, do the multiplication term by term and consider each product in turn in order to generalise. Starting at $n = 0$ gives:

$$For\; n = 0:$$

$$M(t) = M_t J_0(\beta_f)\cos(\omega_s t)$$

$$M(t+\tau) = M_t J_0(\beta_f)\cos(\omega_s t + \omega_s \tau)$$

Multiplying these terms:

$$M(t)M(t+\tau) = \frac{\left[J_0(\beta_f)M_t\right]^2}{2}\left[\cos(\omega_s\tau) + \cos(2\omega_s t + \omega_s\tau)\right]$$

After integration we have:

$$R_{xx}(\tau) = \frac{\left[J_0(\beta_f)M_t\right]^2}{2}\cos(\omega_s\tau) \qquad (5.10)$$

Applying the procedure to the more general terms, n≠0:

$$M(t)M(t+\tau) = K\cos[(\omega_s + n\omega_x)t]\cos[(\omega_s + m\omega_x)(t+\tau)]$$

$$where:$$

$$K = [J_n(\beta_f)\, J_m(\beta_f)\, M_t^2]$$

Integrating this expression in order to find the autocorrelation function:

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} M(t)M(t+\tau)\,dt$$
$$= \lim_{T\to\infty}\frac{1}{2T}\;K\int_{-T}^{T}\cos[(\omega_s + n\omega_x)t]\cos[(\omega_s + m\omega_x)(t+\tau)]\,dt$$

Use trigonometric identities to simplify the integrand:

$$\cos(\omega_s + n\omega_x)t\cos[(\omega_s + m\omega_x)(t+\tau)]$$
$$= \frac{1}{2}\cos(\omega_s\tau + n\omega_x\tau)\,.......for\; n = m$$
$$= \frac{1}{2}\cos(\omega_s\tau + (m-n)\omega_x t + m\omega_x\tau)\,...for\; n \neq m$$

From this it follows that if $n \neq m$ the integral is zero, while if $n = m$ the result is:

$$\frac{[J_n(\beta_f)\,M_t]^2}{2}\cos(\omega_s\tau + n\omega_x\tau) \qquad (5.11)$$

The final result for the autocorrelation of a frequency modulated wave, in this case the flutter signal, is found by adding Equation 5.10 and equation 5.11 as shown below:

$$R_{xx}(\tau)=\frac{[J_0(\beta_f)M_t]^2}{2}cos(\omega_s\tau)+\sum_{n=-\infty}^{n=\infty}\frac{[J_n(\beta_f)M_t]^2}{2}cos(\omega_s\tau+n\omega_x\tau) \qquad (5.12)$$

From this it can be concluded that:

* With no flutter present the autocorrelation is that of the originally recorded sinusoid.
* With flutter on the recording medium the autocorrelation takes on the form of a sum of sinusoids, at frequencies in the neighbourhood of the frequency of the original (desired) signal.
* If the amplitude of the flutter increases the number of sidebands increase.
* If the frequency of the desired signal increases the number of sidebands increase.
* The autocorrelation waveform is the sum of all these components.
* Depending on the severity of the flutter and the frequency of the desired signal a situation could arise where there is no sideband component at the flutter frequency but only at multiples of the flutter frequency. This will arise for β around values of 3.832, 7.1, 10.1 and so on, since at these values of β, $J_1(\beta)=0$.

This analysis of the autocorrelation of the sinusoid affected by flutter seems to contradict what can be expected from reasoning about the nature of the autocorrelation. Firstly, the autocorrelation of a signal will highlight periodicities in the signal. A sine wave signal with flutter distortion is similar to a frequency modulated signal. The frequency of the sine wave signal is periodically deviated from its undistorted frequency. The periodicity of the deviation is at the frequency of the flutter and the autocorrelation should therefore indicate a periodicity at that frequency. The analysis, however, indicates that under certain conditions the autocorrelation function will have no component at the flutter frequency. This is because, for certain values of signal frequency and deviation frequency, the value of $J_1(\beta)$ will be zero. At these points the autocorrelation function will have no component at the fundamental flutter frequency.

When considering the time domain representation of a frequency modulated signal it may be expected that a periodicity equal to the frequency of the modulating signal would be present in the autocorrelation signal. As the deviation (or $\beta$) increases, the shape of the autocorrelation should change since the frequency of the signal is being changed during the same time interval by a greater amount, and the peaks and dips in the autocorrelation would be more pronounced.

Reasoning about the appearance of the autocorrelation function does not necessarily lead to a better understanding about the appearance of the signal. Returning to the expression for the autocorrelation function (equation 5.12) and analysing this expression mathematically, very quickly leads to the correct answer. Taking another look at equation 5.12, repeated below:

$$R_{xx}(\tau)=\frac{[J_0(\beta_f)M_t]^2}{2}cos(\omega_s\tau)+\sum_{n=-\infty}^{n=\infty}\frac{[J_n(\beta_f)M_t]^2}{2}cos(\omega_s\tau+n\omega_x\tau)$$

This shows that there will be a component at the ‘carrier’ frequency ($\omega_s$). This component will decrease in amplitude as the deviation increases, i.e. as the speed variations due to mechanical imperfections increase. Apart from that it is merely a sinusoid at a particular frequency.

More interesting is the effect of the ‘sideband frequencies’. These appear separated from the carrier at multiples of the distortion frequency ($\omega_x$). Closer examination of these components reveals that in general the expression for the angle of any of these components can be written as:

$$Angle=\omega_s\tau\pm n\omega_x\tau \qquad (5.13)$$

Ignoring the carrier component and noting that:

$$\omega_x=2\pi f_x$$

Case 1: When $\tau$ is an odd multiple of the distortion frequency period. The angle will be an even (integer) multiple of 2π for even multiples of the distortion frequency and will at the same time be an odd multiple of π for odd multiples of the distortion frequency.

When this is the case ($\tau = n / f_x$, n odd) the autocorrelation function at these times will consist of the in phase sum of all sideband frequencies at even multiples of the distortion frequency and the ($180^0$) out of phase sum of the sideband frequencies at odd multiples of the distortion frequency. Depending on the relative amplitude of the components, the autocorrelation may show a maximum at times equal to odd multiples of the distortion frequency period.

Case 2 The angle will be an integer multiple of 2π whenever $n\tau$ is an even multiple of the reciprocal of the distortion frequency $f_x$.

When this is the case the autocorrelation will be the in phase sum of all the sideband components. The autocorrelation function will thus have a clear maximum at this point in time.

A clear visualisation of the effect of taking the autocorrelation of a frequency modulated signal can be obtained by simulation. This is the subject of the next section.

### 5.4 Simulation of Flutter Distortion

In order to demonstrate the feasibility of using autocorrelation to detect the presence of flutter in an audio signal, a MATLAB program was written. This program simulates the autocorrelation of a frequency modulated wave. The program uses the analysis above as its basis.

The result of running the program and plotting the resultant autocorrelation function show that a very pronounced periodicity exists at the flutter frequency (corresponding to a spacing of 521 samples between the peaks) for values of FM deviation ratio, $\beta$, from 0.5 to 22. Higher deviation ratios were not investigated.

Note: At a deviation ratio of 0.5 the measured flutter distortion would be 0.675 % and at a deviation ratio of 22 it would be 29.7%, assuming a test frequency of 1 kHz and a flutter frequency of 13.5 Hz.

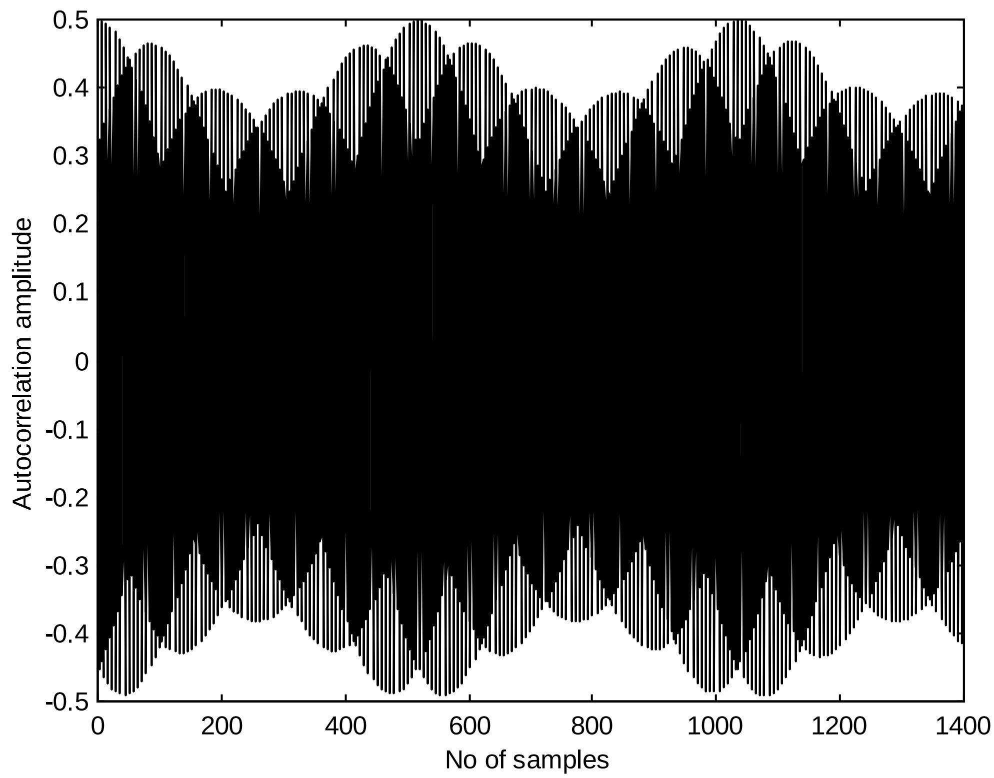


**Figure 5.1:** Simulated autocorrelation with deviation ratio of 0,5

In Figure 5.1 the simulated autocorrelation of an FM signal is shown. The parameters for the simulation are:

Carrier frequency: 2000Hz

Modulating frequency: 13.5 Hz

Sampling frequency: 7034 Hz

Flutter distortion: 0,3375 %

(The parameters for the distortion were chosen to be similar to those of commercial tape recorders. The tape recorder used in this study to simulate flutter distortion was a Revox B77. It would generate a flutter frequency of 13.5 Hz if there was one speed variation per revolution of the capstan.)

Peaks in the autocorrelation are evident at a spacing of approximately 521 samples. The period of these peaks correspond to the frequency of the deviation introduced in the simulation of flutter distortion.

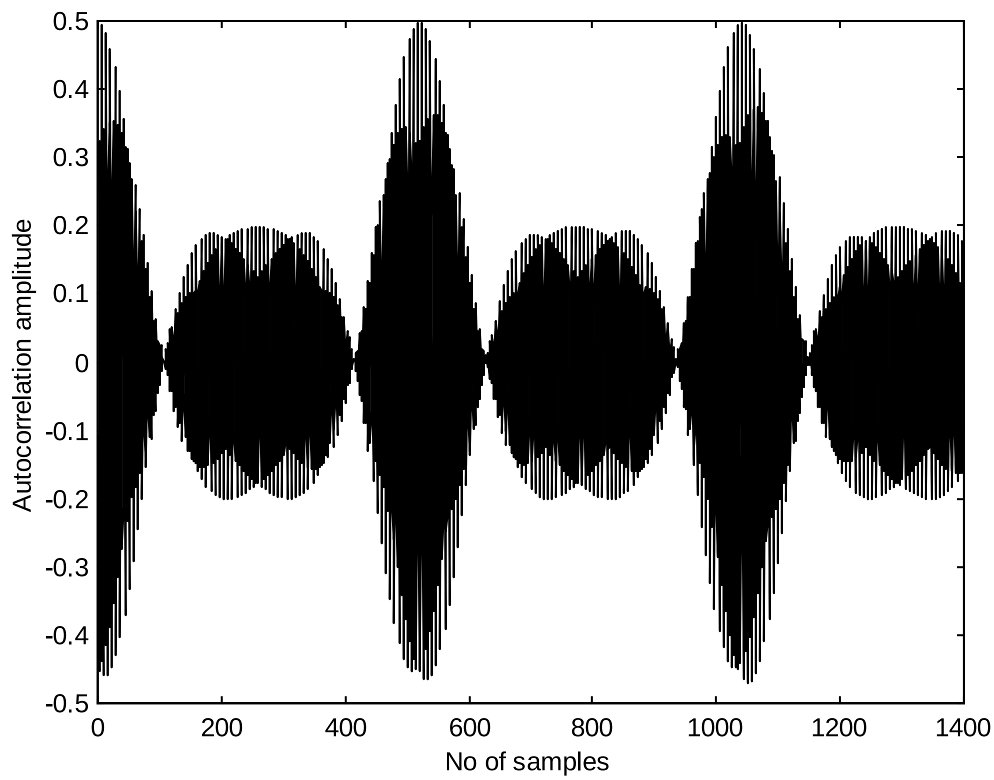


**Figure 5.2:** Autocorrelation of FM with deviation ratio of 2

When the frequency deviation ratio is increased to 2 (flutter distortion 1.35 %), see Figure 5.2, the result is similar to the previous (Figure 5.1) in that the peaks occur at a spacing of approximately 521 samples. The peaks are much more pronounced due to the larger deviation.

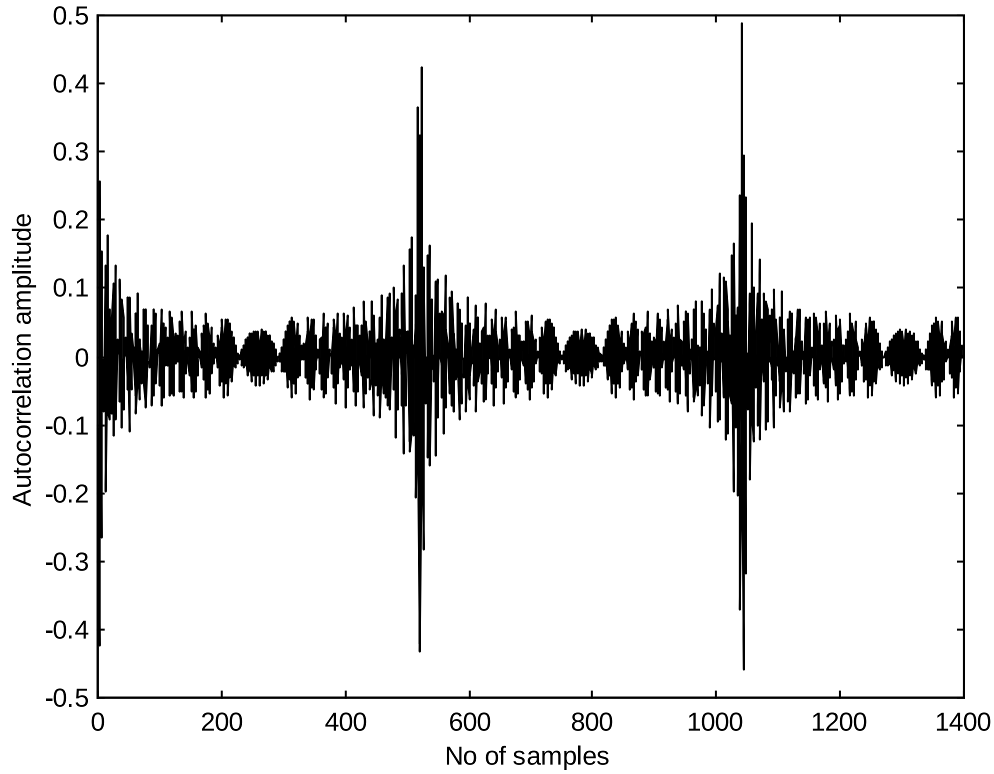


**Figure 5.3:** Autocorrelation of FM with deviation ratio of 22

When the frequency deviation ratio is further increased, to 22 (flutter distortion 14,85 %), the result of the autocorrelation again shows the peaks at a spacing of 521 samples (Figure 5.3).

These results show that it may be possible to detect frequency variations in a simple case where only one ‘carrier’ component i.e. a single sine wave, representing the audio signal, is present. In a real world audio signal this is not likely to be the case as the audio signal is likely to have more than one frequency component. Simulation of more complex signals is undertaken in the next section.

### 5.5 Simulation of Flutter Distortion: A more complex case

The MATLAB program was extended to simulate the case where two ‘carriers’, i.e. an audio signal made up of two components, were frequency modulated by the same modulating signal. From the theoretical analysis it is known that the deviation ratio depends on the frequency of the signal being modulated. This aspect was incorporated in the simulation as shown below. The parameters of the simulated signal were:

Carrier frequency 1: 2000 Hz
Deviation ratio: 0.5
Carrier frequency 2: 4000 Hz
Deviation ratio: 1.0
Modulating frequency: 13.5 Hz
Flutter distortion: 0.3375 %

Here the simulated audio signal is the sum of two harmonically related sinusoids. One is at a frequency of 2 kHz and one at 4 kHz. The deviation of the two sinusoids is appropriately set to 0.5 and 1.0 as the deviation ratio is proportional to the 'carrier' frequency.

The result is shown in Figure 5.4. The peak at 521 sample spacing remains clearly identifiable.

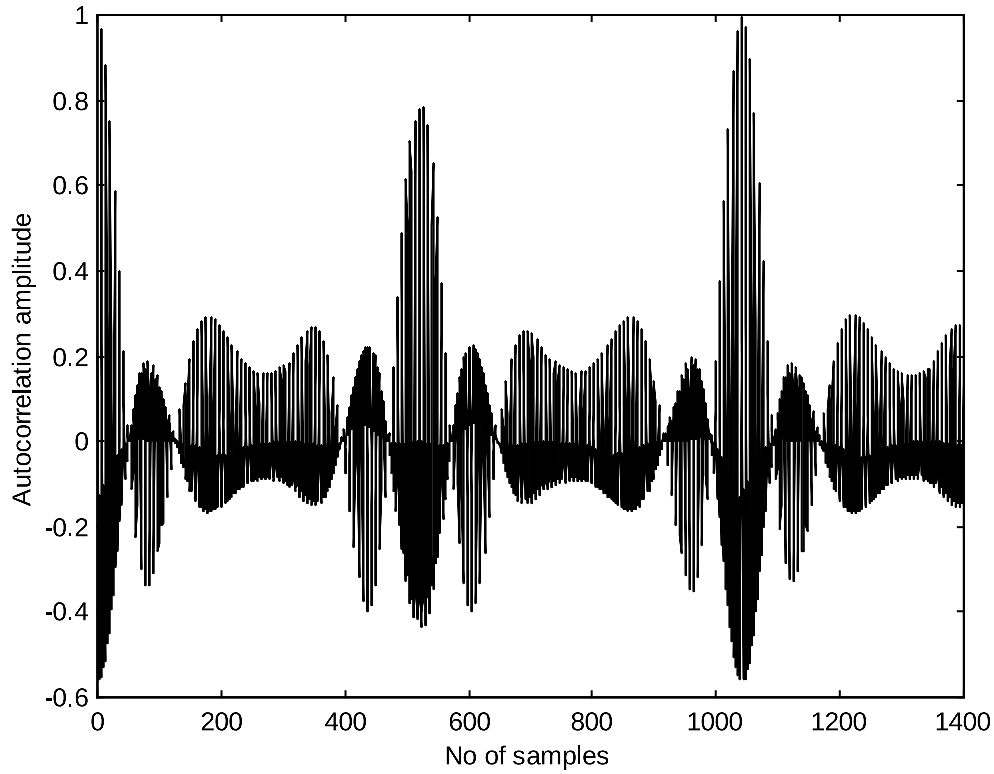


**Figure 5.4:** Autocorrelation of two FM ‘carriers’ which are harmonically related

The simulation is repeated with two carriers which are not harmonically related. The parameters for the simulation are:

Carrier frequency 1: 2000 Hz

Deviation ratio: 2.0

Carrier frequency 2: 3250 Hz

Deviation ratio: 3.25

Modulating frequency: 13.5 Hz

Flutter distortion: 1.35 %

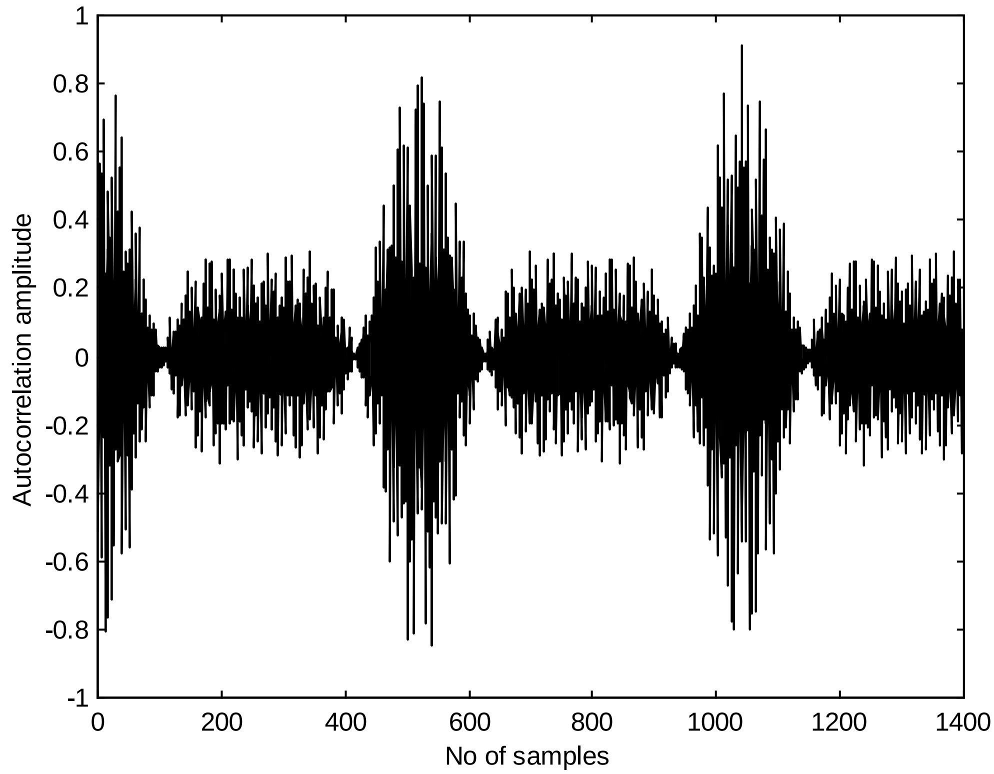


**Figure 5.5:** Autocorrelation of unrelated FM carriers

The peak at 521 samples is still clearly identifiable. Notice how the autocorrelation has taken on a ‘hairy’ appearance. In this simulation the carriers are no longer harmonically related.

A third ‘carrier’ was added to the simulated signal with the signal parameters now being:

Carrier frequency 1: 1310 Hz
Deviation ratio: 1.31
Carrier frequency 2: 2000 Hz
Deviation ratio: 2
Carrier frequency 3: 3250 Hz
Deviation ratio: 3.25
Flutter distortion: 1.35 %

The result of the autocorrelation for this signal is shown in Figure 5.6. The periodicity at 521 samples is still evident.

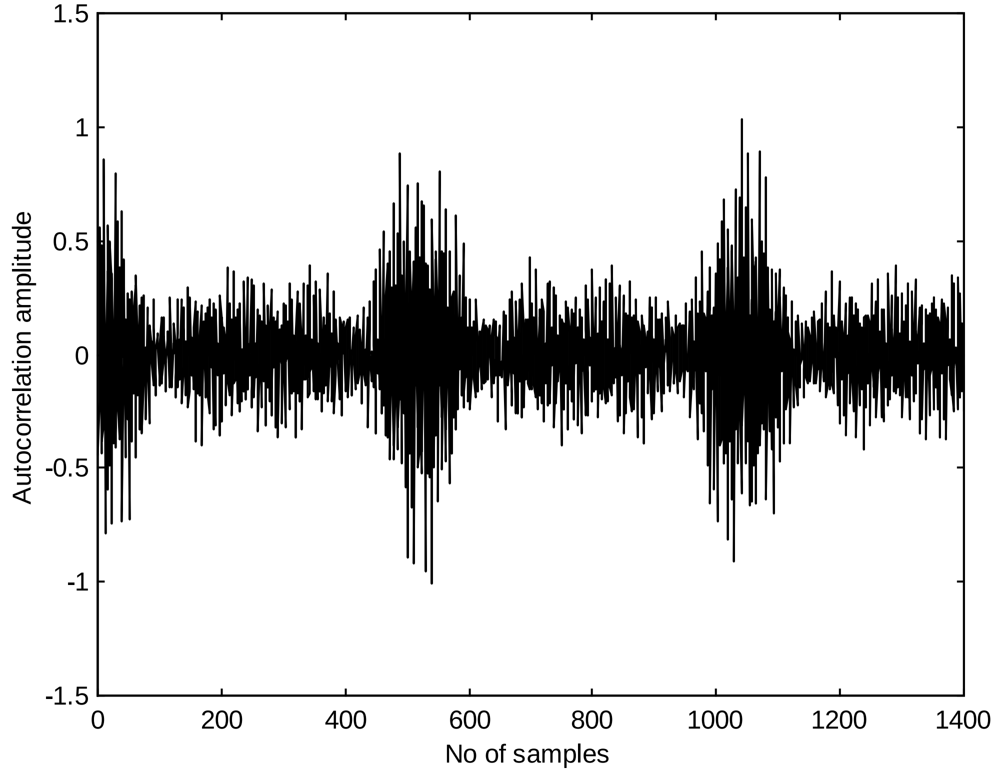


**Figure 5.6:** Autocorrelation of FM with three unrelated carriers

From the examples shown it is clear that the effectiveness of detecting the presence of a periodic frequency perturbation is decreased if the number of 'carriers' is increased and if the carriers are not harmonically related to each other.

Figure 5.7 shows the simulated signal with three 'carriers' before the autocorrelation is carried out. While each of the individual carriers with their frequency modulation is of constant amplitude the sum of the three is not. In an attempt to improve the performance of the autocorrelation method for detection of frequency variations in complex signals the following steps were taken:

a) The complex signal was amplitude limited, and

b) the signal was bandpass filtered to isolate one of the carriers.

