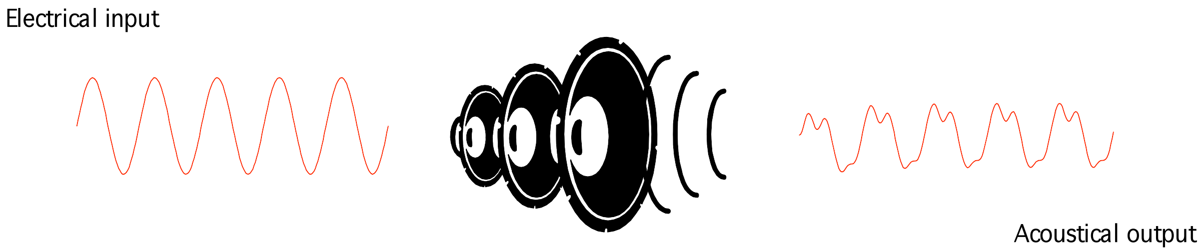
CHALMERS



Assessment of Nonlinearities in Loudspeakers

Volume dependent equalization

Master's Thesis in the Master's programme in Sound and Vibration

VIKTOR GUNNARSSON

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers Room Acoustics Group
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2010

Master's Thesis 2010:27

MASTER'S THESIS 2010:27

Assessment of Nonlinearities in Loudspeakers

Volume dependent equalization

VIKTOR GUNNARSSON

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Roomacoustics Group

CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2010

Assessment of Nonlinearities in Loudspeakers
Volume dependent equalization

Master's Thesis 2010:27

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Roomacoustics Group
Chalmers University of Technology
SE-41296 Göteborg
Sweden

Tel. +46-(0)31 772 1000

Reproservice / Department of Civil and Environmental Engineering
Göteborg, Sweden 2010

Assessment of Nonlinearities in Loudspeakers
Volume dependent equalization
Master's Thesis in the Master's programme in Sound and Vibration
VIKTOR GUNNARSSON
Department of Civil and Environmental Engineering
Division of Applied Acoustics
Roomacoustics Group
Chalmers University of Technology

Abstract

Digital room correction of loudspeakers has become popular over the latest years, yielding great gains in sound quality in applications ranging from home cinemas and cars to audiophile HiFi-systems. However, loudspeaker nonlinear distortion is often not taken into account by room correction products and this thesis investigates how distortion can be assessed. The theory of how distortion is produced in loudspeakers and the criteria that governs whether it is perceptible or not is reviewed. Two Matlab graphical user interfaces were developed, one for measuring distortion in loudspeakers and one implementing an algorithm that varies the low frequency cut-off of a loudspeaker depending on the volume level, thus increasing the bass extension possible with small loudspeakers. This method gives good subjective results with small loudspeakers.

Keywords: Loudspeaker equalization, Nonlinear control, Distortion perception, Digital Signal Processing

Contents

Acknowledgments

I would like to thank the people at Dirac Research who gave me the opportunity to do my thesis in this interesting subject. Special thanks to Fredrik Hagman and Elis Johansson who reviewed the report and to Adrian Bahne who was my supervisor at Dirac and to Mendel Kleiner who was my supervisor at Chalmers.

1 Introduction

Digital Room Correction (DRC) has become a popular technology in the last few years as a means to compensate for some of the "linear" distortion caused by loudspeakers and rooms. By *linear* distortion is meant, that the sound pressure response of the audio system is non-constant as a function of input signal frequency. This adds undesired coloration to the sound. DRC is basically very advanced linear equalization that takes into account how the hearing system interacts with the listening room and produces an "inverse coloration" that is applied to the input signal of the loudspeaker. DRC products exist for applications ranging from home cinemas or audiophile HiFi-systems to car audio systems.

The purpose of this master thesis is to investigate how *nonlinear* distortion[1] should be taken into account in DRC products. This is important when applying DRC to for example flat panel TV-sets or laptops which often have small, low-quality loudspeakers that produce a lot of nonlinear distortion at moderate sound levels. There is a risk when DRC is applied that frequency ranges where the loudspeaker locally produce high levels of distortion are boosted, resulting in unpleasant sound.

One of the goals with the thesis is to develop a software for measuring the amount of nonlinear distortion produced in a perceptually relevant way. This information can then be used to modify the DRC algorithm to ensure that disturbing distortion is not produced. In practice it is often the amount of bass fed to the speaker and how loud it plays that decides how much audible distortion it will produce. A second goal with this thesis is therefore to develop an algorithm that varies the amount of bass fed to the loudspeaker as a function of the playback level so that the best possible sound is obtained at each playback level.

The assessment of nonlinear distortion must start with an understanding of how it is produced, and the criteria that makes it perceptible to the auditory system. The thesis starts therefore with a brief introduction to nonlinear systems and the history of distortion measurements in audio (Chapter 2). Thereafter follows the psychoacoustical rules that governs the perception of distortion (Chapter 3). Modern metrics of distortion are then introduced, these have been developed in recent years since the classical metrics of distortion widely used in the audio industry fail to quantify distortion in a perceptually relevant way (Chapter 4).

[1] From this point, *distortion* will always refer to *nonlinear* distortion.

The basic mechanisms responsible for the generation of distortion in common dynamical loudspeakers are described in Chapter 5. Chapter 6 deals with the development of a distortion measurement tool in MATLAB and chapter 7 describes the implementation and evaluation of a level dependent DRC equalization.

2 The history of Distortion Measurements in Audio

This chapter introduces the standard distortion metrics THD and IMD. But before it is possible to delve into the history of distortion measurement, a background is needed on some basic theory about nonlinear systems.

2.1 Properties of Nonlinear Systems

Some basic properties and different types of nonlinear systems will be introduced in this section. The reader is assumed to be familiar with basic linear Signal & Systems theory.

A system can be defined in the mathematical sense by a rule that maps an excitation $x(t)$ to a response $y(t)$ as

$$y(t) = \mathbf{T}[x(t)].$$

The *operator* $\mathbf{T}[\cdot]$ could denote either a linear or a nonlinear system. Linear system theory is mostly concerned with LTI-systems and can be applied if $\mathbf{T}[\cdot]$ fulfills the principles of superposition (linearity) and if a certain input signal $x(t)$ yields the same output $y(t)$ regardless of *when* the input signal is applied (time invariance). These are familiar concepts to anyone who has experience with linear system theory.

The operator $\mathbf{T}[\cdot]$ is said to have *memory* if it is a function of the values of previous output and/or input signals $y(t)$ and $x(t)$. This is what gives a system its *dynamical* properties – frequency dependence in other words, if the system is analyzed in the frequency domain.

Nonlinear systems are often categorized into systems exhibiting *memoryless* nonlinearities, also called *static nonlinearities* and systems exhibiting nonlinearities with memory, also called *dynamical nonlinearities*. Static nonlinearities are much simpler to analyze mathematically since the operator $\mathbf{T}[\cdot]$ is then a simple one-dimensional function.

Figure 2.1 plots $\mathbf{T}[\cdot]$ for some static nonlinearities. The examples are cross-over distortion, which occurs in solid state amplifiers; clipping, which also occurs in amplifiers and also easily in digitally sampled systems; a 2nd order polynomial nonlinearity ($y(t) = x(t) - 0.2x(t)^2$) and a linear system with an amplification of 0.5. These systems affect sinusoidal input signals of all frequencies the same because of the memoryless property.

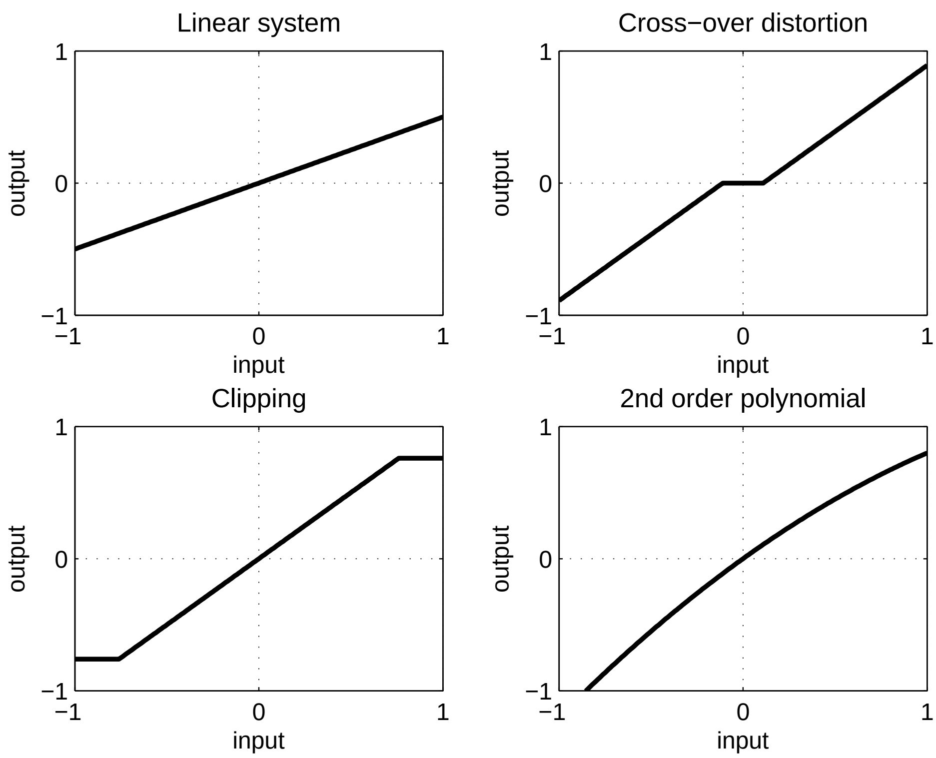


Figure 2.1: Examples of static nonlinearities.

2.2 Static Nonlinearity and Sinusoidal Stimulus

The case when the input signal to the nonlinear system is a sinusoid is interesting, because this is how the distortion of audio equipment is very often measured. A sinusoid is applied to the input, and by Fourier's theorem it is known that the output will be a periodic signal with the same fundamental frequency as the input sinusoid but with harmonics added to the signal. By comparing the magnitude of the harmonics with the fundamental tone in the output signal, the degree of nonlinearity of the device under test can be measured.

The assumption of a static nonlinearity is a highly idealized one when it comes to loudspeakers (which this thesis is mainly about) but it is nevertheless illustrative to derive the output signal to a sinusoidal stimuli. All input-output curves of static nonlinearities can be expanded in a Taylor-series to desired accuracy provided that **T[·]** is a smooth function so that the series converges:

$$\mathbf{T}[x(t)] = \sum_n a_n x^n(t)$$

The n-th power of x is called an n-th order nonlinearity. Consider, as an example, a second order polynomial nonlinearity like that shown in Figure 2.1:

$$y(t) = a_1 x(t) + a_2 x^2(t)$$

Let the input signal be $x = A\cos(\omega t) = \frac{A}{2}(e^{i\omega t} + e^{i\omega t})$. Then the output is

$$\begin{aligned} y(t) &= a_1 \frac{A(e^{i\omega t} + e^{-i\omega t})}{2} + a_2 \frac{A^2(e^{i\omega t} + e^{-i\omega t})^2}{4} \\ &= a_1 \frac{A(e^{i\omega t} + e^{-i\omega t})}{2} + a_2 \frac{A^2(e^{i2\omega t} + e^{-i2\omega t})^2}{4} + a_2 \frac{A^2}{2}. \end{aligned}$$

The output contains the input scaled by a_1, a harmonic of the input at 2ω scaled by $a_2 A^2/2$ and finally a DC offset of magnitude $a_2 A^2/2$. If this analysis is carried out for a polynomial nonlinearity of arbitrary order, the following conclusions can be made [Cze 01]:

- An even order nonlinearity of order n produces harmonics of order 2,4,6...n and a DC offset.
- An odd order nonlinearity of order n produces harmonics of order 3,5,7...n and no DC offset.

Nonlinear distortion can be defined to occur whenever a system adds frequency components to the output signal that were not present in the input signal. If the added signal components are independent of the instantaneous input signal amplitude, then these components are classified as noise. A distortion metric that has for long been very popular in measurements of both loudspeakers and electronics is **THD** – Total Harmonic Distortion. It is based on a single sinusoidal stimulus and is defined as the ratio of the total level of the harmonics to the level of the fundamental in the output signal:

$$THD = \sqrt{\frac{\sum_{k=2}^{N+1} H_k^2}{H_1^2}}$$

where H_k denotes the amplitude of the k:th harmonic, H_1 the fundamental and N the number of harmonics. THD is usually expressed in %. Sometimes the square-root is omitted and it then becomes a power ratio rather than an amplitude ratio.

2.3 Static Nonlinearity and Multitone Stimulus

Music, of course, does not contain only one frequency. A single sinusoid is not likely to excite the nonlinearities of a loudspeaker at all in the same way as a complex music signal, that has a rich spectrum of frequencies and completely different statistical properties than a sinusoid. A multitone signal that contains multiple sinusoids covering a wide part of the audible spectrum has been suggested as a testing signal by Czerwinski Et. Al. [Cze 01] among others.

Let us first examine what happens when two tones are input to a static nonlinearity of the second order. The output $y(t)$ for $x(t) = \sin(\omega_1 t) + \sin(\omega_2 t)$ will be

$$\begin{aligned} y(t) &= a_1(\sin\omega_1 t + \sin\omega_2 t) + a_2(\sin\omega_1 t + \sin\omega_2 t)^2 \\ &= a_1(\sin\omega_1 t + \sin\omega_2 t) + a_2(1 - 0.5\cos 2\omega_1 t - 0.5\cos 2\omega_2 t \\ &+ cos(\omega_1 - \omega 2)t - cos(\omega_1 + \omega_2)t). \end{aligned}$$

In addition to harmonics at $2\omega_1$ and $2\omega_2$, *intermodulation* products are generated at $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$. The amplitude of the intermodulation products is twice the amplitude of the harmonics.

The situation becomes much more complicated when more tones are added and the order of the nonlinearity is increased. The result is that a larger number of combinations of the input signals is obtained, a third order nonlinearity for example will also produce intermodulation products on the form $2\omega_i \pm \omega_j$. All intermodulation products are together called *intermodulation distortion* or IM-distortion/IMD.

Czerwinski Et. Al. [Cze 01] investigated how the number and energy of intermodulation products changes in relation to harmonic products for different number of initial tones and orders of nonlinearity. Tables 5-8 in [Cze 01] lists the number and energy of IM-products and harmonics for orders of nonlinearity from two to five and for multitone stimuli with one to twenty tones. The dominance of intermodulation products is striking, a fifth order nonlinearity for example with 20 initial tones produces close to half a million IM-products but only 40 harmonics of the initial tones. Some conclusions that are drawn in [Cze 01] are:

- An n:th order nonlinearity needs at least *n* input tones to reveal all possible intermodulation combinations of the input signals.
- The number and energy of the intermodulation products vastly exceeds that of the harmonic products when multiple input tones are present. The disproportion grows rapidly with the order of nonlinearity.
- High-order nonlinearity generates very low levels of harmonic distortion of this order, compared to the levels of IM-products and lower order harmonics.

It is clear that testing of distortion with only one tone, like the THD-metric employs, cannot reveal all information about the nonlinearities of a loudspeaker. Czerwinski Et. Al. argues that several authors have tried to find a universal relationship between the level of harmonic distortion and IM-distortion without success. It is quite possible that a loudspeaker may show apparent low harmonic distortion but produce high IM-distortion. Testing with one tone inherently gives the largest weight to low order nonlinearity, but the higher order nonlinearities are more responsible for production of IM-products and these are also more likely to be audible for psychoacoustical reasons, as will be explained in the next chapter.

A further point is that although the harmonic components are a symptom of nonlinearity inevitably giving IM-distortion, the harmonic components themselves may not even be perceived as distortion. Assuming music that contains instruments with already present overtones, the harmonic distortion will more likely alter the levels of the overtones, giving coloration, rather than being perceived as distortion. IM-components on the other hand lies in dissonance with the fundamental and overtones and are likely to cause a large audible degradation.

Two methods that have historically been used to measure IM-distortion using a two tone stimulus are the CCIF method and the SMPTE method. In short, the SMPTE method uses one tone fixed at a low frequency and a lower level high frequency tone (also called the *modulation method*). The second order IM-product is recorded and may be plotted over frequency if the higher frequency tone is swept over frequency. The CCIF method uses two closely spaced tones that are swept over frequency while the second order intermodulation product $\omega_1 - \omega_2$ is recorded (also called the *frequency difference method*). Neither a one-tone or a two-tone test gives complete information about the nonlinearities of a loudspeaker though, since as many tones as the highest order of nonlinearity present is required to excite all kinds of IM-products.

