Section of Acoustics
Institute of Electronic Systems
Aalborg University

# Perception & Thresholds of Nonlinear Distortion using Complex Signals

Group 1061:
Eric Mario de Santis
Simon Henin

Submitted: June 7th 2007

Department of Electronic Systems
Aalborg University

AALBORG UNIVERSITY

**Section of Acoustics**

**THEME:**
Acoustics

**SUBJECT:**
Investigation of thresholds of nonlinear distortion using improved metrics

**TITLE:**
Perception & Thresholds of Nonlinear Distortion using Complex Signals

**PROJECT PERIOD:**

1. February 2007 -
7. June 2007

**GROUP:**
1061

**GROUP MEMBERS:**
Eric Mario de Santis
Simon Henin

**SUPERVISOR:**
Per Rubak

**NUMBER OF DUPLICATES:** 5.

**NUMBER OF PAGES IN REPORT:** 65.

**NUMBER OF PAGES IN APPENDIX:** 17.

**TOTAL NUMBER OF PAGES:** 82.

Characterizing the perceptual effects of nonlinear distortion by means of conventional metrics such as Total Harmonic Distortion and Intermodulation Distortion has proven to be rather ineffectual. Conventional metrics have also proven unable to characterize the perception of nonlinear distortion in complex signals, and thresholds for the perception of nonlinear distortion have been limited to simple sinusoidal stimuli.

The use of improved metrics based on psychoacoustic principles is studied from the perspective of determining the threshold of perception of nonlinear distortion in complex signals. The Distortion Score (DS) and $R_{nonlin}$ metrics are implemented and investigated by means of a verification experiment to study their correlation to subjective perception of nonlinear distortion. Once verified, the metrics are used to determine the threshold of nonlinear distortion by means of another listening experiment.

Nonlinear distortion thresholds for four types of nonlinear devices are obtained using classical and jazz music samples. The thresholds for clipping distortion are found to be much lower than second or third order distortion systems. The clipping distortion types are also nearly independent on the music type. For the second and third order distortion systems, the obtained thresholds are dependent on the characteristics of the music sample.

# PREFACE

This report is written by Group 1061 at the Section of Acoustics at Aalborg University (AAU) and completed during the spring semester of 2007. The report provides documentation pertaining to the group's Master's thesis. The report investigates the audibility of nonlinear distortion and threshold estimates obtained using new nonlinear distortion metrics. The report itself is addressed to the staff and students at the Section of Acoustics at AAU and to anyone who has an interest in the perception of nonlinear distortion.

The report is divided into six chapters which include an introduction, problem analysis, implementation of the new nonlinear distortion metrics, design and analysis and experiment 1, design and analysis of experiment 2 and a final chapter containing both a discussion and conclusion. Graphs, measurement reports and other analysis not directly related to the report are included in the appendix.

A CD is provided along with the report containing:

- MATLAB code for listening test interfaces and selected simulations
- MATLAB code for the implemented nonlinear distortion metrics.
- Audio samples.
- The report in PDF format.

Eric Mario de Santis    Simon Henin

# CONTENTS

CHAPTER 1

# INTRODUCTION

## 1.1 Background

The inadequacy of traditional nonlinear distortion metrics and the audio industry's never ending pursuit of perfect sound reproduction has motivated research dedicated to the development of an appropriate metric describing the human perception of nonlinear distortion. Conventional methods of nonlinear distortion measurement only partly correlate with the perceived quality of reproduced sound. The task of providing a metric of nonlinear distortion is not a simple one. Such a metric must take many parameters into account. These parameters would include the dependence of distortion detectability on the temporal characteristics and the frequency content of the signals used in listening evaluations, and the correlation between the physical effects causing nonlinear distortion and their corresponding detectability [3].

The conventional methods of measuring nonlinear distortion are based on the measurement of distortion products excited by a sinusoid or two or more sinusoids. These metrics are commonly known as Total Harmonic Distortion (THD) and intermodulation distortion (IMD). They are typically expressed as a ratio between the distortion by- products to the total system output [10]. There are many problems with these metrics and to list them all is pointless as their main flaw is that these metrics are not at all correlated with subjective ratings of nonlinear distortion. That is, they do not describe how we as humans perceive nonlinear distortion and to what extent we perceive nonlinear distortion.

A metric is a typically a single value parameter that facilitates the quantification of the characteristics of a particular system. For instance, sound pressure can be a metric in the context of human sound perception or temperature can be a metric for human perception of heat. The audio industry has the need for a proper metric relating to the human sound perception of nonlinear distortion. Building amplifiers or loudspeakers with distortion products far below audibility is unnecessary and expensive. Many investigations by many researchers often suggest different thresholds of distortion audibility [3]. However, a proper metric which is correlated with subjective ratings of distortion can be used to properly quantify in some way the point where a listener can or cannot hear nonlinear distortion. The main theme of this thesis is the utilization of the new nonlinear distortion metrics proposed by Moore et. al. [25][26] to find this point, or threshold of nonlinear distortion audibility. These metrics, DS and $R_{nonlin}$ , have been found to be highly correlated with subjective ratings of nonlinearly distorted speech and music signals for a variety of distortion types.

## 1.2 Problem Statement

Proper metrics which are well correlated with subject ratings of nonlinear distortion have been made available as outlined in [25][26]. Obtaining a threshold in terms of these metrics would offer the audio industry assistance in the development of high quality products. Such thresholds of nonlinear distortion could improve the manner in which manufacturers prove the quality of their products and reduce costs by ensuring that those systems do not have distortion products which are far below audibility.

### Project Goal

The goal of this project is to obtain threshold estimates using the new nonlinear distortion metrics, DS and $R_{nonlin}$. These estimates will be found for different types of music and for different nonlinear distortion types. The following list details the main goals which this thesis will evaluate:

1. The influence of the music sample on the obtained threshold.
2. The influence of the distortion type on the obtained threshold.

### Project Scope

In order to arrive at the nonlinear distortion threshold using the new metrics the following steps must be taken:

1. Understanding and Implementation of the nonlinear distortion metrics DS and $R_{nonlin}$.
2. Verification of the new metrics' correlation with subject ratings of nonlinear distortion.
3. Design of a listening experiment to obtain the nonlinear distortion thresholds.
4. Analysis of the obtained thresholds in relation to the project goal.

CHAPTER 2

# PROBLEM ANALYSIS

*This section presents the fundamental theories behind nonlinear systems as well as important psychoacoustical principles relating to the perception of nonlinear distortion. Finally, an overview of conventional metrics used in evaluating nonlinear distortion is presented along with alternate distortion metrics that aim at developing a metric that relates nonlinear distortion to subjective perception.*

## 2.1 Nonlinear Systems and Distortion

Signal distortion resulting from acoustical transducers and transmission channels can be classified as being either linear or nonlinear. Linear distortion affects the amplitudes and phases of the frequency components present in a complex signal. This type of distortion can be compensated for by applying linear filtering methods. As an example, an equalizer can be used to compensate for the undesirable frequency response caused by a certain loudspeaker. In contrast to linear distortion, nonlinear distortion injects frequency components that were not present in the original signal. The effects of nonlinear distortion are difficult and sometimes impossible to compensate for [25].

A linear system is described as having the following mathematical properties:

1. Additivity: $f(x+y) = f(x) + f(y)$
2. Homogeneity: $f(ax) = a * f(x)$

Together, these two properties of linear systems are referred to as the principle of superposition. A system is said to be nonlinear if its input and output characteristics are not linearly related mathematically. That is, the system does not obey the principle of superposition.

The input-output relationships shown in figure 2.1 illustrate three distortion types: linear (top), asymmetrical clipping (middle), and symmetrical clipping (bottom). The output of the first linear function , $f_1(x)$, is described by

$$f_1(x) = 0.5x \tag{2.1}$$

where $x$ is the input signal.

The function, $f_2(x)$, describing the asymmetrical clipping is:

$$f_2(x) = \begin{cases} x & \text{if } x < 0.5 \\ 0.5 & \text{if } x \geq 0.5 \end{cases} \tag{2.2}$$

This system is referred to as asymmetrical nonlinear distortion as the clipping is only applied to half of the waveform (positive cycle).

The function, $f_3(x)$, describing the symmetrical nonlinear distortion is defined by:

$$f_3(x) = \begin{cases} x & \text{if } 0.5 < x < 0.5 \\ 0.5 & \text{if } x \geq 0.5 \\ -0.5 & \text{if } x \leq -0.5 \end{cases} \tag{2.3}$$

As the clipping is applied to both positive and negative cycles, the distortion is referred to as being symmetrical.

A discrete time input signal, $x(n) = sin(2\pi 1000nT)$, is plotted at the top of figures 2.2, 2.3 and 2.4 with a sampling frequency of $fs = 44100$ Hz. The output signal resulting from the signal, $x(n)$, passing through the distortion systems described above is plotted in both the time and frequency domain.

The output signal from the linear distortion system shown in figure 2.2 has the same phase as the original signal differing only in amplitude. The change in amplitude is also evident in the frequency domain as shown in the bottom graph of figure 2.2. As expected, the frequency spectrum from this linear system contains the same frequency component as the original signal. The original frequency is known as the fundamental.

The output signal plotted in figure 2.3 has the same amplitude as the original signal on the negative cycle but has been clipped to 0.5 on the positive cycle. This type of distortion injects both even and odd order harmonics throught the spectrum decreasing in magnitude with increasing frequency. An even harmonic occurs at even multiples of the fundamental frequency, $2f, 4f, 6f, 8f$..... Odd order harmonics occur at odd multipes of the fundamental frequency, $1f, 3f, 5f, 7f$..... The presence of frequency components not existing in the original input signal characterizes the systems nonlinear behavior.

The output signal from shown in figure 2.4 has been clipped to 0.5 on both the positive and negative cycles. Overdriven solid state devices exhibit this type of nonlinear distortion. In contrast with the asymmetrical distortion, the added frequency components occur at only odd multiples of the input frequency.

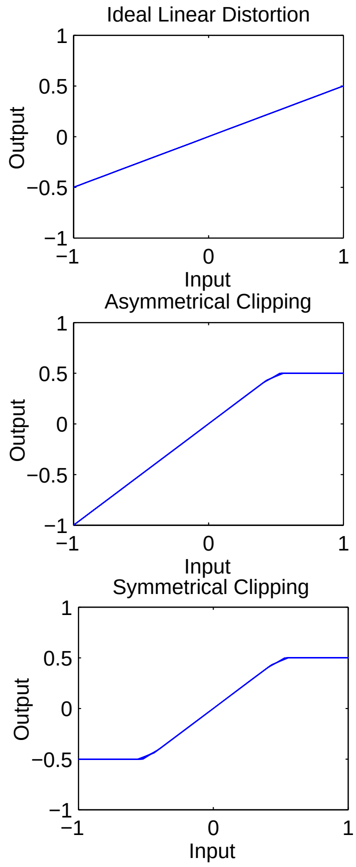



***Figure 2.1:*** *Examples of nonlinear distortion. (1) Ideal linear distortion. (2) Asymmetrical clipping. (3) Symmetrical clipping.*

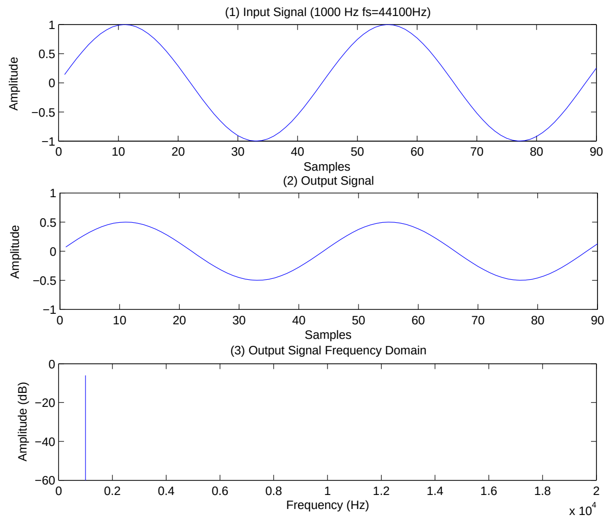



***Figure 2.2:*** *Linear Distortion. (1) Input signal x(n). (2) Output signal from linear distortion. (3) Output signal frequency response.*

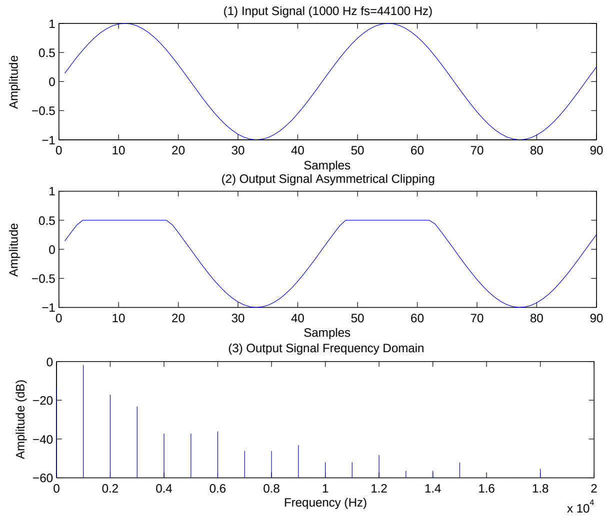



***Figure 2.3:*** *Asymmetrical Distortion. (1) Input signal x(n). (2) Output signal from asymmetrical nonlinear distortion. (3) Output signal frequency response.*

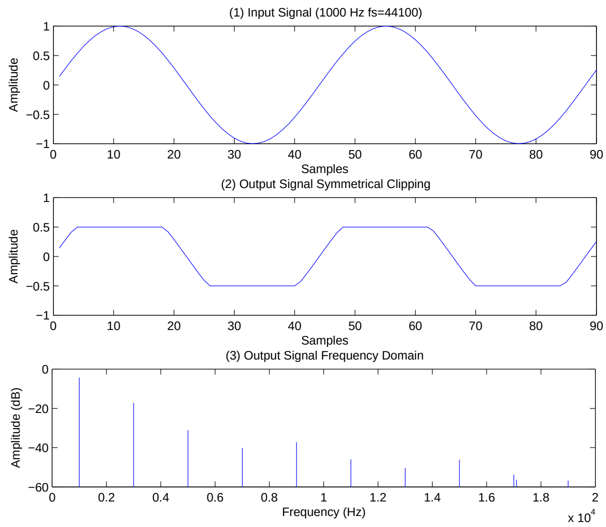



***Figure 2.4:*** *Symmetrical Distortion. (1) Input signal x(n). (2) Output signal from symmetrical nonlinear distortion. (3) Output signal frequency response.*

## 2.2 Modeling Nonlinear Systems

The input-output relationships in figure 2.1 can be expanded in many different ways [10]. The more involved mathematical expansions would represent the curves as Legendre Polynomials, Chebycheff Polynomials, Laguerre Polynomials and also as a Fourier series. For the sake of simplicity, simple polynomial expansion is presented. The polynomial expansion is of the form:

$$f(x) = \sum_{n=0}^{N} a_n x^n \tag{2.4}$$

where the first term $a_0$ is the offset or DC term. The second term, $a_1 x$ is the gain of the system which is also the linear portion of the expansion. The third term, $a_2x^2$, is the second order (quadratic) nonlinearity. The third term, $a_3x^3$ is the third order (cubic) nonlinearity. Even order polynomials will contribute only even order harmonics and odd order polynomials odd harmonics. For example, the polynomial, $x^5$ will have harmonic components at multiples of $1, 3, 5$ times the fundamental frequency. For a pure tone, the output of the nonlinear system will contain only harmonic components of the input frequency. However, the output resulting from a more complex tone being passed through the system will not only contain the harmonic components of the frequencies present in the tone, but also sum and difference components between those frequencies. These sum of difference components are known as intermodulation products.

To illustrate mathematically how the output of a nonlinear system produces both harmonic and intermodulation products, a general equation for cubic nonlinearity is considered in equation 2.5. An input signal consisting of two frequencies is shown in equation 2.6.

$$f(x) = a_1x + a_3x^3 \tag{2.5}$$

$$x = bsin(\omega_1 t) + csin(\omega_2 t) \tag{2.6}$$

$$f(x) = \begin{cases} a_1\Big(bsin(\omega_1 t) + csin(\omega_2 t)\Big) + \\ a_3\Big((\frac{3b^3}{4} + \frac{3bc^2}{2})sin(\omega_1 t) + (\frac{3c^3}{4} + \frac{3b^2c}{2})sin(\omega_2 t)\Big) - \\ \frac{a_3}{4}\Big(b^3sin(3\omega_1 t) + c^3sin(3\omega_2 t)\Big) - \\ \frac{3a_3b^2c}{4}\Big(sin((\omega_2 - 2\omega_1)t) + sin((2\omega_1 + \omega_2)t)\Big) - \\ \frac{3a_3bc^2}{4}\Big(sin((\omega_1 - 2\omega_2)t) + sin((2\omega_2 + \omega_1)t\Big) \end{cases} \tag{2.7}$$

The output signal shown in the above equation can be described as follows: the first line in equation 2.7 represents the linear term which is simply a gain determined by the coefficient $a_1$. The second line in equation 2.7 is the first order distortion product and the third line in equation 2.7 represents the third order harmonic distortion products. The fourth and fifth lines show the intermodulation distortion products resulting from the interaction between the two input frequencies.

To illustrate the use of these expansions, polynomial expansion was applied to the nonlinear distortion systems described in the above section. The frequency components present in the output signals from both the asymmetrical and symmetrical nonlinear distortion systems contain high order harmonic components spanning up to 18th harmonic. Intuitively, a high order expansion will be able to model the nonlinear system more accurately.

The MATLAB function, $polyfit$, was used to generate an nth order polynomial based on two data vectors. These data vectors contain both the input and output signals from each system. The function fits the data $p(x)$ to $y$ in the least squared sense. The function returns an nth order polynomial according the to the equation below.

$$p(x) = p(1)x^n + p(2)x^{n-1} ... p_n x + p_{n+1} \tag{2.8}$$

Both nonlinear distortion system were modeled with two polynomials of different order. The first expansion order was selected as a third order polynomial as this would be the minimum order possible to represent both odd and even order harmonics. The higher order polynomial expansion was of 7th order. The input sequence was the same 1000 Hz discrete time signal used in the previous section.

Figure 2.5 plots the time and frequency domain output from the 3rd and 7th order polynomial expansions of the asymmetrical nonlinear distortion. The clipping in the 7th order model is more defined than that of the lower order representation. The 7th order model has more frequency components than the 3rd order model and in both cases even and odd order harmonics are present. The amplitude of the represented harmonic components is nearly equal to the equivalent harmonic components in the actual frequency response of the asymmetrical system in figure 2.3.

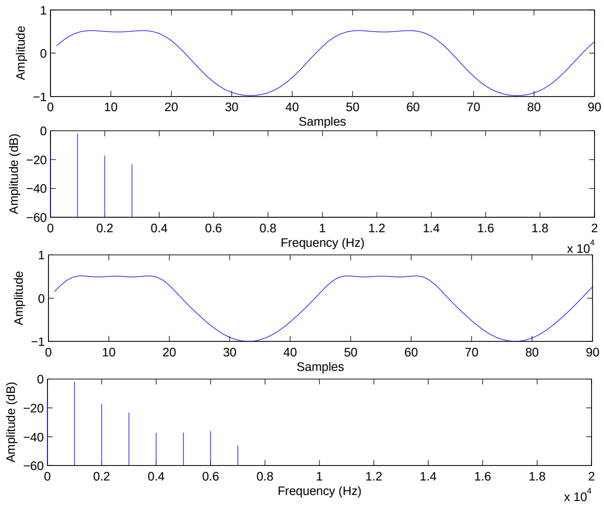



***Figure 2.5:*** *Polynomial expansion of asymmetrical type distortion. (Top: Output signal from 3rd order model. Second from Top: Frequency spectrum of output signal. Third from top: Output signal from 7th order model. Bottom: Frequency spectrum of output signal.)*

The results of polynomial expansion representing the symmetrical clipping is analagous to that of the asymmetrical model as shown in figure 2.6. It is highlighted that the even order coefficients of the polynomial expansion are zero and only odd order harmonic components are present.

## 2.3 Psychoacoustics of Distortion Perception

This section presents a brief overview of the psychoacoustic concepts relevant to the human perception of distortion.

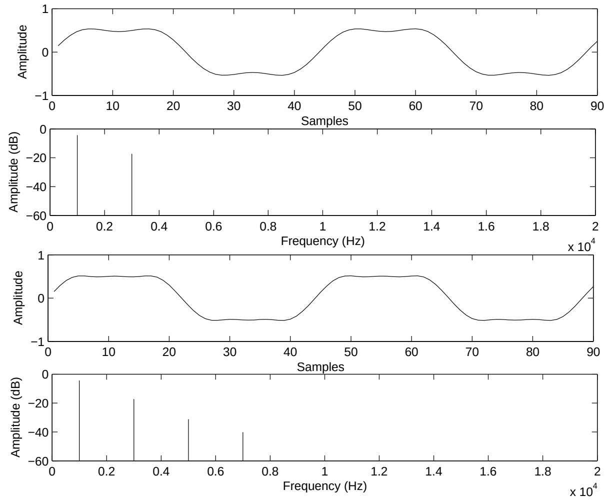



*__Figure 2.6:__ Polynomial expansion of symmetrical type distortion. (Top: Output signal from 3rd order model. Second from Top: Frequency spectrum of output signal. Third from top: Output signal from 7th order model. Bottom: Frequency spectrum of output signal.)*

## Hearing Thresholds

The threshold of human hearing has been extensively studied and is described in a number of hearing threshold curves by Fletcher & Munson and Robinson & Dadson, to name a few. Along with providing the absolute threshold of hearing, the curves provide equal loudness contour curves which show the relative sensitivity of the human ear with respect to frequency in a free field as shown in figure 2.7. The hearing threshold is dependent on listening environment and various curves are needed to describe these different environments fully.

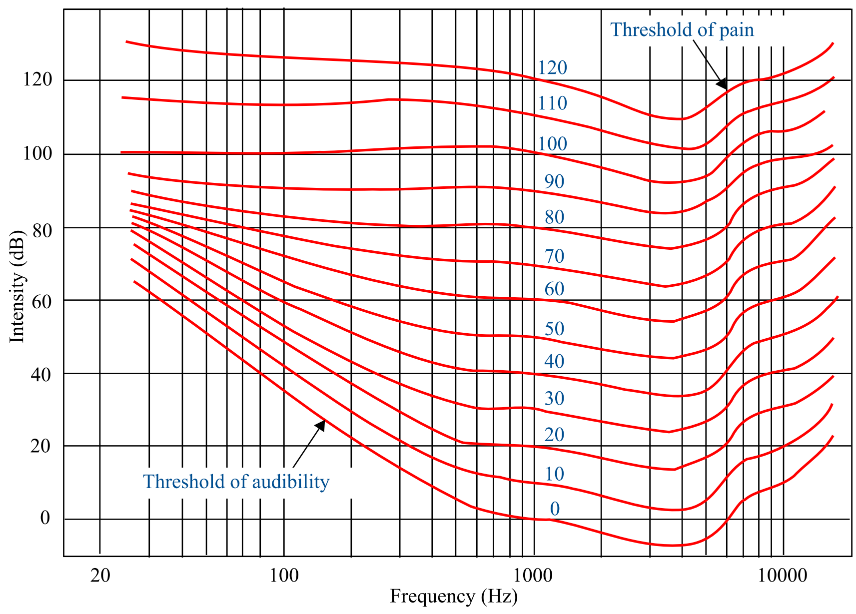



*__Figure 2.7:__ Fletcher-Munson curves. Numbers above each curve represent the phon value of the curve.*

The equal-loudness curves, such as the ones shown in figure 2.7, are derived from empirical data from a large number of subjects. Subjects are presented with a 1 kHz reference tone at a specified dB (phon) level and asked to match the loudness of other frequencies to the loudness of the 1 kHz tone. For example, the 40 phon curve shows that a tone at 100 Hz will need to have a level of approximately 60 dB to be perceived to be as loud as a 1 kHz tone at 40 dB. Equal loudness curves are important as they show that the human ear is less sensitive to low and high frequencies. Therefore, distortion products residing in these frequency regions will not contribute as much to the perception of distortion if there level is not sufficient.

### Auditory Filtering and the Critical Band Concept

Beginning with Fletcher (1940), the notion of auditory filtering was introduced. Through experiments determining the threshold of audibility of a sinusoidal signal in the presence of a noise masker, Fletcher found that at a certain point increasing the bandwidth did not effect the threshold. To explain this, the concept of auditory filters was developed which suggested that the ear can be thought of as consisting of a set of auditory filters, or critical bands. These filters are a function of the movement of the basilar membrane (BM) where specific frequency components will excite different regions on the BM. This frequency dependent movement can be thought of as the auditory system's frequency analyzer. These filters have been termed the critical bands of the auditory system.

The movement of the BM in response to a pure tone is not precisely located, and the movement will be spread over a small area of the BM. This area can be thought of as the frequency resolution of the auditory system. This frequency resolution is directly related to the concept of critical bands which is defined empirically as the effective bandwidth of an auditory filter. Therefore, the auditory system can be thought of as having a series of filters, or critical bands, with a certain bandwidth.

The concept of critical bands is important for understanding the interaction of sounds and the human perception of sounds. In general, energies of sounds with the same critical band will interact, causing beating, roughness, and masking effects. Whereas the energies of sounds in different critical bands will not interact, but will contribute to the perceived loudness.

Much research has been devoted to determining the bandwidth of these critical bands based on empirical data. Improvements to Fletcher's initial experiments [5] using band-widening experiments led to the 'critical bandwidth' estimation of the bandwidth of the auditory filters, also known as the Bark scale. Patterson [14] developed a method termed the 'notched-noise method', which uses a noise masker with a notch centered about a sinusoidal center frequency. The width of the notch is then varied and the threshold of the sinusoidal signal is determined. As the spectral notch is increased, less and less noise passes through the auditory filter centered around the sinusoidal signal and at a certain point increasing the bandwidth of the notch has no effect on the sinusoidal threshold. This indicates that the bandwidth of the auditory filter has been exceeded. The disadvantage of this method is that it assumes that the auditory filter is symmetric. This method was used by Glasberg & Moore [6] to develop the Equivalent Rectangular Bandwidth (ERB) model for the bandwidth of the auditory filters. The ERB is a measure of the bandwidth of the auditory filters assuming a rectangular shape as an approximation.

Both methods of determining the bandwidth of the auditory filters have shown to be consistent with one another and with empirical data. The main difference is at low frequencies where the Bark scale estimates the bandwidth to be constant below 500 Hz, whereas the ERB scale estimates a decrease in the bandwidth with decreasing frequency below 500 Hz. Figure 2.8 shows the difference between the two models.