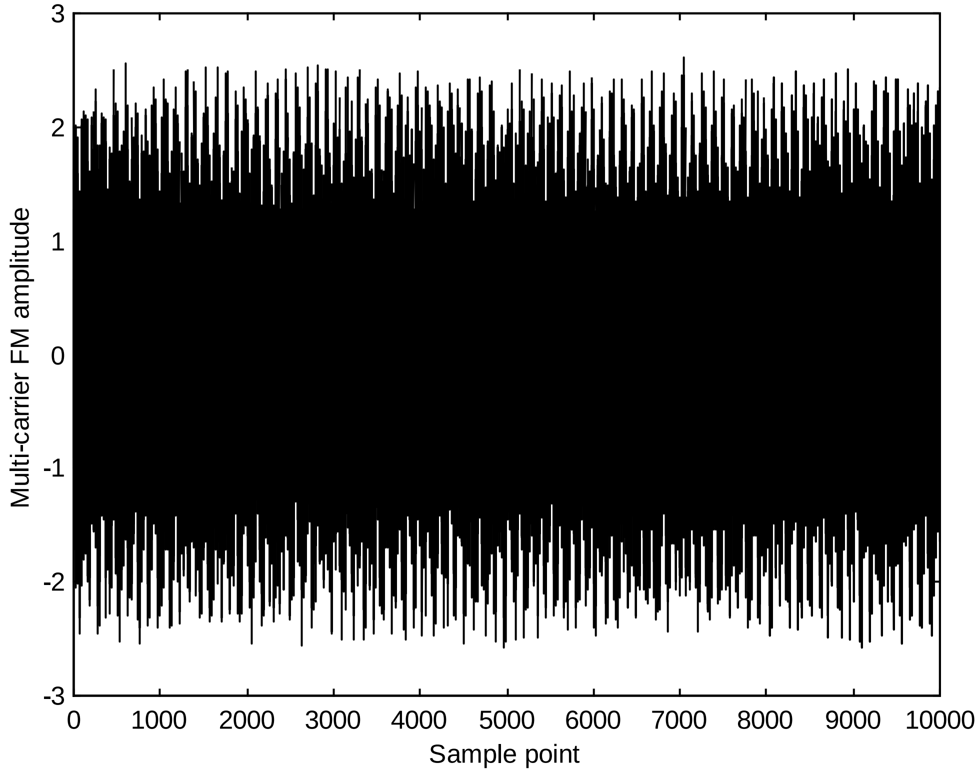


**Figure 5.7:** Three carrier FM signal

The result of this operation is shown in Figure 5.8 and Figure 5.9. In Figure 5.8 the complex signal has been amplitude limited by setting all sample values greater than the limit value to the limit value before the autocorrelation was calculated. In Figure 5.9 the limited signal was filtered using a bandpass filter to isolate the carrier at 2 kHz

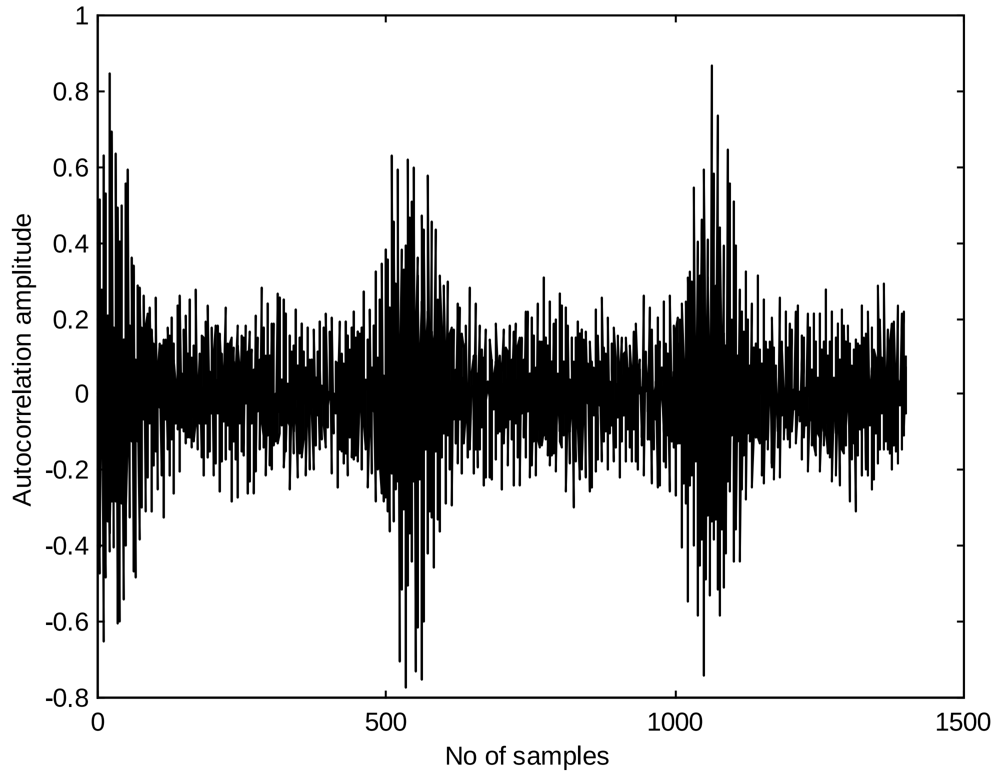


**Figure 5.8:** Autocorrelation of three carrier FM signal after limiting

Comparing Figure 5.6 to Figure 5.8 (signal with amplitude limiting) we see that limiting a complex signal does not improve the ability to detect the presence of frequency variations over that possible with a signal with no amplitude limiting. In both Figures the presence of periodicity at approximately 521 sample spacing is visible.

The plot of Figure 5.9 shows the result of performing the autocorrelation after the signal has been limited and two of the ‘carrier’ signal components have been removed by filtering. The only component left is the 2 kHz sinusoid with deviation ratio of 2.

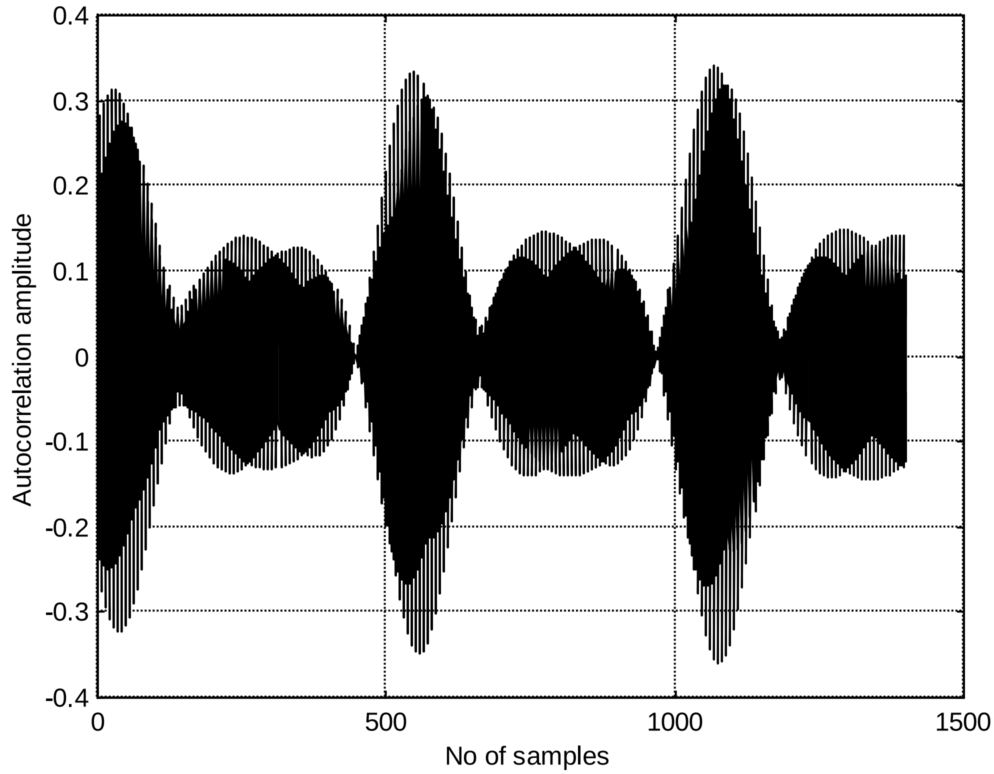


**Figure 5.9**: Autocorrelation of three carrier FM after limiting and filtering

Figure 5.9 should be compared to Figure 5.2. Both show autocorrelation for a single carrier with deviation ratio of 2. The similarity between the two plots indicates that good results can be obtained by limiting and filtering to isolate a single carrier.

The results obtained from the simulations discussed above indicate that the autocorrelation of an FM signal may be used as the basis for identifying the presence of frequency variations (wow and flutter) in a transmitted audio signal. The results are improved if the simulated audio signal is pre-processed by filtering to isolate strong sinusoidal components in the signal. The use of limiting and band pass filtering was investigated and improved results obtained using these techniques.

The limitations on the analysis and simulation are that both the audio signal and the variation are sinusoidal. This is not generally the case. The technique will fail if the audio signal is purely random e.g. an unvoiced sound.

## 5.6 Results of Tests on Recorded Audio Signals

The theoretical possibility discussed above was put to the test using a Revox B77 audio tape recorder to introduce flutter distortion into a real audio signal. The tape transport mechanism in this machine is a capstan with diameter 0.176 inches (~4.5 mm). At a tape speed of 7.5 inches per second (the standard speed for high quality audio recording/playback) the capstan rotates at 13.475 revolutions per second. Any flutter frequencies should therefore be at 13.475 Hz, assuming that the capstan has a simple mechanical imperfection which causes one perturbation per revolution.

The capstan circumference was distorted by sticking a small piece of tape on the capstan. The piece of tape did not wrap around the capstan but covered only a portion of the circumference. This introduced a change in tape speed. A 1 kHz tone was recorded using the tape machine with the distortion present. The tape was then played back without the distorted capstan. The resultant audio signal was sampled at 44100 samples/s and an autocorrelation performed on the data. The result of performing an autocorrelation on this signal is shown in Figure 5.10.

From Figure 5.10 it can be seen that the autocorrelation peaks are separated by approximately $0.01667 \times 10^5$ sample points. At a sampling frequency of 44100 samples/s this translates to a frequency of 26.46 Hz, approximately double the frequency that a simple irregularity would impart. The doubling in disturbance frequency is due to the effect of the tape edges on the capstan. Since there are two edges the actual imperfection occurred twice per revolution.

This result shows that the simulated results can successfully be extended to a simple real case involving a single sinusoid.

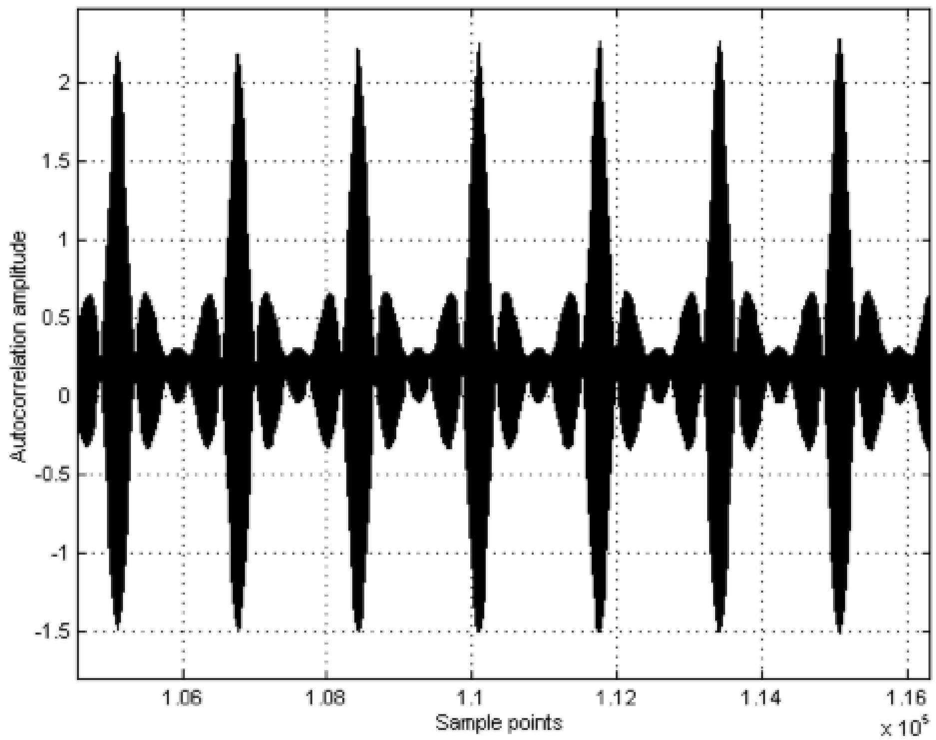


**Figure 5.10:** Autocorrelation of a recorded sine wave

The procedure for detection of flutter distortion was then extended to segments of typical sound programme material. Recordings from a local radio station were obtained, and played back on the modified tape machine. The modified machine would introduce flutter distortion in the reproduced audio signal. The results are shown in the following figures. An autocorrelation was calculated for 204 800 samples i.e. a 4.6 second segment of recorded audio signal. The figures show only a portion of the complete autocorrelation so that the details can be clearly observed.

In Figure 5.11 the periodicity of the mechanical imperfection can be seen in the envelope of the autocorrelation. There are higher frequency variations present as well. These are assumed to be due to periodicities, other than those due to the flutter distortion.

The data in Figure 5.11 were obtained from a filtered version of the original sampled audio signal. The sampled data was filtered using a bandpass filter centred on 4.2 kHz. The bandpass filter was chosen to isolate a strong frequency component centred on 4.2 kHz. The signal was not limited to remove amplitude variations. The low frequency variations, spaced at slightly less than 2000 samples, represent the effect of the speed variations in the modified tape machine. This corresponds to the periodicity detected in a recorded sine wave (Figure 5.10).

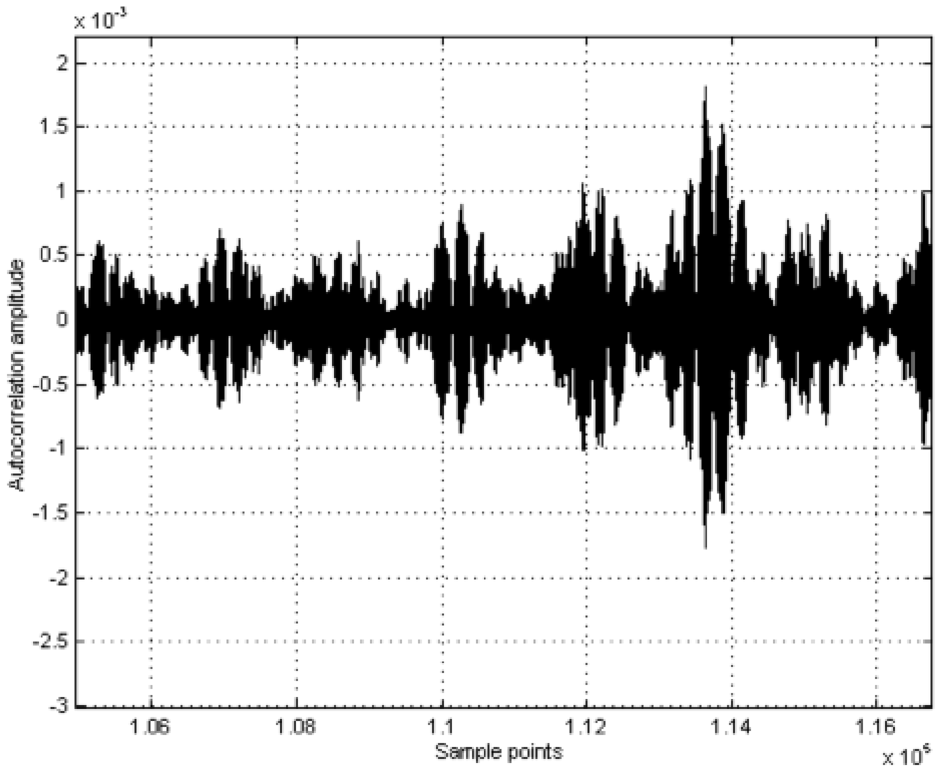


**Figure 5.11**: Autocorrelation of band pass filtered audio signal, modified tape machine

A different segment of audio was used to obtain the results presented in Figures 5.12 and 5.13. The segment without any pre-processing before performing the autocorrelation, is shown in Figure 5.12. The audio signal segment showed a strong frequency component at about 1 kHz. In Figure 5.13 the result of performing an autocorrelation on a filtered audio signal is shown. The filter was centred at 1 kHz.

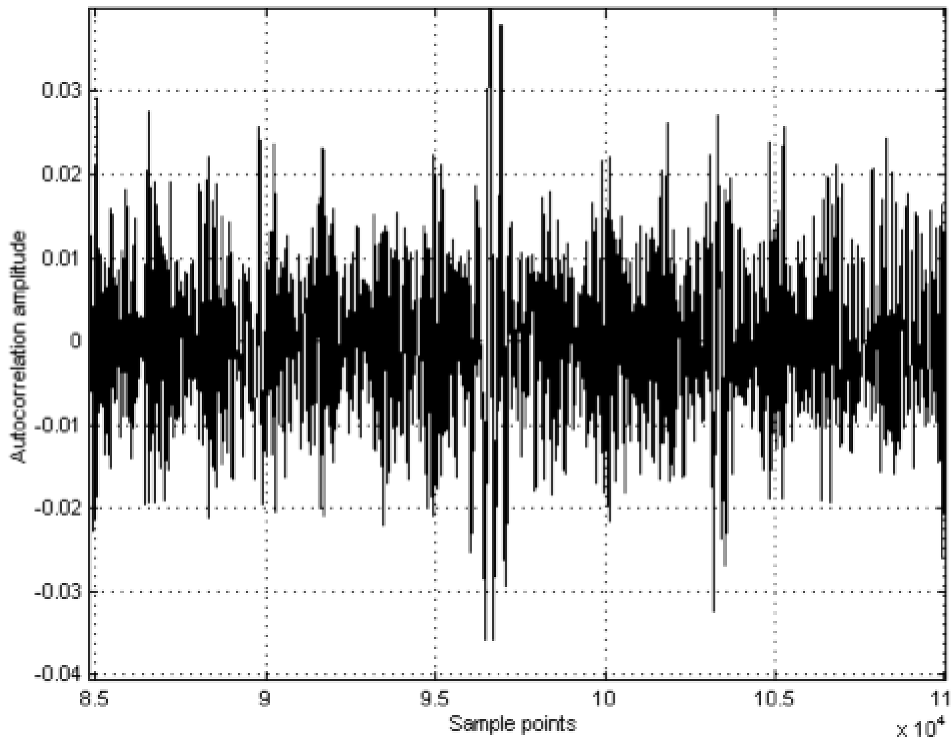


**Figure 5.12:** Autocorrelation of unfiltered audio signal, modified tape machine

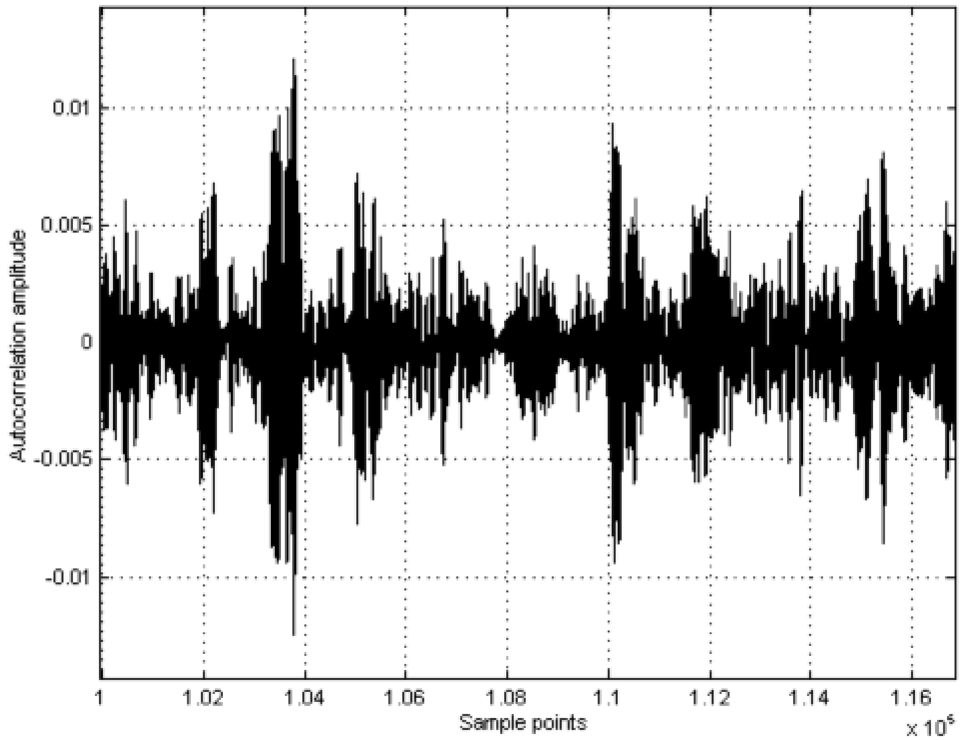


**Figure 5.13:** Autocorrelation of filtered audio, modified tape machine

The process of detection of flutter distortion using the autocorrelation function can be enhanced by extracting the envelope of the autocorrelation. This is done by squaring the result and low pass filtering. The result of this enhancement, applied to the segment shown in Figure 5.11, is shown in Figure 5.14. The component at approximately 2000 sample spacing is more prominent and easily distinguished from the background noise.

A comparison of a segment of an audio signal with and without the addition of flutter distortion was done. The results are shown in Figures A1.1 and A1.2, in the Annexure 1, the Appendix to this chapter, respectively. The flutter distortion can be seen in Figure A1.1 while in Figure A2.2 it is not present.

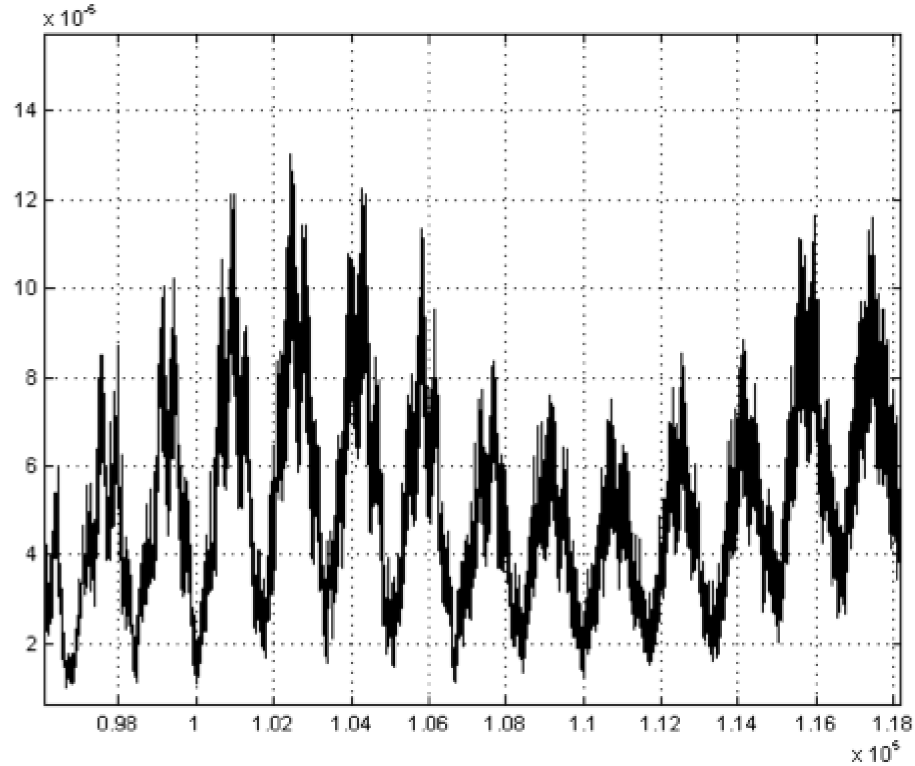


**Figure 5.14**: Envelope of autocorrelation shown in Figure 5.11

## 5.7 Correlation with conventional measurements

An attempt was made to find a correlation between the results obtained using the time domain autocorrelation technique and conventional measurements (see Annexure 1). The autocorrelation peaks do not increase in amplitude in a way that can be related to conventional measurements. This is probably due to the non-linear nature of frequency modulation and the resultant amplitudes of the sidebands.

It is concluded that the autocorrelation technique can be used to detect the presence of wow and flutter distortion but that it cannot be used to quantify the distortion.

The correlogram technique of pitch detection in audio signals [Slaney and Lyon, 1990] makes use of similar techniques. Pitch (and signal frequency) is a parameter that changes with time as the audio signal changes while wow and flutter distortion will yield the same frequency no matter what the frequency of the audio signal is. It will therefore be possible to distinguish between periodicities due to signal structure and periodicities due to wow and flutter distortion.

### 5.8 Summary

Periodic deviations in frequency from the original frequency of recorded audio signals are caused by variations in the speed of the recording medium. Speed variations in the recording medium, either during playback or recording of a signal, are caused by mechanical imperfections in the apparatus. These frequency deviations are sources of audible distortion in the signal. The distortion is known as ‘wow and flutter’.

It has been shown that the presence of periodic frequency variations in a recorded audio signal can be identified using the autocorrelation technique. The technique can be enhanced by pre-processing the audio signal, to remove large amplitude variations, and to isolate any strong frequency components in the audio signal. Extracting the signal envelope after autocorrelation further enhances the effectiveness of this technique.

The results cannot provide a quantification of the magnitude of the distortion but do identify the presence of the distortion and can be used to determine the frequency of the distortion.

A method has been developed and publicised, and confirmed through mathematical analysis, simulation and application to audio signals, for the detection of the distortion caused by mechanical imperfections in recording and playback apparatus.

# CHAPTER 6

## DETECTION OF IMPAIRMENTS IN FREQUENCY CONTENT AND DYNAMIC RANGE

### 6.1 Introduction

Traditional measurements of dynamic range and frequency response involve sine wave test signals [King, 1979]. These tests provide objective, quantitative data about the performance of audio equipment. They do not provide any information about the signals that are likely to be transmitted, or recorded, using the equipment. In this chapter techniques used to extract information regarding frequency content and dynamic range from the transmitted signal, are discussed. These techniques involve the statistical characterisation of the audio signal [Maré, 1985; Maré, 1986].

The *frequency content* of a signal is determined by the collection of spectral components in the signal. Recording artists and broadcasters manipulate the frequency content of an audio signal to shape the sound of their product e.g. emphasizing high frequency content, results in a 'bright' sound.

The *dynamic range* of a signal is defined to be the amplitude difference between the smallest and largest signal amplitudes, the smallest amplitude generally being such that the signal components are not 'lost' in the noise. The largest amplitude signal is that signal which can be passed without excessive distortion.

Frequency content and dynamic range of a signal play an important role in the perceived quality of the signal. A signal with adequate high frequency components will sound more 'bright' than the same signal with high frequency components removed. Not only the presence but also the amplitude of the components and their prevalence is of importance. While it may be possible to have high frequency content present in the signal, since the equipment has adequate bandwidth, this is not enough. For adequate perceived sound quality, the signal must have the high frequency components present.