Another interesting aspect of a testing signal for measuring distortion is its statistical distribution of amplitude over time. The amplitude distributions of an example music track, a sinusoid and a multitone signal comprised of twenty logarithmically spaced tones from 100-10000 Hz are shown in Figure 2.2. Music typically has a Gaussian-like distribution which means that the waveform "spends more time" at low amplitudes than at high amplitudes. A sinusoid has quite the opposite distribution and spend much of its time near full amplitude which results in that testing signals based on a single sinusoid gives more weight to nonlinearities that occurs at high signal levels than at low levels. That is a bad property since music is more affected by nonlinearities occurring at low signal levels. An example of this is given in [Voi 06]. In an experiment, two signals were created from a sampled piece of music with an amplitude in the range ± 1. One was hard limited (clipped) to 50% signal amplitude, everything outside ± 0.5 was "chopped off". The other had all parts of the signal with an amplitude lower than 0.05 set to zero. The THD corresponding to the clipping operation was 22.6 % and the THD corresponding to the zero crossing operation was only 2.9 %. When listening to the music sequences, "the hard clipping produced only rare unpleasant, but tolerable effects" whereas, "the zero crossing produced intolerable deterioration of sound quality". The amplitude distribution of a multitone signal is similar to that of typical music and should therefore as a testing signal give a much better weighting between low and high level nonlinearities than a sinusoidal testing signal.

2.4 Dynamic Nonlinearity and Multitone Stimulus

The analysis of dynamical nonlinearities require much more sophisticated and intricate mathematical tools than simple static nonlinearities. There exists a large knowledge base on the analysis of nonlinear systems. This section will introduce, very briefly, the method of **Volterra series** since it is often used in the analysis of *weakly* nonlinear time-invariant systems such as loudspeakers. The assumption of weak nonlinearity means that higher order nonlinearities are negligible and that effects such as bifurcation, dead-zone effects, saturation and hysteresis do not occur. Volterra series are covered in detail in [Sch 89].

The Volterra series representation of a nonlinear system has a nice intuitive interpretation in that it can be seen as a Taylor series with memory. The form of the Volterra series is

$$\begin{aligned} y(t) = &\int_{-\infty}^{\infty} h_1(\tau)x(t-\tau)d\tau \\ &+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \\ &+ \int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} h_n(\tau_1,...,\tau_n)x(t-\tau_1)...x(t-\tau_n)d\tau_1...d\tau_n \end{aligned} \tag{2.1}$$

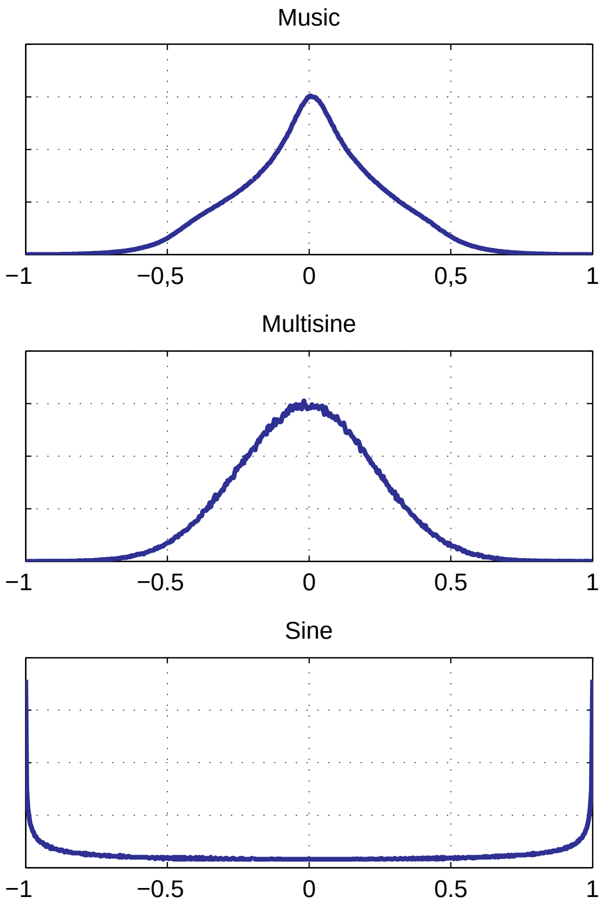


Figure 2.2: The amplitude distribution of an example music track, sinusoid & multitone signals.

where $x(t)$ is the input signal and $h_1...h_n$ are called the *Volterra kernels* of the system. It is seen that a Volterra series is an infinite series of multidimensional convolution integrals. The first order kernel is the linear impulse response of the system and the first term of the Volterra series is a conventional convolution integral. The n:th order kernel is n-dimensional and describes the n:th order nonlinearity of the system through an n-dimensional convolution. The higher order kernels could be called "higher order impulse responses".

Volterra series is a nonparametric, "black-box" model of a system. Applications could be system identification, modeling and simulation of a nonlinear system.

Equation (2.1) can alternatively be represented in the frequency domain. The n:th order kernel then transforms to an n:th order frequency response that depends on n frequency variables. The point here is that it takes n frequencies to be able to sample all nonlinear reactions of an n:th order nonlinearity; a one-tone measurement like THD merely samples a line in the multidimensional space of each kernel. One way to measure Volterra kernels is actually using multitone signals.

A drawback of the Volterra series approach is that an increasingly large amount of data is needed to sample higher order kernels, making the method impractical for systems that are dominated by high order nonlinearities.

2.5 Conclusions

Conclusions of this chapter are that a single or two-tone measurement cannot fully describe the nonlinear properties of loudspeakers. However, measurements of harmonic and intermodulation distortion are still valuable since there exists a large experience in interpreting these measurements and relating them to certain physical properties of loudspeakers. It is important to remember though that these kind of measurements do not reveal the whole truth about *subjective* aspects of distortion.

One common misconception that can be disregarded is that the presence of mostly second order harmonic distortion in a one-tone measurement would indicate that the distortion is subjectively benign. It is true that a second order harmonic in itself is not a big problem but a second harmonic is always a sign of the presence of second order nonlinearity – which will always yield intermodulation distortion in any signal more complex than a single sinusoid. Such distortion can never be subjectively benign because the IM-products are not harmonic to the fundamentals in the musical phrase. Nonlinearities of higher order yields increasingly more IM-distortion.

3 Basic Principles behind the Perception of Distortion

Describing the perception of nonlinear distortion is a vast subject that delves deep into the field of psychoacoustics. A good overview is given in ”Measurements and Perception of Nonlinear Distortion - Comparing Numbers and Sound Quality” by Alex Voishvillo [Voi 07]. A brief overview is given here as well.

The concept of masking helps to explain when distortion becomes audible. The distorted sound pressure signal from a loudspeaker can be logically divided into an undistorted part and a distorted part. If the distorted part is fully masked by the undistorted part, the distortion is inaudible.

The simplest case of masking is that of a pure tone that masks simultaneous tones close in frequency and lower in level. An illustration (though employing a narrow-band noise masker) is given in figure 3.1. All sound that falls below the masking threshold indicated in the figure is rendered inaudible by the masker. The shape of the masking curve in the figure is quite idealized for illustration purposes. The masking curve actually has a different shape for maskers of different frequency and level. The higher in level the masker is, the more asymmetric the masker curve becomes – frequencies above the masker center frequency are masked considerably more than frequencies below the masker center frequency. This is illustrated in figure 3.2. A more comprehensive discussion of masking can be found in a standard text on psychoacoustics such as *”Psychoacoustics: Facts and Models”* by Zwicker and Fastl [Zwi 06]. Here it also indicated that low frequency narrow-band maskers gives a masking curve that extends over a broader frequency range than high frequency maskers, seen on a logarithmic frequency scale.

The frequency asymmetry in the masking curve resulting from a single frequency component implies that distortion components that occur below the signal components in frequency are more audible than distortion components occurring above the signal components. In one-tone harmonic distortion tests, the harmonics occur above the signal component so the lowest harmonics are likely to be masked to a large degree. Thus the audibility of distortion in one-tone signals is not representative for the audibility of distortion in more complex signals.

Existing models of masking makes it possible to construct masking curves for arbitrary masker spectra. This is exploited in audio compression algorithms based on perceptual coding, where frequency bands that are likely to be masked are stored with

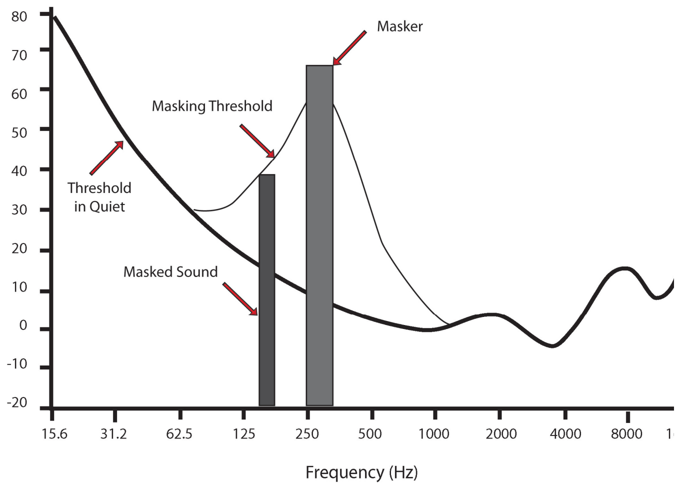


Figure 3.1: Example of simultaneous masking.
(Picture available at http://commons.wikimedia.org/wiki/File:Audio_Mask_Graph.jpg)

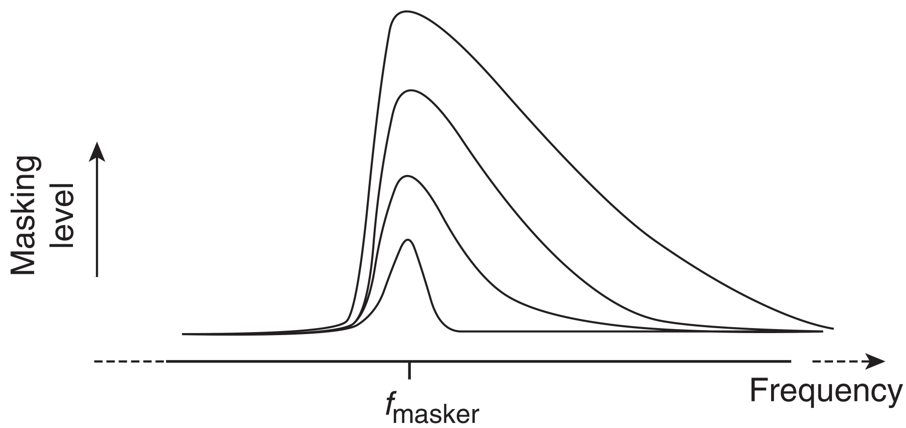


Figure 3.2: Illustration of the (idealized) level dependence of simultaneous masking (from [How 06]).

fewer bits of precision to save space. See e.g. [Spa 07].

Temporal aspects of the masker are also important. *Forward masking* refers to when a strong sound masks an immediately following weaker sound. Further, there is *Backward masking* where a weak sound followed by a stronger sound is masked by the stronger sound, although this effect not as strong as forward masking.

It can be concluded that distortion components that are separated far in frequency to the strongest signal components are more likely to be above the masking curve, and thus audible. High order nonlinearities produces distortion components that extends farther from the instantaneous signal components than low order nonlinearities and this explains why higher order nonlinearities are subjectively much worse than lower order nonlinearities.

The listening environment can have some influence on the audibility of distortion. Long reverberation and background noise can mask weak signal components for example and the frequency response of the acoustic path from the loudspeaker to the listener may accentuate or attenuate critical frequency bands. A loudspeaker may also show different distortion behavior in different directions so the direct and reflected sound may contain different levels of distortion.

The instantaneous signal spectrum will yield an instantaneous distortion spectrum and an instantaneous masking curve. Different pieces of music will reveal different kinds of nonlinearities to a different degree. Nonlinearities that occur at low levels are more audible than nonlinearities that only occur at high levels. An explanation to this may be that the statistical distribution of the signal amplitude of typical music is peaking around zero (see figure 2.2). High level signals are also better maskers. Voishvillo suggests further that a dense signal spectrum tends to make nonlinearities more audible. A dense signal spectrum excites a more dense distortion spectrum which could account for this.

4 Modern Metrics for Distortion Measurement

This chapter summarizes some of the methods for assessing the nonlinear properties of loudspeakers that are available. These methods can be coarsely divided into *Perceptual methods* which are based on properties of the auditory system, *Physical metrics* that use various signals for quantifying the physical nonlinearity of the loudspeaker and lastly *Nonlinear system identification and simulation* as a means of fully characterizing the properties of a loudspeaker which makes it possible to simulate any distortion measurement off-line.

4.1 Perceptual Methods

It has been recognized for a long time that the usual distortion metrics like harmonic and IM-distortion tests are not sufficient for describing the complex relationship between the audibility of distortion and the excitation of a complex nonlinear system (a loudspeaker) with a complex signal (music) [Voi 06]. Methods taking properties of the auditory system into account have been developed and some of them are presented here.

4.1.1 DS and Rnonlin

Tan, Moore, Zakharov, and Mattila have developed two of the more interesting distortion metrics based on psychoacoustical principles; the *DS* and R_{nonlin} metrics [Tan 04].

DS or *Distortion Score* is the simpler of the two metrics. A multitone signal is used as an excitation to the nonlinear system. Slightly simplified, the input and output signals to the nonlinear system are analyzed separately in 30 ms frames. Each frame is transformed to the frequency domain and split into 40 non-overlapping frequency bands emulating auditory filters. The absolute value of the average amplitude differences in each frequency band between the input and output signals are summed over all frequency bands. The obtained amplitude difference for each input/output pair of frames are then summed for all time frames. This gives the Distortion Score.

A good correlation between the Distortion Score and subjective ratings of simulated static nonlinearities was found. However, the correlation became much worse when the Distortion Score was applied to artificial dynamic nonlinearities or real nonlinear transducers. Therefore, another distortion metric – R_{nonlin} – was developed to overcome this limitation.

The R_{nonlin} metric uses a more sophisticated perceptual model and also applies real music and speech as a testing signal. The analysis in frames and frequency bands is similar to the DS-metric, but a filter is applied to the distorted and undistorted signals that emulate the frequency response of the outer and middle ear. And instead of comparing the spectrum of the input and output of the nonlinear system, the maximum cross-correlation between each distorted and nondistorted frame in each frequency band is calculated. The cross-correlations are summed over the frequency bands by weighting them by the level in each band, and then summed over all frames. The result is the R_{nonlin} metric. A low value for R_{nonlin} indicates a low correlation between the input and output of the nonlinear system and thus high distortion.

A high correlation between R_{nonlin} and subjective ratings of sound quality was found even for real nonlinear transducers, especially with music as a testing signal. But, the R_{nonlin} rating is not *linearly* correlated to the subjective ratings. That is, if R_{nonlin} is plotted versus subjective ratings, the regression line through the data points is not a straight line. A nonlinear curve fit is needed and this curve is different for different kinds of stimuli (music) and also different for different kinds of nonlinear systems. This makes this method presumably hard to apply in practice to loudspeakers since different loudspeakers show very different nonlinear behavior. But the R_{nonlin} metric is promising and this type of metric may lead the path to a psychoacoustically superior distortion metric that can replace the old IM/HD-distortion tests.

4.1.2 Perceptual Codec Evaluation Methods

Perceptual audio coders are algorithms for compressing audio without losing significant sound quality. The MP3 format is an example that utilizes this technology. Perceptual audio coding is a well researched area which involves very advanced psychoacoustics.

Methods to evaluate the sound quality of perceptual coders have been developed and these methods could possibly be applied to evaluate loudspeaker sound quality as well. One of the more straight-forward methods is based on the concept of *noise-to-mask ratio* [Spa 07]. Simply put, the masking curve (discussed in the previous chapter) obtained in a time frame by an excitation signal is calculated using a psychoacoustical model of the auditory system. This masking curve is compared to the level of an error signal obtained by taking the difference between the input and the output of the codec algorithm (cascading the codec's encode/decode algorithms). The level difference between the masking curve and the error signal is called the noise-to-mask ratio. The quality of the codec depends on how far below the masking curve the error signal is.