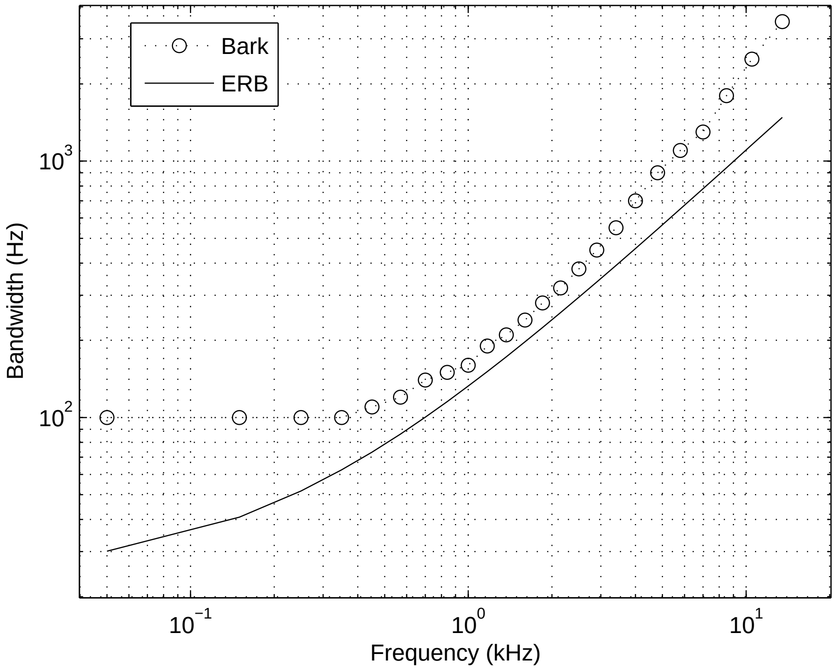



***Figure 2.8:*** *Bark scale and ERB scale*

Besides the noise-widening and notched-noise methods used to investigate the bandwidth of the auditory filters, a number of other studies have revealed other interesting details about the auditory filter. Using a notched-noise method with varying masker level and various signal frequencies, Rosen as well as Glasberg & Moore [6], found that the shape of the auditory filter was level dependant. That is, the shape of an auditory filter centered around a particular frequency varies depending on the input level of a test signal as shown in figure 2.9. As the level of an input signal is increased, the slope on the low frequency side of the auditory filter becomes more shallow. In the higher frequency side, only very small variations were found.

## Auditory Modeling

Obtaining a model of the peripheral auditory system has been of particular concern to researchers for some time. As not all mechanisms involved in the auditory system's processing of acoustic stimuli are known, much of this research has been devoted to modeling the overall process in two main steps; namely, the filtering process produced on the basilar membrane (BM), and the conversion of these filtered signals into representative neural activity produced by the auditory system. As the biological mechanisms

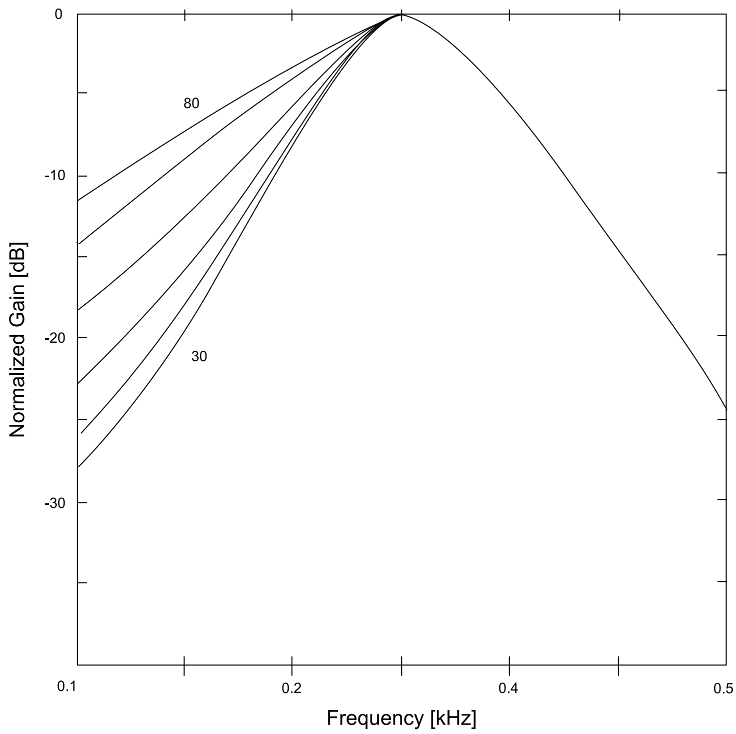



***Figure 2.9:*** *Shape of the auditory centered around 250Hz for sinusoidal input levels between 30-80dB [6].*

of the auditory system are not of concern in this report, but rather the signal processing representation of signals involved in the perception of sound, the latter of these two steps is omitted from further discussion.

Estimating the magnitude response, or shape, of the auditory filters in the filter bank often used to describe the auditory filtering process has been the aim of this research. Most of this modeling has attempted to fit psychoacoustic measurements, such as the thresholds derived using the notched-noise method, to obtain a function describing the shape of the auditory filters. Many researchers, such as Patterson, Meddis and Lyon, to name only a *few*, have developed models and only a few will be discussed here.

The simplest technique would be to use the $\text{ERB}_N$'s described in the previous section to develop a filter bank of ideal rectangular filters. However, this would only present an idealized model of the peripheral auditory system. This simplified model is based in the assumption that power is summed in 1-$\text{ERB}_N$ wide bands.

In [14], Patterson evaluated thresholds obtained using various techniques and found that a good approximation to the central portion of the auditory filter's shape could be defined as,

$$|H(f)|^2 = e^{-\pi(\Delta f/f_0 ERB)^2} \tag{2.9}$$

The main assumption was that when the "notch" of notched-noise centered around a particular test tone was varied, the threshold (in dB) of the test tone decreased nearly linearly with an increasing notch width. This indicated that the shape of the auditory could be described approximately by an exponential function.

Patterson's initial model shown in equation 2.9 was further revised into the concept of the "rounded-exponential", or $roex$, function [20]. The $roex$ model assumed that the filter shape could be represented by a pair of back-to-back exponential functions. The simplest form of one of these exponential functions is described by,

$$W(g) = (1 + pg)e^{-pg} \tag{2.10}$$

where $g$ is the distance from the filter center frequency, $f_c$, to the evaluation point, $f$, normalized with respect to the signal frequency such that $g = |f - f_c|/f_0$ [19], and $p$ is a function parameter used to tune the bandwidth and the slopes of the skirts. Equation 2.10 is termed the $roex(p)$ function as its only parameter is $p$ and is used primarily to model the passband of the auditory filter. The level-dependent asymmetry of the auditory filter could be implemented by using different parameter values on the lower and upper halves of the model ($roex(p_l; p_u)$).

The $roex$ function was further developed into a family of $roex$ functions used to model different parts of the auditory filter such as the tail of the auditory filter $roex(t)$ [19]. The auditory filter could then be modeled as a linear combination of $roex(p)$ and $roex(t)$ as the $roex(p,t)$. Additional roex models include the $roex(p,w,t)$ which uses a component weight, $w$, to weight the slopes of the passband and tail functions, as well as the $roex(p,r)$ model which uses a fixed floor used to model the relatively flat thresholds of older listeners [19]. A plot of a few typical $roex$ models is shown in figure 2.10.

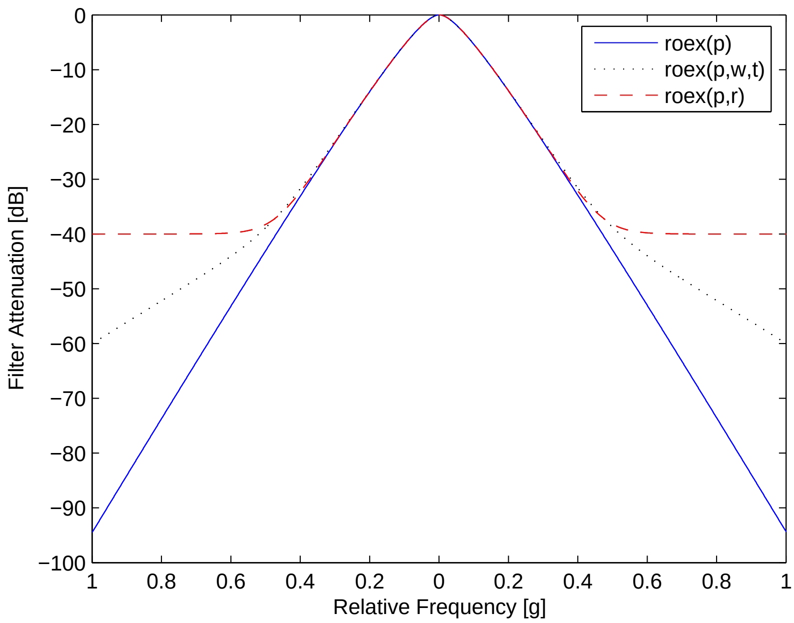



***Figure 2.10:*** *A comparison of the $roex$ family of auditory filters. All filters created using typical parameter values as proposed by Patterson [19]. [$p = 25$, $w = 0.002$, $t = 10$, $r = 0.0001$]*

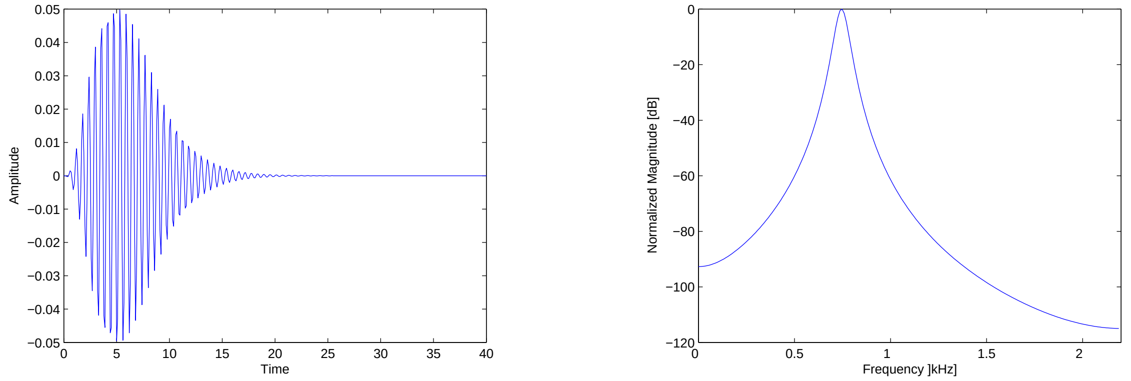
Although the $roex$ models were found to fit measured data very well, the obvious limitation is that they only offer a frequency domain model. Roex models are used solely to filter stimuli in the frequency domain by specifying the relative attenuation of the modeled filters. Methods implementing an inverse-fft to create the time domain representations failed to capture correctly the phase response of the auditory filters and were not defined for discontinuous models, such as the $roex(p, w, t)$ [29].

Boer [1] first proposed the gammatone function for modeling the shape of the impulse response function of the auditory system as estimated by a reverse correlation ("revcor") function of neural firing times in cats [15]. The shape of the magnitude response was found to be very similar to that of the $roex$ model at moderate sound levels, and so it was developed into a time-domain model of the auditory filter by Patterson [17, 16, 15]. The gammatone filter is described as,

$$g(t) = \alpha t^{n-1} e^{-2nbt} cos(2nf_c t + \phi),\ for\ t > 0 \tag{2.11}$$

where $\alpha$ is an arbitrary factor that is typically used to normalize the peak magnitude of the transfer function to unity, where $n$ is the filter order, $b$ is the impulse duration, $f_c$ is the center frequency, and $\phi$ is the phase. The main parameters are $b$ and $n$, where $b$ mainly determines the duration of the impulse, and therefore the bandwidth of the filter, and where $n$ determines the slope of the filter skirts [15]. The gammatone filter provides a good model of the spectral analysis in humans at moderate sound levels where the shape of the auditory filter is relatively symmetric on a linear frequency scale [8].

Typically, fourth order gammatone filters are used which are said to be an appropriate model to use for simulating the cochlear filtering of broadband sounds such as speech and music, especially when the sound level is in the broad middle range of hearing [16]. An example of a fourth order gammatone filter's impulse response and its corresponding frequency response are shown in figure 2.11.




(a) Example of the impulse response of a gammatone filter

(b) Example of the frequency response of a gammatone filter

***Figure 2.11***

More recently, modified versions of the gammatone filter have been developed to take into account the level dependence of the auditory filters by Patterson and Irino [18].

## Auditory Masking

The concept of auditory masking refers to the psychoacoustic phenomenon of one stimulus, audible in isolation, being rendered inaudible in the presence of a masking stimulus. A typical example would be the masking of speech by background noise such as road traffic.

In general, masking depends on loudness where a loud sound will mask a soft sound. Masking also depends on frequency. In general, sounds mask frequencies higher than the frequency of the sound itself more than frequencies below the masker. Additionally, sounds within a critical band mask each other more than those more than a critical bandwidth apart. Complex tones are more difficult to mask than pure tones.

The primary mechanisms thought to be involved in the process of masking are swamping and suppression. Swamping suggests that the neural activity evoked by the masker may render the neural activity produced by the target stimulus undetectable. Suppression suggests that the masker stimulus may suppress the neural response of the target stimulus.

Masking is a principle concept in the perception of distortion, as distortion products will only contribute to the percept of distortion if they are not masked by the primary stimulus or other distortion products. Figure 2.12 shows the masking threshold for a pure tone in the presence of narrow band noise as determined by Zwicker.

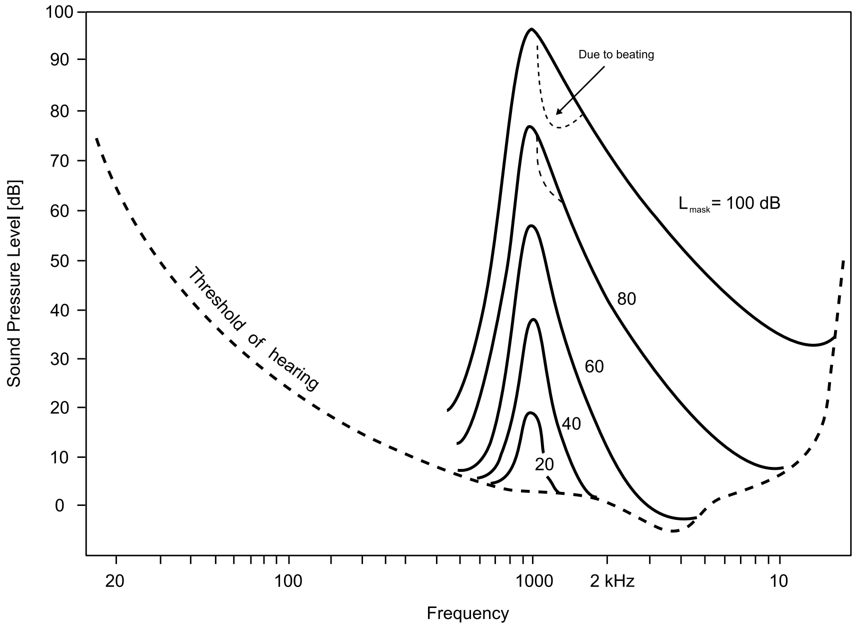



***Figure 2.12:*** *Zwicker Masking Threshold for a test tone in the presence of narrow band noise centered at 1kHz. [after Zwicker] [5, p.63]*

Figure 2.12 is derived from empirical data, showing the masking threshold of a test tone in the presence of narrow band noise centered at 1 kHz at various levels of the narrowband noise. The threshold is at its peak at 1 kHz. As the test tone decreases in frequency, the threshold decreases rapidly, whereas, when the test tone frequency is increased past 1 kHz, the threshold decreases more slowly. From figure 2.12, two things should be noted; first, masking predominately effects higher frequencies rather than lower frequencies [5]. Second, that the masking effect increases nonlinearly with an increase in the masker level. As an example, an increase from 60 dB to 80 dB in the noise level of the masker will produce a 30 dB increase in the masking threshold at 3 kHz, as seen in the Zwicker masking threshold curve.

In the case of harmonic distortion, this masking threshold could have an effect on different harmonics with relation to their distance to the fundamental. As an example, given a nonlinear system producing 2nd and 3rd harmonics with equal amplitudes, the 2nd harmonics would be masked more than the 3rd harmonic. This principle can also be applied to higher order harmonics with the idea being that higher order harmonics will be perceived more than lower order harmonics and will therefore be more perceptible.

### Factors Effecting Distortion Perception

Of particular note in the psychoacoustics of distortion perception are frequency discrimination of distortion by-products as well as temporal effects of distortion perception.

It has been found that harmonic distortion below 400 Hz is harder to detect than harmonic distortion above 400 Hz [12, 27]. This can be partially explained by the fact that the threshold of hearing increases at low frequencies.

Additionally, temporal effects have an impact on the perception distortion due to the finite time resolution of the ear. In studies conducted by Moir [12], it was found that the "just detectable" distortion decreased with increased presentation time. Specifically, it was found that for a 4ms tone burst distorted by clipping, the just detectable distortion reached approximately 10%, while increasing the presentation time to 20ms reduced the just detectable distortion level to 0.3% [12] .

## 2.4 Conventional Distortion Metrics

The classic distortion metric for harmonic distortion is known as total harmonic distortion (THD). The THD is defined as the ratio of the square of the root-mean-square (rms) values of the harmonics to that of the fundamental. The THD can be expressed mathematically as:

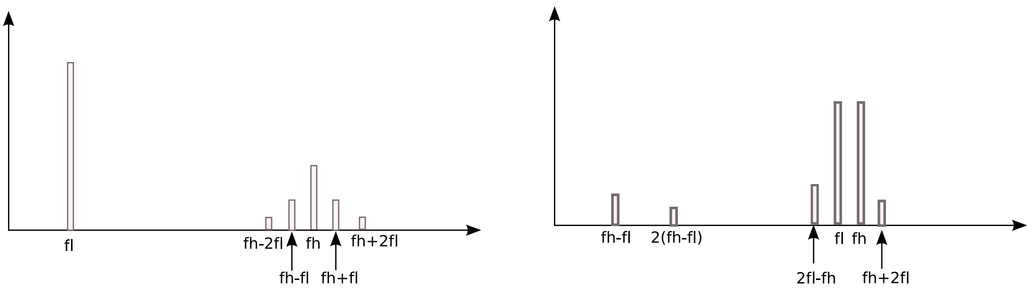
$$\%THD = 100\frac{\sqrt{V_2^2 + V_3^2 + V_4^2 + ...V_n^2}}{V_T} \tag{2.12}$$

where $V_n$ is the rms value of each harmonic component and $V_T$ is the rms value of the fundamental. Typically, a high purity sine wave is used as an input to the nonlinear system to excite the harmonic components. A frequency analyzer can be used to measure the rms value of all the harmonic components and the THD can be found using the above equation. This procedure is rather tedious and is it quite often quite difficult to measure harmonic distortion products with precision (especially since most solid-state audio devices have such low distortion).

The THD+N metric is often provided instead of the THD. This method is based on the same procedure as for the THD, however, this parameter also includes any other noise present in the system. The numerator of the above ratio is determine by removing the fundamental from the output of the nonlinear device by means of a notch filter. The total rms voltage of this signal includes all the harmonic components plus any other noise. The denominator of the above equation is the rms level of the entire signal including the fundamental, harmonics and additional noise. Removing the fundamental by subtracting the output of the nonlinear device by the input signal could also be performed instead of using a notch filter. However, many systems will provide some phase shifting and as a result the simple subtraction would not work as the output and input fundamental would not be in phase.

Using a pure tone input to a nonlinear device does not excite intermodulation products. A metric, intermodulation distortion (IMD), quantifies distortion products not related harmonically to the fundamental. Two standard methods for measurement and evaluation of the IMD will be discussed. These standards are the SMPTE test (Society of Motion Picture and Television Engineers) and the CCIF. For the SMPTE IMD method two standard frequencies at $f_l$=60 Hz and $f_h$=7 kHz with a 4:1 amplitude difference (12dB) are mixed together for the input signal to the nonlinear system. The upper intermodulation components are spaced at multiples of the lower frequency component as shown in Figure 2.13(a). The rms sum between the distortion products is evaluated and expressed as a ratio against the rms value of the upper frequency component (7 kHz).

In contrast to the SMPTE method, the CCIF (difference frequency distortion) method uses two frequencies of equal amplitude with a 1 kHz difference. The distortion products can be found in Figure 2.13(b). The rms sum between the distortion products is evaluated and expressed as a ratio against the rms value of the input signal. Even order distortion produces the lower difference frequency components and the odd order the higher difference frequency components closer to the input signals . Most applications of this test only measure the lower even order distortion products [2]. Typical input frequencies are often 14 kHz and 15 kHz, which effectively eliminate any harmonic contribution to the measurement.

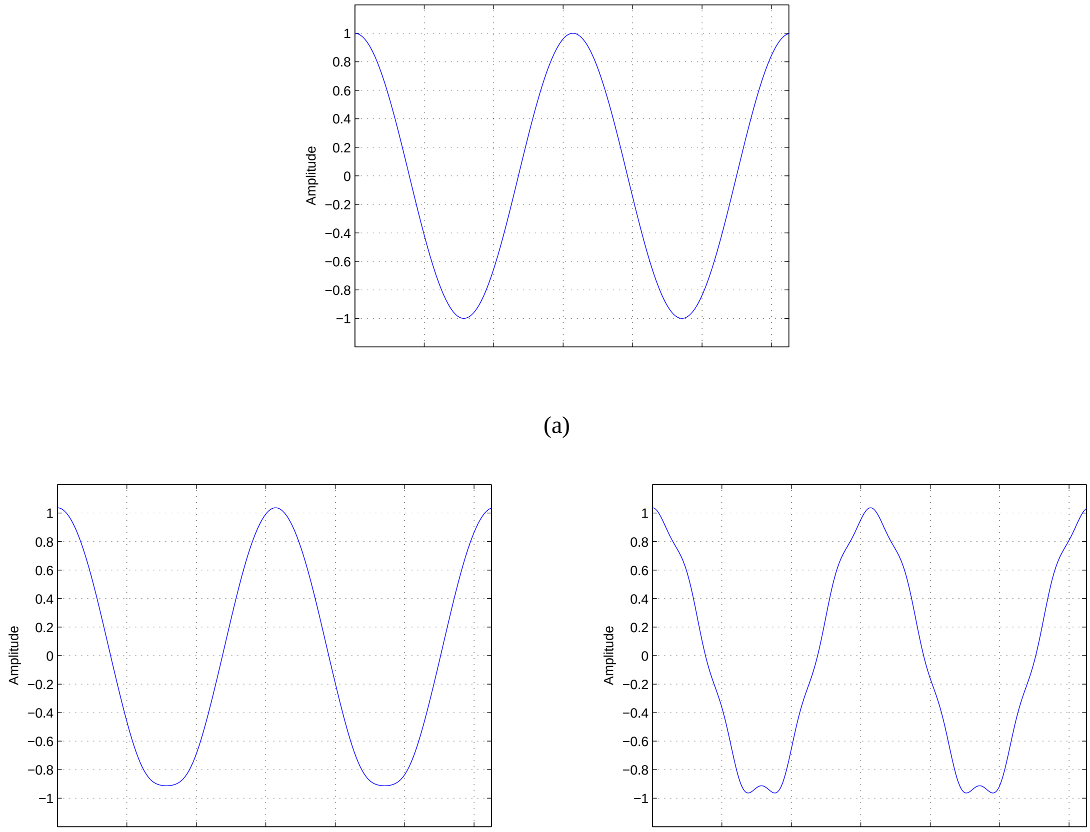



(a) SMPTE Test for IMD (b) CCIF Test for IMD

***Figure 2.13:*** *Comparison of the SMPTE and CCIF methods for calculating the IMD metric.*

## Problems with THD Metric

As described above, the THD provides a measure for the total contribution of harmonic distortion by-products in a resultant output signal. As a metric, it provides a good indication of the amount of harmonic distortion produced by an amplifier, however it says nothing of the type (i.e. order, distribution) of the distortion.




(a) (b) (c)

***Figure 2.14:*** *a) Undistorted Signal b) Low-Order Harmonic Distortion [5% 2nd, 2% 3rd, 1% 4th, 0.5% 5th, 0.2% 6th] c) High-Order Harmonic Distortion [0.2% 2nd, 0.5% 3rd, 1% 4th, 2% 5th, 5% 6th]*

Figure 2.14 shows the problem with the THD metric. Figures 2.14(b) and 2.14(c) show two output signals with the same overall THD. However, the output signal in figure 2.14(b) contains lower order distortion products, whereas the signal in figure 2.14(c), contains higher order distortion products.

It is easy to see that the signal shown in figure 2.14(c) appears graphically to be more distorted than the signal shown in figure 2.14(b). This can be noticed from the evident ripples present in the signal's time waveform shown in figure 2.14(c). This example was provided to highlight that two signals having the same THD *may* not sound equally distorted.

## 2.5 Multitone Test Stimulus

The inadequacy of the current metrics such as THD and IMD is primarily a result of the test signals used in deriving the metrics. This consists of the use of single sinusoids or sweeping sinusoids in the case of the THD metric and a two tone signal in the case of the IMD metric. Although adequate in describing the specific contribution of harmonic or intermodulation components in a distorted signal, the test signals do not provide an accurate picture of the distortion introduced in more realistic signals, such as music signals.

The THD and IMD metrics are based on the output of a nonlinear device to a single tone or two tone input test signal. These test signals only excite either harmonic distortion products or intermodulation products. This fails to encapsulate the complete interaction of all distortion products making it difficult correlate these metrics with a subjective relevant quantity. By using a multitone stimulus, a more complete picture of the nonlinear distortion can be realized.

The concept of using a multitone test stimulus was introduced by Czerwinski et. al. [3, 4] as means of capturing an increased amount of information as to the type and content of the distortion introduced by a nonlinear system. The multitone test stimulus can be described as,

$$x(t) = \sum_{i=1}^{N} A_i sin(\omega_i t + \phi_i) \tag{2.13}$$

where $\omega_i$ and $\phi_i$ are the frequency and the starting phase of the $ith$ tone, respectively.