Signal dynamic range can affect the perceived loudness of a signal and, together with the signal frequency content, is used by program controllers to shape the ‘image’ of their programme. For example, compressing the signal dynamic range to increase the mean loudness of the signal is a practice that is commonly used by the creators of sound broadcast advertisements.

## 6.2 The Cumulative Spectral Amplitude Distribution (CSAD)

Due to its non-stationary nature, an audio signal cannot be characterised by examining short time duration segments in isolation. Statistics need to be gathered over a period of time long enough for accumulating a representative sample of the signal [Maré, 1985].

The technique that has been developed for monitoring the spectral content and dynamic range of an audio signal is called the Cumulative Spectral Amplitude Distribution (CSAD). The CSAD is a graphical representation obtained by repeatedly calculating the discrete Fourier transform of successive, short, time segments of the audio signal. The amplitude of each spectral component is used to increment an entry in a matrix. The position of an entry in the matrix is determined by the spectral component’s frequency and amplitude. The matrix row index corresponds to the signal frequencies and the column index corresponds to signal amplitudes. The number of rows has been set to 70 corresponding to a signal dynamic range of 70 dB and the number of columns is set to 256 corresponding to 256 frequency points between 0 Hz and 22.5 kHz, half the sampling frequency. The choice of 256 points is a compromise between the need for speed in the calculation of the CSAD and the need to accumulate enough samples to accurately characterise the signal properties. More samples will increase the frequency resolution, which is not of prime concern.

The lines in the CSAD, Figure 6.1, indicate the cumulative amplitude distribution of the audio signal at different frequencies. In the example shown in Figure 6.1 the uppermost line indicates the amplitude, at each frequency, exceeded 1% of the time, the second line amplitudes exceeded 10% of the time and from there the lines are spaced apart by 10 percentage points for 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. This means that the CSAD can be used to determine how signal amplitude is distributed over frequency. If the lines are close together then the signal has been compressed in amplitude and its dynamic range has been reduced. This will have the effect of changing the mean amplitude of the signal. If the bunched lines are at a relatively high amplitude then the mean amplitude has been increased, the signal will sound loud.

When sufficient entries have been accumulated, say, over a ten minute segment of audio signal, the cumulative distribution for each column in the matrix is calculated. The result is plotted as lines of amplitude versus frequency. The amplitude lines represent the amplitudes that are exceeded a certain percentage of the time (see Figure 6.1).

The cumulative spectral amplitude distribution of a signal can be used to observe the quality of the signal in terms of:

a) The spectral content of the signal. This can be determined by examining the top most line in the CSAD. Normal broadcast audio would have a slope of approximately -6dB/decade. Deviations from this slope would indicate that the signal has been processed in some way.
b) The amplitude distribution vs frequency. Readings taken along a vertical line drawn on the CSAD at any frequency provide a measure of the signal cumulative amplitude distribution at that frequency.
c) Relative loudness. If the amplitude vs frequency lines on the CSAD are bunched together vertically, then the signal has been compressed and this will affect the relative loudness of the signal.

d) The presence of any interfering tones. Interfering tones show up as strong spectral components with very small spacing between the lowest and highest amplitude lines on the CSAD.
e) The extent of amplitude compression of the signal vs frequency. See (c) above.
f) Consistent loudness and spectral content of items that are presented within the same programme segment.

An example of a CSAD is shown in Figure 6.1.

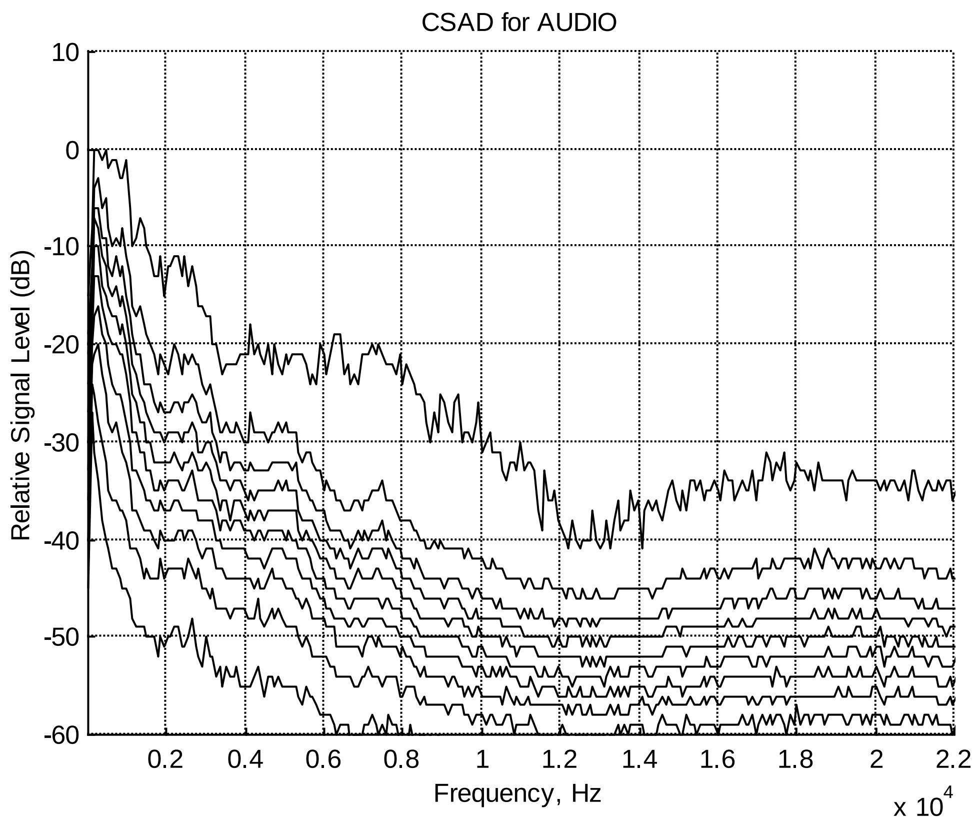


**Figure 6.1**: Example of CSAD for an audio signal

Advertisement inserts into a programme may have been processed to have a high mean amplitude to increase the impact of the advert. Classical music will tend to have lines that are evenly spaced over the whole amplitude range. This improves the dynamic range of the signal.

### 6.2.1 The spectral content of the signal.

The CSAD is an estimate of the statistical distribution of the amplitude of the audio signal at different frequency points. It shows the spectral content of the *signal* and from this deductions can be made as to the quality of the transmitted signal.

The frequency range of the signal is shown in the CSAD display. The frequency range is the span from the lowest frequency to the highest frequency between which the CSAD has identified the presence of significant spectral content. The frequency range can be used to identify problems in the frequency response of equipment e.g. high frequency roll off. The term frequency range is used rather than frequency response since the frequency response refers to the effect a system would have on a signal passing through it while the frequency range refers to the spectral content of the signal itself.

The concept of frequency range is illustrated with reference to Figures 6.1 and 6.2. In Figure 6.1 the frequency range extends to 22.5 kHz while in Figure 6.2 the range has been restricted by low pass filtering to just over 15 kHz. This means that in Figure 6.2 the signal has no frequency components at frequencies higher than about 15 kHz. Lack of frequency range could indicate frequency response problems in equipment or it could alert the sound engineer to the fact that the original signal has a defect.

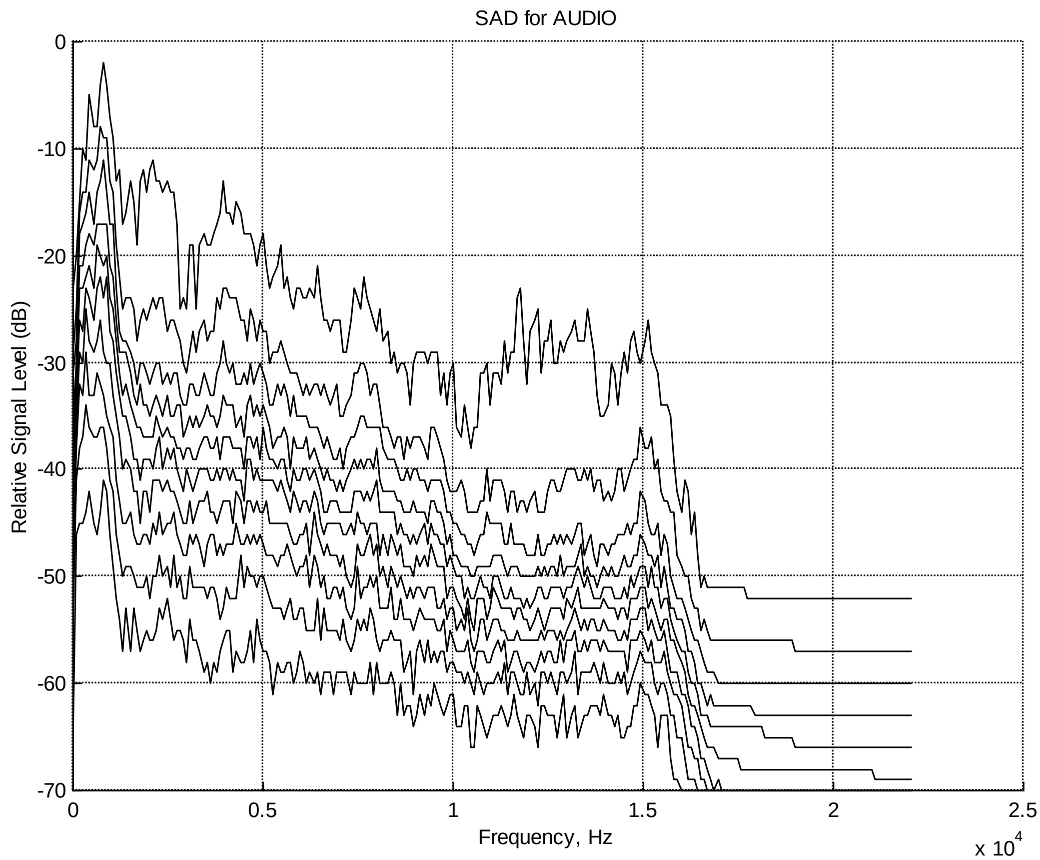


**Figure 6.2:** Low pass filtered audio signal

### 6.2.2 The amplitude distribution vs frequency.

A spectrogram [Rabiner & Schafer, 1978] is a display of the time varying spectral characteristics of an audio signal (see Figure 6.4). In a spectrogram the vertical axis corresponds to frequency and the horizontal axis to time. The intensity of the patterns produced on the spectrogram is proportional to the signal energy. The spectrogram cannot show the amplitude of individual spectral components and the dynamic range cannot be determined from a spectrogram. The dynamic range of a signal is of interest to a sound engineer as it is related to the perceived quality of an audio signal.

A comparison of a spectrogram and a CSAD for the same 10 second audio segment is shown in Figures 6.2 and 6.3 below.

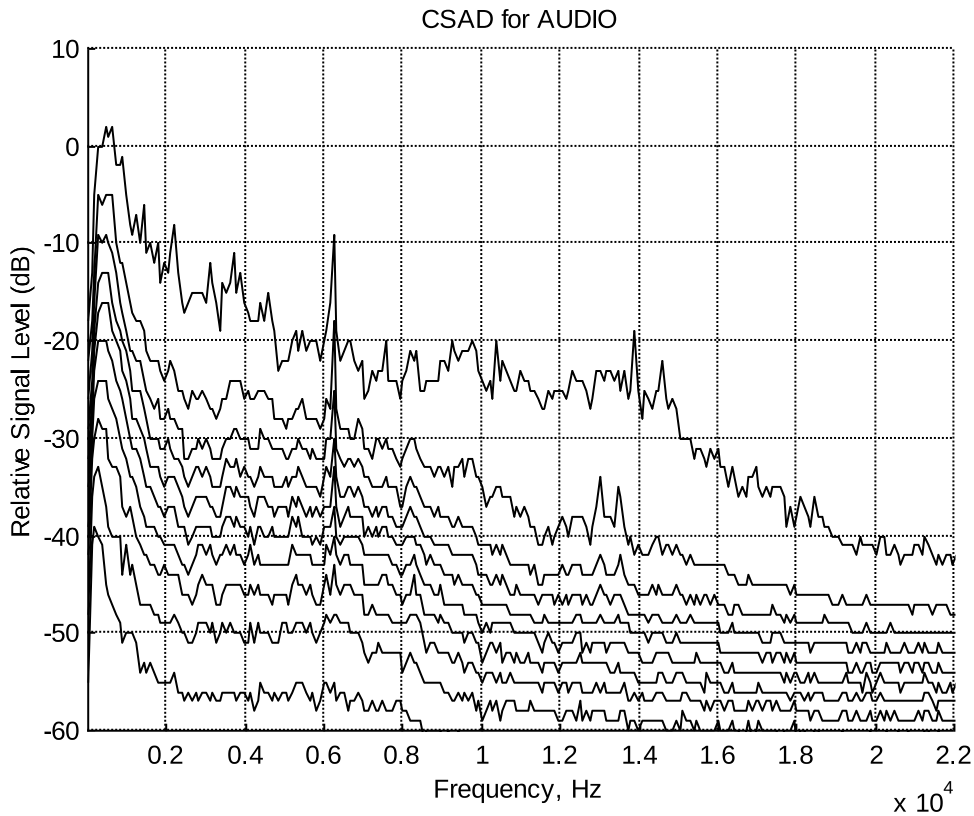


**Figure 6.3:** CSAD for comparison to Spectrogram

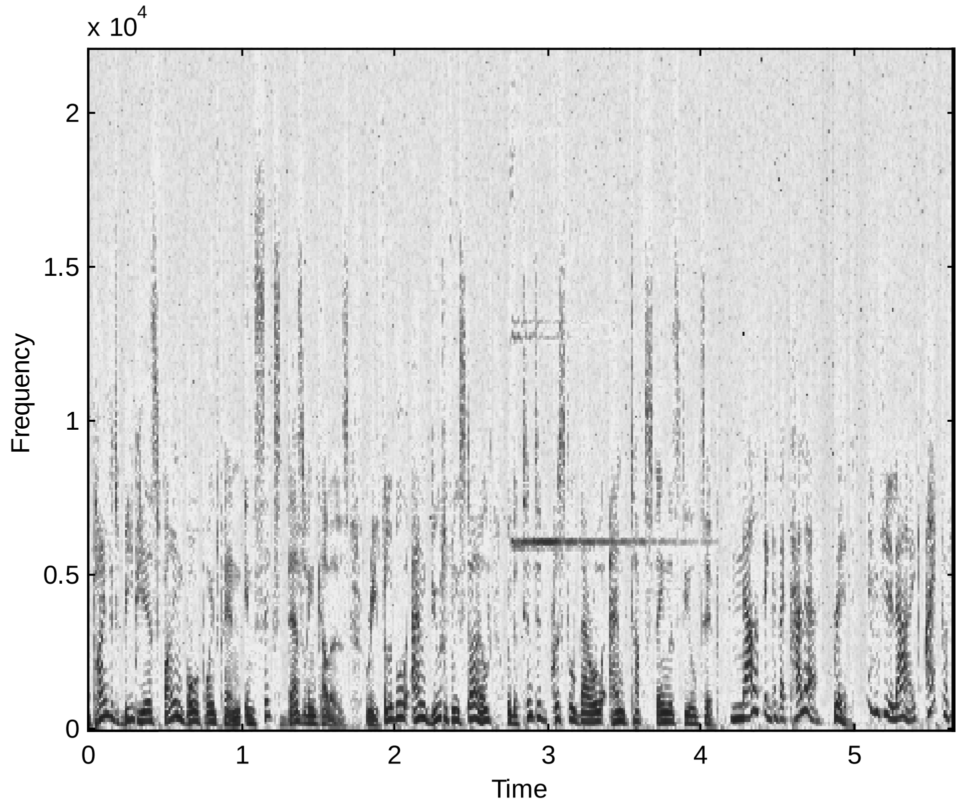


**Figure 6.4:** Spectrogram for comparison to CSAD

Note how the two representations have both identified the presence of a prolonged spectral component, known as a ‘tone’, at about 6.4 kHz with (possibly) a harmonic at 12.8 kHz. On the spectrogram this is shown as a horizontal line, at a frequency of 6.4 kHz, starting at about 2.7 seconds and lasting until just past four seconds. The harmonic can be seen as a less intense, parallel line higher in frequency. The spectrogram indicates that the tone and harmonic start at the same time and it can therefore be assumed that they originated from the same source.

The same spectral component is seen on the CSAD display as a vertical line at 6.4 kHz (Figure 6.3). Identifying the time of occurrence of a signal component is not possible with the CSAD. On the other hand, the CSAD indicates that the 6.4 kHz tone is at an amplitude that would make it audible, its peak extends to 10 dB below the highest signal component.

### 6.2.3 Signal dynamic range

The CSAD can be used to visualise the dynamic range over the frequency range of the signal [Maré, 1985] while this is not possible with the spectrogram. The dynamic range of two audio signals can be compared by plotting the amplitude difference between the 1% line and the 90% line versus frequency for the two signals. This technique has been used in determining the cause of a difference in the 'sound' of two broadcast signals [Maré, 1985]. A significant difference in dynamic range over a range of frequencies will indicate that the two signals will sound different to a listener.

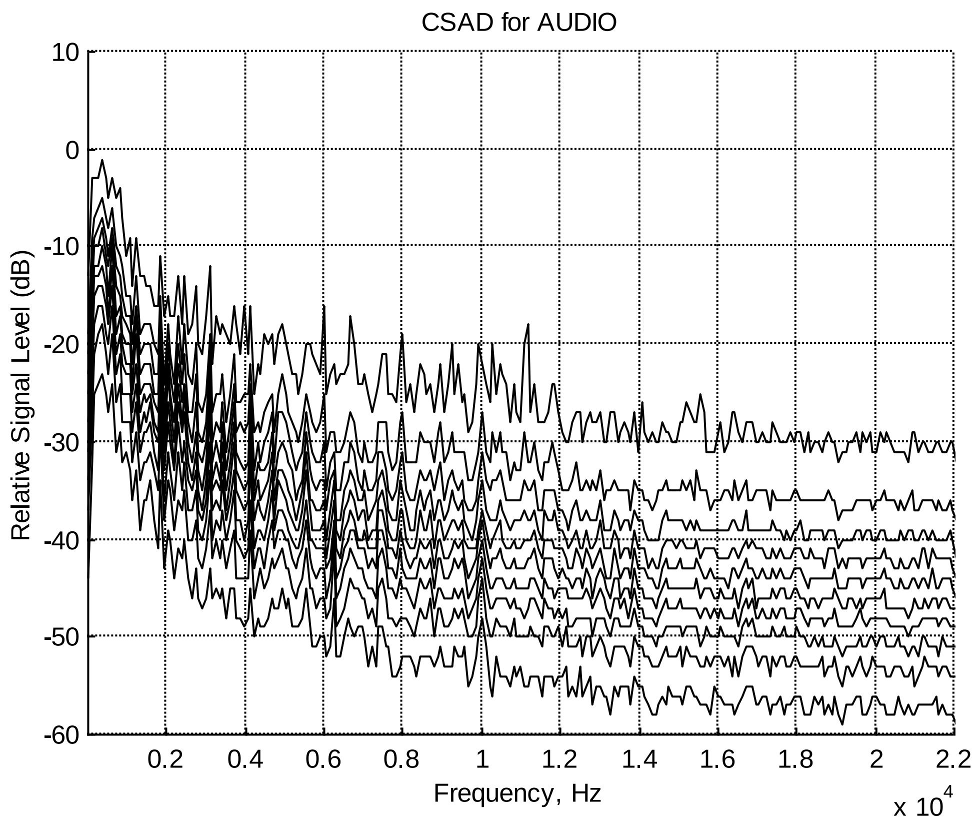


**Figure 6.4:** CSAD for compressed audio

Radio broadcast managers will shape the sound of their programmes to suit the image that they wish to convey e.g. 'pop', classical and so on. The CSAD, as a display of amplitude distribution versus frequency, can be used in the setting up of equipment like dynamic sound level controllers, to achieve the desired ‘sound’.

Figure 6.4 can be compared to Figure 6.1. The difference in amplitude distribution between the two signals can be seen most clearly at low frequencies. The difference is due to the amount of compression applied to the two signals. The signal analysed in Figure 6.4 will sound louder than that displayed in Figure 6.1. (when played back at the same maximum level).

### 6.2.4 Relative Loudness.

Audio processors are routinely used in sound studios. These machines are primarily intended to limit the peak amplitude of the audio signal before it reaches the broadcast transmitter. An audio processor will treat different frequency segments of the signal in different ways and the processing power of such a machine is often used by sound engineers to create a particular ‘sound’. For example a classical music programme requires large dynamic range and virtually no amplitude compression, while a ‘pop’ programme may be processed to increase the mean loudness of the signal.

The CSAD can be used to compare two audio signals. The amount of amplitude compression can be compared. A segment where the lines are closer together has been compressed to increase the perceived loudness of the signal. If this compression is frequency dependent the ‘sound’ of the audio signal can be shaped.

Figures A2.12 and A2.13 (Annexure 2) give an example of an audio signal that has been compressed [Dutilleux and Zölzer, 2002] compared to a signal that has not been processed. The dynamic range (1% to 95%) in Figure A2.12 (unprocessed) is larger across all frequencies than the dynamic range of the signal in Figure A2.13 (compressed)

.

### 6.2.5 The Presence of any Interfering Tones.

Continuous tones are readily identified in the CSAD as shown in Figure 6.3 above. The tone can be seen at a frequency of just more than 6 kHz.

## 6.3 Summary

The concept of the CSAD has been developed and publicized by the author. The CSAD is a graphical display of signal spectral content, signal dynamic range, signal loudness, signal amplitude versus frequency distribution, the extent of amplitude compression and the presence of prolonged spectral components. All of these parameters are indicators of audio signal quality.

The CSAD can be used by a trained observer to detect a change in the audio signal parameters mentioned above. While this may not indicate that the signal has been distorted, it could indicate that the signal has been changed in a way that would change the ‘image’ of the signal. This is of importance to the controllers of audio production facilities who may wish to project a particular image.

The CSAD is derived from the audio signal itself. It is therefore a ‘blind’ technique for monitoring quality related parameters of an audio signal.

Annexure 2 provides a comparison of a number of excerpts from the same and different audio sources. These short time CSAD plots show the variation in dynamic range, taken to be the difference between the 1 % and the 95 % lines on a CSAD plot. It is interesting to note, for example, the difference between Figures A2.5 and A2.6. Clearly the audio in Figure A2.5 has been compressed.

In Figures A2.3, female voice, compression has been used to increase the perceived loudness, but in Figure A2.4, male voice, compression is not evident.

# CHAPTER 7

## DETECTION OF NON-LINEAR DISTORTION

### 7.0 Introduction

An audio transmission system that is non-linear will introduce distortion into a signal passing through it. A characteristic of non-linear distortion is that new frequency components, that were not present in the input signal, are created in the output signal.

In this chapter it is shown that non-linear distortion changes the probability density function (pdf) of the signal so that the probability density of the output signal is not the same as that for the input signal. This is illustrated with an example.

Conventional methods for the measurement of non-linear distortion are discussed. These, stimulus-response methods, are compared to methods that make use of signal processing techniques. Some of the signal processing based methods can be classified as being blind identification methods.

The method for detection of non-linear distortion, in an audio signal, developed by the author, uses the difference between the pdf of the signal, at the output of a non-linear system, compared to the pdf of the signal, at the input to the system. The classification of a signal as being distorted or undistorted is done using a probabilistic neural network (PNN). Once the PNN has been trained it no longer requires access to the input signal in order for it to classify a signal as being distorted or undistorted.

## 7.1 Linear Systems

A system, *H*, is linear [Bendat, 1990:1] if for any inputs $x_1(t)$and $x_2(t)$, and for any constants $c_1$ and $c_2$, the following holds:

$$H[c_1x_1(t)+c_2x_2(t)] = c_1H[x_1(t)] + c_2H[x_2(t)] \tag{7.1}$$

Where:

*H* is the system transfer function:

$x_1(t)$ and $x_2(t)$ are time domain input signals

$H[x(t)]$ is the output from the system for input x(t)

$c_1$ and $c_2$ are constants

From this the two fundamental properties of linear systems follow:

a) The additive property:

$$H[x_1(t)+x_2(t)] = H[x_1(t)] + H[x_2(t)] \tag{7.2}$$

b) The homogeneity property:

$$H[cx(t)] = cH[x(t)] \tag{7.3}$$

If any of these two properties does not hold for a system then the system is non-linear. The properties of linear systems lead to the convolution integral relationship between the input and output of a linear system and its frequency domain equivalent, the frequency response:

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau \quad \textit{in the time domain, and :}$$

$$\textit{(7.4)}$$

$$Y(f) = H(f)X(f) \quad \textit{in the frequency domain.}$$

Where:

*x(t)* is the time domain input signal

*y(t)* is the time domain output signal

*h(t)* is the system impulse response

*X(f)* is the Fourier transform of x(t)

*Y(f)* is the Fourier transform of y(t)

*H(f)* is the system frequency response function

From the frequency domain expression in Equation 7.4, we see that if a particular frequency domain component is zero in the input, *X(f*), then it will also be zero in the output, *Y(f*), i.e. there are no frequency components in the output that were not present in the input. This is not the case for a non-linear system.

A linear system will not change the character of the probability density function (pdf) of the signal. If the input pdf, $p_i(x)$, is Gaussian, then the output pdf, $p_0(y)$, will also be Gaussian [Bendat,1990]. If the input/output relationship is non-linear then a Gaussian input pdf will result in a non-Gaussian output pdf. This relationship can be used for the detection of non-linearity's within the system. The bispectrum of a Gaussian distribution will be zero [Brillinger & Irizarry, 1998], hence if it is known that the input process is Gaussian and if the output bispectrum is not zero then the system between input and output must be non-linear. This relationship only applies to signals having Gaussian pdfs.

## 7.2 Non-Linear Systems

A non-linear system will have an input, *x(t)*, to output, *y(t),* relationship given by:

$$y(t) = g(x(t)) \qquad (7.5)$$

Where:

*g( )* is a single valued non-linear function of the input

*x(t)* is the time domain input signal

*y(t)* is the system time domain output signal

For a non-linear system the output pdf is related to the input pdf by [Bendat, 1990; Peebles, 1980]:

$$p_o(y) = \frac{p_i(x)}{|dy/dx|} = \frac{p_i(x)}{|g'(x)|} \qquad (7.6)$$

Where:

$p_0(y)$ is the output probability density function, pdf of *y(t)*

$p_i(x)$ is the input probability density function, pdf of *x(t)*

*g(.)* is the non-linear system function relating output to input

*The way in which a non-linear system function changes the probability distribution function of a signal is shown in the following example.*

***Application example*** *[Bendat, 1990]*

Let *x(t)* be the input signal and *y(t)* be the output signal, assume that the system is non-linear and that it has a square law response, i.e.