There exist some differences between the sound degradation of a loudspeaker and a codec though, so it is not obvious that this method can be directly applied to loudspeakers. A codec usually has a flat frequency response which a loudspeaker does not have. A codec also has quite different distortion mechanisms compared to loudspeakers, like pre-echoes and various artifacts from the processing of the signal in blocks.

An advanced codec evaluation method worth mentioning is the PEAQ (Perceptual Evaluation of Audio Quality) method. It was developed by a Task Group convened by the International Telecommunications Union (ITU). Several research groups prominent in the area contributed with suggestions for codec evaluation schemes and the most promising features of them all were combined into a single codec evaluation scheme, which became the ITU-R BS.1387 standard [ITUR 98]. In addition to the noise-to-mask ratio, PEAQ takes more factors into consideration such as the codec's effect on the excitation patterns of the basilar membrane, effects on temporal envelope, proportion of frames containing audible distortion and probability of detection.

A set of model output values (MOVs) are produced that describes different features of distortion audibility. A neural network is used to map these MOVs to a single figure, the objective difference grade (ODG). The mapping from MOVs to the ODG is optimized using data from subjective listening experiments so that the ODG accurately predicts the subjective ratings.

PEAQ could possibly be used to evaluate loudspeaker distortion. A distorted signal and a reference signal would be needed for the evaluation. The distorted signal would be the measured sound pressure from the loudspeaker and the reference signal the input signal to the loudspeaker. To only assess the nonlinear distortion of the loudspeaker, the reference signal would need to be equalized with the small-signal loudspeaker frequency response before application of PEAQ.

4.1.3 The GedLee Metric

Another metric worth mentioning here is the GedLee metric, developed by Geddes and Lee [Ged 03]. This metric is based on research trying to find a correlation between subjective sound quality and physical properties of static nonlinear functions. This resulted in the following metric that is a mathematical operation on the static nonlinearity $\mathbf{T}(x)$:

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

A high curvature of $\mathbf{T}(x)$ produces high order distortion and is therefore weighted higher by the second derivative term. The cosinus term weights nonlinearities that occur at low levels higher than nonlinearities that occurs at high levels.

Geddes and Lee showed that the metric gives a higher correlation to subjective ratings of sound quality than IM/harmonic distortion measurements. This proves further that high order nonlinearity and low-level nonlinearity are the most severe forms of nonlinearity.

The GedLee metric is however not directly applicable to dynamic nonlinearity, which characterizes loudspeakers.

4.2 Physical Metrics

Metrics that more directly measure the physical properties of a loudspeaker and do not directly take psychoacoustics into account could be called physical metrics. Examples would be the harmonic and intermodulation distortion tests as well as multitone testing, as described in chapter 2.

An extreme case of multitone testing would be to use combfiltered noise as a testing signal and measure the distortion in the notches. This approach was described in [Haw 05].

Another similar testing signal is noise with a single notch in its spectrum. The instantaneous distortion spectrum at the notch frequency can then be measured. By adjusting the frequency of the notch in real time, the distortion spectrum can be effectively measured at all frequencies. This approach was described by Farina under the name ”Silence Sweep” [Far 09].

A method mostly adopted for measuring distortion in hearing aids is to use a noise signal and measure the coherence spectrum.

Methods using a broad band excitation are objectively better than one- or two-tone measurements, as discussed in chapter 2, and they have been reported to correlate better with the subjective sound quality impression but so far no standards exist that relates the subjective sound quality impression to these kinds of measurements.

One possibility to get an indication of the subjective severity of distortion could be to measure the distortion spectrum, for example employing a silence sweep, and the masking curve of the signal spectrum as done in the perceptual codec evaluation methods. The level of the distortion at a particular frequency could then be compared to the level of the masking curve at that frequency, which would give important information about the subjective distortion level. The spectrum of the measurement signal, which is noise in the case of a silence sweep, could be shaped to resemble an average music spectrum. A problem though with this method is that music can take any spectral or temporal structure so it is hard to take all ”worst cases” into account with a single measurement spectrum.

An implementation of the silence sweep method was done in this thesis work, but it proved to be difficult to measure the distortion spectrum with any accuracy due to background noise which seems to be a problem with the silence sweep method. The notched noise method mentioned above could possibly be used instead with better results.

4.3 Nonlinear System Identification and Simulation

By creating a nonlinear digital model of a loudspeaker, any distortion measurement could be simulated off-line. It would also be possible to separate the linear and nonlinear response to the excitation. Theoretically, by simulating an auditory model of the masking properties of the ear, it would be possible to calculate the degree of audibility of distortion to an arbitrary excitation. Music would be possible to use as a testing signal which is a great advantage since music is what the loudspeaker after all is supposed to reproduce.

Nonlinear system identification of loudspeakers is a rather advanced field. Volterra series could be used as a nonlinear model of the loudspeaker, or a physical model based on physical parameters of the loudspeaker. There exists a considerable amount of literature on the subject, but it will not be discussed further here. A notable commercial product for loudspeaker nonlinear identification is the Klippel Analyzer [Kli 10].

5 Overview of Nonlinearities in Dynamical Loudspeakers

In this chapter the physical causes of nonlinearity in dynamical transducers will be briefly reviewed. It will also be discussed how distortion can be avoided or compensated using various techniques.

5.1 Physical Causes

The dynamical loudspeaker is a transducer from a voltage $u(t)$ to sound pressure $p(t)$. If the displacement $x(t)$ of the loudspeaker membrane is known then the sound pressure can be calculated with knowledge of the membrane area, see e.g. [Bla 00].

For the nonlinear analysis of the driver it is sufficient to analyze the relation between input voltage $u(t)$ and membrane displacement $x(t)$. This relation is given by two coupled nonlinear differential equations [Shu 97]:

$$
\begin{aligned}
\text{electric equation} \quad u(t) &= i(t)R_e(T_c) + \frac{dL_e(x(t),i(t))i(t)}{dt} + Bl(x(t),i(t))\frac{dx(t)}{dt} \\
\text{mechanic equation} \quad F(t) &= Bl(x(t),i(t))i(t) + \frac{i^2(t)}{2}\frac{dL_e(x(t),i(t))}{dx(t)} \\
&= M_{md}\frac{d^2x(t)}{dt^2} + R_{ms}\frac{dx(t)}{dt} + \frac{x(t)}{C_{ms}(x(t))} + Z_{mr}(x(t))
\end{aligned}
$$

The electric equation determines what the current $i(t)$ will be in the voice coil for a certain applied voltage $u(t)$. The mechanical equation determines the relationship between the current in the voice coil, the corresponding force $F(t)$ developed by the voice coil and the resulting displacement $x(t)$ of the loudspeaker cone.

In the mechanical equation it is assumed that the membrane moves as a rigid piston and that there are no acoustic modes in the air inside the box. This limits the validity of this model to low frequencies.

M_{md} is the mass of the moving parts of the driver. Bl, L_e, R_e, C_{ms}, Z_{mr} and R_{ms} are nonlinear parameters which are described below.

5.1.1 Force Factor

The *force factor* $Bl(x(t), i(t))$ describes how many Newtons of force are developed for each ampere of current that flows in the voice coil due to the interaction of the permanent (DC) magnetic field and the (AC) magnetic field created by the voice coil.

The force factor depends on the position of the voice coil in the magnetic gap because when the voice coil moves away from its resting position and out of the gap it moves away from the region of maximum magnetic field strength. The force factor also depends on the current in the voice coil. This is because the AC magnetic field, that is created by the voice coil, modulates the working point (the *permeability*) of the permanent magnet and thus modulates the permanent field. This is commonly referred to as flux modulation.

The force factor nonlinearity with excursion is typically stronger than the nonlinearity in current [Kli 06]. The force factor nonlinearity affects all frequencies as long as there is sufficient excursion to excite it.

5.1.2 Voice Coil Inductance

The inductance $L_e(x(t), i(t))$ of the voice coil is also nonlinear in voice coil position and current. This nonlinearity gives rise to a *magnetic reluctance force*, seen as an extra term in the mechanical equation, that acts on the voice coil.

The voice coil nonlinearity affects high frequencies more than low frequencies since the impedance due to inductance is proportional to frequency.

It should be noted that modeling the voice coil impedance as a resistance R_e in series with an inductance L_e as done in the electrical equation above is only valid for low frequencies. At high frequencies the current in the voice coil induces eddy currents in metal parts close to the voice coil. This decreases the inductance at high frequencies and leads to resistive losses [Van 89].

5.1.3 Voice Coil Resistance

The resistance $R_e(T_c)$ of the voice coil depends on the temperature T_c of the voice coil. As the voice coil gets warm, its resistance increases. This makes the loudspeaker draw less current and its sensitivity decreases, *compression* occurs. The temperature of the voice coil is continuously modulated by the musical program.

The modulation of R_e is too slow to create distortion at audible frequencies [Zuc 09], but the frequency response of the loudspeaker changes with R_e, mostly around the fundamental resonance frequency.

5.1.4 Suspension Compliance

$C_{ms}(x(t))$ is the compliance (inverse of spring stiffness) of the driver suspension. The driver suspension has its most pronounced nonlinearity in the cone excursion – the suspension gets stiffer as it is stretched away from its resting position. The nonlinear behavior of the suspension is quite complicated. It can show hysteresis behavior, and creep which means that it gets looser after some time of heavy excursion. It is typically slightly temperature dependent as well.

Nonlinearity in the suspension creates more distortion at low frequencies than at high frequencies. This is because above the fundamental resonance frequency of the loudspeaker, the force needed to accelerate the mass of the membrane is much larger than the force needed to stretch out the suspension.

5.1.5 Suspension Damping

The damping of the suspension, R_{ms}, depends mainly on temperature and it mainly affects frequencies around the resonance frequency of the box-driver system.

5.1.6 Radiation impedance

The term $Z_{mr}(x(t))$ in the mechanic equation represents a collection of terms that gives the force on the membrane due to the interaction with the acoustical surroundings. It includes the force on the membrane from the radiated sound pressure and the force from the pressure inside the box the loudspeaker is mounted in, so Z_{mr} depends on the type of box modeled.

Taking a closed box as an example, the air in the box acts as a spring in parallel with the driver suspension. The sound pressure inside the box can get very high. Air compresses nonlinearly at high pressures and air nonlinearity can be significant in especially small closed subwoofers with high excursion capability.

5.1.7 Other Sources of Distortion

Examples of other sources of distortion are nonlinear membrane stiffness, doppler distortion, port turbulence and propagation nonlinearity.

The membrane of a loudspeaker is typically not perfectly stiff. The stiffness of the membrane might be nonlinear and at high frequencies the membrane motion will be characterized by resonant behavior which might excite nonlinearities further.

In vented box loudspeakers, turbulence and compression from the port are large sources of distortion at high playback levels.

Doppler distortion arises when a low frequency tone is reproduced at a large excursion amplitude together with a high frequency tone. The frequency of the high frequency tone will then be slightly modulated by the large movements of the membrane.

And finally, nonlinearity in the propagation medium, the air itself, is important in some applications. When the sound pressure becomes very high the air compresses nonlinearly, as noted previously. In addition to small boxes with high excursion subwoofer drivers, very high sound pressure levels are found for example inside horn loudspeakers close to the horn throat. Also, distortion can develop over distance when high level sound propagates through the air.

5.2 Avoiding Distortion

It can be concluded from the discussion above that the cone excursion is a very important factor regarding how strongly the nonlinearities in a loudspeaker driver are excited. The excursion needs thus to be kept low. Low frequencies cause the most excursion so in loudspeaker systems this is often taken care of by using multi-way speakers with dedicated bass drivers. Splitting the frequency range decreases excursion in the higher frequency drivers and prevents intermodulation distortion between low and high frequencies.

Another way to decrease distortion is to feed the loudspeaker using a high impedance source. A current source linearizes the electric equation for the loudspeaker, making the current flowing through the voice coil distortion free.

5.2.1 Techniques for Compensation of Distortion

Many different techniques have been developed to try to reduce the distortion produced by an existing nonlinear loudspeaker using various kinds of control topologies. The success using negative feedback for reducing distortion in amplifiers has encouraged many attempts to apply this technology to loudspeakers. The feedback signal can come from a microphone sensing pressure, an accelerometer on the cone sensing the cone acceleration or some more elaborate scheme using for example extra turns on the voice coil sensing the voice coil velocity, or some kind of current feedback using a voice coil current sensor.

Although some feedback systems have showed a good distortion reduction, most feedback systems have had limited commercial success. This might be due to that the increased complexity and cost of the system could simply be replaced by a higher cost, more linear transducer.

Another approach is to use a digital *feed forward* controller. The audio signal is then *predistorted* in a way such that the distortion in the sound pressure output is reduced. This approach is potentially less complex and cheaper than the feedback approach. A drawback is that since the controller must be based on some kind of model of the transducer, its performance will deteriorate over time as the driver age and its parameters change. Additionally there is always some spread in parameters between different production samples. These sources of error can be remedied to some extent by making the controller adaptive using some kind of sensor that gives information about the current state of the driver. For more information about distortion reduction and loudspeaker nonlinearities, see [Shu 97], [Bri 02], [Ped 08], and [Kli 06].

6 The Development of a Matlab Distortion Evaluation Tool

This chapter presents the different types of distortion measurements that were incorporated in a Matlab user interface. These are the *Logarithmic sine sweep*, *Multitone test*, *Silence sweep* and *Compression test*. The chapter is concluded by a discussion about the goal of finding a perceptually relevant metric for distortion in loudspeakers.

6.1 Logarithmic Sine Sweep

This is a method of simultaneously measuring the impulse response and harmonic distortion of a device using a sine sweep. The method was first described in detail by Angelo Farina [Far 00], and more recent papers by Farina [Far 07] and Swen Müller [Mül 04] elaborate further on the topic.

Different types of noise signals have previously been very common as stimuli in acoustical measurements. Logarithmic sine sweeps have however several advantages. The main advantage is that the nonlinear part of the measured impulse response can be separated from the linear part, thus giving a higher signal-to-noise ratio if the loudspeaker that plays the excitation signal is slightly nonlinear. Another advantage is that logsweeps are not very sensitive to time-variance in the measured system whereas measurements with noise signals, where averaging over several blocks of data is often employed, require that the system can be considered to be time-invariant. For a more comprehensive discussion of the pro's and con's of sine sweeps, see Müller's paper.

The complete procedure implemented in Matlab for taking measurements with logsweeps will now be detailed. A simple block diagram representation of the signal processing involved is shown in figure 6.1. The first step is to generate the sweep. Logarithmic sine sweeps actually employ a sine sweep where the instantaneous frequency is increased exponentially over time, so the name 'Logarithmic sine sweep' is a bit misleading. The equation that describes a logsweep that starts at frequency ω_1 and ends at frequency ω_2 given in [Far 00] is:

$$x(t) = \sin\left(\omega_1 T \frac{\exp\left(\frac{t}{T}\log_2\left(\frac{\omega_1}{\omega_2}\right)\right)}{\log_2\left(\frac{\omega_1}{\omega_2}\right)}\right)$$

where T is the duration of the sweep in seconds.

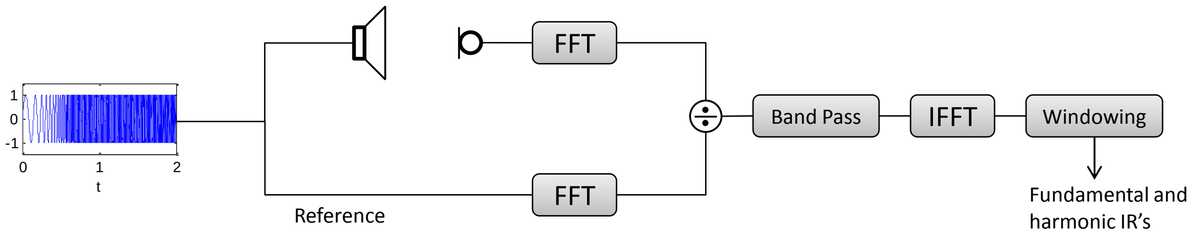


Figure 6.1: Impulse response measurement with logsweeps.

The sweep is played back through the system under test, and the result recorded. The next step is to transform the recording to the frequency domain with an FFT operation. The recorded spectrum is then compared to a reference spectrum by division in the frequency domain, as shown in figure 6.1. This gives the frequency response of the measured system.

The reference sweep may be either the same as the original sweep $x(t)$, or a loopback recording of $x(t)$ where one output channel of the sound card is connected directly to an input channel. If a loopback recording is used for the reference sweep, the measured frequency response is corrected for any latency or frequency response anomalies added by the sound card.