The frequency components of the multitone signal are logarithmically spaced in order to place the fundamental tones into a non-harmonic relationship so as to avoid any distortion components being hidden within the primary signal. For example, if a two component multitone signal has frequencies at 1 and 2 kHz then second harmonic components could be produced at 2 and 4 kHz. The 2 kHz harmonic distortion product would interact with the 2 kHz component of the multitone signal. Additionally, the logarithmic spacing avoids periodicity of the resulting signal [3].

Determining the number and frequency distribution of the multitone components is primarily determined by the application. Increasing the number of tones in the multitone signal makes the signal resemble a noise or musical signal more closely. However, increasing the number of tones also increases the crest factor of the signal which is not desired. Minimization of the crest factor in the multitone signal is desirable since it increases the dynamic range of measurement by decreasing signal peaks, increasing the rms level of the signal and increasing the signal-to-noise ratio [3]. The crest factor is defined as,

$$CF = \frac{|A|_{max}}{A_{rms}} \tag{2.14}$$

In general, it is desired to have the smallest crest factor possible in the multitone signal in order to increase the dynamic range possible with the measurement. This is due to the fact that signal peaks that are far above the overall rms level of a signal do not permit the dynamic range to be excited evenly since energy in the signal is not equally spread. In order to achieve a proper balance of the crest factor, a tradeoff between the number, spacing, and starting phases of the multitone components must be configured properly based on the application.

An example of a multitone stimulus is shown in figure 2.15 along with the distorted output resulting from the nonlinear system described by $y = \pm|x|^2$.

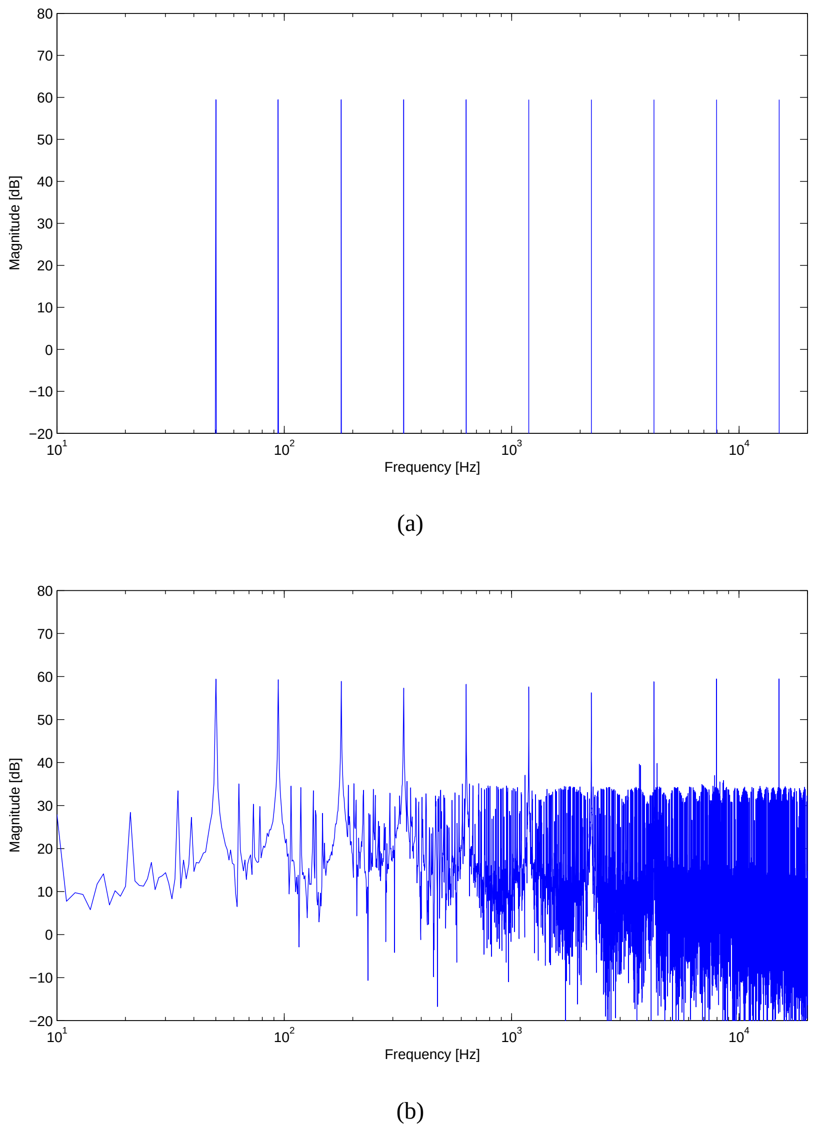



***Figure 2.15:*** *(a)10-component multitone test stimulus used to calculate DS metric. Components are spaced approximately 1.88x apart. (b) Distorted Multitone test stimulus [Distortion system: $y = \pm|x|^2$]*

## 2.6 Alternate Distortion Metrics

More recent literature on the subject of distortion perception shows an attempt at finding an improved distortion metric that correlates the distortion products with the overall subjective perception of distortion. This attempt has focused on exploiting psychoacoustic principles of human sound perception and applying them to derive metrics that more accurately reflect the effects of distortion and the perception of the distortion. Such metrics allow for more meaningful measurements that not only quantify the amount of nonlinear distortion, but also relate these quantities to a certain subjective perception. Three important metrics are the Gedlee metric proposed by Geddes & Lee [10, 9], the DS metric proposed by Moore et. al [25], and the $R_{nonlin}$ metric proposed by Moore et. al [26].

### GedLee Metric

The GedLee ($G_m$) metric was derived primarily based on the psychoacoustic properties of masking, as described in Section 2.3. The authors proposed a metric to take into account two principle effects of masking. The first being that higher order harmonics are more perceptible due to the tendency of lower order harmonics to be masked. The second, that nonlinear distortion products will be more audible at low signal levels since the masking threshold at low signal levels is lower than at higher levels. (refer to figure 2.12 on page 15, masking threshold) Additionally, the proposed metric is to be immune to offset and gain characteristics of the output signal, since these effects are either linear contributions to the input signal or imperceptible.

The GedLee metric is defined as,

$$G_m = \sqrt{\int_{-1}^{1} \left(\cos\left(\frac{x\pi}{2}\right)\right)^2 \left(\frac{d^2}{dx^2}T(x)\right)^2 dx} \tag{2.15}$$

where $x$ is the input signal, and $T(x)$ is the nonlinear transfer function of the system in question. By taking the second derivative of the transfer function, the metric gives more weight to higher order distortion products. Additionally, by taking the second derivative of the transfer function, the metric eliminates any gain or offset biases since the second derivative will eliminate all components up to the second harmonic components. The cosine term is applied to weight the level of the signal as described above. For small input levels the cosine term will approach unity, whereas for larger input signal levels the cosine term will approach zero. Therefore the cosine term provides more weight to small input levels where the masking threshold is small. Finally, the equation is integrated over the range -1 to +1 (the range of the output signal), to produce a single valued metric.

To illustrate the use of the metric, a simple third-order transfer $T(x) = x + x^3$ is applied. The Gedlee metric, $G_m$, reduces to,

$$G_m = \sqrt{\int_{-1}^{1} \left(\cos\left(\frac{x\pi}{2}\right)\right)^2 (6x)^2\, dx} \tag{2.16}$$

Of particular note is that equation 2.16 is only in terms of the second derivative of the third order component. Solving equation 2.16 results in $G_m = 1.5$.

In listening tests, the authors found a moderate correlation between this metric and the subjective impressions of artificially applied distortion on the magnitude of 0.67 [9]. The GedLee metric is not applicable to nonlinear systems which are frequency dependent since the metric assumes that the transfer function,

$T(x)$, is valid for all frequencies. The nonlinear distortion in real transducers often varies with frequency. As such, it was recommended in [26] that the GedLee metric be extended into frequency bands and applying the metric separately in each band.

### Distortion Score (DS) Metric

The Distortion Score (DS) is a more elaborate metric that attempts to take into account the psychoacoustic process of auditory filtering in deriving the metric. As described in section 2.3, psychoacoustic models for the human auditory system assume that energies within a critical band interact to produce the overall perception of the sound within that band, such as the masking of a tone by narrowband noise [5]. The DS metric attempts to model the auditory system using the Equivalent Rectangular Bandwidths ($\mathrm{ERB}_N$) developed by Moore [13]. The Difference Score relates to the difference between the overall spectrum produced by the input signal as compared to the overall spectrum produced by the output signal including its distortion products.

The steps involved in determining the value of DS of a given signal are crudely given as follows: First, an input signal is passed through a nonlinear system giving rise to an output signal. The input and output signals are time aligned so as to remove any time delays introduced by the nonlinear system. Next, the input and output signals are analyzed in a series of 30ms frames. The 30 ms time frame is typically chosen to ensure that the Discrete Fourier Transform (DFT) is performed over a stationary signal. The DFT is performed over each frame, and the relative peaks of the output signal are normalized to the input signal to remove any offset or gain bias. This ensures that any linear distortion is removed from the output signal. The signal spectra are grouped into appropriate $\mathrm{ERB}_N$ frequency bands so as to have a representation of the signal's "frequency analysis" in the auditory system. The overall power of the input and output signals in each band is then calculated and converted to decibels. Finally, the absolute value of the difference in each band is then summed across all bands resulting in the DS value.

According to the authors, this gives a *'perceptually relevant measure of the difference spectrum between the input and the output'* [26]. Using subjective ratings of distortion obtained for a variety of test stimuli and nonlinear distortion systems, the authors found the metric to be highly correlated with the subjective perception of distortion perception with correlation values up to 0.97 for music and speech signals. However, more moderate values of correlation, 0.60-0.67, were found when distortion produced by real transducers was used. The lower correlation of the DS metric with subject ratings of real transducers was pointed out in [26] to be a result of the crude modeling of the frequency analysis in the peripheral auditory system used in the DS metric's algorithm.

### $R_{\text{nonlin}}$ Metric

The $R_{\text{nonlin}}$ metric was developed as an extension of the DS metric developed by Moore et. al. in [25] as described above. As the performance of the DS metric using real distortion produced by real transducers was only moderate, the $R_{\text{nonlin}}$ metric was proposed using a different approach to analyzing the difference between the input test signal and its distorted output. Instead of calculating a difference score based on the input and output spectrum, a coherence analysis was performed by taking the cross-correlation between the input and distorted output waveforms. Additionally, the metric algorithm uses a more comprehensive model of the frequency analysis performed in the peripheral auditory system including the filtering produced by the outer and middle ear. Figure 3.2 shows a block diagram of the steps involved in deriving the $R_{nonlin}$ metric.

Similar to the DS metric, an input signal is passed through a nonlinear system resulting in a distorted output signal. Additionally, the input and output waveforms are time aligned to remove any unwanted delays caused by the nonlinear system. Next, both waveforms are filtered to mimic the response of the outer and middle ear by 4097 FIR filters, as described by Glasberg and Moore [7]. Next, both waveforms are filtered by an array of 40 gammatone filters with a bandwidth of 1-$ERB_N$. This filtering provides a more elaborate modeling of the auditory filtering mechanism as described by Patterson et. al. [17].

Next, the input and output signals are split into 30ms frames for further processing. The maximum value of the normalized cross-correlation between the input and output signals, Xmax, is calculated. For each frame, the Xmax values are summed across all filters. Finally, the Xmax values are averaged over all the frames resulting in the single valued metric, $R_{nonlin}$ .

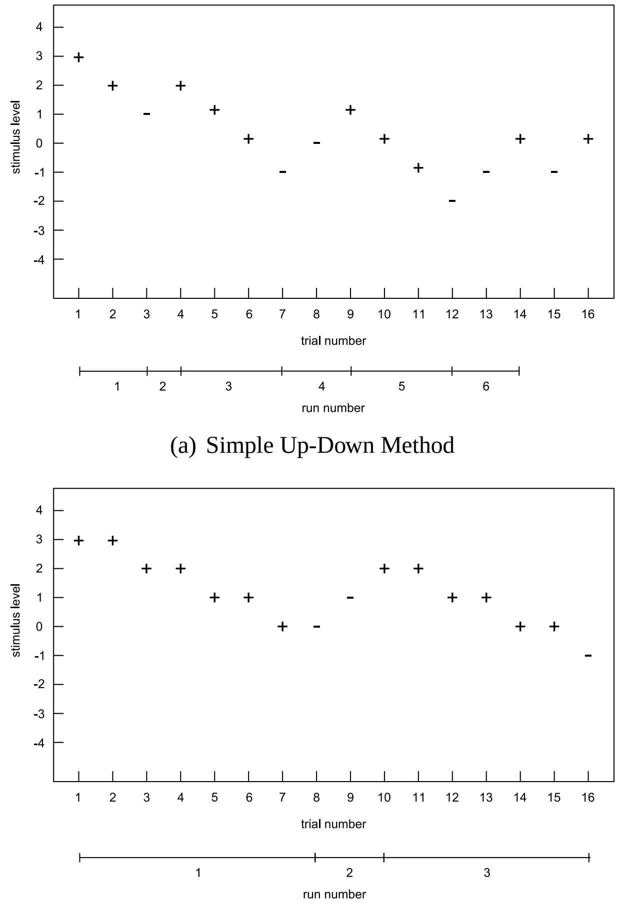
## 2.7 Methods of Threshold Estimation

The principle aim of this project relates to the application of the new nonlinear distortion metrics, DS and $R_{nonlin}$ , to nonlinear distortion thresholds. Obtaining a threshold implies increasing or decreasing some independent variable to find the point at which the subjective response to an auditory event changes. Hearing thresholds, for instance, vary the sound pressure level of a pure tone signal to the point, or threshold, where the subject can no longer hear the signal.

There are various methods available to experimenters to arrive at a given threshold. One such method, the method of constants, presents pre-determined samples with a varying parameter to a subject in random presentation order and the subject is asked to determine which samples are audible and which are not with respect to the parameter in question. Once the test is finished, the distribution of positive and negative responses is calculated and the threshold may be determined. While this method provides a way to ensure an unbiased result, it requires a larger number samples and may take an excessive amount of time if only the threshold is of interest [11].

Another method labeled the method of limits, presents a sample with a high probability of a positive response to a subject. Based on the response of the subject, the subsequent sample presentation is either increased or decreased in level accordingly until the threshold is reached, indicated by a negative response, or reversal. At this point the test is over and the threshold is determined as the reversal point. This may be done in either an ascending or descending fashion, where the threshold is approached from either below or above the threshold, respectively. While simple in its implementation, it provides no safeguard against false-positives and may yield misleading results.

The simple up-down method is similar to the method of limits, however the test is not finished after the first reversal. The simple up-down procedure, within the context of distortion, would decrease the amount of applied distortion after a positive response or decrease the amount of distortion after a negative response. The amount by which the distortion is increased or decreased is the step size. The accuracy of the experiment depends heavily on the selected step size as a large step size would yield an imprecise determination of the threshold and a small step size would require many presentations to arrive at the threshold. The simple up-down method typically uses the same step size throughout the experiment, however, variations of the method utilize a variable step size. A common practice is to change the step size after the 1st, 3rd or 7th reversal. The test then terminates after at least six to eight reversals [11]. The positive response, assuming that the subject is asked to indicate if the distortion is audible, refers to the subject responding that he or she can hear the distortion.




(a) Simple Up-Down Method

(b) Transformed Up-Down Method

***Figure 2.16:*** *Example data sets in the Up-Down methods.*

Figure 2.17 plots the distribution of the percentage of correct responses for a given stimulus which is also known as the psychometric function. The simple up-down procedure converges to the limit which corresponds to a 50 % probability of a correct response.

The transformed up-down method operates on the same basis as at the simple up-down procedure. However, a DOWN (decreased in the amount of applied distortion) would require two positive responses from the subject and an UP (increase in distortion) would requires a negative response from the subject.

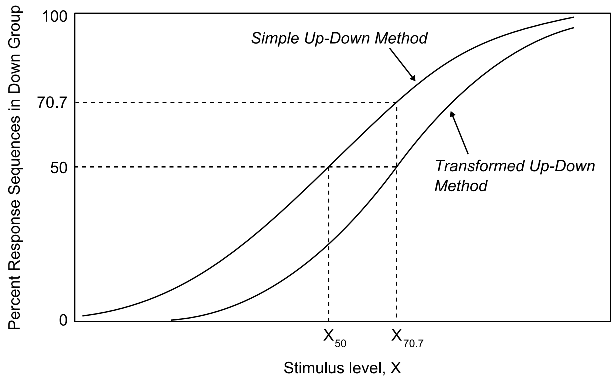



***Figure 2.17:*** *Psychometric functions for the simple Up-Down Method and the Transformed Up-Down Method [11].*

The transformed up-down sequence converges to the point where the probabilty of an UP or DOWN sequence are equal, or 0.5. The probabilty of a correct response at convergence is then, $P(x) = 0.707$ since $[P(x)]^2 = 0.5$.

## 2.8 Summary

A definition of nonlinear distortion was presented from a mathematical perspective and also within the scope of sound reproduction systems. The asymmetrical and symmetrical clipping examples can be considered nonlinear distortion types that often result from amplifiers. The added frequency components injected from these nonlinear distortion systems were also discussed. It was concluded that asymmetrical clipping injects both even and odd order harmonic distortion products and symmetrical clipping only odd order.

Psychoacoustic concepts were discussed as a means of understanding the perception of nonlinear distortion components. Nonlinear distortion products residing in the lower frequency range would not be as apparent as distortion products in the higher frequency range. This results from the human ear being less sensitive to low frequencies tones. Auditory masking also plays an important role in the perception of distortion products where higher distortion products could be masked to a lesser extent than lower distortion products.

Conventional nonlinear distortion metrics were also detailed, namely the THD and IMD metrics. Measurement descriptions of these parameters were also presented. Problems related to the THD metric were highlighted showing that two signals although having the same amount of THD *may* not be perceived as being equally distorted. The concept of the multitone test stimulus was described. The development of the multitone stimulus was motivated by the inability of the signal tone or two tone test signal used in conventional metrics to properly describe the characteristics of nonlinear distortion.

Several alternate distortion metrics were described. These are the Gedlee metric, the DS metric and the $R_{nonlin}$ metric. These metrics include psychoacoustic principles in their derivation with the intent of being able to predict the subject perception of distortion with a single value. The Gedlee metric was found to have only moderate correlation to subject ratings of distortion. The DS and $R_{nonlin}$ metrics were found to have quite high correlation with subjective ratings of distortion. For the purposes of the nonlinear distortion thresholds, it is most necessary to have a highly correlated metric. As such, the Gedlee parameter is not considered for future implementation and analysis.

Experimental methods for obtaining a threshold within the scope of human experimental psychology were detailed. These methods included the method of constants, the method of limits and the simple up-down method. Finally, the transformed up-down method was arrived at which improves the efficiency and threshold estimation of the simple up-down procedure.

CHAPTER 3

# IMPLEMENTATION OF METRICS

*This chapter provides a detailed description of the DS and $R_{nonlin}$ metrics. The algorithms for both metrics were implemented in MATLAB. The methods for calculating the the THD+N metric and the IMD metric according to the CCIF standard are also presented. These metrics will be used in Chapter 4 to verify the correlation of these metrics to subjective ratings of distortion.*

## 3.1 Implementation of the DS Metric

This section presents the implementation of the DS metric developed by Moore et. al. [25] The program was developed in MATLAB and a block diagram of the metric is shown in figure 3.1.

### Description of the DS Model

The underlying idea in deriving the DS metric is to find the difference between the input and output spectrum of a signal after undergoing nonlinear distortion. Additionally, the metric aims at taking into account the peripheral auditory filtering process in its derivation.

The steps involved in determining the value of DS of a given signal are as follows: First, an multi-tone input signal is passed through a nonlinear system giving rise to an output signal. The input and output signals are time aligned so as to remove any time delays introduced by the system. Next, the input and output signals are analyzed in a series of 30ms frames. A 1323 point Discrete Fourier Transform (DFT) is performed over each frame, $i$, and the relative peaks of the output signal are scaled to the input signal to remove any offset or gain bias. The is accomplished by finding the maximum value of the 1323 frequency bins from both the output and input signals. The peak value of the output signal is then scaled down to the peak value of the input signal. The signal spectra are then grouped into 40 non-overlapping $\mathrm{ERB}_N$ frequency bands covering the center frequencies from 50-19739 Hz. This provides a perceptually relevant representation of the signal processed by the auditory system. The overall power of the input and output signals in each band is calculated and converted to decibels. Finally, the absolute value of the difference in each band is then summed across all bands, resulting in the DS metric.

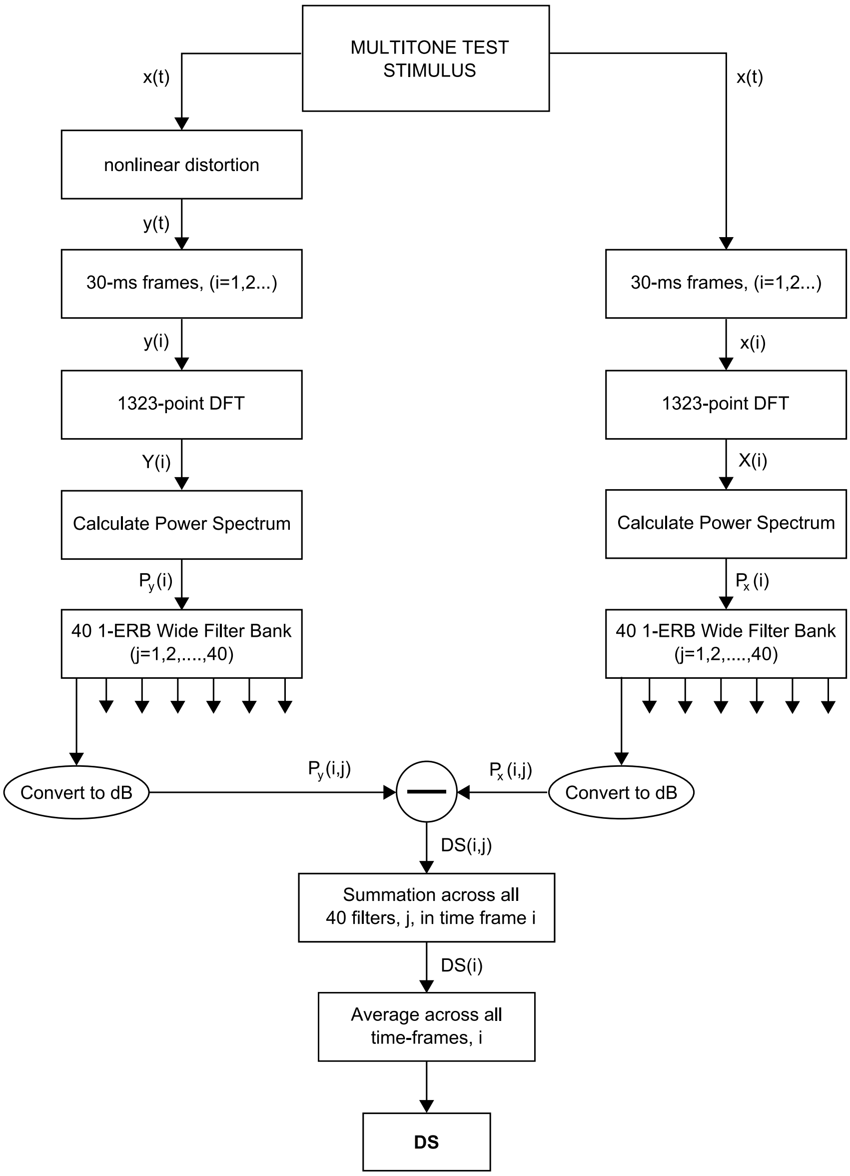



***Figure 3.1:*** *Block Diagram for the calculation of the DS Metric.*

## Multitone Test Stimulus

As mentioned in Section 2.5, determining the number of tones needed for the multitone signal greatly depends upon the application. In [25], Moore et. al. used subjective ratings of distortion to find the best correlation between the number tones and relative phases in the multitone signal to the ratings. They found that a 10-component multitone stimulus with a spacing of approximately $1.88f$ resulted in the greatest correlation between DS and subjective rating.

## Deriving the ERB Filter Bank

Using the notched noise method described in section 2.3, Moore & Glasberg [6], determined the Equivalent Rectangular Bandwidth (ERB) of the auditory filters. The mean values of the ERB's measured using moderate sound levels for young people with normal hearing, denoted ERB$_N$, is given by,

$$ERB_N = 0.108 f_c + 24.7 \tag{3.1}$$

where, $f_c$ is the ERB center frequency in Hz. Using equation 3.1, a filter can be constructed having a specific bandwidth (ERB$_N$) centered around a certain frequency ($f_c$). A filter bank spanning the audible frequency range can then be created providing a rough model of the auditory filtering process.

In determining the DS metric, a filter bank of 40 rectangular non-overlapping frequency bands, each 1-ERB wide, is used covering the range from 50 to 19739 Hz. [26].In order to calculate the center frequencies, $f_c$, and the ERB$_N$ values it is necessary to divide the frequency range of interest into the appropriate number of bands by manipulating equation 3.1.

Equation 3.1 can be rewritten as,

$$ERB_N = \frac{f_c}{Q} + BW_{min},\ where\ Q = 9.26, BW_{min} = 24.7 \tag{3.2}$$

where $f_c$ is the center frequency in Hz, $Q$ is the quality factor of the filter, and $BW_{min}$ is the minimum bandwidth of an auditory filter in Hz. The quality factor is measure that represents the sharpness of the filter by the relation $Q = \frac{f_c}{\Delta f}$. The quality factor is a constant for all ERB auditory filters and is related to the bandwidth ERB$_N$. To solve for $f_c$ we can rewrite equation 3.2 as,

$$f_c = ERB_N Q - BW_{min} Q \tag{3.3}$$

Additionally, if we are to have N equally spaced, 1-ERB wide filters, over the frequency range $f_{c_N}$ to $f_{c_1}$, than we can write each ERB$_N$ as,

$$\log_{10}(ERB_N(n)Q) + Kn \quad = \quad \log_{10}(ERB_N(f_{c_1})Q) \tag{3.4}$$

where $K$ is some integer constant describing the seperation between ERB$_N$ center frequencies. For consistency, we can rewrite as $K = \frac{k}{Q}$. It should be noted that each ERB$_N(n)$ decreases in frequency with increasing n (i.e. ERB(1) is the highest frequency band and ERB(N) is the lowest frequency band).