$y(t) = g(x(t) = x^2(t).$

Then:

$$\frac{dy}{dx} = g'(x) = 2x$$

The inverse function $g^{-1}(x) = \pm\sqrt{y}$ is bi-valued. From Equation 7.6 the output probability density function will be:

$$p_o(y) = \frac{p_i(x)}{\frac{dy}{dx}} = \frac{2p_i(x)}{|2x|} = \frac{p(\sqrt{y})}{\sqrt{y}} \quad for\ y > 0$$

For a Gaussian input pdf:

$$p_i(x) = p(\sqrt{y}) = \frac{1}{\sigma_x\sqrt{2\pi}} exp\left(\frac{-y}{2\sigma_x^2}\right) \quad for\ y > 0$$

It follows that, for the Gaussian input pdf, the output pdf will be :

$$p_o(y) = \frac{1}{\sigma_x\sqrt{2\pi y}} exp\left(\frac{-y}{2\sigma_x^2}\right) \quad for\ y > 0$$

This shows that the pdf of the output signal from the system is not the same as the pdf of the signal at the input to the system

.

## 7.3 Methods of Measuring Non-Linear Distortion

### 7.3.1 Conventional methods

Conventional methods rely on the testing of the audio equipment to ensure that the equipment does not introduce any distortion into the signal passing through. Standard tests are described in [King, 1979]. These tests are all based on the stimulus-response paradigm and provide information about the effect of the system on the signal. The results of these tests have been related to the audibility of the distortion through listening tests.

Early tests for non-linear distortion in audio equipment were ‘static’ tests. Static distortion depends only on the amplitude of the signal.

More sophisticated tests measure the dynamic response of the system. Dynamic distortion is dependent on both the amplitude and the frequency of the signal [Cordell, 1981]. A variety of test signals have been designed to measure various types of dynamic distortion, and to extract information about the form of the distortion [Hirata et al. 1981; Hueber et al. 1976; Leinonen et al. 1977]. These test have been related to the audibility of the distortion [Petri-Larmi et al. 1980].

### 7.3.2 Methods based on signal processing

These have been discussed in Chapter 4 section 4.5. The methods that have been reported in the literature involve stimulus-response testing and are not appropriate for blind detection of distortion in an audio signal.

### 7.3.3 Methods based on modelling

There are two approaches to modelling the non-linear system. The first method relies on a physical system being available for measurements to be done. The second method applies when only a distorted signal is available. In the case where a physical system is not available, time series identification techniques can be applied to identify model parameters.

The first method is illustrated in the paper by Jang and Kim (1994). A method of identifying non-linearities in loudspeakers using a NARMAX (non-linear autoregressive moving average with exogeneous input) model is described. The NARMAX model takes the form [Jang & Kim, 1994]:

$$y(t) = F^n\left\{y(t-1),\cdots,y(t-p),x(t-d),\cdots x(t-d-q),e(t-1),\cdots,e(t-r)\right\} + e(t) \tag{7.7}$$

Where:

- $x(t)$ is the input
- $y(t)$ is the output
- $e(t)$ is the prediction error
- $d$ is the system time delay
- $p$, $q$, and $r$ are the orders of input, output and error respectively
- $F^n(\ )$ is a non-linear function of order $n$

The model relies on knowledge of both the input and the output signals i.e. a physical system must be available for measurement. This method is therefore not appropriate for distortion identification when only the output signal is available.

The second approach to modelling of the non-linearity, used when a physical system is not available, is described in [Godsill et al. 1998]. Their approach relies on creating a non-linear model flexible enough to simulate the various forms of non-linear distortion likely to be present at various times in an audio signal [Godsill et al. 1998: 183]

To avoid the complexities and computational burden of the Volterra series or NARMA (non-linear autoregressive moving average) modelling approaches, Godsill et al. (1998: 184) proposed the following model, Figure 7.1. The audio signal is modelled as an autoregressive (AR) process. This process is followed by a non-linear system model. Two non-linear models are discussed in Godsill et al. (1998: pp186 -8). The first is the Autoregressive-Memoryless Non-linearity (AR-MNL) and the second is a non-linearity with memory, referred to as an Autoregressive Non-linear Autoregressive (AR-NAR) model.

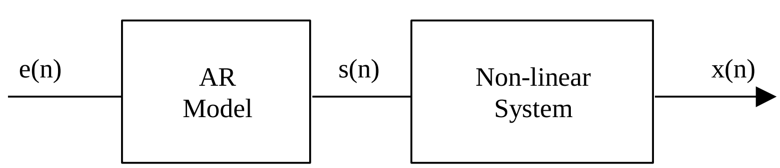


**Figure 7.1**: Model of distortion process [Godsill et al. 1998].

This technique is still in development and it may turn out that a single model for the distortion is not adequate to simulate all possible distortion mechanisms that are likely to affect a real audio signal.

## 7.4 Detection of the Presence of Non-Linear Distortion Based on the Probability Density Function (pdf) of Distorted Audio Signals

The expression for the relationship between the pdf of the signal at the input and the pdf of the signal at the output of a non-linear system, equation 7.6 [Bendat, 1990], shows that a difference will be apparent between the two pdfs.

$$p_o(y) = \frac{p_i(x)}{|dy/dx|} = \frac{p_i(x)}{|g'(x)|} \tag{7.6}$$

The difference will depend on the nature of the non-linearity. There will be no difference between input and output pdf if the system is linear.

The relationship of Equation 7.6 was used to develop a new method to identify the presence of non-linear distortion in an audio signal. The new method is based on the Probabilistic Neural Network (PNN) as described in the next section.

A neural network approach to the problem was chosen because a neural network can be trained using undistorted audio (the input signal) and distorted signals (the output from the 'distortion generator', the non-linear system) with the distorted signal being the 'target'. Once the network has been trained the input signal is no longer required.

Effectively the relationship between the undistorted audio signal and the distorted audio is memorized in the neural network weights.

### 7.4.1 Procedure for identification of the presence of non-linear distortion

The procedure must use the signal itself to detect the presence of non-linear distortion. The procedure developed here uses a property of the distorted and undistorted signals, the pdf, to distinguish between the two. In order to do this it has to be established that the difference between input and output pdf for a signal passed through a non-linear system is large enough for the difference to be detected reliably. The process is illustrated in Figure 7.2.

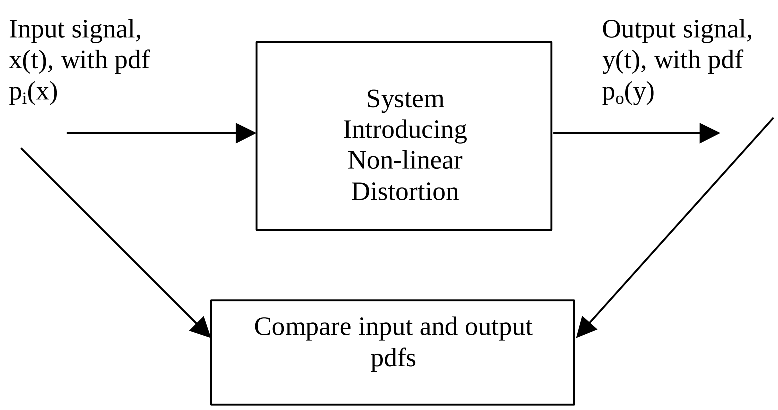


**Figure 7.2:** Process model

The non-linear distortion introduced by the system model was generated by a non-linear transfer function chosen from the limiter family [Jeruchim et al. 1992]. A limiter type of non-linearity is chosen since this is most likely to be the source of non-linearity in an audio transmission chain. The non-linearity is described by:

$$y(n) = \frac{M\, sgn(x(n))}{\left(1 + \left(u / |x(n)|\right)^{k}\right)^{\frac{1}{k}}} \qquad (7.8)$$

Where:

$y(\,)$ is the instantaneous output, generated by an input, $x(\,)$.

$n$ is the sample number.

$M$ is the asymptotic output level as $|x|$ tends to infinity

$u$ is the output limit level

$k$ is the knee sharpness parameter.

The severity of the distortion is controlled by the knee sharpness parameter, $k$, which controls the sharpness of the break from linearity. Input/output transfer curves for different values of the distortion parameters are shown in Figures A3.1 and A3.2 in Annexure 3; Appendix to Chapter 7.

From Equation 7.6 we see that in order to estimate the effect of the non-linearity on the probability distribution function, we need to find the derivative of the function describing the non-linearity, Equation 7.8. This is given by:

$$\frac{dy}{dx} = \frac{M\, sgn(x)}{\left(x^{k} + u^{k}\right)^{\frac{1}{k+1}}} \qquad (7.9)$$

The input and output pdfs can be compared if a method of describing them has been established. If a suitable describing pattern, or vector, is established an artificial neural network can be used to classify the pattern as belonging to either an undistorted audio signal or to a distorted audio signal

The approach chosen was to approximate the audio signal pdfs using histograms. The approximate distribution functions were determined from 65536 samples of the audio signal, sampled at 44.1 kHz. This is a rather arbitrary choice and any other number of samples large enough to give a good approximation to the pdf could be used. The approximate pdfs were plotted and a fifth order polynomial was fitted to the resulting curve. A fifth order polynomial was chosen since it was the lowest order polynomial giving a good fit to the estimated pdf.

The coefficients of the polynomial for the various data files are listed in Table 2.1 in Annexure 2: The appendix to Chapter 7. The coefficients were used as the describing vector for the pdfs and was input to a Probabilistic Neural Network (PNN) [Demuth and Beale, 1998; Katagiri, 2000; Fausett, 1994 ]. The PNN was used to classify the coefficient vectors as belonging to either undistorted or distorted audio.

A PNN is a neural network for pattern classification, using ideas from classical probability theory. It will classify input vectors into classes in a Bayesian optimal manner. The Bayesian decision rule states that a vector should be classified as belonging to class A if [Fausett, 1994],

$$P_A C_A p_A(x) > P_B C_B p_B(x) \qquad (7.10)$$

Where:

$P_A$ is the a priori probability of occurrence of patterns in class $A$

$C_A$ is a cost function associated with incorrectly classifying vectors

$p_A(x)$ is the probability density function for class A

$P_B$ is the a priori probability of occurrence of patterns in class $B$

$C_B$ is a cost function associated with incorrectly classifying vectors

$p_B(x)$ is the probability density function for class B

The MATLAB Neural Network Toolbox [Demuth and Beale, 1998] was used to create a PNN. The network was trained using coefficient vectors where the difference between input (undistorted) and output (distorted) was readily apparent. The PNN was then used to classify a number of data sets. Those data sets that had been incorrectly classified as either undistorted, when they were distorted. or vice versa, were then used as additional training inputs in training the PNN. This refinement of the training set resulted in the PNN's classification of data sets improving.

The PNN was further tested with data from the same audio samples but with more severe distortion present.

As a measure of the amount of distortion generated by the simulation, the Total Harmonic Distortion (THD) content of an equal amplitude sine wave signal was calculated using a formula similar to the formula specified by King (1979):

$$THD = \frac{\sqrt{(V_2^2 + V_3^2 + V_4^2 + \ldots)}}{\sqrt{(V_1^2 + V_2^2 + V_3^2 + V_4^2 + \ldots)}} \; X \quad 100\% \tag{7.11}$$

Where:

$V_n$ is the amplitude of the component at the frequency of harmonic $n$

This formula actually calculates the ratio of distortion products power to the power of input plus distortion products, it is more correctly called the 'distortion to signal-plus-distortion' ratio. The formula used was the same as Equation 7.11, except for using the distortion power and also not including the distortion components, i.e. the harmonics generated by the distortion, in the calculation below the line. This is shown in Equation 7.12. [ITT, 1973]:

$$THD = \frac{\sqrt{V_2^2 + V_3^2 + \cdots}}{\sqrt{V_1^2}} X \quad 100\% \tag{7.12}$$

## 7.5 Results

Two distortion curves were used in preparing the data. They were similar except in the knee sharpness factor, *k*, see Equation 7.8. Two different values of *k* were used, *k*=0.5 resulting in a 1.9 % distortion, and *k*=0.9, resulting in 0.5% distortion, as calculated using equation 7.12. Note that the distortion calculation is only valid for an input sine wave signal at one set amplitude. If the sine wave amplitude is changed the % THD will either increase or decrease.

The PNN was trained using 24 examples of undistorted data and 24 examples of distorted data. After training new data was presented to the PNN and the results were as follows:

| **Data Source** | **Distortion factor, k** | **Number of data segments** | **Data segment length (samples)** | **Number correctly classified as distorted** | **Number correctly classified as undistorted** |
|---|---|---|---|---|---|
| **A.** Audio tape | 0.9 | 64 | 65536 | 21 (66%) | 24 (75%) |
| **B.** Audio tape | 0.9 | 64 | 131072 | 25 (78%) | 24 (75%) |
| **C.** Audio tape | 0.5 | 38 | 65536 | 31 (81.5%) | N/A |
| **D.** Audio tape | 0.5 | 38 | 131072 | 35 (92%) | N/A |
| **D.** Audio tape | 0.5 | 38 | 196608 | 36 (94.7%) | N/A |
| **E.** Video tape | - | 8 | 196608 | 6 (75%) | N/A |

**Table 7.1:** Results obtained in classification of distorted audio signals.

Details of the data reproduced in Table 7.1 are given below.

***Data distorted with k=0.9***

**A. Data segments representing approximately 1.5 seconds of audio signal (65536 samples)**

Of the 64 examples presented to the PNN half were from undistorted data and half were from distorted data, *k*=0.9. Of the data samples 8 were wrongly classified as distorted and 11 were incorrectly classified as undistorted.

### B. Data segments representing approximately 3.0 seconds of audio signal (131072 samples)

The data samples that had been incorrectly classified as undistorted were extended in length to 131072 samples and presented to the PNN. This resulted in an additional 4 segments being correctly classified as distorted.

**Data distorted with *k*=0.5**

### C. Data segments representing approximately 1.5 seconds of audio signal (65536 samples)

Of the 38 examples presented to the PNN all were from distorted data, $k$=0.5. Of the data 7 were incorrectly classified as undistorted.

### D. Data segments representing approximately 3.0 and 4.5 seconds of audio signal (131072 and 196608 samples)

The data samples that had been incorrectly classified as undistorted were extended in length to firstly 131072 samples and presented to the PNN. This resulted in an additional 4 segments being correctly classified as distorted. When the data segment was further extended to 196608 samples an additional segment was correctly classified as distorted.

### E. Data segment from a VHS video tape with distorted audio track

Eight data segments, 196608 samples long, were taken from the videotape. When presented to the PNN, 6 were correctly classified as distorted and two were incorrectly classified.

## 7.6 Discussion of Results

The probability density of signal amplitude (pdf) and the signal cumulative spectral amplitude distribution (CSAD) will vary depending on which time segment of the audio signal is being analysed, an audio signal is not stationary. Both the pdf and CSAD will also vary depending on the source of the audio signal, e.g. speech or music. The shape of these density curves will indicate whether the signal can be classified as having undergone some form of non-linear distortion or not. The figures below show typical histogram (approximation to the pdf) curves for two signal examples.

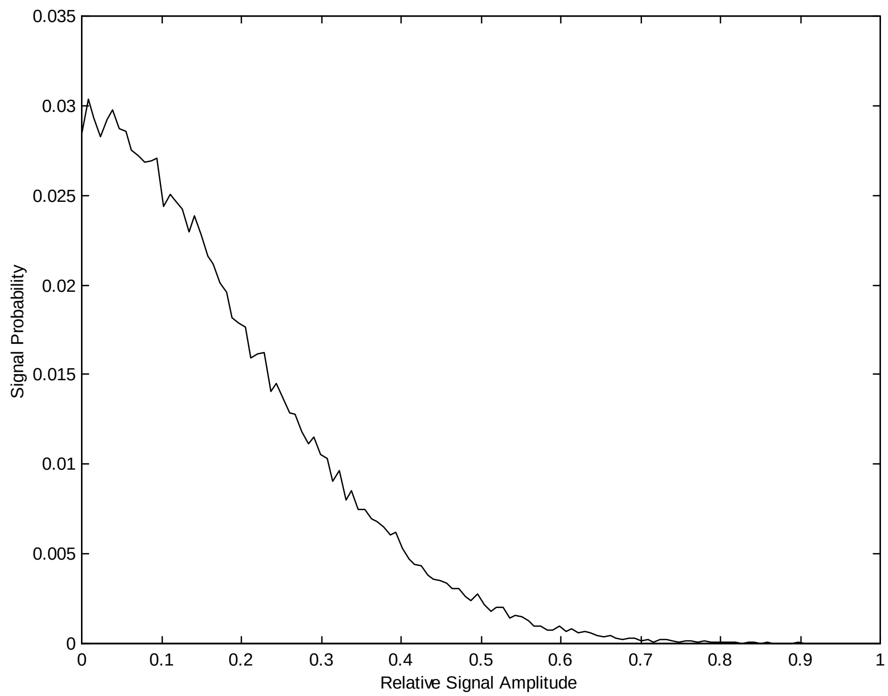


**Figure 7.3:** Signal amplitude histogram curve for signal with adequate amplitude spread.

In Figure 7.3 the histogram (approximation to the pdf) of a segment, 65536 samples long, of an input signal, i.e. a signal that has not been distorted, is shown. After passing this signal through the simulated limiter (Section 7.4.1 Equation 7.8), the Probabilistic Neural Network (PNN) correctly classified the output signal as having been distorted.

The histogram for a second segment of input signal is shown in Figure 7.4. After passing this signal through the limiter the PNN could not correctly classify the signal as having been distorted.

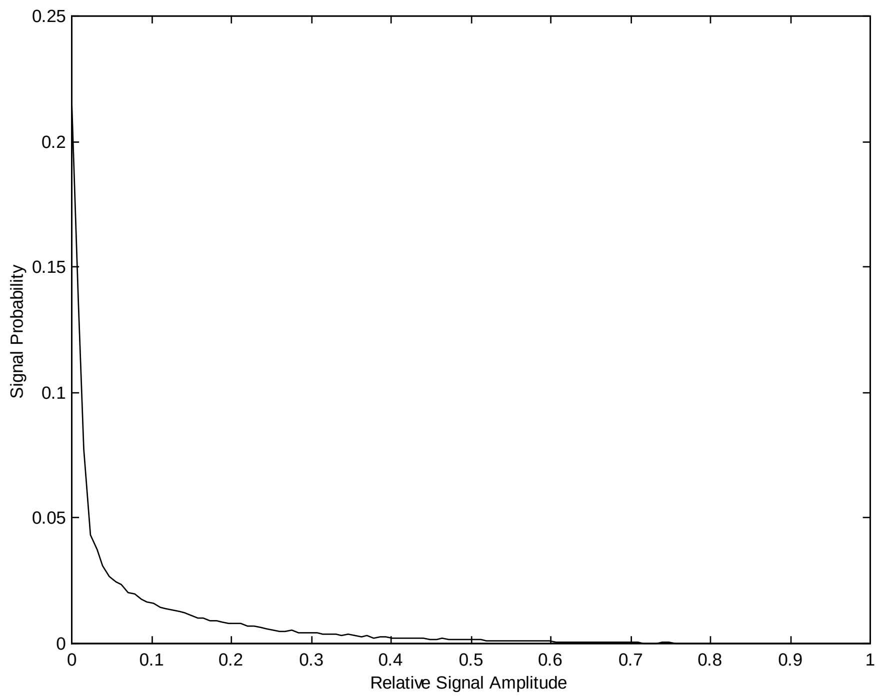


**Figure 7.4:** Amplitude histogram curve for signal with inadequate amplitude spread.

Comparing the histogram of Figure 7.3 to that of Figure 7.4 it is clear that the audio signal in Figure 7.4 is less likely to have relatively large amplitude components. The amount of distortion present in this signal after passing through the distorter will therefore be less than that fro the signal shown in Figure 7.3 and hence detection of distortion may fail simply because there is no distortion to be detected.

The histogram shown in Figure 7.4 indicates that the signal segment did not have adequate high amplitude components for the distortion to be detectable. The model of a limiter (Equation 7.8 and Figures A3.1 and A3.2 in Annexure 3, page 139 - 140) shows that low amplitude components are passed without distortion while high amplitude components are distorted. The input/output transfer curves of figures A3.1 and A3.2 in Annexure 3 indicate that there is a deviation from the linear curve which increases progressively as the input amplitude increases.

This is further explained by reference to the following figures, Figure 7.9, Figure 7.10, Figure 7.11 and Figure 7.12.

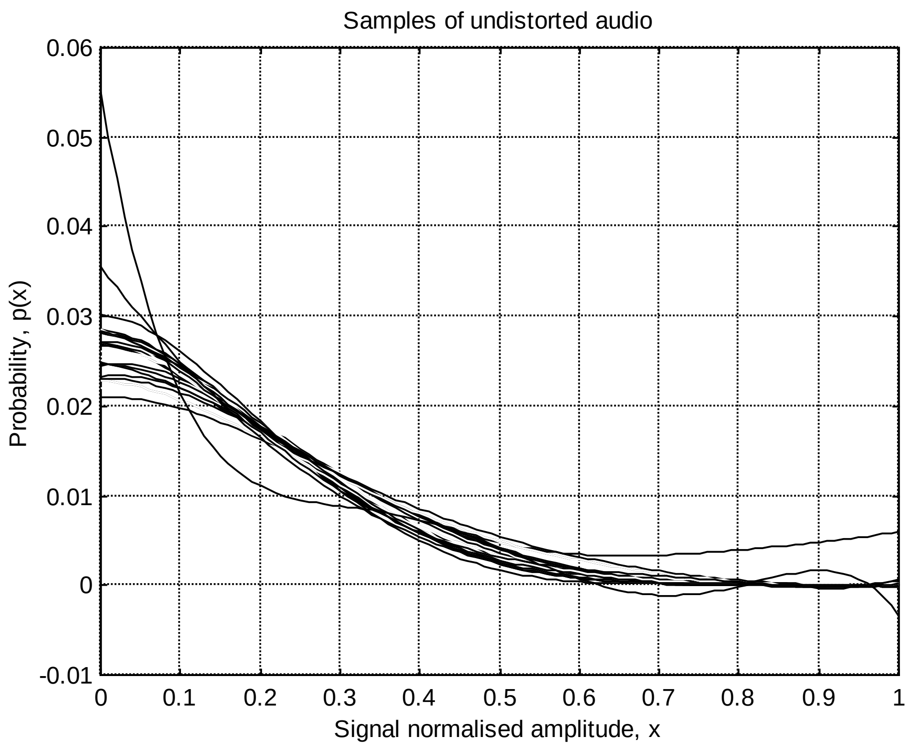


**Figure 7.9**: Histograms of undistorted audio.

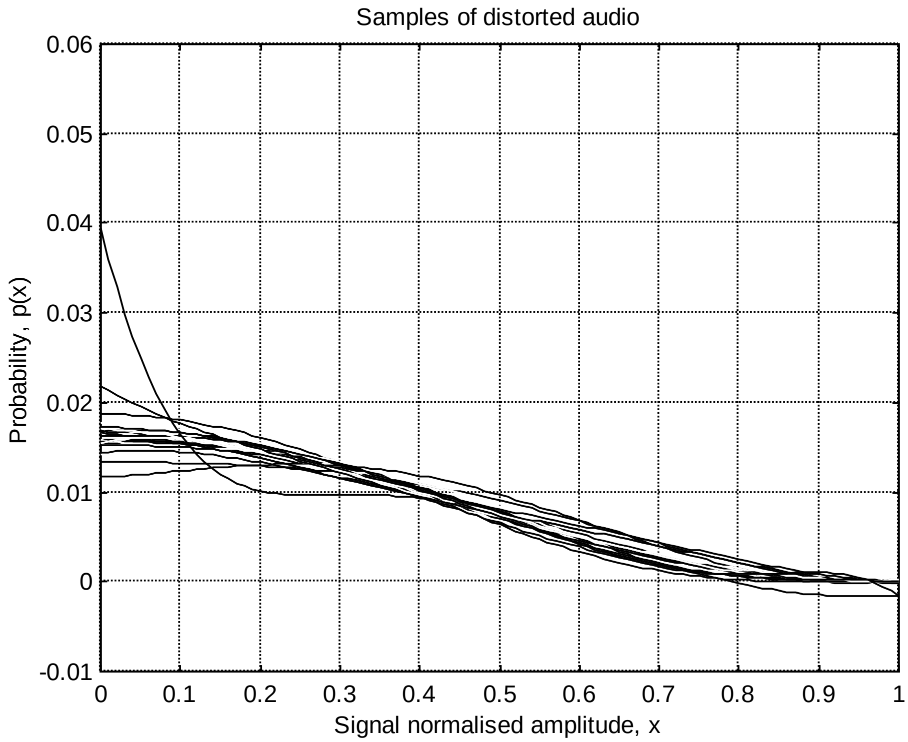


**Figure 7.10:** Histograms of distorted audio.

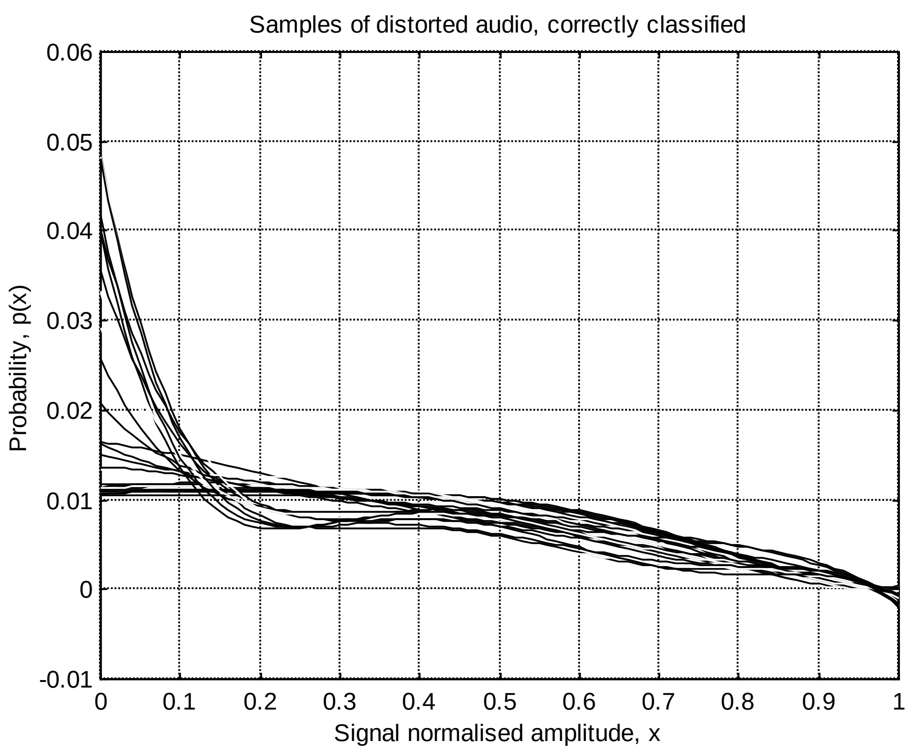


**Figure 7.11:** Histograms of distorted audio correctly classified as distorted.

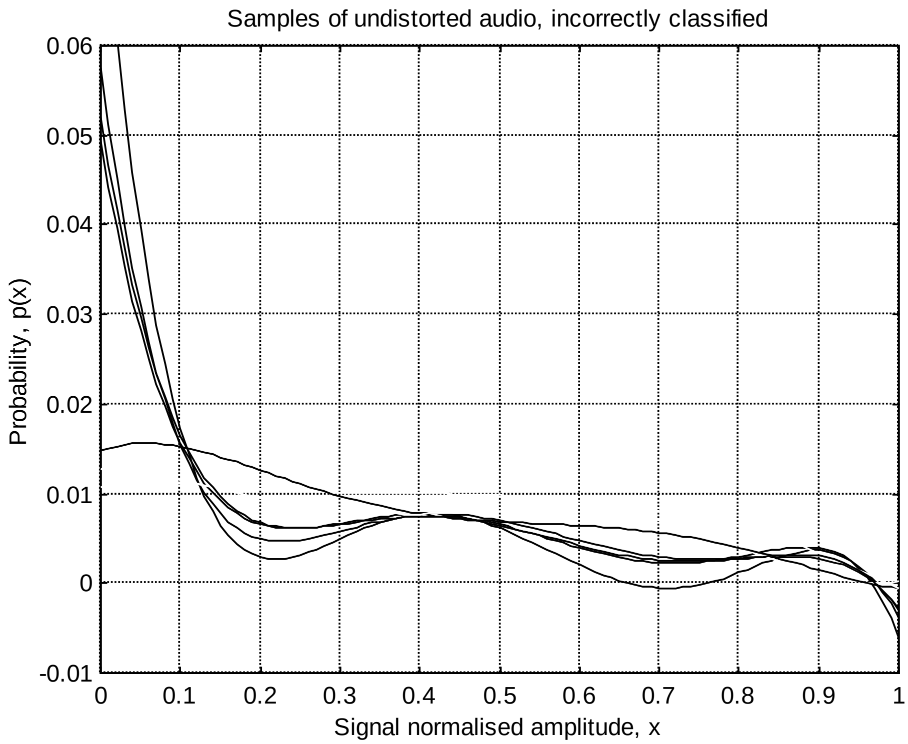


**Figure 7.12:** Histograms of audio incorrectly classified as distorted.

Comparing Figures 7.9 to 7.12 it becomes clear that the histogram for signal amplitudes (normalized) between about 0.3 and 0.7 is crucial to the classification process. A bulge in this area leads to a classification of the signal as being distorted. Audio signals have an approximately exponential distribution, the histogram would therefore tend to be concave. Redistribution of signal amplitudes due to distortion changes this shape.