The next block in figure 6.1 is a band-pass filter. The reason why this is needed is understood if we look at an example measurement. A nonlinear and noisy system with a slight bandpass characteristic was simulated. The result of a logsweep measurement is shown in figure 6.2. In this example, a sampling frequency of 44.1 kHz and a logsweep with start frequency 100 Hz and end frequency 10 kHz is used. Often the sweep is chosen to cover the whole frequency spectrum but in some applications it is desired to energize the system under test with only part of the spectrum, hence these start and stop frequencies were chosen for illustrative purposes.

The upper left plot shows the recorded logsweep spectrum, and as can be seen it is dominated by recording noise above and below the border frequencies of the logsweep. The upper right plot shows the spectrum of the logsweep and the inverse of the reference spectrum. The inverse reference spectrum shows a large gain below and above the border frequencies of the sweep so when it is multiplied with the recorded spectrum, the resulting spectrum shown in the lower left plot contains a lot of noise above and below the sweep passband, and this noise of course also contaminates the corresponding impulse response. This is why a bandpass filter is needed and the resulting frequency response with the bandpass filter applied is shown in the lower right plot.

When a logsweep signal is calculated, a fade in/out is often applied to the beginning and end of the signal. This avoids transients (clicks) when playing the sweep. This

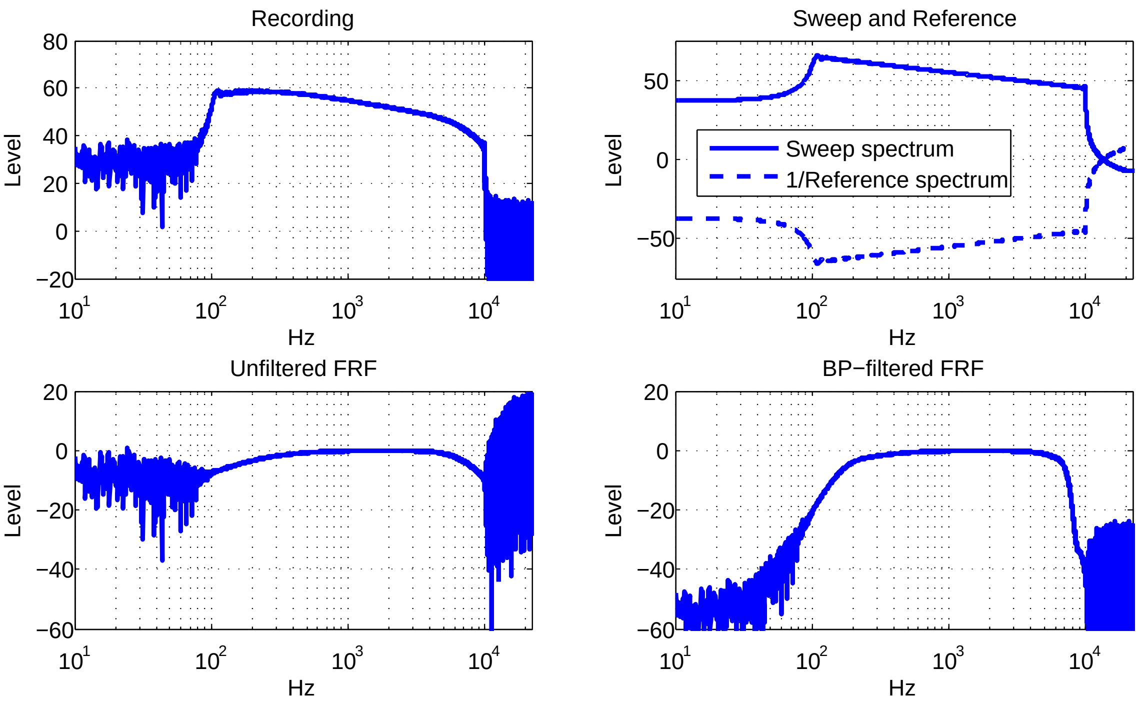


Figure 6.2: Example logsweep measurement. *Upper left*: recorded signal spectrum. *Upper right*: logsweep and inverse reference spectra. *Lower left*: recorded spectrum multiplied with inverse reference spectrum. *Lower right*: as lower left, after BP-filter.

was however not done in the current implementation because the inverse reference spectrum then gets a very large gain outside the sweep passband, and the resulting noise amplification outside the passband is difficult to attenuate with a bandpass filter.

In the current implementation, desired lower and upper frequencies $f_{low,desired}$ and $f_{high,desired}$ are set for the measured frequency range. The sweep start and stop frequencies f_{start}, f_{stop} and the bandpass filter corner frequencies f_{hp}, f_{lp} are then calculated to be

$$\begin{aligned} f_{start} &= f_{low,desired}/2, \\ f_{stop} &= \min(2 * f_{high,desired},\ fs/2), \\ f_{lp} &= \min(1.5 * f_{high,desired},\ 0.97 * fs/2), \\ f_{hp} &= f_{low,desired}/1.5. \end{aligned}$$

The filters used for the bandpass filter are a high-pass and a low-pass FIR-filter in cascade. The filters used in the example are shown in figure 6.3. Linear phase FIR-filters were chosen to avoid distorting the phase of the measured responses, as correct phase information is essential in some applications like loudspeaker cross-over design. The FIR-filters are designed using the window method, the high-pass filter uses a Bohman window and the low-pass filter uses a triangular window. These windows were chosen in order to avoid nulls in the filter stop band, since that gives a phase jump of 180 degrees and it was wished to keep the phase information accurate in a frequency range as large as possible. The filter orders N_{hp} and N_{lp} were after some experimentation chosen to be approximately

$$\begin{aligned} N_{hp} &= 3 * fs/f_{hp}, \\ N_{lp} &= 500 * \sqrt{100/f_{lp}} \end{aligned}$$

in order to keep the filter slope approximately equal regardless of what corner frequencies are chosen for the filters. These formulas might need to be fine tuned to fit a particular application, and N_{lp} should ideally depend on fs as well. It was optimized for a sampling frequency of around 44.1 kHz. The delay that is introduced by the FIR filters ($N_{hp}/2 + N_{lp}/2$ samples) must be removed from the measured responses.

Referring again to figure 6.1, after the bandpass filter the result is transformed back to the time domain. This yields the impulse response of the system. The fine thing about exponential sine sweeps is that the impulse response is composed of a main impulse, that is the fundamental impulse response, and several smaller impulses that comes before the main impulse. These are the IR:s of the harmonic distortion products. The one closest to the main impulse is the 2:nd harmonic IR, the one before that is the 3:rd

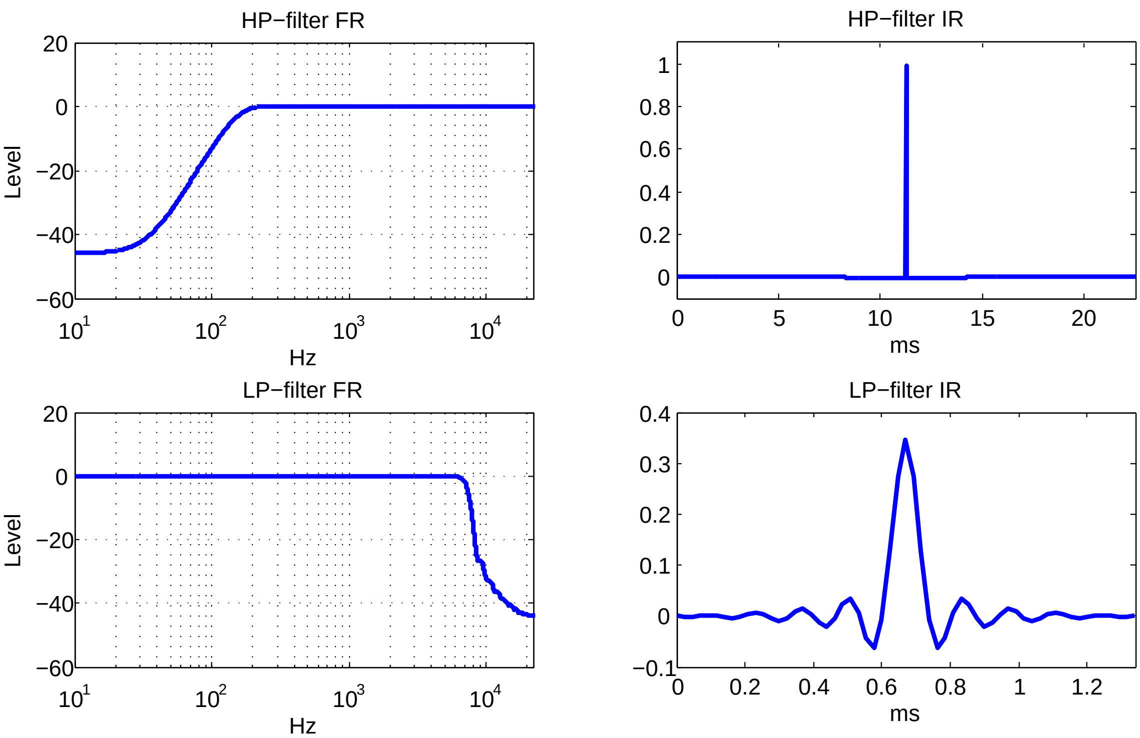


Figure 6.3: Magnitude and impulse responses of the high-pass and low-pass filters used in the example logsweep measurement.

harmonic IR and so on. Farina [Far 00] gives a formula for the time advance of the N:th harmonic IR in relation to the linear IR so that they can be individually windowed out:

$$\Delta t = T * \frac{\log_2 N}{\log_2 f_{stop}/f_{start}}.$$

Windows for the linear and harmonic impulse responses can be set in the Matlab Gui. The windows used in the example are shown in figure 6.4. The next figure, number 6.5, shows the calculated cut-points for the impulses in the top right plot and the top left plot shows the frequency responses of the fundamental and harmonics.

Note that the N:th harmonic frequency response has been moved down on the frequency axis a factor of N, this is because normally you want to compare the level of the harmonics with the level of the frequencies that causes them. So instead of having to compare the magnitude of the fundamental at frequency f with the resulting harmonics at frequencies $2f$, $3f$, ..., the fundamental and harmonics can be compared directly at frequency f in the plot. The lower left plot in figure 6.5 shows the amplitudes of the harmonics in percent relative to the fundamental.

The lower right plot in figure 6.5 shows an estimation of the signal+noise to noise ratio of each harmonic that was obtained in the measurement. This is calculated by recording a short period of background noise before the sweep signal is played. It is recorded before and not after so that the background noise recording is not contaminated by reverberation from the sweep. The recorded background noise is filtered by the inverse reference spectrum and the bandpass filter so that its spectrum becomes comparable to the spectrum of the impulse responses. The signal to noise ratio must be compensated individually for each harmonic for the fact that different window sizes are used for the different impulses and the background noise. For example when taking an FFT of a noise signal a doubling of the window size gives a magnitude response in average 3dB higher. Since the impulses are windowed with Tukey windows in this case and the background noise with a rectangular window, the window energy correction level L_{corr} decibels for the k:th harmonic becomes:

$$L_{corr,k} = 10 * \log_{10} \left(\frac{\sum_n W_k[n]}{\min(N_{bgNoise}, N_{fft})} \right)$$

where $W_k[n]$ is the window for the k:th harmonic, $N_{bgNoise}$ is the number of samples of background noise recorded, N_{fft} is the FFT size used to transform the background noise to the frequency domain (it's included because it may not be equal to $N_{bgNoise}$).

A comparison between the true frequency response and the measured linear response using the logsweep method is shown for the simulated example system in figure 6.6. The desired measurement range is here set to 200 Hz to 5 kHz. The slight overestimation of the magnitude inside the measurement range is due to that distortion products were excluded from the 'true' response and the simulated nonlinear system adds

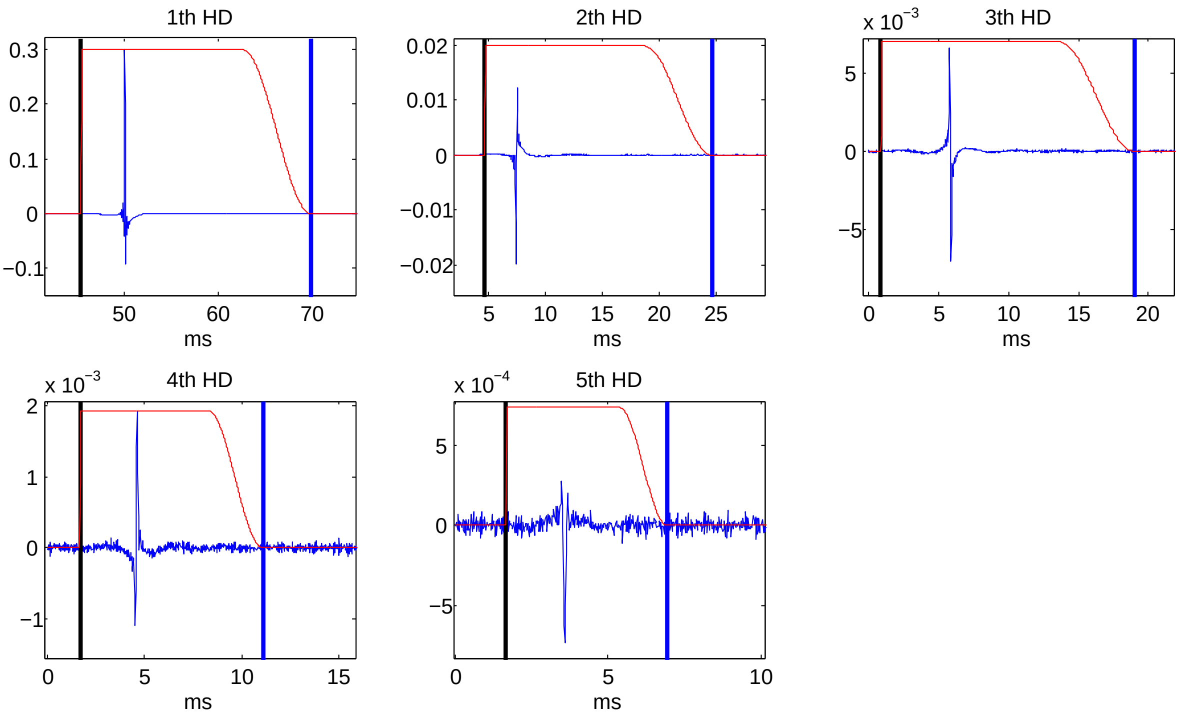


Figure 6.4: The windows used on the linear and harmonic IR's in the example logsweep measurement.

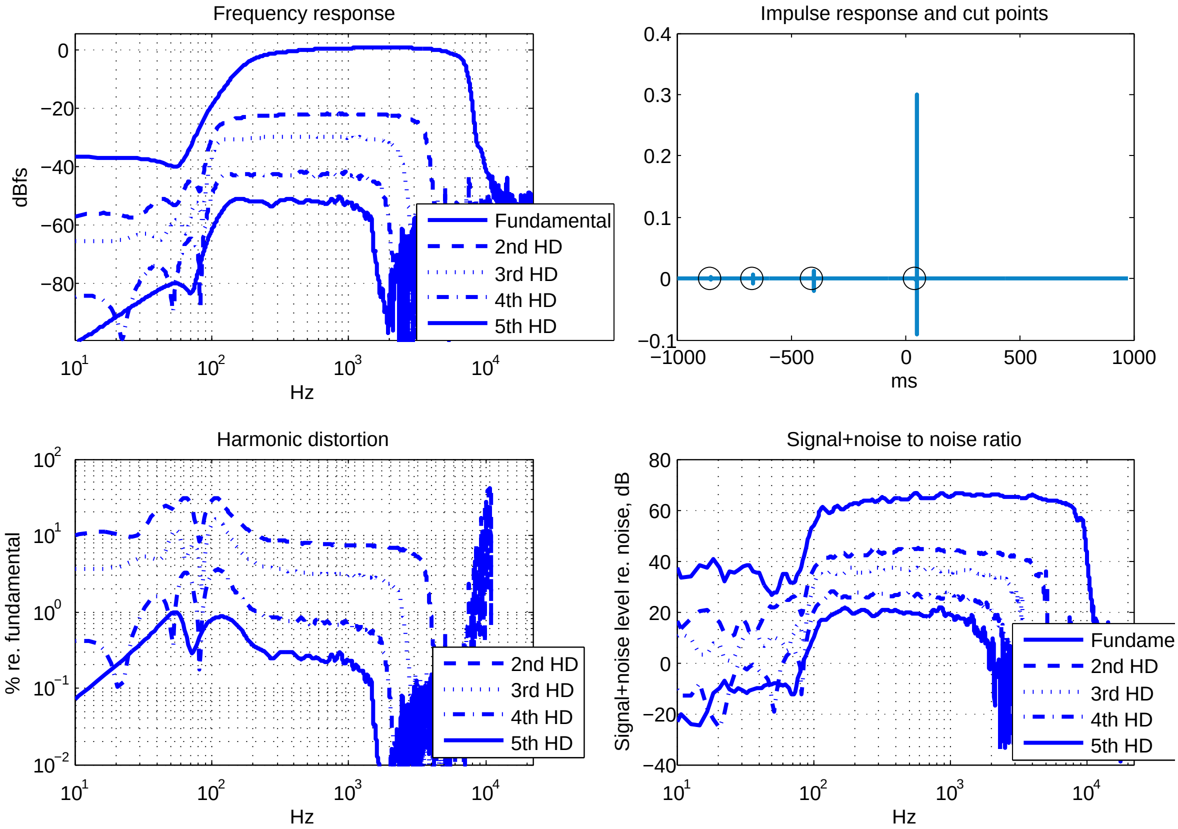


Figure 6.5: Distortion and signal to noise ratio estimation for the example logsweep measurement. *Upper left*: system magnitude response and harmonic distortion magnitude. *Upper right*: Cut points for the harmonic impulses indicated with circles. *Lower left*: harmonic distortion in % re fundamental. *Lower right*: estimated signal+noise to noise level for the linear and harmonic frequency responses.

some distortion products at the fundamental. The phase response shows a good match inside the measurement range.