Rewriting equation 3.4 using equation 3.3 we have,

$$log_{10}(f_c(n) + BW_{min}Q) + \frac{kn}{Q} = log_{10}(f_{c_1} + BW_{min}Q),\ for\ n = 1, 2, ..., N \tag{3.5}$$

Therefore, to create N equally spaced, 1-ERB wide filters between $f_{c_N}$ and $f_{c_1}$, we can solve for the spacing k,

$$k = \frac{Q}{N}[log_{10}(f_{c_N} + BW_{min}Q) - log_{10}(f_{c_1} + BW_{min}Q)] \tag{3.6}$$

Further, we can solve for $f_c$ in equation 3.5 as,

$$f_c(n) = \frac{f_{c_1} + BW_{min}Q}{10^{\frac{kn}{Q}}} - BW_{min}Q, \quad for\ n = 1, 2, ..., N \tag{3.7}$$

By inserting $k$ from equation 3.6 into equation 3.7, we can solve for $f_c(n)$ as,

$$f_c(n) = \frac{f_{c_1} + BW_{min}Q}{10^{\left(\frac{\frac{Q}{N}[log_{10}(f_{c_1} + BW_{min}Q) - log_{10}(f_{c_N} + BW_{min}Q)]}{Q}\right)n}} - BW_{min}Q, \quad for\ n = 1, 2, ..., N \tag{3.8}$$

Using these relationships, the 40 non-overlapping ERB$_N$ frequency bands were calculated as shown in table 3.1.

| **Band** | $f_c$ | **ERB$_N$** | **Band** | $f_c$ | **ERB$_N$** |
|---|---|---|---|---|---|
| 1 | 50 | 30 | 21 | 2264 | 269 |
| 2 | 82 | 34 | 22 | 2552 | 300 |
| 3 | 118 | 37 | 23 | 2874 | 335 |
| 4 | 158 | 42 | 24 | 3233 | 374 |
| 5 | 203 | 47 | 25 | 3634 | 417 |
| 6 | 253 | 52 | 26 | 4081 | 465 |
| 7 | 309 | 58 | 27 | 4580 | 519 |
| 8 | 371 | 65 | 28 | 5136 | 579 |
| 9 | 441 | 72 | 29 | 5757 | 646 |
| 10 | 518 | 81 | 30 | 6450 | 721 |
| 11 | 605 | 90 | 31 | 7223 | 804 |
| 12 | 701 | 100 | 32 | 8085 | 897 |
| 13 | 809 | 112 | 33 | 9048 | 1001 |
| 14 | 929 | 125 | 34 | 10121 | 1117 |
| 15 | 1063 | 139 | 35 | 11319 | 1246 |
| 16 | 1213 | 156 | 36 | 12656 | 1391 |
| 17 | 1379 | 174 | 37 | 14147 | 1552 |
| 18 | 1566 | 194 | 38 | 15811 | 1731 |
| 19 | 1773 | 216 | 39 | 17668 | 1932 |
| 20 | 2005 | 241 | 40 | 19739 | 2155 |

***Table 3.1:*** *ERB$_N$ center frequencies & bandwidths.*

## 3.2 Implementation of The R$_{\text{nonlin}}$ Metric

The R$_{\text{nonlin}}$ metric was developed as an extension of the DS metric developed by Moore et. al. [25]. The R$_{\text{nonlin}}$ metric was proposed using a more comprehensive model of the frequency analysis performed in the peripheral auditory system. Instead of calculating a difference score based on the input and output spectrum, a coherence analysis was performed by taking the cross-correlation between the input and distorted output waveforms. The cross-correlation permits a valid measure of the dissimilarity of the undistorted and distorted test signal. Figure 3.2 shows a block diagram of the steps involved in deriving the $R_{nonlin}$ metric.

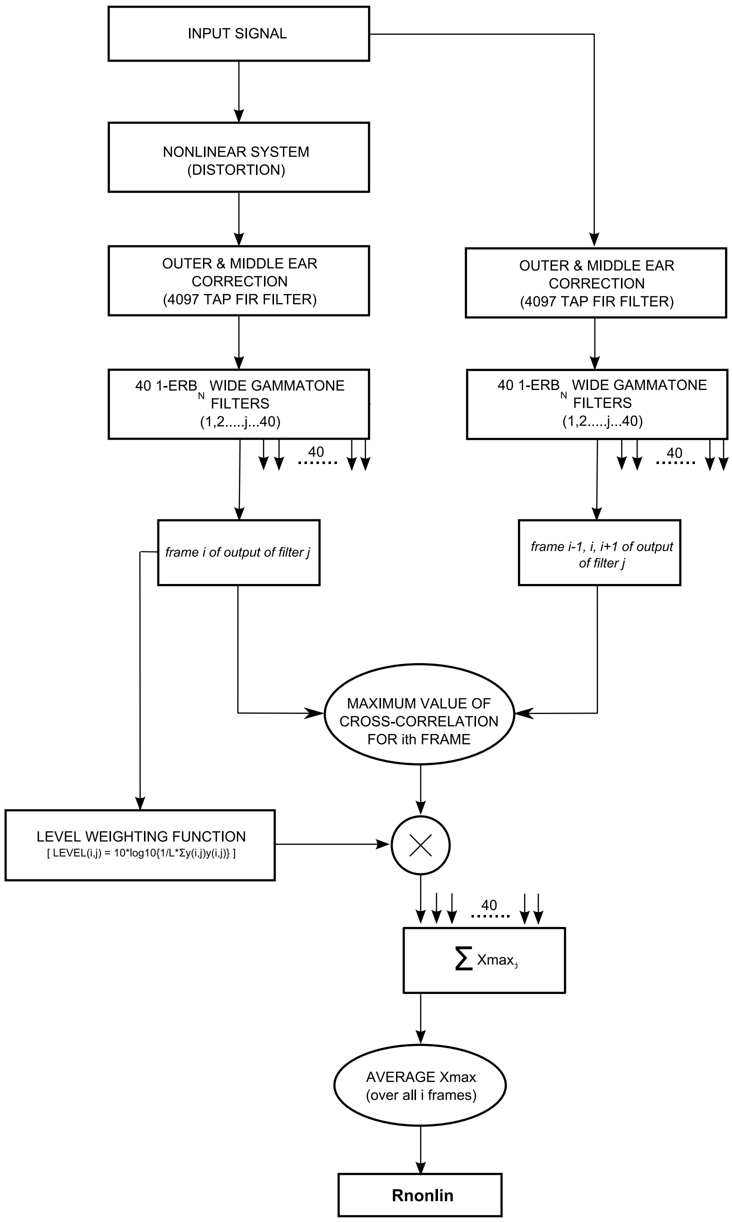



***Figure 3.2:*** *Block Diagram for the calculation of $R_{nonlin}$.*

As shown in the block diagram in figure 3.2, an input signal is passed through a nonlinear system resulting in a distorted output signal. The input and output waveforms are time aligned to remove any unwanted delays caused by the nonlinear system. Next, both waveforms are filtered to mimic the response provided by the outer and middle ear by a 4097 FIR filter, as described by Glasberg and Moore [7]. The relatively high order of this filter was selected to ensure sufficient attenuation in the lower frequency range. Although a filter of this exact order is not necessary, it is thought desirable to follow all the steps and procedures used in [26]. Next, both waveforms are filtered by an array of 40 gammatone filters with a bandwidth of 1-ERB$_N$.

Next, the input and output signals are split into 30ms frames [L=1323 samples] for further processing. The normalized cross-correlation at the $ith$ frame and $jth$ filter of the output signal is calculated with respect to the concatenation of (i-1), i, (i+1) frames of the input signal at the same filter, for lags of -10 to +10 ms [-441 to +441 samples].

$$r_{x,y}(i;j;\eta) = \frac{\sum_{n=(i-1)L+1+\eta}^{iL+\eta} x(n;j)y(n-\eta;j)}{\sqrt{\left(\sum_{n=(i-1)L+1+\eta}^{iL+\eta} x(n;j)x(n;j)\right)\left(\sum_{n=(i-1)L+1+\eta}^{iL+\eta} y(n-\eta;j)y(n-\eta;j)\right)}} \tag{3.9}$$

for $-441 \leq \eta \leq +441$ samples. Theoretically, if there is no time delay in the distorted signal, then the maximum cross correlation will occur when $\eta = 0$. In real transducers, there is often some time delay applied to the distorted signal and this lag parameter ensures that the input and output signals are properly compared. The maximum value, $X_{max} = max(r_{x,y}(i;j;\eta))$, is then found for each frame, $i$, and filter output, $j$.

An additional weighting function is applied to the values of $X_{max}$ calculated in each $ith$ frame. The weighting assumes that the perception of distortion at the output of a given filter is related to the relative magnitude of the output at that filter [26]. For example, for a filter with a relatively low output, the perceived amount of distortion will be small, and vice versa. Therefore, a weighting is applied to each Xmax value across all filters of a particular frame. For each frame i, the power at each output of each filter is calculated and converted to decibels as,

$$Level(i,j) = 10log_{10}\left(\frac{1}{L}\sum_{n=(i-1)L+1}^{iL} y(n;j)y(n;j)\right) \tag{3.10}$$

where L is the size of each 30-ms frame (L=1323 samples). The value of $Level(i,j)$ is used to determine the weight applied to the value of Xmax at a particular filter output. The weighting function was determined empirically by Moore et. al. [26] which provided the best fit to the data. First the maximum Level(i,j) is determined. Levels within 40dB of the maximum value are assigned the same weight. Levels greater than 80dB from the maximum value are assigned a weight of 0. other levels, between 40 to 80 dB of the maximum value are scaled linearly. In such a way, filters with high output are scaled equally, whereas filters with low output are scaled to 0. Filters outputs between the two are scaled proportionally (in decibel scale units). This provides an appropriate compressive scale that relates the output level of the auditory filters to psychoacoustic perception.

For each frame, the Xmax values are summed across all filters yielding a single value for each 30 ms time frame. The weighting function described above also ensures that the summation of all Xmax values across all filters is between 0 and 1. Finally, the Xmax values for each 30 ms time frame are averaged resulting in the single valued metric, $R_{nonlin}$ .

To summarize, the $\mathrm{R_{nonlin}}$ metric calculates a time-averaged cross-correlation coefficient across the 40 non-overlapping 1-$\mathrm{ERB}_N$ filters. In such a manner, a more representative metric is derived, taking into account the effects of the peripheral auditory system.

### Outer-Middle Ear (OME) Filter

The Outer-Middle Ear (OME) filter used in calculating the $\mathrm{R_{nonlin}}$ metric was derived as an approximation to the filter described by Glasberg & Moore [7] using the fir2 function provided by the MATLAB signal processing toolbox. The filter is shown in figure 3.3.

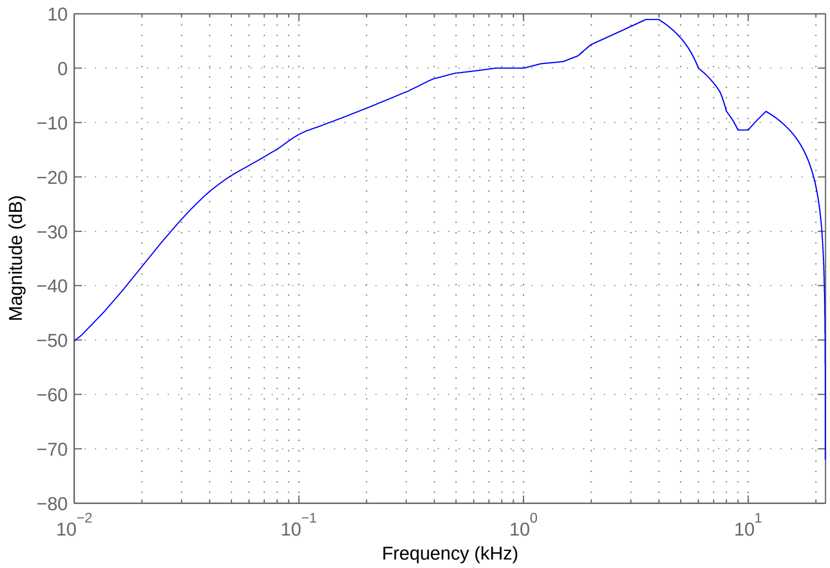



***Figure 3.3:*** *4097 Coefficient Outer-Middle Ear Filter*

The purpose of the OME filter is to mimic the frequency response of the outer and middle ear and to effectively attenuates frequency components below 500 Hz and above 5 kHz.

### Gammatone Filter Bank

Using the $\mathrm{ERB}_N$ model described in Section 2.3, a gammatone auditory filter bank can be created which fairly accurately models the filtering produced by the inner ear.

To implement the gammatone auditory filter bank in determining the $\mathrm{R_{nonlin}}$ metric, the MATLAB function MakeERBFilters, provided by the Auditory Toolbox [23] created by Malcom Slaney, was used along with the center frequencies derived in Section 3.1.

### Predicting Subjective Rating with the $\mathrm{R_{nonlin}}$ Metric

In order to predict the subjective ratings given to a stimulus based on its $\mathrm{R_{nonlin}}$ value, Moore et. al.[26] empirically derived a formula relating the $\mathrm{R_{nonlin}}$ value to subjective rating. This was accomplished by

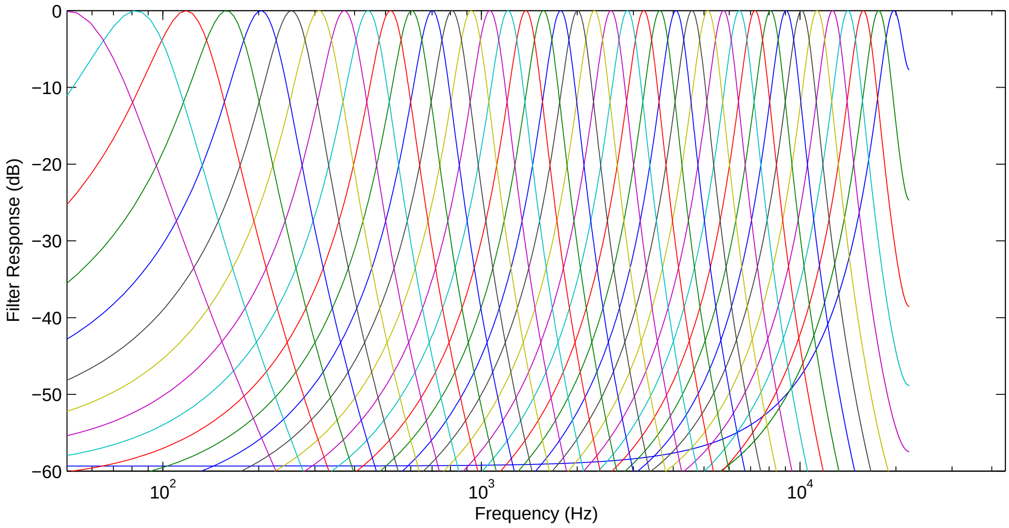



***Figure 3.4:*** *Gammatone auditory filter bank with 40 1-ERB$_N$ wide 4$^{th}$ order filters.*

fitting a curve to the data obtained in their subjective listening tests and is defined as follows,

$$Predicted\ Rating = \frac{a + b(R_{nonlin})^c}{a + (R_{nonlin})^c} \tag{3.11}$$

where a, b, and c are function parameters used to fine tune the curve fit in order to obtain a best fit to the data. An example of the curve fitting is shown in figure 3.5.

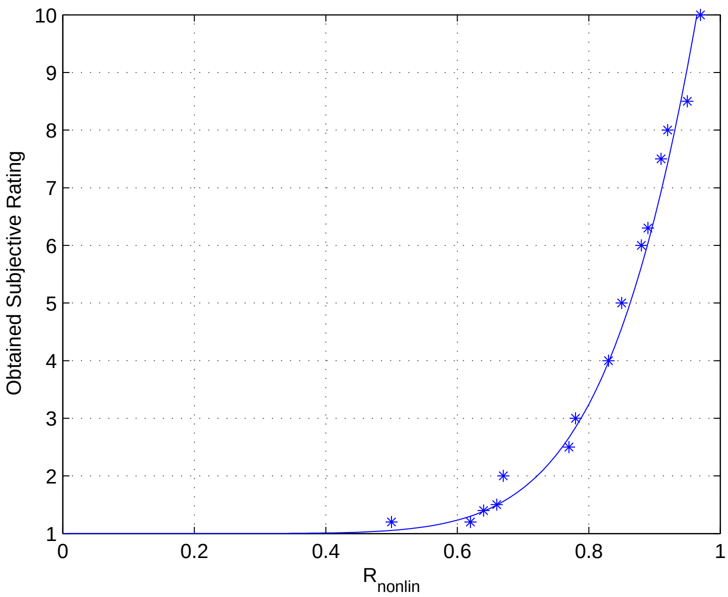



***Figure 3.5:*** *Example of curve fitting using equation 3.11. [* represents examples of obtained subjective rating. solid curve represents example of the fitted curve.]*

Moore et. al. found very high correlations between equation 3.11 and their obtained ratings with correlation coefficients as high as 0.98 [26]. Therefore, by using equation 3.11, a prediction of the subjective rating due to a particular type of distortion may be obtained with relatively high accuracy.

## 3.3 Implementation of The THD & IMD Metrics

The THD and IMD metrics will also be provided in this chapter. This serves to verify the poor correlation of these metrics in comparison with the new DS and $R_{nonlin}$ metrics.

The implementation of these metrics was not trivial. This stems from the fact that the THD and IMD metrics are dependent on the amplitude of the input test signal. Furthermore, most THD and IMD input test signals are selected to be at least 10 dB below clipping levels. This requirement does not suit the needs of this project as clipping has been artificially added to a music sample and it is of interest to obtain the THD or IMD values describing how much distortion these clippings introduce. As such, the amplitude for the input test signal to arrive at appropriate THD and IMD values for the nonlinear systems described in this chapter is the peak value of the undistorted music sample. It should be noted that comparison between the THD and IMD values presented in this report is not valid between THD and IMD values presented by other researchers unless the same input test signal conditions apply.

### Description of the THD+N Algorithm

The THD+N method was implemented in MATLAB as outlined in Section 2.4. The input test signal was fixed at 1 kHz with an amplitude corresponding to the peak value of the undistorted music sample. The sampling frequency of the input test signal was 44.1 kHz with a 1 second duration. The test signal was then passed through each of the nonlinear distortion systems. To compute the numerator of equation 3.12 a notch filter (FIR 1000 taps) centered around 1 kHz was used to remove the fundamental component of the test signal leaving only harmonic components. The first 2000 samples of the output signal from the notch filter were removed to reduce the effects of the filter on the overall rms output. Removing the first 2000 taps is not necessary as only the first 1000 samples would contain effects from the filter. However, the first 2000 samples were removed so as to be well beyond any effect from the filter.

$$\%THD = 100\frac{\sqrt{V_2^2 + V_3^2 + V_4^2 + ...V_n^2}}{V_T} \tag{3.12}$$

### Description of the IMD Algorithm

The IMD method was implemented in MATLAB according to the CCIF method described in Section 2.4. The sampling frequency of the input test signal was 44.1 kHz with a 1 second duration. The two tones of the input signal were set to 14 and 15 kHz. The selection of these frequencies reduces the injection of harmonic components and the higher intermodulation products. The rms sum between the distortion products was evaluated and expressed as a ratio against the rms value of the input signal. The rms sum of the distortion products was calculated by removing the two input frequencies using two cascaded notch filters (FIR 300). The first 1000 samples of the output signal from the two cascaded notch filters were removed to reduce the effect of the FIR filter on the overall rms output.

## 3.4 Summary

The DS metric was presented at the beginning of the chapter. The DS metric is essentially the average of the power level differences between the output spectrum from a nonlinear device and the original undistorted input test signal across 40 non-overlapping ERBs. The input test signal is a 10 component multitone signal. The method of arriving at the appropriate center frequencies for the 40 non-overlapping ERBs required for the DS computation was also detailed.

The implementation of the $R_{nonlin}$ metric was further discussed. The $R_{nonlin}$ metric is based of the cross-correlation between the original undistorted input test signal and the resulting distorted output from a nonlinear device. The metric algorithm models the auditory system by taking into account the filtering produced by the outer and middle ear and the filtering of the auditory system. A weighting factor is also used in the metric which is based on the assumption that the perception of distortion from the output of a given auditory filter is related to the magnitude of the output from that filter. The methods of predicting subjective ratings of nonlinear distortion using the $R_{nonlin}$ metric were also discussed.

The THD+N and IMD algorithms were presented at the end of the chapter and implemented in MATLAB. The input signal used to arrive at both metrics was set to the peak value of an undistorted music sample which is used in the following chapter. The following chapter relates the presented metrics to subjective ratings of nonlinear distortion.

CHAPTER 4

# EXPERIMENT 1: VERIFICATION OF METRICS

*This chapter provides an overview of the listening experiment conducted to verify the correlation of the DS and $R_{nonlin}$ metrics with subjective perception of nonlinear distortion. The design of the listening test is presented along with a description of the test stimuli and the steps involved in their creation. An analysis of the results is then presented along with the verification of the metrics using the obtained data.*

## 4.1 Listening Evaluation to Determine Subjective Ratings of Nonlinear Distortion Systems

The research of [25] and [26] collected subjective ratings of distortion for a wide variety of nonlinear systems using both artificially applied distortion and distortion produced by real transducers. These subjective ratings were used to arrive at the proper algorithm settings for for the DS and $R_{\text{nonlin}}$ metrics. The metric algorithms were implemented as described in the previous sections. However, it was found difficult to verify that they were implemented correctly without checking their correlation with some subjective ratings. Therefore, a small listening experiment was designed to "check" if these metric algorithms were in fact implemented correctly. By correctly implemented, it is meant that these metrics are highly correlated with the subjective ratings. For the purposes of the the distortion threshold evaluation, it is imperative that these metrics are well correlated with subjective ratings.

The purpose of the listening experiment was to be a "check" of the implemented metric algorithms and therefore it was designed to be short for an individual subject. The test was designed to be approximately 6 minutes in duration. Within the short session the subjective would rate the level of distortion on a scale from 1 - 10 where 10 would refer to an undistorted signal and 1 a completely distorted signal. Before allowing the subject to rate the distorted stimuli, the subjects were presented with signals corresponding to a 1 and 10. This rating procedure was same as used in [25].

A total of four nonlinear systems were selected for the purposes of this metric verification listening evaluation. These systems are:

1. Hard asymmetrical clipping.

| Vf Clipping Factor (Vp/Vf) | Asymmetrical Clipping | Symmetrical Clipping |
|---|---|---|
| | DS | DS |
| 1.48 | 13.244 | 25.059 |
| 1.615 | 32.103 | 50.816 |
| 1.74 | 51.846 | 75.109 |
| 1.873 | 72.408 | 100.07 |
| 2.01 | 93.016 | 125.03 |
| 2.155 | 113.21 | 150.09 |
| 2.315 | 133.36 | 175.37 |
| 2.49 | 153.42 | 200.08 |
| 2.92 | 195.18 | 250.69 |
| 3.5 | 237.82 | 300.64 |
| 4.37 | 282.78 | 350.68 |
| 5.91 | 325.81 | 400.2 |
| 10 | 351.6 | 450.57 |

***Table 4.1:*** *Table showing clipping levels and DS values for asymmetrical and symmetrical clippings.*

2. Hard symmetrical clipping.
3. Squared distortion described by $y = \alpha x^2 + x$.
4. Cubic distortion described by $y = \beta x^3 + x$.

A sample of guitar music of 4.4 second duration was selected as the input signal to these distortion systems. This wave file segment was taken from the CD "Sound Quality Assessment Material" (SQAM) produced by the European Broadcasting Union. The music sample has a relatively constant overall signal level without major peaks which could dominate the overall perceived distortion.

The clipping levels for the hard symmetrical clipping were found by dividing the peak value of the input signal, $V_p$, by a factor $V_f$. The clipping levels for the hard symmetrical clipping were set to yield DS values of 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400 and 450. These same clipping levels were used for the asymmetrical clipping level which yielded somewhat different DS values. The clipping levels and corresponding DS values are shown below in table 4.1 for both the asymmetrical and symmetrical distortion systems. During pilot experiments performed on group members it was found that distortion levels between 25-200 DS were often hard to perceive. For this reason, more points between DS values from 25-200 were selected.

For both the squared and cubic distortion systems, the coefficients $\alpha$ and $\beta$ were adjusted to yield DS values of 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400. The coefficients $\alpha$ and $\beta$ are shown in table 4.2 with their the corresponding DS values. While applying the polynomial distortion systems, special care was taken to ensure that peak levels in the signal did not exceed 1 or -1. The function $wavwrite$ provided by MATLAB clips all peaks exceeding 1 or -1.

Figure 4.1 shows the input-output relationships for the above nonlinear systems. For graphical purposes, the input signal used to make these graphs was not the input signal from SQAM. Rather, a signal linearly spaced between -.2 and .2 was used. The 0.2 was chosen as this was the peak value of the guitar signal described above. The asymmetrical clipping is plotted for DS levels of 13, 153, and 351. The symmetrical clipping is plotted for DS levels of 25, 200 and 450. The squared distortion type is plotted for DS

| | Squared Distortion | Cubic Distortion |
|---|---|---|
| **DS** | **Alpha** | **Beta** |
| 25.059 | 0.245 | 2.7 |
| 50.816 | 0.48 | 5.2 |
| 75.109 | 0.75 | 8.3 |
| 100.07 | 1.04 | 11.8 |
| 125.03 | 1.36 | 16 |
| 150.09 | 1.72 | 20.8 |
| 175.37 | 2.1 | 26.6 |
| 200.08 | 2.55 | 33.3 |
| 250.69 | 3.56 | 50.6 |
| 300.64 | 4.98 | 75.6 |
| 350.68 | 6.9 | 114.6 |
| 400.2 | 13.4 | 181.5 |

***Table 4.2:*** *Table showing coefficient values and DS values for squared and cubic distortions.*

values of 25, 200 and 400. At lower DS values, the squared factor is less significant and the resulting input output relationship shows a kind of soft asymmetrical clipping for values below 0. As the coefficient $\alpha$ increases, the squared polynomial dominates the waveform as demonstrated by the parabolic function. The cubic function is somewhat similar. As the coefficient $\beta$ increases the cubic function takes over.