It must be accepted that the shape of the histogram can only change if the undistorted signal is rich in signal components with amplitudes across the range of possible amplitudes. Amplitude distortion rarely affects signals with small amplitudes and hence such signals will pass through a non-linear device without exhibiting any signs of distortion.

Tables 7.2 and 7.3 below support this argument. In these tables the classification of the audio sample is indicated in the column headed Class 6. A 1 in this column indicates an incorrect classification. Audio data files which were incorrectly classified as undistorted when subjected to lower distortion ($k$=0.9, Table 7.2) would either not change (Dat05; Dat06; Dat07>; Dat15; Dat15>; Dat17 ) or would be correctly classified (Dat04; Dat05; Dat06>; Dat17>; Dat18; Dat19; Dat19>) when subjected to higher levels of distortion ($k$=0.5, Table 7.3).

The column headed Class 1 in Tables 7.2 & 7.3 show the results of using the PNN to classify each coefficient of the polynomial individually and then use majority logic to arrive at an overall classification. A 0 indicates an equal vote for distorted and undistorted, a 1 indicates an incorrect classification and a 2 indicates a correct classification. Examination of the tables shows that this method does not improve on the previous method for low distortion and provides a marginal improvement for higher distortion.

It can therefore be concluded that the proposed method for detection of non-linear distortion is capable of detecting non-linear distortion in an audio signal within the limitations discussed above. After training the PNN the method can operate as a blind technique for detecting non-linear distortion.

**Table 7.2:** Audio data files and associated parameters, k=0.9

## Distortion parameters: m=1; u=1.5; k=0.9

| Data File | 5th | 4th | 3rd | 2nd | 1st | Oth | Class 6 | Class 1 |
|---|---|---|---|---|---|---|---|---|
| **Dat01** | 0.0401 | -0.1328 | 0.1939 | -0.1281 | 0.0104 | 0.0167 | 2 | 2 |
| **Dat01 >** | 0.0356 | -0.0772 | 0.0878 | -0.0586 | -0.0044 | 0.0171 | 2 | 2 |
| **Dat02** | -0.1199 | 0.2821 | -0.1802 | 0.0057 | -0.007 | 0.0174 | 2 | 2 |
| **Dat02 >** | 0.0773 | -0.2375 | 0.2949 | -0.1695 | 0.0193 | 0.0155 | 2 | 2 |
| **Dat03** | -0.1756 | 0.4671 | -0.4252 | 0.1673 | -0.555 | 0.0219 | 2 | 2 |
| **Dat03 >** | -1.184 | 3.3789 | -3.5551 | 1.6861 | -0.367 | 0.0395 | 2 | 2 |
| **Dat04** | 0.2091 | -0.5806 | 0.5832 | -0.2468 | 0.019 | 0.0161 | 1 | 1 |
| **Dat04 >** | 0.1293 | -0.3538 | 0.3244 | -0.0826 | -0.0415 | 0.0241 | 2 | 0 |
| **Dat05** | -1.4195 | 4.0756 | -4.3805 | 2.181 | -0.5123 | 0.0534 | 1 | 1 |
| **Dat05 >** | -3.0158 | 8.4416 | -8.7626 | 4.1295 | -0.8682 | 0.0702 | 1 | 1 |
| **Dat06** | -2.4225 | 6.8154 | -7.1087 | 3.3676 | -0.7178 | 0.0619 | 1 | 1 |
| **Dat06 >** | -2.0679 | 5.8763 | -6.228 | 3.0329 | -0.6793 | 0.0627 | 1 | 1 |
| **Dat07** | -1.3462 | 3.7985 | -3.9683 | 1.8849 | -0.4144 | 0.0434 | 2 | 2 |
| **Dat07 >** | -4.4812 | 12.4953 | -12.891 | 6.0228 | -1.2471 | 0.0943 | 1 | 1 |
| **Dat08** | -0.0771 | 0.1926 | -0.1332 | 0.0092 | -0.0077 | 0.0162 | 2 | 2 |
| **Dat08 >** | 0.0025 | -0.0663 | 0.1709 | -0.133 | 0.0075 | 0.0185 | 2 | 2 |
| **Dat09** | -0.0348 | 0.0781 | -0.014 | -0.0493 | 0.0043 | 0.0158 | 2 | 2 |
| **Dat09 >** | -0.1143 | 0.2871 | -0.2045 | 0.0196 | -0.0038 | 0.0159 | 2 | 2 |
| **Dat10** | -0.0163 | 0.0491 | -0.0124 | -0.0357 | -0.005 | 0.016 | 2 | 2 |
| **Dat10 >** | -0.0554 | 0.1686 | -0.1437 | 0.019 | -0.0026 | 0.0143 | 2 | 2 |
| **Dat11** | -0.0702 | 0.1724 | -0.1067 | -0.0071 | -0.0047 | 0.0165 | 2 | 2 |
| **Dat11 >** | -0.1235 | 0.3357 | -0.2841 | 0.0688 | -0.0126 | 0.0156 | 2 | 2 |
| **Dat12** | -0.1063 | 0.18 | -0.011 | -0.0843 | 0.002 | 0.0194 | 2 | 2 |
| **Dat12 >** | -0.0231 | 0.1052 | -0.1413 | 0.0846 | -0.046 | 0.0207 | 2 | 2 |
| **Dat13** | 0.0021 | -0.0277 | 0.0912 | -0.0878 | 0.0056 | 0.0167 | 2 | 2 |
| **Dat13 >** | -0.011 | 0.0901 | -0.1193 | 0.0436 | -0.0197 | 0.0168 | 2 | 2 |
| **Dat14** | -0.4096 | 1.1315 | -1.1709 | 0.5692 | -0.1522 | 0.0284 | 2 | 2 |
| **Dat14 >** | -1.3727 | 3.9424 | -4.2057 | 2.0499 | -0.4638 | 0.048 | 2 | 0 |
| **Dat15** | 0.4439 | -1.0901 | 0.9032 | -0.2548 | -0.0247 | 0.0233 | 1 | 1 |
| **Dat15 >** | 0.1505 | -0.2641 | 0.0639 | 0.0916 | -0.0611 | 0.0196 | 1 | 1 |
| **Dat16** | -0.8032 | 2.3239 | -2.4904 | 1.2122 | -0.2782 | 0.0345 | 2 | 2 |
| **Dat16 >** | 0.0318 | -0.0525 | -0.0216 | 0.084 | -0.0653 | 0.023 | 2 | 2 |
| **Dat17** | -2.4339 | 6.8717 | -7.2108 | 3.4518 | -0.748 | 0.065 | 1 | 1 |
| **Dat17 >** | -1.8064 | 5.2328 | -5.6877 | 2.8642 | 0.67 | 0.0643 | 1 | 1 |
| **Dat18** | -2.0511 | 5.7471 | -5.9323 | 2.7503 | -0.5675 | 0.0502 | 1 | 1 |
| **Dat18 >** | -1.3037 | 3.7227 | -3.9343 | 1.8892 | -0.4199 | 0.0442 | 2 | 2 |
| **Dat19** | -2.3015 | 604728 | -6.7433 | 3.1949 | -0.6881 | 0.0616 | 1 | 1 |
| **Dat19 >** | -2.042 | 5.7437 | -5.9643 | 2.7959 | -0.5896 | 0.0532 | 1 | 1 |

**Table 7.3:** Audio data files and associated parameters

## Distortion parameters: m=1; u=1.5; k=0.5

| Data File | 5th | 4th | 3rd | 2nd | 1st | oth | Class 6 | Class 1 |
|---|---|---|---|---|---|---|---|---|
| **Dat01** | 0.1099 | -0.2415 | 0.1889 | -0.0817 | 0.0132 | 0.112 | 2 | 2 |
| **Dat01 >** | 0.1716 | -0.3837 | 0.2889 | -0.0991 | 0.0108 | 0.0114 | 2 | 2 |
| **Dat02** | 0.0996 | -0.1684 | 0.0717 | -0.0148 | 0.0003 | 0.0117 | 2 | 2 |
| **Dat02 >** | 0.0386 | -0.0847 | 0.0746 | -0.0523 | 0.0129 | 0.0107 | 2 | 2 |
| **Dat03** | 0.046 | -0.0243 | -0.0746 | 0.0631 | -0.0249 | 0.0151 | 2 | 2 |
| **Dat03 >** | -0.4299 | 1.3956 | -1.6941 | 0.9288 | -0.2302 | 0.0293 | 2 | 2 |
| **Dat04** | -0.0275 | 0.0017 | 0.0659 | -0.0634 | 0.0113 | 0.011 | 2 | 2 |
| **Dat04 >** | 0.0204 | -0.0944 | 0.1278 | -0.062 | -0.0092 | 0.0164 | 2 | 2 |
| **Dat05** | -0.8516 | 2.4612 | -2.6923 | 1.3803 | -0.3394 | 0.0401 | 2 | 2 |
| **Dat05 >** | -2.2686 | 6.3565 | -6.6341 | 3.162 | -0.6773 | 0.0576 | 1 | 1 |
| **Dat06** | -1.6313 | 4.6465 | -4.9456 | 2.4107 | -0.5324 | 0.0493 | 1 | 1 |
| **Dat06 >** | -1.2564 | 3.6268 | -3.939 | 1.9869 | -0.4688 | 0.0486 | 2 | 2 |
| **Dat07** | -0.7644 | 2.2436 | -2.468 | 1.2419 | -0.2874 | 0.0322 | 2 | 2 |
| **Dat07 >** | -3.466 | 9.7044 | -10.0772 | 4.7523 | -0.9983 | 0.0787 | 1 | 1 |
| **Dat08** | 0.0769 | -0.1385 | 0.0633 | -0.0133 | 0.0006 | 0.011 | 2 | 2 |
| **Dat08 >** | -0.0789 | -0.1512 | 0.1046 | -0.0517 | 0.0017 | 0.0125 | 2 | 2 |
| **Dat09** | 0.0914 | -0.1683 | 0.0906 | -0.0308 | 0.0062 | 0.0107 | 2 | 2 |
| **Dat09 >** | 0.1377 | -0.2676 | 0.1545 | -0.0401 | 0.005 | 0.0107 | 2 | 2 |
| **Dat10** | 0.1666 | -0.3615 | 0.259 | -0.0861 | 0.0114 | 0.0106 | 2 | 2 |
| **Dat10 >** | 0.1315 | -0.2714 | 0.1666 | -0.043 | 0.066 | 0.0096 | 2 | 2 |
| **Dat11** | 0.0941 | -0.1723 | 0.0878 | -0.0233 | 0.0025 | 0.0111 | 2 | 2 |
| **Dat11 >** | 0.1379 | -0.2628 | 0.1399 | -0.0275 | 0.0022 | 0.0105 | 2 | 2 |
| **Dat12** | -0.0003 | 0.0723 | -0.1149 | 0.0364 | -0.0065 | 0.0133 | 2 | 2 |
| **Dat12 >** | 0.1901 | -0.4359 | 0.3277 | -0.0926 | -0.0028 | 0.0137 | 2 | 2 |
| **Dat13** | 0.0802 | -0.1554 | 0.096 | -0.0381 | 0.0059 | 0.0113 | 2 | 2 |
| **Dat13 >** | 0.2875 | -0.653 | 0.4916 | -0.151 | 0.0143 | 0.0108 | 1 | 0 |
| **Dat14** | -0.2827 | 0.8024 | -0.8517 | 0.412 | -0.1014 | 0.0207 | 2 | 2 |
| **Dat14 >** | -0.5284 | 1.6742 | -2.0037 | 1.107 | -0.2851 | 0.0355 | 2 | 2 |
| **Dat15** | 0.3775 | -1.0344 | 0.991 | -0.3831 | 0.0337 | 0.0148 | 1 | 1 |
| **Dat15 >** | 0.2433 | -0.6513 | 0.5611 | -0.166 | -0.0003 | 0.0127 | 1 | 0 |
| **Dat16** | -0.461 | 1.3622 | -1.5305 | 0.792 | -0.1899 | 0.0257 | 2 | 2 |
| **Dat16 >** | -0.1622 | 0.37 | -0.3224 | 0.1403 | -0.0432 | 0.0163 | 2 | 2 |
| **Dat17** | -1.7496 | 4.9437 | -5.2212 | 2.5326 | -0.5606 | 0.052 | 1 | 1 |
| **Dat17 >** | -0.9405 | 2.773 | -3.1193 | 1.6617 | -0.4248 | 0.0482 | 2 | 2 |
| **Dat18** | -1.2439 | 3.6168 | -3.9204 | 1.9254 | -0.4197 | 0.0398 | 2 | 2 |
| **Dat18 >** | -0.6481 | 1.9674 | -2.2435 | 1.1709 | -0.2811 | 0.0334 | 2 | 2 |
| **Dat19** | -1.3457 | 3.9238 | -4.2751 | 2.1337 | -0.4865 | 0.048 | 2 | 0 |
| **Dat19 >** | -1.2116 | 3.5392 | -3.859 | 1.9151 | -0.4271 | 0.0418 | 2 | 2 |

# CHAPTER 8

## CONCLUSION

### 8.0 Introduction

The conventional approach to quality maintenance in high quality audio signal transmission and production equipment is to use stimulus response testing. The performance of the *equipment*, through which the audio signal passes, is measured using standard test signals and if the performance does not meet laid down criteria the equipment is taken out of service and adjusted or repaired.

The objective of this research project has been to consider the possibility of blind identification of distortions in high quality audio signals. This would enable a network operator to monitor the audio signals passing through the network and to determine whether a signal has been distorted.

Distortion could be caused by the source, e.g. a broadcast studio, or the transmission network. The operator of a broadcast service would monitor the signal at the output of the studio complex. This would enable the operator to identify signal distortions and to initiate corrective measures.

It has been shown that using the techniques developed during this research project the following distortions can be identified:

Distortions resulting from mechanical imperfections in analogue recording and playback apparatus (wow and flutter).

Distortions due to restricted dynamic range, frequency content and frequency range of an audio signal.

The presence of non-linear distortion in an audio signal.

## 8.1 Distortion Due to Mechanical Imperfections

The mechanical imperfections that are of interest are those giving rise to variations in rotational speed in analogue playback and recording equipment. These variations result in frequency modulation of the recorded or reproduced audio signal.

It has been shown that the frequency modulation of the audio signal (known as wow and flutter) can be detected. The procedure for detection of wow and flutter is based on the autocorrelation of the signal with refinements to enhance the visibility of the detected impairment.

Detection of the presence of wow and flutter distortion is not possible for all time segments of an audio signal. When the audio signal does not contain a strong component at a particular frequency the distortion cannot be detected as frequency modulation of a random signal does not lead to clearly defined periodicities, which could be detected using the autocorrelation.

## 8.2 Dynamic Range, Frequency Content and Frequency Range

Dynamic range, frequency content and frequency range refer to parameters of the audio signal which influence the sound and the impression this creates in the human auditory system. The operators of broadcast facilities manipulate these parameters to shape the sound image that they wish to project for their particular broadcast programme. This also applies to recording artists. For example classical music is usually broadcast or recorded with a high dynamic range and little or no processing of the audio signal in terms of manipulation of the dynamic range, frequency content or frequency range. In contrast, sound advertisements are heavily processed to increase the perceived loudness and to create the desired sound image.

By monitoring the dynamic range, frequency content and frequency range of a sound programme, the broadcast programme manager can be assured that the programme projects the desired sound image.

The CSAD (cumulative spectral amplitude distribution) has been developed, by the author, as a tool to aid in monitoring audio signal dynamic range, frequency content and frequency range. The CSAD is generated by creating a plot of the cumulative amplitude distribution of the audio signal at different frequencies. It is similar to the spectrogram in that the statistics of a number of consecutive, short time FFTs are accumulated. However it has the advantage over the spectrogram in that the CSAD shows the amplitude distribution of the audio signal, and hence the dynamic range of the signal at all frequencies can be deduced from the CSAD diagram.

### 8.3 Distortion Due to Non-Linear Effects

Audio signals are often generated by sources which are rich in harmonics and which do not have a Gaussian probability distribution. This means that tests for non-linearity will indicate that there is non-linearity present in the signal, not because the signal has been distorted but because the apparent non-linearity is inherent in the signal.

The approach to detecting non-linear distortion developed in this research project has relied on the change in probability distribution of a signal between the input and output when the signal is passed through a non-linear system. If this difference is large enough an artificial neural network can be trained to detect this difference and classify the signal as either undistorted or distorted.

The parameters describing the probability distribution of the signal that have been chosen as inputs to the neural network are the coefficients of a fifth order polynomial fit to the probability distribution curve. It has been shown that using the five coefficients as an input vector it, is possible to train a probabilistic neural network to classify a signal according to whether it has been distorted or not. When the input signal does not have high amplitude components the neural network will classify the signal as undistorted. This technique classifies the signal and makes no statement about the possible cause of the distortion.

## 8.4 Future Work

The purpose of audio measurements is to enable a system operator/designer to detect distortions, which would detract from the listening experience of a user of the system. Only a small segment of possible distortions which may give rise to impairment of the audio signal have been considered in this research.

The work reported on here, has been limited to the detection of distortion in an analogue audio signal. In a digital transmission system the distortion is limited to the effects of bit errors. However changes in the way a signal sounds to a human listener may result from the signal processing applied to the digitized analogue signal. Such processing may have the aim of reducing redundancy thereby limiting the occupied bandwidth. When signal processors are used in series, e.g. one processing technique may be used over a satellite link and another over a terrestrial microwave link, the combined effect may result in unwanted signal distortion.

Future research should consider advances in psychoacoustics to refine and extend the work that has been reported on here to digital transmission systems.

## REFERENCES CITED

*AES Standard Method for Measurement of Weighted Peak Flutter of Sound Recording Equipment*, AES Standard AES6-1982 (r2003)

Allnatt, J. 1983. *Transmitted-Picture Assessment*, New York: John Wiley & Sons.

Atkin, G. 1991. *Sound Recording and Production*, London: Focal Press.

Beerends, J.G. 1998*. Audio quality determination based on perceptual measurement techniques. In:* Kahrs, M. and Brandenburg, K. (eds). *Applications of Digital Signal Processing to Audio and Acoustics.* Boston: Kluwer Academic Publishers, pp 1-38.

Beerends, J.G. and Stemerdink, J. 1992 A. "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation", *J. Audio Eng. Soc.*, vol 40, pp 963-978.

Beerends, J.G. and Stemerdink, J. 1992 B. "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation", *J. Audio Eng. Soc.*, vol 40, page 968

Bendat, J.S. 1990. *Non-linear System Analysis and Identification from Random Data*. New York: John Wiley & Sons.

Bendat, J.S. and Piersol, A.G. 1971.*Random Data: Analysis and measurement Procedures,*New York: Wiley-Interscience.

Benson, D.J. Mathematics and Music (Web version).
Available: http://www.math.uga.edu/¬djb/html/math-music.html

Brillinger, D.R. & Irizarry, R.A. 1998. “An investigation of the second- and higher-order spectra of music”, *Signal Processing*, Vol 65, part 2, pp 161-179.

Carroll, T. (undated). “ DTV and ‘The Annoyingly Loud Commercial problem’.” Available: http://www.tvtechnology.com/features/audio_notes/f-TC-dtv.shtml

Cartwright, H.E., González, D.L. and Piro, O. 1999. “A New Nonlinear Model for Pitch Perception.” In Reguera, D. Rubi, R. and Vilar, J. (eds) *Statistical Mechanics of Biocomplexity,* Lecture Notes in Physics, vol. 527, pp 205 – 216, Springer, 1999.

CCIR, *Characteristics of Signals Sent Over Sound-Programme Circuits*. CCIR Report 491-2.

Cohen, M.A., Grossberg, S. and Wyse, L.L. 1995. “A spectral network model of pitch perception.” J. Acoust. Soc. Am., 98, pp 862 – 878. Cited in: Cartwright, H.E., González, D.L. and Piro, O. 1999. “A New Nonlinear Model fro Pitch Perception.” In

Reguera, D. Rubi, R. and Vilar, J. (eds) *Statistical Mechanics of Biocomplexity,* Lecture Notes in Physics, vol. 527, pp 205 – 216, Springer, 1999.

Cordell, Robert R. 1983. “A Fully In-Band Multitone Test for Transient Intermodulation Distortion”, *J.Audio Eng. Soc*., Vol. 29, No. 9, pp 578 – 586.

Demuth, H. & Beale, M. 1998. *Neural Network Toolbox User’s Guide: For Use with Matlab* (Fifth Printing-Version 3), The MathWorks

Duetsche Welle. (Accessed September 2004), *Quality Parameters in Audio*. Available: http://www9.dw-world.de/rtc/infotheque/qual-parameter/qualpar.pdf

Dutilleix, P. and Zölzer, U. 2002, “Nonlinear Processing” Chapter 5 in Zölzer, U. (ed) 2002. *DAFX Digital Audio Effects*. Chichester, John Wiley & Sons.

Ehara, S. January 1977. *Some Applications of Programme-Model Signals*. NHK Laboratories, Note No. 207.

Ehara, S. January 1982. *Spectral Amplitude and Power Distribution of Sound Programme Signals*. NHK Laboratories, Note No. 269.

Fausett, L. 1994. *Fundamentals of Neural networks: Architectures, algorithms, and applications.* New Jersey: Prentice Hall.

Fletcher, H. 1942. “Hearing, the Determining Factor for High Fidelity Transmission”, Proc. IRE, vol. 30, pp266 – 277, June 1942 [Cited in: Fielder, L.D. *Dynamic Range Issues in the Modern Digital Audio Environment.*
Available: http://www.zainea.com/Dynamic%2Orange.htm]

Fielder, L.D. *Dynamic Range Issues in the Modern Digital Audio Environment.*
Available: http://www.zainea.com/Dynamic%2Orange.htm

Godsill, S. and Rayner, P. 1998. *Digital Audio Restoration*. Berlin: Springer-Verlag.

Godsill, S., Rayner, P. & Cappé, O. 1998 . *Digital Audio Restoration*. *In:* Kahrs, M and Brandenburg K (eds). *Applications of Digital Signal Processing to Audio and Acoustics.* Boston: Kluwer Academic Publishers, pp 133 – 194.

Gold, B. and Morgan, N. 2000. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, New York, John Wiley & Sons, Inc.

Graham, G, “Whatever happened to dynamic range on compact discs?”
Available: http://georgegraham.com/compress.html

Granqvist, S. and Hammarberg, B. (undated). “The Correlogram: A visual display of periodicity.”

Available: http://www.speech.kth.se/~svante/Thesis/paperll.pdf

Gray, P., Hollier, M.P. and Massara, R.E. 2000, "Non-intrusive Speech Quality Assessment Using Vocal Tract Models", IEE Proc. Visual Image Signal Process., Vol. 147, No. 6, pp493 – 502. ( Cited in: Picovici, D. and Mahdi, A.E. (undated), "Towards Non-intrusive Speech Quality Assessment for Modern telecommunications".
Available: http://telecomms.eeng.dcu.ie/symposium/papers/F2.pdf)

Hartmann, W.M. 1998. *Signals, Sound, and Sensation*, New York: Springer-Verlag Inc., page 455

Haykin, S. 2001, *Communication Systems (4th ed.)* New York: John Wiley & Sons.

Heyser, R.C. 1976 A. "Perspectives in Audio Analysis: Changing the Frame of Reference, Part 1", *J. Audio Eng. Soc.,* vol 24, pp 660 - 667.

Heyser, R.C. 1976 B. "Perspectives in Audio Analysis: Changing the Frame of Reference: Part 2", *J. Audio Eng. Soc.,* vol 24, pp 742 - 751.

Hinich, M.J. & Patterson, D.M. 1985. "Evidence of Nonlinearities in Daily Stock Returns", *J. Business & Economic Statistics",* vol 3, pp69-77.

Hirata, Y., Ueki, M. & Kasuga, T. 1981. "Nonlinear Distortion Measurement Using Composite Pulse Waveform", *J. Audio Eng. Soc*., Vol. 29, No. 4, pp243 – 248.