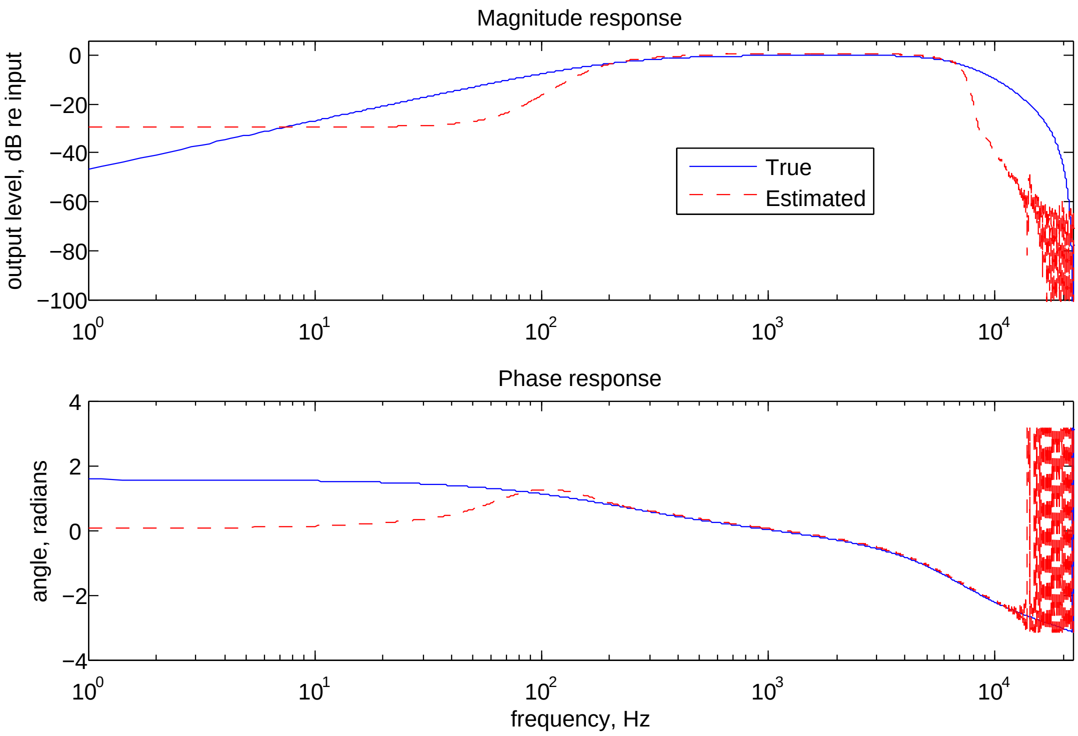


Figure 6.6: Comparison between estimated and true frequency response of the example system.

6.2 Multitone Test

The benefits of using a broad band excitation signal for distortion measurements were outlined in chapter 2. In a multitone test, the excitation signal consists of a number of tones spread out in frequency. The signal is fed to the device under test and recorded. The recording is then analyzed in the frequency domain and any signal components that lies between the original tones is either distortion, or noise.

A multitone test gives a quick overview of the linearity of a loudspeaker, a high signal to noise ratio is easily achieved and the presence and magnitude of deleterious high order intermodulation distortion is revealed.

The properties of the multitone signal depends a lot on how it is constructed, this is dealt with in [Ale 01]. The frequencies of the tones should *not* be linearly spread out over frequency, because in that case many of the distortion products will overlap with the signal. A logarithmic spacing of the tones is a good starting point.

The *crest factor* of the resulting signal – the peak to average ratio – should be as low as possible so that maximum energy can be fed to the system under test, and maximum signal to noise ratio obtained. Minimization of the crest factor can be done by optimizing the starting phase of the individual tones. A good starting point is to give the phase for each tone a random distribution. Further optimization is possible, but that was not attempted in the current implementation.

When generating the multitone signal, a block size for the signal is first chosen. This should be a power of two in length so that an FFT can be applied effectively on the recorded signal. It is very important that the frequencies of the tones are chosen so that an integer number of cycles of each tone fits into the block, otherwise there will be leakage across the DFT bins which will mask the distortion in the recording. For a block size of N_{block} this gives the possible frequencies $\omega(n)$:

$$\omega(n) = \frac{2\pi * fs * n}{N_{block}}, \quad n = 1 \dots N_{block}/2$$

where fs is the sampling frequency.

In practice, multiple blocks are always played back and recorded. Since the blocks are periodic this takes away the need for exact synchronization of the recording to the playback which usually includes an unknown delay. Averaging over several blocks also increases the signal to noise ratio (+3 dB for every doubling of the number of blocks averaged), although since loudspeakers are slightly time-variant the averaging time should not be excessively long.

There is a big benefit in making the FFT block size for the analysis twice as large as the synthesis block size N_{block}. By doing this, the even DFT bins will contain the signal, distortion and noise, but the odd bins will theoretically only contain noise. This phenomenon can be understood by considering the distortion products that a static nonlinearity would produce, as discussed in chapter 2. The distortion components will only appear at frequencies that are combinations of the signal frequencies involving addition, subtraction and multiples of the signal frequencies. A DFT of size N_{block} covers all the possible frequencies so a FFT buffer twice as large can be used to simultaneously measure the noise in the system.

An example measurement is shown in figure 6.7 that displays the signal, noise and distortion components separately. The magnitude of the frequency response is normalized by $2/N_{fft}$ so that the actual amplitudes of the sinusoidal components in the signal are displayed. The y-axis scale of the figure is thus in dB re. full scale (dBfs) where 0 dB represents full digital scale.

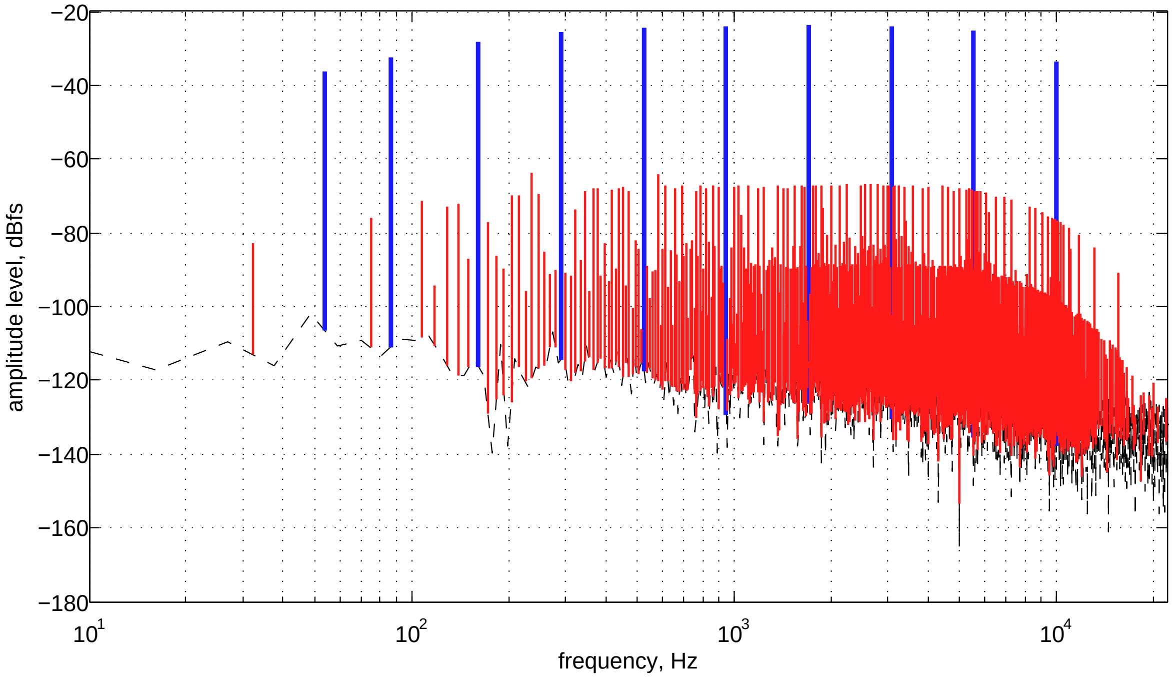


Figure 6.7: Example multitone measurement. *Thick lines:* signal. *Thin lines:* distortion. *Lower curve:* noise floor.

6.3 Silence Sweep

The silence sweep is a method which was initially considered very interesting for this work, but it turned out to be difficult to apply in practice due to problems with background noise. The idea is however conceptually interesting so it will shortly be described here.

The silence sweep was described in a paper by Farina [Far 09]. The principles behind the silence sweep are demonstrated by figure 6.8 that shows a series of spectrograms (a spectrogram is a plot of frequency contents as a function of time).

A signal containing white noise is first created, with a gap of silence in the middle (plot 1). This signal is convolved with an exponential sine sweep (plot 2). The sweep has a group delay that increases with frequency and the result is that the silence gap is swept over frequency in the noise signal. The spectrum of the noise signal also becomes pink, which is often preferred over a white spectrum in the testing of loudspeakers since it closer approximates the spectrum of music and also the background noise. The final measurement signal is shown in plot 3, which has been cut in the beginning and the end and had a gap of silence added to it in the beginning so that the spectrum of the background noise can be estimated from the recording.

The swept silence signal is played back by the loudspeaker system under test. In the recorded signal (plot 4, recording of background noise cut out) the silence sweep will at each time contain a sample of the underlying distortion spectrum that the excitation signal gives rise to in the loudspeaker, in the frequency region that the silence is currently swept over.

The recorded signal is next convolved by an inverse exponential sweep (plot 5). This restores the silence gap in the middle of the noise sequence but now it contains the sampled distortion spectrum in the form of noise. This can be plotted together with the spectrum of the excitation noise signal.

The Matlab Gui contains an implementation of the silence sweep. Parameters that are set are the length of the silence sweep T_{sweep} [s], the width of the swept silence gap in the frequency domain given in octaves $N_{oct,gap}$, and the estimated time it takes for the reverberation to drop below the noise floor in the acoustical space where the measurement is carried out T_r [s].

The larger that T_{sweep} and $N_{oct,gap}$ are chosen, the longer the final gap containing the distortion (as seen in figure 6.8, plot 5) will be and the better the frequency resolution of the corresponding distortion spectrum will be. There is a trade-off between accuracy and measurement time involved here, however, because $N_{oct,gap}$ should ideally be as small as possible to sample the true distortion spectrum that would result if the noise signal would cover the whole spectrum simultaneously without the swept gap.

The formula for calculating the length of the gap containing the distortion is:

$$T_{gap} = T_{sweep} * \frac{N_{oct,gap}}{N_{oct,sweep}}$$

where T_{sweep} and $N_{oct,gap}$ are set by the user and $N_{oct,sweep}$ is the number of octaves covered by the logsweep used in the construction of the silence sweep signal, given by

$$N_{oct,sweep} = \frac{\log_2(f_u/f_l)}{\log_2(2)}$$

where f_l and f_u are the start and end frequencies of the logsweep.

The gap that contains the distortion will also contain reverberation from the measurement space in the beginning of the gap. The beginning of the gap thus needs to be cut away before the distortion spectrum can be estimated, and that's why the user sets the parameter T_r, the estimated reverberation time.

An example measurement of a laptop is shown in figure 6.9 that shows the spectrum of the noise excitation signal, the distortion spectrum and the background noise spectrum. Even though the laptop was operated at the maximum volume setting the distortion spectrum is hardly above the noise floor. This makes this method difficult to use in practice. In a test case with a simulated noise free but nonlinear system, the method gave good results close to the true distortion spectrum when $N_{oct,gap}$ was set in the order of 0.1 and the sweep length at least 30 seconds.

6.4 Compression Test

A test method was implemented to investigate the 'dynamic linearity' of a loudspeaker by measuring the compression taking place at high levels relative to low levels. Compression in loudspeakers was described in chapter 5.

The frequency response of the loudspeaker is measured at a number of different levels set by the user. The frequency response measurements are done with the logsweep method described above. The nonlinear part of the impulse responses is discarded as only the compression of the fundamental is of interest.

Before each frequency response measurement, a period of pink noise is played, which length is set by the user. This is done to increase the temperature of the voice coil to working conditions. The user selects the number of measurements and the corresponding levels of the noise in dB re. full output scale. The level of the logsweeps is normalized to the same RMS level as the pink noise sequences since the noise and logsweep signals have quite different peak to RMS ratio.

An example of a compression test done on a laptop for two playback levels with 34 dB difference is shown in figure 6.10. The left plot shows the recorded frequency responses and background noise level. The right plot shows the compression of the high

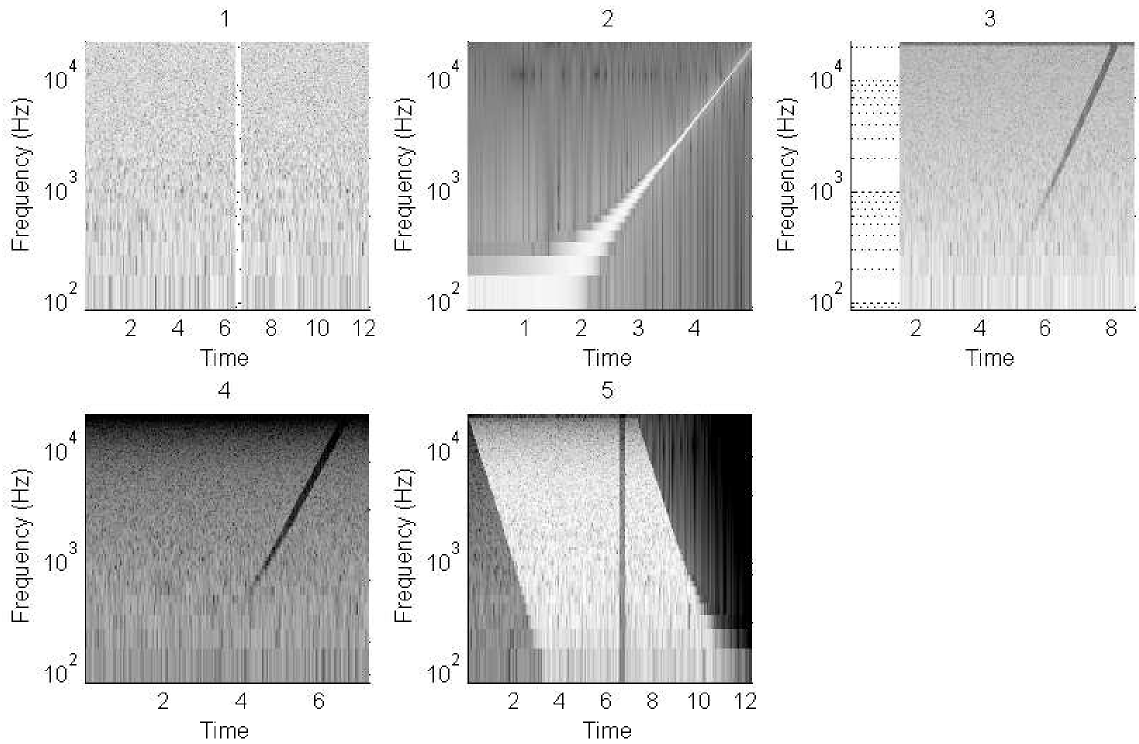


Figure 6.8: Example of silence sweep creation and measurement. *Plot 1:* A noise signal with a gap of silence in the middle. *Plot 2:* An exponential sine sweep. *Plot 3:* Signals in plot 1 and 2 convolved, result trimmed in the end and beginning, and silence added in the beginning for measurement of background noise. *Plot 4:* Signal in plot 3 played through a loudspeaker (silence in beginning excluded). *Plot 5:* Signal in plot 4 convolved with an inverse exponential sine sweep.