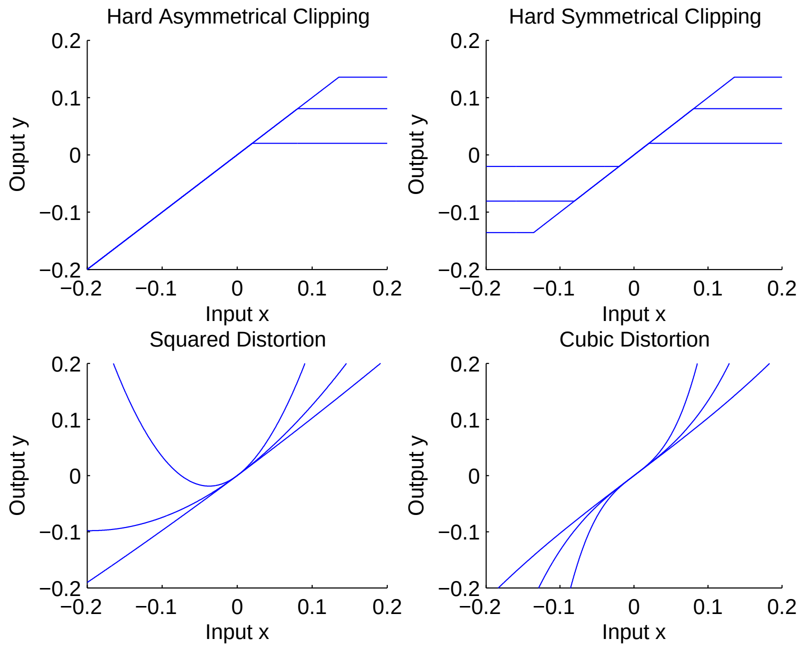



***Figure 4.1:*** *Input-Output relationships for nonlinear systems using input signal described above.*

The input signal was distorted by each of the above systems and the output was stored as a stereo wave file where each channel contained the same signal. A total of 50 stimuli were created for the purposes of the subjective listening evaluation.

When evaluating a psychophysical percept such as the perceived amount of nonlinear distortion in a stimulus, it is important to isolate the percept. Therefore, in order to eliminate the effects of linear distortion introduced by each of distortion types, a loudness normalization was applied to each of the stimuli. In such a way, any linear gain or attenutation introduced will not effect the subjects responses. This is particularly important when evaluating multiple stimuli on the same subjective scale (i.e. 1 to 10). The level of one stimulus in comparison with the next could influence the subject's response in an undesirable way. By normalizing the loudness of all stimuli, the only difference between the original signal and the distorted signal is the due to the applied nonlinear distortion. For example, clipping the signal by a any factor will reduce the overall level of the signal. Therefore, some gain should be applied to the clipped signal to ensure the distorted signal has the same overall loudness as the original signal, as well as the same overall loudness as all other distorted signals.

The loudness normalization was performed using a MATLAB program that implements the DIN 45631 / ISO532B standard based on Zwicker's loudness model (see Appendix B). Each of the 50 stimuli were normalized to have an overall loudness of approximately 15 sones (plus or minus .5 sone). As the loudness model uses a third-octave band analysis to approximate the ear's frequency selectivity, it is important to take into account the effect that the headphone's frequency response will have on the resulting loudness. As the headphone's frequency response is not flat (Appendix A.2), it will alter certain frequencies more than others. Therefore, before calculating the loudness of the distorted stimuli, they were first filtered by an average of the headphone's left cup impulse response. The resulting loudness was then calculated. If the overall loudness caluculated was not 15 sones, an amplitude scaling factor was then applied to the original stimuli, and the loudness calculation process repeated, until the appropriate sone level was achieved. This scaling factor was then used to scale the original, unfiltered stimulus and the resulting waveform written to a stereo .wav file. The loudness model assumes that the input signal is in Pascals, however the signal is not in Pascals and the sone value specified here is not truly representative of the actual loudness. However, as all stimuli will pass through the same reproduction chain, and the model is used only to normalize the stimuli to the same loudness level, this consideration will not effect the normalization process. Figure 4.2 shows the block diagram of the processing involved in creating all distorted stimuli.

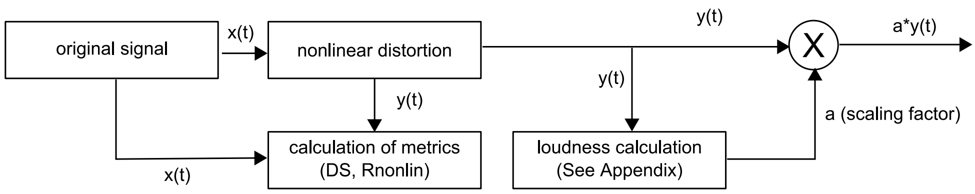



***Figure 4.2:*** *Block diagram showing the processing of the original, undistorted signal to the distorted version used in listening tests.*

A MATLAB graphical user interface (GUI) was created to interface with the subject. As mentioned earlier, the subject is presented with the undistorted and most distorted signal before proceeding with the subjective evaluation. The subject was able to listen to both scale extremes as many times as they wanted before beginning with the test. The most distorted signal (case 1) was selected as the signal having the highest DS value and the lowest the original undistorted signal (case 10). The highest DS value corresponded to the most extreme case of the symmetrical clipping. A screenshot of the GUI used for the evaluation is shown below in figure 4.3. The GUI program presents the 50 stimuli in random order and stores the subject's response and the file name of the presented stimulus for further statistical analysis.

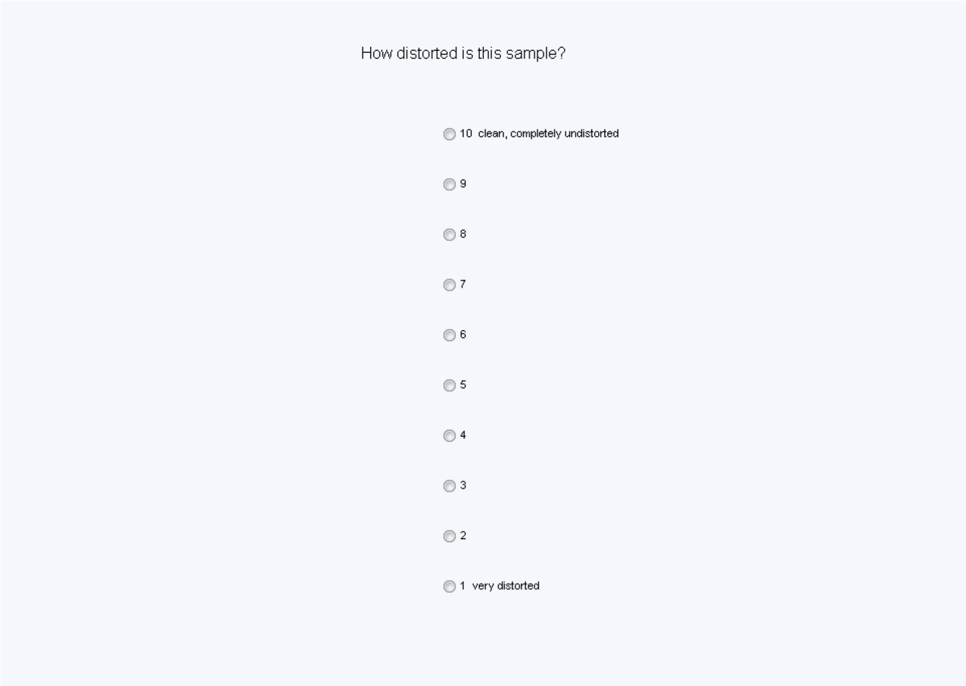



***Figure 4.3:*** *Graphical user interface for DS verification listening test.*

The signals were presented to the subject after clicking on the selected rating. Case 10, the undistorted signal, was never presented to the subject during the experiment. Each stimulus was presented only once within the session since the design of the experiment was to make it relatively short in duration.

The listening evaluation was performed in a small listening cabin in the Acoustic Laboratory at Aalborg University. The subjects ranged in ages from 21 to 26 years. The stimuli were presented to the subjects through a pair of Beyerdynamic DT990 headphones which were connected directly to a PC sound card. Nonlinear distortion measurements were performed on the sound reproduction chain prior to the listening evaluations (Appendix A.1). The overall signal level presented to the subjects was 70 dBA, or appoximtely 20 sones (DIN 45631/ISO532B). The loudness measurement description can be found in Appendix B on page 73 along with relavent equipment information and the instruction sheet given to the subjects before the evaluation.

A total of nine subjects were used for this experiment excluding group members. The subjects were all 10th semester Acoustic's students. It should be noted that one of two subjects felt that certain types of distortion were perceived as being more distorted than the extreme case (DS 450 from symmetrical clipping). This was later found to be of little consequence as most subjects rated this signal as having the highest distortion.

## 4.2 Analysis of Results

### Analysis of Variance

An analysis of variance (ANOVA) was performed on the subjective ratings given by all subjects with respect to the amount of distortion. The ANOVA analysis provides a means of evaluating the variance between groups, or in this case, the variance between the subjects' ratings for the different amounts of distortion. More specifically, it is assumed that the mean subjective rating for different levels of DS are different. The ANOVA analysis is used to either accept or reject the null hypothesis. The null hypothesis for the evaluation of the variance between the subjective ratings at different distortion levels is that their overall mean is the same. For example, if the mean subjective ratings at 50 DS are equal to the overall mean rating at 200 DS, then the null hypothesis would be true. To support the aforementioned assumption, the null hypothesis must be rejected.

The ANOVA analysis yields two parameters, F and p-value, which can be used to evaluate the significance. The F parameter is defined below.

$$F = \frac{(found\ variation\ of\ the\ group\ averages)}{(expected\ variation\ of\ the\ group\ averages)} \tag{4.1}$$

The null hypothesis would be correct is the F value is closer to 1. This F ratio is used to test for a statistical significance, or p-value. The p-value refers to the probability that a variate would have a value greater than or equal to the value observed by chance. For example, a p-value [$p < .01$] would indicate that the means differ by more than what would be expected by chance. In this case, the null hypothesis can be rejected. The null hypothesis, for the purposes of this project, is rejected for [$p < .05$] (95% confidence interval).

The analysis performed on the obtained data showed that the subjective ratings were highly significant [$F = 23.51$, $p < 0.0001$].

Additionally, for each type of distortion (i.e. asymmetrical, symmetrical, quadratic, cubic), a Two-Way ANOVA was performed on the subjective rating with amount of distortion and subject as factors. For each type of distortion it was found that the amount of distortion was significant [$p < 0.0001$ for all types of distortion]. For subjects as a factor there was only one case in which the subject was not significant [$p = .0125$]. This occurred for the asymmetrical clipping group. This indicates that the overall mean subject ratings for all distortion levels within this group where sometimes different between subjects.

The Two-Way analysis is an extended version of the one way analysis which is based on the following null hypotheses:

1. The means of the first factor are equal (level of distortion).
2. The means of the second factor are equal (subject).
3. There is no interaction between the two factors. This is not applicable as there are not repeated observations for both factors.

### Consistency Across Subjects

In order to assess the consistency across the subjects, the mean values of the subjective ratings were calculated for each stimulus and compared to the rating given by a particular subject for the given stimulus. This resulted in a correlation coefficient comparing the subjective ratings given by a particular subject against the overall mean values given by all subjects. Table 4.3 shows these correlation coefficients for all 9 subjects and shows a fairly high consistency across all subjects. The standard deviation, SD, of the subjective ratings across all subjects was 1.4 scale units which is relatively consistent with the SD obtained by Moore et. al. [25].

| **Subject** | ALE | ANT | CAR | CAS | DAV | IRE | LAR | LRM | YES |
|---|---|---|---|---|---|---|---|---|---|
| **Correlation Coefficient** | 0.85 | 0.90 | 0.85 | 0.91 | 0.92 | 0.80 | 0.77 | 0.92 | 0.88 |

***Table 4.3:*** *Correlation of subjective ratings for individual subjects with mean value of subjective ratings across all subjects.*

## 4.3 Verification of DS Metric

The mean values of the subjective ratings for each of the distortion types were computed and are shown in figure 4.4. Additionally, the overall mean values of the subjective ratings were computed for all stimuli and are shown in figure 4.5. The DS metric and the subjective ratings were found to be highly correlated with a negative correlation of $-0.9486$. This correlation is consistent with the correlation found by Moore et. al. [25] of $-0.95$ for artificial distortion applied to a music stimulus.

## 4.4 Verification of $\mathbf{R_{nonlin}}$ Metric

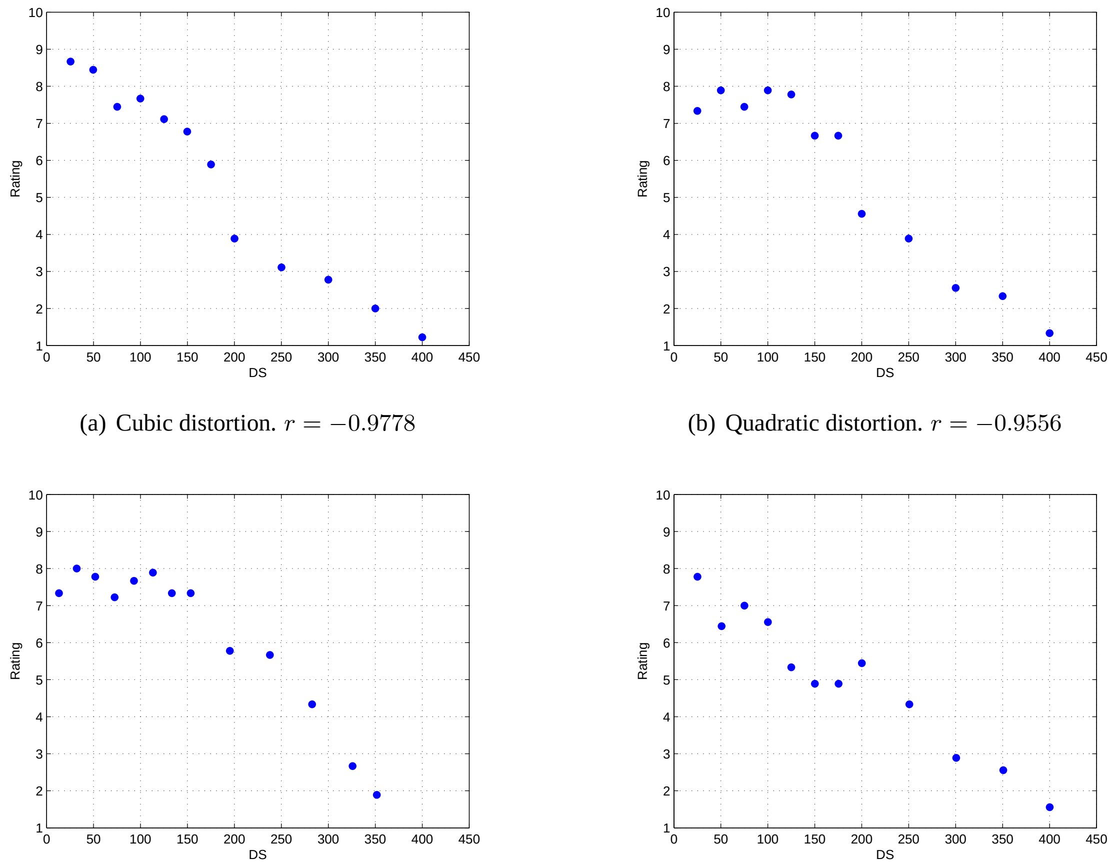
The mean values of the subjective ratings are plotted against the $\mathrm{R_{nonlin}}$ metric in figure 4.6.

As described by Moore et. al. [26], a curve fitting may be applied to the data using the empirically obtained formula,

$$Predicted\ Rating = \left(\frac{a + bR^{c}_{nonlin}}{a + R^{c}_{nonlin}}\right) \tag{4.2}$$

where a,b, and c are function parameters used to fine tune the curve to achieve the best fit, as described in section 3.2. The fitted curve is shown in figure 4.7. The fitted curve was determined to have a reasonable fit, with an adjusted-$R^2$ value of $0.9295$.

Using the curve fitting equation described by equation 4.2, the predicted values of the subjective ratings may be calculated. Figure 4.8 shows a plot of the observed versus the predicted values of the subjective ratings found using equation 4.2. It was determined that the observed versus predicted ratings were




(a) Cubic distortion. $r = -0.9778$

(b) Quadratic distortion. $r = -0.9556$

(c) Asymmetrical clipping. $r = -0.9282$

(d) Symmetrical clipping. $r = -0.9791$

***Figure 4.4:*** *Mean Rating versus DS.*

highly correlated with a correlation coefficient of $0.9547$.

The individual $\mathrm{R}_{\mathrm{nonlin}}$ fittings for each type of distortion along with plots of predicted ratings can be found in Appendix C.

## 4.5 Comparison with THD & IMD Metrics

In order to judge the performance of the DS and $\mathrm{R}_{\mathrm{nonlin}}$ metrics with respect to the traditional metrics, corresponding THD and IMD values were calculated from the data obtained from the listening test. Section 3.3 describes the process used to obtain the corresponding values of THD and IMD. Figures 4.9(a) and 4.9(b) show the resulting mean subjective ratings versus % THD and % IMD, respectively. As before, mean ratings can take the values from 10 (undistorted) to 1 (completely distorted), and increasing values of % THD and % IMD imply increasing levels of distortion.

Not surprisingly, both figures show that the metrics are not correlated with any perceptual rating of distortion. Of particular note, is that for both THD and IMD, similar values of a particular metric result in

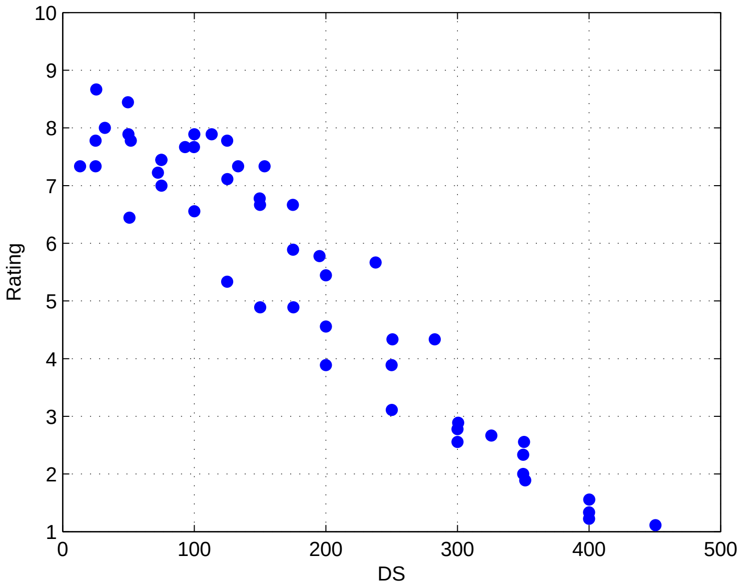



***Figure 4.5:*** *Overall Mean Rating versus DS.* $r = -0.9486$

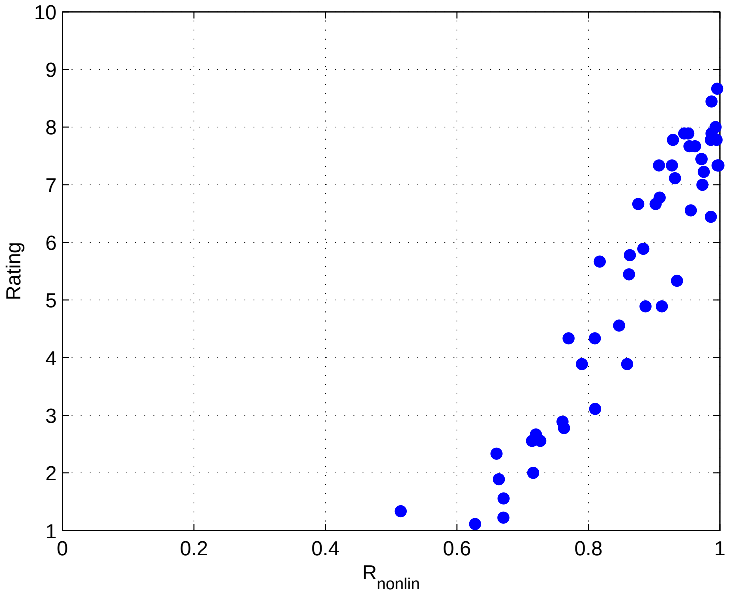



***Figure 4.6:*** *Mean Rating versus* $R_{nonlin}$.

widely varying subjective ratings of the amount of distortion. The correlation coefficients for % THD and % IMD were found to be, $r = -0.4845$, and $r = -0.4466$, respectively, confirming this observation and confirming the initial hypothesis presented in the beginning of this paper.

Additionally, they are useful in showing the vast improvement of the DS and $R_{\text{nonlin}}$ metrics in terms of their ability to provide a perceptually relevant metric to the perception of nonlinear distortion. From

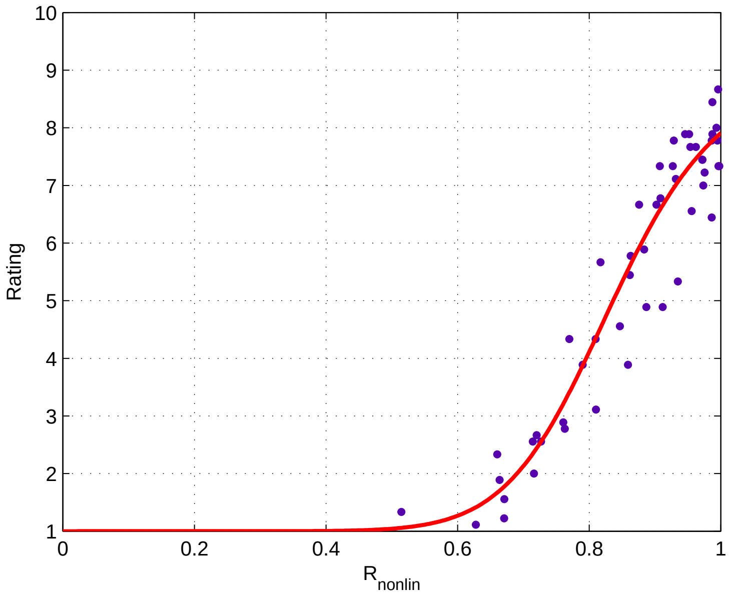



***Figure 4.7:*** *Curve fitting of mean rating versus $R_{nonlin}$. Solid curve represents the fitted curve obtained using the equation* $Predicted\ Rating = \frac{(a+bR^c_{nonlin})}{(a+R^c_{nonlin})}$, *where* $a = 0.1674$, $b = 9.063$, $c = 10.1$

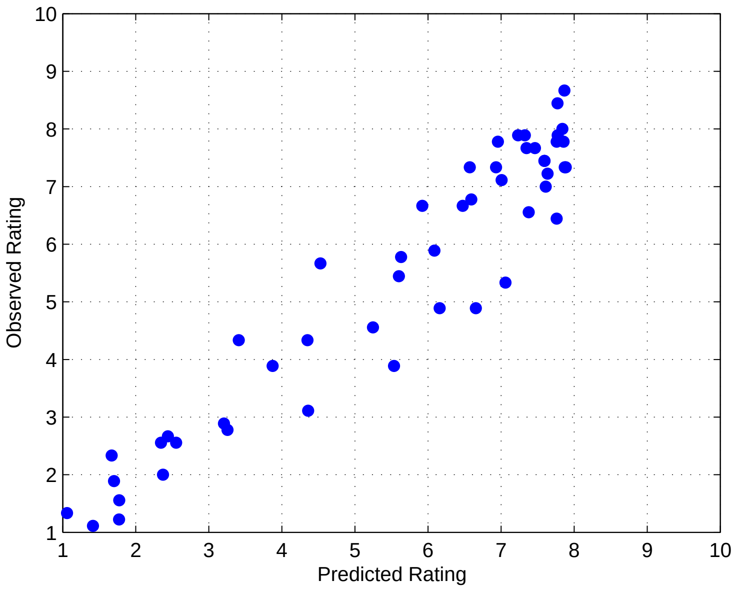
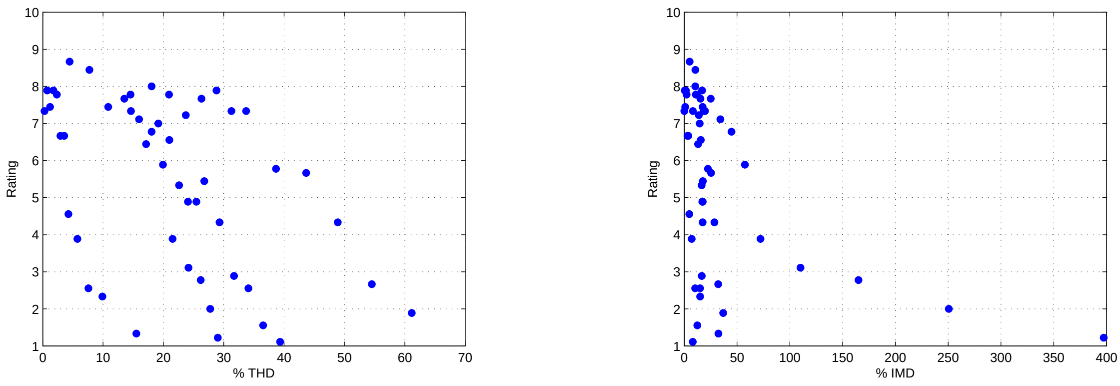



***Figure 4.8:*** *Predicted Rating versus Observed Rating obtained using the fitted curve shown in figure 4.7.* $r = 0.9547$

this, it can be concluded that both the THD and IMD metrics are completely insufficient metrics for describing the perceptual effects of nonlinear distortion and are therefore excluded from any further evaluation in the remainder of this paper.




(a) Mean Rating vs. % THD, $r = -0.4845$

(b) Mean Rating vs. % IMD, $r = -0.4466$

***Figure 4.9:***

## 4.6 Discussion of Results

Both the DS and $R_{nonlin}$ metrics prove to be highly correlated with subjective ratings, and therefore are good metrics in evaluating the perception of nonlinear distortion. While the $R_{nonlin}$ provides a slightly higher correlation to subjective ratings in the evaluation of artificial distortion, the DS metric also proves to be highly correlated. In contrast, the THD & IMD metrics prove to be highly uncorrelated with subjective perception of distortion.

In terms of efficiency of use, the DS metric provides a faster calculation of the metric as fewer steps are involved and less complex filtering is involved in the computation of the metric. However, the $R_{nonlin}$ metric may prove to be more versatile, as it has been shown to be highly correlated to distortions produced by real transducers as well [26]. The $R_{nonlin}$ metric must be fit to subjective ratings of distortion as described above. The parameters used in the fitting equation may not necessarily show the same correlation for other types of distortion and stimuli. Practical use of the $R_{nonlin}$ metric would require typical fitting values pre-determined for a variety of music and speech including different distortion types.

## 4.7 Summary

This chapter presented the design of the listening experiment to verify the correlation of the distortion metrics with subjective perception of nonlinear distortion. In order to determine the correlation between nonlinear distortion and subjective perception of the distortion, an experiment was designed in which the amount of distortion was varied and subjects were asked to rate the perceived amount of distortion. The metric values for the distorted stimuli were then calculated for each distortion metric and plotted versus the subjects' responses.