Hollier, M.P., Hawksford, M.O.J. & Guard, D.R. 1993. "Characterisation of Communication Systems Using a Speechlike Test Stimulus", *J. Audio Eng. Soc.,*vol 41, pp 1008-1020.

Hollier, M.P., Hawksford, M.O.J. & Guard, D.R. 1995. "Algorithms for Assessing the Subjectivity of Perceptually Weighted Audible Errors", *J. Audio Eng. Soc.,* vol 43, pp 1041-1045.

Howard, D.M. & Angus, J. 2001, *Acoustics and Psycoacoustics (2 ed)*, Oxford: Focal Press, page 230.

Hueber, G.A.A.A., Nijholt, B. & Tendeloo, H. 1976. "Twin-Tone Tape Testing", *J. Audio Eng. Soc.,* vol 24, pp 542 - 553.

ITT, 1973, *Reference Data for Radio Engineers (5$^{th}$ Ed.)*, New York, Howards Sams & Co, page 17-12.

Jang , Han-Kee & Kim, Kwang-Joon. 1994. " Identification of Loudspeaker Nonlinearities Using the NARMAX Modeling Technique", *J. Audio Eng. Soc.,* vol. 42, No. 12, pp 50-59.

Jeruchim, M.C., Balaban, P. and Shanmugan, K.S. 1992. *Simulation of Communication Systems,* New York: Plenum Press.

Jin, C and Kubichek, R. 1996: "Vector Quantization Techniques for Output Based Objective Speech Quality", Proc. ICASSP-96, Vol.1, pp 491-494, Atlanta, May 1996. (Cited in: Picovici, D. and Mahdi, A.E. (undated), "Towards Non-intrusive Speech Quality Assessment for Modern telecommunications".
Available: http://telecomms.eeng.dcu.ie/symposium/papers/F2.pdf)

Johns, P.B. and Rowbotham, T.R.1972. *Communication Systems Analysis*, London: Butterworths.

Katagiri, S. (ed). 2000. *Handbook of Neural Networks for Speech Processing*, Boston: Artech House.

Kates, J.M. 1992. "On using coherence to measure distortion in hearing aids", *J.Acoust.Soc. Am.*, vol 91(4), Pt.1, April 1992, pp2236-2244.

Katz,B. "Compression in mastering". Available: http://www.digido,cm/

Katz, B. 2003. *Mastering Audio: The Art and the Science,* Focal Press

King, G.J. 1979. *Audio Equipment Tests,* London: Newness-Butterworth.

Koya, D, “Aural Phase Distortion Detection”, MSc (Music Engineering Technology) dissertation, Dept.of Music, University of Miami, Miami.
Available: http://www.music.miami.edu/programs/Mue/mue2003/research/dkoya/

Kuni, C.1997. “Measure Wow and Flutter”, *Electronics World,* February 1997, pp 163 - 166.

Leinonen, E., Otala, M. and Curl, J. 1977. “A Method for Measuring Transient Intermodulation Distortion”, *J. Audio Eng. Soc.,* vol 25, pp 170 - 177.

Leman, M. 2000. “Visualization and Calculation of the Roughness of Acoustical Musical Signals Using the Synchronization Index Model.” Proc. COST G-6 Conf. on Digital Audio Effects (DAFX-00), Verona, Italy, Dec. 2000.

Lipshitz, S.P. & Vanderkooy, J. 1981. “The Great Debate: Subjective Evaluation”, *J. Audio Eng. Soc.*, vol 29, No 7/8, pp 482 – 491.

Maré, S. 1985. “Non-intrusive testing using programme statistics”, *IEEE Trans. Broadcasting*, Vol BC-31, no 4.

Maré, S. 1986. “Using digital signal analysis to monitor audio programme quality”, *IEEE Conference on Digital Signal Processing*, Rhodes University, Grahamstown, 1986.

Maré, S. 2002. “Detection of Non-Linear Distortion in Audio Signals”, *IEEE Trans. Broadcasting,* vol. BC-48, no. 2, June 2002

Maré, S. 2004. “Detection of Distortion Due to Mechanical Imperfections in Recording and Playback Apparatus”, IPET 2004 Conference, Tswane University of Technology, Pretoria, September 2004.

McKnight, J.G. and Weiss, M.R. 1976. “Flutter Analysis for Identifying Tape Recorders”, *J. Audio Eng. Soc*., vol 24, pp 728 - 734.

Meddis, R. and Hewitt, M.J. 1991. “Virtual pitch and phase sensitivity of a computer model of the auditory reriphery. I: Pitch identification.” J. Acoust. Soc. Am., 89, pp2866 – 2882. Cited in: Cartwright, H.E., González, D.L. and Piro, O. 1999. “A New Nonlinear Model fro Pitch Perception.” In Reguera, D. Rubi, R. and Vilar, J. (eds) *Statistical Mechanics of Biocomplexity,* Lecture Notes in Physics, vol. 527, pp 205 – 216, Springer, 1999.

Mock, T. 2004. “Music Everywhere”, *IEEE Spectrum,* September 2004

Octiv, Inc. 1999. “Applications of Consumer-Side Audio Dynamic Range Processing”, White paper, OCTiMax. Available: http://www.octiv.com/

Ohm, G.S. 1843. “Über die Definition des Tones, nebst dara geknüfter Theorie der Sirene und änlicher tonbildender Vorichtung. Ann. Phys. Chem. 59, 513 – 565. Cited in: Cartwright, H.E., González, D.L. and Piro, O. 1999. “A New Nonlinear Model fro Pitch Perception.” In Reguera, D. Rubi, R. and Vilar, J. (eds) *Statistical Mechanics of Biocomplexity,* Lecture Notes in Physics, vol. 527, pp 205 – 216, Springer, 1999.

Paillard, B., Mabilleau, P., Morisette, S. & Soumagne, J. 1992. “PERCEVAL: Perceptual Evaluation of the Quality of Audio Signals”, *J. Audio Eng. Soc.,* vol 40, pp 21-31.

Petri-Larmi, M., Otala, M. & Lammasniemi, J.1980. “Psychoacoustic Detection Threshold of Transient Intermodulation Distortion”, *J. Audio Eng. Soc.,*Vol. 28, No. 3, pp98 – 104.

Peebles, P.Z. Jr. 1980. *Probability, Random Variables and Random Signal Principles*, New York: McGraw-Hill.

Picovici, D. and Mahdi, A.E. 2003, “Non-intrusive Assessment of Perceptual Speech Quality Using a Self-Organising Map”, Eurospeech-2003, pp2077-2080.

Picovici, D. and Mahdi, A.E. (undated), “Towards Non-intrusive Speech Quality Assessment for Modern telecommunications”.
Available: http://telecomms.eeng.dcu.ie/symposium/papers/F2.pdf

Pierce, J.R. 1983. *The Science of Musical Sound*, New York: Scientific American Books.

Plomp, R. 1966. *Experiments on Tone Perception*. Van Gorcum, Assen. Cited in: Leman, M. 2000. “Visualization and Calculation of the Roughness of Acoustical Musical Signals Using the Synchronization Index Model.” Proc. COST G-6 Conf. on Digital Audio Effects (DAFX-00), Verona, Italy, Dec. 2000.

Preis, D.1976. “Impulse Testing and Peak Clipping”, *J. Audio Eng. Soc.,* vol 25, pp 2 - 14.

Preis, D. 1976. “Linear Distortion”, J. Audio Eng. Soc., vol 24, pp 346 - 367.

Preis, D. 1982. "Phase Distortion and Phase Equalization in Audio Signal Processing — A Tutorial Review," *J. Audio Eng. Soc.,* vol. 30, pp. 774-794

Rabiner, L.R. & Schafer, R.W. 1978. *Digital Processing of Speech Signals*, New Jersey: Prentice Hall.

Robinson, D.J.M. and Hawksford, M.O.J. 1999. "Time-domain auditory model for the assessment of high-quality coded audio", 107 th AES Convention, New York, September 1999, preprint 5017 (G-2)

Robinson, D.J.M and Hawksford, M.O.J. 2000." Psychoacoustic Models and Non-Linear Human Hearing", 109th AES Convention, Los Angeles, September 2000.

Rossing, T.D., Moore, F.R. and Wheeler, P.A. 2002. *The Science of Sound (3rd ed.)*, San Fransisco, Addison Wesley.

Scheirer, E.D. 1998, *Music Perception Systems*, PhD proposal, MIT Media Laboratory. Available: http://web.media.mit.edu/~eds/papers/phdprop

Scheirer,E.D. 1999. "Towards music understanding without separation: Segmenting music with correlogram comodulation." MIT Media Laboratory Perceptual Computing Section Technical Report #492, April 1999.
Available: http://vismod.media.mit/tech-reports/TR_492.pdf

Scheirer, E.D. 2000. *Music Listening Systems,* PhD dissertation, MIT Media Laboratory.
Available:
http://www.idiap.ch/~paiement/references/to_read/music/feature_extraction/eds-dis-full/eds-dis-full.pdf

Schouten, J.F., Ritsma, R.J. and Cardozo, B.L. 1962. “Pitch of the residue”, J. Acoust. Soc. Am., 34, pp 1418 – 1424. Cited in: Cartwright, H.E., González, D.L. and Piro, O. 1999. “A New Nonlinear Model fro Pitch Perception.” In Reguera, D. Rubi, R. and Vilar, J. (eds) *Statistical Mechanics of Biocomplexity,* Lecture Notes in Physics, vol. 527, pp 205 – 216, Springer, 1999.

Schroeder, M.R.1970. “Parameter Estimation in Speech: A Lesson in Unorthodoxy*”, Proc. IEEE*, vol 58, pp 707-712.

Schwartz, D.A., Howe, C.Q. and Purves, D. 2003. ”The Statistical Structure of Human Speech Sounds Predicts Musical Universals.” The Journal of Neuroscience, August 2003, Vol. 23, No. 18, pp7160 – 7168.

Schwartz, M. & Shaw, L. 1975. *Signal Processing: Discrete Spectral Analysis, Detection and Estimation*, Kogakusha: McGraw-Hill.

Seebeck, A. 1843. “Über die Definition des Tones”, Ann. Phys. Chem., 60 , 449 – 481. Cited in: Cartwright, H.E., González, D.L. and Piro, O. 1999. “A New Nonlinear Model fro Pitch Perception.” In Reguera, D. Rubi, R. and Vilar, J. (eds) *Statistical Mechanics of Biocomplexity,* Lecture Notes in Physics, vol. 527, pp 205 – 216, Springer, 1999.

Shlien, S. and Soulodre, G. 1996. “Measuring the Characteristics of “Expert” Listeners”, AES 101st Convention, 8-11 November 1996, Los Angeles.

Slaney, M and Lyon, R.F. 1990. “A Perceptual Pitch Detector.” Proc. International Conf. on Acoustics, Speech and Signal Processing, Vol. 1, pp357 – 360.

Temme, Steve, Undated “Audio Distortion Measurements”, Brüel & Kjær Application Note.

Terhardt, , E. 1974. “Pitch, consonance and harmony.” J. Acoust. Soc. Am., 55, pp1061-1069. Cited in: Scheirer, E.D. 2000. *Music Listening Systems,* PhD dissertation, MIT Media Laboratory.
Available:
http://www.idiap.ch/~paiement/references/to_read/music/feature_extraction/eds-dis-full/eds-dis-full.pdf

Terhardt, E. 2000. “Topics of Research in Retrospect: Spectral Pitch; Pitch Perception; Sensory Consonance”
Available: http://www.mmk.e-technik.tu-muenchen.de/persons/ter/

Totzek, U. and Preiss, D. 1987. “How to Measure and Interpret Coherence Loss in Magnetic Recording”, J. Audio Eng. Soc., vol 35, 11, pp869-887

Tsai, Chen-Gia, 2003. “Perception of subharmonics: Roughness”, Chapter 9 of : The Chinese Membrane Flute (*dizi*): Physics and Perception of its Tones, Ph.D. dissertation under the supervision of Prof. Dr. Wolfgang Auhagen and Dr. René Caussé . Date of defense: 28.11.2003 (Musikwissenschaftliches Seminar, Humboldt-University Berlin, Germany). Available: http://www.yogimont.net/jia/dizi/diss/

Tsai, Chen-Gia, 2004. “Auditory Grouping in the Perception of Roughness Induced by Subharmonics: Empirical Findings and a Qualitative Model”, Proc. International Symposium on Musical Acoustics, ISMA2004, March-April 2004, Nara, Japan

Verhey, J.L. 1999. Psychoacoustics of spectro-temporal effects in masking and loudness perception. PhD dissertation, University of Oldenburg
Available: http://www.bis.uni-oldenburg.de/bisverlag/verspy99/verspy99.html

Von Békésy, G. *Experiments in Hearing*, McGraw-Hill. Cited in: Cartwright, H.E., González, D.L. and Piro, O. 1999. “A New Nonlinear Model fro Pitch Perception.” In Reguera, D. Rubi, R. and Vilar, J. (eds) *Statistical Mechanics of Biocomplexity,* Lecture Notes in Physics, vol. 527, pp 205 – 216, Springer, 1999.

Von Helmholtz, H.L.F. 1863. *Die Lehre von dem Tonempfindungen als physiologische Grundlage für die theorie der Musik*. Braunschweig. Cited in: Cartwright, H.E., González, D.L. and Piro, O. 1999. “A New Nonlinear Model fro Pitch Perception.” In

Reguera, D. Rubi, R. and Vilar, J. (eds) *Statistical Mechanics of Biocomplexity,* Lecture Notes in Physics, vol. 527, pp 205 – 216, Springer, 1999.

Weaver, L.E.1971. *Television Measurement Techniques*, London: Peter Peregrinus.

Wonho Yang, Myron Dixon and Robert Yantoro 1997, “A Modified Bark Spectral Distortion Measure which uses Noise Masking Threshold”, IEEE Speech Coding Workshop, Pocono Manor, pp 55-56.

Wonho Yang, Majid Benbouchta and Robert Yantorno 1998, “Performance of the Modified Bark Spectral Distortion as an Objective Speech Quality Measure”, Proc. ICASSP, pp 541-544.

Zölzer, U. (ed) 2002. *DAFX Digital Audio Effects*. Chichester, John Wiley & Sons.

Zwicker, E. and Fastl, H. 1990. *Psychoacoustics Facts and Models,* Berlin, Springer-Verlag.[Cited in: Shlien, S. and Soulodre, G. 1996. “Measuring the Characteristics of “Expert” Listeners”, AES 101st Convention, 8-11 November 1996, Los Angeles.]

Zwicker, E. & Zwicker, U.T. 1991. “Audio engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory Systems”, *J. Audio Eng. Soc*., vol 39, pp 115-126.

# ANNEXURE 1:

## Appendix to Chapter 5

### A1.1 Comparison of distorted and undistorted signals

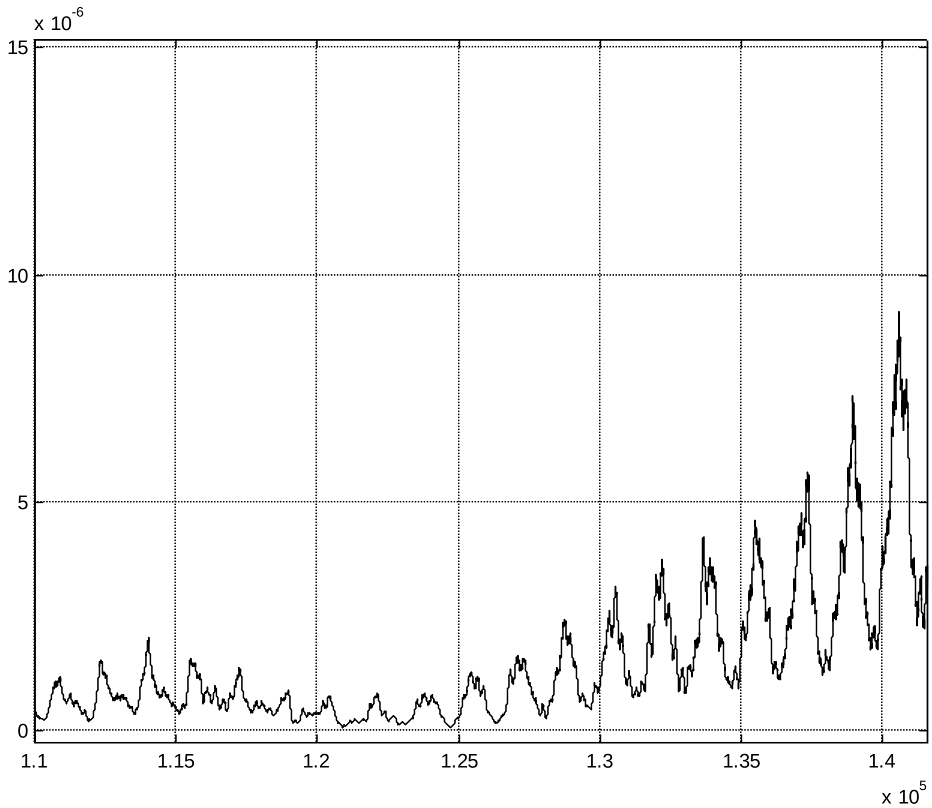


**Figure A1.1:** Envelope of autocorrelation of audio signal with flutter distortion.

The flutter frequency in Figure A1.1 can be calculated to be approximately 26.5 Hz which corresponds to the flutter frequency observed in the audio segment shown in Figure 5.14.

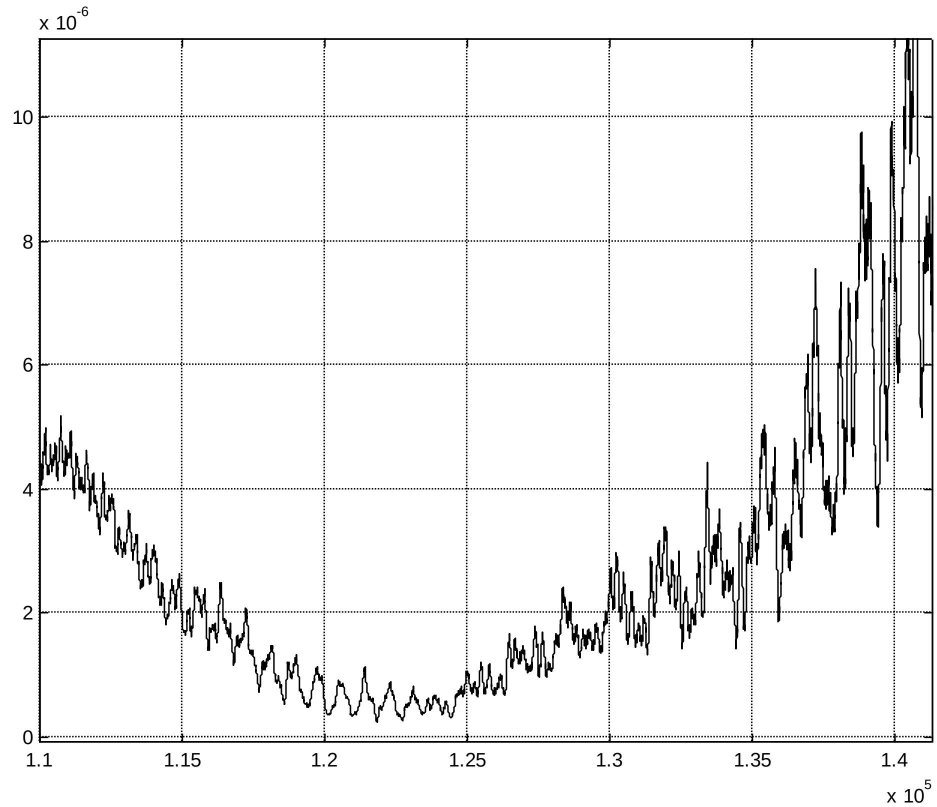


**Figure A1.2:** Audio segment without flutter distortion

The audio segment in Figures A1.1 and A1.2 are the same except for the introduction of flutter distortion in Figure A1.1.

## A1.2 Calibration of the measurement of Wow and Flutter using the time domain autocorrelation technique.

The starting point for the calibration is the three figures below, Figure A1.3, A1.4 and A1.5. These figures show the results of the time domain autocorrelation technique for the specified values of wow & flutter distortion.

The results shown are the output of a simulation of the proposed technique. It is assumed that a single tone has been recorded. In these figures the frequency of the distortion is set to 13.5 Hz, the frequency of the recorded tone is set to 3150 Hz.

The results have been enhanced by squaring the autocorrelation. On examinantion of these figures no correlation between any feature of the squared autocorrelation and the results that would be obtained from a measurement of wow and flutter distortion can be found.

It must therefore be concluded that the proposed technique is able to identify the presence of wow and flutter distortion but that it cannot be used to quantify the magnitude of the distortion.