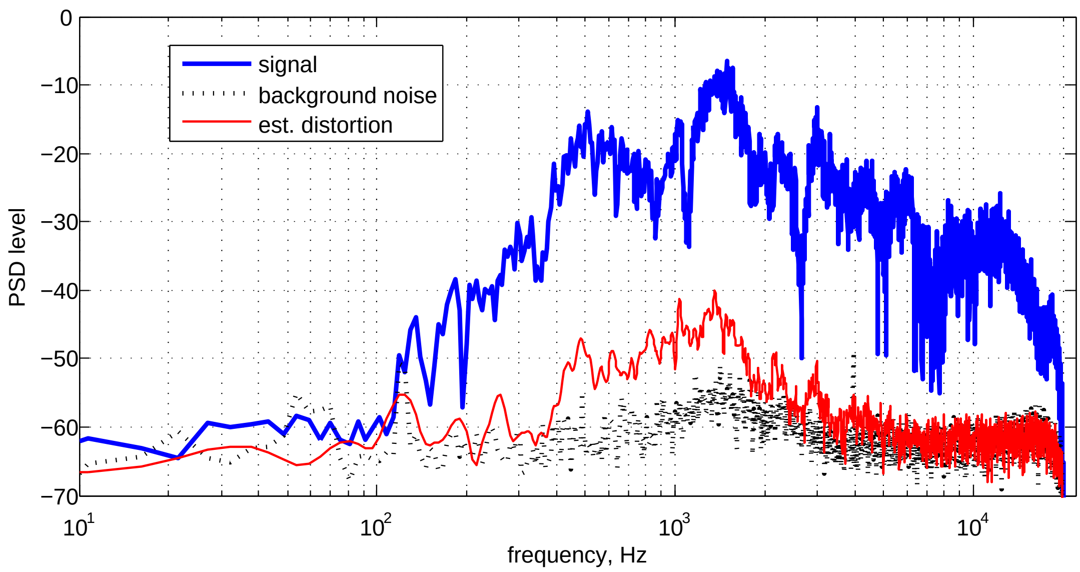


Figure 6.9: Example of a silence sweep measurement done on a laptop with maximum volume.

level frequency response relative to the low level frequency response. The sharp peaks seen in this plot is not actual compression, but due to a small shift in frequency for some of the notches in the frequency response between the two measurements. This could depend on slightly varying air temperature etc. Some compression can be seen especially around 1-2 kHz and below about 250 Hz. The compression reaches about 4 dB just under 200 Hz. For frequencies much below 200 Hz where the output of the small laptop loudspeakers is low, the background noise makes the measurement unreliable and the compression is overestimated.

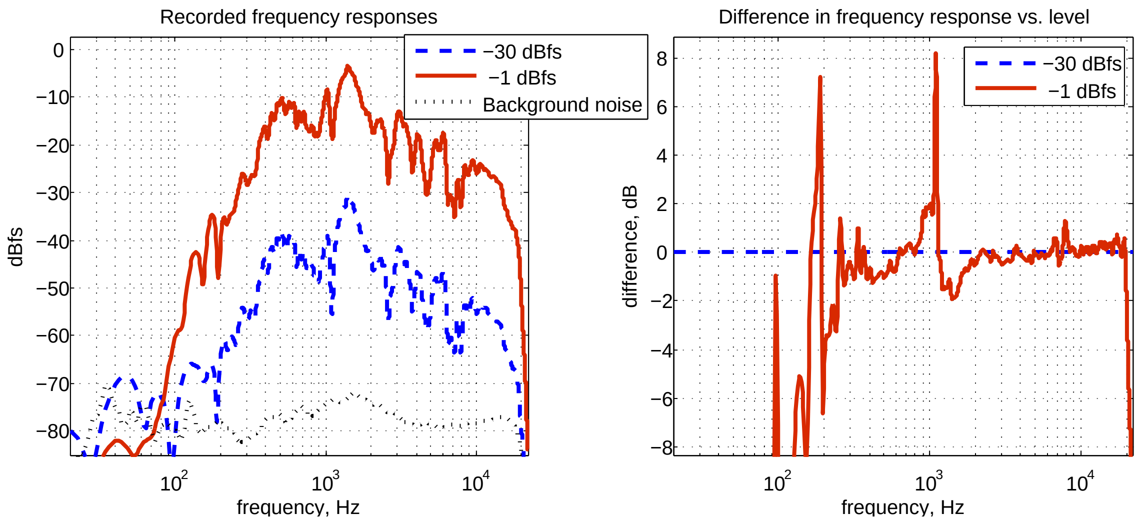


Figure 6.10: Example of a compression test done on a laptop.

6.5 The Graphical Interface

A picture of the record device setup screen and the logsweep measurement screen is shown as an example of the user interface in figure 6.11. An external library called ”Playrec” was used for all audio playback and recording in Matlab[1]. Playrec makes simultaneous playback and recording possible with high precision.

Variable fractional octave smoothing was implemented for smoothing the recorded frequency responses, making the curves easier to interpret in some cases. The smoothing operation uses a sliding average triangular window which size is recalculated for each frequency bin according to the chosen resolution in octaves.

[1]Playrec is written by Robert Humphrey, see http://www.playrec.co.uk/.

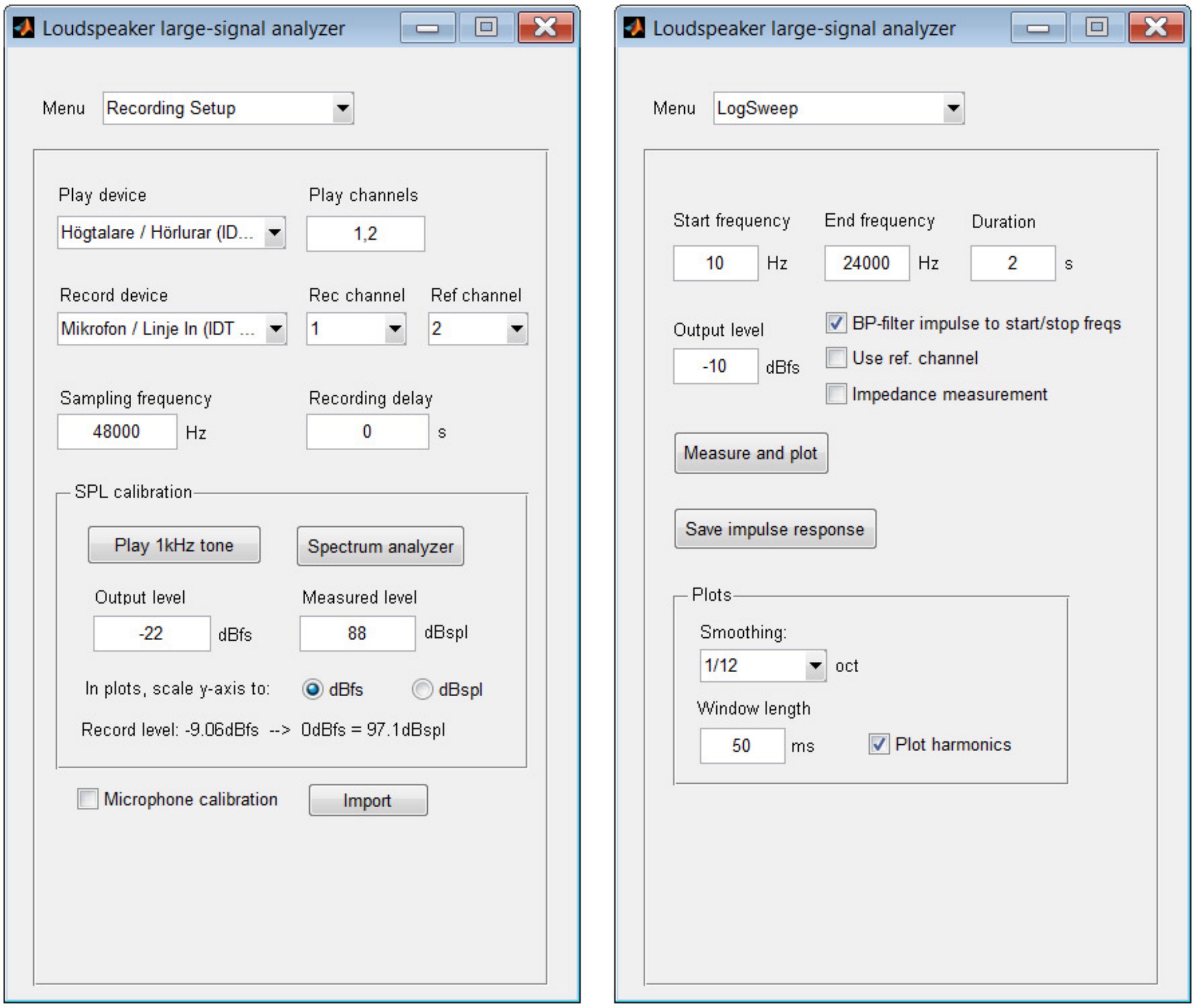


Figure 6.11: The record device setup screen and the logsweep measurement screen from the Matlab Gui for distortion measurements.

6.6 Discussion

None of the implemented metrics gives a direct answer to how the distortion of the measured loudspeaker is subjectively perceived. That would require more advanced metrics involving psychoacoustic models as discussed in chapter 4. Much information about the performance of loudspeakers can be obtained though through the implemented measurement methods. Logsweeps for testing harmonic distortion are good for finding problematic frequencies in the measured loudspeaker where it locally produces much distortion. Also, the level ratio of low order distortion (2:nd and 3:rd harmonic) to high order distortion (roughly 4:th harmonic and higher) gives information about how badly the distortion will affect the sound.

Multitone measurements are good for exciting a wide spectrum and all kinds of intermodulation distortion. This metric is sensitive to nonlinearities occuring at low levels and to high order nonlinearities, so it is objectively a good measure of distortion. It is quick and gives typically a good SNR.

Testing compression can be a good idea especially for loudspeakers that operate close to their limits during normal operation, like very small loudspeakers. Sometimes the high level frequency response can be markedly different from the low level frequency response.

7 Application of a Level Dependent Equalization Target Curve

A method was developed for extending the low frequency response of a loudspeaker using information of the current volume setting on the amplifier. At low to medium volume levels, it is often possible to extend the -3 dB cut-off frequency of the loudspeaker downwards and still be inside its linear working range.

The method is *not* a bass-boost or loudness function, since the level of the bass is not increased as the volume goes down, but rather the lowest frequencies that are not present at high volumes are added at low volumes, giving a higher fidelity experience.

7.1 Loudspeaker Model

The method for low frequency extension was theoretically investigated for the two most common loudspeaker box types, the closed box and the vented box. This section presents the models used for the two kinds of boxes. Some parameters need to be defined:

$$
\begin{aligned}
V_{box} &= \text{physical box volume } [m^3] \\
F_s &= \text{resonance frequency of loudspeaker driver in free air, [Hz]} \\
\omega_s &= 2\pi F_s \\
Q_{ts} &= \text{total Q of the driver at } F_s \text{ (inversely prop. to total damping)} \\
S_d &= \text{surface area of cone } [m^2] \\
Bl &= \text{motor strength [Newton/Ampere]} \\
R_e &= \text{DC resistance [ohm]} \\
Vas &= \text{suspension equivalent air volume } [m^3] \\
\rho_0 &= \text{air density } (\approx 1.2\, Kg/m^3) \\
c &= \text{sound velocity } (\approx 344\, m/s) \\
C_{ms} &= Vas/(\rho_0\, c^2\, S_d^2) \\
M_{ms} &= 1/(\omega_s^2 C_{ms}) \\
Q_{es} &= \omega_s M_{ms} Re/Bl^2 \\
Q_{ms} &= Q_{ts} Q_{es}/(Q_{es} - Q_{ts}) \\
R_{ms} &= \omega_s M_{ms}/Q_{ms} \\
C_{cas} &= C_{ms} S_d^2 \\
C_{cab} &= V_{box}/(\rho_0 c^2) \\
C_{atot} &= C_{cas} C_{cab}/(C_{cas} + C_{cab}) \\
R_{ae} &= 1/(\omega_s Q_{es} C_{cas}) \\
R_{am} &= 1/(\omega_s Q_{ms} C_{cas}) \\
R_{apr} &= \text{port losses (set here to 0.001)} \\
R_{abr} &= \text{box absorption losses (set here to 0.001)} \\
R_{atc} &= R_{am} + R_{abr} + R_{ae} \\
U_{ad} &= Bl\, V_{in}/(S_d R_e) \\
L_{mas} &= M_{ms}/S_d^2 \\
L_{mapr} &= 1/(w_b^2\, C_{cab}) \\
R_{alr} &= Q_{leak}\sqrt{L_{mapr}/C_{cab}} \\
V_{in} &= \text{input voltage [V]}
\end{aligned}
$$

$$
\begin{aligned}
Q_{leak} =& \text{box leakage-loss Q (set here to 10)} \\
\omega_b =& 2\pi f_{box}, \quad f_{box} \text{ is the tuning frequency of the port.} \\
T_b =& 1/\omega_b \\
T_s =& 1/\omega_s \\
\alpha =& Vas/V_{box}
\end{aligned}
$$

7.1.1 Closed Box

The transfer function from voltage U(s) at the input terminals to sound pressure P(s) for the closed box, normalized to one in the pass-band, is given by [Sma 72]:

$$
H_p(s) = \frac{P(s)}{U(s)} = \frac{s^2 T_c^2}{s^2 T_c^2 + sT_c/Q_{tc} + 1}
$$

where

$$
\begin{aligned}
Q_{tc} =& 1/(2\pi F_c\, C_{atot}\, R_{atc}) \\
F_c =& 1/\left(\sqrt{M_{ms}\, C_{atot}/S_d^2} * 2\pi\right) \\
T_c =& 1/(2\pi F_c)
\end{aligned}
$$

Evaluating $H_p(s)$ with $s = j\omega$ gives the frequency response of the loudspeaker. Another interesting transfer function is $H_x(s) = X(s)/U(s)$ where $X(s)$ is the diaphragm displacement. This is given by:

$$
H_x(s) = \frac{X(s)}{U(s)} = \frac{Bl/R_e}{s^2 M_{ms} + s(R_{ms} + Bl^2/Re) + 1/C_{ms}}.
$$

7.1.2 Vented Box

The corresponding pressure transfer function $H_p(s)$ and excursion transfer function $H_x(s)$ for a ported box is given below [Sma 73]. Port losses and box absorption losses are neglected.

$$
\begin{aligned}
H_p(s) =& \frac{s^4 T_b^2 T_s^2}{D(s)} \\
H_x(s) =& \frac{-(U_{ad} * C_{cas} * (s^2 C_{cab} L_{mapr} R_{alr} + s L_{mapr} + R_{alr}))}{E(s)}
\end{aligned}
$$

where

$$
\begin{aligned}
D(s) =& s^4 T_b^2 T_s^2 + s^3 (T_b^2 T_s / Q_{ts} + T_b T_s^2 / Q_{leak}) + s^2((\alpha + 1) T_b^2 + T_b T_s / (Q_{leak} Q_{ts}) + T_s^2) \\
& + s(T_b / Q_{leak} + T_s / Q_{ts}) + 1 \\
E(s) =& S_d (R_{alr} \\
& + s^4 C_{cab} C_{cas} L_{mas} L_{mapr} R_{alr} \\
& + s^3 (C_{cas} L_{mas} L_{mapr} + C_{cab} C_{cas} L_{mapr} R_{ae} R_{alr} + C_{cab} C_{cas} L_{mapr} R_{alr} R_{am}) \\
& + s^2 (C_{cas} L_{mapr} R_{ae} + C_{cas} L_{mas} R_{alr} + C_{cab} L_{mapr} R_{alr} + C_{cas} L_{mapr} R_{alr} + C_{cas} L_{mapr} R_{am}) \\
& + s(C_{cas} R_{ae} R_{alr} + C_{cas} R_{alr} R_{am} + L_{mapr}))
\end{aligned}
$$

7.2 Principles of the Volume Dependent EQ

Since the excursion largely decides how much distortion a loudspeaker produces (see chapter 5), it can be assumed that if the excursion is kept approximately constant (and sufficiently low) at all volume levels then distortion can be kept at a sufficiently low level. So the question then is, how low can the -3 dB frequency (here called f_b) be pushed for a certain decrease in volume? The following discussion will answer this question.