Analysis of the data obtained from the experiment showed that subjects were consistent in their responses and showed a statistical independence between the stimuli presented and subject ratings, indicating that a broad range of distortions were tested. Both the DS and $R_{nonlin}$ metrics were found to be highly correlated with the subjective perception of nonlinear distortion. In contrast, both the THD and IMD metrics were found to be uncorrelated with the subjective perception of nonlinear distortion.

With the DS and $R_{nonlin}$ metrics verified as being correlated with the perception of distortion, the metrics may be used to investigate other perceptual properties of nonlinear distortion. Therefore, the threshold of audibility of nonlinear distortion may be investigated with a metric related to the perception of nonlinear distortion which was never before available with the conventional metrics.

CHAPTER 5

# EXPERIMENT 2: DETERMINATION OF NONLINEAR DISTORTION THRESHOLDS

---

*This chapter documents the methods used to obtain nonlinear distortion thresholds for four types of nonlinear distortion systems. The design of the experiment is first presented detailing the distortion types and music stimuli used. Details of the performed listening evaluations are also given. The chapter also presents a statistical analysis of the results.*

## 5.1 Thresholds of Nonlinear Distortion

A listening test was designed to find the point at which nonlinear distortion is "just audible," or in other words, the threshold of perception of nonlinear distortion. In order to design the listening test, an appropriate metric is needed to serve as the dependent variable which is varied during the test. In the case of finding the thresholds for different kinds of distortion it was selected to use the DS metric as the dependent variable. The DS metric was chosen as it is the most physical metric of the metrics discussed in this report in the sense that it incorporates very little psychoacoustic modeling in deriving the metric. Therefore the DS metric, while an improved metric over such metrics as THD and IMD, still provides a physical measurement of the amount of distortion in a signal.

In experiment 1, the DS metric was found to be linear with respect to subjective perception of distortion. As the DS value increased, the perception of the distortion increases proportionally. Using a linear metric provides the advantage of making it easier to determine the threshold, since a linear interpolation method may be applied in deriving the threshold value.

In such a way, varying the DS metric provides a way of varying the physical amount of distortion in the signal. Once the threshold is determined, the DS value for this threshold can easily be mapped to a corresponding $R_{nonlin}$ value from the transfer function and its corresponding parameters which have been computed a priori. It is highlighted that the DS and $R_{nonlin}$ values are derived directly from a nonlinear transfer function and it is not possible to map a DS value directly to a $R_{nonlin}$ value without knowledge of the transfer function.

### Experimental Design

In order to find the threshold of nonlinear distortion, an appropriate psychometric design is needed to provide both accurate and repeatable results. Traditionally, threshold experiments are conducted using a single stimulus which is varied according to some physical parameter in question. Subjects are then presented with the stimuli and must respond when the subjective response to the parameter in question is perceived by the subject. This results in both positive (correct) and negative (incorrect) responses with respect to the varying parameter.

In this experiment, however, the traditional single stimulus procedure is not desirable as there is no reference stimulus with which subjects can base their responses. Therefore, a 2 Alternative Forced Choice (2AFC) paradigm was used. This procedure is preferred over the traditional one stimulus procedure, as the understanding of the concept of distortion is not always widely understood. In the 2AFC method, a subject is presented with both a distorted stimulus along with the undistorted or reference signal. The subject is then asked to determine which of the two stimuli is the distorted stimulus.

It is desired to design a test that is both accurate and efficient in its implementation, therefore an appropriate algorithm should be selected that fulfills both of these requirements. Several commonly used psychometric methods include the method of constants, the method of limits, simple up-down method, and the transformed up-down method, the latter two being the most common among them.

Traditionally, the methods described above are conducted using stimuli whose parameters are varied by a constant factor, or step size. In such a way, subsequent stimuli are varied according to the step size until the threshold is reached. While this method is accurate, it is not necessarily efficient. If the region where the threshold lies is unknown, this procedure may take a long time to converge to the threshold. Using an adaptive procedure, or a procedure in which the step size is varied based on subject responses, provides a more efficient method of determining the threshold. Using such a method, one can initialize the test using a large step size and as responses are recorded, the step size may vary accordingly. It has been shown that this adaptive method of reducing the step size leads to a maximal rate of convergence to the desired threshold value. [11]

### Algorithm Implementation

An adaptive transformed up-down algorithm with descending distortion scores (i.e. starting with very distorted signals) was selected due to its accuracy and efficiency. The choice between descending or ascending starting points in an up-down method is important due to subject biasing, known as hysteresis. When a subject is presented with stimuli with descending level, it is possible for the a subject to continue perceiving an effect even after it is physically gone. Conversely, when a subject is presented with stimuli with increasing level, a subject may not perceive the increase in level until a marked increase in level has occurred, which may be beyond the threshold. In this fashion, a descending experiment will always yield a lower threshold than an ascending experiment. Up-Down methods have the advantage of limiting this bias to a certain extent due to the fact that it converges to the threshold from both sides. It was decided that beginning the test with an audible example of the effects of distortion on the particular sample was important in allowing the subject time to perceive the effects of distortion on the test signal, and therefore a descending starting point was chosen.

The initial step size of the test was selected to be 25 DS, with subsequent step sizes of 10 DS and 5 DS after the 1st and 3rd reversals, respectively. The resulting resolution of 5 DS was found to be suitable from past experience. Additionally, the cross-correlation coefficient was also calculated for two signals differing by a value of 5 DS and was found to have a maximum value of 0.9985, leading to the assumption that a difference in 5 DS of two given signals provides a high resolution. The test was terminated after 6 reversals. It is recommended to use at least 6-8 reversals, but using 6 has the added advantage of reducing the duration of the experiment which reduces subject fatigue.

The initial starting DS was determined empirically in a pilot test. As stated before, it was thought to be important to give the subjects an initial impression of examples of distorted signals. Therefore, for each test case, a DS value was found where the distortion was clearly evident.

### Experimental Stimuli

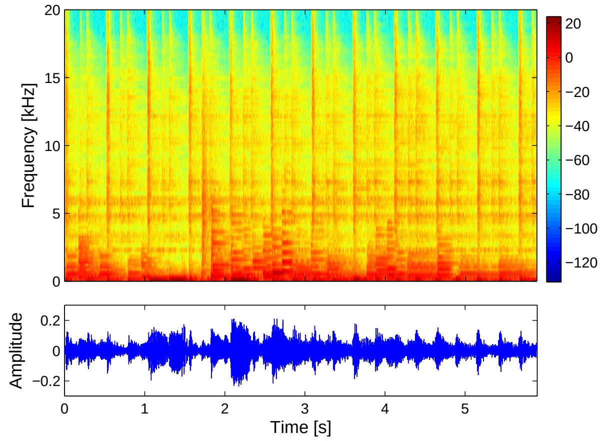
It was shown in the previous experiments, that while the DS and $R_{nonlin}$ metrics are well correlated with subjective ratings for the perceived amount of distortion, it can not be assumed that each type of distortion will have the same threshold. Therefore it was decided to test the threshold for each of the four types of distortion described in section 4.1 independently.

It was desired to investigate the influence of the type of stimulus used on the measured thresholds and therefore two stimuli were used for each of the four distortion types to be investigated, resulting in eight separate cases. The two stimuli were chosen to both have a relatively constant overall level, but with slightly different time and frequency content. The stimuli selected were a jazz excerpt from the song "Charles Christopher" by the Phil Woods Trio off the Chesky Records Audiophile Test CD [21] and a classical excerpt from "Lyric Andante" by Reger taken from the AAU High Fidelity Reference CD#2 [28]. A spectrogram of each excerpt is shown in figure 5.1.

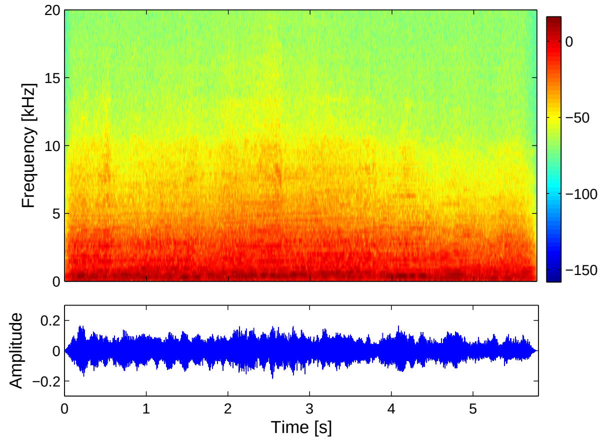
The main difference between the musical excerpts in figures 5.1(a) and 5.1(b) is that the jazz excerpt contains more transients than does the classical excerpt, and that overall time-frequency content of the classical excerpt is more constant.

## 5.2 Test Implementation

As eight cases were to be evaluated during the test, it was decided to split the test into two groups in order to reduce the duration of the complete experiment. The two groups were divided into a polynomial distortion group and a hard clipping group. It was calculated that each group would require approximately 45 minutes from a subject. Each group consisted of four test sessions to evaluate the threshold of distortion for each type of distortion in the group. Table 5.1 shows the two groups and the test sessions involved. Within each group, a database of stimuli were created which contained distortion levels varying in increments of 5 DS from 0 to 200 DS. The maximum level of 200 DS was selected as it was clearly audible as confirmed by pilot test subjects. As in the previous experiment, all stimuli were loud-




(a) Spectrogram for jazz excerpt




(b) Spectrogram for classical excerpt

***Figure 5.1:*** *Spectrograms for music excerpts.*

ness normalized to the same level.

| test session | Group 1 | Group 2 |
|---|---|---|
| 1 | Cubic distortion applied to classical music sample | Symmetric hard clipping applied to classical music sample |
| 2 | Cubic distortion applied to jazz music sample | Symmetric hard clipping applied to jazz music sample |
| 3 | Quadratic distortion applied to classical music sample | Asymmetric hard clipping applied to classical music sample |
| 4 | Quadratic distortion applied to jazz music sample | Asymmetric hard clipping applied to jazz music sample |

***Table 5.1:*** *Division of groups in the listening test.*

A total of twelve subjects participated in the listening experiment, where six evaluated Group 1, and six evaluated Group 2. Each subject was screened with an audiometry prior to the test to ensure that they had no hearing loss greater than 20 dB HL. Each test subject was then given a set of instructions for the test, as well as a questionnaire in order to assess any other possible history of hearing impairments that could adversely affect the test.

A Graphical User Interface (GUI) was programmed in MATLAB to interface with the subjects. A screen-

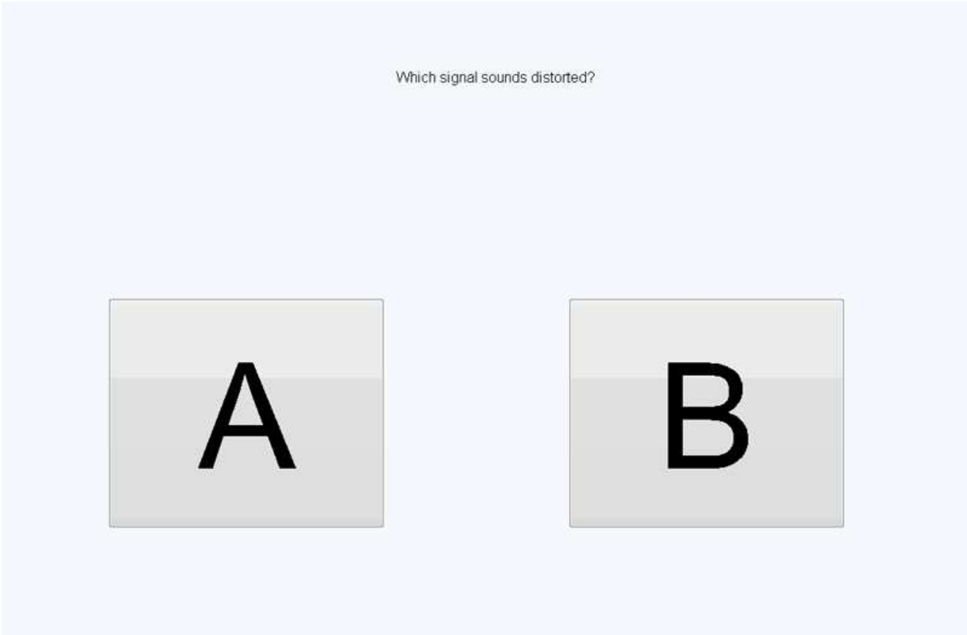



***Figure 5.2:*** *GUI for threshold experiment.*

shot of the GUI is shown in figure 5.2. Prior to the test sessions, a brief training session was conducted where the subjects were presented with two extreme examples of the type of 2AFC comparisons they would hear during the test sessions. Extreme examples were used where the distorted signals presented were the most severe cases they would encounter and the distortion was clearly evident. This was done to ensure that the subjects were familiar with the concept of distortion as well as to familiarize the subject with the types of signals that would be evaluated during each test sessions. After the training session, the main test began. Subjects were allowed breaks between each of the four test sessions.

## 5.3 Results from the Threshold Experiments

Before describing the procedure used to arrive at the estimation of the threshold, $X_{70}$, two examples of the transformed up-down procedure used during the experiments are presented. Figure 5.3 plots the results from experiment A using cubic distortion and figure 5.4 experiment B using symmetrical hard clipping both applied to a sample of classical music. It is re-iterated that a different subject was used for each experiments. In the figures below, descending values values of DS indicate decreasing levels of distortion.

From figure 5.3 it can be noted that the initial DS level is at 120 which decreases in steps of 25 to 45 DS at trial number 7. This point is the first reversal reached during the experiment. The DS level returns to the level preceding the reversal point from which the step size is changed to 10 DS for the next two reverals. The third reversal occurs at trial 16 from which point the step size changes to its final value of 5 DS. Referring to experiment B in figure 5.4 it is noted that the initial DS level is at 100. As was seen with almost all of the hard clipping distortion threshold experiments, the first reversal occured at DS 0.

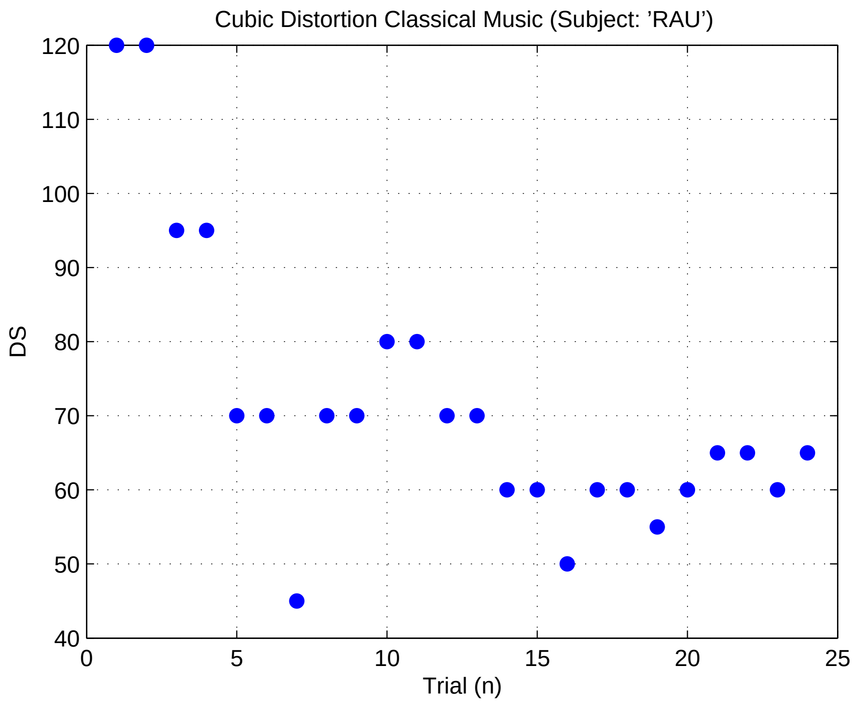



*Figure 5.3: Experiment A: cubic distortion applied to classical music sample.*

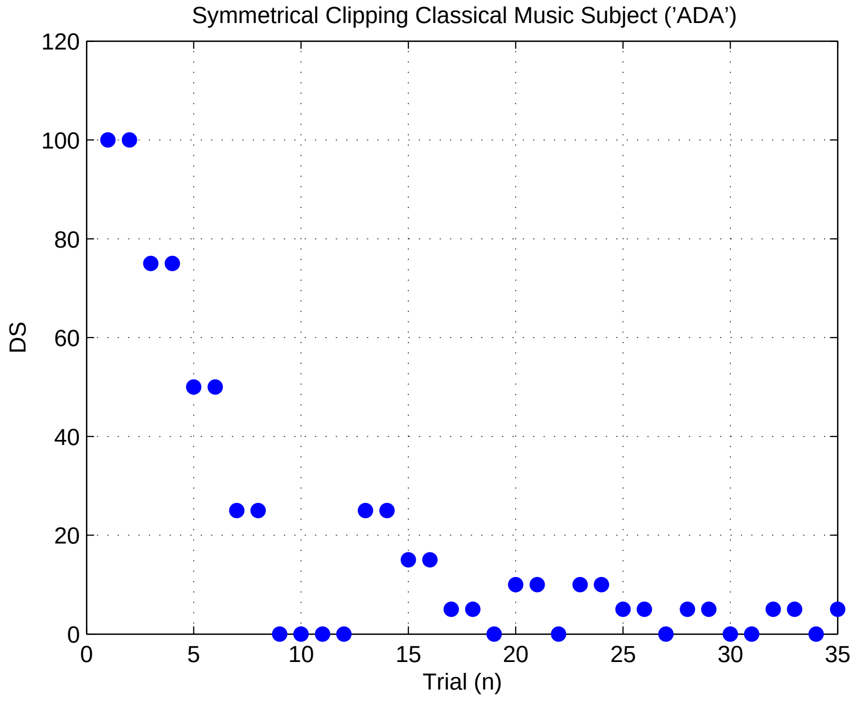



*Figure 5.4: Experiment B: symmetrical clipping applied to classical music sample.*

As the subject reached the 0 DS level, it was only a matter of probability that the subject provided the first negative response as the subject was evaluating an undistorted A and B presentation. Experiment

B carried forward as described above with the only difference being a very clear convergence region between 0 and 5 DS. Experiment B taken from the hard clipping distortion demanded more trials from the subject and therefore required more time. This is due to the fact that this subject was able to provide more positive responses than the subject in experiment A.

Estimation of the threshold, $X_{70}$, was derived from the data plotted above. The method used to arrive at the estimation is based on taking the midpoint of every second run as described in [11], called the mid-run estimate. A run, within the context of the described experiments, is defined as a sequence of distortion level changes in only one direction. For example, the first run in figure 5.3 is between trials 1 and 7. The second run is between trials 7 and 11, the third run between trials 11 and 16, and so on. The midpoint values for every second run were thusly averaged to arrive at a single threshold estimate. It is furthermore described in [11] that taking every second run reduces estimation bias. This is important for the purposes of this experiment as the first run would undoubtedly affect the overall threshold estimation as the initial DS level is far from the convergence level. The mid-run estimates for experiment A for runs 2,4,6 and 8 are 62.5, 55.0, 60.0 and 62.5, respectively. The overall average is then defined as the estimated threshold which is at 60 DS. The mid-run estimates for experiment B for runs 2, 4, 6, 8, 10 and 12 are 12.5, 5.0, 5.0, 2.5, 2.5 and 2.5, respectively.

Each of the 12 subjects had evaluated 4 out of 8 groups of distortion. A subject evaluated distortions from either items 1-4 or items 5-8 as listed below. The eight types are:

1. Cubic distortion applied to classical music sample (denoted as cc in the graphs).
2. Cubic distortion applied to jazz music sample (cj).
3. Quadratic distortion applied to classical music sample (qc).
4. Quadratic distortion applied to jazz music sample (qj).
5. Symmetric hard clipping applied to classical music sample (sc).
6. Symmetric hard clipping applied to jazz music sample (sj).
7. Asymmetric hard clipping applied to classical music sample (ac).
8. Asymmetric hard clipping applied to jazz music sample (aj).

The following analysis of the threshold estimates is based on the midpoint estimates from each of these groups. Each group listed above was evaluated by 6 individuals resulting in 6 estimates of the threshold for that group. figures 5.5 shows a boxplot representation of the collected threshold estimates in terms of the DS metric. The red line represents the median of the six threshold estimates for each group. The top and bottom blue lines plot the upper and lower quartile values and the whiskers extending from the end of each box show the extreme upper and lower values. Outliers are the data plotted in red which are data points beyond the 95% confidence interval. An errorbar plot showing the mean value of each group along with the 95% confidence interval for that group is shown in figure 5.6.

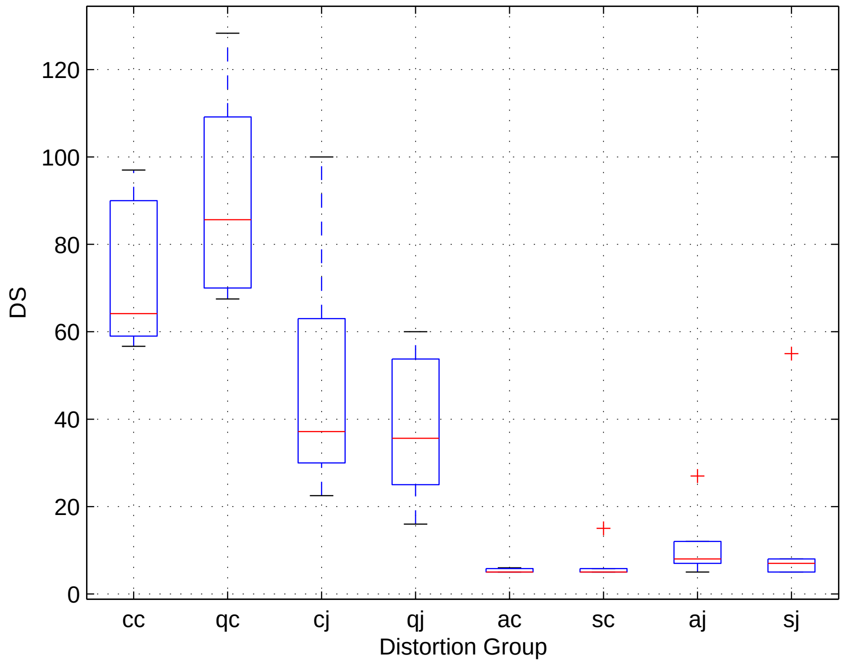



***Figure 5.5:*** *Boxplot representation of collected mean midpoint threshold estimates [95% confidence interval].*

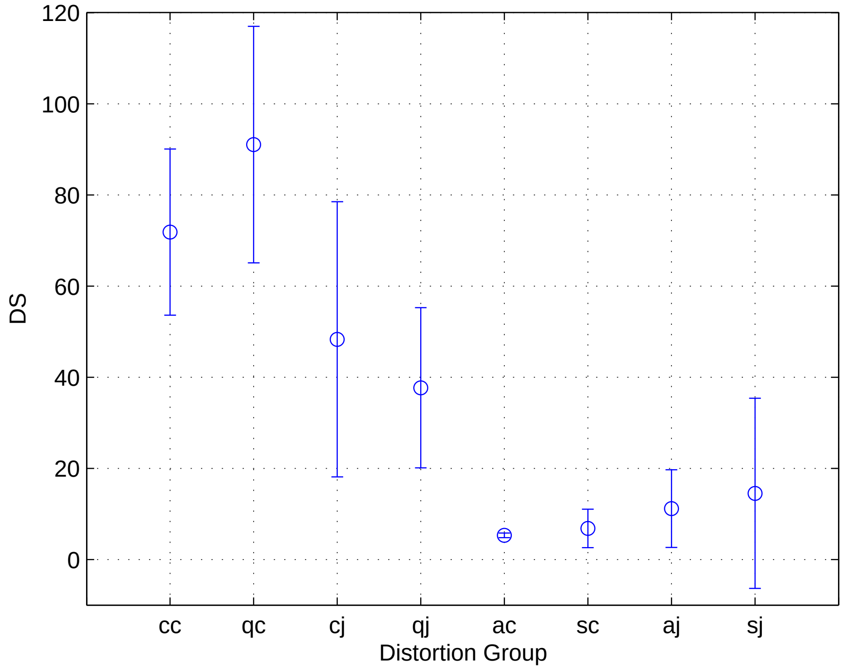



***Figure 5.6:*** *Boxplot representation of collected mean midpoint threshold estimates [95% confidence interval].*

The mean mid-run estimates for each subject's threshold experiment was mapped to the equivalent $R_{nonlin}$ value as described in the above section. The resulting threshold estimates in terms of $R_{nonlin}$ are

| Group | Threshold (DS) | Threshold ($R_{nonlin}$ ) |
|---|---|---|
| cc | 71.8 | 0.974 |
| qc | 91.0 | 0.958 |
| cj | 48.3 | 0.985 |
| qj | 37.7 | 0.992 |
| ac | 5.3 | 0.999 |
| sc | 6.8 | 0.999 |
| aj | 11.2 | 0.998 |
| sj | 14.5 | 0.996 |

***Table 5.2:*** *Mean values for thresholds obtained using DS and $R_{nonlin}$ metrics*

shown below in the boxplot in figure 5.7 and the errorbar plot in figure 5.8. The $R_{nonlin}$ values begin at 1, which would indicate a undistorted signal. Table 5.2 summarizes the obtained mean threshold values for all eight distortion groups.