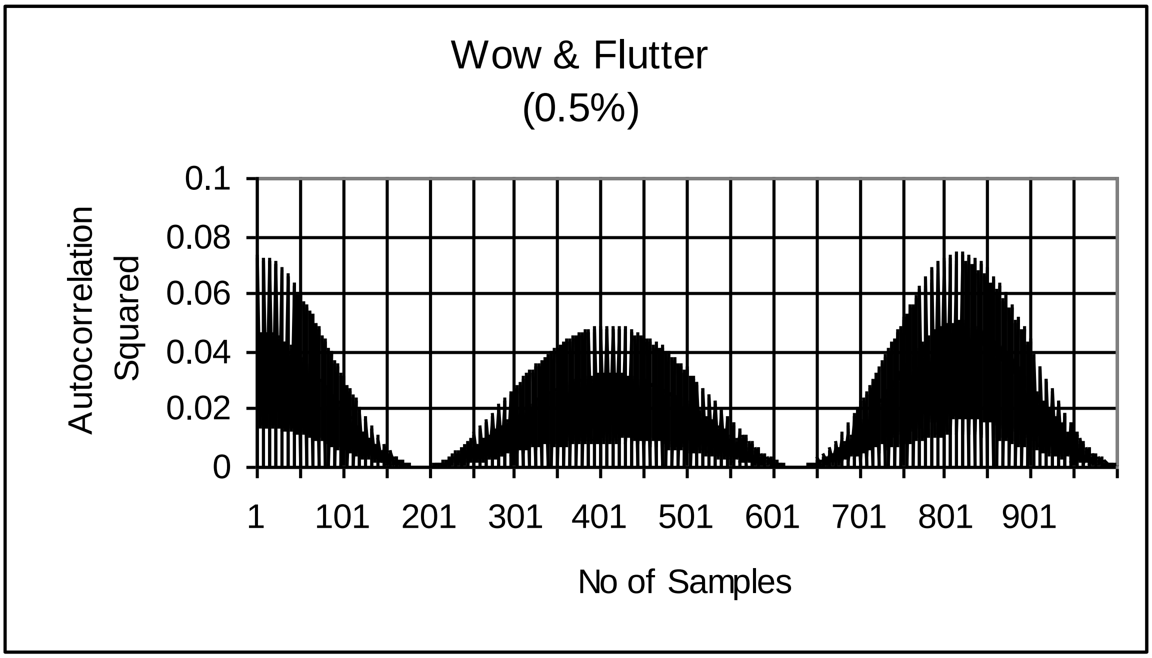


**Figure A1.3:** Output with 0.5% Wow and Flutter

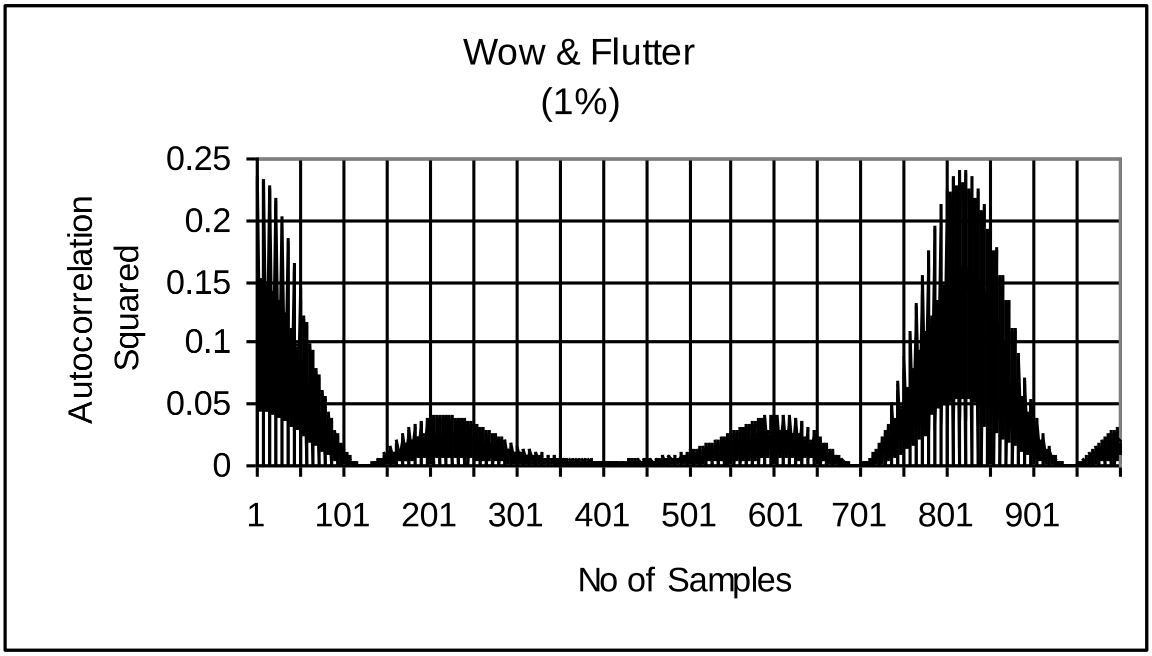


**Figure A1.4:** Output with 1% Wow and Flutter

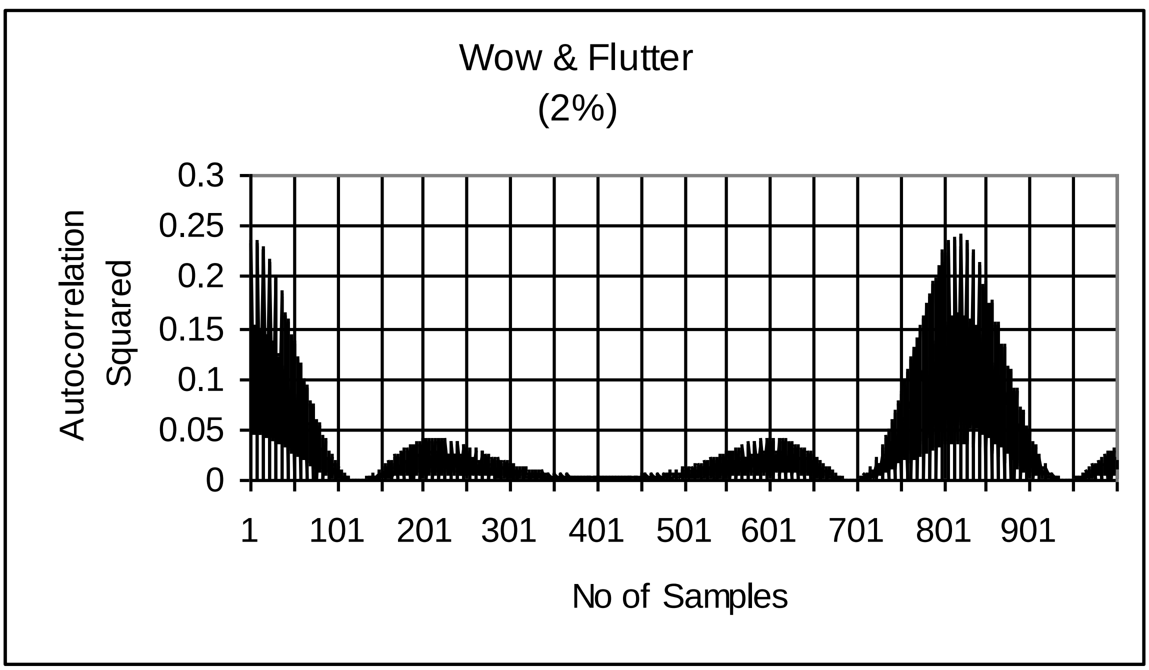


**Figure A1.5:** Output with 2% Wow and Flutter

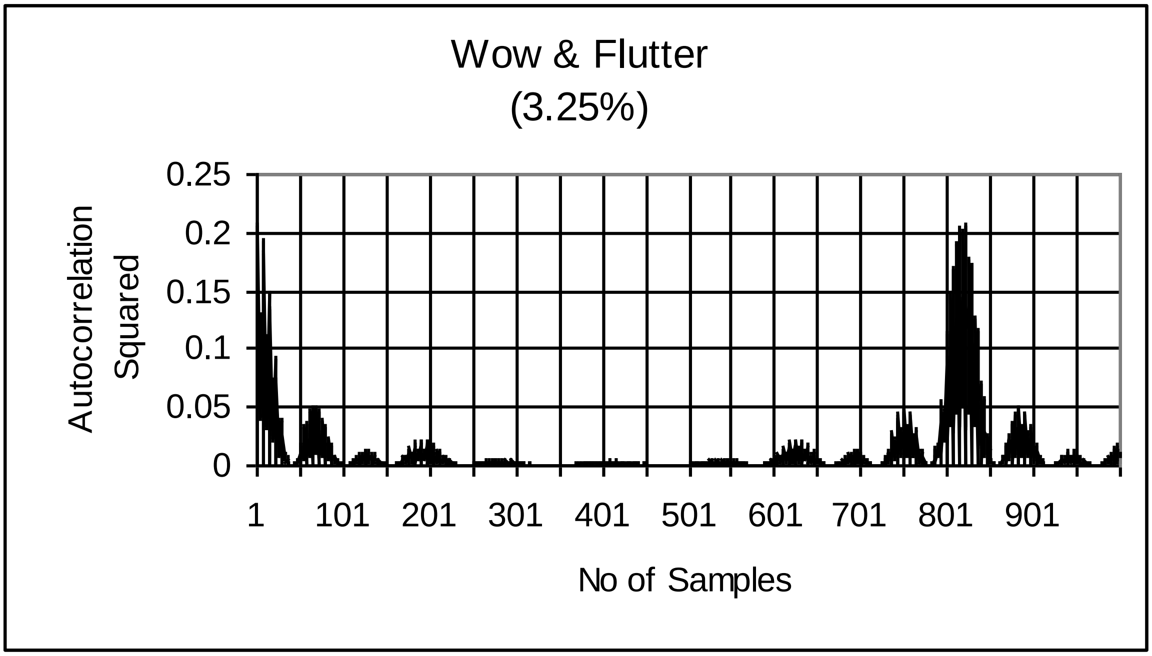


**Figure A1.6:** Output with 3.25% Wow and Flutter

# ANNEXURE 2

## Appendix to Chapter 6

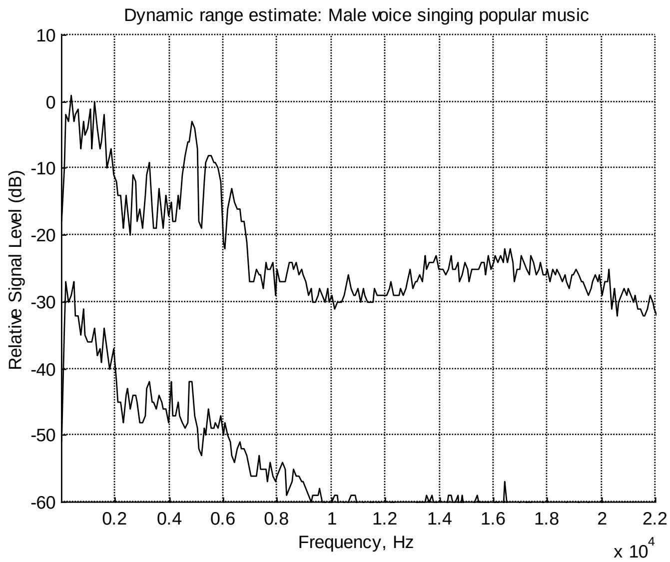


**Figure A2.1**: Data file dat01

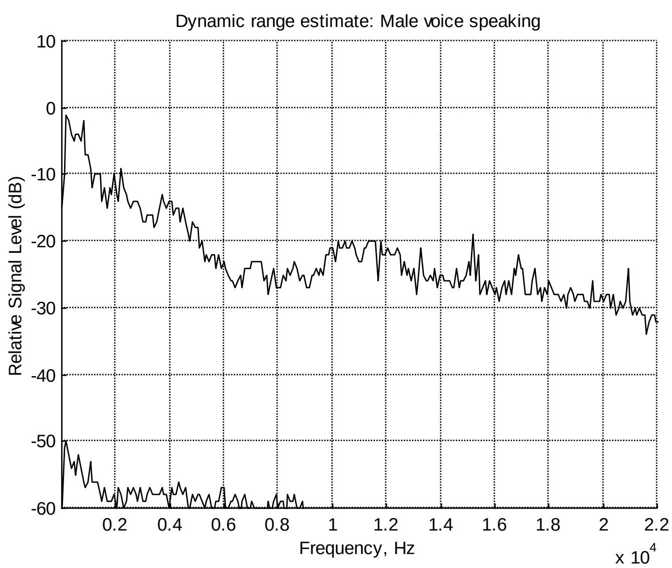


**Figure A2.2:** data file dat03

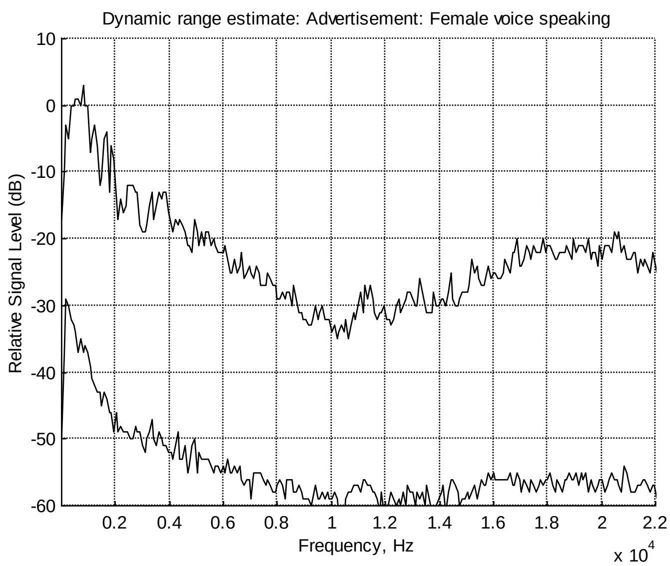


**Figure A2.3:** Data file dat04

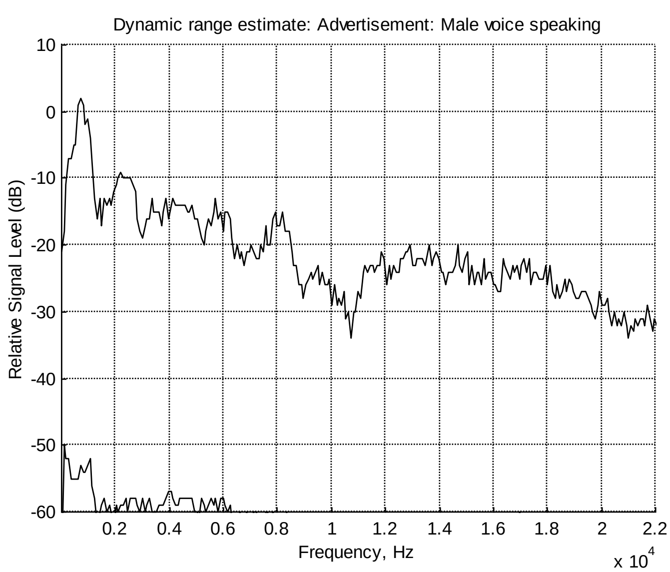


**Figure A2.4:** data file dat06

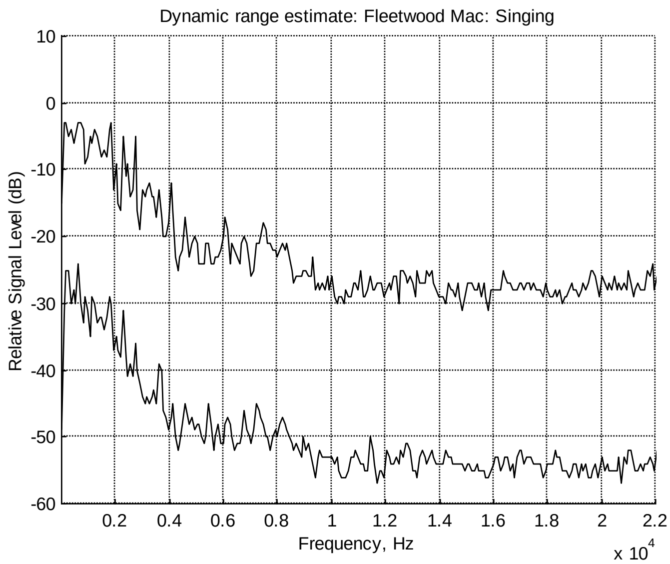


**Figure A2.5:** Data file dat09

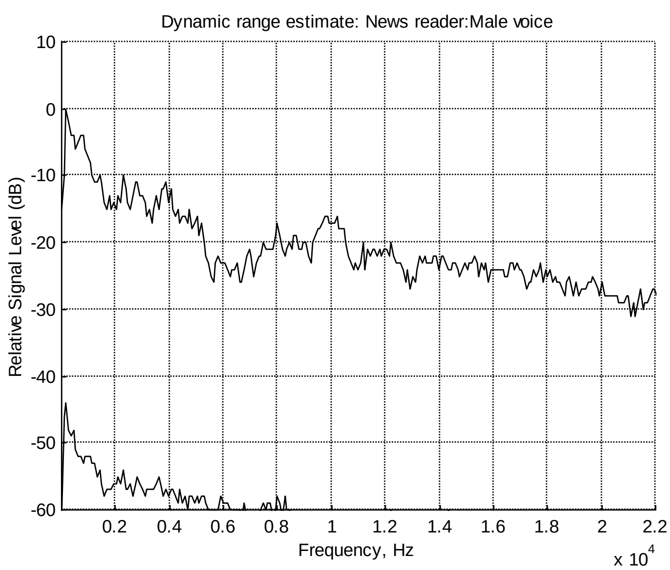


**Figure A2.6:** Data file dat18

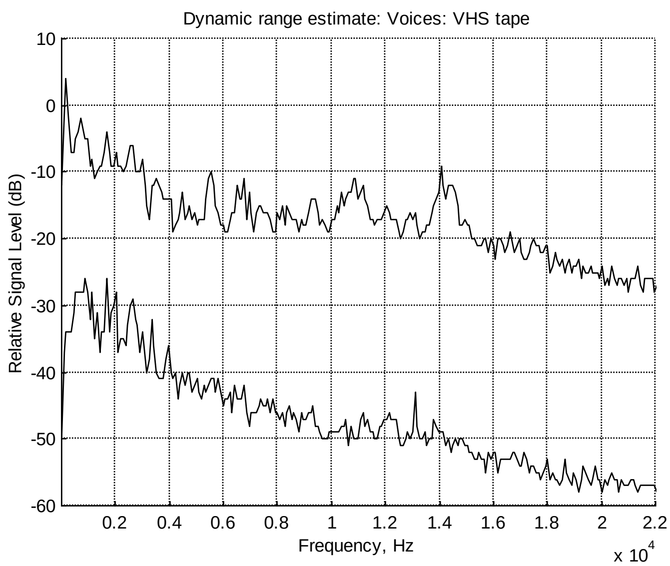


**Figure A2.7:** Data file data100

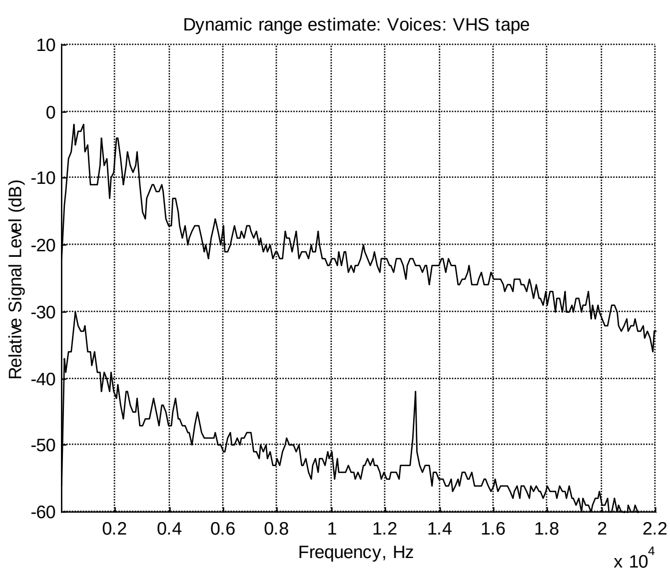


**Figure A2.8:** Data file data106

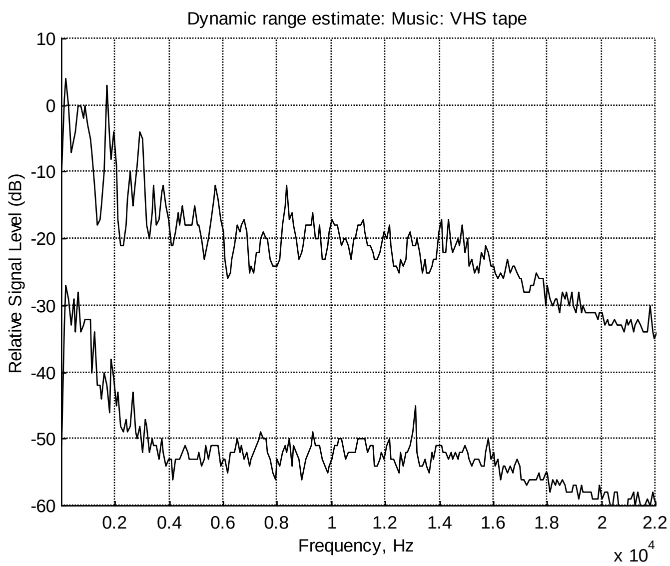


**Figure A2.9:** Data file Data109

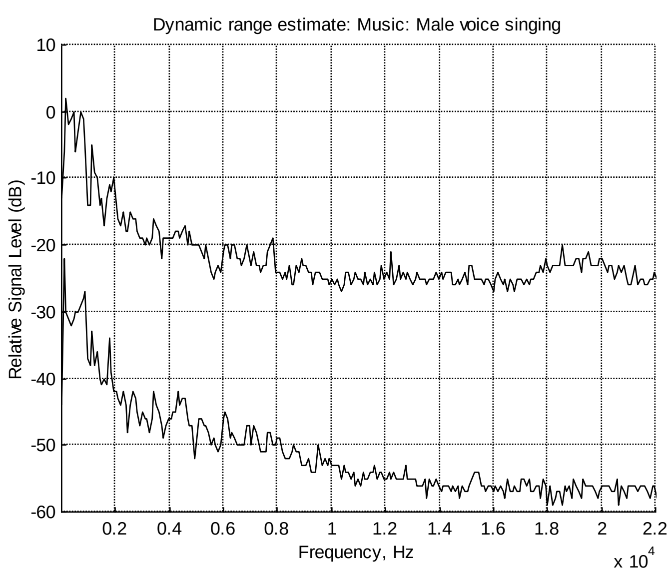


**Figure A2.10:** Data file data117

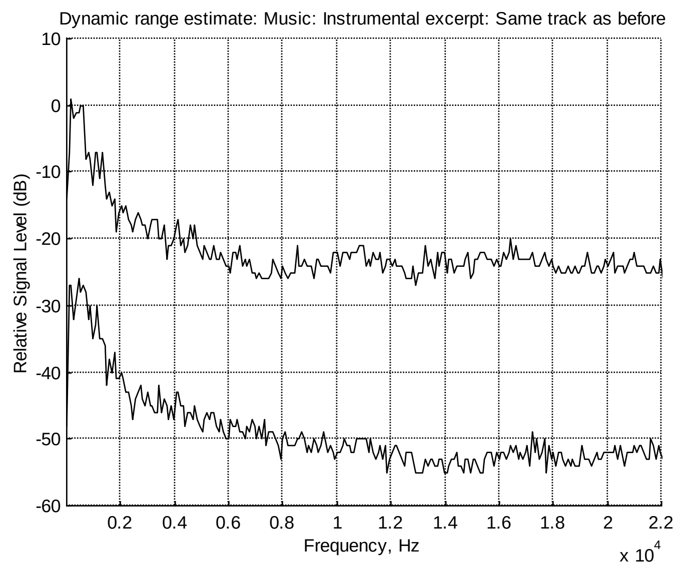


**Figure A2.11:** Data file data118

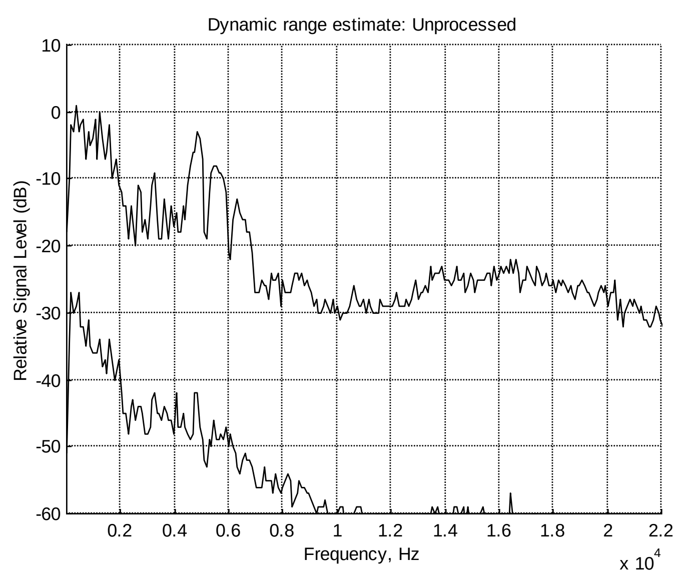


**Figure A2.12:** Data file dat01, unprocessed

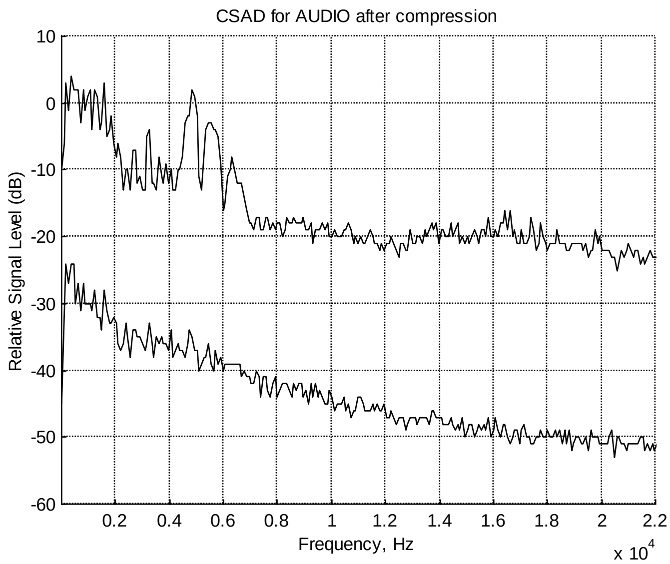


**Figure A2.13:** Data file dat01, after compression

# ANNEXURE 3

## Appendix to Chapter 7

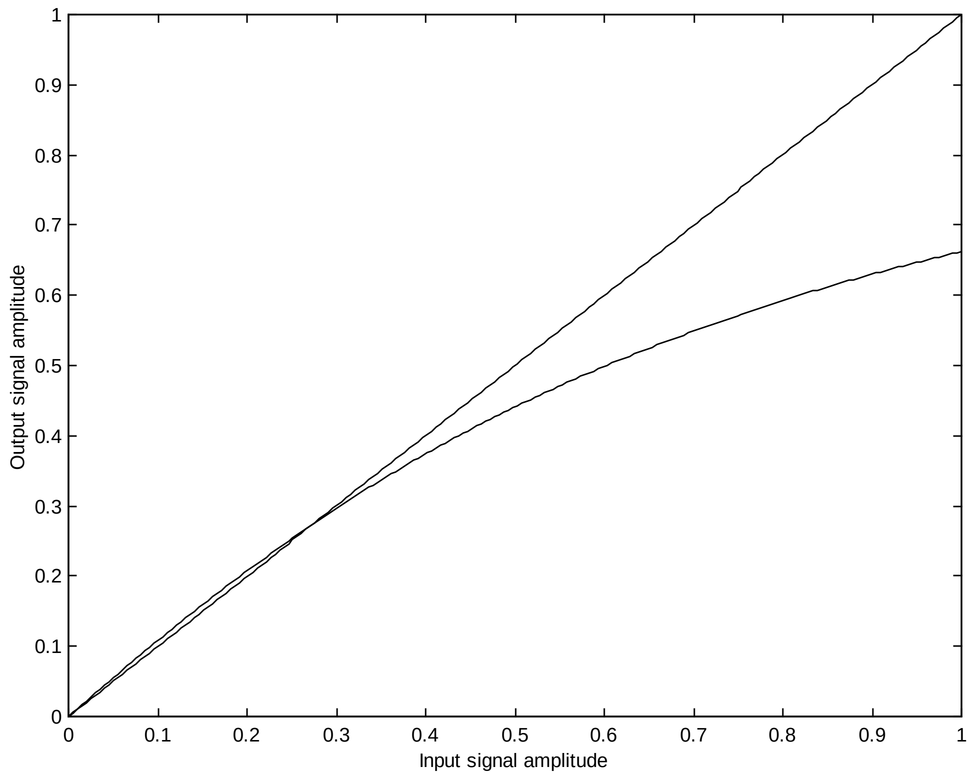


**Figure A3.1:** Distortion curve: M=1.0, u=1.5, k=0.9

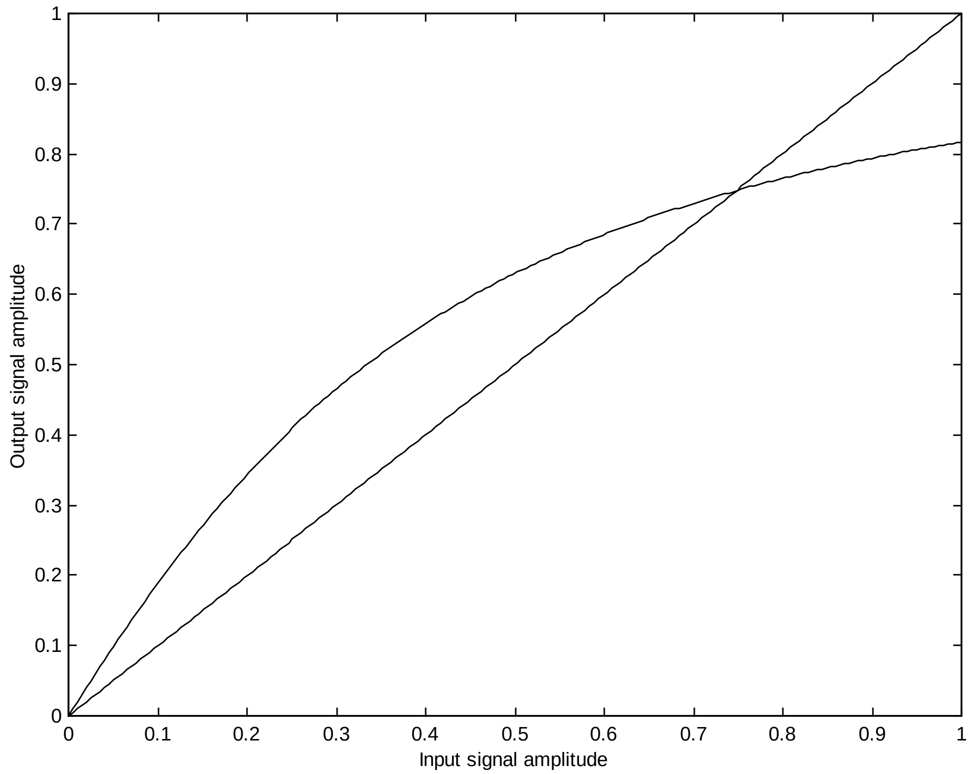


**Figure A3.2:** Distortion curve: M=1.0, u =1.5, k=0.5

## ANNEXURE 4:

**Description of Data files used in Chapters 5, 6 and 7**

The data files are identified by a alpha numeric code followed by a bracketed expression indicating the portion of the (250 kilo sample to 500 kilo sample) original file that was used e.g. (>65k) signifies that the first 65 kilo samples were ignored and the next 65 kilo samples were used.