An example of the frequency dependent properties of bass reflex versus closed box loudspeakers is shown in figure 7.1. The figure shows the frequency response, normalized to one in the passband, and excursion for a 2.83 V input (standard voltage for loudspeaker sensitivity/small-signal measurements) for the same loudspeaker driver put in either a closed or a ported box. It is evident from these plots that it is possible to apply much more boost below the -3 dB frequency f_b for a closed box compared to a ported box, since the excursion increases much more rapidly below this frequency for a ported box. This needs to be taken into account in the level dependent equalizer.

Using the excellent freeware program WinISD from LinearTeam[1] simulations were made on the maximum sound pressure level of various drivers in closed and ported boxes. It can be concluded that the displacement-limited maximum SPL of a closed box decreases asymptotically by 12 dB/oct under f_b whereas for a ported box the maximum SPL decreases by around 24-30 dB/oct under the tuning frequency. This is the basis for

[1] http://www.linearteam.dk/

the operation of the volume dependent EQ. Figure 7.2 demonstrates how a new desired f_b can be calculated from a volume decrease of ΔL decibels, using an asymptotic curve representing the maximum displacement limited SPL. If the loudspeaker operates safely and distortion free at the higher level, then equalizing the response to $f_{b,new}$ at the lower level will not strain the loudspeaker further according to the maximum SPL curve.

The mathematical formula for calculating $f_{b,new}$ is:

$$\begin{aligned} f_{b,new} &= K * f_b \\ K &= 1/\exp(-\Delta L * \log_2(2)/a) \end{aligned} \tag{7.1}$$

where a is the tilt of the maximum SPL curve, about 12 dB/oct for a closed box and 24-30 dB/oct for a ported box. ΔL is a negative number – the current volume level in decibels in relation to the point where f_b should be unchanged.

7.3 Digital filters for the volume dependent EQ

In this section, the filters needed to equalize the frequency response of a closed or a vented box loudspeaker to give a new lower cutoff frequency will be given.

7.3.1 Closed Box

Starting with a closed box, the frequency response is completely described by two parameters – F_c which is the resonance frequency of the loudspeaker driver mounted in the box; and Q_{tc} which is a Thiele/Small parameter that describes the amount of damping at the resonance frequency.

There is a widely spread method called *Linkwitz Transform* [Lin 10] where the response of the loudspeaker is equalized using an active analogue filter before the power amplifier. The filter is designed using desired parameters $Q_{tc,new}$ $F_{c,new}$ that describes the desired frequency response of the total system.

The transfer function of a Linkwitz Transform filter H_{LT} for a closed box is obtained by canceling the poles in the loudspeaker transfer function by putting them in the nominator and putting new desired poles in the denominator, thus obtaining:

$$H_{LT} = (F_c/F_{c,new})^2 \frac{(s^2 T_c^2 + sT_c/Q_{tc} + 1)}{(s^2 T_{c,new}^2 + sT_{c,new}/Q_{tc,new} + 1)}$$

where

$$\begin{aligned} T_c &= 1/(2\pi F_c) \\ T_{c,new} &= 1/(2\pi F_{c,new}) \end{aligned}$$

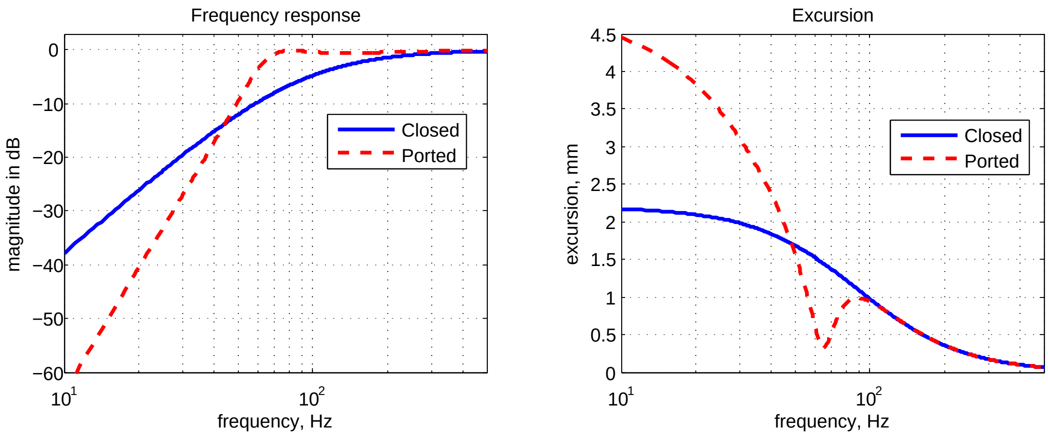


Figure 7.1: Example of normalized frequency response and excursion of a loudspeaker put in a ported versus a closed box.

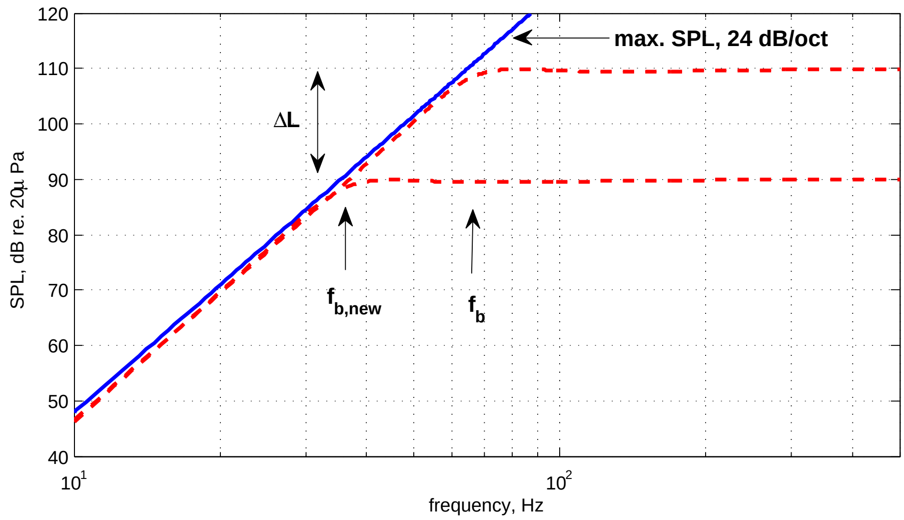


Figure 7.2: Example of calculating a new possible f_b from the asymptotic maximum SPL curve for the loudspeaker.

Now, the volume dependent EQ is implemented digitally so this transfer function needs to be discretized. That can be done using the bilinear transform. By setting $s = 2(z-1)/(T_s(z+1))$ where T_s is the sampling period, the corresponding digital transfer function H_{LTd} becomes:

$$H_{LTd} = \frac{B(z)}{A(z)}$$

where

$$\begin{aligned} B(z) =& (F_c/F_{c,new})^2(z^2(4T_c^2 + 2T_sT_c/Q_{tc} + T_s^2) \\ &+ z(2T_s^2 - 8T_c^2) + (4T_c^2 + Ts^2 - 2T_sT_c/Qtc)) \end{aligned}$$

$$\begin{aligned} A(z) =& z^2(4T_{c,new}^2 + 2T_sT_{c,new}/Q_{tc,new} + T_s^2) \\ &+ z(2T_s^2 - 8T_{c,new}^2) + (4T_{c,new}^2 + T_s^2 - 2T_sT_{c,new}/Q_{tc,new}) \end{aligned}$$

Equation (7.1) is used to calculate the desired $F_{c,new}$ for the current playback volume. $Q_{tc,new}$ does not depend on the volume, it is typically set around 0.7 so that the response around the cutoff frequency becomes as flat as possible.

7.3.2 Vented Box

A similar equalization filter as for the closed box can be developed for a vented box. $Q_{tc,new}$ and $F_{c,new}$ are not relevant parameters for a vented box so instead, desired parameters $f_{s,new}$ and $f_{p,new}$ are specified where $f_{s,new}$ represents the desired resonance frequency of the loudspeaker driver and $f_{p,new}$ the desired tuning frequency of the port. These parameters are calculated as follows:

$$f_{s,new} = K * f_s$$
$$f_{p,new} = K * f_p$$

with K from equation (7.1). The analogue transfer function for the vented box equalizing filter becomes:

$$H_{LT} = K^{-4}\frac{B(z)}{A(z)}$$

where

$$\begin{aligned}
B(z) =& s^4(T_p^2 T_{fs}^2) + s^3(T_p^2 * T_{fs}/Qts + T_p T_{fs}^2/Q_{leak}) \\
&+ s^2((\alpha+1)T_p^2 + T_p T_{fs}/(Q_{leak}Q_{ts}) + T_{fs}^2) \\
&+ s(T_p/Q_{leak} + T_{fs}/Q_{ts}) + 1
\end{aligned}$$

$$\begin{aligned}
A(z) =& s^4(T_{p,new}^2 T_{fs,new}^2) + s^3(T_{p,new}^2 T_{fs,new}/Q_{ts} + T_{p,new} T_{fs,new}^2/Q_{leak}) \\
&+ s^2((\alpha+1)T_{p,new}^2 + T_{p,new} T_{fs,new}/(Q_{leak}Q_{ts}) + T_{fs,new}^2) \\
&+ s(T_{p,new}/Q_{leak} + T_{fs,new}/Q_{ts}) + 1
\end{aligned}$$

$$\begin{aligned}
T_{fs} =& 1/(2\pi f_s) \\
T_{fs,new} =& 1/(2\pi f_{s,new}) \\
T_p =& 1/(2\pi f_p) \\
T_{p,new} =& 1/(2\pi f_{p,new})
\end{aligned}$$

Bilinear transformation of this transfer function was not attempted since the procedure gets quite involved. Instead another method was used to arrive at a discrete impulse response in the MATLAB implementation.

The frequency response of H_{LT} is first sampled at L_h points from $\omega = 0$ to $\omega = 2\pi f_s(L_h - 1)/L_h$ where f_s is the sampling frequency, by setting $s = j\omega$. The resulting frequency response, which we can call H_{LTa}, lacks the symmetry around $f_s/2$ that is needed to obtain a real impulse response through an inverse DFT operation. By using the MATLAB `ifft` command with the 'symmetric' option, the problem is solved since it only uses the first half of the input vector and considers the second half to be conjugate symmetric, thus yielding a real impulse response:
h_{LTd} `= ifft(`H_{LTa}`,'symmetric');`

The length L_h of this filter needs to be chosen sufficiently long so that the impulse response has time to decay to near zero. In the implementation, 8192 samples at a sampling frequency of 44.1 kHz seemed to be appropriate for most cases.

7.4 The Matlab Gui

Figure 7.3 shows the four screens that make up the Gui for testing the level dependent EQ. The first screen – top left – controls playback, volume and handles the loading of a music track and impulse responses for the loudspeaker system and digital room correction filter if one is available. It is also possible to set the parameters *page size* and *page buffer size* (a frame of samples is called a *page* by the Playrec library used for playback), these parameters are for the overlap-add algorithm implemented for realtime convolution of the filters with the music.

The second screen – top right – handles the model parameters for the loudspeaker model. It is also possible to calculate the maximum peak excursion that the loudspeaker exhibits with the loaded music track and the currently set loudspeaker parameters and filter settings.

The third lower left screen is used to set parameters for the level dependent EQ and subsonic filter. The subsonic filter is a high-pass filter that cuts away bass frequencies that the loudspeaker cannot produce, thus lowering the excursion. It is a Butterworth digital filter in this implementation and its cutoff frequency and order can be set.

For the level dependent EQ, the tilt of the maximum SPL curve can be set which is the parameter a in equation 7.1. This controls how much the bass is extended as the volume is turned down. It may also be desired to set limits for the bass extension, this is set by the parameters Min. and Max. by the text ”EQ cut-off multiplier”. If the loudspeaker has a natural -3 dB cutoff frequency of 200 Hz for example and you would like to limit the extension to 100 Hz then you would set the Min. EQ cut-off multiplier to 0.5. The max. setting can be set slightly below 1 if a certain bass extension is always wanted even at maximum volume.

The fourth lower right screen is the plot screen that shows the frequency responses of the loudspeaker model with and without the level dependent EQ applied. The EQ-curve is also shown. If the true impulse response of the loudspeaker is loaded then its true frequency response is plotted as well so that the model parameters can be adapted to give a good fit.

7.4.1 Case Study: a Dell 1557 Laptop

In this section some results will be presented for the internal speakers of a Dell 1557 laptop. Figure 7.4 shows the plot screen when the impulse response of the laptop loudspeakers has been loaded. The frequency response of the laptop is quite irregular with a strong resonant peak around 1.5 kHz, a falling high frequency response and lack of bass below 400 Hz.

A correction filter was created by a simple minimum phase inverse of the laptop impulse response. More advanced, higher quality software could be used for this but

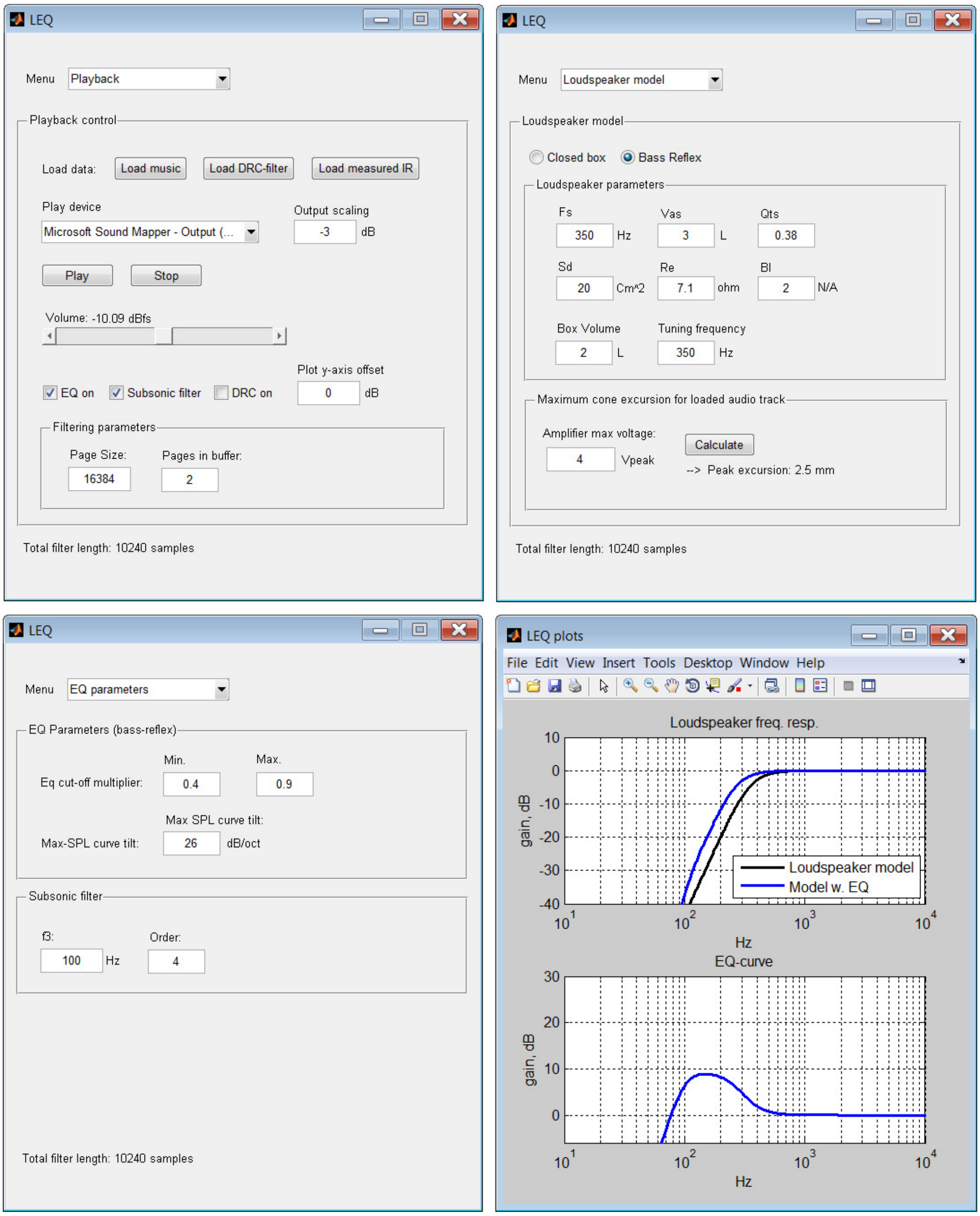


Figure 7.3: The four screens of the graphical user interface for testing the level dependent EQ.

for this illustrative purpose it is sufficient. The frequency response with the correction impulse response loaded is shown in figure 7.5. The correction filter makes the frequency response much flatter and it already applies some bass extension in this case. The model frequency response coincides pretty well with the loudspeaker plus correction filter response.