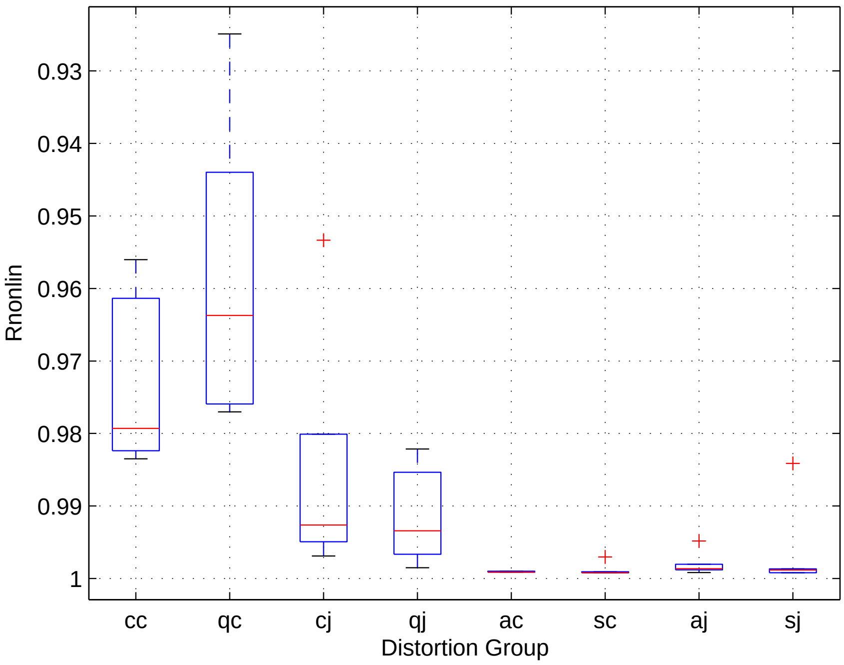



***Figure 5.7:*** *Boxplot representation of collected mean midpoint threshold estimates using $R_{nonlin}$ metric [95% confidence interval].*

## 5.4 Discussion of Results

The most apparent observation to note is the large difference in the obtained thresholds between the hard clipping (ac,sc,aj,sj) and the polynomial (cc,qc,cj,qj) distortion groups. The thresholds in terms of both DS and $R_{nonlin}$ for the hard clipping group was found to be much lower than the polynomial group. The overall variance for the hard clipping groups was also much less. This indicates that the overall subject thresholds are very similar. This was also noted during listening experiments, where each session would converge to a defined region as shown in figure 5.4.

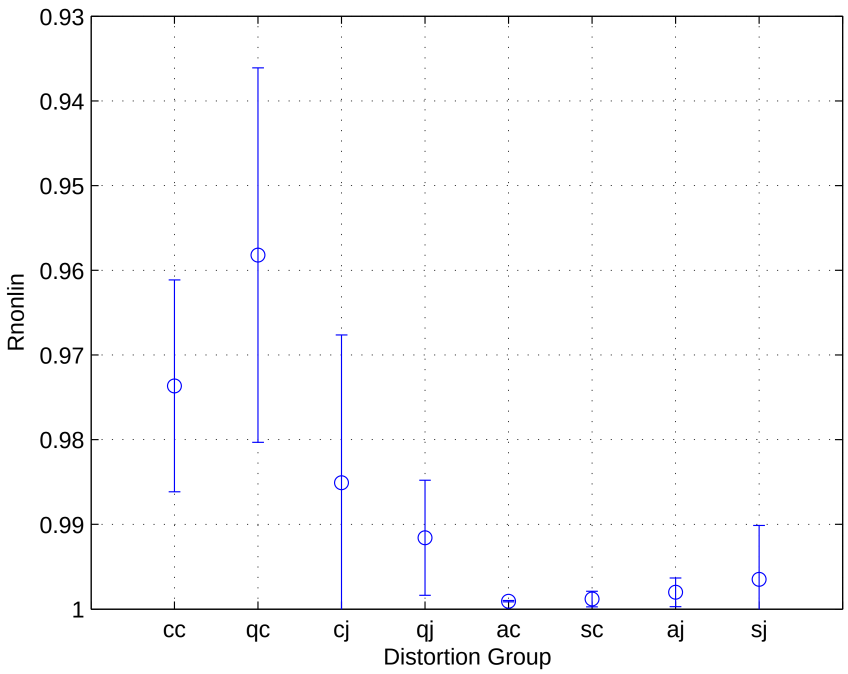



***Figure 5.8:*** *Boxplot representation of collected mean midpoint threshold estimates using $R_{nonlin}$ metric [95% confidence interval].*

Within the polynomial group, the observed thresholds were found to be different not only from the different distortion type, but also between music stimuli. The variance within each of the polynomial groups was also quite large. This results from the different thresholds obtained from the subjects within these groups.

Similar statistical results were found using the corresponding $R_{nonlin}$ metric. Overall, the experiment yielded a large variance for many distortion groups. This results from the relatively small group of subjects used in the experiment. Further experiments might include a larger group of subjects to reduce the variance.

## 5.5 Summary

An experiment to determine the threshold of audibility of four nonlinear distortion types was implemented using a transformed up-down method and a 2 alternative forced choice (2AFC) paradigm. Two types of music stimuli, a jazz stimulus and a classical stimulus, were used in the evaluation to determine the possible effects of music type on the thresholds of audibility for each distortion type.

At total of twelve subjects participated in the experiment and were split into two groups to evaluate either polynomial distortion or clipping distortion. Using a mean mid-run estimate calculation, thresholds were determined for each type of distortion and each type of stimulus.

CHAPTER 6

# DISCUSSION & CONCLUSIONS

## Discussion

From the threshold experiment using the DS metric as an independent variable, it was found that subjects were more sensitive to the hard clipping distortion than to the polynomial type distortion. The hard clipping distortion introduces distortion products which span the entire spectrum of audibility. These higher order products have been found to be more audible [3]. This is due to the fact that higher order distortion products are less likely to be masked by tones present in the music signals. The result of this observation is quite interesting since it implies that one type of nonlinear distortion is more audible than another. This indicates that nonlinear distortion thresholds are somehow dependent on the type of applied distortion.

Another interesting point can be observed from the thresholds within the polynomial distortion group where there is a statistical difference in mean value within the quadratic distortion group. This indicates that the thresholds show a certain dependence on the type of stimulus. This observation was also made in 1995 by Schmidt [3, 22] in which it was noted that audibility thresholds of nonlinear distortion was strongly dependent on the type of stimulus used. It was further suggested in [22] that the influence of the signal's temporal characteristics were stronger than its spectral components. Within the cubic group, it cannot be explicitly said that the overall mean thresholds between music types are different. However, in the authors' opinion a larger data set would reveal a similar dependence on signal type.

Most commercial audio transducers and amplifiers used today seldom operate in their hard clipping regions. Studying the thresholds of nonlinear distortion for a greater variety of transducer models is therefore of more practical importance. Many aspects of real transducers, for example, can be modeled as second or third order systems. The squared and cubic distortion models used in this project can be related to such models. Assuming that a device under test (DUT) is not operating in a clipping region, it is suggested from the obtained data that 37.7 DS (0.992 $R_{nonlin}$ ) is the minimum audible threshold for the conditions used in experiment 2.

Further research is required to investigate the nonlinear distortion thresholds for a wider variety of nonlinear systems for a wide range of stimuli. These thresholds and also subjective ratings could be used within industry as an improved method of determining the nonlinear effects of their products. The minimum nonlinear distortion threshold obtained from this wide variety of conditions could be used as a benchmark for the audio industry.

More issues with the THD and IMD metrics were revealed during course the project work, other than their poor correlation with subjective ratings. Many researchers often state conflicting thresholds of audbility in terms of both THD and IMD [3]. Furthermore, many of these values are expressed purely as a percentage and without the conditions under which those values were obtained. The importance of expressing the conditions of the test signal used for a THD or IMD measurement stems from the fact that the output from a nonlinear device is level dependent. That is, a change in test signal amplitude will change the observed THD value.

## Conclusions

This report investigated the subjective perception and thresholds of nonlinear distortion in complex music signals by means of conventional and newer psychoacoustically based metrics. A large portion of the work presented in this project is related to the understanding and implementation of the new nonlinear distortion metrics. The ultimate goal of the project was to apply these metrics and obtain meaningful nonlinear distortion thresholds.

An overview of the theory of nonlinear distortion was presented along with the relevant theories in psychoacoustics involved in the perception of nonlinear distortion. A comparison of various metrics including conventional metrics such as THD and IMD were presented, along with newer metrics such as the Gedlee, DS and $R_{nonlin}$ metrics. The aim of the newer metrics is to improve on the conventional metrics by incorporating psychoacoustical modeling in order to develop metrics that are correlated with the subjective perception of distortion.

A verification experiment was conducted to confirm the correlation of the conventional metrics, THD and IMD, and the newer metrics, DS and $R_{nonlin}$ , with the perception of nonlinear distortion using one music stimuli and four types of nonlinear distortion. It was found that both conventional metrics were not well correlated with subjective perception of distortion with correlations of -0.4845 and -0.4466, respectively. Both of the newly developed metrics, on the other hand, were found to be well correlated with subjective data obtained with overall correlations of -0.9486 and .9547, respectively. From these correlations, it was concluded that the conventional metrics were not well suited to assess the subjective perception of distortion. In contrast, the DS and $R_{nonlin}$ metrics provide a means for objectively quantifying the perceived amount of distortion in a complex music stimulus.

The aim of the project was to obtain thresholds for the audibility of nonlinear distortion in terms of the subjectively correlated metrics. In doing so, a more revealing threshold may be obtained than previously described by research determining threshold using conventional metrics. Since conventional metrics show widely varying subjective rating of stimuli with the same metric values, determining a threshold using such a metric may be misleading. With highly correlated metrics such as the DS and $R_{nonlin}$ metrics, a perceptually relevant value may be obtained.

A threshold experiment was conducted using the psychoacoustic metrics. The experiment investigated the same four distortion types used in the verification experiment, but with two different music types in order to assess the dependence of stimulus on the obtained thresholds. From the nonlinear threshold experiments, it was concluded that nonlinear distortion thresholds are dependent on the type of applied distortion and on the characteristics of the stimulus.

# BIBLIOGRAPHY


[1] E de Boer. Synthetic whole-nerve action potentials for the cat. *Journal of the Acoustical Society of America*, 58(5):1030–1045, November 1975.

[2] Richard C. Cabot. Fundamentals of modern audio measurement. *Journal of the Audio Engineering Society*, 47(9):738–762, September 1999.

[3] Eugene Czerwinski, Alexander Voishvillo, Sergei Alexandrov, and Alexander Terekhov. Multitone testing of sound system components - some results and conclusions, part 1: History and theory. *Journal of the Audio Engineering Society*, 49(11):1011–1047, November 2001.

[4] Eugene Czerwinski, Alexander Voishvillo, Sergei Alexandrov, and Alexander Terekhov. Multitone testing of sound system components - some results and conclusions, part 2: Modeling and application. *Journal of the Audio Engineering Society*, 49(12):1181–1193, December 2001.

[5] Zwicker E. and Fastl H. *Psychoacoustics, Facts and Models*. Springer-Verlag, 1990.

[6] B.R. Glasberg and Brian C.J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138, 1990.

[7] B.R. Glasberg and Brian C.J. Moore. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50:331–342, May 2002.

[8] Luc Van Immerseel and Stefaan Peeters. Digital implementation of linear gammatone filters: Comparison of design methods. *Acoustics Research Letters Online*, March 2003.

[9] Lidia W. Lee and Earl R. Geddes. Auditory perception of nonlinear distortion. In *Proc. 115th AES Convention, New York, NY*, 2003.

[10] Lidia W. Lee and Earl R. Geddes. Auditory perception of nonlinear distortion - theory. In *Proc. 115th AES Convention, New York, NY*, 2003.

[11] H. Levitt. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49 (2):467–477, 1971.

[12] J. Moir. 'just detectable' distortion levels. *Wireless World*, 1981.

[13] Brian C.J. Moore. *Psychology of Hearing*. Elsevier Academic Press, 5th edition, 2004.

[14] R.D. Patterson. Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America*, 59:640–654, 1976.

[15] R.D. Patterson. The sound of a sinusoid: Spectral models. *Journal of the Acoustical Society of America*, 96(3):1409–1418, September 1994.

[16] R.D. Patterson. A functional model of neural activity patterns and auditory images. *Advances in Speech, Hearing and Language Processing*, 3:547–563, 1996.

[17] R.D. Patterson, M.H. Allerhand, and C. Giguere. Time-domain modeling of the peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, 98:1890–1894, 1995.

[18] R.D. Patterson and Toshio Irino. A time-domain, level-dependent auditory filter: The gammachirp. *Journal of the Acoustical Society of America*, 101(1):412–419, January 1997.

[19] R.D. Patterson, Ian Nimmo-Smith, Daniel L. Weber, and Robert Milroy. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *Journal of the Acoustical Society of America*, 72(6), 1982.

[20] R.D. Patterson and Ian Nirnmo-Smith. Off-frequency listening and auditory-filter asymmetry. *Journal of the Acoustical Society of America*, 67(1):229–245, January 1980.

[21] Chesky Records. Chesky records jazz: Sampler & audiophile test compact disc, vol.1. CD, 1990.

[22] R. Schmitt. Audibility of nonlinear loudspeaker distortion. *Journal of the Audio Engineering Society*, 43(11):402, May 1995.

[23] M. Slaney. Auditory toolbox version 2. Interval Research Corporation 1998-010, 1998.

[24] DS Dansk Standardiseringsrad. Acoustics. method for calculating loudness level. ISO, 07-15 1975. International Standard ISO 532, 1st edition.

[25] Chin-Tuan Tan, Brian C.J. Moore, and N. Zacharov. The effect of nonlinear distortion on the perceived quality of music and speech signals. *Journal of the Audio Engineering Society*, 51(11):1012–1030, November 2003.

[26] Chin-Tuan Tan, Brian C.J. Moore, N. Zacharov, and Ville-Veikko Mattila. Predicting the perceived quality of nonlinearly distorted music and speech signals. *Journal of the Audio Engineering Society*, 52(7):699–711, July 2004.

[27] Steve Temme. Audio distortion measurements. B&K Application Note, May 1992.

[28] Aalborg University. High fidelity reference cd #2. CD.

[29] Masashi Unokia, Toshio Irino, B.C.J. Moore, and R.D. Patterson. Comparison of the roex and gammachirp filters as representations of the auditory filter. *Journal of the Acoustical Society of America*, 120(3):1474–1492, September 2006.

[30] E. Zwicker, H. Fastl, Widnmann U., K. Kurakata, Kuwano S., and S. Namba. Program for calculating loudness according to din 45631 (iso 532b). *Journal of the Acoustical Society of America*, 12 (1):39–42, 1991.

APPENDIX A

# MEASUREMENT REPORTS

## A.1 Harmonic and Intermodulation Distortion Products of Reproduction Chain

### Purpose

Before proceeding with listening experiments related to nonlinear distortion, it is necessary to evaluate the distortion characteristics of the sound reproduction chain. The purpose of this experiment is to investigate the intermodulation and harmonic distortion products in the sound reproduction chain to be used for listening experiments relating to the study of nonlinear distortion. The reproduction chain consists of a soundcard connected to an amplifier connecting to a pair of headphones as shown in Figure A.1. Nonlinear distortion products may result from the soundcard, the amplifier or the headphones.

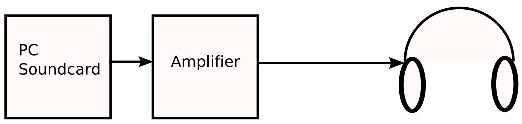



***Figure A.1:*** *Sound reproduction chain related project for listening experiments.*

### Equipment

### Setup and Measurement Description

The complete measurement setup is illustrated in Figure A.2. The soundcard mounted in the desktop PC Akulab13 was used as the sound source for the measurement. The output of the soundcard was initially connected to the Behringer headphone amplifier. The amplifer was connected to the panel connecting the control room to Cabin A in the Acoustics Laboratory at AAU. Within the cabin (small listening room) the right cup of the Beyerdynamic DT990 headphones was placed over the artificial ear. The 01 dB measurement system was used to record the output of the microphone placed within the artificial ear.

MATLAB was used to create the pure tones used in the experiment. To investigate any evidence of harmonic distortion, a single pure tone at 1 kHz was applied to the system and the resulting output spectrum

| Name | Manufacturer | AAU Number |
|---|---|---|
| Harmonie | —- | —- |
| 01dB Measurement System | —- | —- |
| Desktop Computer Akulab13 | Dell | 37726 (monitor) 53118 (tower) |
| Headphones | Beyerdynamic DT 990 | 2036-1 |
| Measurement Mic. | B&K 4134 | 8130 |
| Headphone Amplifier | Behringer HA903 | 33239 |
| Mic. Preamplifier | G.R.A.S.S 26AK | 52665 |
| Artificial Ear | B&K | 07631 |
| Assorted Cables | N/A | N/A |

***Table A.1:*** *Equipment list for evaluating distortion products*

was analyzed from a measurement recording. The intermodulation distortion measurement was made in accordance with the CCIF standard. The CCIF standard recommends using a signal consisting of two tones having equal amplitude differing by 1 kHz. The two tone signal used in this experiment contained frequencies at 2 and 3 kHz.

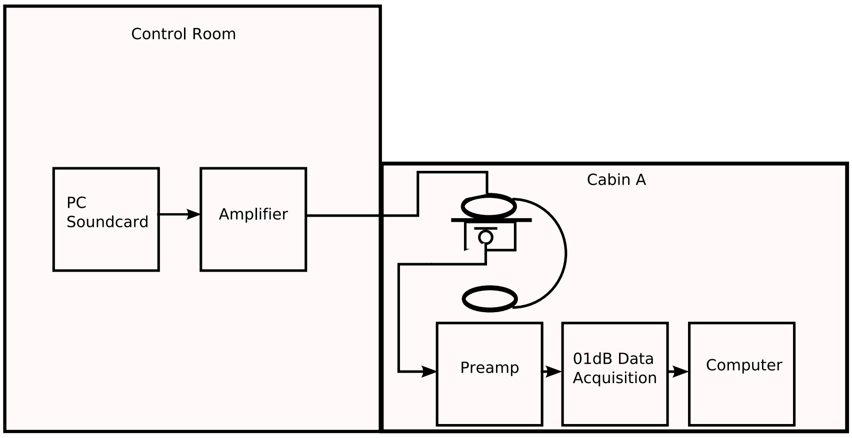



***Figure A.2:*** *Measurement setup for evaluating distortion products.*

## Results and Conclusions

Figure A.3 shows the measured output of the reproduction chain using a 1 kHz tone. The most prominent harmonic distortion products can be found at 2 kHz (29.5 dB SPL), 3 kHz (23 dB SPL) and 5 kHz (29.5 dB SPL). Intermodulation distortion products stimulated by the two tone signals (2 and 3 kHz) were also evident as shown below in Figure A.4. The first intermodulation product is found at 1 kHz which has a sound pressure level at nearly 70 dB. Intermodulation products with relatively high sound pressure level are found throughtout the audible frequency range.

The intermodulation and harmonic distortion products were found to be directly caused by the Behringer headphone amplifier. This device was removed from the reproduction chain and the above procedure

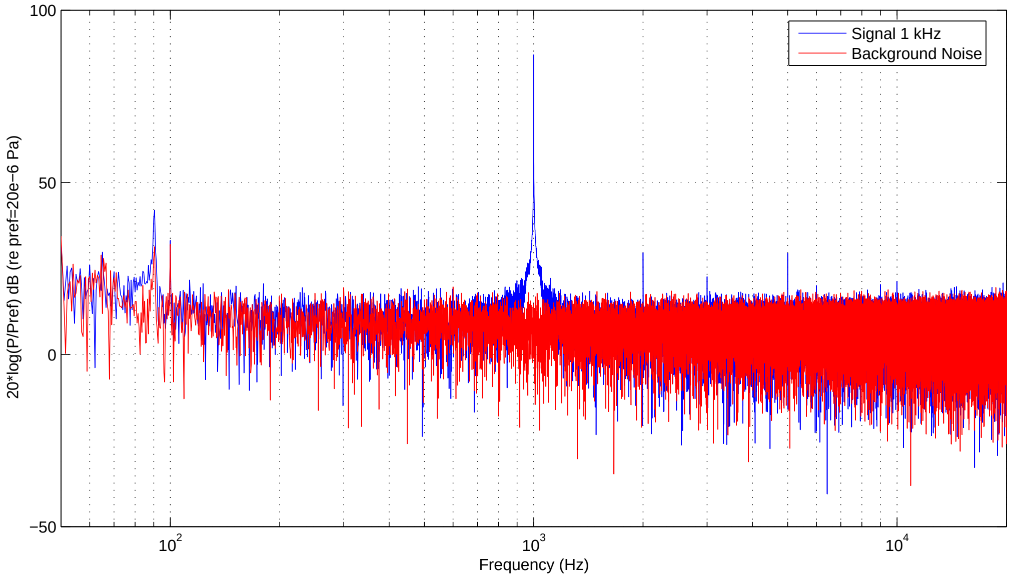



***Figure A.3:*** *Frequency spectrum from 1 kHz input signal.*

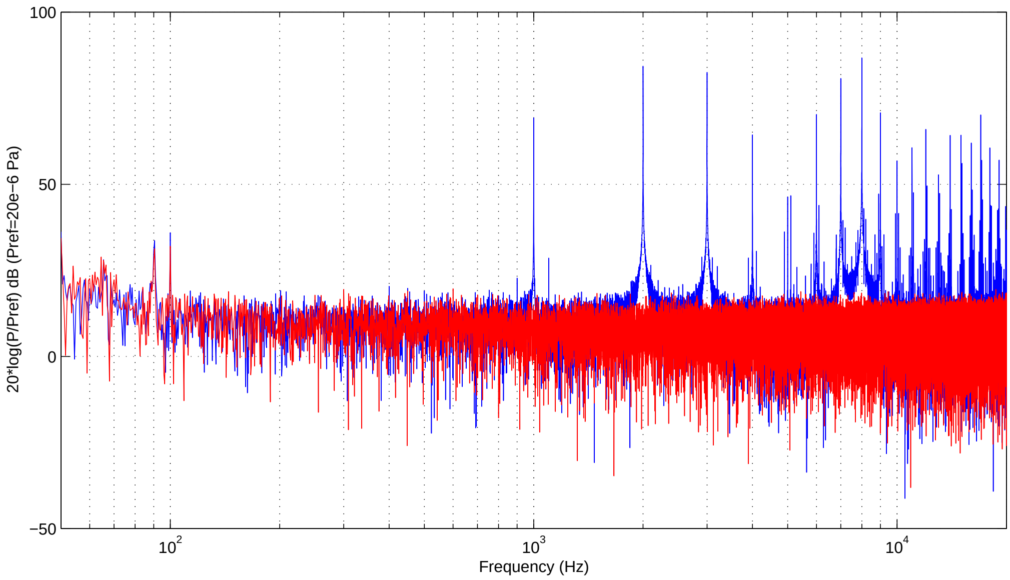



***Figure A.4:*** *Frequency spectrum from 2 and 3 kHz input signal.*

was repeated. Figure A.5 shows the frequency spectrum resultant of the 1 kHz pure tone. There are no apparant distortion products of particular detriment. Figure A.6 shows the frequency spectrum resultant from the 2 and 3 kHz input signal. Once again there are no distortion products of concern. These distortion products are considered to be negligible in comparison with the measurement taken using the amplifier.

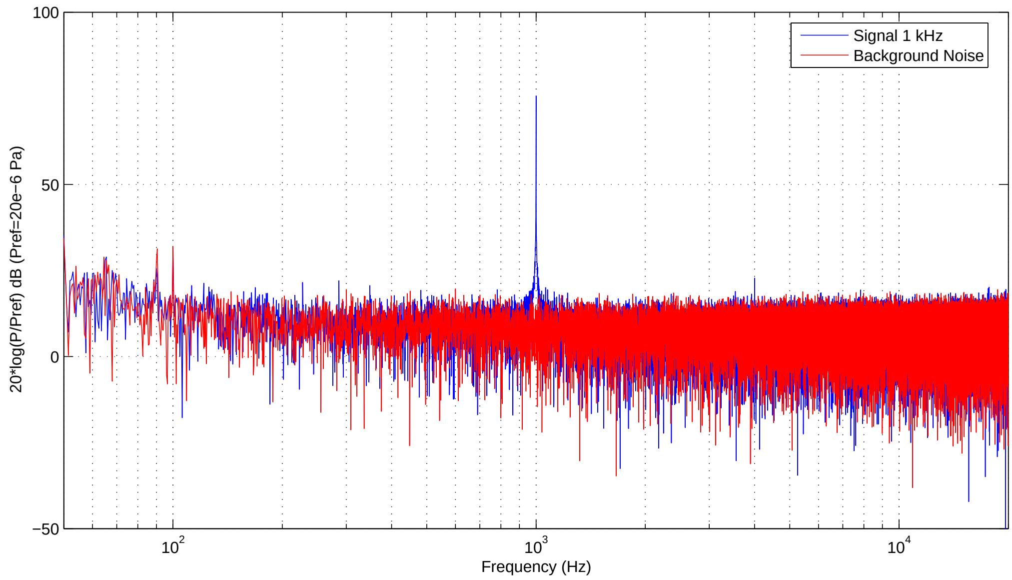



***Figure A.5:*** *Frequency spectrum from 1 kHz input signal with amplifier removed.*

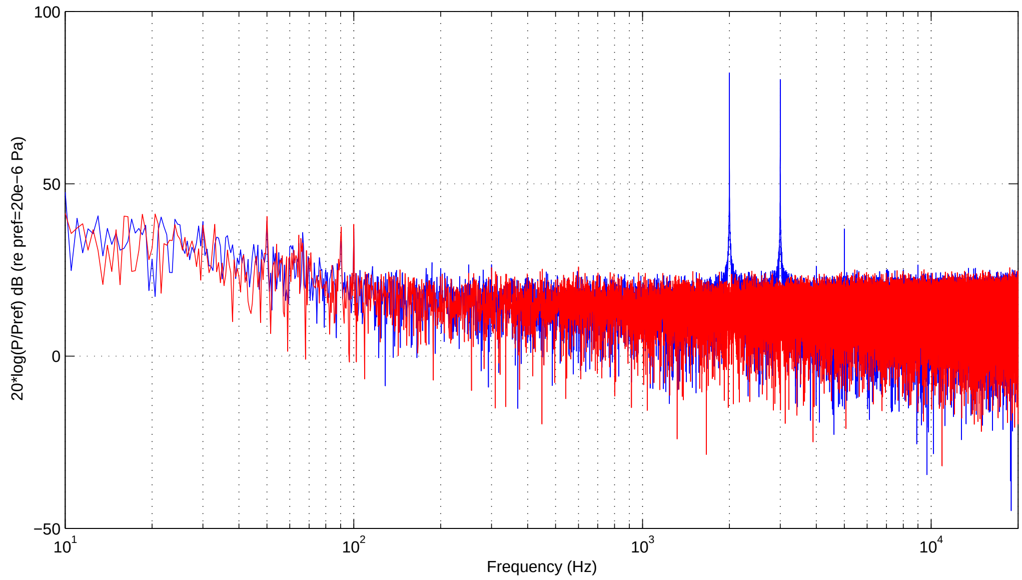



***Figure A.6:*** *Frequency spectrum from 2 and 3 kHz input signal with amplifier removed.*

## A.2 Headphone Impulse Response Measurement

This section describes the measurement setup used to arrive of the average headphone impulse response of the left and right cup of the Beyerdynamic DT-990 headphones. The setup makes use of an artificial dummy head and the MLSSA measurement system in a master/slave configuration enabling acquisition of both dummy head microphones simultaneously.

| Equipment | Type | AAU Number |
|---|---|---|
| MLS system | MLSSA | 37493 |
| MLS system | MLSSA | 26827 |
| Measuring Amp. | B&K | 08022 |
| Measuring Amp. | B&K | 08717 |
| VALDEMAR | AAU | 2150-01 |
| Microphone | GRAS 40AD | −− |
| Microphone | GRAS 40AD | −− |
| Headphones | Beyerdynamic DT990 | 2036-1 |
| Headphone Amp. | Fostex PH-5 | 02092-00 |
| Synchronizing Unit | −− | −− |
| Clock | Philips PM5193 | 02092-00 |
| Mic. Calibrator | B&K 4230 | 08373-00 |

***Table A.2:*** *Equipment used in the measurement of DT990 headphone impulse response.*

Figure A.7 illustrates the setup

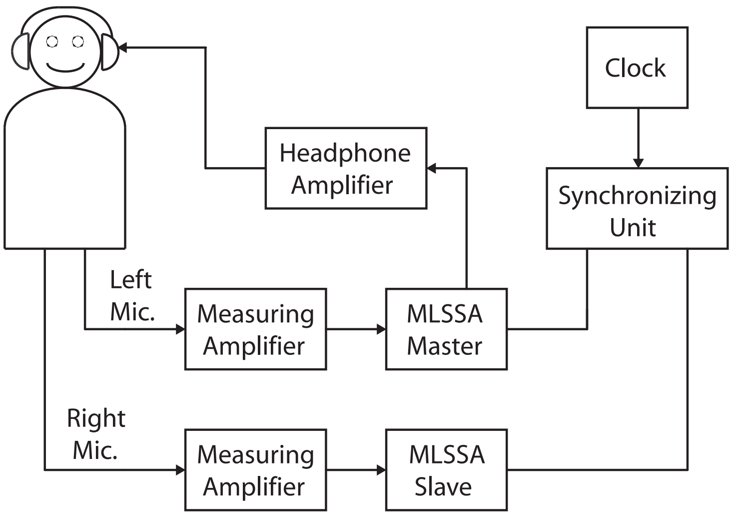



***Figure A.7:*** *Setup for measuring transfer function of Beyerdynamic headphones.*

In the measurement setup two MLSSA measuring systems are setup in a master/slave configuration. This enables the possibility to measure both cups at the same time when using the synchronizing unit. The common clock frequency was set to 48 kHz and the anti-aliasing filter cut-off is set to 20 kHz. The MLS-sequence is set to $\pm 0.5$ V and a length of 4096 samples.