**TABLE A4.1: Initial, unclassified data.**

| Data File | $5^{th}$ Order | $4^{th}$ Order | $3^{rd}$ Order | $2^{nd}$ Order | $1^{st}$ Order | Constant |
|---|---|---|---|---|---|---|
| 21u (>65k) | 0.2422 | -0.7774 | 0.8989 | -0.4003 | 0.0181 | 0.0245 |
| Distorted | -0.0625 | 0.1530 | -0.0831 | -0.0257 | 0.0029 | 0.0155 |
| 22u | 0.3070 | -0.9351 | 1.0103 | -0.4079 | -0.0023 | 0.0281 |
| Distorted | -0/0735 | 0.1544 | -0.0573 | -0.0374 | -0.0043 | 0.018 |
| 23u | 0.2077 | -0.6504 | 0.7225 | -0.2939 | -0.0129 | 0.0270 |
| Distorted | -0.0538 | 0.1349 | -0.0810 | -0.0100 | -0.0068 | 0.0168 |
| 24u | 0.2656 | -0.8363 | 0.9465 | -0.4167 | 0.0159 | 0.0251 |
| Distorted | -0.1107 | 0.2696 | -0.1822 | 0.0102 | -0.0028 | 0.0159 |
| 41u | 0.2226 | -0.6803 | 0.7368 | -0.2907 | -0.0153 | 0.0271 |
| Distorted | -0.0368 | 0.0859 | -0.0330 | -0.0285 | -0.0041 | 0.0166 |
| 40u (>165k) | 0.0597 | -0.2546 | 0.3751 | -0.2045 | 0.0021 | 0.0221 |
| Distorted | 0.0066 | 0.0361 | -0.0596 | 0.0049 | -0.0012 | 0.0134 |
| 47u (>165k) | 0.3591 | -1.0443 | 1.0765 | -0.4132 | -0.0062 | 0.0285 |
| Distorted | -0.0112 | -0.0119 | 0.0967 | -0.0945 | 0.0032 | 0.0178 |
| 1u (>100k) | 0.1661 | -0.4937 | 0.5422 | -0.2344 | -0.0027 | 0.0228 |
| Distorted | 0.1448 | -0.3775 | 0.3680 | -0.1621 | 0.0113 | 0.0156 |
| 2u | 0.0667 | -0.2813 | 0.3975 | -0.1959 | -0.0119 | 0.0248 |
| Distorted | 0.0090 | 0.0038 | 0.0004 | -0.0218 | -0.0078 | 0.0166 |
| 5u | 0.2228 | -0.7360 | 0.8610 | -0.3789 | 0.0044 | 0.0267 |
| Distorted | -0.1635 | 0.4220 | -0.4442 | 0.0698 | -0.0117 | 0.0165 |
| 25u | 0.3197 | -0.9959 | 1.1145 | -0.4882 | 0.0244 | 0.0256 |
| Distorted | -0.1120 | 0.2579 | -0.1485 | -0.0154 | 0.0016 | 0.0163 |
| 5u (>65k) | 0.4075 | -1.2010 | 1.2472 | -0.4789 | -0.0047 | 0.0302 |
| Distorted | -0.1023 | 0.2051 | -0.0715 | -0.0476 | -0.0025 | 0.0188 |
| 5u (>150k) | 0.1766 | -0.6229 | 0.7679 | -0.3528 | 0.0050 | 0.0261 |
| Distorted | -0.2248 | 0.5941 | -0.5048 | 0.1380 | -0.0184 | 0.0159 |
| 25u (>65k) | 0.1314 | -0.4808 | 0.6188 | -0.2985 | 0.0044 | 0.0246 |
| Distorted | -0.0404 | 0.1322 | -0.1047 | -0.002 | -0.0019 | 0.0153 |
| 25u | 0.0859 | -0.3440 | 0.4776 | -0.2465 | 0.0038 | 0.0231 |

| | | | | | | |
|---|---|---|---|---|---|---|
| (>150k) | | | | | | |
| Distorted | -0.0526 | 0.1781 | -0.1705 | 0.0367 | -0.0057 | 0.0142 |
| 33u (>150k) | 0.1390 | -0.4959 | 0.6327 | -0.3111 | 0.0120 | 0.0233 |
| Distorted | -0.0871 | 0.2913 | -0.2867 | 0.0688 | 0.0023 | 0.0117 |

**Distortion parameters:** M=1.0; u=1.5; k=0.9 (See figure 4. For plot)

Table A4.2: **Data classified as per Table 4.4**

| Data File | $5^{th}$ order | $4^{th}$ order | $3^{rd}$ order | $2^{nd}$ order | $1^{st}$ order | Constant |
|---|---|---|---|---|---|---|
| **Dat01** | **0.3392** | **-0.976** | **0.9947** | **-0.3743** | **-0.0115** | **0.0282** |
| **Distorted** | **0.0401** | **-0.1328** | **0.1939** | **-0.1281** | **0.0104** | **0.0167** |
| **Dat01 (>65)** | **0.1683** | **-0.5123** | **0.5465** | **-0.1946** | **-0.0363** | **0.0286** |
| **Distorted** | **0.0356** | **-0.0772** | **0.0878** | **-0.0586** | **-0.0.044** | **0.0171** |
| **Dat02** | **0.2906** | **-0.9021** | **0.9972** | **-0.4179** | **0.0052** | **0.0272** |
| **Distorted** | **-0.1199** | **0.2821** | **-0.1802** | **0.0075** | **-0.007** | **0.0174** |
| **Dat02 (>65)** | **0.3762** | **-1.0421** | **1.0161** | **-0.3606** | **-0.0172** | **0.0283** |
| **Distorted** | **0.0773** | **-0.2375** | **0.2949** | **-0.1695** | **0.0193** | **0.0155** |
| **Dat03** | **0.0188** | **-0.0752** | **0.0441** | **0.0912** | **-0.1146** | **0.0355** |
| **Distorted** | **-0.1756** | **0.4671** | **-0.4252** | **0.1673** | **-0.0555** | **0.0219** |
| **Dat03 (>65)** | **-1.7920** | **4.9328** | **-5.0256** | **2.3522** | **-0.5258** | **0.055** |
| **Distorted** | **-1.1840** | **3.3789** | **-3.5551** | **1.6861** | **-0.367** | **0.0395** |
| **Dat04** | **0.3269** | **-0.818** | **0.6823** | **-0.1602** | **-0.06** | **0.0297** |
| **Distorted** | **0.2091** | **-0.5806** | **0.5832** | **-0.2468** | **0.019** | **0.0161** |
| **Dat04 (>65)** | **0.0299** | **0.0676** | **-0.3403** | **0.4108** | **-0.2102** | **0.0424** |
| **Distorted** | **0.1293** | **-0.3538** | **0.3244** | **-0.0826** | **-0.0415** | **0.0241** |
| **Dat05** | **-2.3545** | **6.6708** | **-7.0538** | **3.444** | **-0.783** | **0.0728** |
| **Distorted** | **-1.4195** | **4.0756** | **-4.3805** | **2.181** | **-0.5123** | **0.0534** |
| **Dat05 (>65)** | **-4.2838** | **11.9318** | **-12.2933** | **5.7349** | **-1.1881** | **0.0909** |
| **Distorted** | **-3.0158** | **8.4416** | **-8.7626** | **4.1295** | **-0.8682** | **0.0702** |
| **Dat06** | **-3.4651** | **9.6518** | **-9.9528** | **4.665** | **-0.9856** | **0.0808** |
| **Distorted** | **-2.4225** | **6.8154** | **-7.1087** | **3.3676** | **-0.7178** | **0.0619** |
| **Dat06 (>65)** | **-3.4755** | **9.755** | **-10.1739** | **4.8494** | **-1.0474** | **0.0865** |
| **Distorted** | **-2.0679** | **5.8763** | **-6.228** | **3.0329** | **-0.6793** | **0.0627** |
| **Dat07** | **-1.9590** | **5.4628** | **-5.6643** | **2.708** | **-0.6118** | **0.0609** |
| **Distorted** | **-1.3462** | **3.7985** | **-3.9683** | **1.8849** | **-0.4144** | **0.0434** |
| **Dat07 (>65)** | **-5.7642** | **16.0276** | **-16.4774** | **7.6620** | **-1.5725** | **0.1142** |
| **Distorted** | **-4.4812** | **12.4953** | **-12.8919** | **6.0228** | **-1.2471** | **0.0943** |
| **Dat08** | **0.2121** | **-0.6582** | **0.7271** | **-0.2975** | **-0.0097** | **0.0263** |
| **Distorted** | **-0.0771** | **0.1926** | **-0.1332** | **0.0092** | **-0.0077** | **0.0162** |
| **Dat08 (>65)** | **0.4242** | **-1.2225** | **1.2436** | **-0.4696** | **-0.005** | **0.0297** |
| **Distorted** | **0.0025** | **-0.0663** | **0.1709** | **-0.133** | **0.0075** | **0.0185** |
| **Dat09** | **0.2835** | **-0.8562** | **0.9256** | **-0.3808** | **0.0017** | **0.0265** |
| **Distorted** | **-0.0348** | **0.0781** | **-0.014** | **-0.0493** | **0.0043** | **0.0158** |
| **Dat09 (>65)** | **0.2205** | **-0.7325** | **0.8719** | **-0.4033** | **0.0188** | **0.0246** |
| **Distorted** | **-0.1143** | **0.2871** | **-0.2045** | **0.0196** | **-0.0038** | **0.0159** |
| **Dat10** | **0.1906** | **-0.6097** | **0.696** | **-0.2957** | **-0.0071** | **0.0261** |
| **Distorted** | **-0.0163** | **0.0491** | **-0.0124** | **-0.0357** | **-0.005** | **0.016** |
| **Dat10 (>65)** | **0.1313** | **-0.4649** | **0.5891** | **-0.2867** | **0.0079** | **0.0232** |
| **Distorted** | **-0.0554** | **0.1686** | **-0.1437** | **0.019** | **-0.0026** | **0.0143** |
| **Dat11** | **0.2362** | **-0.7324** | **0.8084** | **-0.3327** | **-0.0062** | **0.0268** |
| **Distorted** | **-0.0702** | **0.1724** | **-0.1067** | **-0.0071** | **-0.0047** | **0.0165** |
| | | | | | | |
| **Data file** | **$5^{th}$ order** | **$4^{th}$ order** | **$3^{rd}$ order** | **$2^{nd}$ order** | **$1^{st}$ order** | **Constant** |
| Dat11 (>65) | 0.1413 | -0.4995 | 0.6209 | -0.2865 | -0.0014 | 0.0251 |
| Distorted | -0.1235 | 0.3357 | -0.2841 | 0.0688 | -0.0126 | 0.0156 |
| Dat12 | 0.5084 | -1.4798 | 1.5268 | -0.6007 | 0.0163 | 0.0294 |
| Distorted | -0.1063 | 0.18 | -0.011 | -0.0843 | 0.002 | 0.0194 |
| Dat12 (>65) | -0.0754 | 0.2024 | -0.2391 | 0.1989 | -0.1206 | 0.0336 |
| Distorted | -0.0231 | 0.1052 | -0.1413 | 0.0846 | -0.046 | 0.0207 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Dat13 | 0.311 | -0.9166 | 0.9632 | -0.3804 | -0.0042 | 0.0274 |
| Distorted | 0.0021 | -0.0277 | 0.0912 | -0.0878 | 0.0056 | 0.0167 |
| Dat13 (>65) | -0.0483 | 0.0199 | 0.0785 | -0.0387 | -0.047 | 0.0272 |
| Distorted | -0.011 | 0.0901 | -0.1193 | 0.0436 | -0.0197 | 0.0168 |
| Dat14 | -0.5318 | 1.5537 | -1.7494 | 0.9775 | -0.2961 | 0.0452 |
| Distorted | -0.4069 | 1.1315 | -1.1709 | 0.5692 | -0.1522 | 0.0284 |
| Dat14 (>65) | -2.2772 | 6.3439 | -6.5631 | 3.1201 | -0.6936 | 0.0659 |
| Distorted | -1.3727 | 3.9424 | -4.2057 | 2.0499 | -0.4638 | 0.048 |
| Dat15 | -0.2064 | 0.7503 | -1.0443 | 0.7124 | -0.2548 | 0.043 |
| Distorted | 0.4439 | -1.0901 | 0.9032 | -0.2548 | -0.0247 | 0.0233 |
| Dat15 (>65) | -0.4639 | 1.3501 | -1.4604 | 0.7395 | -0.1987 | 0.033 |
| Distorted | 0.1505 | -0.2641 | 0.0639 | 0.0916 | -0.0611 | 0.0196 |
| Dat16 | -1.2831 | 3.5708 | -3.6879 | 1.7629 | -0.2782 | 0.0345 |
| Distorted | -0.8052 | 2.3239 | -2.4904 | 1.2122 | -0.2782 | 0.0345 |
| Dat16 (>65) | -0.1192 | 0.4716 | -0.7058 | 0.5178 | -0.2026 | 0.0384 |
| Distorted | 0.0318 | -0.0525 | -0.0216 | 0.0840 | -0.0653 | 0.023 |
| Dat17 | -3.6065 | 10.0841 | -10.4519 | 4.9305 | -1.0474 | 0.085 |
| Distorted | -2.4339 | 6.8717 | -7.2108 | 3.4518 | -0.748 | 0.065 |
| Dat17 (>65) | -3.3569 | 9.493 | -9.9898 | 4.8127 | -1.052 | 0.0876 |
| Distorted | -1.8064 | 5.2328 | -5.6877 | 2.8642 | 0.67 | 0.0643 |
| Dat18 | -2.6947 | 7.4177 | -7.5355 | 3.4765 | -0.7345 | 0.0655 |
| Distorted | -2.0511 | 5.7471 | -5.9323 | 2.7503 | -0.5675 | 0.0502 |
| Dat18 (>65) | -2.0287 | 5.6273 | -5.7908 | 2.7407 | -0.6128 | 0.0609 |
| Distorted | -1.3037 | 3.7227 | -3.9343 | 1.8892 | -0.4199 | 0.0442 |
| Dat19 | -3.251 | 9.0449 | -9.3335 | 4.4005 | -0.9474 | 0.0807 |
| Distorted | -2.3015 | 6.4728 | -6.7433 | 3.1949 | -0.6881 | 0.0616 |
| Dat19 (>65) | -2.7961 | 7.7240 | -7.889 | 3.6697 | -0.783 | 0.0693 |
| Distorted | -2.042 | 5.7437 | -5.9643 | 2.7959 | -0.5896 | 0.0532 |

**Distortion parameters:** M=1.0; u= 1.5; k=0.9. (See Figure A3.1: Annexure 3. For a plot)

**Table A4.3**: Data file as for Table 4.2 but using different distortion parameters

| Data file | $5^{th}$ order | $4^{th}$ order | $3^{rd}$ order | $2^{nd}$ order | $1^{st}$ order | Constant |
|---|---|---|---|---|---|---|
| Dat01 | 0.1099 | -0.2415 | 0.1889 | -0.0817 | 0.0132 | 0.0112 |
| Dat01 (>65) | 0.1716 | -0.3837 | 0.2889 | -0.0991 | 0.0108 | 0.0114 |
| Dat02 | 0.0996 | -0.1684 | 0.0717 | -0.0148 | 0.003 | 0.0117 |
| Dat02 (>65) | 0.0386 | -0.0847 | 0.0746 | -0.0523 | 0.0129 | 0.0107 |
| Dat03 | 0.0460 | -0.0243 | -0.0746 | 0.0631 | -0.0249 | 0.0151 |
| Dat03 (>65) | -0.4299 | 1.3956 | -1.6941 | 0.9288 | -0.2302 | 0.0293 |
| Dat04 | -0.0275 | 0.0017 | 0.0659 | -0.0634 | 0.0113 | 0.011 |
| Dat04 (>65) | 0.0204 | -0.0944 | 0.1287 | -0.062 | -0.0092 | 0.0164 |
| Dat05 | -0.8516 | 2.4612 | -2.6923 | 1.3803 | -0.3394 | 0.0401 |
| Dat15 (>65) | -2.2686 | 6.3565 | -6.6341 | 3.1620 | -0.6773 | 0.0576 |
| Dat06 | -1.6313 | 4.6465 | -4.9456 | 2.4107 | -0.5324 | 0.0493 |
| Dat06 (>65) | -1.2564 | 3.6268 | -3.939 | 1.9869 | -0.4688 | 0.0486 |
| Dat07 | -0.7644 | 2.2436 | -2.468 | 1.2419 | -0.2874 | 0.0332 |
| Dat07 (>65) | -3.466 | 9.7044 | -10.0772 | 4.7523 | -0.9983 | 0.0787 |
| Dat08 | 0.0769 | -0.1385 | 0.0633 | -0.0133 | 0.0006 | 0.011 |
| Dat08 (>65) | -0.0789 | -0.1512 | 0.1046 | -0.0517 | 0.0071 | 0.0125 |
| Dat09 | 0.0914 | -0.1683 | 0.0906 | -0.0308 | 0.0062 | 0.0107 |
| Dat09 (>65) | 0.1377 | -0.2676 | 0.1545 | -0.0401 | 0.005 | 0.0107 |
| Dat10 | 0.1666 | -0.3615 | 0.2590 | -0.0861 | 0.0114 | 0.0106 |
| Dat10 (>65) | 0.1315 | -0.2714 | 0.1666 | -0.043 | 0.0066 | 0.0096 |
| Dat11 | 0.0941 | -0.1723 | 0.0878 | -0.0233 | 0.0025 | 0.0111 |
| Dat11 (>65) | 0.1379 | -0.2628 | 0.1399 | -0.0275 | 0.0022 | 0.0105 |
| Dat12 | -0.0003 | 0.0723 | -0.1149 | 0.0364 | -0.0065 | 0.0133 |
| Dat12 (>65) | 0.1901 | -0.4359 | 0.3277 | -0.0926 | -0.0028 | 0.0137 |
| Dat13 | 0.0802 | -0.1554 | 0.096 | -0.0381 | 0.0059 | 0.0113 |
| Dat13 (>65) | 0.2875 | -0.653 | 0.4916 | -0.151 | 0.0143 | 0.0108 |
| Dat14 | -0.2827 | 0.8024 | -0.8517 | 0.412 | -0.1014 | 0.0207 |
| Dat14 (>65) | -0.5284 | 1.6742 | -2.0037 | 1.107 | -0.2851 | 0.0355 |
| Dat15 | 0.3775 | -1.0344 | 0.991 | -0.3831 | 0.0337 | 0.0148 |
| Dat15 (>65) | 0.2433 | -0.6513 | 0.5611 | -0.166 | -0.0003 | 0.0127 |
| Dat16 | -0.461 | 1.3622 | -1.5305 | 0.7920 | -0.1899 | 0.0257 |
| Dat16 (>65) | -0.1622 | 0.37 | -0.3224 | 0.1403 | -0.0432 | 0.0163 |
| Dat17 | -1.7496 | 4.9437 | -5.2212 | 2.5326 | -0.5606 | 0.052 |
| Dat17 (>65) | -0.9405 | 2.773 | -3.1193 | 1.6617 | -0.4248 | 0.0482 |
| Dat18 | -1.2439 | 3.6168 | -3.9204 | 1.9254 | -0.4197 | 0.0398 |
| Dat18 (>65) | 0.6481 | 1.9674 | -2.2435 | 1.1709 | -0.2811 | 0.0334 |
| Dat19 | -1.3457 | 3.9238 | -4.2751 | 2.1337 | -0.4865 | 0.048 |
| Dat19 (>65) | -1.2116 | 3.5392 | -3.8590 | 1.9151 | -0.4271 | 0.0418 |

**Distortion parameters:** M=1.0; u= 1.5; k=0.5 (See Figure A3.2.Annexure 3 For a plot)

Table A4.4: **Description of data files**

| Data File | Description |
|---|---|
| Dat01 | Male voice singing: ‘The one that you love’ Air Supply |
| Dat02 | Male voice singing: ‘The one that you love’ Air Supply |
| Dat03 | Male voice speaking: Announcer |
| Dat04 | Advert: Female voice: ‘Your family magazine’ |
| Dat05 | Advert: Male voice ‘Dale Carnegie’ |
| Dat06 | Advert: Male voice ‘Dale Carnegie’ |
| Dat07 | Male voice: Announcer |
| Dat08 | Music: Intro ‘I Do’ Fleetwood Mac |
| Dat09 | Music Singing: ‘I Do’ Fleetwood Mac |
| Dat10 | Music Singing: ‘I Do’ Fleetwood Mac |
| Dat11 | Music, instruments only: ‘I Do’ Fleetwood Mac |
| Dat12 | Music, male voice: ‘Satellite’ |
| Dat13 | Music, male voice: ‘Satellite’ |
| Dat14 | Male voice: Announcer |
| Dat15 | Female voice, advert: ‘garden & Home’ |
| Dat16 | Male voice, advert: “M-Net’ |
| Dat17 | Male voice, advert: ‘Goodwill Foundation’ |
| Dat18 | Male voice: news |
| Dat19 | Male voice: news |
| Data100 | Segment form VHS video: Striptease |
| Data101 | Segment form VHS video: Striptease |
| Data102 | Segment form VHS video: Striptease |
| Data103 | Segment form VHS video: Striptease |
| Data104 | Segment form VHS video: Striptease: Pussy cat sequenc |
| Data105 | Segment form VHS video: Striptease: Talking to kids |
| Data106 | Segment form VHS video: Striptease: Starting: “Gotta run…” |
| Data107 | Segment form VHS video: Striptease: Music |
| Data108 | Segment form VHS video: Striptease: Music |
| Data109 | Segment form VHS video: Striptease: Music: “Come to me…” |
| Data110 | Segment form VHS video: Striptease: Music “ Come to me…” |
| Data111 | Segment form VHS video: Striptease: Music: “Come to me…” |
| Data112 | East Coast Radio: Female singing: “The sweetest days” Vanessa Williams |
| Data113 | East Coast Radio: Female singing: “The sweetest days” Vanessa Williams |
| Data114 | East Coast Radio: Female singing: “The sweetest days” Vanessa Williams |
| Data115 | East Coast Radio: Female singing: “The sweetest days” Vanessa Williams |
| Data116 | East Coast Radio: Intro to: “Jesus to a child”, George Michael |
| Data117 | East Coast Radio: Male voice: “Jesus to a child”, George Michael |

| | |
|---|---|
| Data118 | East Coast Radio: Instrumental: “Jesus to a child”, George Michael |
| Data119 | East Coast Radio: Instrumental: “Jesus to a child”, George Michael |
| Data120 | East Coast Radio: Intro: “The sweetest days” With flutter |
| Data121 | East Coast Radio: Intro to “ The sweetest days” No flutter |

**Table A4.5: data files with 131072 samples**

| Data file | 5th order | 4th order | 3rd order | 2nd order | 1st order | Constant |
|---|---|---|---|---|---|---|
| Dat04 | 0.1802 | -0.4934 | 0.4754 | -0.1711 | -0.0114 | 0.0203 |
| Dat05 | -2.5002 | 7.0497 | -7.384 | 3.5285 | -0.7649 | 0.0667 |
| Dat06 | -2.3776 | 6.7125 | -7.042 | 3.3714 | -0.733 | 0.0645 |
| Dat07 | -3.0906 | 8,623 | -8.9026 | 4.1713 | -0.8796 | 0.0731 |
| Dat12 | -0.1109 | 0.2489 | -0.1651 | 0.0436 | -0.0397 | 0.0231 |
| Dat13 | -0.0199 | 0.0666 | -0.0407 | -0.0139 | -0.009 | 0.0171 |
| Dat14 | -0.8999 | 2.5768 | -2.7479 | 1.3551 | -0.326 | 0.406 |
| Dat15 | 0.2801 | -0.6182 | 0.4105 | -0.0403 | -0.054 | 0.0226 |
| Dat16 | -0.3573 | 1.089 | -1.246 | 0.6667 | -0.1835 | 0.0306 |
| Dat17 | -2.169 | 6.1885 | -6.5889 | 3.2223 | -0.722 | 0.0655 |
| Dat18 | -1.7514 | 4.9276 | -5.1141 | 2.3953 | -0.5085 | 0.0484 |
| Dat19 | -2.1852 | 6.1433 | -6.3867 | 3.0092 | -0.6415 | 0.0576 |

(Distortion parameters: M=1.0, u=1.5, k=0.9)

**Table A4.6: Data files with 131072 samples**

| Data file | 5th order | 4th order | 3rd order | 2nd order | 1st order | Constant |
|---|---|---|---|---|---|---|
| Dat04 | 0.0068 | -0.0721 | 0.1202 | -0.0709 | 0.0017 | 0.0138 |
| Dat05 | -1.7962 | 5.074 | -5.5322 | 2.5903 | -0.5723 | 0.0531 |
| Dat06 | -1.5444 | 4.4188 | -4.7341 | 2.3346 | -0.5283 | 0.0508 |
| Dat07 | -2.3019 | 6.4855 | -6.7816 | 3.2228 | -0.6884 | 0.0595 |
| Dat08 | 0.0652 | -0.1138 | 0.0624 | -0.0284 | 0.0022 | 0.0125 |
| Dat13 | 0.18 | -0.391 | 0.2799 | -0.0889 | 0.0088 | 0.0113 |
| Dat15 | 0.3406 | -0.9083 | 0.8214 | -0.2827 | 0.0144 | 0.0144 |
| Dat17 | -1.3741 | 3.9433 | -4.2617 | 2.1415 | -0.5023 | 0.0508 |
| ***Data files with 196608 samples*** | | | | | | |
| Dat05 | -1.5145 | 4.3026 | -4.5759 | 2.2422 | -0.507 | 0.0496 |
| Dat06 | -1.3085 | 3.8301 | -4.1936 | 2.1108 | -0.4907 | 0.0499 |
| Dat07 | -1.5871 | 4.4576 | -4.6482 | 2.2061 | -0.4777 | 0.0462 |
| Dat17 | -1.8194 | 5.1635 | -5.4854 | 2.6857 | -0.6043 | 0.0564 |
| ***Data files from VHS video with distorted sound track*** | | | | | | |
| Data100 | 0.0996 | -0.2621 | 0.2213 | -0.0775 | 0.0074 | 0.0105 |
| Data101 | -0.0208 | -0.0084 | 0.0151 | 0.0089 | -0.0051 | 0.0088 |
| Data102 | 0.145 | -0.3342 | 0.2571 | -0.0822 | 0.0009 | 0.0132 |
| Data103 | 0.0240 | -0.189 | 0.2373 | -0.0953 | 0.0135 | 0.0073 |
| Data104 | -0.0125 | 0.0267 | -0.696 | 0.0856 | -0.0492 | 0.018 |
| Data105 | -0.1659 | 0.4884 | -0.4609 | 0.1457 | -0.0228 | 0.0159 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Data106 | 0.2035 | -0.5927 | 0.6045 | -0.2185 | -0.0218 | 0.025 |
| Data107 | 0.0339 | -0.3676 | 0.7020 | -0.4437 | 0.0530 | 0.0216 |

(Distortion parameter: M=1.0, u=1.5, k=0.5)

# ANNEXURE 5

## Recent publications resulting from this research project


1. Maré, S. (2002). “Detection of non-linear distortion in audio signals”, IEEE Trans. Broadcasting, 8, No. 2, June 2002.
2. Maré, S. (2004). “Detection of Distortion Due to Mechanical Imperfections in Recording and Playback Apparatus”, IPET Conference 2004, Tswane University of Technology, Pretoria, 28- 29 September 2004.