The previous plots were with the volume at max and thus without bass extension. When the volume is turned down by 15 dB from max, the bass is extended as seen in figure 7.6. A fourth order subsonic filter with cutoff at 100Hz is used as the laptop has no chance to reproduce frequencies lower than this. The result of the bass extension and correction filter is a great increase in sound quality, as judged subjectively by the 10-15 persons that the program has been demonstrated to.

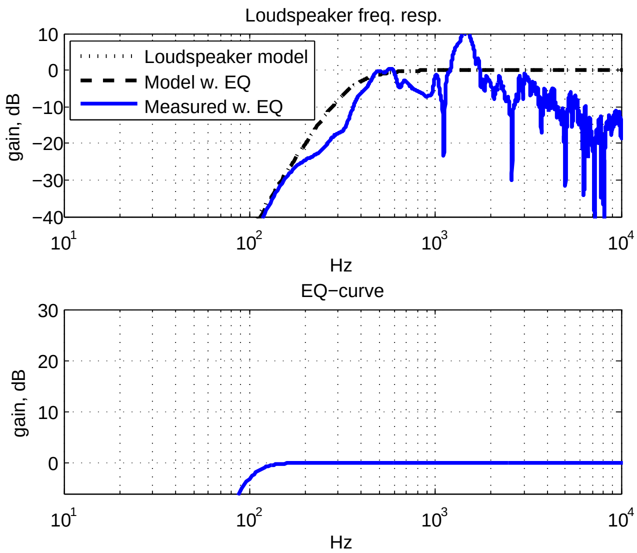


Figure 7.4: The plot screen with the laptop impulse response loaded.

7.5 Simulation of Cone Excursion

Table 7.5 below shows the calculated maximum peak excursion for two different music tracks, three different volume levels and both with and without the level dependent bass extension. The used tracks were chosen because they have a rich low frequency content and the artists are well known.

The excursion is calculated from the laptop loudspeaker model. Some parameters in the model had to be guessed so the absolute value of the excursion calculated is not completely true (especially since a linear loudspeaker model model is used and not a

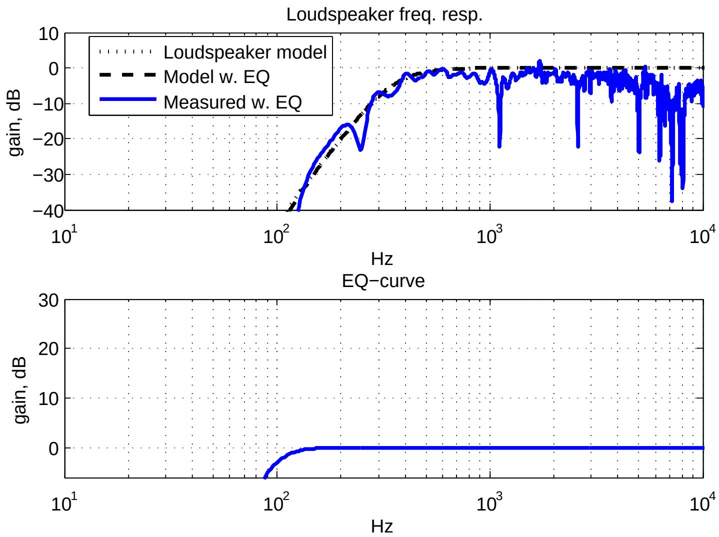


Figure 7.5: The plot screen with the correction impulse response loaded.

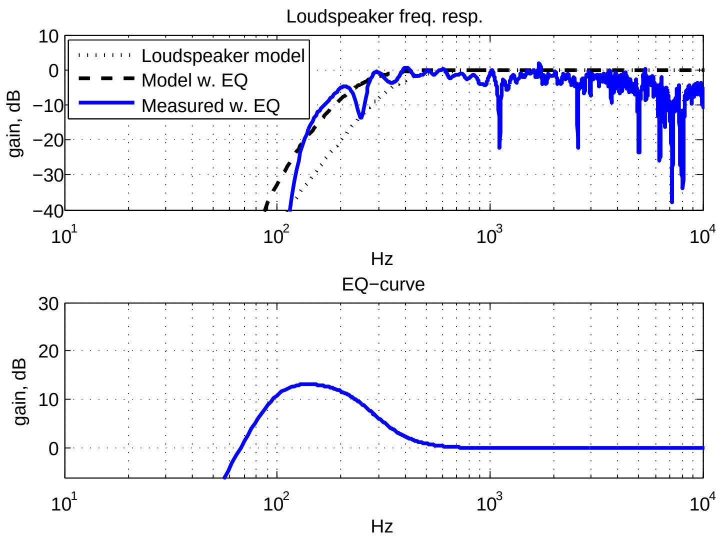


Figure 7.6: The plot screen with the volume turned down 15 dB, a fair amount of bass extension is seen.

nonlinear one), but it is the relative differences in excursion for the different cases that is of interest anyhow and these are approximately correct.

Artist - Title	EQ setting	Volume:	0 dB	-10 dB	-20 dB
ATB – The Autumn Leaves	On		3.3 mm	2.5 mm	1.9 mm
	Off		3.3 mm	1 mm	0.33 mm
Massive Attack - Safe from harm	On		2 mm	1.9 mm	1.3 mm
	Off		2 mm	0.63 mm	0.2 mm

With the EQ turned off, the excursion decreases linearly with decreasing volume as expected. With the EQ on, the excursion decreases only moderately just as desired.

7.6 Discussion and Subjective Impressions

The usual trade-off when equalizing the frequency response with a target curve that is independent of volume level is to weight the amount of low frequency extension versus the highest possible playback level where distortion is likely to occur.

With a volume dependent target curve, distortion can be avoided at all playback levels while at the same time maximizing the amount of low frequency sound energy. The resources of the loudspeaker and amplifier are utilized better at every volume level. No extra headroom is required from the amplifier, the volume dependent equalization only utilizes more of the headroom that is available at low to medium volume levels.

The level dependent EQ gave nice results when tested on a Dell 1557 laptop. The laptop originally has a -3 dB cutoff frequency in the range of 400 Hz and the extension that is possible down to around 200 Hz at medium listening levels adds a lot to the listening experience.

The method was also tested on a Samsung 46” flat panel TV. The TV had an original in room cut-off frequency around 100 Hz. Some bass extension was possible in the 75-100 Hz frequency range, though the subjective difference was not as large as for the laptop. This could depend on that this frequency range may not contain as much important information as the 200-400 Hz range which was boosted for the laptop, at least not for the music tested. Also, the ear is not as sensitive at low frequencies $<$ 200 Hz according to the ear’s equal loudness contours. Another factor may be that some bass heavy music contains very strong energy at bass frequencies compared to higher frequencies and in those cases it may not be possible to extend the bass very far down in frequency without distortion.

8 Conclusion

The audibility of distortion is dependent on the interaction between a complex nonlinear system – a loudspeaker – with a complex signal – music – and the very advanced auditory system. The theory of nonlinear systems is complex. Music is dynamic and has a complex temporal structure as well as rich frequency content. The auditory system is still not fully understood. With this in mind, it is evident that finding a universal method to evaluate the subjective performance of a loudspeaker to all kinds of signals is a complex task.

The distortion metrics implemented in the Matlab Gui still gives information about the character of nonlinearities in a loudspeaker, and knowledge and experience of what to look for in these measurements gives information about how subjectively bad the distortion is likely to be perceived. The first three chapters of the thesis deals with finding these relationships.

Some conclusions are that nonlinearities occurring at low levels are more audible than high level nonlinearities, because of the statistical distribution of the amplitude of music signals and the level dependent masking properties of the ear. High order nonlinearities are more severe than low order nonlinearities, since high order nonlinearities are less likely to be masked.

A broad-band signal is preferred for measuring distortion because a single tone cannot excite all kinds of nonlinearities, and in addition a single tone emphasizes low order and high level distortion, quite contrary to what is desired. Single tone measurements are useful though to identify objective problems in loudspeakers.

The distortion measurements included in the Matlab Gui that was developed include single tone measurement of harmonic distortion using exponential sine sweeps, a multitone distortion test and measurement of compression.

Different metrics of distortion that aim at giving a higher correlation to subjective sound quality were discussed. The most well-developed methods are made for evaluating the sound quality of perceptual audio codecs. These methods could perhaps be applied to loudspeaker evaluation with some modifications, this is an area of further research.

The level dependent equalizer implemented can give a good bass extension in some applications at low to medium listening levels, mostly for small loudspeakers, and it also protects the loudspeaker from damage at high playback levels by limiting the bass extension at high levels. It builds on a simple model of the loudspeaker's maximum capability and adjusts the bass extension to take advantage of the full loudspeaker headroom at each volume level.

8.1 Suggestions for Further Work

There are several techniques which could be combined with the volume dependent equalizer that could yield a better result. Firstly it could be made signal dependent and not only volume dependent. By monitoring the frequency contents in the input signal, the bass extension possible could be calculated both from the capabilities of the loudspeaker and from properties of the input signal.

Another possibility is to predistort the signal to lower distortion and extend the capabilities of the loudspeaker. This would probably require some kind of adaptive controller to account for the production spread in loudspeaker parameters.

A method for bass extension that has been the focus of some research is to generate synthetical harmonics to musical tones that occurs below the loudspeakers pass-band. This method relies on the theory of ”the missing fundamental” which has to do with that the auditory system perceives the fundamental tone in a harmonic series even if the fundamental is physically missing. This could possibly be combined with the volume dependent EQ, although it is more of a sound effect than high fidelity reproduction.

A drawback with these approaches is that more resources are needed in a digital signal processor implementation compared to a simple volume dependent EQ.

The future in perceptually relevant distortion measurements could be to use non-linear system identification of the loudspeaker and simulate the loudspeaker off-line with music as a testing signal and use a model of the auditory system to be able to predict accurately the audibility of distortion for a certain piece of music. Another possibility could be to adapt one of the available codec evaluation methods to measure loudspeaker distortion.

References

[Ale 01] Czerwinski E., Voishvillo A., Alexandrov S., and Terekhov A.: *Multitone Testing of Sound System Components'Some Results and Conclusions, Part 2: Modeling and Application*, J. Audio Eng. Soc., vol. 49, No. 12, pp. 1181-1192, (2001, December).

[Bla 00] Blackstock D. T.: *Fundamentals of Physical Acoustics*, Wiley-Interscience, (2000)

[Bri 02] Bright, A.: *Active Control of Loudspeakers: An Investigation of Practical Applications*, PhD thesis, Technical University of Denmark, 2002

[Cze 01] Czerwinski E., Voishvillo A., Alexandrov S., and Terekhov A.: *Multitone Testing of Sound Systems - Some Results and Conclusions, Part 1: History and Theory*, J. Audio Eng. Soc., vol. 49, No. 11, pp. 10111027, (2001, November).

[Far 09] Farina, A.: *Silence Sweep: a novel method for measuring electro-acoustical devices*, Presented at the 126th Convention of the Aud. Eng. Soc., Paper 7700 (2009, May)

[Far 00] Farina, A.: *Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique*, Presented at the 108th Convention of the Aud. Eng. Soc., Paper 5093 (2009, Feb)

[Far 07] Farina, A.: *Advancements in impulse response measurements by sine sweeps*, Presented at the 122nd Convention of the Aud. Eng. Soc., Paper 7121 (2007, May)

[Ged 03] Geddes E., Lee L.: *Auditory Perception of Nonlinear Distortion - Theory*, Presented at the 115th Convention of the Aud. Eng. Soc., Paper 5890 (2003 Dec.)

[Haw 05] Hawksford, M. O. J.: *System Measurement and Identification Using Pseudorandom Filtered Noise and Music Sequences*, J. Audio Eng. Soc., Vol. 53, No. 4, pp. 275296 (2005, April)

[How 06] Howard D. M., Angus J.: *Acoustics and Psychoacoustics*, Focal Press, (2006) [ITUR 98] International Telecommunications Union, Radiocommunications Sector (ITU-R), Recommendation BS.1387: *Method For Objective Measurements of Perceived Audio Quality*, 2001

[Kli 06] Klippel, W.: *Loudspeaker Nonlinearities - Causes, Parameters, Symptoms*, J. Audio Eng. Soc., vol. 54, No. 10, pp. 907-939 (2006, October)

[Kli 10] Klippel, W.: Specification of the KLIPPEL Analyzer System, Klippel GmbH, www.klippel.de, 2010

[Lin 10] Linkwitz, S: *Active Filters*, Internet resource, Available at http://www.linkwitzlab.com/filters.htm, (2010-05-12)

[Mor 04] Tan C., Moore B., Zacharov N., and Mattila V.: *Predicting the Perceived Quality of Nonlinearly Distorted Music and Speech Signals*, J. Audio Eng. Soc., vol. 52, No. 7/8, pp. 699711 (2004, July/August)

[Mül 04] Müller S.: *Transfer-Function Measurement with Sweeps*, J. Audio Eng. Soc., vol. 49, No. 6, pp. 443471 (2001, June)

[Ped 08] Pedersen, B. R.: *Error Correction of Loudspeakers*, PhD thesis, Aalborg University, 2008

[Sch 89] Schetzen, M.: *The Volterra and Wiener Theories of Nonlinear Systems*, Krieger Publishing Co., Malabar, Florida, 1989

[Shu 97] Schurer, M.: *Linearization of Electroacoustic Transducers*, PhD thesis, University of Twente Enschede, 1997

[Sma 72] Small, R. H.: *Closed-box loudspeaker systems, part i: Analysis*, J. Audio Eng. Soc., Vol. 20, No 10, pp. 798808 (1972, December)

[Sma 73] Small, R. H.: *Vented-Box Loudspeaker Systems–Part 1: Small-Signal Analysis*, J. Audio Eng. Soc., Vol. 21, No 5, pp. 363-372 (1973, June)

[Spa 07] Spanias A., Painter T. and Atti V.: *Audio Signal Processing and Coding*, Wiley-Interscience, 2007

[Tan 04] Tan, Moore, Zakharov, and Mattila: *Predicting the Perceived Quality of Nonlinearly Distorted Music and Speech Signals*, J. Audio Eng. Soc., Vol. 52, No. 7/8, pp. 699-711 (2004, July/August)

[Van 89] Vanderkooy, J: *A Model of Loudspeaker Driver Impedance Incorporating Eddy Currents in the Pole Structure*, J. Audio Eng. Soc., Vol. 37, No. 3, pp. 119-128 (1989, March)

[Voi 06] Voishvillo, A.: *Assessment of Nonlinearity in Transducers and Sound Systems - from THD to Perceptual Models*, Presented at the 121st Convention of the Aud. Eng. Soc., Paper 6910 (2006 Dec.)

[Voi 07] Voishvillo, A.: *Measurements and Perception of Nonlinear Distortion - Comparing Numbers and Sound Quality*, Presented at the 123rd Convention of the Aud. Eng. Soc., Paper 7174 (2007 Oct.)

[Zuc 09] Zuccatti, C and Bandiera, M.: *Dynamics Distortion: Loudspeaker Sensitivity Modulation Generated by Audio Signals*, J. Audio Eng. Soc., Vol. 57, No 5, pp. 338-348 (2009, May)

[Zwi 06] Zwicker E. and Fastl H.: *Psychoacoustics: Facts and Models*, Springer, 3rd edition, 2006