### Measurement description

Before running the actual measurements the microphones have to be calibrated. The left microphone sensitivity was found to be 35.07 mV/Pa and the right microphone to 31.71 mV/Pa.
The level inside the cups was then measured to yield an absolute sound pressure level around 70 dB. The 70 dB absolute SPL is calculated from the impulse response with a built-in function in the MLSSA system. Five different positions of the headphones were measured. The GRAS 40AD microphones are

flat up to 10 kHz and the therefore the microphones were not corrected.

### Frequency Response of Headphones

The average of the measurements was carried out in the frequency domain. All the five frequency response magnitudes were added together and then divided by 5. Figure A.8 illustrates the left cup average frequency response and Figure A.9 the right cup frequency response. The impulse response is divided by the microphone sensitivity to get the output in Pa/V.

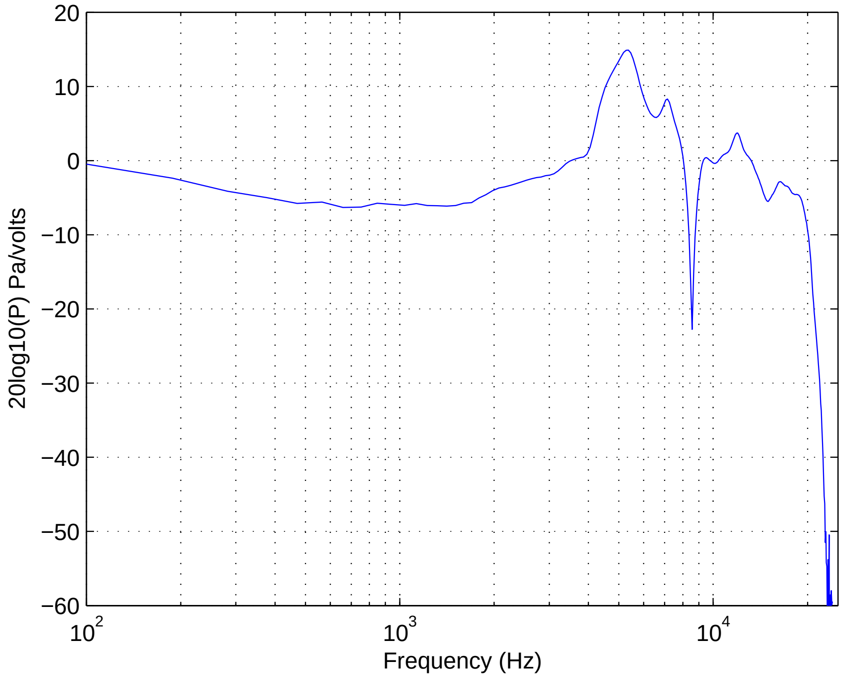



***Figure A.8:*** *Average of left cup's frequency responses.*

The left and right responses are quite similar. However, the right cup appears to have less attenuation up to the 3 kHz gain present in both responses. Both responses have a large dip at about 8 kHz which is more pronounced in the left cup response. Present also is a very steep roll-off at around 20 kHz.

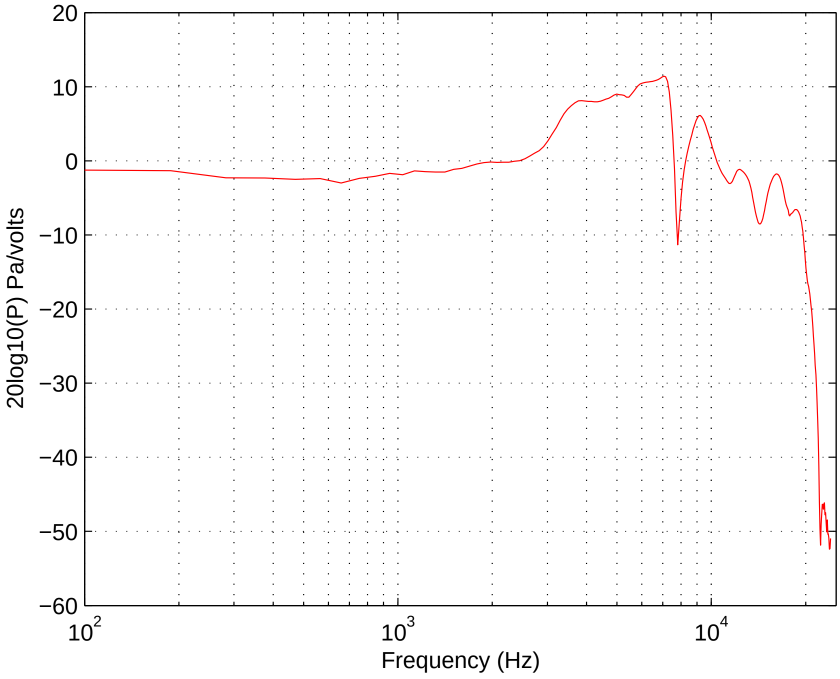



***Figure A.9:*** *Average of right cup's frequency responses.*

APPENDIX B

# LOUDNESS NORMALIZATION OF STIMULI USING LOUDNESS MODEL DIN 45631 (ISO 532B)

*This appendix chapter describes the loudness model developed by Zwicker [5]. The model's use for the loudness normalization of the listening experiment stimuli used for this project is also detailed.*

The sensation which most characterizes the sound intensity of a stimulus is known as loudness. By asking a subject to compare how much louder or softer one sound is heard relative to a standard sound, the sensation stimulus relation of loudness can be measured [5]. Loudness normalization is applied in this project to all the stimuli used in listening experiments. This results from the project's definition of nonlinear systems, in which the nonlinear system contributes frequency components not contained in the original input signal. These added frequency components are of interest to this study, and not any gain or attenuation that may also result from a nonlinear system.

## B.1 Zwicker's Loudness Model

There are three essential stages used to arrive a given loudness level which are depicted in Figure B.1. First, a signal's frequency spectrum is transformed into an excitation pattern. This pattern represents the distribution of excitation at different points along the basilar membrane [13]. The excitation pattern can be found by calculating the output from all auditory filters as a function of their centre frequency. Figure B.2 shows how an excitation pattern can be calculated from 1 kHz tone. The following stage transforms the previous excitation pattern into specific loudness, $N^{'}$, which can be calculated from Equation B.1 where $E_{TQ}$ is excitation at threshold in quiet and $E_0$ is the excitation corresponding to a reference intensity of $I_0 = 10^{-12} W/m^2$ . The overall loudness, $N$, in sones is then calculated from the area under the specific loudness pattern as in Equation B.2.

$$N^{'} = 0.08(\frac{E_{TQ}}{E_0})[(0.5 + \frac{E}{2E_{TQ}})^{.23} - 1] \quad sone/Bark \tag{B.1}$$

$$N = \int N^{'}(z)dz \tag{B.2}$$

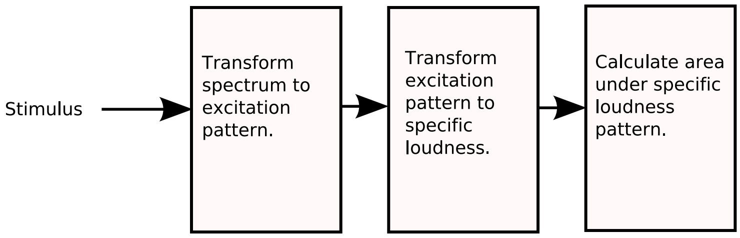
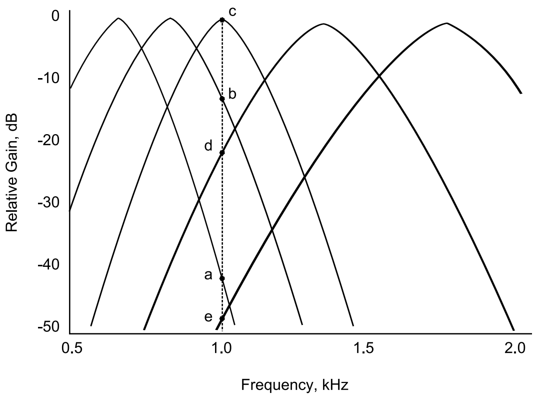
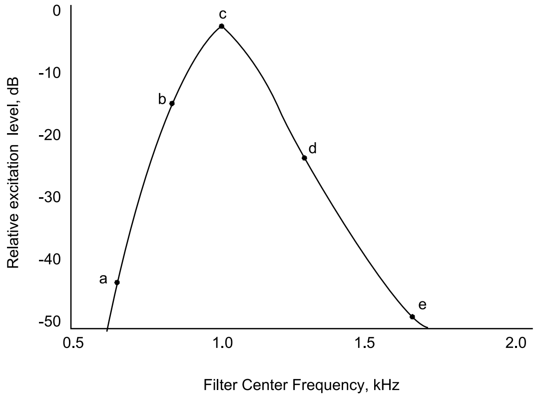



***Figure B.1:*** *Stages used in loudness calculation.*




(a) Bank of auditory filters.




(b) Resulting excitation pattern for a 1 kHz tone.

***Figure B.2:*** *Excitation pattern of 1 kHz sinusoid derived from output of auditory filters [13, p.90].*

A practical implementation of the above loudness model (ISO 532) also proposed by Zwicker yields a single number corresponding to the loudness level of a given sound from an available spectrum analysis of that sound taken from a physical measurement [24]. A loudness level, expressed in sones can be calculated from either an octave or third octave band analysis of a sound signal. Within the ISO 532 standard, the octave band procedure is referred to as Method A and the third octave band procedure as Method B.

The original procedure is based on a graphical procedure utilizing a set of graphs provided by the ISO 532 standard. The graphs are are selected according to the level of the sound under consideration and the type of sound field being used. The sound type can be either diffuse or front incidence. For the purposes of this project, the Beyerdynamic DT-990 headphones used in the experiments are labeled as being diffuse field headphones and therefore the loudness model assumes a diffuse field.

An example of the graphical procedure calculate the loudness level in sones is shown in Figure B.3. The measured third octave band levels are the horizontal lines outlined in yellow. On the left side of the horizontal bar a straight line is drawn downwards. If the adjacent (right side) third octave band level is lower than the left side third octave band level, then a downward slope is added running parallel to the curves outlined in the graph provided in the standard. The area within these boundaries is shaded in black. To calculate the loudness level in sones, an equivalent rectangular area is drawn on the graph having the same width as the graph. The height of the equivalent rectangular area corresponds to the loudness level as read from the right or left side of the graph.

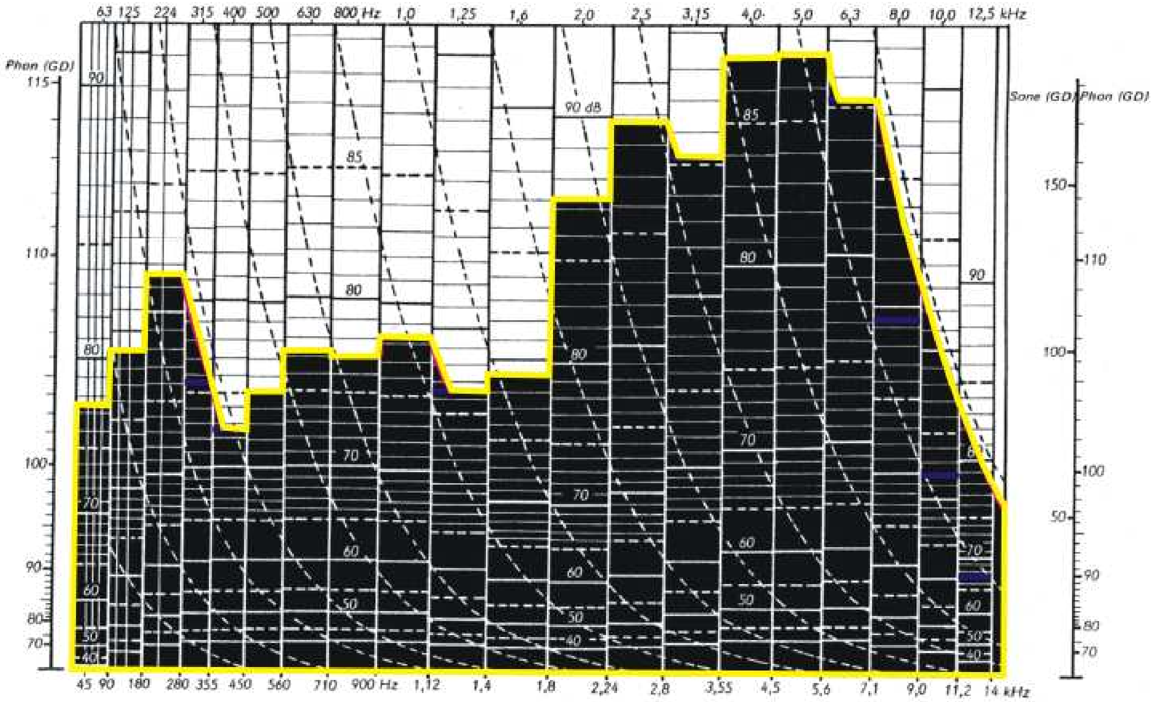



***Figure B.3:*** *Example of loudness calculating according to ISO 532B.*

This graphical procedure is rather tedious and a FORTRAN and BASIC program was presented in [30] which gives the exact values as those calcuted manually from the graphical procedure. Adapted from this program, is a MATLAB implementation provided by Aaron Hastings at Herrick Labs, Purdue University. This program was used to calculate the loudness level in sones of the stimuli.

APPENDIX C

# $R_{NONLIN}$ CURVE FITTING

As discussed in section 4.4 on page 43, the subjective ratings obtained from the listening test can be used and a curve fit to the data using equation C.1. In such a way, a predictor for the subjective perception of nonlinear distortion is obtained. Section 4.4 on page 43 obtained the overall predictor for all types of distortion. Below, individual predictors for each type of distortion and their correlation to the subjective perception of nonlinear distortion are obtained.

$$Predicted\ Rating = (\frac{a + bR^c_{nonlin}}{a + R^c_{nonlin}}) \tag{C.1}$$

Figure C.1 on the facing page shows the curve fitting obtained for all types of distortion and their respective parameter estimates for equation C.1.

From the curve fittings and their respective parameter estimates, plots of the the observed rating versus predicted ratings were obtained and are shown in figure C.2 along with their respective correlations.

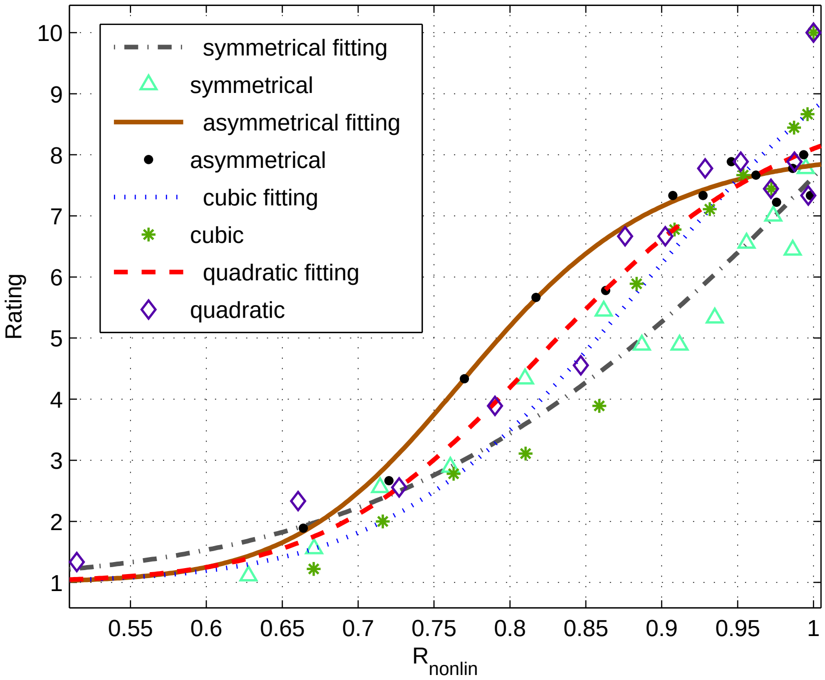
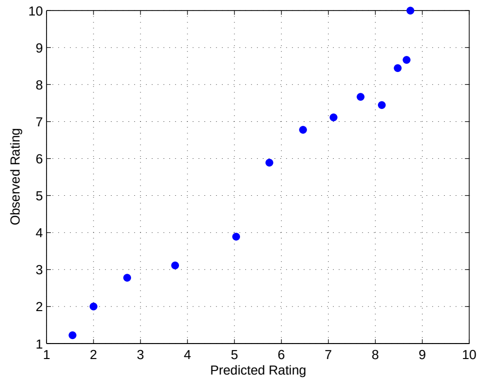
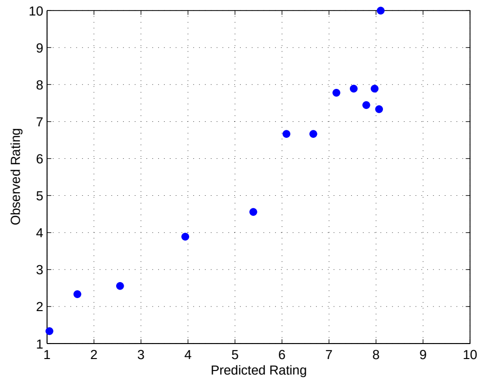


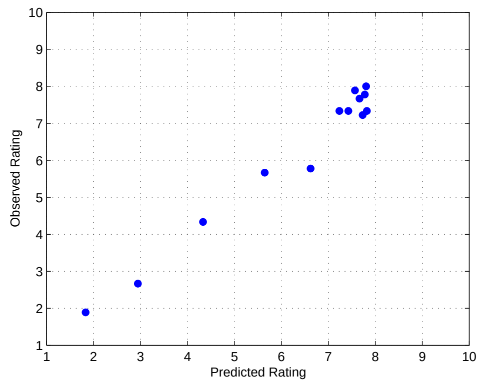

***Figure C.1:*** *Predicted rating curve fitting for different distortion types. [cubic: a = 0.3671, b = 11.59, c = 9.775; quadratic: a= 0.1505, b = 9.169, c = 10.48; asymmetrical: a = 0.03996, b = 8.101, c = 12.79; symmetrical: a = 2.201, b = 22.26, c = 5.624]*



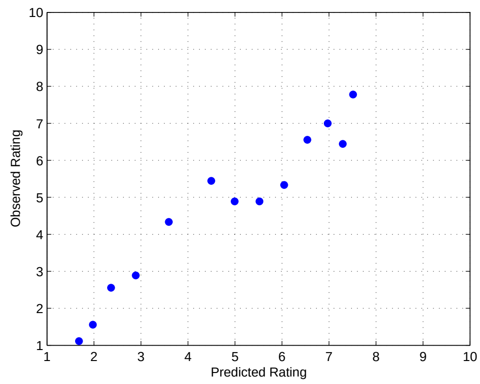

(a) Cubic Distortion, r = 0.9817




(b) Quadratic Distortion, r = 0.9642




(c) Asymmetrical Clipping, r = 0.9874




(d) Symmetrical Clipping, r = 0.9658

***Figure C.2:*** *Observed vs. Predicted Rating*

APPENDIX D

# LISTENING EXPERIMENTS

## D.1 Description of Listening Test Setup For Experiments 1 and 2

### Purpose

The purpose of experiment 1 was to collect subjective ratings of the amount of applied nonlinear distortion from a group of listeners. The subjective ratings will be used to check if the implemented nonlinear distortion metrics were done correctly by evaluating the correlation between these metrics and the collected subjective ratings.

Experiment 2 was performed in order to find a threshold in DS for two types of music signals and 4 types of nonlinear distortion systems. An interactive program which presented signals according to the transformed up-down procedure was utilized.

All subjects had hearing tests performed prior to proceeding with the above experiments. All subjects had hearing threshold above 20 dBHL.

### Equipment

| Name | Manufacturer | AAU Number |
|---|---|---|
| Harmonie | —- | —- |
| 01dB Measurement System | —- | —- |
| Desktop Computer Akulab13 | Dell | 37726 (monitor) 53118 (tower) |
| Headphones | Beyerdynamic DT 990 | 2036-1 |
| Measurement Mic. | B&K 4134 | 8130 |
| Headphone Amplifier | Behringer HA903 | 33239 |
| Mic. Preamplifier | G.R.A.S.S 26AK | 52665 |
| Artificial Ear | B&K | 07631 |
| Assorted Cables | N/A | N/A |

***Table D.1:*** *Equipment list for evaluating loudness and carrying out listening experiment.*

### Listening Experiment Description

Before presenting the stimuli to the subjects, the original undistorted signals were set to a level of 70 dBA as produced by the left cup of the headphones. To measure this level, the left cup of the headphones was placed over an artificial ear. The measurement (or recording) diagram can be found in A.2. The 01 dB system was used to record the undistorted music sample. The measurement acquisition was then performed over the duration of the signal. The signal volume was then varied until 70 dBA was achieved. A time recording was also saved from the measurement system which was used to evaluate the overall loudness level in sones. The loudness model implemented in accordance with the DIN 45631/ISO532B standard predicted a loudness of 22 sones from the measurement acquisition.

The listening evaluation in Experiment 1 was designed to be approximately 6 minutes in duration. Within each session the subject was presented with 50 stimuli. The nonlinear distortion systems described in Section 4.1 were used to apply varying amounts of distortion to a sample of guitar music. The evaluation was carried out in Cabin A at the Acoustic Laboratories of Aalborg University. Headphones were used to present the related stimuli to the subjects which were directly connected to the soundcard of a PC. The subjects then rated the amount of distortion from a scale from 1-10 where 1 corresponds to a completely distorted signal and a 10 refers to a clean undistorted signal. The subjects responded to each stimuli via a GUI implemented in MATLAB.

There were four sessions per subject in the listening evaluation used in Experiment 2. Each session lasted approximately 7-9 minutes depending on each subject. The evaluations were carried out in the same location as in Experiment 1 using the same sound reproduction system. During each threshold evaluation the subject was presented with two signals A and B. The subject was then asked to select the stimulus which sounded the most distorted.

## D.2 Instruction Sheet for Subjects

The following set of instruction were give to each of the subjects before beginning the listening evaluation:

### Listening Test Instructions for Experiment 1

The purpose of this listening test is to determine the perception of distortion. You will be presented with a series of sound samples, each with a varying amount of distortion. Your task will be to rate these samples based on how distorted the sound sample appears to be on a scale from 1 to 10, **where 10 means the sample represents a "clean, completely undistorted" sound sample and 1 represents a "very distorted" sound sample**.

The test will take approximately 15 minutes. The main test will be preceded with a brief training session where you will be presented with the two limits of the scale (i.e. a sample that would rated a 10 and a sample that would be rated a 1). Please use this training session as reference as to how to rate the sound samples during the rest of the test.

### Listening Test Instructions Experiment 2

The purpose of this listening test is to determine the point at which distortion becomes audible in music. The test is divided into 4 sections, each with a duration of approximately 8 minutes. After each section there will be the option to take a 5 minute break to walk around and have some cookies and coke.

During each of the sections, you will be presented with a pair of sound samples. Your task will be to determine which sample out of the two sounds distorted. This will not always be evident, and at times it will be hard to tell the two samples apart. However, there are no "right" or "wrong" answers in this test, and therefore you will be forced to make a choice between the two samples in order to continue, even if you are unsure.

Before beginning the experiment a brief audiometry (hearing test) will be performed to ensure that your hearing is suitable to perform the test.

Please take a second to fill out the questionnaire below. All information will remain confidential.

Name: ___________________________
Age: ___________________________
Gender: male:☐ female: ☐

Have you ever been diagnosed with a hearing impairment? yes:☐ no: ☐
If yes, please briefly describe the impairment:

_______________________________________________________________________________
_______________________________________________________________________________
_______________________________________________________________________________

Have you ever participated in listening test? yes:☐ no: ☐
If yes, please briefly describe the test:

_______________________________________________________________________________
_______________________________________________________________________________
_______________________________________________________________________________

Additional comments:

_______________________________________________________________________________
_______________________________________________________________________________
_______________________________________________________________________________